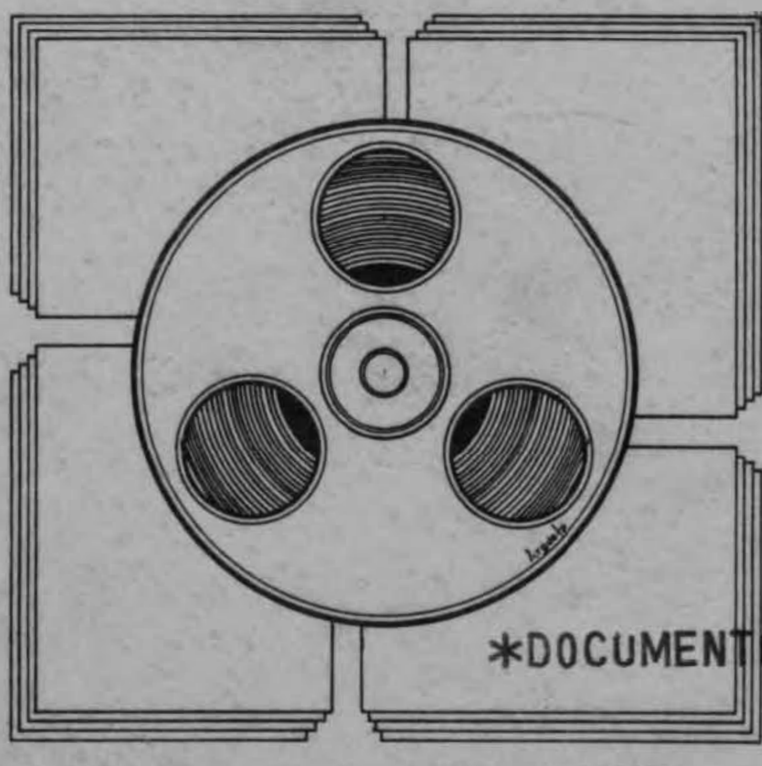
MANEJO Y ANALISIS DE DATOS DE INVESTIGACION

Curso corto nacional



DOCUMENTO de TRABAJO

Del 17 al 21 de octubre de 1983

AGRICULTURA

CENIP

W. K. Kellogg Foundation

CATIE

CENTRO AGRONOMICO TROPICAL DE INVESTIGACION Y ENSEÑANZA

CONTENIDO

NOTA: El presente material fue preparado por el Ing. MS. Rolando Piskulich, con base a revisiones técnicas de diferentes publicaciones, que tratan sobre la estadística.

Asimismo, este documento debe ser considerado como material de trabajo en el evento para el cual fue preparado.

ESTADISTICA

Definición

La estadística es una ciencia pura y aplicada que se ocupa de la colección, análisis e interpretación de datos, que sirven de base para tomar decisiones.

Conceptos básicos:

Observación

Es el registro de una característica. Ejemplo: Juan es muy inteligente, en 1980 la producción de leche en Turrialba fue de 500 T.M.

Variable

Cuando nos referimos a una característica, la misma que puede tomar diferentes valores en cada observación, decimos que es una variable. Así, en los ejemplos anteriores nos referímos primero a la variable inteligencia y después a la variable producción de leche en Turrialba.

Población

Es el conjunto de observaciones que tienen una característica común. Ejemplo: La colección de observaciones de ganado vacuno en Turrialba.

Muestra

Es el conjunto de observaciones extraídas de la población. Ejemplo: Si la población es el ganado vacuno en Turrialba, una muestra podría ser el ganado vacuno de la finca del CATIE

Muestreo Aleatorio

Cuando cada elemento de la población tiene una oportunidad igual e independiente de ser elegida para una muestra, la muestra se dice que es una muestra aleatoria

Parámetro

Son medidas constantes que caracterizan a una población. Los parámetros más usados para caracterizar una población son la media de la población, simbolizada por la letra griega μ (mu) y la varianza representada por σ^2 (sigma al cuadrado), existen otros parámetros como la mediana, la moda, coeficiente de variación, etc.

Valor estadístico

Son las medidas determinadas en una muestra que "estiman" a los parámetros y se caracterizan por cambiar de valor de una muestra a otra, lo que no ocurre con los parámetros que son constantes. Los valores estadísticos más usados son el promedio de la muestra representado por $\bar{X}$ y la varianza de la muestra representada por S^2

CLASIFICACION DE VARIABLES

Las variables se clasifican en:

Variables cualitativas

Las que representan una cualidad, por ejemplo el sexo, el color, el gusto de un animal, etc. Este tipo de variables se mide en dos escalas:

a) Escala nominal: En esta escala las observaciones se clasifican en categorías, de manera que la única relación que puede establecerse entre ellas es la de equivalencia. Cada escala se identifica con un nombre, y en los casos en que se usan números para identificarlos, estos son arbitrarios y no tienen ninguna relación de orden entre ellos. Ejemplo: El sexo: tiene dos posibles resultados: macho o hembra y podríamos decir que si es macho se clasifica como 0 y si es hembra como 1, o inversamente, si es hembra como 0 y si es macho como 1.

b) Escala ordinal: En esta escala las observaciones de tipo cualitativo se pueden ordenar de menor a mayor, sin embargo, no puede definirse una distancia entre las observaciones. Las relaciones en esta escala son $<$, $>$, $=$. Ejemplo: Presentar a una persona 3 sabores de helados y pedirle que los enuncie por orden de preferencia. En esta escala pueden asignarse números a las observaciones, pero la distancia entre dos de ellas no se conoce. Por ejemplo, en el caso de los sabores podrían asignarles 3, 2, 1 (tres al favorito, 2 al siguiente y así), pero el resultado es el mismo si se les asignan valores 8, 5, 2 o cualquier otra tercia de valores.

Nota: En el caso de la escala nominal, el único estadístico que puede calcularse es la frecuencia de cada clase. En el caso de la escala ordinal puede calcularse la mediana, puesto que pueden ordenarse las observaciones.

Variables cuantitativas

Son aquellas que pueden representarse en forma numérica, ejemplo: kg/ha de arroz, kg/leche por vaca, etc. Este tipo de variable se mide en dos escalas:

a) **Escala de intervalo:** En este tipo de escala se puede establecer el orden y la distancia entre observaciones, para lo cual se requiere de un cero y una unidad de distancia, aunque ambos serán arbitrarios. Por ejemplo, la temperatura se mide en una escala de intervalo, en este caso la unidad de medida y el punto cero son arbitrarios y por eso existe mas de una escala como la Fahrengeit y centigrados.

b) **Escala de proporción:** En esta escala se dan la igualdad, el orden y la distancia entre observaciones, y la escala un cero que no es arbitrario. Ejemplo, el peso de un individuo, la estatura en centímetros, el rendimiento de una hectárea de maíz.

Nota: En las dos últimas escalas pueden calcularse la media y la varianza. La mayoría de las técnicas paramétricas suponen una de estas dos escalas.

Fuente: García, A. y Salhuana, W. Fundamentos de Estadística y sus aplicaciones. Lima, Perú, 1970. 215 p.

CLASIFICACION Y REPRESENTACION GRAFICA DE DATOS

Introducción.

Para una correcta clasificación y representación gráfica de datos es importante determinar si la variable que está en estudio es cualitativa ó cuantitativa, ésto permitirá además una adecuada representación gráfica de los datos que tiene como objeto presentar las características de la población ó muestra en forma de dibujos o gráficos, que atraigan la atención de los lectores.

Clasificación y representación gráfica de variables cualitativas.

La clasificación consiste en determinar las cualidades que están en estudio y una vez establecidas éstas, hallar el número de observaciones que pertenecen a cada cualidad. Así por ejemplo, tenemos la distribución de la superficie cultivada publicada por el Ministerio de Agricultura del Perú en 1966, que se dá en el Cuadro 4.1.

Para las variables cualitativas no existen reglas establecidas de como deben ser estos gráficos y la única recomendación es que deben ser lo más ilustrativos. Así por ejemplo, presentamos los datos del Cuadro antes mencionado en el Dibujo 4.1.

Ilustramos la clasificación y representación gráfica de variables cualitativas con otro ejemplo. Según información publicada por el Banco Central de Reserva del Perú, correspondiente al

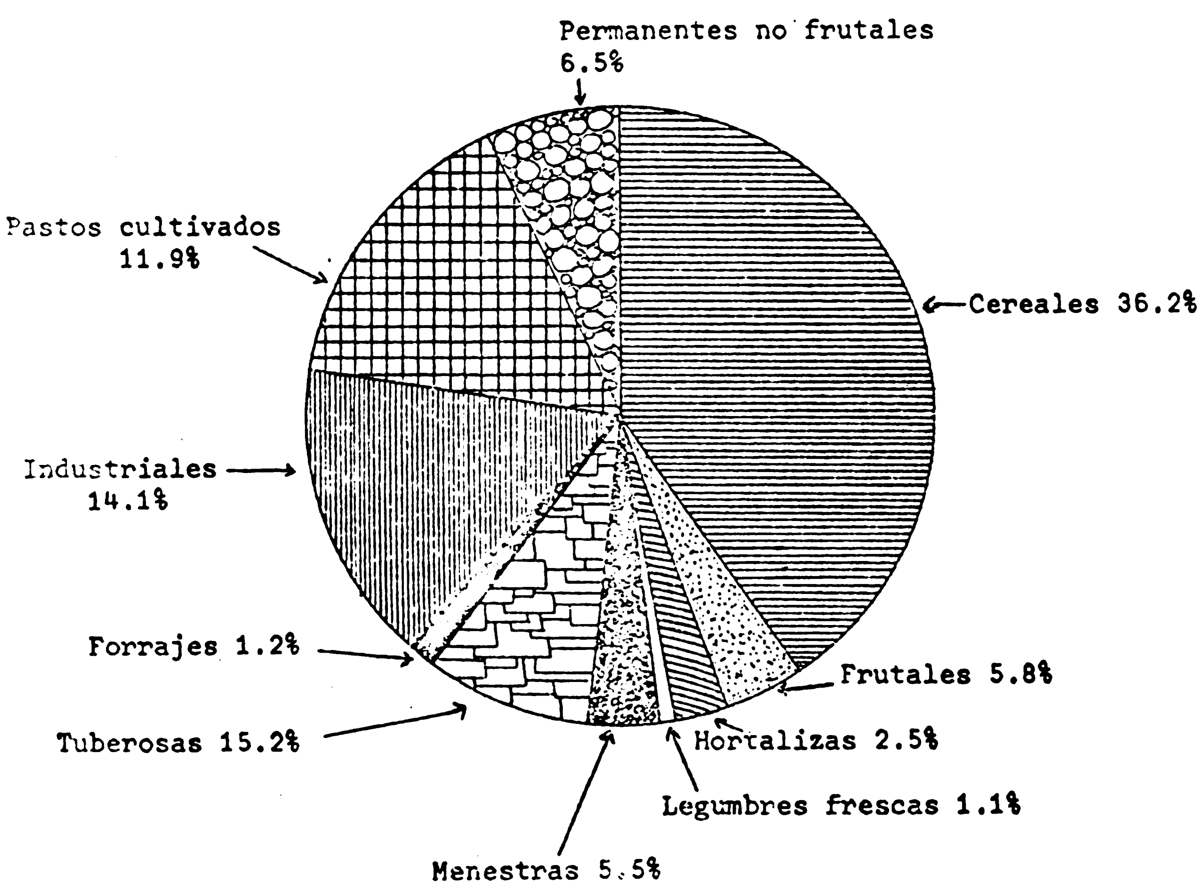
CUADRO 4.1

DISTRIBUCION DE LA SUPERFICIE CULTIVADA DEL PERU

Cultivo	Distribución de Superficie cultivada (%)
Cereales	36.2
Tuberosas	15.2
Industriales	14.1
Pastos cultivados	11.9
Permanentes no frutales	6.5
Frutales	5.8
Menestras	5.5
Hortalizas	2.5
Forrajeras	1.2
Legumbres frescas	1.1

DIBUJO 4.1

DISTRIBUCION DE LA SUPERFICIE CULTIVADA DEL PERU



CUADRO 4.2

EXPORTACION DE LOS PRINCIPALES PRODUCTOS
(ENERO - MARZO 1968)

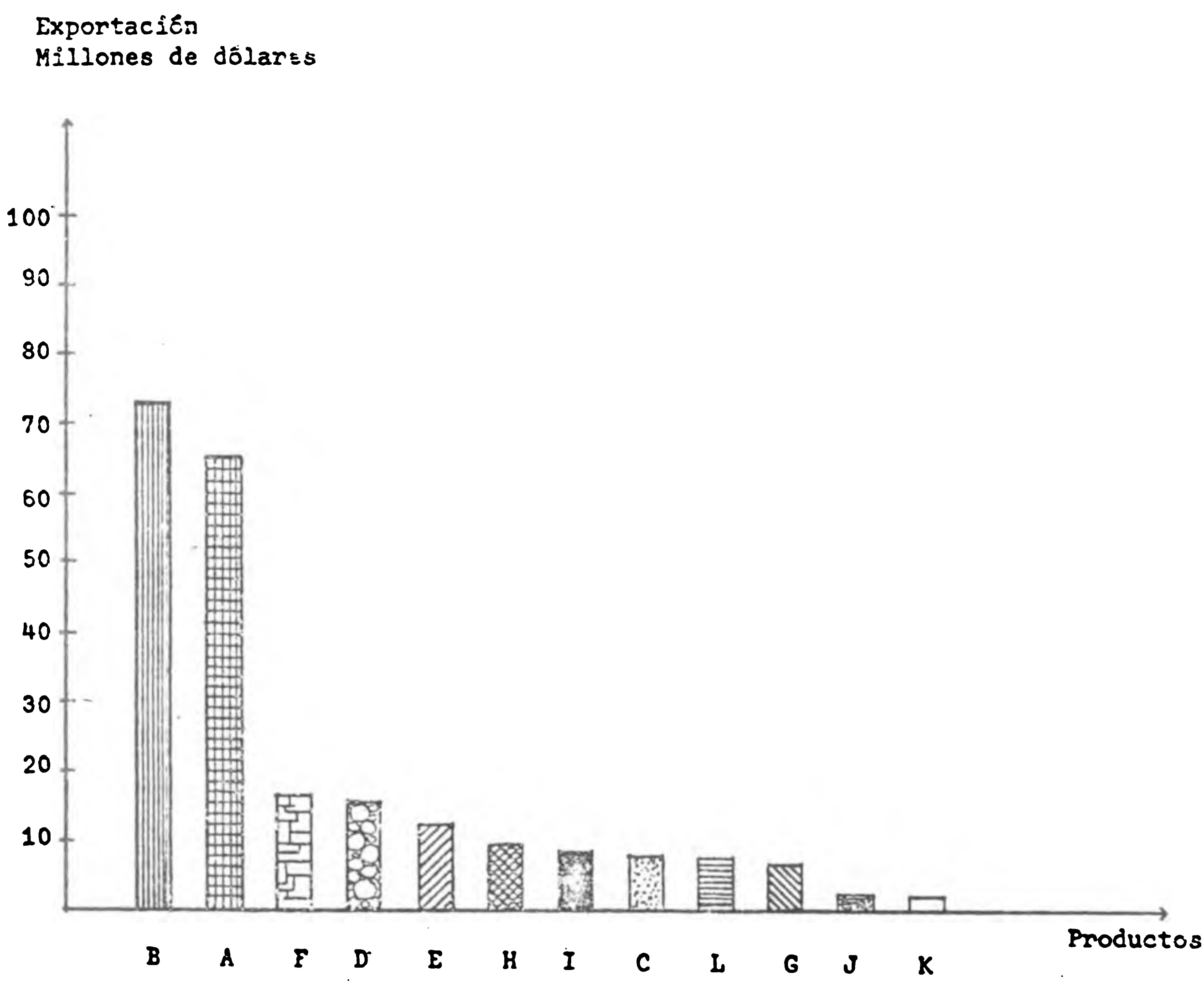
Productos	Exportación (millones de dólares)	Código
Productos pesqueros	65.1	A
Cobre	73.0	B
Algodón	8.4	C
Hierro	15.8	D
Azúcar	12.5	E
Plata	16.8	F
Plomo	6.4	G
Zinc	9.5	H
Café	8.6	I
Petróleo y derivados	2.6	J
Lanas	2.1	K
Otros	8.3	L

Fuente: Superintendencia Nacional de Aduanas.

DIBUJO 4.2

EXPORTACIONES DE LOS PRINCIPALES PRODUCTOS PERUANOS

(ENERO - MARZO 1968)



Primer Trimestre de 1968, las exportaciones de los principales productos pueden verse en el Cuadro 4.2.

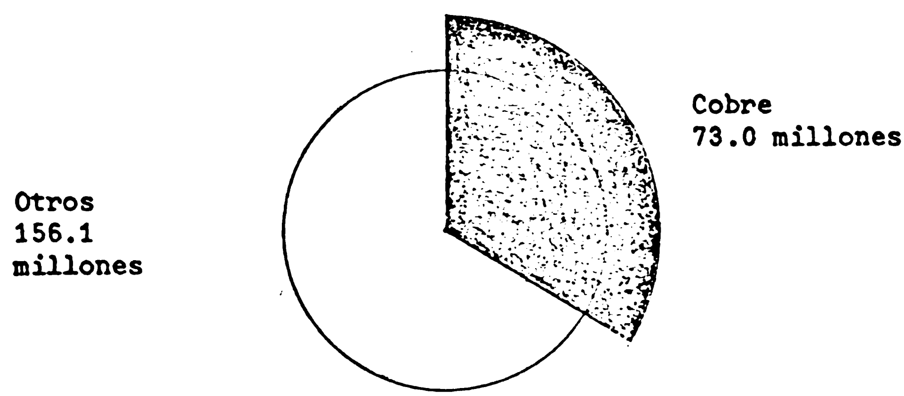
Esta información podemos representarla graficamente, usando el sistema cartesiano y representando por rectángulos las exportaciones de cada producto expresadas en millones de dólares, como puede verse en el Dibujo 4.2

La información gráfica de la exportación de cobre en relación con el total de exportaciones puede verse en el Dibujo 4.3

DIBUJO 4.3

EXPORTACION DE COBRE EN RELACION A LOS DEMAS

PRODUCTOS EN MILLONES DE DOLARES



4.3 Clasificación y representación gráfica de variables cuantitativas.

Las variables cuantitativas generalmente se agrupan en una TABLA DE FRECUENCIA, cuya representación gráfica recibe el nombre de HISTOGRAMA DE FRECUENCIA y a partir del histograma puede hallarse el POLIGONO DE FRECUENCIA.

Utilicemos los datos de las investigaciones realizadas por Bermúdez, sobre el efecto citoplasmático en diferentes características de maíz para hacer la presentación de la clasificación y representación gráfica de los datos. Bermúdez pesó 100 granos de 20 cruzas de maíz de invierno y encontró los siguientes resultados expresados en gramos:

45	45	37	34	36	41	41	44	40	40
44	47	46	37	42	42	43	41	44	41
41	43	38	37	42	43	45	39	39	38
44	42	36	42	43	40	44	40	40	38

La simple observación de los datos no nos ofrece mucha información, resultando dificil poder llegar a alguna conclusión a partir de ellos. Ante esta situación nuestro objetivo será poder determinar como se distribuyen los datos, y a partir de ellos poder estimar los parámetros.

Para la clasificación de datos, damos las siguientes recomendaciones:

a) Antes de empezar a agrupar los datos, es necesario especificar si se ha efectuado o no el redondeo de datos. El sistema de redondeo de datos recomendado puede verse en los siguientes ejemplos:

34.4 se redondea a 34

34.5 se redondea a 34

34.6 se redondea a 35

35.4 se redondea a 35

35.5 se redondea a 36

35.6 se redondea a 36

De la observación del método de redondeo que se ha seguido en los ejemplos, podemos establecer:

1) Si en nuestros datos deseamos aproximar a enteros, es necesario establecer el valor del último dígito.

2) Si el último dígito es menor de 5, el dígito anterior quedará ....

3) Si el último dígito es mayor de 5 el dígito anterior quedará ....

4) Si el último dígito es 5 debemos fijarnos en el dígito anterior, si este es par quedará......................., y si es impar quedará ....

b) Una vez concluído el redondeo de datos, se procede a establecer cuales son las observaciones máxima y mínima. En nuestros datos de 4.3, ellas son 47 y 34 respectivamente. Luego se procede a establecer la diferencia entre ellas.

$(47) - (34) = 13$

c) La diferencia recibe el nombre de AMPLITUD O RANGO (A) y es una medida de variación.

d) La amplitud (A) se divide entre un número entero positivo (k) que será escogido entre 5 y 20. La división A/k nos dará la amplitud del intervalo de clase. Así, si k = 7 tendremos para nuestros datos 13/7 = 1.8. Como nuestros datos son enteros podemos redondear, sin grave consecuencia 1.8 a 2. Una vez establecido que cada intervalo de clase tiene una amplitud de 2, el primer intervalo comprenderá los valores 34, 35; el segundo intervalo comprenderá los valores 36, 37, y así sucesivamente como puede verse en el Cuadro 4.3, en el que se establece el intervalo de clase y el número de observaciones en cada intervalo, es decir la frecuencia, recibiendo por esta razón el nombre de Tabla de Frecuencia. Una vez establecidos los intervalos de clase, el punto medio recibe el nombre de MARCA DE CLASE, que es de gran utilidad para hallar el promedio y la variancia de los datos.

CUADRO 4.3

TABLA DE FRECUENCIA

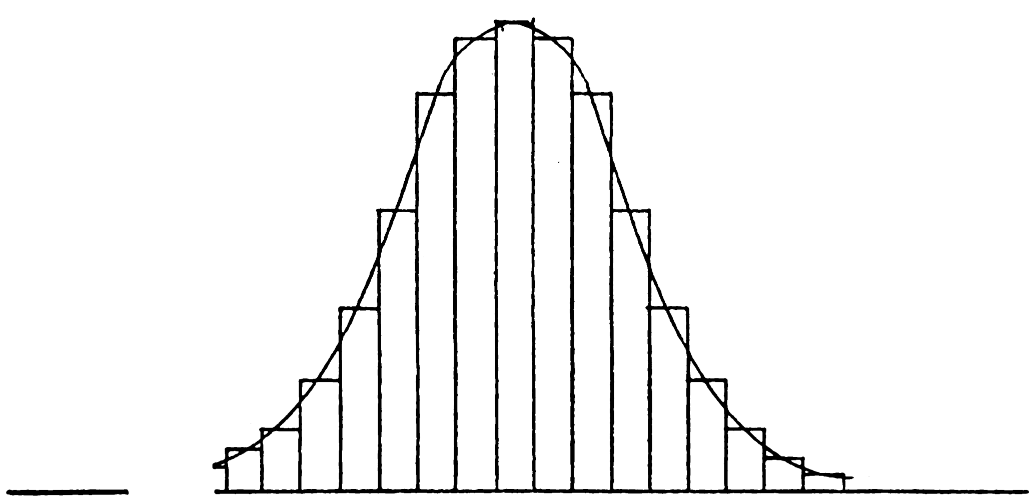
Intervalo de Clase			Marca de Clase	Frecuencia
Redondeo	Sin redondeo		(X_i')	
34-35	33.5 -	35.4	34.5	1
36-37	35.5	37.4	36.5	5
38-39	37.5	39.4	38.5	5
40-41	39.5	41.4	40.5	10
42-43	41.5	43.4	42.5	9
44-45	43.5	45.4	44.5	8
46-47	45.5	47.4	46.5	2

En el primer intervalo de clase están las observaciones cuyos valores van de 33.5 a 35.4 (según la regla de redondeo). Estos valores reciben el nombre de BORDES DE CLASE. Igual ocurre con los otros intervalos de clase.

La selección de k y consecuentemente del intervalo de clase deberá hacerse de acuerdo a las normas siguientes:

a) Establecer este valor de acuerdo a la experiencia previa de la variable en estudio.

b) Muchas de las variables siguen una distribución de frecuencia en forma acampanada con una mayor concentración de datos alrededor del promedio y una menor frecuencia para valores que se alejan de él, como puede verse en el siguiente dibujo.



Esta distribución en forma acampanada llamada normal servirá de referencia para la selección de un adecuado intervalo de clase. El dibujo que se obtenga con la Tabla de frecuencia hallada (Histograma de frecuencia) deberá mantener esta forma acampanada.

Para los datos de Bermúdez (4.3) consideremos un intervalo de frecuencia igual a 3 y la Tabla de frecuencia con dicho intervalo se presenta en el Cuadro 4.4.

CUADRO 4.4

TABLA DE FRECUENCIA

Intervalo de clase	Marca de Clase	Frecuencia
34-36	35	3
37-39	38	8
40-42	41	15
43-45	44	12
46-48	47	2

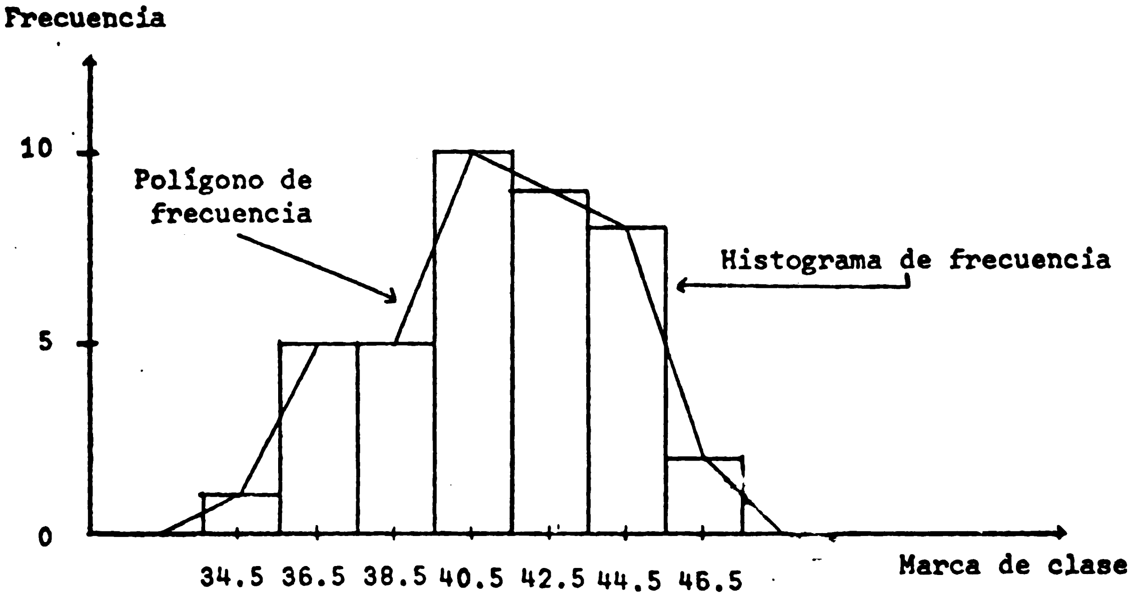
Como puede observarse al comparar los Cuadros 4.3 y 4.4 conforme aumentan las dimensiones del intervalo de clase se va perdiendo el detalle de la información individual de las observaciones, encontrándose lo contrario cuando la longitud del intervalo de clase va disminuyendo.

En general, no existe una regla fija para determinar la longitud del intervalo de clase. Existe una serie de recomendaciones que no exponemos para no agotar a nuestros lectores.

El Histograma de frecuencia se construye en un sistema de ejes cartesianos como se ilustra en el Dibujo 4.4, que es la representación gráfica de la Tabla de Frecuencia dada en el Cuadro 4.3.

DIBUJO 4.4

HISTOGRAMA Y POLIGONO DE FRECUENCIA



Si en la parte superior de cada rectángulo que corresponde al intervalo de clase ubicamos la marca de clase, y unimos los puntos con líneas habremos construído un dibujo llamado POLIGONO DE FRECUENCIA; para concluir el Polígono de frecuencia se añaden dos marcas de clase adicionales, en este caso 32.5 y 48.5 respectivamente, tal como puede verse en el Dibujo 4.4.

4.4 Tabla, Histograma y Polígono de Frecuencia Relativa.

Si en una Tabla de frecuencia, dividimos cada frecuencia entre el número total de observaciones, tendremos la Tabla de frecuencia relativa, como puede verse en el Cuadro 4.5, obtenida a partir del Cuadro 4.3.

CUADRO 4.5

TABLA DE FRECUENCIA RELATIVA

Intervalo de Clase	Frecuencia Relativa	Frecuencia Relativa (%)
34-35	0.025	2.5
36-37	0.125	12.5
38-39	0.125	12.5
40-41	0.250	25.0
42-43	0.225	22.5
44-45	0.200	20.0
46-47	0.050	5.0
	1.000	100.0

Podemos decir que la frecuencia relativa de cada clase mide en forma relativa el área del rectángulo correspondiente. Así, por ejemplo, el área relativa del primer rectángulo es 0.025, en otras palabras que el 2.5% de las observaciones están comprendidos entre 34 y 35. Igual razonamiento se aplica para las frecuencias relativas de las otras clases. De esta manera, resulta que el área total del histograma tiene un valor relativo igual a la unidad, como puede verse en el Cuadro 4.5.

Con estos datos puede elaborarse el HISTOGRAMA DE FRECUENCIA RELATIVO Y EL POLIGONO DE FRECUENCIA RELATIVO cuya forma no cambia en relación a las figuras obtenidas anteriormente, cambiando unicamente la escala del eje de ordenadas.

4.5 Tabla de Frecuencia Acumulativa, Absoluta y Relativa y su representación gráfica.

Como su nombre lo indica, es una tabla donde las frecuencias se acumulan. Tomando como base los Cuadros 4.3 y 4.5 y considerando las marcas de clase establecemos la siguiente Tabla de Frecuencia Acumulativa Absoluta y Relativa.

CUADRO 4.6

TABLA DE FRECUENCIA ACUMULATIVA ABSOLUTA Y RELATIVA

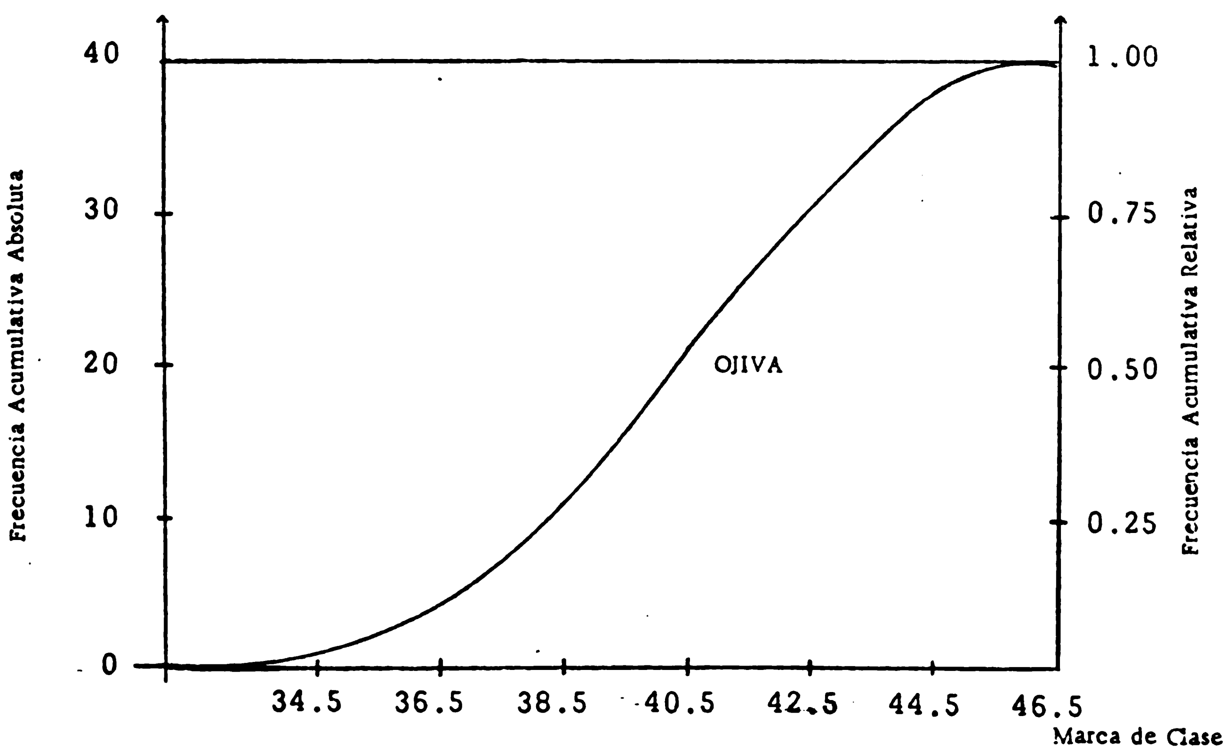
Marca de Clase	Frecuencia Acumulativa Absoluta	Frecuencia Acumulativa Relativa
34.5	1	0.025
36.5	6	0.150
38.5	11	0.275
40.5	21	0.525
42.5	30	0.750
44.5	38	0.950
46.5	40	1.000

Para la elaboración del Cuadro 4.6 se asume que todas las observaciones de un intervalo de clase tienen un valor igual a la respectiva marca de clase. La frecuencia acumulativa absoluta se obtiene sumando la frecuencia que corresponde a la marca de clase de interés más las frecuencias que corresponden a marcas de clase inferiores. Así la frecuencia acumulativa absoluta para la marca de clase 36.5 se halla sumando la frecuencia de la marca de clase 34.5 y la frecuencia de la marca de clase 36.5, es decir 1+5 = 6. Igual procedimiento se usa para hallar la frecuencia acumulativa relativa.

El gráfico de los datos obtenidos recibe el nombre de OJIVA que puede verse en el Dibujo 4.5.

DIBUJO 4.5

REPRESENTACION GRAFICA DE LA FRECUENCIA ACUMULATIVA ABSOLUTA Y RELATIVA

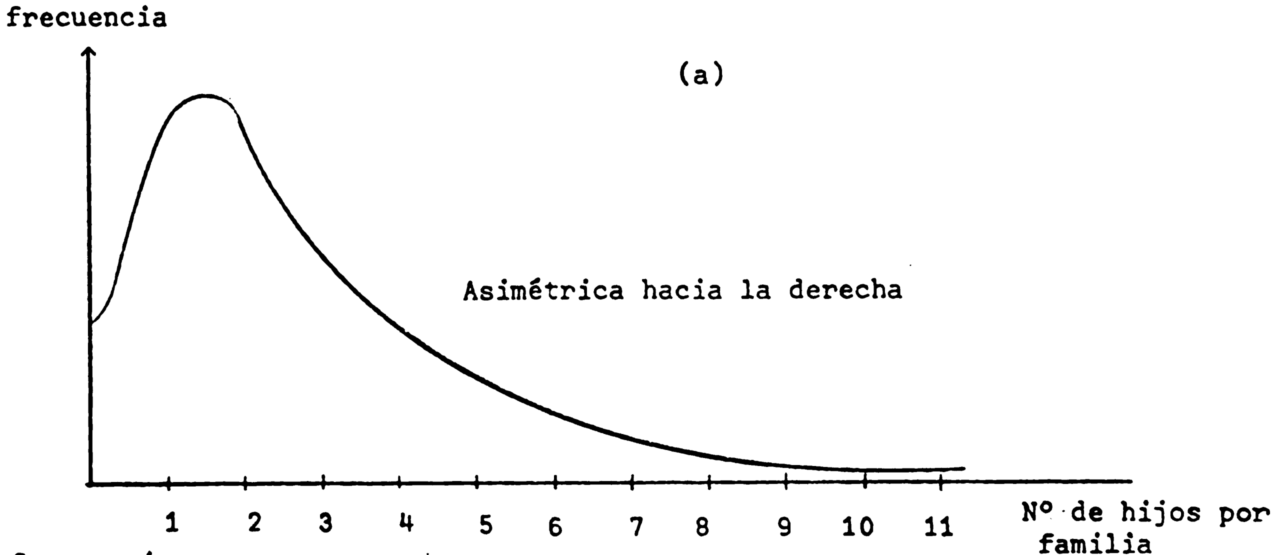


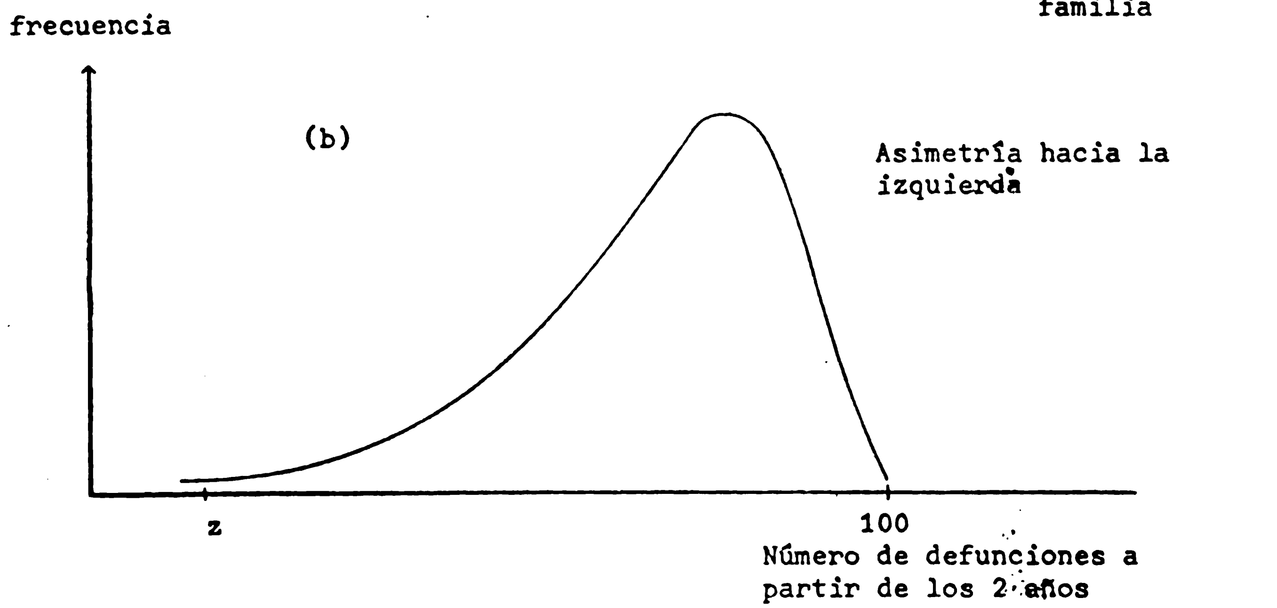
4.6 Distribuciones asimétricas.

Algunas de las variables de estudio en Biología, Industria, Comercio, etc., no tienen una distribución normal que se caracteriza por su simetría, sino más bien tienen una forma asimétrica o torcida, como puede verse en el Dibujo 4.6

DIBUJO 4.6

DISTRIBUCIONES ASIMETRICAS





EJERCICIOS

1. Los siguientes son los datos en Kilogramos de consumo de sal doméstica e industrial para el año 1966.

Departamento	Doméstica	Industrial
Amazonas	650,091	-------
Ancash	3'586,164	211,396
Apurimac	1'795,506	377,340
Arequipa	3'821,545	375,680
Ayacucho	3'456,038	4,058
Cajamarca	7'058,599	75,124
Cerro de Pasco	741,119	-------
Cuzco	5'254,913	336,071
Huancavelica	1'709,412	36,200
Huánuco	2'205,214	-------
Ica	2'461,190	268,214
Junín	4'760,378	46,859
Lambayeque	5'722,029	145,312
Loreto	2'597,537	-------
La Libertad	4'644,697	303,306
Lima-Callao	38'669,431	1'079,455
Madre de Dios	98,152	395
Moquegua	422,820	31,500
Piura	6'060,000	3'696,203
Puno	5'122,728	178,342

Tacna	312,472	-------
Tumbes	391,660	1,580

Fuente: Memorias del Estanco de la Sal

Establezca un gráfico que represente estos datos.

2. La producción de los principales países productores de mercurio para el año 1965 fué:

	Frascos (76 lbs)
España	82,760
Italia	57,291
Rusia	40,000
China	26,000
Estados Unidos	19,582
México	18,000
Yugoeslavia	16,419
Japón	4,820
Perú	3,280
Otros países	6,848

Represente estos datos en forma gráfica.

3. Con el objeto de determinar el número de horas diarias que las amas de casa dedican a ver televisión, se llevó a cabo una encuesta obteniéndose los siguientes resultados expresados en horas.

4.4	6.3	4.0	2.8	3.7
4.8	4.5	4.2	4.0	4.3
4.9	4.4	4.1	3.1	4.2
5.7	4.5	4.2	4.1	3.8
4.9	4.5	4.1	4.1	4.3
5.6	4.5	4.1	3.3	3.8
5.0	4.8	4.2	4.1	3.9
5.2	4.6	4.3	3.4	4.3
5.0	4.6	4.2	3.5	3.9
5.4	4.7	4.3	3.6	4.3
5.3	4.7	4.4	4.3	4.0
5.0	4.7	4.3	3.6	3.4
5.3	4.7	4.3	4.2	4.0

a) Construya una tabla de frecuencia con siete intervalos de clase y especifique la marca de clase y luego construya el Histograma de Frecuencia y el Polígono de frecuencia correspondiente.

b) Con los datos del problema construya una Tabla de frecuencia con 5 intervalos de clase y luego construya el Histograma y el polígono de frecuencia correspondiente.

c) Compare los Histogramas de frecuencia de las preguntas a y b ¿Cuáles son sus conclusiones?

d) Construya las Tablas de frecuencia acumulativa relativa de las dos Tablas de frecuencia y compare sus ojivas.

4. En un establo de 54 vacas se tiene la siguiente producción de litros de leche diaria:

10.7	14.6	19.6	22.2	24.8	16.7
21.2	25.1	17.7	14.7	17.5	27.2
14.1	19.9	13.8	19.4	9.7	16.3
23.2	11.2	10.5	17.4	21.7	15.5
16.4	9.8	8.6	2.3	17.9	5.4
24.3	5.9	18.8	13.0	13.7	17.1
18.5	16.5	15.3	16.5	10.3	11.4
13.3	13.6	12.2	18.4	15.8	10.2
9.2	12.7	7.5	15.2	20.4	15.4

Construya:

a) Una Tabla de Frecuencia con siete clases, mostrando además la frecuencia relativa y la frecuencia acumulativa relativa.

b) Histograma, polígono de frecuencia y ojiva.

c) Escoja otro intervalo de clase y halle la Tabla de Frecuencia, la frecuencia relativa y la frecuencia acumulativa relativa, los gráficos respectivos y compare estos resultados con (a) y (b).

5. Los siguientes datos corresponden a los jornales de 210 obreros de una hacienda:

52.50	57.40	61.50	58.40	48.50	56.60	55.90
53.40	48.30	54.50	54.20	51.40	53.10	55.60
56.40	51.50	56.30	56.60	51.70	54.00	56.50
55.40	55.60	55.10	57.00	52.20	56.60	51.40
48.60	48.70	46.00	53.20	58.60	47.00	49.50
54.20	54.60	57.20	48.50	57.10	59.80	57.10
52.30	57.20	55.80	48.20	53.60	53.10	55.10
55.90	50.60	50.60	54.90	54.30	50.50	52.60
54.60	55.60	51.60	50.70	54.60	54.60	50.00
53.50	55.20	55.60	53.00	55.10	48.20	50.30
60.50	50.70	57.20	49.20	64.00	52.60	55.80
59.10	55.10	52.60	51.50	56.30	53.60	60.30
51.50	58.60	54.20	51.50	61.20	54.00	53.60
53.20	46.20	50.30	52.60	46.80	57.60	55.10
52.60	51.10	55.00	53.50	50.50	52.60	57.10
53.10	55.10	57.30	55.80	55.50	56.50	51.20
61.40	55.80	49.10	57.20	56.60	52.60	55.50
53.80	50.50	46.00	56.20	53.20	49.50	55.20
53.10	62.00	57.20	58.20	54.60	51.50	54.00
56.20	48.00	54.30	60.20	56.70	55.60	54.50
48.20	59.10	51.50	49.00	47.00	56.70	56.00
56.30	53.50	51.60	45.70	59.20	55.10	51.10
58.00	60.00	53.50	53.60	54.00	57.80	50.50
52.20	49.20	54.60	48.20	49.00	52.60	50.30
49.10	52.50	56.20	57.70	54.90	54.30	56.20

53.70	52.60	52.00	55.10	52.30	55.80	52.00
57.20	54.00	59.50	61.10	60.00	54.30	59.60
52.70	54.80	55.60	59.00	57.10	58.60	55.10
52.80	52.60	54.80	49.20	49.10	47.20	51.60
58.00	48.20	54.20	55.60	53.10	52.60	52.50

a) Elabore una Tabla de Frecuencia, el histograma de frecuencia y el Polígono de frecuencia.

b) Construya la Tabla de Frecuencia acumulativa relativa y la ojiva correspondiente.

MEDIDAS DE TENDENCIA CENTRAL Y DE DISPERSION

7.1 Introducción a las Medidas de Tendencia Central.

Es característica común de las gentes, el tratar de resumir la información de un conjunto de observaciones en un sólo valor, así hablamos de la altura promedio del hombre peruano, de nuestros ingresos promedios mensuales, del promedio de notas de los alumnos, del promedio de la producción mensual de cobre, etc. Al condensar toda la información en un sólo valor, podemos pensar que todas las observaciones poseen ese valor.

A las medidas de tendencia central también se le conocen como medidas de posición ó localización.

Las principales medidas de tendencia central que estudiaremos son el promedio aritmético ó media, la mediana, la moda, el promedio geométrico y la media armónica.

La aplicación de cada una de estas medidas de tendencia central tienen sus ventajas y desventajas y deberá hacerse un uso adecuado de ellas.

7.1.1 Media.

La media se define como la suma de todas las observaciones dividida entre el número de ellas, así si tenemos cuatro observaciones X_1, X_2, X_3, X_4 la media será:

$$\bar{x} = \frac{X_1 + X_2 + X_3 + X_4}{4} = \frac{\sum_{i=1}^{4} X_i}{4}$$

En forma general si tenemos n observaciones x_1, x_2, x_3, ... x_n la media será:

$$\bar{x} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

La media de la muestra, $\bar{x}$, es un valor estadístico que estima a la media de la población (μ), de tamaño N

$$\mu = \frac{X_1 + X_2 + \ldots + X_N}{N} = \frac{\sum_{i=1}^{N} X_i}{N}$$

Ejemplo 1. Si 9 niños tienen las siguientes alturas 50, 62, 75, 55, 53, 62, 71, 85, 90 cm. la media de sus alturas será $\frac{603}{9} = 67$ cm.

Debe observarse que el resultado se expresa en las mismas unidades que las observaciones.

Si los datos son presentados en una Tabla de frecuencia y deseamos calcular la media, asumimos que todas las variables en un intervalo de clase tienen un valor igual a la marca de clase.

Así en una tabla de frecuencia en que tenemos k clases y las marcas de clase de cada uno de ellos es X'_1, X'_2, X'_3,..., X'_k y sus frecuencias f_1, f_2, f_3, ..., f_k respectivamente la media se halla según la siguiente ecuación.

$$\bar{x} = \frac{f_1X'_1 + f_2X'_2 + f_3X'_3 + \ldots + f_kX'_k}{f_1 + f_2 + f_3 + \ldots + f_k} = \frac{\sum_{i=1}^{k} f_iX'_i}{\sum_{i=1}^{k} f_i}$$

La suma de las frecuencias indicará el tamaño de la muestra, n, es decir $\sum_{i=1}^{k} f_i = n$ entonces tendremos:

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i X'_i}{n}$$

Ejemplo 2. Consideremos los datos del Cuadro 4.3 que se dá a continuación.

Intervalo de Clase	Marca de clase (X'_i)	Frecuencia f_i	$f_i X'_i$
34-35	34.5	1	34.5
36-37	36.5	5	182.5
38-39	38.5	5	192.5
40-41	40.5	10	405.5
42-43	42.5	9	382.5
44-45	44.5	8	356.0
46-47	46.5	2	93.0
		$\sum_{i=1}^{7} f_i = 40$	$\sum_{i=1}^{7} f_i X'_i = 1646.5$

La media de la muestra será:

$$\bar{x} = \frac{\sum_{i=1}^{7} f_i X'_i}{40} = \frac{1646.5}{40} = 41.1625$$

La media es muy facil de calcular y muy útil, su principal limitación es que los valores extremos de las observaciones la afectan notablemente. Consideremos, por ejemplo, la renta promedio por departamento de un edificio con 10 departamentos. Nueve departamentos tienen una renta de S/.5,000 cada uno, pero el décimo posee una renta de S/.50,000. Si queremos hallar la renta promedio por departamento usando la media tendremos.

$$\frac{9\ (5{,}000) + 1\ (50{,}000)}{10} = 9{,}500$$

Como hemos visto, ningún departamento tiene una renta de S/. 9,500 y en este caso la información dada por la media no es muy adecuada. Es más aconsejable calcular la media sin el valor extremo, indicando este valor aparte.

7.1.2 Mediana.

La mediana de un conjunto de observaciones, ordenados de acuerdo a su magnitud en orden ascendente ó descendente es el valor de la observación ubicada en punto medio del conjunto cuando el número de observaciones es impar y la media de las dos observaciones ubicadas en el punto medio cuando ésta es par. Esta medida divide al conjunto de observaciones en dos partes con igual número de observaciones.

Ejemplo 1. En una cuenta de bacterias hecha en el microscopio en áreas iguales, el número de bacterias fué: 10, 8, 5, 20, 25, 6, 3.

Ordenando las observaciones de menor a mayor tenemos:

3, 5, 6, 8, 10, 20, 25. El número de observaciones es impar, n = 7, la mediana será el valor de la cuarta observación, 8. De esta manera se tienen 3 observaciones mayores y 3 menores de 8.

Ejemplo 2. Si la muestra es la siguiente: 3, 5, 6, 8, 10, 11, 20, 25. El número de observaciones es par, la mediana será la media de las dos observaciones ubicadas en el punto medio, 8 y 10, ó sea 9. Cualquier valor entre 8 y 10 podría ser llamado la mediana porque divide al grupo en igual número de observaciones.

Uno de los atributos de la mediana es que no es afectada por valores extremos como lo es la media, como puede observarse en el ejemplo anterior, referente a la renta de los departamentos de un edificio, en el cual la mediana vendría a ser S/. 5,000 que representa mejor los ingresos promedios por departamento.

Siendo la media y la mediana medidas de tendencia central expresan conceptos diferentes como puede verse en el siguiente ejemplo:

En un muestreo realizado para conocer el número de horas semanales que las personas dedican a ver televisión, se obtuvo una mediana de 16 horas y una media de 14 horas. La mediana nos informa que la mitad de las personas ven televisión más de 16 horas semanales y la otra mitad menos de 16 horas semanales. La media se interpreta en el sentido de que todas las personas ven televisión durante 14 horas semanales.

El método para encontrar el valor de la mediana, cuando los datos se agrupan en una tabla de frecuencia se explicará usando los datos del Cuadro 4.3, que se repite a continuación

Intervalo de clase	Marca de clase (X_i')	Frecuencia
34-35	34.5	1
36-37	36.5	5
38-39	38.5	5
40-41	40.5	10
42-43	42.5	9
44-45	44.5	8
46-47	46.5	2
TOTAL:		40

Al número total de observaciones se le suma uno y se divide entre dos, según nuestros datos $\frac{40 + 1}{2} = 20.5$, este valor nos informa que la mediana de este conjunto de datos está comprendido en el cuarto intervalo de clase (40-41), y corresponde a la observación 20.5. El intervalo de clase en el cual está ubicada la mediana tiene un total de 10 observaciones, y podemos decir que a cada una de las observaciones le corresponde un incremento de $\frac{41 - 40}{10} = 0.1$ entonces a la observación 20.5, que es la 9.5 observación del intervalo de clase en la cual se encuentra la mediana, {20.5 - (1 + 5 + 5)} = 9.5, le corresponde un aumento de 9.5 x 0.1 = 0.95. La mediana se hallará sumando al límite inferior del intervalo de clase el incremento de la 9.5 observación en esta clase, es decir 40 + 0.95 = 40.95.

Según la definición de la mediana, la mitad de las observaciones pesan menos de 40.95 gr. y la otra mitad pesan más de 40.95 gr. Las operaciones realizadas pueden resumirse en la siguiente ecuación:

$$M \cong L_M + \left[\frac{\frac{n+1}{2} - T}{f_M}\right] \quad A$$

M: mediana

$\cong$: aproximadamente igual

L_M : límite inferior de la clase donde se encuentra la mediana

n : número de observaciones de la muestra

T : suma de las frecuencias en todas las clases anteriores a la clase en la cual se encuentra la mediana

f_M : frecuencia de la clase donde se encuentra la mediana

A : amplitud de la clase donde está la mediana

Aplicando esta fórmula a los datos, tenemos

$$M \cong 40 + \left[\frac{\frac{40 + 1}{2} - 11}{10}\right] \; 1 = 40.95$$

La mediana de las observaciones originales de Bermúdez (4.3) es 41 y la mediana encontrada en estos datos agrupados en una tabla de frecuencia es 40.95, que es un valor muy próximo a la mediana de las observaciones originales.

7.1.3 La Moda.

Esta se define como el valor que ocurre más frecuentemente en un conjunto de observaciones.

Ejemplo 1. Si se tienen los siguientes datos de edades en años: 38, 35, 40, 25, 35, 38, 37, 36, 38; la observación que es más frecuente, llamada moda, es 38.

En un conjunto de observaciones puede haber más de una moda, luego la moda no es única como lo son la media y la mediana.

Ejemplo 2. En el siguiente conjunto de datos: 34, 35, 38, 37, 35, 34, 36, 39; los valores que son más frecuentes son 34 y 35, en este caso se dice que el conjunto es bimodal.

Como puede verse en este ejemplo, la moda puede ser un valor extremo.

Ejemplo 3. En los datos de Bermúdez (4.3) existen ... modas luego el conjunto de observaciones es ....

Para hallar la moda de un conjunto de datos agrupados en una tabla de frecuencia se utiliza la siguiente ecuación

$$M_o = L_{M_o} + \left[\frac{d_1}{d_1 + d_2}\right] A$$

M_o : moda

L_{M_o} : límite inferior del intervalo de clase con mayor frecuencia

d_1 : diferencia, sin considerar signo, entre la frecuencia del intervalo de clase con mayor frecuencia y la frecuencia del intervalo de clase anterior.

d_2 : diferencia, sin considerar signo, entre la frecuencia del intervalo de clase con mayor frecuencia y la frecuencia del intervalo de clase siguiente.

A : amplitud del intervalo de clase con mayor frecuencia.

Aplicando esta fórmula a los datos de Bermúdez agrupados en el Cuadro 4.3 tenemos:

$$M_o \doteq 40 + \left[\frac{5}{5+1}\right] 1$$

$$M_o \doteq 40.83$$

Utilizando este método para hallar la moda, en una colección de datos agrupados en una tabla de frecuencia nos dará una sola, mientras que en las observaciones originales sin agrupar puede haber más de una moda, ésto se debe a que al agrupar las observaciones en una tabla de frecuencia, se pierde información sobre el valor particular de las observaciones.

7.1.4 Media Geométrica.

La media geométrica, $\bar{x}_G$, de un grupo de n observaciones positivas X_1, X_2, ..., X_n, es definida como la raíz enésima del producto de todos ellos.

$$\bar{x}_G = \sqrt[n]{X_1, X_2, X_3 \ldots X_n} = \text{antilog}\left[\frac{1}{n}\sum_{i=1}^{n} \log X_i\right]$$

La media geométrica es usada principalmente en problemas referentes a promedios de proporciones. Nunca debe ser usada si uno de los números que van a intervenir en su cálculo es cero ó negativo porque obtendríamos cero ó un número imaginario.

Ejemplo 1. Una ciudad X tuvo en 1930 una población de 2000 habitantes, en 1940 el número de habitantes fué 8000 y en 1950 fué de 128,000. Deseamos conocer cuál ha sido el promedio de cambio en la población. Vemos que de 1930 a 1940 el número de habitantes de la ciudad se cuadruplicó y que de 1940 a 1950 se multiplicó por 16.

¿Cuál ha sido el promedio de estos incrementos de la población? Si hallamos la media de los incrementos obtendremos $\frac{4 + 16}{2} = 10$; este valor nos indicaría que cada 10 años la población se incrementa 10 veces. De acuerdo a estos resultados, la población en 1940 sería 20,000 habitantes (2,000 x 10) y en 1950 sería 200,000 habitantes (20,000 x 10). Los valores 20,000 y 200,000 difieren marcadamente de la realidad.

Ahora bien, si calculamos la media geométrica $\bar{x}_G$ de 4 y 16

$$\bar{x}_G = \sqrt{4 \times 16} = 8$$

y si aplicamos este promedio de incremento, veremos que para 1940 nos dá 2,000 x 8 = 16,000 habitantes muy superior a 8,000 pero inferior al valor hallado mediante la media, y para 1950: 16,000 x 8 = 128,000 valor que más se ajusta a la realidad.

7.2 Medidas de dispersión.

Como hemos visto las medidas de tendencia central nos permiten hallar un valor que caracteriza a la población. Es razonable pensar que las observaciones del conjunto tienen valóres diferentes, y que es necesario informarse de la dispersión ó variabilidad que existe entre las observaciones del conjunto.

Observemos los siguientes conjuntos de datos.

A 11, 12, 12, 15, 15, 15, 18, 18, 19, 20

B 11, 12, 13, 14, 15, 16, 17, 18, 19, 20

Ambos conjuntos tienen la misma media 15.5, pero puede observarse que en el conjunto B existe mayor variabilidad entre las observaciones que en el conjunto A.

Las medidas de variabilidad que estudiaremos son: la Amplitud, la Variancia y el Coeficiente de Variación.

7.2.1 Amplitud.

Es la diferencia entre la observación que tiene valor más alto y la observación de valor más bajo, es fácil de calcular y para muestras mayores de 2 observaciones es ineficiente, porque no toma en cuenta los valores intermedios de la muestra y sólo es influenciada por los valores extremos.

En los conjuntos A y B dados en 7.2, ambos tienen igual amplitud, 9, sin embargo puede notarse que en B, existe mayor variabilidad entre las observaciones que en A, se ratifica entonces la ineficiencia de la amplitud para medir la variabilidad.

7.2.2 Variancia.

Es una medida de dispersión ó variabilidad, que a diferencia de la amplitud considera todas las observaciones. Como hemos visto, la variancia está definida por la siguiente ecuación:

$$\sigma^2 = E\left[X_k - \mu\right]^2 = \sum_{i=1}^{N} (X_k - \mu)^2 \; p_k$$

Si todas las observaciones de la población tienen igual probabilidad $(\frac{1}{N})$, la variancia se hallará según la siguiente ecuación:

$$\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$$

En la práctica se utiliza la siguiente ecuación para el cálculo de la variancia.

$$\sigma^2 = \frac{\sum_{i=1}^{N} X_i^2 - \frac{(\sum_{i=1}^{N} X_i)^2}{N}}{N}$$

Note que esta ecuación es el desarrollo algebraico de la anterior, donde el numerador recibe el nombre de SUMA DE CUADRADOS DE LA POBLACION. Al término, $\sum_{i=1}^{N} X_i^2$, se le conoce como SUMA DE CUADRADOS SIN CORREGIR y al término, $\frac{(\sum_{i=1}^{N} X_i)^2}{N}$ se le conoce como TERMINO DE CORRECCION.

Ejemplo 1. Sea 1, 2, 3, 4, 5, 6 las observaciones de una población. La media, variancia y desviación standard de la población son:

a) $E(X) = \mu = \frac{21}{6} = 3.50$

b) $$\sigma^2 = \frac{91 - \frac{(21)^2}{6}}{6} = 2.917$$

c) $\sigma = 1.707$

Cuando las observaciones de la población se agrupan en una tabla de frecuencia, la variancia puede hallarse mediante la siguiente ecuación

$$\sigma^2 = \frac{\sum_{i=1}^{k} f_i X_i'^2 - \frac{\left(\sum_{i=1}^{k} f_i X_i'\right)^2}{N}}{N}$$

donde k es el número de intervalos de clase y X_i' es la marca de clase.

En los casos en los que se desconoce la población, será necesario estimar los parámetros μ y σ^2mediante una muestra.

La variancia de la muestra s^2, se hallará mediante la siguiente ecuación

$$s^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{x})^2}{n - 1}$$

$$s^2 = \frac{\sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n}}{n - 1}$$

Al numerador de estas ecuaciones se le conoce como SUMA DE CUADRADOS DE LA MUESTRA y al denominador como GRADOS DE LIBERTAD.

Ejemplo 2. Una muestra de 3 observaciones extraída al azar de la población enunciada en el Ejemplo 1, dió los siguientes resultados: 2, 3, 6

La media, variancia y desviación standard de la muestra son:

$\bar{x} = 3.667$

$$s^2 = \frac{49 - 40.333}{2} = 4.334$$

$s = 2.081$

7.2.3 <u>Coeficiente de Variación.</u>

Es una medida de dispersión ó variabilidad que se determina según la siguiente ecuación.

$$C.V. = \frac{s}{\bar{x}} \times 100$$

El coeficiente de variabilidad mide el número de veces que la desviación standard contiene al promedio de la muestra expresado en porcentaje. Como toda razón está expresada en unidades abstractas. Esta medida de variación es utilizada cuando se comparan experimentos expresados en las mismas ó diferentes medidas.

Ejemplo 1. La muestra 2, 3, 6 tiene como promedio

$\bar{x} = 3.667$ y $s = 2.081$ entonces

$$C.V. = \frac{2.081}{3.667} \times 100$$

$$C.V. = 56.75\%$$

EJERCICIOS

1. Consideremos la siguiente población de resistencia de alambres medida en ohmios:

0.139, 0.136, 0.141, 0.145, 0.137, 0.133, 0.141, 0.143, 0.134

Halle y diga en qué unidades está expresada cada una de las siguientes medidas

a) La media

b) La mediana

c) La moda

d) La amplitud

e) La variancia y desviación standard

f) El coeficiente de variación

2. Para la siguiente población de alturas de plantas de maíz, expresadas en metros

1.40, 1.36, 1.75, 1.82, 2.10, 2.11, 1.74, 1.82, 2.45, 1.82

Halle:

a) La media

b) La mediana

c) La moda

d) La amplitud

e) La variancia y desviación standard

f) El coeficiente de variación

3. Las siguientes observaciones de una muestra corresponden a la producción de gasolina expresada en porcentaje de aceite crudo

13.2, 11.0, 15.2, 26.8, 14.0, 15.7, 15.2, 7.4

Halle:

a) La media

b) La mediana

c) La moda

d) La amplitud

e) La variancia y desviación standard

f) El coeficiente de variación

4. Se quiere medir la duración de una nueva mezcla de asfalto y para el objeto se pavimentó una avenida. Al cabo de cierto tiempo se midió el área de los daños causados en el asfalto por el tráfico, éstos fueron en centímetros cuadrados:

0.0600, 0.4000, 1.6200, 0.4225, 0.7396, 15.2100, 24.0100, 0.4225

Halle:

a) La media

b) La mediana

c) La moda

d) La amplitud

e) La variancia y desviación standard

f) El coeficiente de variación

5. En las siguientes muestras halle la media, la mediana, la moda, la amplitud, la variancia y la desviación standard y el coeficiente de variación.

a) -3, -1, 0, 1, 3

b) - 8, -5, -2, -1

c) 8, 5, 2, 1

d) -10, 8, -5, 4, -6

e) 15, -15, 10, -10

6. Los siguientes datos corresponden a la exportación de cobre en toneladas métricas del Perú, según la Superintendencia General de Aduanas.

Año	Exportación Total T.M. (contenido fino)
1950	27,211
1955	41,292
1960	167,992
1965	179,847

¿Cuál ha sido el incremento promedio de las exportaciones?

7. Un agente viajero va a Chiclayo que está a 750 Km., a 20 Km. por hora y regresa a 100 Km. por hora. ¿Cuál ha sido el tiempo empleado en el recorrido?

8. Los siguientes fueron los resultados en libras de una prueba para medir la "ternura" de carne en vacunos:
(mediante el tenderómetro Warner - Bratzler)
14.25, 12.00, 11.50, 7.00, 6.00, 17.25, 8.75, 11.75, 6.25, 12.75, 9.50, 7.50, 11.50, 5.50, 11.00, 11.50, 12.00
Calcule:
a) La media, moda y mediana
b) La variancia, la desviación standard y el coeficiente de variación

9. En una fábrica se instala una nueva máquina y para estimar el tiempo que demora la máquina en hacer una operación, se toma al azar una muestra, expresándose el tiempo en segundos:

18.2	14.2	13.2	16.0	15.8
15.6	15.3	18.0	15.3	16.5

Halle:
a) La media, mediana y moda
b) La amplitud, la variancia, la desviación standard y el coeficiente de variación.
c) Compare este coeficiente de variación con el del ejercicio anterior y establezca sus conclusiones.

10. En el ejercicio N°5 del Capítulo IV, considere a los obreros de la hacienda como la población:

a) Calcule la media y variancia de la población

b) Usando la Tabla de números al azar obtenga 10 muestras de 5 observaciones cada una y calcule su media y variancia en cada una de ellas.

c) ¿Los valores estadísticos calculados en las muestras son iguales? Explique que ha ocurrido.

11. El consumo total de cobre en el Perú, según el Instituto Nacional de Investigación y Fomento Mineros es:

Año	1950	1955	1960	1965
Total en T.M.	160	595	1,237	2,164

Utilizando logaritmos halle la media geométrica y diga qué significa.

12. El consumo de algodón en término de hilados en el Perú expresados en toneladas métricas han sido:

1952	10,303
1953	11,628
1954	13,527
1955	12,623
1956	12,611
1957	13,372
1958	12,153
1959	14,214
1960	15,240
1961	14,762

Fuente: Cámara Algodonera del Perú y Estadísticas del Comercio Exterior.

Halle la media, la variancia, la desviación standard y el coeficiente de variación de esta muestra.

13. Para los ejercicios 3, 4, y 5 del Capítulo IV, halle

a) La media

b) La mediana

c) La moda

d) La amplitud

e) La variancia y la desviación standard

f) El coeficiente de variación

14. Una compañía desea determinar el rendimiento de sus obreros y para el efecto observa la producción diaria de sus obreros, los resultados obtenidos son los siguientes: 38.5, 36.0, 42.0, 39.0, 36.1, 37.6

a) Halle la media de la muestra

b) Halle la variancia, la desviación standard y el coeficiente de variación de la muestra.

15. Un laboratorio desea probar un nuevo producto para disminuir la presión sanguínea y según las especificaciones debe ser suministrado cuando la presión llega a 20. Para el efecto se decide realizar el estudio y se suministra el producto a 50 pacientes de un

hospital (muestra), encontrándose los siguientes datos.

Tabla de frecuencia

Intervalo de frecuencia	Frecuencia
8- 9	5
10-11	8
12-13	15
14-15	12
16-17	10

a) Halle la media, mediana y moda

b) Halle la variancia, la desviación standard y el coeficiente de variación.

16. Una fábrica recién instalada necesita 150 obreros para iniciar la producción y para el efecto contrata a una empresa para que le seleccione el personal. La empresa toma una prueba de aptitud a 400 postulantes y presenta los resultados de los 150 postulantes que ocuparon los primeros lugares en la siguiente tabla.

Tabla de frecuencia

Notas	Frecuencia
50-54	6
55-59	10
60-64	16
65-59	22
70-74	36

75-79	25
80-84	17
85-89	12
90-94	4
95-100	2

a) Halle la media, mediana, moda

b) Halle la variancia, la desviación standard y el coeficiente de variación.

17. Los siguientes datos corresponden a los coeficientes de congelación de 100 carcasas de vacunos agrupados en la siguiente Tabla de frecuencia.

Tabla de Frecuencia

Intervalo de frecuencia	Frecuencia
326-350	8
351-375	12
376-400	20
401-425	30
426-450	25
451-475	10
476-500	5

a) Halle la media, mediana, moda

b) Halle la variancia, la desviación standard y el coeficiente de variación.

UTILIZACION DE SAS PARA DISTRIBUCION DE FRECUENCIAS Y GRAFICA DE HISTOGRAMA

El SAS, es un sistema de cómputo para análisis de datos. Provee las herramientas necesarias para un completo análisis de datos. Así, SAS comprende programas para:

a) Almacenamiento y recuperación de información

b) Modificación o transformación de datos

c) Reportes escritos

d) Análisis estadístico

e) Manipulación de archivos

PROC FREQ.

Dentro de los programas o "PROC" que ofrece SAS para análisis estadístico, se encuentra el PROC FREQ que sirve para crear tablas de distribución de frecuencias, tanto en un sentido de clasificación como en n-sentidos de clasificación. En cada caso, nos calculará las frecuencias absolutas y relativas, así como las frecuencias acumuladas absoluta y relativa de valores dentro de cada clase. Ver Pag. 1 del listado.

PROC CHART.

Adicional al procedimiento PROC FREQ, el SAS brinda el PROC CHART. Con este procedimiento podremos hacer gráficas de histogramas y polígonos de frecuencia en 2 y 3 dimensiones. Los histogramas pueden ser para la frecuencia absoluta acumulativa, relativa, relativa acumulativa, las que aparecen en las páginas 2,3,4,5 y 6 del listado de computador.

El PROC. CHART ofrece además, la opción de hacer gráficas de pastel; como se observa en la pag. 7 del listado.

Para mayor referencia, ver el manual de operación SAS, SAS USER'S GUIDE.

(SAS USER'S GUIDE: 1982 Edition. Cary N.C., SAS Institute, 1982. 494 p.)

Inferencia estadística

DISTRIBUCIONES CONTINUAS

Distribución normal.

Una de las más importantes distribuciones contínuas de probabilidad es la DISTRIBUCION NORMAL y muchas variables en estudio de Economía, Administración, Industria, Ingeniería, Educación, Medicina, Bilogía, Agricultura, etc. siguen esta distribución.

En el Dibujo 9.1 puede verse la forma acampanada de la distribución normal, la curva es simétrica con respecto a la perpendicular levantada en la media, μ. En esta curva el eje de las abscisas representa los valores de la variable y la altura representa la probabilidad y como puede verse la probabilidad es mayor para aquellas observaciones próximas a la media μ.

Existen muchas curvas normales, pudiendo identificarlas por su media, μ; y desviación standard, σ. En el Dibujo 9.2 se muestra tres curvas con diferentes medias $\mu_1 < \mu_2 < \mu_3$, pero con la misma desviación standard, σ. El Dibujo 9.3 presenta tres curvas normales con igual media, μ, pero desigual desviación standard: $\sigma_1 < \sigma_2 < \sigma_3$

La ecuación de la curva normal es:

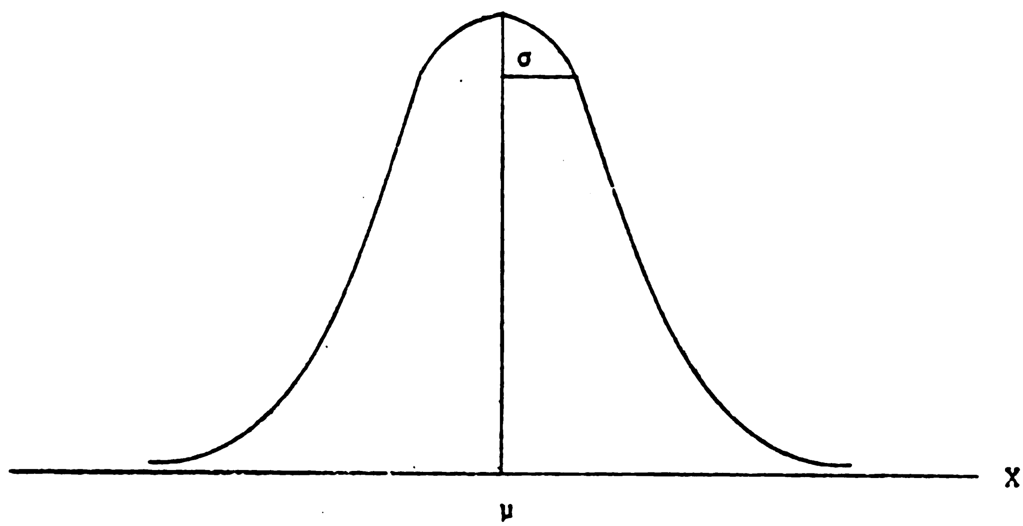
$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} \; e^{-\frac{(X-\mu)^2}{2\sigma^2}} \qquad -\infty < X < \infty$$

Donde:

X : es la variable contínua en estudio

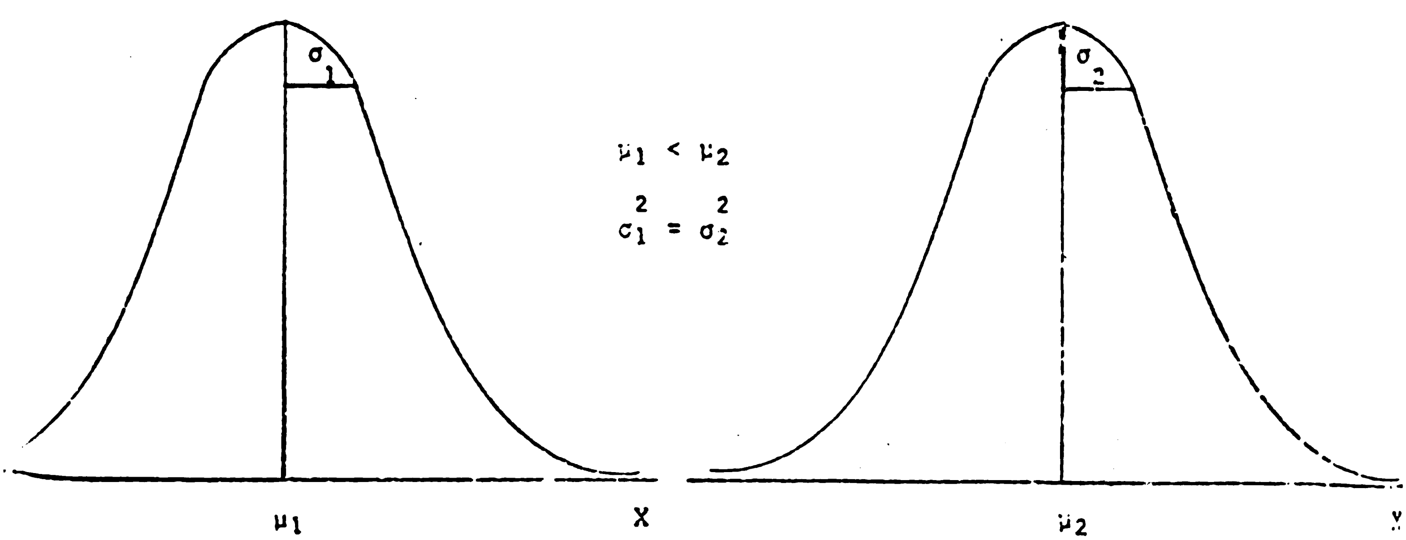
DIBUJO 9.1

CURVA NORMAL



DIBUJO 9.2

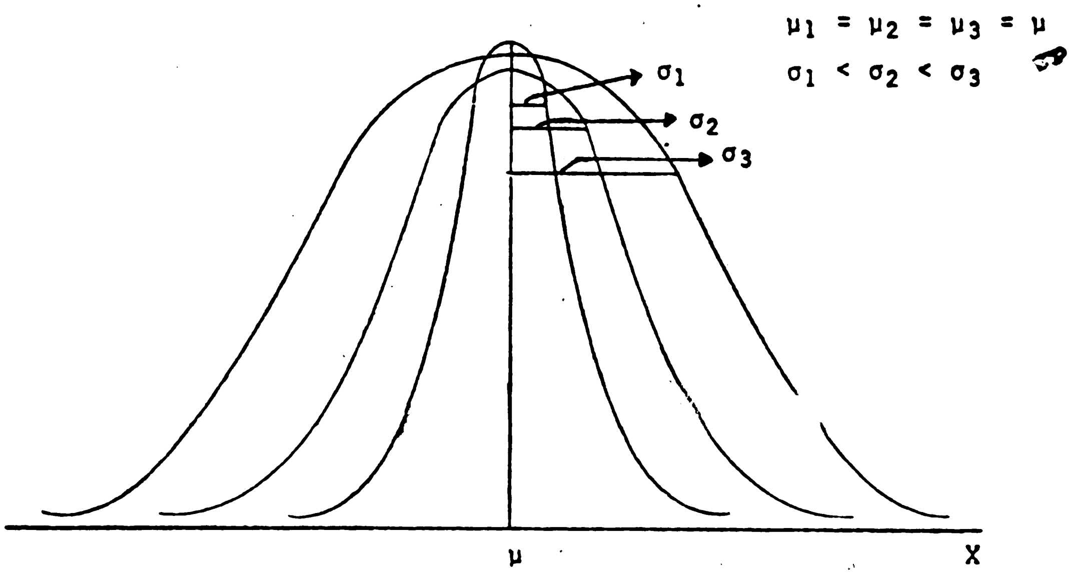
DISTRIBUCIONES NORMALES CON MEDIAS DIFERENTES Y VARIANCIAS IGUALES



DIBUJO 9.3

DISTRIBUCION NORMAL CON DISTINTAS DESVIACIONES STANDARD

CON LA MISMA MEDIA



μ : es la media

σ^2 : es la variancia

σ : es la desviación standard

e : es la constante 2.7183

π : 3.1416

El área debajo de la curva normal es igual a la unidad ó 100%, y el 50% del área corresponde a observaciones con valores inferiores a la media y el otro 50% corresponde a observaciones con valores mayores que la media. Esta curva normal es asintótica con respecto al eje de las abscisas.

La curva normal tiene la propiedad de que el intervalo $(\mu - \sigma)$ a $(\mu + \sigma)$ encierra aproximadamente el 68% de las observaciones, el intervalo (μ - 2σ) a (μ + 2σ) encierra aproximadamente el 95% de las observaciones, y el intervalo (μ - 3σ) a (μ + 3σ) encierra aproximadamente el 99% de las observaciones.

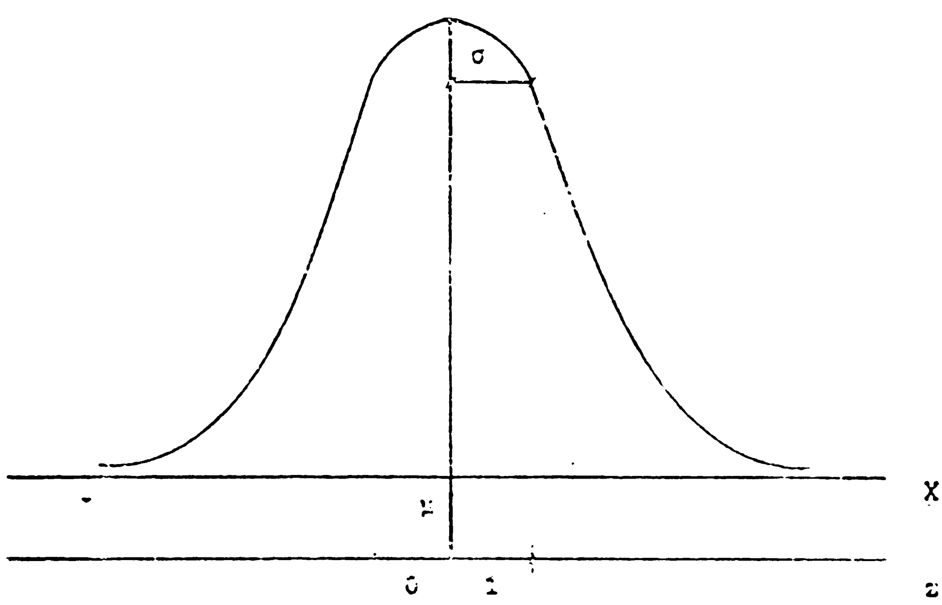
Es importante notar que σ es la distancia que existe de la ordenada levantada en la media, μ, al punto de inflexión de la curva, como puede verse en el Dibujo 9.1.

Distribución Normal Standard.

Determinar un área ó probabilidad en la distribución normal demanda la integración de la ecuación de la curva normal. Para simplificar este trabajo se utiliza la distribución normal standard ó Distribución de z. En el Dibujo 9.4 se vé la correspondencia entre la distribución normal y la distribución de z.

DIBUJO 9.4

DISTRIBUCION NORMAL CON VALORES ORIGINALES Y TRANSFORMADOS



La distribución de z se obtiene restando a las observaciones originales su media y dividiendo el resultado entre su desviación standard como puede verse en la siguiente ecuación:

$$z_i = \frac{X_i - \mu}{\sigma}$$

Esta transformación de los datos en z dan una distribución que tiene como media 0 y variancia 1, como se demuestra a continuación:

	Observaciones originales	Valores transformados por z
	X_1	$\frac{X_1 - \mu}{\sigma}$
	X_2	$\frac{X_2 - \mu}{\sigma}$
	$\vdots$	$\vdots$
	X_N	$\frac{X_N - \mu}{\sigma}$
Media	μ	0
Desviación standard	σ	1

La media de la distribución normal standard, $\bar{z}$, es:

$$\bar{z} = \frac{\sum_{i=1}^{N} \left[\frac{X_i - \mu}{\sigma}\right]}{N} = \frac{\frac{1}{\sigma}\left[\sum_{i=1}^{N} X_i - N\mu\right]}{N}$$

$$\bar{z} = \frac{\sum_{i=1}^{N} X_i - N\mu}{N\sigma} = \frac{1}{\sigma}\left[\frac{\sum_{i=1}^{N} X_i}{N} - \frac{N\mu}{N}\right]$$

$$\bar{z} = \frac{\mu - \mu}{\sigma} = 0$$

La variancia de la distribución normal standard, σ_z^2 es

$$\sigma_z^2 = \frac{\sum_{i=1}^{N} \left[\left(\frac{X_i - \mu}{\sigma}\right) - \bar{z}\right]^2}{N} = \frac{\sum_{i=1}^{N} \left(\frac{X_i - \mu}{\sigma}\right)^2}{N}$$

$$\sigma_z^2 = \frac{1}{\sigma^2}\left[\frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}\right] = \frac{\sigma^2}{\sigma^2} = 1$$

La tabla 2 corresponde al área ó probabilidad de la distribución de z ó normal standard. La primera columna y primera fila de la tabla corresponden a los valores de z debiendo leerse primero la columna y luego la fila y las probabilidades constituyen el cuerpo de la tabla.

La aplicación de la distribución de z en el cálculo de probabilidades ó área de una distribución normal se ilustra a través de los siguientes ejemplos:

Ejemplo 1. Hallár el área ó probabilidad de encontrar al azar valores de z entre:

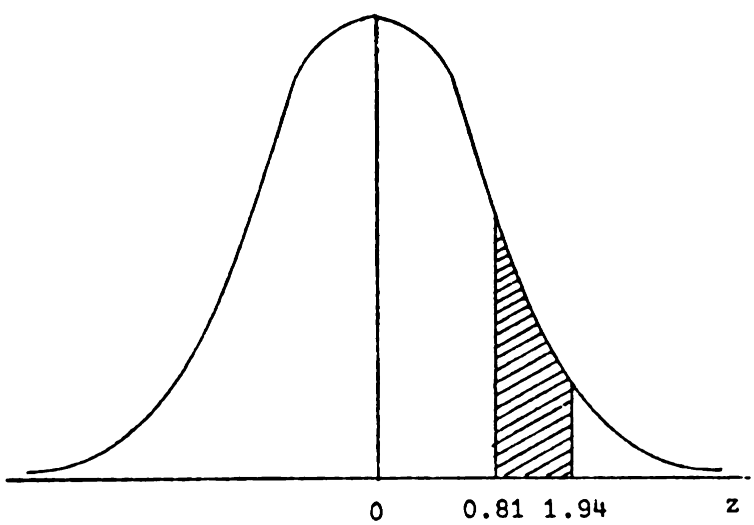
a) 0.81 y 1.94 (Area sombreada)

Area requerida = (Area entre z = 0 y z = 1.94)

- (Area entre z = 0 y z = 0.81)

= 0.4738 - 0.2910 = 0.1828

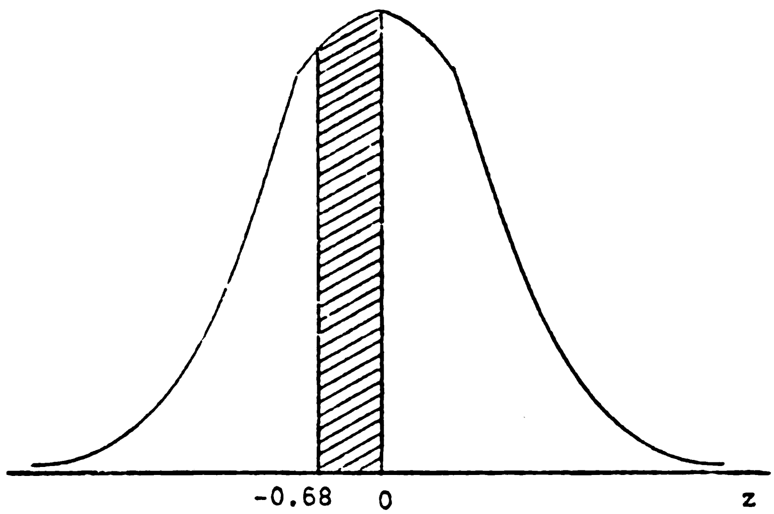
$P(0.81 < z < 1.94) = 0.1828$



b) 0.68 y 0 (Area sombreada)

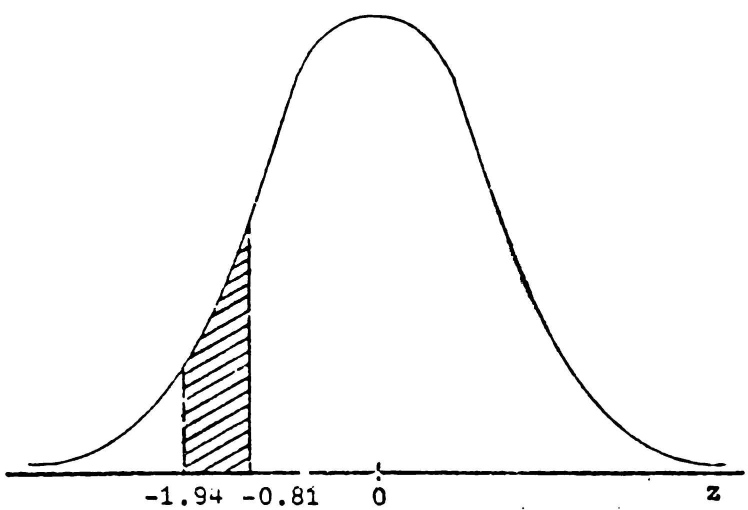
Por simetría, equivale a encontrar el área de z = 0 a z = 0.68

$P(-0.68 < z < 0) = 0.2518$



c) -1.94 y -0.81

Por simetría el área corresponde a valores de z entre 0.81 y 1.94 calculada en (a)

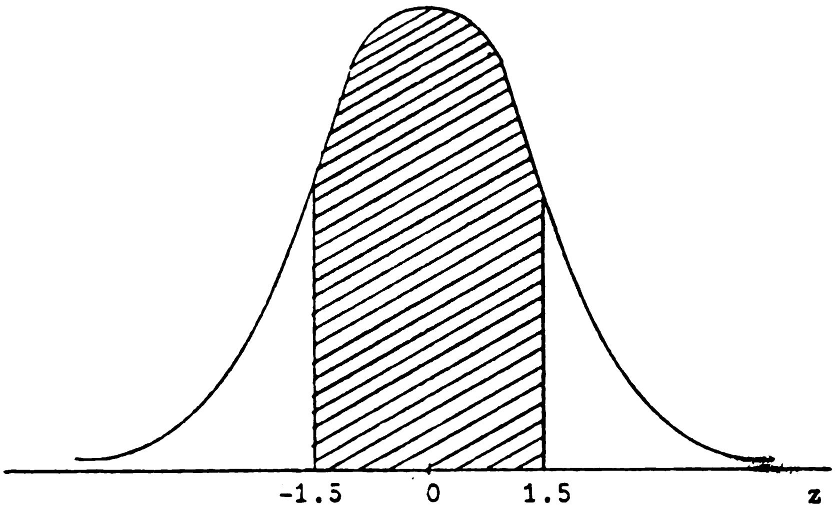


d) -1.5 a 1.5

Por simetría el área podría encontrarse de la siguiente manera:

2 (área entre z = 0 y z = 1.5)

= 2 (0.4332) = 0.8664



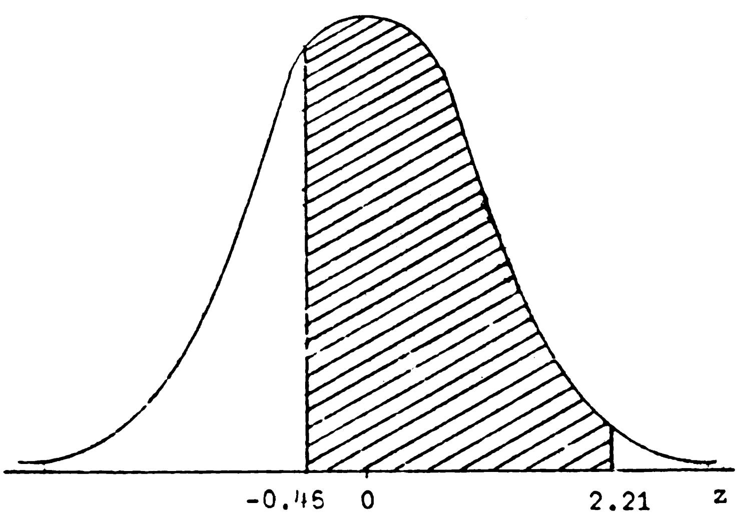
e) -0.46 y 2.21

Area requerida = (Area entre z = -0.46 y z = 0)

+ (Area entre z = 0 y z = 2.21)

= 0.1772 + 0.4864 = 0.6636

P (-0.46 < z < 2.21) = 0.6636



f) 1.94 y ∞

En el gráfico se establece el área deseada

Area requerida = (Area entre z = 0 y z = ∞)

- (Area entre z = 0 y z = 1.94)

= 0.5000 - 0.4738

= 0.0262

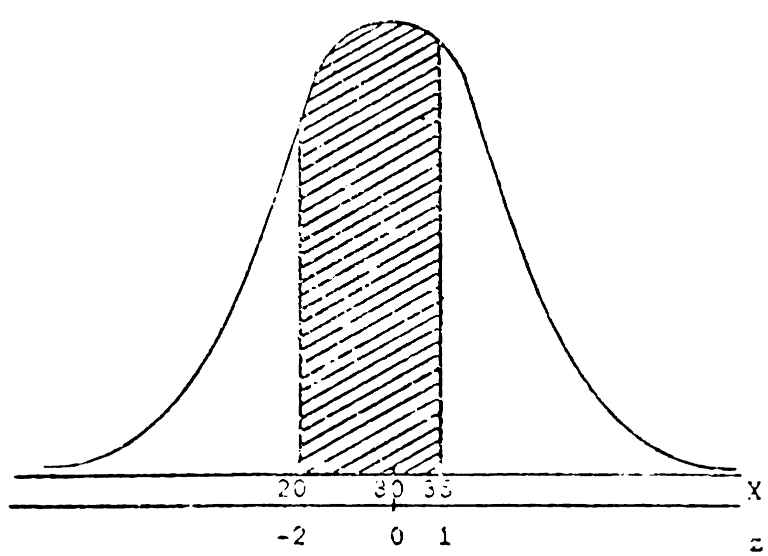
Igual procedimiento se sigue para valores entre $-\infty$ y cualquier otro valor negativo.

Muchos problemas de la distribución normal pueden resolverse mediante la distribución de z como puede verse a través de los siguientes ejemplos.

Ejemplo 1. En una población normalmente distribuída con μ = 30 y σ = 5

a) ¿Qué porcentaje del número total de observaciones estarán entre 20 y 35?

Estamos interesados en encontrar el área sombreada en el siguiente dibujo:



El problema planteado corresponde al evento $(20 < X < 35)$ y si a todos los términos de la desigualdad se le resta una constante (μ) y se divide entre otra constante (σ) la desigualdad no se altera. Es decir:

$$(20 < X < 35) = \left(\frac{20 - 30}{5} < \frac{X - \mu}{\sigma} < \frac{35 - 30}{5}\right)$$

$$(20 < X < 35) = (-2 < z < 1)$$

los eventos son equivalentes y por lo tanto tendrán igual probabilidad; luego

$$P(20 < X < 35) = P(-2 < z < 1) = 0.8185$$

Luego se concluye que 81.85% de las observaciones de la población están entre 20 y 35, en otras palabras diremos que existe 0.8185 probabilidades de que al seleccionar una observación al azar, ésta tenga un valor entre 20 y 35

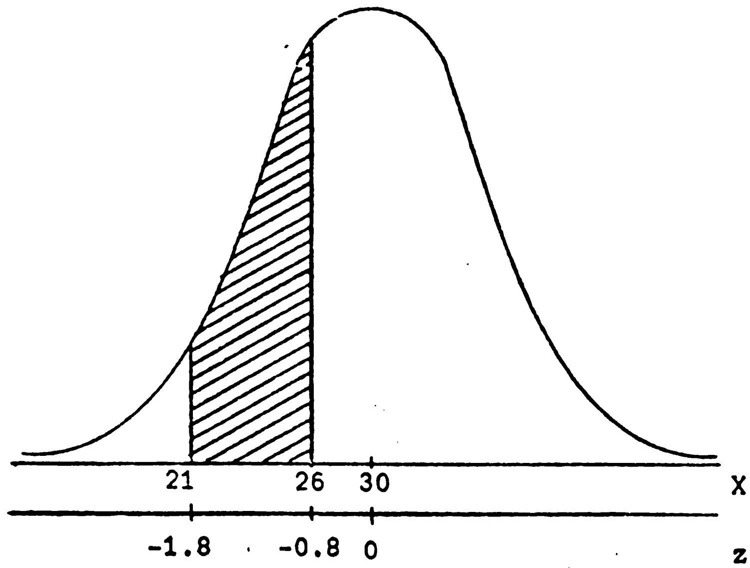
b) ¿Cuál es la probabilidad de que al seleccionar al azar un valor de X, éste tenga un valor entre 21 y 26?

Estamos interesados en encontrar $P(21 < X < 26)$. El evento $(21 < X < 26)$, por la transformación de z, equivale al evento $(-1.8 < z < -0.8)$ luego ambos eventos tienen igual probabilidad:

$$P(21 < X < 26) = P(-1.8 < z < -0.8) = 0.1760$$

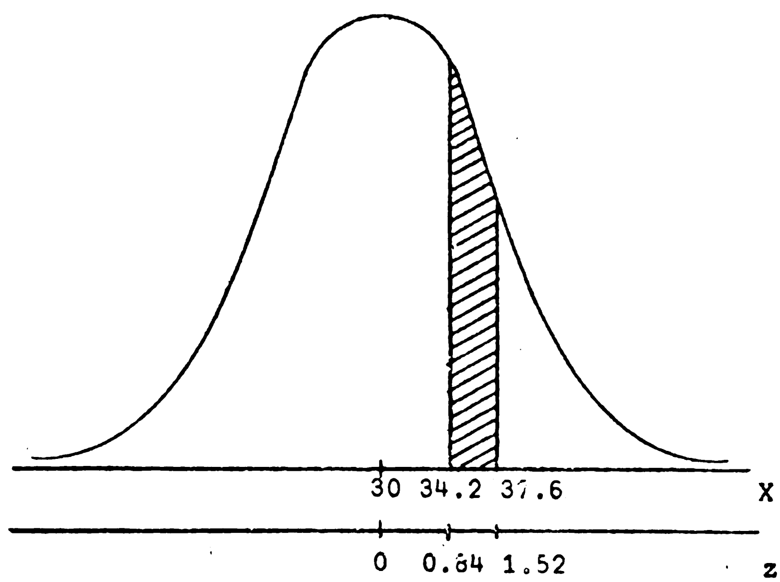
La probabilidad de que un individuo seleccionado al azar tenga un valor entre 21 a 26 es 0.1760.

A continuación se dá el dibujo del ejemplo.



c) ¿Qué porcentaje del número total de observaciones estarán entre 34.2 y 37.6?

El área requerida se vé en el siguiente dibujo.



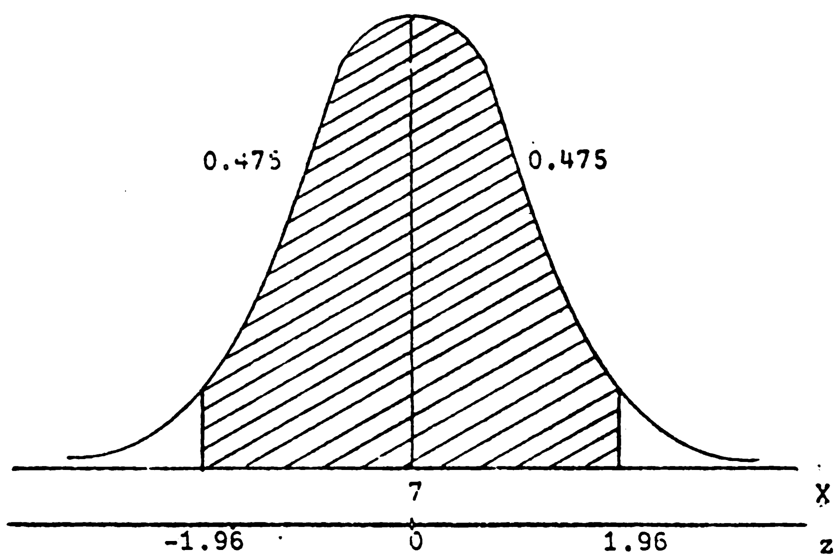
La probabilidad es la siguiente:

$P\ (34.2 < X < 37.6) = P\ (0.84 < z < 1.52) = 0.1361$

Concluiremos diciendo que, 13.61% de las observaciones de la población tienen valores entre 34.2 y 37.6

Ejemplo 2. En una población normalmente distribuida con $\mu = 7$ y $\sigma = 2$ y si el área es repartida por igual a ambos lados de la media y mide en total 0.95 ¿Cuáles son los valores correspondientes de z?

Como la curva normal es simétrica, la mitad del área ó sea 0.475 está a la derecha de μ y la otra mitad 0.475 a la izquierda, como puede verse en el siguiente dibujo.



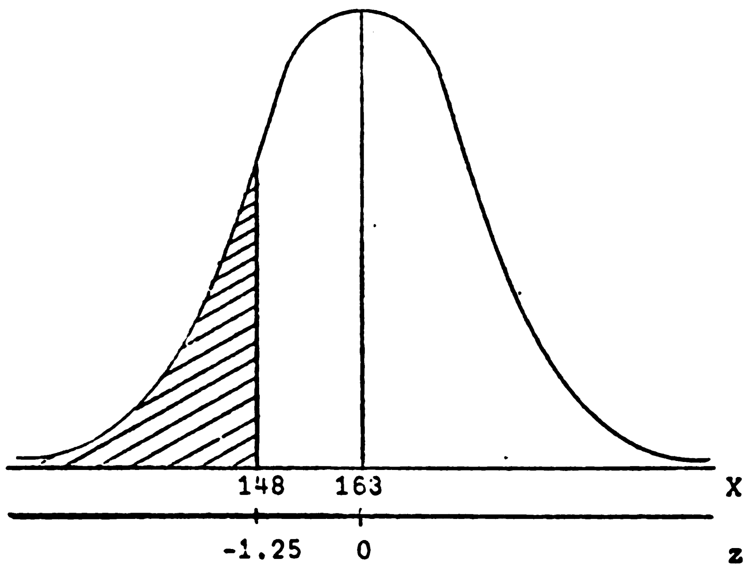
El problema consiste en encontrar el valor de z que corresponde a una probabilidad 0.475 en la tabla y éste es 1.96.

Por simetría el otro valor de z será -1.96.

Ejemplo 3. En una población normalmente distribuída con $\mu = 163$ y $\sigma = 12$

a) ¿Cuál es la probabilidad que un individuo seleccionado al azar tenga un valor menor que 148?

La probabilidad requerida puede verse en el siguiente dibujo



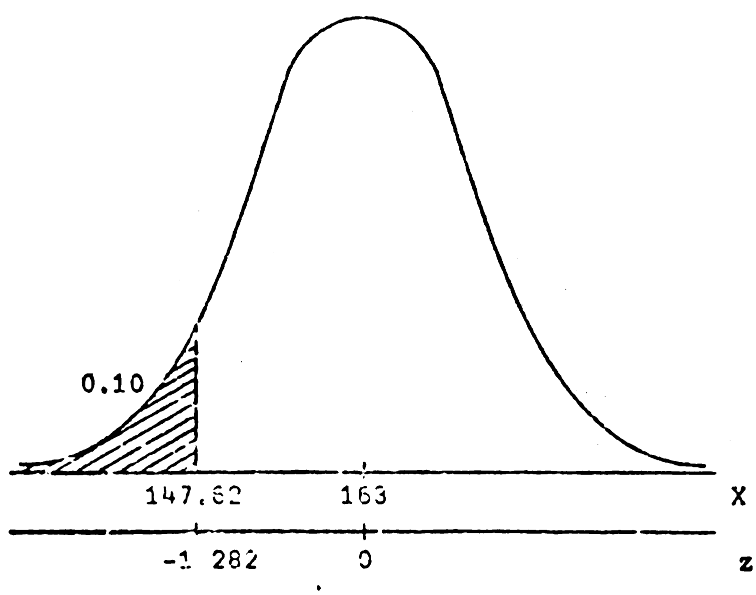
El evento (X < 148) equivale al evento (z < -1.25)

luego P (X < 148) = P (z < -1.25) = 0.1056

Entonces la probabilidad que una variable seleccionada al azar tenga un valor menor de 148 es 0.1056

b) ¿Debajo de qué valor de X se encuentran el 10% de las observaciones?

El área requerida puede verse en el siguiente dibujo.



Si el área entre $-\infty$ y z es 0.10, el área entre z y μ será 0.50 - 0.10 = 0.40. En la tabla de z mediante interpolación encontramos z = 1.282 y de la ecuación $z = \frac{X - \mu}{\sigma}$ tendremos:

$X = 1.63 - (1.282)(12) = 147.62$

Concluiremos que el 10% de las observaciones tienen valores menores de 147.62.

Distribución de medias.

Si de una población se extraen todas las muestras posibles, de tamaño n y en cada muestra se calcula la media, habremos formado una POBLACION DERIVADA DE MEDIAS DE MUESTRAS.

La población de medias es una de los más importantes tópicos de la Estadística, por la relación existente entre las características de la población original y las características de la población derivada, como puede verse a continuación.

Consideremos la siguiente población: $X_1 = 1$, $X_2 = 2$, $X_3 = 3$, $X_4 = 4$, $X_5 = 5$ y $X_6 = 6$.

En esta población: $\mu = 3.5$, $\sigma^2 = 2.917$ y $\sigma = 1.707$. Si de esta población extraemos con reemplazo todas las muestras posibles de tamaño 2, tendremos 36 muestras. La constitución de las muestras y sus medias se dan en el Cuadro 9.1.

La población derivada de medias tendrá su media $\mu_{\bar{x}}$ y su variancia $\sigma^2_{\bar{x}}$. El valor de estos parámetros los hallaremos agrupando

CUADRO 9.1

DISTRIBUCION DE MEDIAS

Población Original	Selección de la Muestra		Constitución de la Muestra	Población derivada de medias ($\bar{x}$)
	Primera observación	Segunda observación		
1	1	1	1, 1	1
2		2	1, 2	1.5
3		3	1, 3	2
4		4	1, 4	2.5
5		5	1, 5	3
6		6	1, 6	3.5
	2	1	2, 1	1.5
		2	2, 2	2
		3	2, 3	2.5
		4	2, 4	3
		5	2, 5	3.5
		6	2, 6	4
	3	1	3, 1	2
		2	3, 2	2.5
		3	3, 3	3
		4	3, 4	3.5
		5	3, 5	4
		6	3, 6	4.5
	4	1	4, 1	2.5
		2	4, 2	3
		3	4, 3	3.5
		4	4, 4	4
		5	4, 5	4.5
		6	4, 6	5
	5	1	5, 1	3
		2	5, 2	3.5
		3	5, 3	4
		4	5, 4	4.5
		5	5, 5	5
		6	5, 6	5.5
	6	1	6, 1	3.5
		2	6, 2	4
		3	6, 3	4.5
		4	6, 4	5
		5	6, 5	5.5
		6	6, 6	6

làs medias en la siguiente tabla de frecuencia.

Tabla de Frecuencia	
Población derivada de medias $\bar{x}$	Frecuencia
1	1
1.5	2
2	3
2.5	4
3	5
3.5	6
4	5
4.5	4
5	3
5.5	2
6	1

La media de esta población derivada de medias es $\mu_{\bar{x}} = 3.5$ y es igual a la media, μ, de la población original, de la cual se seleccionaron las muestras y la variancia de esta población derivada, de medias es igual a $\sigma^2_{\bar{x}} = 1.458$ y viene a ser igual a

$$\frac{\sigma^2}{n} = \frac{2.917}{2} = 1.458$$

El tamaño de la muestra es un factor muy importante, conforme el tamaño se incrementa la distribución de medias se aproxima a la distribución normal, propiedad expresada en el importante TEOREMA DEL LIMITE CENTRAL que se da a continuación.

Teorema.- Si una población tiene una media μ y una variancia σ^2 conforme el tamaño de las muestras (n) aumenta la distribución de las medias de las muestras se aproxima a la distribución normal con media μ y variancia $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$

Según este teorema, la distribución de las medias se aproximará a la distribución normal conforme n aumenta, sin importar cual fué la distribución de la población original.

La definición de z, también podemos aplicarla a la población de medias según la siguiente ecuación:

$$z = \frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

La distribución de z también puede usarse para calcular el área entre dos medias $\bar{x}_1$ y $\bar{x}_2$, como puede verse en el siguiente ejemplo.

De una población normal con $\mu = 50$ y $\sigma^2 = 100$ se extrae al azar una muestra de 25 observaciones. ¿Cuál es la probabilidad de que la media de esta muestra, $\bar{x}$, tenga un valor entre 48 y 52?

Aplicando la transformación de z al evento tendremos:

$$(48 < \bar{x} < 52) = \left[\frac{48 - 50}{\sqrt{\frac{100}{25}}} < \frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}} < \frac{52 - 50}{\sqrt{\frac{100}{25}}}\right] = (-1 < z < 1)$$

luego

$P(48 < \bar{x} < 52) = P(-1 < z < 1) = 0.6826$

Concluimos diciendo que la probabilidad de que la media de esta muestra se encuentre entre 48 y 52 es 0.6826.

Distribución de "t"

Es una distribución importante llamada así por W. S. Gosset quien la desarrolló y la publicó por primera vez en la revista Biometrika, bajo el seudónimo de Student.

La distribución de z para medias según la ecuación:

$$z = \frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

es muy útil cuando se conoce la variancia de la población (σ^2), no así cuando sólo podemos estimar esta variancia, resultando muy útil en estos casos la distribución de t para hallar probabilidades. Esta distribución está definida por la siguiente ecuación:

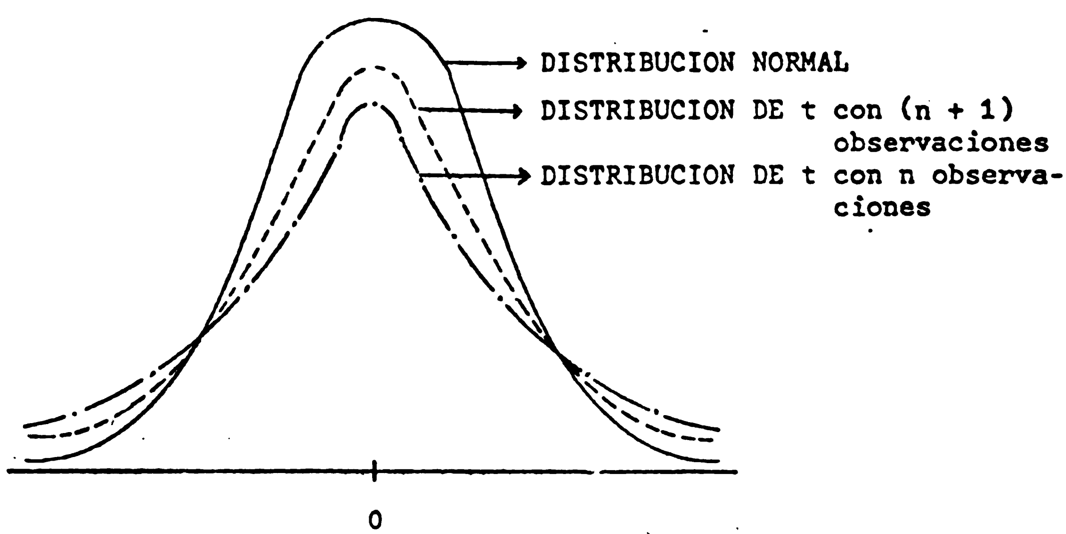
$$t = \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{n-1}}}$$

La distribución de t tiene forma acampanada muy parecida a la distribución normal. La distribución de t es en realidad un conjunto de curvas, dependiendo cada una de ellas del tamaño de la muestra.

Cuando el tamaño de muestra aumenta, la distribución de t se aproxima más a la normal y cuando n se aproxima a infinito, la distribución de t es igual a la normal, como puede verse en el Dibujo 9.5

DIBUJO 9.5

DISTRIBUCION DE "t" PARA DISTINTOS TAMAÑOS DE MUESTRA



Las probabilidades de la distribución de t está dada en la Tabla 3, en la cual cada fila representa un número particular de grados de libertad (n - 1), y en el cuerpo de la Tabla están dados los valores absolutos de t para distintas probabilidades.

El uso de la Tabla 3 se ilustra en los siguientes ejemplos:

Ejemplo 1. Para una muestra de tamaño 7, es decir con 6 grados de libertad, cual es la probabilidad de encontrar al azar un valor absoluto de t mayor de 2.447, es decir $P(|t| > 2.447)$

Con 6 grados de libertad buscamos en la sexta fila el valor de 2.447 que se encuentra en la columna de 0.05 probabilidades. De la ecuación anterior podemos concluir, debido a la simetría de la distribución de t que:

$P(t > 2.447) = 0.025$ y $P(t < -2.447) = 0.025$

Ejemplo 2. En una muestra de tamaño 23 y una probabilidad de 0.2 ¿Cuál es el valor de t_o tal que se cumpla la siguiente ecuación?

$P(|t| > t_o) = 0.2$

En la intersección de la vigésima segunda fila y la columna de probabilidades de 0.2 se encuentra el valor 1.321, es decir t_o

Ejemplo 3. Para una muestra de tamaño 11.

a) La probabilidad de encontrar un valor de t mayor que 4.587 es:

$P(t > 4.587) = 0.0005$

b) La probabilidad de encontrar un valor de t menor que -1.812, es:

$P(t < -1.812) = 0.05$

c) La probabilidad de encontrar un valor de t mayor que -2.764, es:

$(t > -2.764) = 0.99$

Este resultado ha sido obtenido de la siguiente manera:

$P(t < -2.764) = 0.01$

Luego la probabilidad del complemento será $1 - 0.01 = 0.99$ que es la probabilidad anteriormente dada.

Distribución de χ^2

Otra importante distribución de probabilidades es la de JI - CUADRADA, simbolizada por la letra griega ji al cuadrado, χ^2. La distribución de Ji-cuadrada está definida por

$$\chi^2 = \sum_{i=1}^{n} z_i^2$$

donde n representa el número de observaciones y varía de 1 a ∞ y z es una variable normalmente distribuída con media cero (0), y variancia uno (1).

El valor de χ^2, obtenido a partir de una muestra variará al cambiar de muestra. Estos cambios se ilustran en el siguiente ejemplo:

De una población normal con $\mu = 40$ y $\sigma^2 = 100$

a) Se extrajo una muestra de 5 observaciones, obteniendo los siguientes resultados 53, 48, 45, 33, 50, el valor de χ^2 será:

$$\chi^2 = \sum_{i=1}^{5} z_i^2 = \left(\frac{53-40}{10}\right)^2 + \left(\frac{48-40}{10}\right)^2 + \left(\frac{45-40}{10}\right)^2 + \left(\frac{33-40}{10}\right)^2 + \left(\frac{50-40}{10}\right)^2$$

$$= \frac{(53-40)^2 + (48-40)^2 + (45-40)^2 + (33-40)^2 + (50-40)^2}{100}$$

$$= \frac{13^2 + 8^2 + 5^2 + (-7)^2 + 10^2}{100} = \frac{407}{100} = 4.07$$

b) En una segunda muestra de 5 observaciones se obtuvo 30, 35, 58, 60, 70 y el valor de χ^2 será:

$$\chi^2 = \sum_{i=1}^{5} z_i^2 = \left(\frac{30-40}{10}\right)^2 + \left(\frac{35-40}{10}\right)^2 + \left(\frac{58-40}{10}\right)^2 + \left(\frac{60-40}{10}\right)^2 + \left(\frac{70-40}{10}\right)^2$$

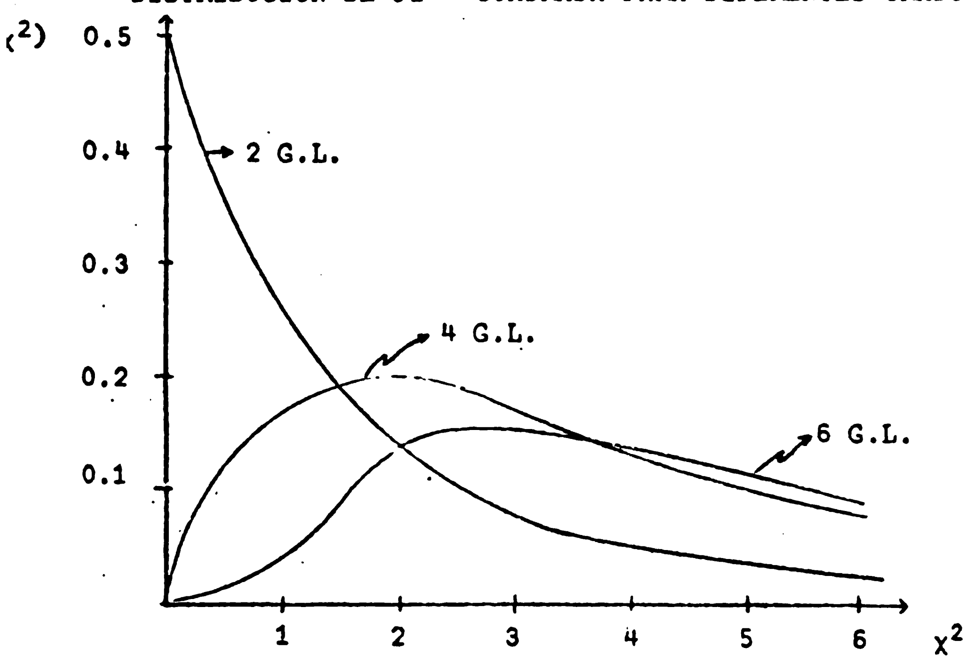
$$= \frac{(30-40)^2 + (35-40)^2 + (58-40)^2 + (60-40)^2 + (70-40)^2}{100}$$

$$= \frac{(-10)^2 + (-5)^2 + 18^2 + 20^2 + 30^2}{100} = \frac{1749}{100} = 17.49$$

Como puede verse dos muestras del mismo tamaño extraídas de la misma población tienen diferentes valores de χ^2 y para cada tamaño de muestra habrá una distribución de χ^2.

DIBUJO 9.7

DISTRIBUCION DE JI - CUADRADA PARA DIFERENTES GRADOS DE LIBERTAD



El Dibujo 9.7 muestra las formas de la distribución de χ^2 para diferente número de observaciones.

La Tabla 4 muestra la distribución de probabilidades para diferentes tamaños de muestra que nos da la probabilidad de encontrar valores más altos de χ_o^2. Por ejemplo para 18 grados de libertad, la probabilidad de encontrar al azar un valor de χ^2 mayor de 9.390 es 0.95, es decir

$$P(\chi^2 > 9.390) = 0.95$$

Distribución de $(n-1)\,s^2/\sigma^2$

Según hemos establecido por definición

$$\chi^2 = \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{\sigma^2}$$

donde: n es el número de observaciones.

Esta ecuación se puede igualar a:

$$\frac{\sum_{i=1}^{n} (X_i - \mu)^2}{\sigma^2} = \frac{\sum_{i=1}^{n} \left[(X_i - \bar{x}) + (\bar{x} - \mu)\right]^2}{\sigma^2}$$

$$= \frac{\sum_{i=1}^{n} (X_i - \bar{x})^2 + n(\bar{x} - \mu)^2 + 2(\bar{x} - \mu) \sum_{i=1}^{n} (X_i - \bar{x})}{\sigma^2}$$

$$= \frac{\sum_{i=1}^{n} (X_i - \bar{x})^2}{\sigma^2} + \frac{n(\bar{x} - \mu)^2}{\sigma^2}$$

Finalmente:

$$\frac{\sum_{i=1}^{n} (X_i - \bar{x})^2}{\sigma^2} = \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{\sigma^2} - \frac{(\bar{x} - \mu)^2}{\frac{\sigma^2}{n}}$$

Donde:

$$\frac{\sum_{i=1}^{n} (X_i - \mu)^2}{\sigma^2}$$ está distribuído como χ^2 con n grados de libertad.

$$\frac{(\bar{x} - \mu)^2}{\frac{\sigma^2}{n}}$$ está distribuído como χ^2 con 1 grado de libertad.

Luego $$\frac{\sum_{i=1}^{n} (X_i - \bar{x})^2}{\sigma^2}$$ seguirá una distribución de χ^2 con (n-1) grados de libertad.

Finalmente, teniendo en cuenta que

$s^2 = \dfrac{\Sigma(X_i - \bar{x})^2}{n-1}$, de donde $\Sigma(X_i - \bar{x})^2 = (n-1)s^2$, tendremos que:

$$\frac{\sum_{i=1}^{n} (X_i - \bar{x})^2}{\sigma^2} = \frac{(n-1)s^2}{\sigma^2}$$

Distribución de F.

Sea W una variable distribuída como χ^2 con r_1 grados de libertad y V otra variable distribuída como χ^2 con r_2 grados de libertad. La ecuación:

$F = \dfrac{W/r_1}{V/r_2}$ define la importante distribución de probabilidades conocida como DISTRIBUCION DE F. Esta distribución es muy útil para prueba de hipótesis sobre variancias como veremos en el siguiente capítulo.

Sea $W = \dfrac{(n_1-1)s_1^2}{\sigma^2}$ distribuída como χ^2 con $r_1 = (n_1 -1)$ grados de libertad, y $V = \dfrac{(n_2-1)s_2^2}{\sigma^2}$ distribuído como χ^2 con $r_2 = (n_2-1)$ grados de libertad, luego:

$$F = \frac{\dfrac{(n_1-1)s_1^2}{(n_1-1)\sigma^2}}{\dfrac{(n_2-1)s_2^2}{(n_2-1)\sigma^2}} = \frac{s_1^2}{s_2^2}$$

En el Dibujo 9.8 se ilustra la forma de la distribución de F y como puede verse, varía de acuerdo a los grados de libertad del numerador (n_1-1) y a los grados de libertad del denominador (n_2-1).

La Tabla 5 da los valores de F para diferentes grados de libertad del numerador y denominador para probabilidades de 0.05 y 0.01.

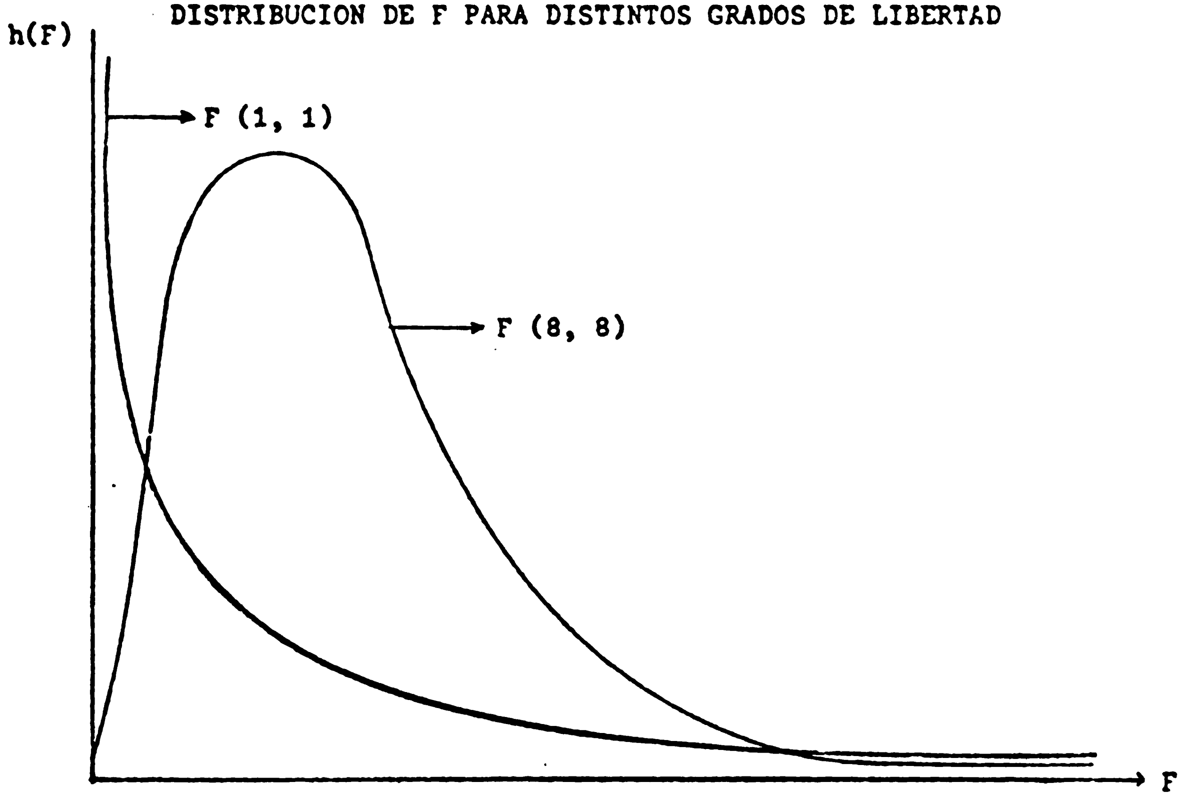
La Tabla 5 da la probabilidad de encontrar valores mayores de F_o. Ejemplo: con 10 grados de libertad del numerador y 22 grados de libertad de denominador

$p\ (F > 2.30) = 0.05$ y

$p\ (F > 3.26) = 0.01$

DIBUJO 9.8

DISTRIBUCION DE F PARA DISTINTOS GRADOS DE LIBERTAD



EJERCICIOS

1. Halle la probabilidad para los siguientes valores de z.

a) $P(z > 1.43)$

b) $P(z < -1.43)$

c) $P(z < 2.50)$

d) $P(z > -2.89)$

e) $P(1.33 < z < 1.85)$

f) $P(-1.98 < z < -1.25)$

g) $P(-0.66 < z < 1.24)$

h) $P(1.19 < z < \infty)$

i) $P(-\infty < z < -0.33)$

j) $P(|z| > 1.96)$

k) $P(|z| < 1.65)$

2. Halle los valores de z_o para las siguientes probabilidades

a) $P(z > z_o) = .0987$

b) $P(z < z_o) = .4793$

c) $P(z < z_o) = .4015$

d) $P(-z_o < z < z_o) = 0.01$

e) $P(|z| > z_o) = 0.05$

3. En una población normal con $\mu = 25$ y $\sigma^2 = 49$ halle:

a) $P(X > 25)$

b) $P(X > 27)$

c) $P(X < 2)$

d) $P(25 < X < 27)$

e) $P(27 < X < 29)$

f) $P(20 < X < 22)$

g) $P(22 < X < 26)$

4. De una población normal con $\mu = 50$ y $\sigma^2 = 36$ se extrae una muestra de 16 observaciones, halle:

a) $P(\bar{x} > 52)$

b) $P(\bar{x} < 48)$

c) $P(48 < \bar{x} < 52)$

d) $P(\bar{x} < 55)$

e) $P(\bar{x} > 46)$

f) $P(55 < \bar{x} < 60)$

g) $P(46 < \bar{x} < 48)$

5. Una planta embotelladora envasa leche en depósitos descartables de cartón de 1.20 litros de capacidad. La variancia en el llenado automático es de 0.000004 $(\text{litros})^2$ y envasa en un día 10,000 cartones.

a) ¿Cuál es la probabilidad de encontrar al azar un cartón con más de 1.005 litros de leche?

b) ¿Cuál es la probabilidad de encontrar al azar un cartón con más de 0.996 litros de leche?

c) Si se escoge al azar un lote de 100 cartones ¿cuál es la probabilidad de que ese lote tenga un promedio entre 1.005 y 0.986 litros?

6. Si la producción de leche de las vacas Holstein en una ciudad tiene como $\mu = 12$ litros diarios y $\sigma^2 = 25$ litros y se extrae al azar una muestra de 100 vacas, halle:

a) $P(\bar{x} > 14)$

b) $P(12 < \bar{x} < 16)$

c) $P(\bar{x} > 20)$

d) $P(\bar{x} < 11)$

e) $P(\bar{x} > 12)$

f) $P(10 < \bar{x} < 11)$

g) $P(\bar{x} > 11.5)$

7. Si el peso promedio (μ) de las terneras Holstein al nacimiento es de 35 Kgs. y la variancia es de $\sigma^2 = 49$. Si se extrae una muestra de 9 pesos halle:

a) $P(X > 36)$

b) $P(\bar{x} > 36)$

c) $P(33 < X < 38)$

d) $P(33 < \bar{x} < 38)$

e) $P(33 < X < 35)$

f) $P(33 < \bar{x} < 35)$

g) $P(X > 33)$

h) $P(\bar{x} < 33)$

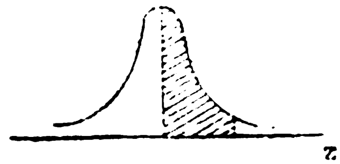
i) $P(X < 35)$

j) $P(\bar{x} < 35)$

8. Un empleado maneja su automovil diariamente de su casa a la oficina por varios años y encuentra que demora en llegar a su trabajo en promedio $\mu = 12.5$ minutos con una desviación standard de 4.5 minutos. Si el empleado sale de su casa a las 8.15 a.m. y debe estar en su trabajo a las 8.30 a.m. ¿Cuál es la probabilidad de que llegue tarde el día Martes?

TABLA 2

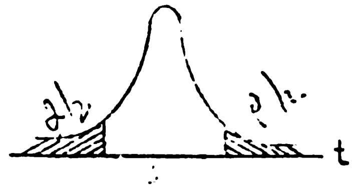
DISTRIBUCION NORMAL STANDARD



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	0000	0040	0080	0120	0160	0199	0239	0279	0319	0359
.1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
.2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
.3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
.4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
.5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
.6	2257	2291	2324	2357	2389	2422	2454	2486	2517	2549
.7	2580	2611	2642	2673	2704	2734	2764	2794	2823	2852
.8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
.9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1.0	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1.1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1.2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1.3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1.4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1.5	4332	4345	4357	4370	4382	4394	4406	4418	4429	4441
1.6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1.7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1.8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1.9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2.0	4772	4778	4783	4788	4793	4798	4803	4808	4812	4817
2.1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2.2	4861	4864	4868	4871	4875	4878	4881	4884	4887	4890
2.3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2.4	4918	4920	4922	4925	4927	4929	2931	4932	4934	4936
2.5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2.6	4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
2.7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2.8	4974	4975	4976	4977	4977	4978	4979	4979	4980	4981
2.9	4981	4982	4982	4983	4984	4984	4985	4985	4986	4986
3.0	4987	4987	4987	4988	4988	4989	4989	4989	4990	4990
3.1	4990	4991	4991	4991	4992	4992	4992	4992	4993	4993
3.2	4993	4993	4994	4994	4994	4994	4994	4995	4995	4995
3.3	4995	4995	4995	4996	4996	4996	4996	4996	4996	4997
3.4	4997	4997	4997	4997	4997	4997	4997	4997	4997	4998
3.6	4998	4998	4999	4999	4999	4999	4999	4999	4999	4999
3.9	5000									

TABLA 3

DISTRIBUCION DE t

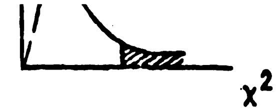


G.L.	0.5	0.4	0.3	0.2	0.1	0.05	0.02	0.01	0.001
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.598
3	.765	.978	1.250	1.638	2.353	3.182	4.541	5.841	12.941
4	.741	.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	.727	.920	1.156	1.476	2.015	2.571	3.365	4.032	6.859
6	.718	.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.998	3.499	5.405
8	.706	.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.821	3.250	4.761
10	.700	.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	.688	.861	1.066	1.328	1.729	2.092	2.539	2.861	3.863
20	.687	.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.500	2.807	3.767
24	.685	.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
60	.679	.848	1.046	1.296	1.671	2.000	2.390	2.660	3.460
120	.677	.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373
∞	.674	.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

Reproducido de la Tabla III de las Tablas Estadísticas para investigaciones en Biología, Agricultura y Medicina, de Fisher y Yates, publicada por Oliver y Boyd Co. Edimburgo, con gentil permiso de los autores y editores.

TABLA 4

DISTRIBUCION DE χ^2



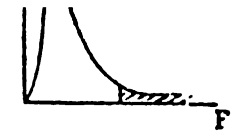
G.L.	Probabilidad de un valor más alto χ^2												
	0.99	0.98	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.02	0.01
1	0.000	0.001	0.004	0.016	0.064	0.148	0.455	1.074	1.642	2.705	3.841	5.412	6.635
2	0.020	0.040	0.103	0.211	0.446	0.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210
3	0.115	0.185	0.352	0.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	9.837	11.341
4	0.297	0.429	0.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	11.668	13.277
5	0.554	0.752	1.145	1.610	2.343	3.000	4.351	6.064	7.289	9.236	11.070	13.388	15.086
6	0.872	1.134	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	15.033	16.812
7	1.239	1.564	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475
8	1.646	2.032	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	18.168	20.090
9	2.088	2.532	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	19.679	21.666
10	2.558	3.059	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209
11	3.053	3.609	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725
12	3.571	4.178	5.226	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	24.054	26.217
13	4.107	4.765	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.688
14	4.660	5.368	6.571	7.790	9.467	10.821	13.339	16.222	18.151	21.064	23.685	26.873	29.141
15	5.229	5.985	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578
16	5.812	6.614	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000
17	6.408	7.255	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.769	27.587	30.995	33.409
18	7.015	7.906	9.390	10.865	12.857	14.440	17.338	20.601	22.760	25.989	28.869	32.346	34.805
19	7.633	8.567	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.144	33.687	36.191
20	8.260	9.237	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566
21	8.897	9.915	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932
22	9.542	10.600	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289
23	10.196	11.293	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638
24	10.856	11.992	13.848	15.659	18.062	19.943	23.337	27.096	29.553	33.196	36.415	40.260	42.980
25	11.524	12.697	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.652	41.566	44.314
26	12.198	13.409	15.379	17.292	19.820	21.792	25.336	29.246	31.795	35.563	38.885	42.856	45.642
27	12.879	14.125	16.151	18.114	20.703	22.719	26.336	30.319	32.912	36.741	40.113	44.140	46.963
28	13.565	14.847	16.928	18.939	21.588	23.647	27.336	31.391	34.027	37.916	41.337	45.419	48.278
29	14.256	15.574	17.708	19.768	22.475	24.577	28.336	32.461	35.139	39.087	42.557	46.693	49.588
30	14.953	16.306	18.493	20.599	23.364	25.508	29.336	33.530	36.250	40.256	43.773	47.962	50.892

Reproducido de la Tabla III del libro Methods for Research Workers, con gentil autorización de la R. A. Fisher y sus editores, Oliver and Boyd, Edimburgo.

TABLA 5

DISTRIBUCION DE F

5% (Número superior) 1% (Número inferior)



Grados de libertad del denominador	Grados de libertad del numerador: 1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
1	161	200	216	225	230	234	237	239	241	242	243	244	245	246	248	249	250	251	252	253	253	254	254	254
	4,052	4,999	5,403	5,625	5,764	5,859	5,928	5,981	6,022	6,056	6,082	6,106	6,142	6,169	6,208	6,234	6,258	6,286	6,302	6,323	6,334	6,352	6,361	6,366
2	18.51	19.00	19.16	19.25	19.30	19.33	19.36	19.37	19.38	19.39	19.40	19.41	19.42	19.43	19.44	19.45	19.46	19.47	19.47	19.48	19.49	19.49	19.50	19.50
	98.49	99.01	99.17	99.25	99.30	99.33	99.34	99.36	99.38	99.40	99.41	99.42	99.43	99.44	99.45	99.46	99.47	99.48	99.48	99.49	99.49	99.49	99.50	99.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78	8.76	8.74	8.71	8.69	8.66	8.64	8.62	8.60	8.58	8.57	8.56	8.54	8.54	8.53
	34.12	30.81	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	27.13	27.05	26.92	26.83	26.69	26.60	25.50	26.41	26.35	26.27	26.23	26.18	26.14	26.12
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	5.91	5.87	5.84	5.80	5.77	5.74	5.71	5.70	5.68	5.66	5.65	5.64	5.63
	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.54	14.45	14.37	14.24	14.15	14.02	13.93	13.83	13.74	13.69	13.61	13.57	13.52	13.48	13.46
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74	4.70	4.68	4.64	4.60	4.56	4.53	4.50	4.46	4.44	4.42	4.40	4.38	4.37	4.36
	16.26	13.27	12.06	11.39	10.97	10.67	10.45	10.27	10.15	10.05	9.96	9.89	9.77	9.68	9.55	9.47	9.38	9.29	9.24	9.17	9.13	9.07	9.04	9.02
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.96	3.92	3.87	3.84	3.81	3.77	3.75	3.72	3.71	3.69	3.68	3.67
	13.74	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72	7.60	7.52	7.39	7.31	7.23	7.14	7.09	7.02	6.99	6.94	6.90	6.88
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.63	3.60	3.57	3.52	3.49	3.44	3.41	3.38	3.34	3.32	3.29	3.28	3.25	3.24	3.23
	12.25	9.55	8.45	7.85	7.46	7.19	7.00	6.84	6.71	6.62	6.54	6.47	6.35	6.27	6.15	6.07	5.98	5.90	5.85	5.78	5.75	5.70	5.67	5.65
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34	3.31	3.28	3.23	3.20	3.15	3.12	3.08	3.05	3.03	3.00	2.98	2.96	2.94	2.93
	11.26	8.65	7.59	7.01	6.63	6.37	6.19	6.03	5.91	5.82	5.74	5.67	5.56	5.48	5.36	5.28	5.20	5.11	5.06	5.00	4.96	4.91	4.88	4.86
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.13	3.10	3.07	3.02	2.98	2.93	2.90	2.86	2.82	2.80	2.77	2.76	2.73	2.72	2.71
	10.56	8.02	6.99	6.42	6.06	5.80	5.62	5.47	5.35	5.26	5.18	5.11	5.00	4.92	4.80	4.73	4.64	4.56	4.51	4.45	4.41	4.36	4.33	4.31
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.97	2.94	2.91	2.86	2.82	2.77	2.74	2.70	2.67	2.64	2.61	2.59	2.56	2.55	2.54
	10.04	7.56	6.55	5.99	5.64	5.39	5.21	5.06	4.95	4.85	4.78	4.71	4.60	4.52	4.41	4.33	4.25	4.17	4.12	4.05	4.01	3.96	3.93	3.91
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.86	2.82	2.79	2.74	2.70	2.65	2.61	2.57	2.53	2.50	2.47	2.45	2.42	2.41	2.40
	9.65	7.20	6.22	5.67	5.32	5.07	4.88	4.74	4.63	4.54	4.46	4.40	4.29	4.21	4.10	4.02	3.94	3.86	3.80	3.74	3.70	3.66	3.62	3.60
12	4.75	3.88	3.49	3.26	3.11	3.00	2.92	2.85	2.80	2.76	2.72	2.69	2.64	2.60	2.54	2.50	2.46	2.42	2.40	2.36	2.35	2.32	2.31	2.30
	9.33	6.93	5.95	5.41	5.06	4.82	4.65	4.50	4.39	4.30	4.22	4.16	4.05	3.98	3.86	3.78	3.70	3.61	3.56	3.49	3.46	3.41	3.38	3.36

TABLA 5

DISTRIBUCION DE F

5% (Número superior) 1% (Número inferior)

Grados de libertad del denominador	Grados de libertad del numerador																							
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
13	4.67	3.80	3.41	3.18	3.02	2.92	2.84	2.77	2.72	2.67	2.63	2.60	2.55	2.51	2.46	2.42	2.38	2.34	2.32	2.28	2.26	2.24	2.22	2.21
	9.07	6.70	5.74	5.20	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96	3.85	3.78	3.67	3.59	3.51	3.42	3.37	3.30	3.27	3.21	3.18	3.16
14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60	2.56	2.53	2.48	2.44	2.39	2.35	2.31	2.27	2.24	2.21	2.19	2.16	2.14	2.13
	8.86	6.51	5.56	5.03	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.70	3.62	3.51	3.43	3.34	3.26	3.21	3.14	3.11	3.06	3.02	3.00
15	4.54	3.68	3.29	3.06	2.90	2.79	2.70	2.64	2.59	2.55	2.51	2.48	2.43	2.39	2.33	2.29	2.25	2.21	2.18	2.15	2.12	2.10	2.08	2.07
	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.56	3.48	3.36	3.29	3.20	3.12	3.07	3.00	2.97	2.92	2.89	2.87
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.45	2.42	2.37	2.33	2.28	2.24	2.20	2.16	2.13	2.09	2.07	2.04	2.02	2.01
	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.61	3.55	3.45	3.37	3.25	3.18	3.10	3.01	2.96	2.89	2.86	2.80	2.77	2.75
17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55	2.50	2.45	2.41	2.38	2.33	2.29	2.23	2.19	2.15	2.11	2.08	2.04	2.02	1.99	1.97	1.96
	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.45	3.35	3.27	3.16	3.08	3.00	2.92	2.86	2.79	2.76	2.70	2.67	2.65
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.29	2.25	2.19	2.15	2.11	2.07	2.04	2.00	1.98	1.95	1.93	1.92
	8.28	6.01	5.09	4.58	4.25	4.01	3.85	3.71	3.60	3.51	3.44	3.37	3.27	3.19	3.07	3.00	2.91	2.83	2.78	2.71	2.68	2.62	2.59	2.57
19	4.38	3.52	3.13	2.90	2.74	2.63	2.55	2.48	2.43	2.38	2.34	2.31	2.26	2.21	2.15	2.11	2.07	2.02	2.00	1.96	1.94	1.91	1.90	1.88
	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30	3.19	3.12	3.00	2.92	2.84	2.76	2.70	2.63	2.60	2.54	2.51	2.49
20	4.35	3.49	3.10	2.87	2.71	2.60	2.52	2.45	2.40	2.35	2.31	2.28	2.23	2.18	2.12	2.08	2.04	1.99	1.96	1.92	1.90	1.87	1.85	1.84
	8.10	5.85	4.94	4.43	4.10	3.87	3.71	3.56	3.45	3.37	3.30	3.23	3.13	3.05	2.94	2.86	2.77	2.69	2.63	2.56	2.53	2.47	2.44	2.42
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.20	2.15	2.09	2.05	2.00	1.96	1.93	1.89	1.87	1.84	1.82	1.81
	8.02	5.78	4.87	4.37	4.04	3.81	3.65	3.51	3.40	3.31	3.24	3.17	3.07	2.99	2.88	2.80	2.72	2.63	2.58	2.51	2.47	1.42	2.38	2.36
22	4.30	3.44	3.05	2.82	2.66	2.55	2.47	2.40	2.35	2.30	2.26	2.23	2.18	2.13	2.07	2.03	1.98	1.93	1.91	1.87	1.84	1.81	1.80	1.78
	7.94	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12	3.02	2.94	2.83	2.75	2.67	2.58	2.53	2.46	2.42	2.37	2.33	2.31
23	4.28	3.42	3.03	2.80	2.64	2.53	2.45	2.38	2.32	2.28	2.24	2.20	2.14	2.10	2.04	2.00	1.96	1.91	1.88	1.84	1.82	1.79	1.77	1.76
	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07	2.97	2.89	2.78	2.70	2.62	2.53	2.48	2.41	2.37	2.32	2.28	2.26
24	4.26	3.40	3.01	2.78	2.62	2.51	2.43	2.36	2.30	2.26	2.22	2.18	2.13	2.09	2.02	1.98	1.94	1.89	1.86	1.82	1.80	1.75	1.74	1.73
	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.25	3.17	3.09	3.03	2.93	2.85	2.74	2.66	2.58	2.49	2.44	2.36	2.33	2.27	2.23	2.21

Reproducido del libro Statistical Methods, con gentil autorización de G. W. Snedecor y sus editores, The Iowa State College, Ames, Iowa

TABLA 5

DISTRIBUCION DE F
5% (Número superior) 1% (Número inferior)

Grados de libertad del denominador	Grados de libertad del numerador: 1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
25	4.24	3.38	2.99	2.76	2.60	2.49	2.41	2.34	2.28	2.24	2.20	2.16	2.11	2.06	2.00	1.96	1.92	1.87	1.84	1.80	1.77	1.74	1.72	1.71
	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.21	3.13	3.05	2.99	2.89	2.81	2.70	2.62	2.54	2.45	2.40	2.32	2.29	2.23	2.19	2.17
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.10	2.05	1.99	1.95	1.90	1.85	1.82	1.78	1.76	1.72	1.70	1.69
	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.17	3.09	3.02	2.96	2.86	2.77	2.66	2.58	2.50	2.41	2.36	2.28	2.25	2.19	2.15	2.13
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.30	2.25	2.20	2.16	2.13	2.08	2.03	1.97	1.93	1.88	1.84	1.80	1.76	1.74	1.71	1.68	1.67
	7.68	5.49	4.60	4.11	3.79	3.56	3.39	3.26	3.14	3.06	2.98	2.93	2.83	2.74	2.63	2.55	2.47	2.38	2.33	2.25	2.21	2.16	2.12	2.10
28	4.20	3.34	2.95	2.71	2.56	2.44	2.36	2.29	2.24	2.19	2.15	2.12	2.06	2.02	1.96	1.91	1.87	1.81	1.78	1.75	1.72	1.69	1.67	1.65
	7.64	5.45	4.57	4.07	3.76	3.53	3.36	3.23	3.11	3.03	2.95	2.90	2.80	2.71	2.60	2.52	2.44	2.35	2.30	2.22	2.18	2.13	2.09	2.06
29	4.18	3.33	2.93	2.70	2.54	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.05	2.00	1.94	1.90	1.85	1.80	1.77	1.73	1.71	1.68	1.65	1.64
	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.08	3.00	2.92	2.87	2.77	2.68	2.57	2.49	2.41	2.32	2.27	2.19	2.15	2.10	2.06	2.03
30	4.17	3.32	2.92	2.69	2.53	2.42	2.34	2.27	2.21	2.16	2.12	2.09	2.04	1.99	1.93	1.89	1.84	1.79	1.76	1.72	1.69	1.66	1.64	1.62
	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.06	2.98	2.90	2.84	2.74	2.66	2.55	2.47	2.38	2.29	2.24	2.16	2.13	2.07	2.03	2.01
32	4.15	3.30	2.00	2.67	2.51	2.40	2.32	2.25	2.19	2.14	2.10	2.07	2.02	1.97	1.91	1.86	1.82	1.76	1.74	1.69	1.67	1.64	1.61	1.59
	7.50	5.34	4.46	3.97	3.66	3.42	3.25	3.12	3.01	2.94	2.86	2.80	2.70	2.62	2.51	2.43	2.34	2.25	2.20	2.12	2.08	2.02	1.98	1.96
34	4.13	3.28	2.88	2.65	2.49	2.38	2.30	2.23	2.17	2.12	2.08	2.05	2.00	1.95	1.89	1.84	1.80	1.74	1.71	1.67	1.64	1.61	1.59	1.57
	7.44	5.29	4.42	3.93	3.61	3.38	3.21	3.08	2.97	2.89	2.82	2.76	2.66	2.58	2.47	2.38	2.30	2.21	2.15	2.08	2.04	1.98	1.94	1.91
36	4.11	3.26	2.86	2.63	2.48	2.36	2.28	2.21	2.15	2.10	2.06	2.03	1.98	1.93	1.87	1.82	1.78	1.72	1.69	1.65	1.62	1.59	1.56	1.55
	7.39	5.25	4.38	3.89	3.58	3.35	3.18	3.04	2.94	2.86	2.78	2.72	2.62	2.54	2.43	2.35	2.26	2.17	2.12	2.04	2.00	1.94	1.90	1.87
38	4.10	3.25	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02	1.96	1.92	1.85	1.80	1.76	1.71	1.67	1.63	1.60	1.57	1.54	1.53
	7.35	5.21	4.34	3.86	3.54	3.52	3.15	3.02	2.91	2.82	2.75	2.69	2.59	2.51	2.40	2.32	2.22	2.14	2.08	2.00	1.97	1.90	1.86	1.84
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.07	2.04	2.00	1.95	1.90	1.84	1.79	1.74	1.69	1.66	1.61	1.59	1.55	1.53	1.51
	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.88	2.80	2.73	2.66	2.56	2.49	2.37	2.29	2.20	2.11	2.05	1.97	1.94	1.88	1.84	1.81
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.02	1.99	1.94	1.89	1.82	1.78	1.73	1.68	1.64	1.60	1.57	1.54	1.51	1.49
	7.27	5.15	4.29	3.80	3.49	3.26	3.10	2.96	2.86	2.77	2.70	2.64	2.54	2.45	2.35	2.26	2.17	2.08	2.02	1.94	1.91	1.85	1.80	1.78
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98	1.92	1.88	1.81	1.76	1.72	1.66	1.63	1.58	1.56	1.52	1.50	1.48
	7.24	5.12	4.26	3.78	3.46	3.24	3.07	2.94	2.84	2.75	2.68	3.62	2.52	2.44	2.32	2.24	2.15	2.06	2.00	1.91	1.86	1.82	1.78	1.75
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.14	2.09	2.04	2.00	1.97	1.91	1.87	1.80	1.75	1.71	1.65	1.62	1.57	1.54	1.51	1.48	1.46
	7.21	5.10	4.24	3.76	3.44	3.22	3.05	2.98	2.82	2.73	2.65	2.60	2.50	2.42	2.30	2.22	2.18	2.04	1.98	1.90	1.86	1.80	1.76	1.72

Reproducido del libro Statistical Methods, con gentil autorización de G. W. Snedecor y sus editores. The Iowa State Coll...

TABLA 5

DISTRIBUCION DE F (Número superior) 1% (Número inferior)

Grados de libertad del denominador	Grados de libertad del numerador 1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
48	4.04	3.19	2.80	2.56	2.41	2.30	2.21	2.14	2.08	2.03	1.99	1.96	1.90	1.86	1.79	1.74	1.70	1.64	1.61	1.56	1.53	1.50	1.47	1.45
	7.19	5.08	4.22	3.74	3.42	3.20	3.04	2.90	2.80	2.71	2.64	2.58	2.48	2.40	2.28	2.20	2.11	2.02	1.95	1.83	1.84	1.78	1.73	1.70
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.02	1.98	1.95	1.90	1.85	1.78	1.74	1.69	1.63	1.60	1.55	1.52	1.48	1.46	1.44
	7.17	5.06	4.20	3.72	3.41	3.18	3.02	2.88	2.78	2.70	2.62	2.56	2.46	2.39	3.26	2.18	2.10	2.00	1.94	1.86	1.82	1.76	1.71	1.68
55	4.02	3.17	2.78	2.54	2.38	2.27	2.18	2.11	2.05	2.00	1.97	1.93	1.88	1.83	1.76	1.72	1.67	1.61	1.58	1.52	1.50	1.46	1.43	1.41
	7.12	5.01	4.16	3.68	3.37	3.15	2.98	2.35	2.75	2.66	2.59	2.53	2.43	2.35	2.23	2.15	2.06	1.96	1.90	1.82	1.78	1.71	1.66	1.64
60	4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.86	1.81	1.75	1.70	1.65	1.59	1.56	1.50	1.48	1.44	1.41	1.39
	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.40	2.32	2.20	2.12	2.03	1.93	1.87	1.79	1.74	1.68	1.63	1.60
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.02	1.98	1.94	1.90	1.85	1.80	1.73	1.68	1.63	1.57	1.54	1.49	1.46	1.42	1.39	1.37
	7.04	4.95	4.10	3.62	3.31	3.09	2.93	2.79	2.70	2.61	2.54	2.47	2.37	2.30	2.18	2.09	2.00	1.90	1.84	1,76	1.71	1.64	1.60	1.56
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.01	1.97	1.93	1.89	1.84	1.79	1.72	1.67	1.62	1.56	1.53	1.47	1.45	1.40	1.37	1.35
	7.01	4.92	4.08	3.60	3.29	3.07	2.91	2.77	2.67	2.59	2.51	2.45	2.35	2.28	2.15	2.07	1.98	1.88	1.82	1.74	1.69	1.62	1.56	1.53
80	3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	1.99	1.95	1.91	1.88	1.82	1.77	1.70	1.65	1.60	1.54	1.51	1.45	1.42	1.38	1.35	1.32
	6.96	4.88	4.04	3.56	3.25	3.04	2.87	2.74	2.64	2.55	2.48	2.41	2.32	2.24	2.11	2.03	1.94	1.84	1.78	1.70	1.65	1.57	1.52	1.49
100	3.94	3.09	2.70	2.46	2.30	2.19	2.10	2.03	1.97	1.92	1.88	1.85	1.79	1.75	1.68	1.63	1.57	1.51	1.43	1.42	1.39	1.34	1.30	1.28
	6.90	4.82	3.98	3.51	3.20	2.99	2.82	2.69	2.59	2.51	2.43	2.36	2.26	2.19	2.06	1.98	1.89	1.79	1.73	1.64	1.59	1.51	1.46	1.43
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.95	1.90	1.86	1.83	1.77	1.72	1.65	1.60	1.55	1.49	1.45	1.39	1.36	1.31	1.27	1.25
	6.84	4.78	3.94	3.47	3.17	2.95	2.79	2.65	2.56	2.47	2.40	2.33	2.23	2.15	2.03	1.94	1.85	1.75	1.68	1.59	1.54	1.46	1.40	1.37
150	3.91	3.06	2.67	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.82	1.76	1.71	1.64	1.59	1.54	1.47	1.44	1.37	1.34	1.29	1.25	1.22
	6.81	4.75	3.91	3.44	3.14	2.92	2.76	2.62	2.53	2.44	2.37	2.30	2.20	2.12	2.00	1.91	1.83	1.72	1.66	1.56	1.51	1.43	1.37	1.33
200	3.89	3.04	2.65	2.41	2.26	2.14	2.05	1.98	1.92	1.87	1.83	1.80	1.74	1.69	1.62	1.57	1.52	1.45	1.42	1.35	1.32	1.26	1.22	1.19
	6.76	4.71	3.88	3.41	3.11	2.90	2.73	2.60	2.50	2.41	2.34	2.28	2.17	2.09	1.97	1.88	1.79	1.69	1.62	1.53	1.48	1.39	1.33	1.28
400	3.86	3.02	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.81	1.78	1.72	1.67	1.60	1.54	1.49	1.42	1.38	1.32	1.28	1.22	1.16	1.13
	6.70	4.66	3.83	3.36	3.06	2.85	2.69	2.55	2.46	2.37	2.29	2.23	2.12	2.04	1.92	1.84	1.74	1.64	1.57	1.47	1.42	1.32	1.24	1.19
1000	3.85	3.00	2.61	2.38	2.22	2.10	2.02	1.95	1.89	1.84	1.80	1.76	1.70	1.65	1.58	1.53	1.47	1.41	1.36	1.30	1.26	1.19	1.13	1.08
	6.66	4.62	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.31	2.26	2.20	2.09	2.01	1.89	1.81	1.71	1.61	1.54	1.44	1.38	1.28	1.19	1.11
	3.84	2.99	2.60	2.37	2.21	2.09	2.01	1.94	1.88	1.83	1.79	1.75	1.69	1.64	1.57	1.52	1.46	1.40	1.35	1.28	1.24	1.17	1.11	1.00
	6.64	4.60	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.24	2.18	2.07	1.99	1.87	1.79	1.69	1.59	1.52	1.41	1.36	1.25	1.15	1.00

Reproducido del libro Statistical Methods, con gentil autorización de G. W. Snedecor y sus editores, The Iowa State College, Ames Iowa

INFERENCIA ESTADISTICA

Introducción.

Muchos de los problemas de Estadística consisten en estimar algunos de los parámetros que caracterizan a la población, principalmente la media y la variancia, y además existen hipótesis referentes a estos parámetros que es necesario confirmar.

Si fuera posible conocer toda la población, por simples procedimientos matemáticos, podemos determinar los parámetros, pero las disponibilidades presupuestales ó el tiempo disponible ó dificultades materiales pueden hacer imposible conocer toda la población y consecuentemente sus características.

Ante esta situación el procedimiento recomendado a seguir es el seleccionar una muestra de la población y calcular en ella los valores estadísticos, que serán utilizados para decir algo referente a los parámetros de la población, lo que constituye la ESTIMACION.

En el caso en que queremos conocer los parámetros μ y σ^2 de una población, seleccionaremos una muestra y en ella calcularemos: $\bar{x}$ y s^2. Al obtener estos dos valores estadísticos $\bar{x}$ y s^2 sentiremos la tentación de pensar que corresponden a μ y σ^2, y diremos que $\bar{x}$ es un estimado de μ y s^2 es un estimado de σ^2 ¿Cuán adecuados son estos estimados? Para contestar esta pregunta recurriremos a los conceptos de Esperanza matemática y al de Variancia mínima que constituyen requisitos de un buen estimado.

Para que un estimado sea considerado el mejor, es necesario que el promedio de todos los valores posibles que puede tomar, sea igual al parámetro, es decir:

$$E(\bar{x}) = \mu \qquad y \qquad E(s^2) = \sigma^2$$

Entonces, decimos que el estimado NO ES VICIADO. Además de la condición mencionada es necesario que el estimado tenga VARIANCIA MINIMA. Consideremos dos estimados de μ: la media ($\bar{x}$) y la mediana de la muestra. La distribución de medias tiene la propiedad de tener menor variancia que la distribución de medianas de muestras, por esta razón decimos que la media de la muestra ($\bar{x}$) es un mejor estimado de μ.

Con los valores estadísticos calculados en la muestra es posible establecer conclusiones acerca de los parámetros que caracterizan a la población, expresando en probabilidades la veracidad de nuestras conclusiones lo cual constituye la INFERENCIA ESTADISTICA.

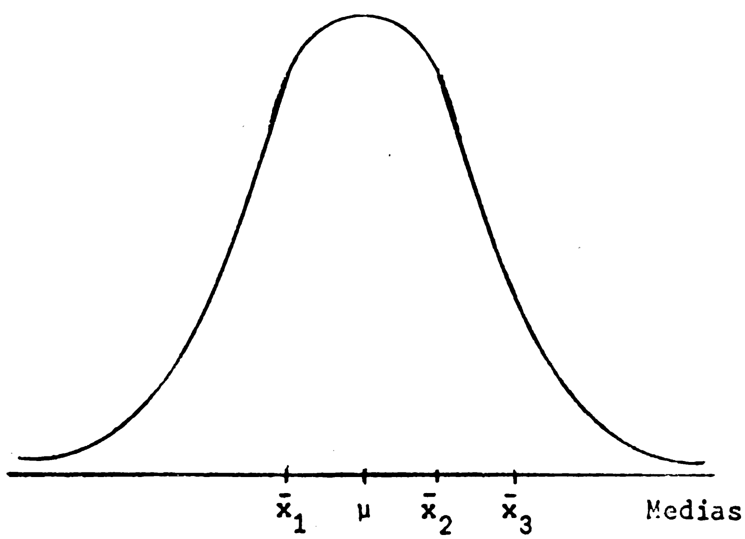
Estimación puntual y Estimación de intervalo.

Como $\bar{x}$ y s^2 se determinan en una muestra sería realmente sorprendente que sus valores sean exactamente iguales a los respectivos parámetros, y más bien nos inclinamos a pensar que sus valores son muy próximos a ellos. Como puede verse para el caso de $\bar{x}$ en el Dibujo 10.1

Como $\bar{x}_1$ fija un punto, decimos que $\bar{x}$ es un estimado puntual de μ.

Como nuestro interés consiste en obtener información respecto a μ, podemos pensar que existiendo escasas probabilidades de que

DISTRIBUCION DE MEDIAS



$\bar{x}$ sea igual a μ, es más aconsejable dar dos valores que establecen un intervalo y expresar el grado de confianza de que el parámetro se encuentre en ese intervalo en probabilidades.

Al explicar la distribución de z, hemos visto que conocidos los parámetros μ y σ^2 podemos hallar la probabilidad de obtener al azar una media ($\bar{x}$) en un intervalo dado. Pero si sólo conocemos los estimados $\bar{x}$ y s^2 ¿Cómo podemos establecer un intervalo en el que se encuentre el parámetro y cuál será la probabilidad de que ello ocurra? Para resolver esta pregunta, recordemos que en la distribución de t establecimos:

$$p\left[-t_o \leq t \leq t_o\right] = 1 - 0.05$$

donde t_o es el valor hallado en la Tabla 3 para los correspondientes grados de libertad con una probabilidad de 0.05. En la ecuación anterior reemplazamos el valor de t por

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}}$$

$$P\left[-t_o \leq \frac{\bar{x} - \mu}{s_{\bar{x}}} \leq t_o\right] = 0.95$$ y finalmente,

$$P\left[\bar{x} - t_o\; s_{\bar{x}}\; s_{\bar{x}} \leq \mu \leq \bar{x} + t_o\; s_{\bar{x}}\right] = 0.95$$

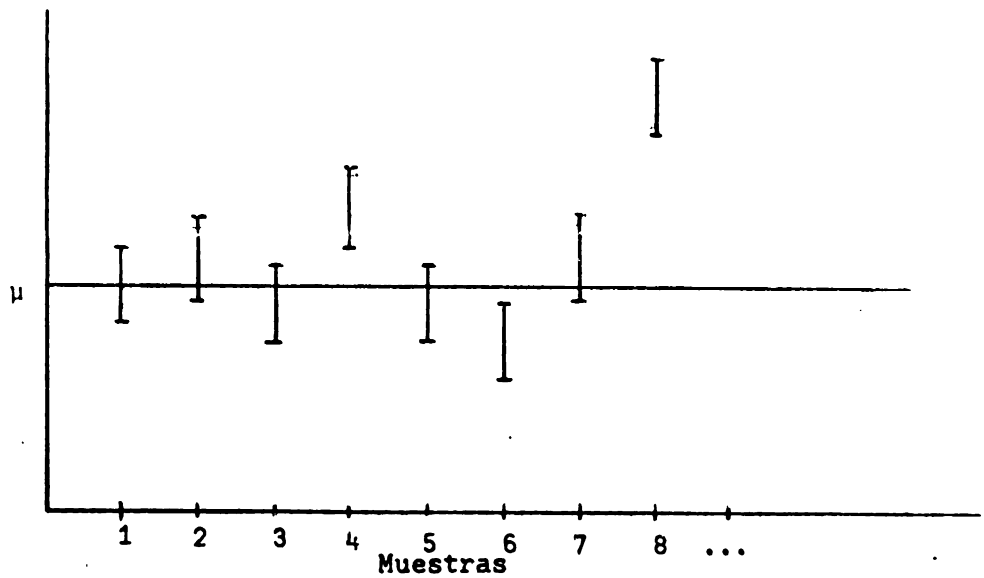
Los límites: $\bar{x} \pm t_o\, s_{\bar{x}}$ constituyen los valores de la estimación de intervalo ó INTERVALO DE CONFIANZA. La probabilidad expresada en estos límites de confianza debe interpretarse de que existe una probabilidad de $\frac{95}{100}$ de encontrar al azar una muestra cuyos limites encierran al parámetro μ de la población. En otras palabras podría decirse que de 100 muestras escogidas al azar, 95 darán intervalos de confianza que encierran al verdadero parámetro de la población y 5 muestras darán intervalos de confianza que no encierran al parámetro μ de la población. El Dibujo 10.2 ilustra el concepto de intervalo de confianza.

Hipótesis.

Una hipótesis es una suposición que puede estar basada en observaciones preliminares ó ser recién propuesta.

La veracidad de la hipótesis será necesaria comprobarla a través de una nueva y objetiva colección de observaciones, las que

INTERVALO DE CONFIANZA



constituyen la muestra y el procedimiento estadístico de comprobación constituye la PRUEBA DE HIPOTESIS.

Si la hipótesis y la realidad que nos da la muestra no están de acuerdo, la hipótesis es descartada o rechazada. En el caso de que la hipótesis está de acuerdo con la realidad determinada en la muestra, la hipótesis es aceptada.

El ejemplo referente al tratamiento de una enfermedad resulta muy ilustrativo para la prueba de hipótesis. A un hospital acude un paciente para recibir tratamiento de una dolencia, el médico que lo atiende luego de escuchar los síntomas del paciente es de opinión de que el mal que le aqueja es una infección renal, para verificar su diagnóstico recomienda al paciente una serie de análisis, y con los resultados de estos análisis el médico determinará cuan acertado estuvo en su diagnóstico.

Del análisis del problema podemos decir que existe una hipótesis planteada por el médico (Hp), y es que el paciente sufre de infección renal. Existirá una hipótesis alternante (Ha), y es que el paciente no sufre de infección renal. Para el efecto, el médico prescribe al paciente una serie de análisis, que constituye la muestra. Por sus estudios universitarios el médico sabe cuáles son los límites de los análisis, entre los cuales se considera al paciente con infección renal. Si los resultados del análisis del paciente se encuentran entre estos límites, el médico aceptará su diagnóstico (Hp). Estos límites reciben el nombre de zona de aceptación de la hipótesis planteada. Si los resultados del análisis se encuentran fuera de estos límites el médico rechazará su diagnóstico.

Tipos de Errores.

Cualquiera que sea la decisión tomada a partir de una prueba de hipótesis, ya sea de aceptación de la Hipótesis planteada o de la Hipótesis alternante puede incurrirse en error, como puede verse en el siguiente Cuadro.

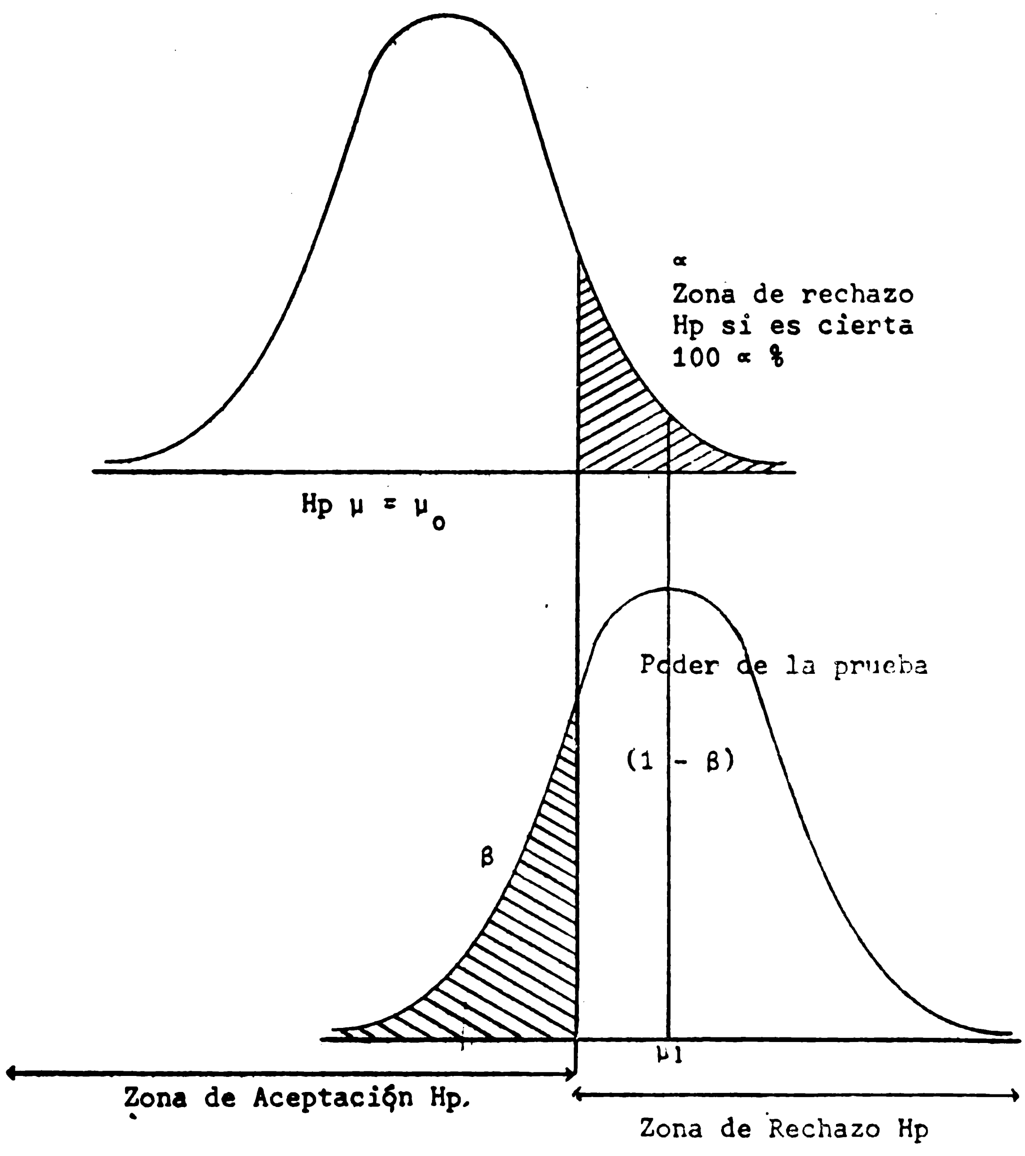
En la Población	La muestra seleccionada conduce a	
	Aceptar Hp	Rechazar Hp
Hp es cierta	No hay error	Error Tipo I ó α
Hp es falsa	Error Tipo II ó β	No hay Error

Es necesario disminuír la probabilidad de cometer cualquiera de los Errores Tipo I y II.

Utilicemos el Dibujo 10.3 para explicar estos errores.

DIBUJO 10.3

TIPOS DE ERRORES



Si la Hipótesis planteada, Hp: $\mu = \mu_0$, es cierta la zona de rechazo, α, medirá la probabilidad de que se rechace dicha hipótesis siendo cierta, incurriendo en Error Tipo I ó α.

Supongamos que la Hipótesis planteada es falsa, Hp: $\mu = \mu_0$, y que la alternante Ha: $\mu = \mu_1$ es verdadera y si los resultados de la muestra nos conducen a aceptar la Hipótesis planteada, estamos cometiendo el Error Tipo II ó β.

La magnitud del Error β depende de la magnitud del error α y de la discrepancia entre μ_0 y μ_1. En el Dibujo 10.3 puede verse la existencia de una relación inversa entre la magnitud de los errores α y β, conforme α aumenta β disminuye. Esto nos obliga a establecer con cuidado el valor de α para las pruebas estadísticas. Lo ideal sería establecer α y β. En la práctica se establece el nivel α y para disminuir el Error β se incrementa el número de observaciones en la muestra, pues así se acortan los límites de confianza respecto a la Hipótesis planteada.

La meta de las pruebas estadísticas es rechazar la hipótesis planteada. En otras palabras es deseable aumentar la probabilidad de aceptar la hipótesis alternante cuando ésta es verdadera ó sea incrementar lo que se llama el PODER DE LA PRUEBA $(1 - \beta)$. La aceptación de la Hipótesis planteada debe interpretarse como que la información disponible no permite detectar la falsedad de esta Hipótesis.

Nivel de significación.

La probabilidad de cometer el Error Tipo I ó α recibe el nombre de NIVEL DE SIGNIFICACION. El nivel de significación que se usa depende del investigador, por lo general se usa el nivel de $\alpha = 0.05$ y $\alpha = 0.01$.

La selección del nivel de significación, α, dependerá del riesgo que desee correr el investigador en el rechazo de su hipótesis planteada.

Procedimiento de pruebas de hipótesis.

El procedimiento general para realizar las pruebas de hipótesis es el siguiente:

1° En base a la experiencia y para una población o distribución se elabora una hipótesis razonable para lo cual se establece una prueba estadística. La hipótesis recibe el nombre de Hipótesis planteada, y en algunos textos llamada Hipótesis nula. Al establecer ésta será necesario establecer una Hipótesis alternante que será aceptada al rechazar la Hipótesis planteada.

2° Escoger un nivel de probabilidad ó significación llamado α (Error Tipo I). Para ilustrar la importancia en el establecimiento de la probabilidad de cometer este tipo de error se dá el siguiente ejemplo: Si un investigador por sus estudios preliminares está casi seguro que Hp es cierta, deseará disminuir la probabilidad de rechazarla, es decir reducirá el error α ó Tipo I. Un criterio usado frecuentemente es establecer que los resultados son significativos si la hipótesis planteada es rechazada con $\alpha = 0.05$ y altamente significativa si es rechazada con $\alpha = 0.01$.

Los pasos 1 y 2 corresponden al planeamiento de las investigaciones.

3° Determinar la prueba estadística que se va a realizar y las asunciones necesarias a esta prueba para que sea válida.

4° Establecer las regiones de aceptación y de rechazo de la hipótesis planteada. Las regiones de rechazo recibe el nombre de regiones críticas.

5° Realizar la prueba estadística. Si el resultado de la prueba estadística se encuentra en la región crítica se rechazará la Hp aceptando la alternante, en caso contrario la Hp será aceptada.

Si se rechaza la Hp se piensa que se tiene una razón definida a no creer en la hipótesis, mientras que si se acepta la Hp se dice que no hay suficiente evidencia para dudar de ella.

El rechazo de la Hp es muy exigente, y este rechazo puede deberse a que la hipótesis planteada no es cierta ó a que la muestra obtenida y utilizada para realizar la prueba estadística pertenece al grupo de muestras que nos van a hacer cometer el Error Tipo I ó α. Sin embargo, como en la selección de la muestra hemos tomado especial cuidado de que ésta sea adecuada concluíremos de que se rechaza la hipótesis planteada debido a que es incorrecta.

Las principales pruebas de hipótesis que trataremos son las siguientes:

1. Referente a la media de una población, μ

2. Referente a la igualdad de 2 medias, $\mu_1 = \mu_2$

3. Referente a la variancia de una población, σ^2

4. Referente a la homogeneidad de 2 variancias, $\sigma_1^2 = \sigma_2^2$

<u>Prueba de hipótesis de la media de una población</u>:

Muchos problemas de estadística consisten en probar una hipótesis respecto a la media, μ, y para el efecto se selecciona una muestra de la población, en base a ella se verifica ó rechaza la Hp siguiendo los pasos establecidos para la prueba de hipótesis.

Si en una población se plantea la hipótesis que la media de la población es μ_o y no existe información acerca de que $\mu > \mu_o$ ó $\mu < \mu_o$ estableceremos la hipótesis planteada y alternante en la siguiente forma:

Hp: $\mu = \mu_o$		Hp: $\mu = \mu_o$
	equivalente	
Ha: $\mu \neq \mu_o$		Ha: $\mu < \mu_o$ ó $\mu > \mu_o$

La prueba estadística de esta hipótesis será estableciendo dos regiones críticas, ésto es lo que se llama la prueba de dos colas ó prueba bilateral.

Existen otra clase de problemas prácticos en el que se tiene de que la media no es igual al valor postulado en la hipótesis, y entonces la alternante podrá ser que la media sea mayor que el valor postulado ó menor. Las hipótesis serán las siguientes:

Aquí se tiene que establecer una región crítica y es por ésto que se llama prueba de una cola ó prueba unilateral.

Para la prueba de hipótesis de la media cuando σ^2 es desconocido se usa la prueba de t. Los siguientes ejemplos ilustran la prueba de hipótesis referente a la media de la población.

Ejemplo 1. Prueba bilateral ó de dos colas.

El gerente de una fábrica de pinturas está interesado en conocer cual es la superficie promedio que se cubre con un litro de pintura. Según los técnicos de la fábrica, un litro de pintura debe alcanzar para 20 m^2. Para confirmar estos informes el gerente lleva a cabo un experimento con 10 pintores, realizado el trabajo se mide en metros cuadrados la superficie pintada por cada uno de ellos, que son: 17, 15, 16, 20, 22, 24, 28, 28, 16, 20.

Desarrollo:

1. Hipótesis: La hipótesis planteada ó nula es que la media de la población es igual a 20. La hipótesis alternante es que la media de la población es diferente a 20 porque no hay ninguna razón para creer que el rendimiento de la pintura sea inferior ó superior a 20 m^2. Luego:

   Hp: $\mu = 20$

   Ha: $\mu \neq 20$

2. Nivel de significación: Este es escogido arbitrariamente por el experimentador al momento de planear el experimento. Sea $\propto = 0.05$ el nivel de probabilidades escogido.

3. Determinar la prueba estadística y sus asunciones. En este caso es la prueba de "t"

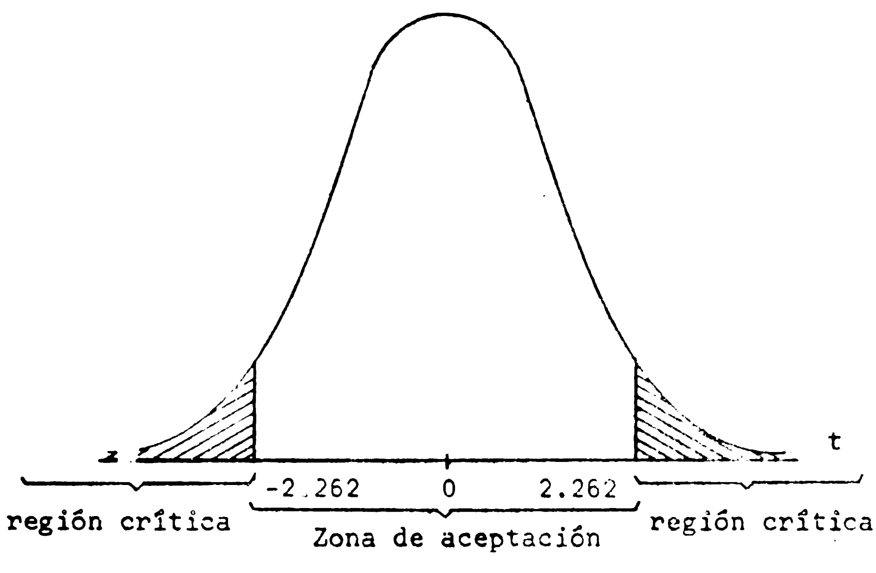
$$t = \frac{\bar{x} - \mu_o}{\sqrt{\frac{s^2}{n}}}$$

Las asunciones son:

a) La población es normal

b) La muestra ha sido extraída al azar

4. Establecer la región crítica: Las regiones críticas para $\alpha = 0.05$ con 9 grados de libertad son: $t < -2.262$ y $t > 2.262$



5. Realizar la prueba estadística:

$\mu_o = 20$

$n = 10$

$\Sigma X = 206$

$\bar{x} = 20.6$

$$\Sigma X^2 = 4456$$

$$(\Sigma X)^2/n = 4243.6$$

$$\Sigma X^2 - \frac{(\Sigma X)^2}{n} = S.C. = 210.4$$

$$s^2 = \frac{S.C.}{n-1} = 23.3$$

$$t = \frac{\bar{x} - \mu_o}{\sqrt{\frac{s^2}{n}}} = \frac{20.6 - 20}{\sqrt{\frac{23.3}{10}}} = \frac{0.6}{1.53} = 0.392$$

Conclusión: Como 0.392 se encuentra entre -2.262 y 2.262 ó sea en la zona de aceptación, la conclusión es que la media de la población es 20. Luego se puede decir que un litro de pintura cubre 20 m^2 de superficie.

Ejemplo 2. Prueba unilateral ó de una sola cola.

En una ciudad el tamaño promedio de los niños al nacer es de 0.40 m. Un médico sustenta que si se suministra una dieta a la madre en gestación, este promedio se incrementará. El cree que las probabilidades de obtener este incremento es de 0.95. Para probar esta hipótesis se escoge al azar a 15 madres, a las cuales se les suministró la dieta. Los resultados en metros del tamaño de los niños al nacer fueron: 0.45, 0.53, 0.68, 0.40, 0.38, 0.58, 0.61, 0.53, 0.40, 0.45, 0.54, 0.56, 0.55, 0.48, 0.51.

Desarrollo

1. Hipótesis: La hipótesis nula es que la media de la población es 0.40 m. La alternante es que la media de la población sea mayor de 0.40 m., ya que mediante la

dieta existe la hipótesis que la media aumentará.

Entonces:

Hp: $\mu = 0.40$

Ha: $\mu > 0.40$

2. El nivel de significación será $\alpha = 0.05$.

3. La prueba estadística es la de t:

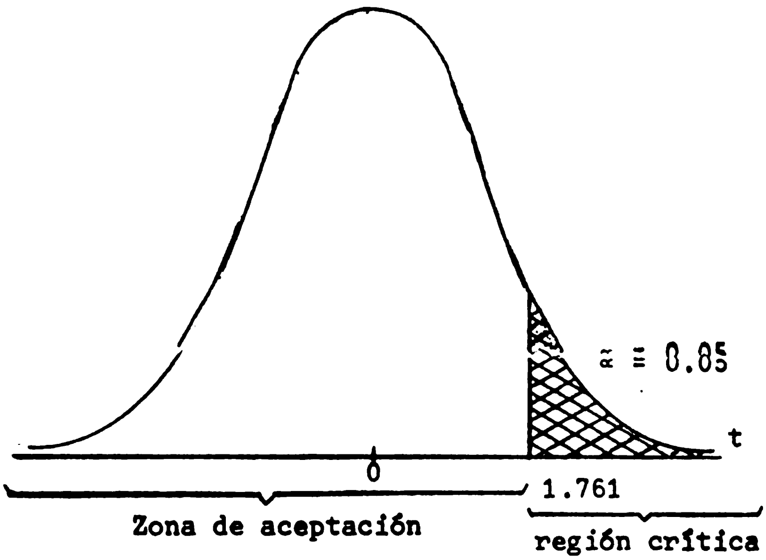
$$t = \frac{\bar{x} - \mu_o}{\sqrt{s^2/n}}$$

y sus asunciones son:

a) Población es normalmente distribuída

b) La muestra ha sido extraída al azar

4. Regiones críticas: En este caso, la hipótesis alternante es que la media de la población es mayor de 0.40 y la región crítica se encontrará a la derecha de la curva. Además, como la Tabla 3 es de dos colas, las probabilidades dadas en esta son la suma de las probabilidades de las dos colas. Luego si queremos la probabilidad para una sola cola, debemos doblar el nivel de probabilidades, es decir, hallar el valor tabular con 2α.
En este caso en que tenemos $\alpha = 0.05$, $2\alpha = 2\ (0.05) = 0.10$ y para 14 grados de libertad el valor de t es 1.761.
Luego la región crítica será $t > 1.761$.



5. Realizar la prueba estadística: Esto es

$\mu_o = 0.40$

$n = 15$

$\Sigma X = 7.65$

$\bar{x} = 0.51$

$\Sigma X^2 = 4.0003$

$(\Sigma X)^2/n = 3.9015$

$S.C. = 0.0988$

$$s^2 = \frac{S.C.}{n-1} = \frac{0.0988}{14} = 0.007057$$

$$t = \frac{\bar{x} - \mu_o}{\sqrt{\frac{s^2}{n}}} = \frac{0.51 - 0.40}{\sqrt{\frac{0.007057}{15}}} = \frac{0.11}{0.0217} = 5.07$$

Conclusión: Como 5.07 se encuentra en la zona de rechazo ó sea es superior a 1.761 rechazaremos la hipótesis nula ó sea que la media de la población es 0.40 m. y aceptamos la alternante ó sea que la media de la población es mayor que 0.40 m.

Prueba de hipótesis concerniente a las medias de dos poblaciones.

A menudo tenemos 2 muestras, cada una proveniente de una población y deseamos conocer si las medias de las dos poblaciones difieren entre sí. Los subíndices 1 y 2 serán usados para valores de μ, σ^2, $\bar{x}$, s^2 y n para indicar la población a la que pertenecen.

Si una muestra al azar de n_1 observaciones es extraída de una población con media μ_1 y variancia σ_1^2 y si otra muestra de n_2 observaciones es extraída de una población con media μ_2 y variancia σ_2^2, podemos hallar la siguiente diferencia $d_1 = (\bar{x}_1 - \bar{x}_2)$, y si igual procedimiento lo realizamos repetidas veces habremos generado una DISTRIBUCION DERIVADA DE DIFERENCIA DE MEDIAS DE MUESTRAS. Los parámetros de esta distribución son:

$$\mu_d = \mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$$

$$\sigma_d^2 = \sigma^2_{\bar{x}_1 - \bar{x}_2} = \sigma^2_{\bar{x}_1} + \sigma^2_{\bar{x}_2}$$

$$= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Si las poblaciones muestreadas son normales la distribución derivada de diferencia de medias de muestras también será normal.

Aplicando la definición de z para la población derivada de diferencia de medias de muestras se tiene:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Prueba de hipótesis concerniente a las medias de dos poblaciones.

A menudo tenemos 2 muestras, cada una proveniente de una población y deseamos conocer si las medias de las dos poblaciones difieren entre sí. Los subíndices 1 y 2 serán usados para valores de μ, s^2, $\bar{x}$, s^2 y n para indicar la población a la que pertenecen.

Si una muestra al azar de n_1 observaciones es extraída de una población con media μ_1 y variancia σ_1^2 y si otra muestra de n_2 observaciones es extraída de una población con media μ_2 y variancia σ_2^2, podemos hallar la siguiente diferencia $d_1 = (\bar{x}_1 - \bar{x}_2)$, y si igual procedimiento lo realizamos repetidas veces habremos generado una DISTRIBUCION DERIVADA DE DIFERENCIA DE MEDIAS DE MUESTRAS. Los parámetros de esta distribución son:

$$\mu_d = \mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$$

$$\sigma_d^2 = \sigma^2_{\bar{x}_1 - \bar{x}_2} = \sigma^2_{\bar{x}_1} + \sigma^2_{\bar{x}_2}$$

$$= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Si las poblaciones muestreadas son normales la distribución derivada de diferencia de medias de muestras también será normal.

Aplicando la definición de z para la población derivada de diferencia de medias de muestras se tiene:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Si las poblaciones no son normales pero las muestras son bastante grandes, podemos aplicar el teorema de límite central y tendremos que, la distribución derivada de diferencia de medias de muestras sigue la distribución normal.

La hipótesis planteada usual en estos casos es de que las medias de las dos poblaciones son iguales y se pueden expresar como:

Hp: $\mu_1 = \mu_2$ ó $\mu_1 - \mu_2 = 0$

pero no siempre la hipótesis planteada es de que la diferencia de las dos medias sea cero, ya que la hipótesis puede ser que la diferencia de dos medias es igual a un número dado. Ejemplo $\mu_1 - \mu_2 = 20$.

La hipótesis alternante de una y dos colas se ilustra en los ejemplos.

Para comparar las medias de las poblaciones μ_1 y μ_2, se toman muestras de n_1 observaciones de la primera población y n_2 observaciones de la segunda población y se calculan sus respectivas medias de muestra $\bar{x}_1$ y $\bar{x}_2$ y usamos los valores estadísticos $\bar{x}_1 - \bar{x}_2$ para probar la hipótesis.

La prueba estadística a usarse para probar la hipótesis $\mu_1 = \mu_2$, cuando las variancias σ_1^2 y σ_2^2 son desconocidas es la de t. En dicha prueba existen dos casos:

Caso I. Cuando existe homogeneidad de variancias $(\sigma_1^2 = \sigma_2^2)$ y éstas son desconocidas (la homogeneidad es determinada por la prueba de F que trataremos más adelante)

La prueba de hipótesis utilizada en este caso es:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_{\bar{x}_1 - \bar{x}_2}}$$

Donde $s^2_{\bar{x}_1 - \bar{x}_2}$ es la variancia de una muestra de diferencia de medias que está definida por:

$$s^2_{\bar{x}_1 - \bar{x}_2} = s^2_p \left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

La variancia conjunta ó combinada simbolizada por s^2_p es la estimación de σ^2 y se obtiene sumando la suma de cuadrados de las dos muestras y dividiéndolas por la suma de los grados de libertad de las dos muestras:

$$s^2_p = \frac{\Sigma(X_{1i} - \bar{x}_1)^2 + \Sigma(X_{2i} - \bar{x}_2)^2}{(n_1 - 1) + (n_2 - 1)}$$

Se puede observar también que la fórmula general de variancia es:

$$s^2 = \frac{\Sigma(X - \bar{x})^2}{n - 1}$$

Luego la suma de cuadrados es igual a

$$\Sigma(X - \bar{x})^2 = (n - 1)\, s^2$$

Reemplazando las sumas de cuadrados en la ecuación de la variancia conjunta por los valores de $(n - 1)\, s^2$ en sus respectivas poblaciones, tendremos que la variancia conjunta será igual a

$$s^2_p = \frac{(n_1 - 1)\, s^2_1 + (n_2 - 1)\, s^2_2}{n_1 + n_2 - 2}$$

Note que en esta ecuación s^2_p es el promedio ponderado de s^2_1 y s^2_2.

Caso II. Cuando no existe homogeneidad de variancias ($\sigma_1^2 \neq \sigma_2^2$) y estas son desconocidas (la no homogeneidad de variancias es determinada por la prueba de F que trataremos más adelante), se usa como prueba estadística para la comparación de promedios la ecuación:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad (10.8.1)$$

Al no existir homogeneidad de variancias la ecuación anterior no sigue una distribución de t y no se puede determinar las regiones críticas, para establecer la significación estadística de la hipótesis con los valores tabulares de t. Cochran estableció valores tabulares llamados t' para indicar significación utilizando el promedio ponderado de dos valores de t, t_1 correspondiente a $(n_1 - 1)$ y t_2 correspondiente a $(n_2 - 1)$, los factores de ponderación son

$\frac{s_1^2}{n_1}$ y $\frac{s_2^2}{n_2}$ como se muestra en la siguiente ecuación

$$t' = \frac{\frac{s_1^2}{n_1} t_1 + \frac{s_2^2}{n_2} t_2}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad (10.8.2)$$

Si ambas muestras son de igual tamaño ($n_1 = n_2 = n$), entonces

$t_1 = t_2 = t$ y la ecuación será:

$$t' = \frac{t\left(\frac{s_1^2}{n} + \frac{s_2^2}{n}\right)}{\left(\frac{s_1^2}{n} + \frac{s_2^2}{n}\right)} = t$$

donde t ha sido encontrado con (n - 1) grados de libertad. El valor de la prueba estadística t (ecuación 10.8.1) se compara con el valor de t' (ecuación 10.8.2) que establece las regiones críticas para aceptar ó rechazar la hipótesis planteada.

Ejemplos de Prueba de hipótesis concerniente a las medias de dos poblaciones, cuando existe homogeneidad de variancias:

Ejemplo 1. Con el objeto de determinar cual es la mejor dieta para el engorde de lechones se alimentó a 12 lechones con maíz opaco-2 de alto contenido de lisina y a otros 12 con maíz corriente. Los siguientes fueron los resultados de ganancia de peso en Kilos por día.

Opaco-2 (1):	.58	.59	.64	.66	.65	.66	.67	.66	.61
	.59	.68	.69						
Corriente (2):	.33	.42	.36	.31	.43	.38	.35	.32	.42
	.40	.30	.30						

<u>Desarrollo</u>.

1. Hipótesis: Hp: $\mu_1 = \mu_2$

   Ha: $\mu_1 \neq \mu_2$

2. Nivel de significación: $\alpha = 0.05$

3. La prueba estadística a usarse es la de t. Existe homogeneidad de variancias (Prueba de F)

$$t = \frac{(\bar{x}_1 - \bar{x}_2) = (\mu_1 = \mu_2)}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Como $n_1 = n_2 = n$ tendremos:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{2 s_p^2}{n}}}$$

Las asunciones para que esta prueba sea válida son:

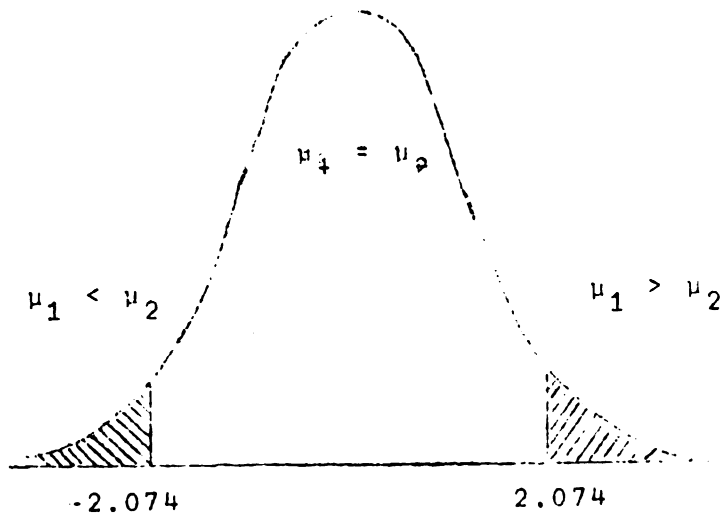
a) Las poblaciones están normalmente distribuídas

b) Las muestras son tomadas al azar independientemente.

c) La homogeneidad de variancias $(\sigma_1^2 = \sigma_2^2)$

4. Regiones críticas. Con $n_1 + n_2 - 2 = 22$ grados de libertad los valores de t que definen las regiones críticas con $\alpha = 0.05$ se encuentran en la Tabla 3

$t < -2.074$ y $t > 2.074$



5.. Realizar la prueba estadística

	Opaco-2	Corriente
n	12	12
ΣX	7.68	4.32
$\bar{x}$	.64	.36
ΣX^2	4.9310	1.5816
$(\Sigma X)^2/n$	4.9152	1.5552
S.C.	0.0158	0.0264
s^2	0.0014	0.0024

$$s_p^2 = \frac{0.0264 + 0.0158}{11 + 11} = 0.001918$$

$$t = \frac{0.64 - 0.36}{\sqrt{\frac{2(0.001918)}{12}}} = \frac{0.2800}{0.0178} = 15.73$$

Como el valor calculado 15.73 está dentro de la región crítica, rechazamos la hipótesis planteada , aceptamos la alternante de que la dieta en base a maíz opaco-2 es mejor que la de maíz corriente.

Ejemplo 2. El anuncio de una empresa en dos diarios diferentes tuvo como finalidad la mayor venta de sus productos. El gerente de la empresa queriendo saber si la propaganda en el períodico 1 le produjo mayor venta que la propaganda en el periódico 2, contrató los servicios de un estadístico. Este mediante un hábil muestreo determinó las ventas de los productos (en soles) de cada uno de los días en que aparecieron los anuncios.

Periódico	
1	2
133	15
111	104
50	35
76	31
24	18
	17

Desarrollo.

1. Hipótesis: La hipótesis es que el promedio de ventas debido a la propaganda de los dos períodicos son iguales y la alternante es de que la media de ventas, mediante la propaganda del periódico 2 es mayor que la del periódico 1. Luego:

$$Hp: \mu_1 = \mu_2 \qquad \mu_1 - \mu_2 = 0$$

ó

$$Ha: \mu_1 > \mu_2 \qquad \mu_1 - \mu_2 > 0$$

5. Calcular la prueba estadística

	Periódico	
	1	2
ΣX	394	220
$\bar{x}$	78.80	36.67
ΣX^2	38862	13840.00
$(\Sigma X)^2/n$	31047.2	8066.67
S.C.	7814.8	5773.33
s^2	1953.7	1154.67

$$s_p^2 = \frac{7814.8 + 5773.33}{9} = 1509.79 \quad \text{luego}$$

$$t = \frac{98.80 - 36.67}{\sqrt{1509.79(\frac{1}{6}+\frac{1}{5})}} = 2.642$$

Concluimos rechazando la hipótesis planteada y aceptamos que la publicidad en el periódico 1 produce más ventas que en el periódico 2.

Prueba de hipótesis concerniente a las medias de dos poblaciones cuando no existe homogeneidad de variancia.

Ejemplo 3. En una empresa se tramitan expedientes a diario. Existe en estudio un nuevo horario de trabajo. El gerente de la empresa desea saber si el nuevo horario es más eficiente ó menos eficiente que el antíguo y decide hacer una investigación al respecto. Para el efecto toma 20 días escogidos al azar y determina el número de expedientes tramitados por día en ambos horarios (los datos han sido redondeados a enteros). Los resultados obtenidos son los siguientes:

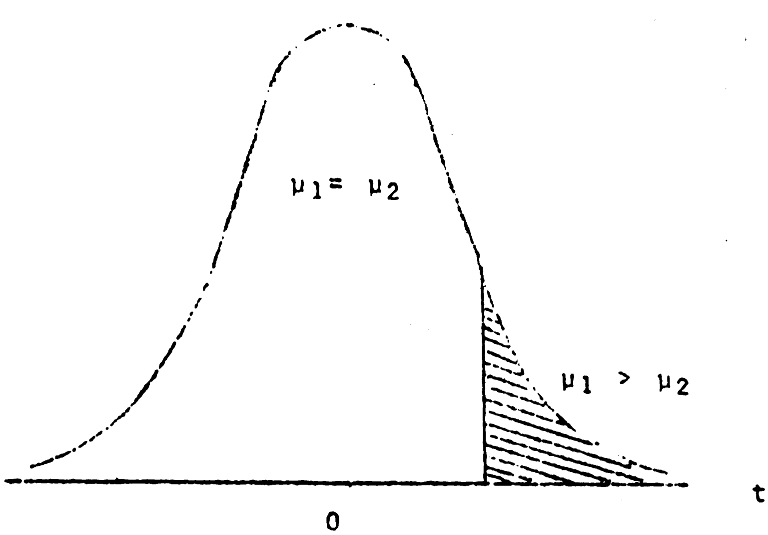
2. El gerente quiere correr un riesgo de 5%. Luego $\alpha = 0.05$.
3. La prueba estadística será:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Las asunciones para que esta prueba sea válida son:

a) Las poblaciones están normalmente distribuídas
b) Las muestras son tomadas al azar independientemente
c) La homogeneidad de variancias ($\sigma_1^2 = \sigma_2^2$)

4. Regiones críticas: Como la prueba estadística es de una sola cola, observamos en la Tabla 3 para $n_1 + n_2 - 2 = 9$ grados de libertad con una probabilidad de $2\alpha = 0.10$ el valor tabular de t es 1.833. Luego la región crítica es $t > 1.833$



Nuevo (1):	3	2	5	8	4	6	4	5	10	14	11	14	12	15
	18	12	18	16	16	15								
Antiguo (2):	1	4	3	1	5	6	3	1	4	2	7	3	4	2
	1	5	8	2	1	3								

Desarrollo:

1. Hipótesis: La hipótesis es de que en promedio el número de expedientes tramitados son iguales en los dos horarios y como no existe ninguna referencia de que el nuevo horario aumentará ó disminuirá el número de expedientes tramitados tendremos una hipótesis alternante de dos colas, es decir:

   Hp: $\mu_1 = \mu_2$ ó $\mu_1 - \mu_2 = 0$

   Ha: $\mu_1 \neq \mu_2$ ó $\mu_1 - \mu_2 < 0$, $\mu_1 - \mu_2 > 0$

2. El gerente escoge $\alpha = 0.01$

3. La prueba estadística será:

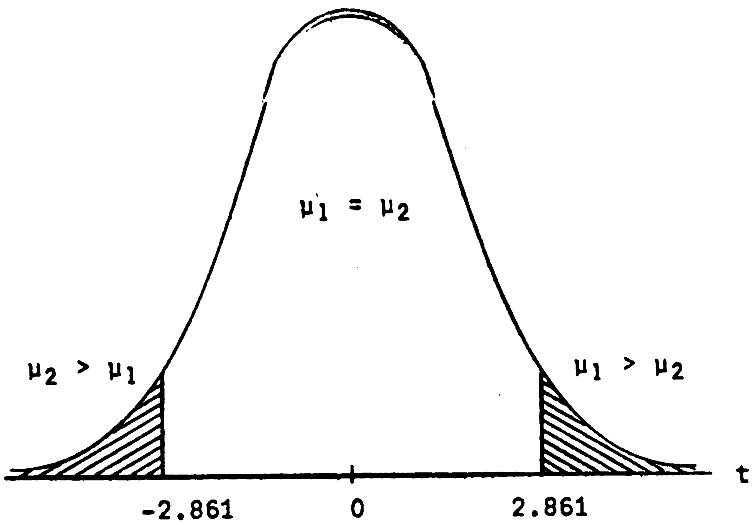
$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

   Las asunciones son:

   a) Las poblaciones están normalmente distribuídas

   b) Las muestras son tomadas al azar independientemente

4. Las Regiones críticas son determinadas con el valor t' que será igual al valor de t con 19 grados de libertad, que es $(n - 1)$, y con una probabilidad de $\alpha = 0.01$. El valor de t' es 2.861



5. Calcular la prueba estadística:

	Horario	
	Nuevo (1)	Antíguo (2)
ΣX	205	65
$\bar{x}$	10.25	3.25
ΣX^2	2697	299
$(\Sigma X)^2/n$	2101.25	211.25
S.C.	595.75	87.75
s^2	31.35	4.62

Luego:

$$t = \frac{(10.25 - 3.25) - 0}{\sqrt{\frac{4.62}{20} + \frac{31.35}{20}}} = 5.22$$

Conclusión: Como la t calculada (5.22) se encuentra dentro de la región crítica se rechaza la hipótesis de que las dos medias de la población son iguales y se acepta la alternante de que en promedio el nuevo horario permitirá mayor trámite de expedientes que el horario antiguo ($\mu_1 > \mu_2$).

Ejemplo 4. Una compañía vende sus productos en dos locales de la ciudad y desea saber si en promedio las ventas de las dos tiendas son iguales. Para el efecto, lleva a cabo un muestreo de ventas obteniendo los siguientes resultados expresados en miles de soles.

Tienda 1	Tienda 2
76	30
71	80
69	48
72	94
75	50
66	40
	65
	38

Desarrollo:

1. Hipótesis: Hp: $\mu_1 = \mu_2$

   Ha: $\mu_1 \neq \mu_2$

2. Nivel de significación: $\alpha = 0.05$

3. La prueba estadística a usarse es la de t. No existe homogeneidad de variancias.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Las asunciones para que esta prueba sea válida son:

a) Las poblaciones están normalmente distribuídas

b) Las muestras son tomadas al azar independientemente

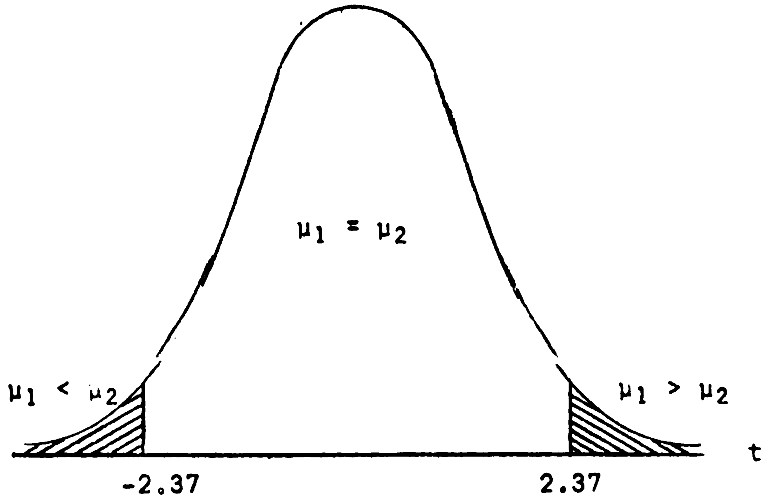
4. Regiones Críticas: Las regiones críticas se determinarán de acuerdo a la ecuación:

$$t' = \frac{\frac{s_1^2}{n_1} t_1 + \frac{s_2^2}{n_2} t_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

no es raíz

$$t' = \frac{2.317\ (2.571) + 61.712\ (2.365)}{2.317 + 61.712} = 2.37$$

Las regiones críticas con $\alpha = 0.05$ serán



5. Realizar la prueba estadística

	Tienda 1	Tienda 2
n	6	8
$\bar{x}$	71.50	55.625
ΣX^2	30743	28209
$(\Sigma X)^2/n$	30673.5	24753.125
S.C.	69.5	3455.875
s^2	13.9	493.6964

$$t = \frac{(71.50 - 55.625) - (0)}{\sqrt{\frac{493.6964}{8} + \frac{13.9}{6}}} = 1.984$$

Como el valor calculado se encuentra en la región de aceptación de la hipótesis planteada, decimos que no hay evidencia para afirmar que existen diferencias en el promedio de ventas de las dos tiendas.

Pruebas de hipótesis concerniente a la igualdad de medias de dos poblaciones mediante el apareamiento de variables.

A veces sucede que al muestrear dos poblaciones, factores extraños influyen en la diferencia de medias, no obstante no haber diferencia en los efectos que se tratan de medir. Por ejemplo, si se quiere estudiar dos métodos de enseñanza se pueden tomar dos grupos de estudiantes y a uno de ellos se le enseñará por el primer método y al otro por el segundo método, pero si el primer grupo fué de los mejores estudiantes, los resultados no medirán la efectividad de los métodos de enseñanza. Si en cambio, se hallan pares de estudiantes que tengan el mismo grado de inteligencia y a uno de los estudiantes de cada par se le enseña por el primer método y al otro estudiante por el segundo método, la efectividad del método será determinada sin factores extraños. En esta forma lo que se está tratando de hacer es que los pares sean iguales en todos los aspectos excepto en lo que deseamos medir.

Una ventaja de este método es que no se necesita asumir que las variancias de las dos poblaciones sean iguales ($\sigma_1^2 = \sigma_2^2$).

El procedimiento es trabajar con las diferencias para hacer un estimado de σ^2 y aplicar la prueba estadística de t con n - 1 grados de libertad (n = número de pares).

Para ilustrar el procedimiento veremos el caso de dos métodos de enseñanza 1 y 2 aplicados a un grupo de alumnos escogidos en pares según el grado de inteligencia. Los resultados de los métodos de enseñanza se juzgaron a través de un examen comun a los estudiantes y las notas obtenidas fueron:

<u>Número de pares</u>

	1	2	3	4	5	6	7	8	9	10
Método 1	20	8	17	15	24	12	11	16	14	9
Método 2	15	2	13	15	20	7	7	10	11	6
1 - 2 = D	5	8	4	0	4	5	4	6	3	3

<u>Desarrollo</u>.

1. Hipótesis: Hp: $\mu_1 = \mu_2$ ó $\mu_1 - \mu_2 = \mu_D = 0$

   Ha: $\mu_1 \neq \mu_2$ ó $\mu_1 - \mu_2 > 0$

   $\mu_1 - \mu_2 < 0$

2. Nivel de significación: $\alpha = 0.01$

3. La prueba estadística es:

$$t = \frac{\bar{D} - \mu_{\bar{D}}}{\sqrt{\frac{s_D^2}{n}}}$$

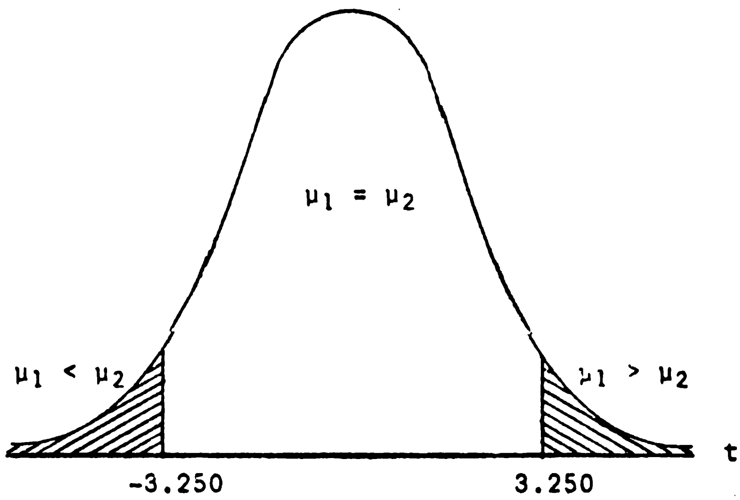
donde: $\bar{D}$ es la media de las diferencias entre los pares, s_D^2 es la variancia de la diferencia de cada par $D_i = (X_{1i} - X_{2i})$

Las asunciones para que esta prueba sea válida son:

1) La población de diferencias son normalmente distribuídas

2) Las muestras son tomadas al azar

4. Regiones críticas para 9 grados de libertad (n - 1) y $\alpha = 0.01$ son: $-3.250 < t < 3.250$



5. Calcular la prueba estadística

Número de Pares	Método 1	Método 2	D_i Diferencia 1-2
1	20	15	5
2	8	2	6
3	17	13	4
4	15	15	0
5	24	20	4
6	12	7	5
7	11	7	4
8	16	10	6
9	14	11	3
10	9	6	3
Suma	146	106	40
Medias	14.6	10.6	4

$$n = 10$$

$$\Sigma D_i^2 = 188$$

$$(\Sigma D_i)^2/n = 160$$

$$\Sigma D_i^2 - \frac{(\Sigma D_i)^2}{n} = S.C. = 28$$

$$s_D^2 = \frac{S.C.}{n-1} = \frac{28}{9} = 3.11$$

$$t = \frac{4}{\sqrt{\frac{3.11}{10}}} = 7.14$$

Como conclusión rechazamos la hipótesis planteada y aceptamos la hipótesis que el método de enseñanza 1 es estadísticamente superior al método de enseñanza 2, pues el valor 7.14 se encuentra dentro de la región crítica.

Prueba de hipótesis concerniente a la variancia de una población.

Muchos problemas de Estadística consisten en determinar la variabilidad de la población, que es tan importante como conocer la media de ésta.

La hipótesis planteada es de que la variancia de la población es igual a una constante, es decir:

Hp: $\sigma^2 = \sigma_o^2$

y las hipótesis alternantes pueden ser:

a) de una sola cola: Ha: $\sigma^2 < \sigma_o^2$, $\sigma^2 > \sigma_o^2$, ó

b) de dos colas: Ha: $\sigma^2 \neq \sigma_o^2$

El procedimiento para probar esta hipótesis es extraer una muestra de la población y calcular la variancia de la muestra s^2. Si el valor de s^2 discrepa notoriamente del valor σ_o^2 y ésta no puede atribuirse al azar, se rechaza la hipótesis planteada.

La prueba estadística que se utiliza para ello es la de χ^2 establecida en la siguiente ecuación:

$$\chi^2 = \frac{(n - 1)\hat{s}^2}{\sigma_o^2}$$

En el siguiente ejemplo, ilustramos el uso de esta prueba de hipótesis.

El gerente de una fábrica productora de clavos, desea que la variabilidad (σ^2) de éstos sea a lo más 0.0005 pulgadas cuadradas y para el efecto, decide tomar una muestra de su producción escogiéndola al azar obteniendo los siguientes resultados expresados en pulgadas: 1.13, 1.12, 1.15, 1.10, 1.11, 1.18, 1.20, 1.14, 1.12, 1.19, 1.10, 1.14, 1,13. La probabilidad de cometer Error Tipo I escogido por el fabricante es de 0.01.

<u>Desarrollo</u>.

1. Hipótesis: Hp: $\sigma^2 = 0.0005$

   Ha: $\sigma^2 > 0.0005$

2. Nivel de significación: $\alpha = 0.01$

3. La prueba estadística es:

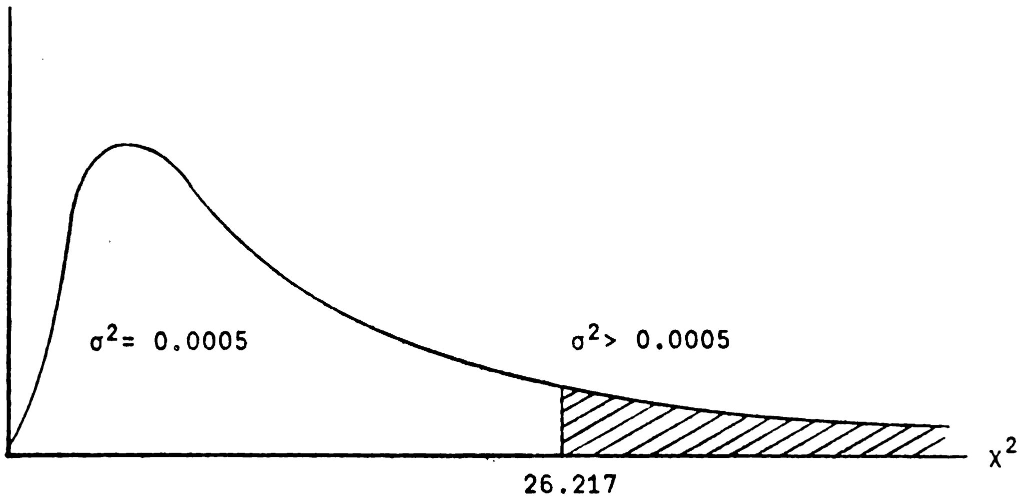
$$\chi^2 = \frac{(n - 1)\, s^2}{\sigma^2}$$

Las asunciones son:

a) La población está normalmente distribuída

b) La muestra es tomada al azar

4. Regiones críticas: El valor tabular de χ^2 para 12 grados de libertad con un nivel de significación de 0.01 es 26.217.



5. Calcular la prueba estadística:

$\Sigma X = 14.81$

$\Sigma X^2 = 16.8849$

$(\Sigma X)^2/n = 16.8720$

S.C. $= 0.0129$

$$\chi^2 = \frac{0.0129}{0.0005} = 25.80$$

Como este valor es menor que el valor tabular concluímos que no hay evidencia para rechazar la hipótesis planteada.

Prueba de hipótesis de dos variancias.

Al tratar la prueba de hipótesis de dos medias vimos que una de las asunciones era de que las variancias de las dos poblaciones eran homogéneas ($\sigma_1^2 = \sigma_2^2$); entonces antes de proceder con la prueba de hipótesis de dos medias es deseable probar si la asunción es correcta. La hipótesis planteada será Hp: $\sigma_1^2 = \sigma_2^2$ y las hipótesis alternantes pueden ser:

a) de una sola cola: Ha: $\sigma_1^2 < \sigma_2^2$, $\sigma_1^2 > \sigma_2^2$ ó

b) de dos colas: Ha: $\sigma_1^2 \neq \sigma_2^2$

El procedimiento para probar esta hipótesis consiste en extraer una muestra de cada población determinando sus variancias y se halla $F = \frac{s_1^2}{s_2^2}$.

La Tabla 5 permite encontrar directamente la región crítica para una sola cola. Para la prueba de dos colas ó sea cuando la alternante es $\sigma_1^2 \neq \sigma_2^2$ las regiones críticas se encontrarán de la siguiente manera:

$$F < F_{1 - \frac{\alpha}{2}} \quad y \quad F > F_{\frac{\alpha}{2}}$$

Para determinar la primera región crítica, tendremos que hacer un pequeño cálculo utilizando la ecuación:

$$F_{(1 - \frac{\alpha}{2})}\,(n_1 - 1, n_2 - 1) = \frac{1}{F_{\frac{\alpha}{2}}\,(n_2 - 1, n_1 - 1)}$$

Los siguientes ejemplos ilustran el uso de esta prueba de hipótesis.

Ejemplo 1. Utilicemos el mismo ejemplo 3 dado en 10.8, en que se utilizan diferentes horarios de trabajo en una oficina; se desea saber si la variabilidad del número de expedientes en los dos horarios es igual, sabiendo que puede existir mayor variabilidad en el horario nuevo.

1. Hipótesis: Hp: $\sigma_1^2 = \sigma_2^2$

   Ha: $\sigma_1^2 > \sigma_2^2$

2. Nivel de significación: $\alpha = 0.05$

3. La prueba estadística a usar será la de F.

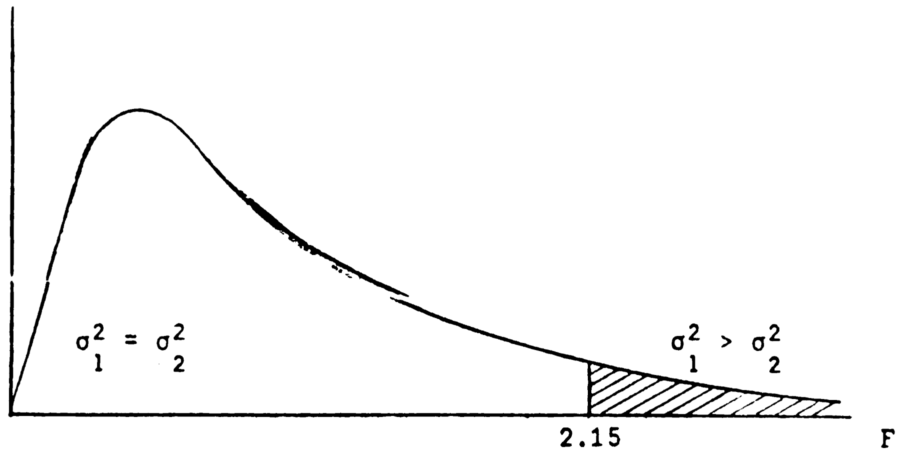
$$F = \frac{s_1^2}{s_2^2}$$

Las asunciones para que esta prueba sea válida son:

a) Las poblaciones están normalmente distribuídas

b) Las muestras son tomadas al azar independientemente

4. Regiones críticas: Como la prueba estadística en este caso es de una sola cola, observamos en la Tabla 5 para 19 y 19 grados de libertad, que son los grados de libertad para la variancia del numerador y denominador respectivamente, con una probabilidad de 0.05 cuyo valor aproximado es de 2.15



5. Calcular la prueba estadística. Primero debemos calcular cuales son las variancias de las muestras para los dos horarios lo que nos dá

	Horario	
	Nuevo	Antíguo
s^2	31.35	4.62

Teniendo estos valores procederemos a calcular:

$$F = \frac{s_1^2}{s_2^2} = \frac{31.35}{4.62} = 6.79$$

Como este valor de F calculado (6.79) es superior a F tabulado (2.15) diremos que rechazamos la hipótesis de que existe homogeneidad de variancias, y concluímos que la variancia en el trámite de expedientes es mayor en el horario nuevo que en el antíguo.

Ejemplo 2. Para ilustrar el uso bilateral ó de dos colas de la prueba utilicemos el ejemplo del anuncio publicado en dos periódicos, utilizado en el ejemplo 2 (10.8), pero con distinto tamaño de muestra. Se desea probar con una probabilidad de 10% si existe homogeneidad de variancias, no teniéndose ninguna experiencia hasta el momento sobre las variancias.

Desarrollo.

1. Hipótesis: $H_0: \sigma_1^2 = \sigma_2^2$

   $Ha: \sigma_1^2 \neq \sigma_2^2$

2. Nivel de significación: $\alpha = 0.10$
3. Determinar la prueba estadística y sus asunciones:

   La prueba estadística es la de F.

$$F = \frac{s_1^2}{s_2^2}$$

   y las asunciones para que esta prueba sea válida son:

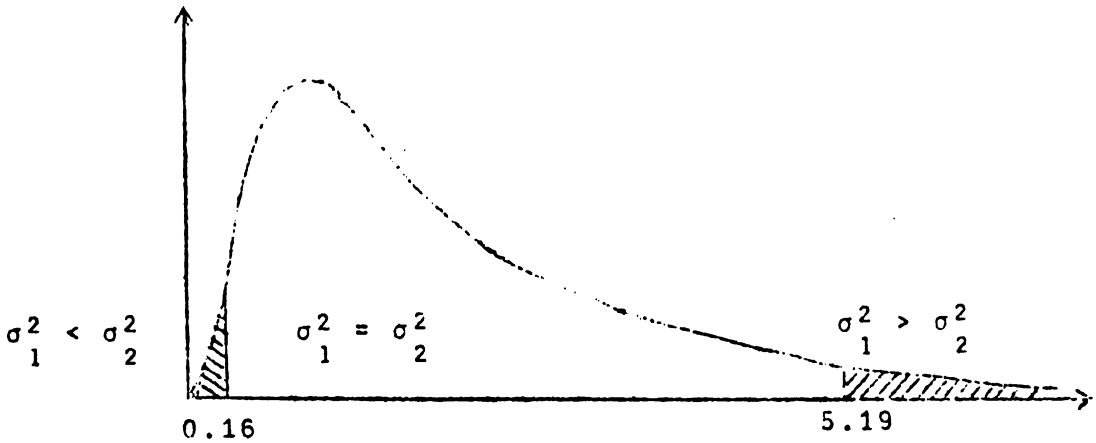
   a) Las poblaciones están normalmente distribuídas

   b) Las muestras son tomadas al azar

4. Regiones críticas: como ésta es una prueba de dos colas se establecen los límites en la siguiente forma:

$$F_{0.05(4,5)} = 5.19$$

$$F_{0.95(4,5)} = \frac{1}{F_{0.05(5,4)}} = \frac{1}{6.26} = 0.16$$



5. Calcular la prueba estadística: Se determinan las variancias para el periódico 1 y 2 que son:

	Periódico	
	1	2
s^2	1953.70	1154.67

y se calcula la prueba de F

$$F = \frac{1953.70}{1154.67} = 1.692$$

Como vemos el valor de F calculado está dentro de la región de aceptación, luego concluiremos que no existe suficiente evidencia como para rechazar la hipótesis planteada.

EJERCICIOS

1. Un laboratorio prepara cápsulas de antibióticos cuyo peso es 250 miligramos. El llenado de las cápsulas es automático y se realiza a continuación un control para verificar el llenado, si éste presenta fallas muy notorias se regresan para el relleno, elevándose así los costos de producción. Para verificar si el llenado (incluyendo el rellenado) se está realizando satisfactoriamente se toma al azar una muestra de 9 observaciones encontrándose los siguientes resultados:

250.5	249.5	250.0	249.0	251.0
248.9	251.0	248.5	247.9	

Si $\alpha = 0.05$

a) ¿Acepta o rechaza la hipótesis de que en promedio el llenado pesa 250 miligramos?

b) Establecer los límites de confianza de la media con una probabilidad de 0.01. Interprete sus resultados.

2. Un nutricionista halla una nueva ración para vacas lecheras y de acuerdo a los estudios preliminares, esta nueva ración es más cara que la ya establecida y unicamente será adecuada si la producción promedio de leche es 20 litros diarios y si la desviación standard es menor de 4 litros. Para tomar una decisión final referente a esta nueva ración, decide llevar a cabo un experimento con 64 vacas en producción lechera, obteniendo los siguientes datos presentados en una Tabla de Frecuencia.

Tabla de Frecuencia

Intervalo de clase (litros)	Frecuencia
11-13	8
14-16	9
17-19	10
20-22	22
23-25	8
26-28	4
29-31	3

Si el nutricionista escoge α = 0.10 ¿Cuál será la decisión del nutricionista?

3. Con el objeto de determinar que tipo de gasolina, A ó B, es más conveniente por su rendimiento promedio para un tipo de automóvil, se lleva a cabo un experimento usando 10 automóviles para cada tipo de gasolina, encontrándose los siguientes resultados:

Tipo de gasolina

	A	B
$\bar{x}$	26.5 Km/galón	32.0 Km/galón
s	2.4 Km.	5.0 Km.

Si α = 0.10 ¿Existen diferencias estadísticas en cuanto al rendimiento promedio de los dos tipos de gasolina?

4. El gerente de la Compañía A desea conocer si las baterías que produce son de mejor calidad que las baterías elaboradas por la Compañía B. El gerente de la Compañía A dirá que sus baterías son mejores si tienen mayor "vida activa" que los fabricados por la Compañía B; y si sus baterías tienen igual ó menor variabilidad en su "vida activa" que las baterías de la Compañía B. Para el efecto, el gerente escoge 10 baterías de la Compañía A y 8 baterías de la Compañía B y con ellas lleva a cabo un experimento, midiendo la "vida activa" de las baterías, con un nivel de significación de 5%. Los resultados obtenidos son los siguientes:

A =	800	950	870	1050	700	650	775	800	785	900
B =	750	800	850	700	800	700	750	650		

¿Cuáles son las conclusiones del gerente?

5. Una compañía productora de sogas desea conocer las características de un nuevo tipo de soga que va a introducir al mercado. Las características de interés son: el peso promedio máximo que puede levantar la soga y la variabilidad de los pesos que ocasionan la ruptura de la soga. Para el efecto, realiza una prueba con 16 sogas y encuentra que el promedio de rompimiento de esas sogas es de 150 Kgs. con una desviación standard de 12 Kgs. Pruebe la hipótesis de que la variancia de los pesos de rompimiento de las sogas es menor de 60 Kgs.

6. Una compañía productora de rodajes encarga a dos de sus fábricas 5,000 rodajes, con un diámetro de 2 cm. Para determinar si existen diferencias estadísticas en el diámetro promedio de los rodajes de las dos fábricas, decide tomar una muestra al azar de 20 rodajes en la Fábrica A y 10 rodajes de la Fábrica B; y mide los diámetros de los rodajes de las muestras obteniendo los siguientes resultados:

	Fábrica A	Fábrica B
n	20	10
Media $\bar{x}$	2.0090 cm.	2.0015 cm.
Variancia (s^2)	38.44×10^{-6}	1×10^{-6}

Si $\alpha = 0.05$ ¿Cuáles serían sus conclusiones?

7. En un estudio realizado para comparar el rendimiento de un híbrido nuevo de maíz, con un híbrido ya establecido se llevó a cabo un experimento encontrándose los siguientes resultados expresados en Kilos por hectárea.

Híbrido nuevo	Híbrido antíguo
6,500	4,200
6,700	4,300
7,000	6,000
5,800	5,300
6,000	6,100
6,300	4,000

Si $\alpha = 0.05$ ¿El rendimiento promedío del híbrido nuevo es superior al rendimiento promedio del híbrido antíguo?

Título: **NOTAS SOBRE MANEJO DE INFORMACION EN EXPERIMENTACION CON ANIMALES CON ENFASIS EN GANADO BOVINO DE LECHE**

Introducción:

Las notas que a continuación se presentan, constituyen una recopilación de experiencias en el manejo de información y conducción de diseños experimentales, en el área de Producción Animal con énfasis en ganado de leche.

La intención es presentar al lector los aspectos más relevantes que deben tenerse en cuenta cuando se trabaja en ensayos con animales, de modo que estos sirvan de elementos de juicio para el planeamiento de ensayos experimentales.

Probablemente el lector encuentre en estas notas algunos errores gramáticales y de estructura, ya que el documento aún no ha sido revisado ni corregido, por esta razón las presentamos en forma preliminar y concientes de que en el futuro deben mejorarse y completarse.

Ing. Mg.Sc. Rolando Piskulich J.

NOTAS SOBRE MANEJO DE INFORMACION EN EXPERIMENTACION CON ANIMALES CON ENFASIS EN GANADO BOVINO DE LECHE

1. Generación y flujo de información en producción animal.

En el campo de la Producción Animal, la generación de información tiene dos fuentes con características propias, esto es, la generación de información en estaciones experimentales y en explotaciones comerciales. En lo que se refiere a montar ensayos para generar información básica, la generación de información en estaciones experimentales se caracteriza porque:

- No necesariamente se tiene en cuenta el valor de los productos dentro de un ensayo (carne, leche, etc.), lo que permite someter a los animales a las pruebas necesarias como por ejemplo fistulaciones para recopilar la información requerida.

- Las condiciones en que se conducen los experimentos permite un mejor control, tanto en la toma de información como en el cumplimiento de supuestos requeridos para un ensayo.

- El número de animales disponibles para ensayos es generalmente limitado, lo que crea limitaciones en lo que se refiere a inferencia estadística y a la utilización de diseños experimentales de alta precisión.

Mientras que la generación de información en explotaciones comerciales se caracteriza porque:

- Generalmente en cualquier ensayo se debe tener especial cuidado en que el mismo no afecte los niveles de producción ni a los animales, lo que limita la realización de ensayos.

- Existe gran dificultad en el control de la toma de información como en el cumplimiento de supuestos requeridos por un ensayo.

- El número de animales disponibles permite utilizar diseños que requieren un mayor número de repeticiones y realizar inferencias sobre poblaciones más grandes.

Como se puede observar, las ventajas de la recopilación de información de estaciones experimentales, son las desventajas de la recopilación de información en explotaciones comerciales y viceversa.

Lo expuesto anteriormente, hace que en la práctica se complemente la información proveniente de ambas fuentes y por lo general, las explotaciones comerciales son utilizadas como fuentes de información para la identificación de problemas, para validación de tecnologías o normas recomendadas que son producto de la investigación en estaciones experimentales y para investigación en el campo de mejoramiento animal, en donde definitivamente las pruebas de hipótesis requieren de ensayos con gran número de animales y a lo largo de períodos que puede incluir varios años. Mientras que las estaciones experimentales están más orientadas hacia el tipo de investigación básica, tratando en la mayoría de los casos, de orientar su investigación a la solución de problemas que surgen en explotaciones comerciales, así en la Figura 1. podemos notar que la decisión sobre el tipo de experimentación a realizar tiene como estímulo los problemas en producción, lo cual ya constituye información que no siempre es fácil de recopilar. Una vez identificado el problema, la experimentación genera información que tiene como objetivo resolver el problema, sin embargo, muchas veces los primeros resultados son parciales y solo son concluyentes cuando son probados en condiciones de producción comercial (llámese validación). El proceso de validación muchas veces lleva a la conclusión de que aún existen vacíos de información y que por lo tanto, se debe continuar la fase de experimentación. De otro modo, la validación podría indicar que el problema esta resuelto, y por lo tanto, las recomendaciones son aplicables. Como se puede apreciar, las dos fuentes de información, estación experimental y producción comercial, están íntimamente relacionadas y, si bien es cierto que la primera fuente sirve a los fines de la segunda, ambas pueden considerarse como componentes de un sistema en donde el flujo de información interrelaciona ambos componentes, de allí que un manejo adecuado de la información de ambas fuentes es de vital importancia para el funcionamiento del sistema, como un ejemplo más,

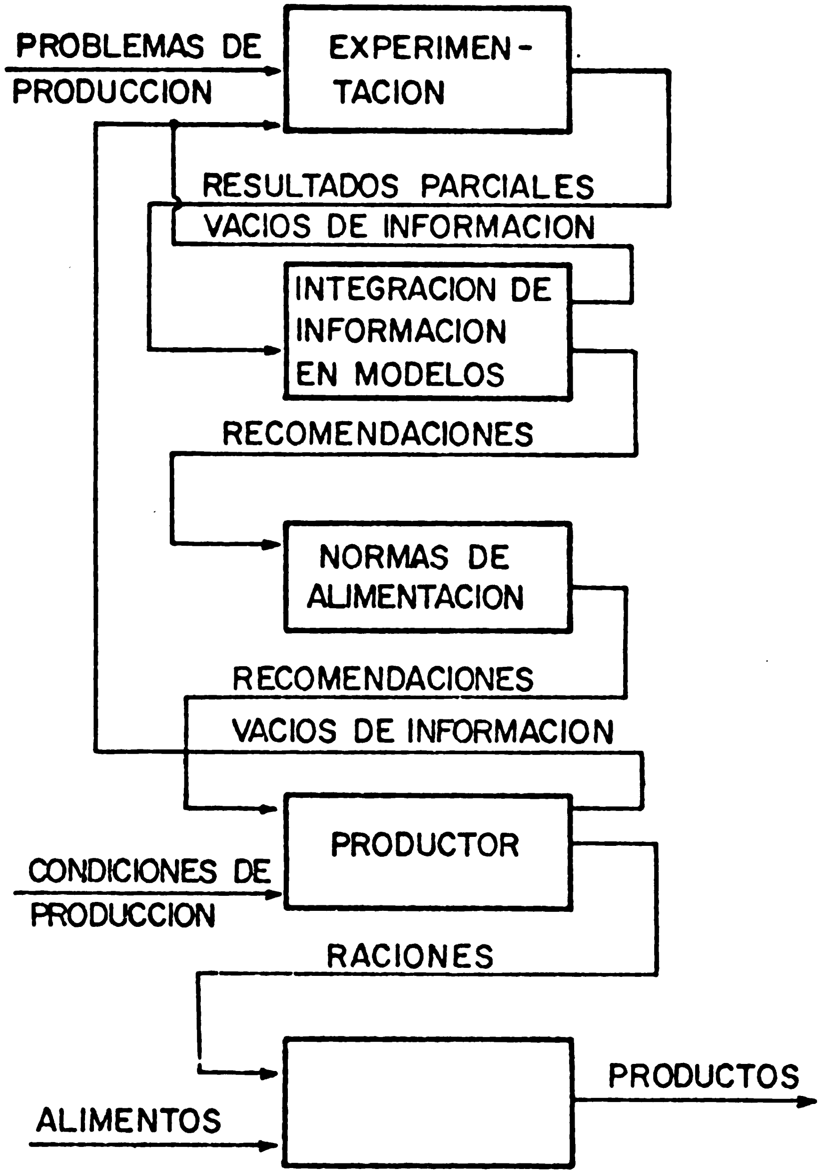
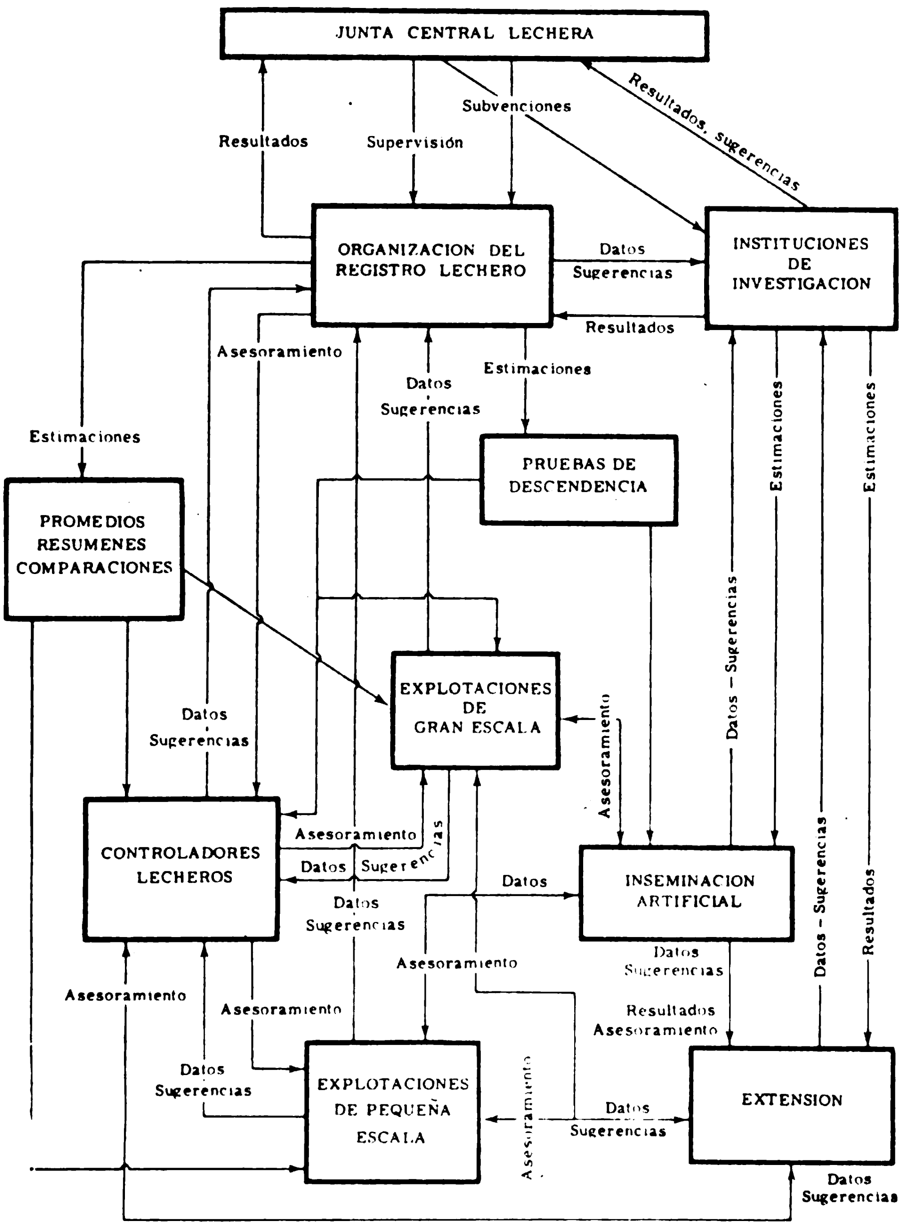


FIG. 1. NORMAS DE ALIMENTACION COMO RESULTADO DE UN PROCESO DE INTEGRACION DE INFORMACION



Gráfica 1. *Diagrama de la corriente de información que hace falta en el registro lechero y actividades conexas*

Fuente: Lindstron (5)

veamos en el diagrama de la Figura 2. en donde se presenta el diagrama de la corriente de información que hace falta en el registro lechero y actividades conexas cuando se tiene en mente programas de mejoramiento genético.

2. Consideraciones de aspectos fisiológicos en diseños experimentales en Producción Animal

Como se sabe, en fisiología animal, la medición de cualquier variable de respuestas, arroja variaciones entre individuos cuyas fuentes de variación son factores genéticos, factores ambientales y la interacción de factores genéticos X factores ambientales; esto implica que la información que se recopile debe permitir la explicación de la variación de una variable de respuesta en función de los factores mencionados, lo cual cuando la limitante es el número de animales no es tarea fácil. Por ello es de vital importancia conocer la fisiología misma del animal, de modo que contemos con elementos de juicio que nos permita elaborar diseños adecuados para la recopilación y análisis de información. En el presente documento, mencionaremos los aspectos más relevantes sobre fisiología de bovinos con énfasis en bovinos de leche, que deben considerarse previamente al diseño de cualquier ensayo experimental, en haras de lograr el mejor control sobre las fuentes de variación que afectan las variables de respuesta más comunes cuando se trabaja con esta especie, es decir producción de leche y componentes de la misma, así como crecimiento corporal. Cabe mencionar que en términos generales, muchos de los aspectos que se tocarán, son aplicables a otras especies.

2.1 Descripción de la curva de lactancia

Al analizar la curva de lactancia debemos tomar en cuenta una serie de factores, tales como: la máxima producción de leche diaria o pico de lactancia, la persistencia, el período de lactancia y la producción total de leche.

2.1.1 Máxima producción de leche diaria

En general se ha encontrado que la máxima producción de leche diaria, se alcanza durante el segundo mes de lactancia Molina (7), esto concuerda con lo reportado por Lucas (6) quien indica que la máxima producción se alcanza entre la cuarta y sexta semana de producción. Este período en que se alcanza el pico de producción no depende ni de la estación ni de la edad al primer parto Molina (7). En el tiempo indicado, el animal se ha recuperado de la tensión causada por el parto y la gestación anterior.

La producción pico va en aumento hasta la tercera lactancia, declinando luego ligeramente. La edad en que los animales alcanzan la producción máxima oscila entre sesenta y noventa y seis meses, cuando se ha alcanzado el máximo desarrollo fisiológico a causa de los partos y lactancias precedentes, Molina (7).

La edad de los animales influye directamente sobre la producción pico. Las vacas de mayor edad y altamente productoras requieren mayor tiempo para alcanzar la máxima producción de leche a través de la lactancia, al respecto se reporta medias de treinta y nueve días para la raza Holstein y cuarenta y tres días para la raza Jersey, Molina (7).

Se han encontrado correlaciones altamente significativas de 0.82 entre la producción pico y la producción total de una lactancia, Molina (7).

2.1.2 Persistencia

La persistencia se expresa como la reducción porcentual de la producción de leche de un mes con respecto al mes anterior, y es afectada por varios factores tales como: la estación del año, la nutrición animal y la edad del animal al primer parto, Molina (7). Al respecto, varios investigadores convergen en que animales paridos en malas condiciones climáticas o que por primera vez tengan cría a temprana edad serán menos persistentes. Cuando se mejoran

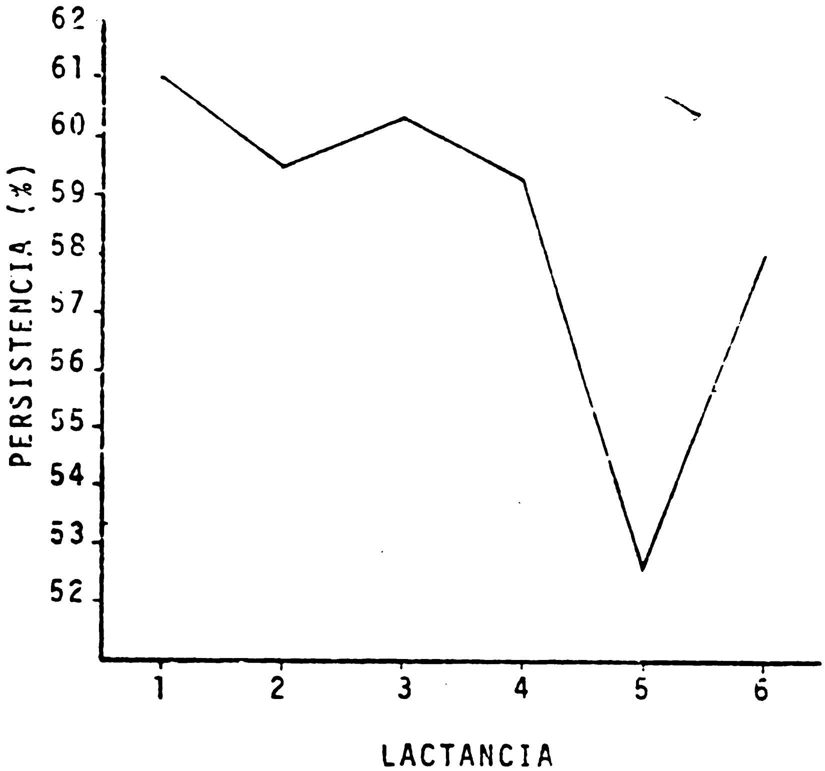


FIGURA 3. DISTRIBUCION DE LOS PROMEDIOS DE PERSISTENCIA POR LACTANCIA

Fuente: Molina (7)

factores como el ambiente se pueden obtener incrementos en la persistencia pero no en la producción pico. Molina (7).

Los máximos valores de persistencia se dan en la primera lactancia, siendo luego cada vez menos hasta experimentar incrementos en la lactancia sexta y octava, ver Figura 3. Conforme aumenta la edad, la persistencia decrece y hay una variación de 0.8 por ciento a 0.4 por ciento de la primera a la segunda lactancia. También el número de días del parto a la concepción producen una variación de un 7% a un 5%, Molina (7).

Entre persistencia y pico de producción la correlación es baja, Molina (7). La curva de producción que describe la lactancia desciende en forma lineal desde el día pico hasta el día doscientos veinte más o menos, para luego decaer bruscamente debido a la influencia que ejerce el crecimiento fetal. Molina (7).

Lucas (6) reporta que la variabilidad en persistencia entre lactaciones de una misma vaca, es menor que la variabilidad de la persistencia entre vacas.

2.1.3 Período de lactancia

La duración óptima para un período de lactancia es de diez meses, de modo que el animal pueda dar la mayor producción dentro de un marco rentable, no obstante, las primeras lactancias en muchos casos no alcanzan esa longitud. Se reportan efectos no significativos de la edad al primer parto sobre la longitud de la lactancia, mientras que esta si es afectada significativamente por la época de parición, Molina (7).

Sistemas de alimentación irregulares o insufiencientes afectan significativamente, tanto la producción total de leche como el período de lactancia, Molina (7), Lucas (6).

2.1.4 Descripción de la curva de lactancia

La producción de leche está delimitada por el área comprendida bajo la curva que representa una función curvilínea de tendencia variable, la cual ha sido dividida en tres regiones, a saber: la primera línea que va desde el día del parto hasta el día de máxima producción de leche o día pico, la segunda que comprende el área entre el día pico y el día doscientos setenta, mostrando un descenso lineal en la producción y por último una región comprendida entre el día doscientos setenta y el final del período de lactancia, donde la curva experimenta una marcada caída por efecto de la gestación.

Se ha encontrado que la función $Y = A\, e^{-Bt}\, t^{C}$

donde: Y : Producción de leche
t : Tiempo
ABC : Parámetros

(Gamma) tiene una buen rango de ajuste de la curva de producción de leche, excepto para la primera lactancia donde el grado de confiabilidad (R^2) es bajo respecto a otras lactancias; debido a una irregularidad en la producción de los animales primerizos que no han alcanzado el completo desarrollo y por lo tanto utilizaron parte de la energía para producción en otras funciones. A pesar de la variabilidad individual, la función Gamma estima la producción total de leche promedio con un coeficiente de determinación cercano a uno ($R^2 = 0.8$ para la primera lactancia y 0.99 para las lactancias siguientes), además en base a ella se puede determinar la persistencia y se reduce el error de cálculo de la producción total.

Lucas (6) recomienda un modelo $Y_t = Y_0\, e^{-kt}$

donde: Y_t : Producción de leche al momento t
Y_0 : Producción de leche en el pico de lactancia
t : Tiempo a partir del pico de producción
k : Persistencia

que resulta especialmente útil para estudiar características de persistencia.

La producción a diferentes tiempos en una misma lactancia tiene alta correlación $0.95 \leq r \leq 0.99$ Lucas (6). La producción a un mismo momento de la lactancia entre diferentes lactancias de una vaca tiene baja correlación $0.5 \leq r \leq 0.8$, donde $r = 0.8$ se alcanza

CONTENIDO

Título: **NOTAS SOBRE MANEJO DE INFORMACION EN EXPERIMENTACION CON ANIMALES CON ENFASIS EN GANADO BOVINO DE LECHE** en lactancias sucesivas, Lucas (6).

2.2 Factores que afectan la curva de lactancia y su consideración en diseños experimentales

En el presente capítulo, se mencionarán los factores que afectan la curva de lactancia y las consideraciones más importantes que se deben hacer, de modo que en los diseños experimentales para conducir ensayos sobre producción de leche, se reduzca el error experimental al mínimo y las fuentes de variación a la que se atribuye la variación en las variables de rendimiento, sea correctamente incluídas dentro del diseño.

2.2.1 Factores genéticos

Es sabido que cada raza tiene características propias de tasas de producción y composición de leche, Cuadro N° 1. Sin embargo, es común observar que individuos dentro de una misma raza varían ampliamente al respecto, aunque gran parte de esta variación es atribuible al ambiente, no se puede decir que el efecto genético no está presente.

En experimentos que incluyen una sola lactancia, la influencia de factores genéticos sobre la persistencia es la relación mas importante a considerar. En el estudio que incluye muchas lactancias, los efectos de factores genéticos sobre la relación entre edad y tasa inicial son las más importantes.

Datos de Maynard y Myers (citado por Lucas (6)) indican que la correlación entre la tasa inicial y tasas subsecuentes de producción, en la misma lactación y por lo tanto, el error experimental esperado en ensayos que envuelven la misma lactancia, no varía por efecto de razas. Sin embargo, la persistencia si tiene una variación en exceso con respecto a la variación en tasa inicial de producción, por lo que, en experimentos en que la distribución de los tratamientos a los animales no se toma en cuenta la raza como factor, se espera

Cuadro 1. Promedio de edad, peso, producción y variables según grupo racial.

VARIABLES	GRUPO RACIAL			
	JERSEY	CRIOLLO	CRIOLLO x JERSEY	AYRSHIRE x F1
Número de lactancias	125	[illegible]	177	[illegible]
VARIABLES INDEPENDIENTES				
Edad (días)	1449 ± 574	2377 ± 1274	[illegible] ± 998	1724 ± [illegible]
Peso vivo (kg)	265 ± 42	360 ± 65	[illegible] ± 49	[illegible] ± 47
Peso metabólico	58 ± 7	73 ± [illegible]	[illegible] ± 7	[illegible] ± 7
VARIABLES DEPENDIENTES				
Producción de leche por lactancia (kg)	1494 ± 527	1352 ± 690	[illegible] ± 499	[illegible] ± 577
Producción de grasa por lactancia (kg)	63 ± 23	56 ± 29	[illegible] ± 24	66 ± 23
Producción relativa (kg)				
Leche/kg de peso vivo	5,58 ± 1,31	3,69 ± 1,84	5,86 ± 1,57	6,35 ± 1,72
Leche/kg de peso metabólico	25,24 ± 8,16	18,17 ± 9,03	28,00 ± 7,26	30,71 ± 8,05
Grasa/kg de peso vivo	0,23 ± 0,08	0,15 ± 0,08	0,24 ± 0,07	0,25 ± 0,07
Grasa/ kg de peso metabólico	1,07 ± 0,36	0,75 ± 0,38	1,19 ± 0,35	1,22 ± 0,32

F1 = Criollo X Jersey

Fuente: Bejarano (1)

un error alto, comparado con aquellos experimentos en que si se considera el factor raza. Barteli (citado por Lucas) sugiere, sin embargo que, cuando el efecto de los tratamientos es pequeño, la consideración de raza no es muy efectivo en reducir el error experimental.

Si la interacción de raza X tratamiento existe, entonces la consideración del factor raza en el diseño es deseable, no solo por aquello de reducir el error experimental, sino también con el fin de calcular la respectiva interacción.

En conclusión, es recomendable considerar el efecto de razas en la distribución de los tratamientos a los animales y en el análilisis de los resultados, cuando se esperan diferencias marcadas entre tratamientos y cuando las razas y edades de los animales que se usan son muy diferentes entre sí.

En estudios que incluyen muchas lactancias, la curva que describe la producción de leche en función de la edad es importante. En general, las tasas promedio de producción diaria y producción inicial son menores durante la primera lactación y se incrementa en sucesivas lactancias, hasta la quinta o sexta lactación, cuando comienza a decrecer, Figura 4. Al respecto, existe una alta variabilidad entre individuos, pero no es clara la diferencia observada entre razas. Por ello, se puede pensar que la curva de producción vs edad es la misma para todas las razas. Sin embargo, desde que esto no se ha demostrado, es aparentemente apropiado, sobre todo en ensayos que envuelven varias lactancias, hacer la consideración del efecto de raza para la producción en función de la edad.

2.2.2 Influencia nutricional

Los factores nutricionales son, por supuesto, de primera importancia en ensayos de alimentación con ganado lechero, resulta obvio que el factor nutricional afectará tanto la tasa inicial como la

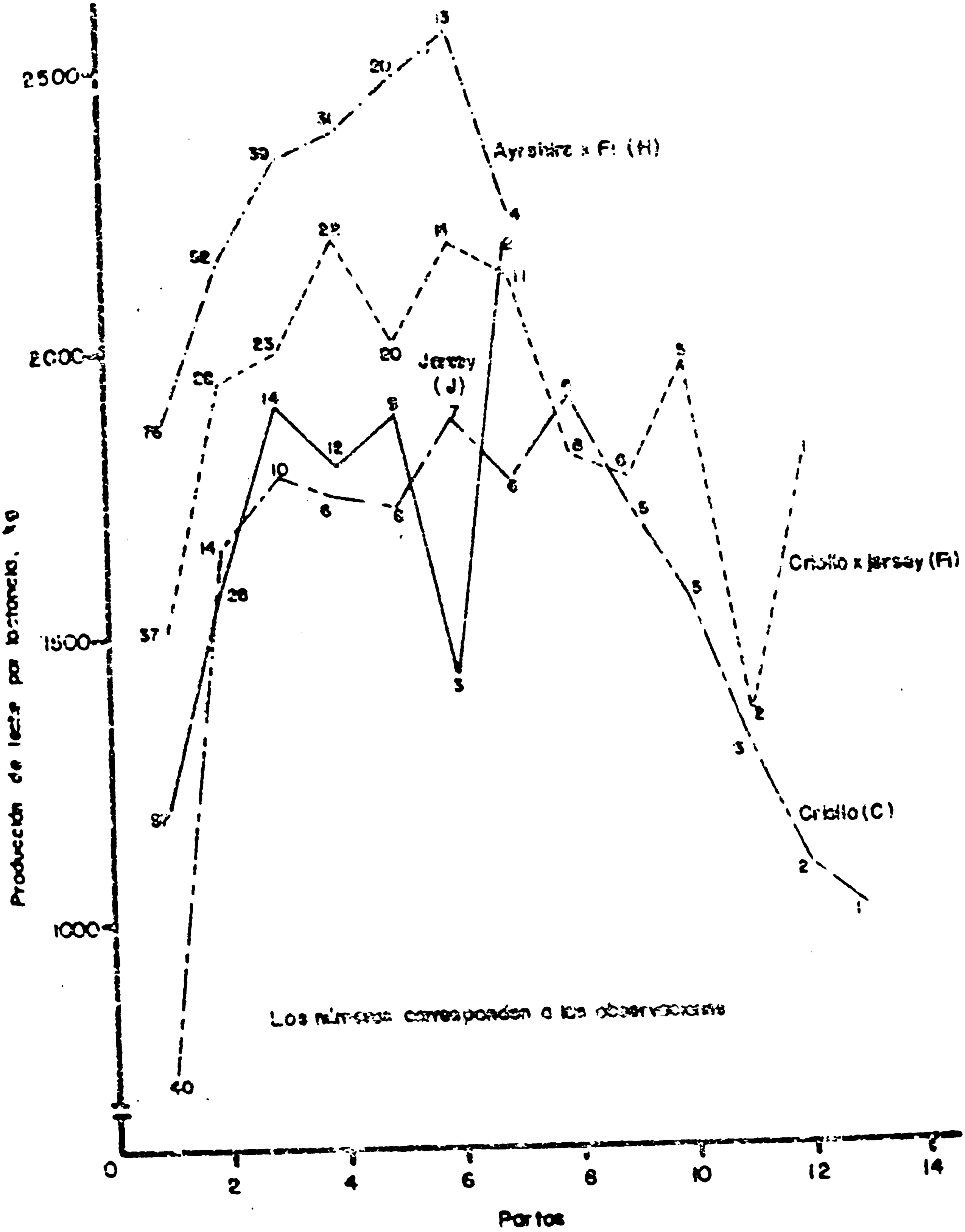


Fig.4. Producción de leche por lactancia según los partos en los distintos grupos raciales

Fuente: Bejarano (1)

persistencia, en este caso hay que diferenciar dos tipos de ensayos: a) Ensayos en los que la variable cantidad consumida no es una variable, b) Ensayo en que la cantidad consumida es una variable.

a) Ensayos en los que la cantidad consumida no es una variable -
En estos ensayos las raciones son usualmente formuladas con una porción de forrajes y otra de concentrados y dentro de estos ensayos el control del factor consumo, juega un papel importante para poder detectar diferencias entre tratamientos. Así, en algunos experimentos la alimentación con forrajes es ad-libitum y, dentro de estas condiciones es muy difícil calcular la eficiencia en producción, porque no se conoce el forraje consumido y este varía ampliamente. Considerando esta dificultad, muchos investigadores llevan a cabo sus ensayos dentro de condiciones de "consumo controlado de forraje", donde la tasa de forraje proporcionada a cada vaca es determinada por el investigador, principalmente sobre la base del peso corporal y atendiendo además, a condiciones peculiares como edad, condición y apetito. Una vez determinada la tasa para un animal dado, esta es usualmente mantenida constante durante el experimento. En algunos casos, la cantidad de forraje asignado es usualmente suficiente para completar satisfactoriamente los requisitos de mantenimiento de el animal, en otros, se cubre parte de los requisitos de producción. Los restantes requerimientos de producción y algunos excesos de rendimiento deseado en ganancia de peso, son suplidos por concentrados. Esta forma de manejar los ensayos, obedece al hecho de que, en explotaciones comerciales, es costumbre mantener el consumo de concentrados proporcional a la tasa de producción a través de la lactancia. Los ajustes en consumo de concentrado son hechos periódicamente ya sea semanal, quincenal o mensualmente, considerando el factor de ajuste como una fracción constante de la producción semanal, quincenal o mensual inmediata anterior, en forma respectiva. La práctica indicada de controlar el consumo, es criticada por el hecho de

que en un momento dado podría ocurrir que no se produzcan cambios en producción de leche, pero si estan ocurriendo pérdidas en tejidos que estaban almacenados. En otros casos, la producción podría sufrir adelantos o retrasos. Por lo que, un animal podría estar sujeto al castigo de un menor consumo, en perjuicio de movilizar tejidos de reserva. Esto indica que si el método se utiliza, debe ponerse especial cuidado a los cambios de peso del animal. Como alternativa, puede usarse la técnica de apareamiento. Así en estudios de crecimiento, los cambios en energía retenida se reflejan en el peso corporal, el cual es usado como medida de crecimiento. En estos casos, el criterio de comparación es altamente dependiente del total de energía consumida y de la eficiencia de la ración y cuando dos animales del mismo peso inicial y edad son alimentados con idéntica energía consumida, la varianza del error podría ser automáticamente limitada, alcanzándose comparaciones precisas. En el caso de animales lactando, sin embargo, el criterio de comparación (producción de leche y grasa) podría ser sustancialmente independiente de la energía consumida y la eficiencia de la ración en un momento dado, porque podría ocurrir que los tejidos almacenados podrían aumentar o decrecer mientras se mantiene una tasa de producción relativamente constante. Por lo tanto, si dos animales van a ser identicamente alimentados (en cuanto a consumo de energía), se deben usar muchos criterios de apareamiento, llámese peso corporal, producción de leche y grasa, edad y condición. Si estos criterios no son usados, los ajustes necesarios al set de consumo de nutrientes, podrían introducir variaciones adicionales, sin embargo, para el apareamiento, siempre estaremos limitados por el N°de animales, lo que no permitirá un apareamiento equitativo en todos los criterios de apareamiento mencionados, luego la técnica debe usarse solo cuando se pretende determinar diferencias muy grandes entre tratamientos.

Con la finalidad de superar las eventualidades del método de control de consumo en base al peso corporal, edad, condición y apetito y de superar las limitaciones de trabajar con animales apareados bajo el criterio de la mayor igualdad posible entre la pareja, Lucas (6) propone un método de alimentación equitativa (A.E), el mismo que puede ser aplicado a la mayor parte de los diseños experimentales. El método A.E consiste en dividir el período de control de consumo en tres etapas:

- Período preliminar
- Período de estandarización
- Período de comparación

Período preliminar - Durante este período, todos los animales son alimentados con la misma ración, con la finalidad de hacer una estimación de los requerimientos del animal considerando: producción, peso corporal, condición, edad y apetito.

Período de estandarización - Este período sirve a dos propósitos: a) Permite ajustar los errores cometidos en el período preliminar en la estimación de los requerimientos de los animales, b) Permite estabilizar los animales a un nivel de producción estandard, esto es de considerable importancia en ensayos contínuos, porque la producción y peso durante el período de estandarización podrían servir como variables independientes en subsecuentes análisis de covarianza.

Período de comparación - La utilidad del período de comparación es obvia, es el período de aplicación de tratamientos. En diseños de sobrecambio, el período de comparación es dividido en sub-períodos dependiendo del número de tratamientos a ser comparados y del diseño que se utilice.

Durante el período preliminar cada animal es individualmente alimentado y se toma especial cuidado en ajustar el consumo a peculiaridades individuales. Es mejor considerar la tasa de concentrado consumido que se desea durante el período experimental y suplir con forraje el resto, de tal modo que cada animal mantenga escencialmente constante su producción cuando todos los animales están recibiendo aproximadamente la cantidad correcta de concentrado por unidad de leche producida. Luego, podrían requerirse muchos ajustes en el forraje suministrado.

Lucas (6) recomienda que el período preliminar podría comenzar cerca del momento en que los animales alcanzan su pico de producción y ser bien alimentados. La longitud del período preliminar es determinada por la longitud de tiempo requerido para establecer la cantidad de forraje y concentrado necesarios para que el animal alcance un rendimiento normal, es probable que se requiera de 2 a 4 semanas

La cantidad de alimento determinada en la segunda semana del período de estandarización es mantenida a lo largo del resto del período de estandarización y durante todo el período de comparación.

Considerando que los ajustes se van a realizar sobre el consumo de concentrado y que el forraje se proporciona en cantidad suficiente, al nivel que se considere adecuado, entonces la cantidad por la cual el consumo de concentrado va a ser periódicamente cambiado podría ser determinada arbitrariamente por el investigador. Por ejemplo, al comienzo de cada semana, el concentrado diario usado podría ser reducido al 98% de aquellos usados durante la semana previa. Es costumbre, sin embargo, que el concentrado durante todo el tiempo guarde consistencia con el nivel de producción. Si es así, entonces los cambios en el consumo de concentrado podría ser basado sobre el promedio de producción alcanzado por todas las vacas que permanecen bajo una ración. Cuando este procedimiento es seguido, es necesario considerar, en adición al consumo inicial de concentrado, los niveles iniciales de producción, es decir, la producción de la última parte del período preliminar.

Para determinar el consumo de concentrado para alguna semana (quincena o mes) por este método, son necesarios los siguientes cálculos: La suma de la producción de todas las vacas durante la última semana (quincena o mes) del período preliminar, es dividida por la suma de producción total de las vacas durante la primera semana del período de estandarización; esto rinde un factor promedio de persistencia. El consumo durante la última semana del período inicial es multiplicado por este factor, para obtener el concentrado diario necesario para cada animal para la semana (quincena o mes) particular deseada.

CUADRO 2. Una ilustración de los cálculos necesarios para ajustar el consumo de concentrado, considerando la producción y el consumo de la última semana del período preliminar.

Vaca N°	Ultima semana del período preliminar Producción Inicial	Período de estandarización 1° semana requisitos de T.N.D. de acuerdo a la producción del período anterior 0.324 **	Producción	2°semana requisitos de T.N.D. (concentrado) 1°semana X 0.977***	Producción	3°semana requisitos de T.N.D. (concentrado) 2°semana X 0.959****
1	60.0	19.4	58.5	19.0	57.0	18.5
2	52.0	16.8	51.0	16.4	50.0	16.0
3	45.0	14.6	44.0	14.2	43.5	14.0
SUMA	157.0		153.5		150.5	

* Leche corregida %/grasa

** 0.324 Libras de T.N.D. X libra de leche producida según Morrison citado por Lucas (6)

*** 0.977 cociente de 153.5/157.0

**** 0.959 cociente de 150.5/153.5

Fuente: Lucas (6)

Los cálculos indicados en el Cuadro 1 se pueden utilizar durante el período de estandarización y el período de comparación, siempre usando el concentrado inicial consumido y la producción inicial como base de cácülo.

b) Ensayos en los que la cantidad consumida es una variable - En los casos en que el total de alimento consumido es una variable en estudio, como por ejemplo cuando se desea comparar concentrados consumidos al 70, 90 y 110 por ciento de la tasa estandard usada en el ejemplo del cuadro 1, en este caso, durante el período de estandarización todos los cálculos proceden como se mostró, pero para llegar a los valores de consumo de concentración durante el período de comparación, habrá que calcular la cantidad de T.N.D. en la primera semana del período de estandarización, como % de la última semana de consumo del período inicial, así, si la vaca N°1 recibirá el 70% del nivel en el período de comparación, entonces el nuevo consumo en la 1°semana del período de estandarización será 19.4 X 0.70 = 13.6 y si la tercera semana mostrada en la tabla pasa a ser período de comparación, entonces en la tercera semana el consumo deberá ser 13.6 X 0.759 = 13.0.

Si en el caso en discusión, se desearía hacer el ajuste sobre bases absolutas, entonces simplemente se adhiere o sustrae algunas cantidades (constante para todas las vacas bajo tratamiento) a la cantidad con la cual podría ser alimentada una vaca dada si el consumo de nutrientes no es una variable en estudio.

En experimentos que han sido conducidos usando el método propuesto por Lucas de A.E., se puede esperar que la magnitud del error experimental sea menor que aquellos experimentos donde se usa otra técnica como control de consumo.

2.2.3 Efecto de la edad

Como se anotó previamente, la edad tiene un efecto marcado sobre la tasa inicial de producción y subsecuentemente, sobre la producción de una lactación dada. En ensayos que comprenden una lactancia, la edad podría no ser un factor importante. Es conocido que vacas jóvenes exhiben una persistencia más alta que vacas viejas y si bien es cierto,que rara vez hay muchas vacas muy viejas o muchas vacas muy jóvenes dentro de un experimento, cuando el N°de animales es apreciable en cada extremo, se debe dar atención al factor edad en el experimento.

Cuando en un experimento se consideran muchas lactancias, la edad es un factor importante a considerar en el diseño, así, Bejarano (1) trabajando con cuatro grupos raciales, encontró que la edad, antes que el peso, fue el factor de mayor influencia en la producción real y relativa de los grupos raciales, encontrándo que el ganado criollo la edad tiene una influencia de 62% sobre la producción real en ganado criollo y de 48% en cruce de criollo X Jersey.

2.2.4 Influencia del estado de gestación

Es conocido que el rendimiento total de una lactancia está negativamente relacionada a la longitud del tiempo de preñez durante la lactancia. Sin embargo, se ha observado que la persistencia es escasamente afectada hasta cerca de la mitad de la preñez (5 meses después de la concepción), después de este tiempo la persistencia decrece con una tasa creciente y quizás en alguna manera esto varía entre vacas. Por lo tanto, en experimento que envuelven una sola lactancia, se deben hacer esfuerzos por usar vacas que no vayan a pasar de la mitad de la preñez al finalizar el experimento. Si eventualmente, algunas vacas podrían pasar de la mitad de la preñez, se deberá considerar la restricción correspondiente en la asignación de tratamientos o, en otro caso, el estado de gestación podría ser considerado como una covariable en el análisis de los resultados.

En hatos mantenidos con fines de experimentación, una época de monta controlada es un modo simple de controlar el efecto de preñez.

En experimentos que envuelven muchas lactaciones, el problema relativo a gestación es primero, encontrar grupos de vacas que sean lo más uniformes posibles en sus fechas de concepción, y segundo, adoptar un programa de manejo tal que la futura concepción en los animales sea lo más uniformes posibles.

2.2.5 Influencia del estado de lactación

El período normal de lactación de una vaca dentro de explotaciones comerciales es cerca de 10 a 13 meses. Si la vaca permanece vacía, sin embargo, ella podría lactar satisfactoriamente por períodos de tiempo más largos.

Cuando se planea experimentos, es ventajoso que todas las vacas que se van a usar, alcancen su pico de producción en la misma fecha, pero estas condiciones solo pueden lograrse en hatos mantenidos en condiciones experimentales.

Definitivamente, lo ideal es evaluar el efecto de un tratamiento sobre el total de la lactancia, sin embargo, esto no siempre es posible, luego, en experimentos en que solo se va a evaluar los tratamientos durante un período de la lactancia, lo ideal es iniciar el ensayo cuando los animales hallan alcanzado su pico de producción y terminar el ensayo antes de que los animales alcancen la mitad de la preñez. Al respecto, Villegas (12) quién trabajó con suplementación con niveles de banano verde a vacas lecheras en pastoreo, atribuye como una de las razones por las que no encontró diferencias significativas entre tratamientos, al hecho de que las vacas utilizadas en su experimento se encontraban en promedio, en 180 días de lactancia. anotando: "Es bien conocido que las vacas responden más eficientemente a la suplementación con concentrado, cuando se encuentran en las primeras etapas de lactancia (Burt, W.A., Laird, R. y Warker-Love, J., Pirper, R. y Holmes, W.,(citados por Villegas (12)). Pasada la etapa ascendente de la curva de lactancia, aparentemente los cambios nutricionales detrimentales producen grandes cambios en la producción de leche. En contraste, un mejoramiento nutricional podría producir cambios más tenues en dicha producción".

Cerdas (3), también realizó un experimento de suplementación de banano verde a 4 niveles (0, 0.3, 0.7 y 1.2 kg MS/100 y PV/DIA) iniciando la suplementación en 5 etapas de la lactancia (-1, 0, + 1, + 3, + 5 meses respecto al parto), resultando 20 tratamientos a compararse, las tendencias a los diferentes niveles de banano se muestran en la Figura 5. los resultados corroboran lo indicado por Villegas (12) y ponen de manifiesto la importancia de conducir el ensayo comenzado en el pico de producción. Sin embargo, las diferencias en respuesta, algunas de ellas no significativas estadísticamente nos indican también que lo ideal sería evaluar el efecto del tratamiento a lo largo de toda la lactancia.

2.2.6 Influencia de otros factores

El peso corporal. - Al respecto, Bejarano (1) evaluando 4 grupos raciales en trópico húmedo, reporta con respecto a la producción de leche en función del peso corporal, lo siguiente: "solamente en el grupo racial criollo X jersey se detectó efecto lineal y cuadrático significativo ($P \leq 0.01$) del peso al parto sobre la producción de leche. La tendencia de producción Figura 6 que se inicia con 276 kg de peso corporal y 1747 kg de leche, se incrementa progresivamente hasta la máxima producción de 2072 kg de leche a los 357 kg de peso corporal. Luego declina paulatinamente con el aumento de peso corporal. Estos resultados indican que la producción de leche en el grupo racial criollo X jersey puede incrementarse en 4 kg por cada kg de aumento de peso corporal hasta llegar al peso de máxima producción, que es 357 kg. Esto también indica que no siempre las vacas más pesadas son aquellas que más producen, puesto que a mayores pesos la producción tiende a declinar. Los mayores pesos corporales corresponden generalmente a edades mayores; por lo

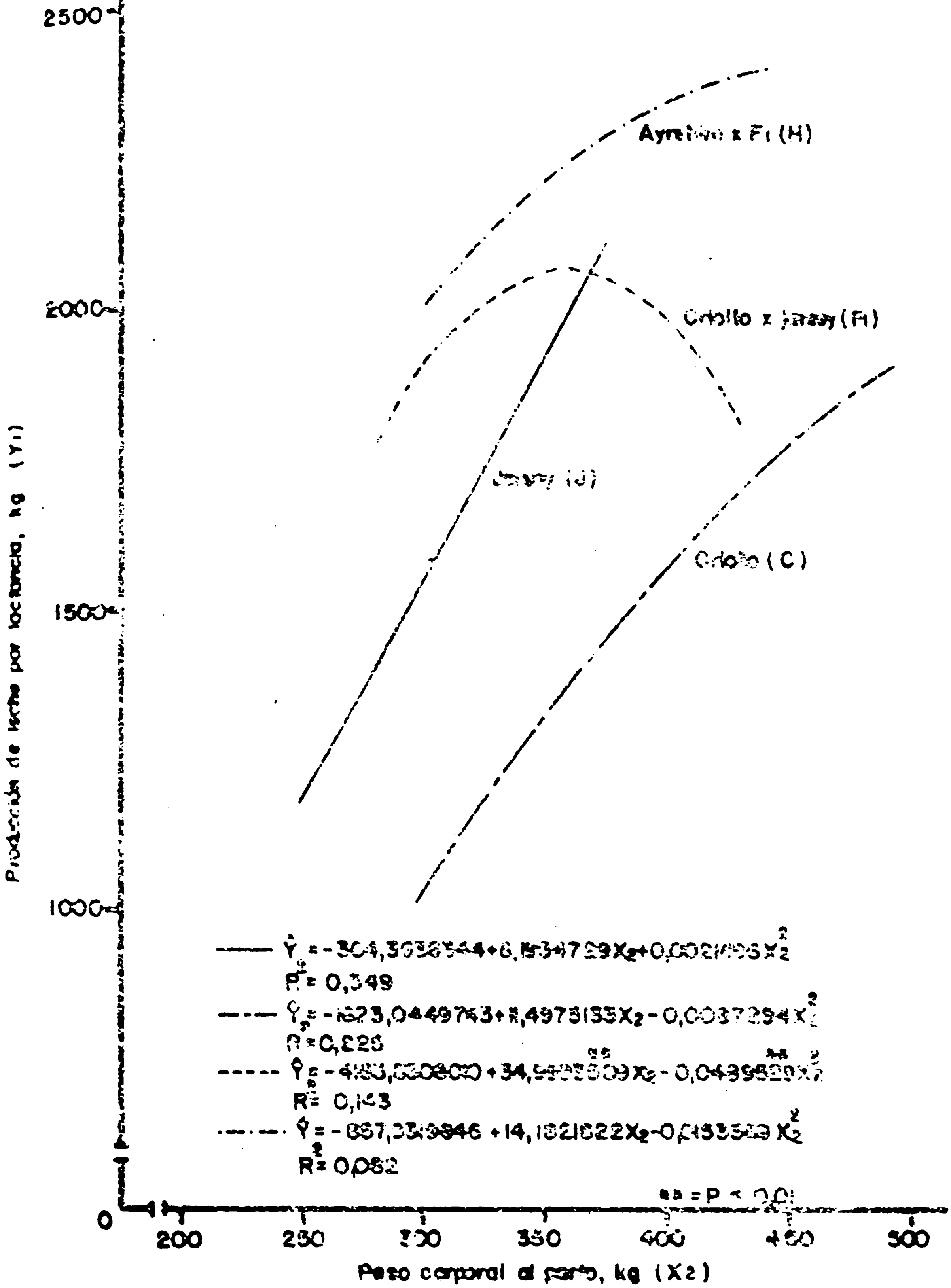


Fig. 6: Producción de leche en función del peso corporal en los distintos grupos raciales (análisis entre vacas)

Fuente: Bejarano (1)

tanto, los efectos de edad y peso están confundidos y son difíciles de separar. Incrementos de 4 kg de leche por cada kg de aumento del peso corporal son reportados por Clara y Touch Berry (citado por Bejarano (1)) en ganado Holstein.

Las tendencias de la producción de leche en función del peso en los restantes grupos raciales, cuyos análisis no dieron significativos, difieren en cada uno de los grupos raciales. En la figura 6. , puede apreciarse el efecto positivo y muy marcado en los grupos raciales puros y el menos intenso en los cruzados".

Al respecto, Lucas (6) anota que en producción de leche muy poco o nada se puede lograr, tomando en consideración el peso para distribuír los tratamientos en las unidades, cuando ya se está usando la edad como criterio de agrupamiento.

En experimentos con ganado de carne, y en algunos experimentos con ganado de leche se usa el peso como covariable, así, Ruiz (9) en un ensayo para probar la utilización de gallinazo en la alimentación de bovinos usó como covariables sobre la variable de crecimiento, la ganancia de peso postdestete y el peso al inicio del experimento, resultando ambos ajustes significativos.

La condición del animal.-Lucas indica que en casos en que se trabaje sobre la misma lactancia, no es necesario hacer consideraciones sobre la condición del animal a menos que algunas vacas se encuentren en condiciones deplorables, de ser así, estos animales no deben ser usados porque esto afectaría la curva normal de lactancia. En caso de ganado de carne, la curva de crecimiento también es afectada.

Lucas (6) indica que usando la técnica A.E. se puede disminuir la variación por efecto de la condición inicial del animal.

Período seco. - Lucas (6) indica que se ha observado que la producción inicial de leche varía directamente (aunque no linealmente) con la longitud de período seco precedente. Este comportamiento podría ser consecuencia más de un aspecto dė manejo y alimentación,que de un aspecto biológico propiamente. Los efectos de un período seco probablemente se reflejan en la condición del animal. Por lo tanto, la longitud del período seco podría no ser considerados en ensayos que comprende la misma lactancia.

En experimentos que envuelven muchas lactancias, el experimentador debería controlar la longitud del período seco entre las lactaciones experimentales, sin embargo, desde que la concepción no puede ser controlada estrictamente, el control del período seco podría ser usado para asegurar longitudes uniformes de la lactancia.

Enfermedades. - Los animales que presentan enfermedades crónicas, o que han sufrido enfermedades en un pasado cercano, no deben ser usados, a menos de que se este seguro de que la enfermedad no causará persistencias anormales o afectará la relación normal entre edad y producción. Como ejemplo, Wheelock, J.V. et al, citado por Cerdas (3), reporta que la mastitis subclínica afectan tanto la concentración de grasa como la producción de leche.

2.2.7 Criterio de producción de leche para formar grupos homogéneos

Como se anotó, la producción a diferentes tiempos en una misma lactancia tienen alta correlación $0.95 \leq r \leq 0.99$, pero la producción entre diferentes lactancias de una vaca tiene baja correlación $0.5 \leq r \leq 0.8$. Se ha encontrado Lucas (6), que cuando el rendimiento durante la lactancia inmediata a la lactancia usada en un experimento, es usada como base para agrupar vacas para el estudio de dicha lactancia, el error experimental podría ser dos veces más grande que cuando se agrupan sobre la base de dos semanas de producción justo antes del experimento.

2.2.8 Exactitud de mediciones parciales para estimar la producción total de una lactancia

Dentro de la problemática de estimar la producción total de una lactancia, así como para determinaciones de grasa y proteína, el problema es la decisión del intervalo que debe hacer entre pesadas y la frecuencia de muestras para determinaciones de grasa y proteína. En ensayos de corta duración y con un número limitado de animales, es posible hacer mediciones diarias. Sin embargo, cuando las mediciones deben hacerse sobre un gran número de animales y además, en explotaciones comerciales, el período entre pesadas, por ejemplo, puede ser semanal, bisemanal, mensual o bimensual, dependiendo de la disponibilidad de tiempo, personal y equipo. Al respecto Lindstron (5) reporta que, estudios realizados en Europa y en América del Norte han demostrado que, para efectos de mejoramiento genético, registrar la leche una vez al mes resulta suficientemente exacto para la selección individual y una vez cada dos meses para la prueba de progenie. En el Cuadro 3 se dan algunos resultados obtenidos en un estudio de 10 hatos en Kenya. El error medio es pequeño: incluso para la prueba bimensual solo es de aproximadamente 1.5 por ciento del promedio real. No obstante, debido a las amplias desviaciones típicas los errores individuales para la prueba bisemanal son del orden de -9 a +12 por ciento del rendimiento real; por ejemplo -251 a 337 kg. La prolongación del intervalo entre controles a un mes aumenta el error de $\pm$ 400 a $\pm$ 500 kg. Por consiguiente, parecería que en Kenya es preferible no exceder de un intervalo de 14 días para los controles si se quiere mantener una exactitud relativamente grande en la selección de vacas individuales. En lo que se refiere a la toma de muestras para determinaciones de la grasa y la proteína, varios estudios, (McDaniel, citado por Lindstrom (5)), han demostrado que los registros mensuales, en general (por lo menos en cuanto al porcentaje de grasa), son ligeramente menos exactos que los registros mensuales de los rendimientos lecheros. A fines de selección, sin embargo, son los suficientemente fiables aunque bien convendría, de ser posible, registros más frecuentes.

Cuadro 3. Exactitud de los rendimientos estimados por lactación de 1097 vacas de cuatro razas europeas y de la raza Sahiwal obtenidos en 10 explotaciones de Kenya, 1969-72

Registro	Correlación con el rendimiento real (dentro de la explotación y temporada de pastos)	Diferencia media: estimada menos rendimiento real = d		95% del intervalo de seguridad para los rendimientos individuales estimados - d = 1.96 X desviación típica	
		kg	%	kg	%
Semanal	0.994	39	1.41	-201 a +279	-7.3 a +10.1
Bisemanal	0.991	43	1.55	-251 a +337	-9.1 a +12.2
Mensual	0.978	37	1.54	-422 a +496	-15.2 a +17.9
Bimensual	0.943	42	1.52	-689 a +773	-24.9 a +27.9

Los rendimientos reales se basaban en las pesadas diarias de la leche......Porcentaje del rendimiento medio real (2 767 kg).

Fuente: Lindstron (5)

2.2.9 Formas adicionales de controlar el error experimental

Existen algunos aspectos de manejo de un experimento con animales, que ayuda a controlar el error experimental, así:

1) Si el ensayo es en estabulación, mantener los animales bajo galpones iguales en las mismas condiciones a lo largo de todo el experimento.

2) Mantener el mismo orden tanto en la asignación de raciones como en el ordeño a través de todo el experimento. Ejemplo, dadas las vacas numeradas del 1, 2, 3, 4 y 5 las cuales son alimentadas en el orden 5, 2, 1, 3, 4 en el primer día del período de estandarización, entonces se debe mantener el mismo orden a lo largo de todo el experimento.

3) Mantener el máximo de cuidado en la pesada de los alimentos y chequear el consumo de los mismos ya que las fluctuaciones en el consumo de alimentos podría reflejarse algunas veces en el rendimiento..

4) Tratar de mantener la vaca ordeñada por la misma persona durante el experimento.

3. Diseños experimentales en producción animal bovina

Como es conocido, cada diseño experimental llámese irrestricto al azar, bloques completos, cuadrado latino, etc. Exigen un N° mínimo de repeticiones por tratamiento, de estos, el más flexible en cuanto a N°de repeticiones es el diseño irrestricto al azar. El N° de repeticiones es una de las formas más eficientes de controlar el error experimental, sin embargo, este es un requisito difícil de llenar en producción animal, otra forma es el control apropiado de fuentes de variación, sin embargo, esta segunda forma de controlar el error experimental implica el uso de diseños más completos y un número mayor de repeticiones. Existe una forma de controlar el error experimental que no necesariamente implica un aumento del N° de repeticiones, nos referimos al empleo del análisis de covarianza. Por ello el uso de covarianza es muy difundido en diseños experimentales con animales y las covariables más usadas con peso al destete o peso al inicio del experimento cuando se trabaja con ganado de carne, con ganado de leche se usa la producción de leche al inicio del experimento. El nivel de consumo de alimentos también es una covariable que se usa en ganado de carne como en ganado de leche. Los requisitos que debe reunir una covariable son:

Estar altamente correlacionada con la variable de respuesta y de preferencia que la variable no sea afectada por los tratamientos; al respecto Lucas (6) reporta que la medición del nivel de producción antes de iniciar el ensayo es de primera importancia, estas medidas son la mayoria de las veces efectivas en reducir el error experimental.

También con la finalidad de aumentar el N° de grados de libertad del error, se idearon ensayos en los que un animal proporcionaba más de una observación o repetición durante el ensayo. En estos diseños, denominados Diseños de Reversión o Sobre Cambio, un animal recibe más de un tratamiento a lo largo del ensayo. El uso de estos diseños también es muy difundido en ensayos con animales.

En lo que resta del presente documento, se explicará el uso de covarianza en diseños contínuos (en los que el animal recibe el mismo tratamiento en todo el ensayo) y el uso de diseños de reversión más comunes.

3.1 Uso de covarianza en diseños experimentales contínuos

Concepto de análisis de covarianza

Para realizar análisis de covarianza se toma en los experimentos, además de los datos propios del estudio que se realiza, otros datos concomitantes con ellos. Así, la producción de leche al inicio de la lactancia es concomitante con la producción total, los pesos iniciales de los animales al comienzo de un experimento son concomitantes con los pesos finales.

En el análisis de covarianza se combinan los análisis de varianza y el de regresión. El análisis de covarianza es posible aplicarlo a los más diversos diseños experimentales, incluso cuando se tiene desigual número de unidades por tratamiento y cuando se dispone de sub-unidades por unidad.

Los usos más importantes de la covarianza, Steel y Torrie (11) son:

- Disminuir el error experimental con el consiguiente aumento en la precisión del experimento.
- Ajustar los promedios de los tratamientos por la diferencia entre los promedios de la variable independiente.
- Hacer una mejor interpretación de los resultados de los experimentos, especialmente en cuanto se relaciona con la naturaleza de los efectos de los tratamientos.

Para que el análisis de covarianza sea válido, es necesario aceptar o asumir que se cumple una serie de requisitos que le dan validez al análisis, Calzada (2). Estas asunciones son:

- Que las X's (covariable) no son afectadas por los tratamientos

- Tanto la variable X como la variable Y (dependiente), deben tener varianzas homogeneas
- La variable X y Y deben tener distribución normal.
- La regresión de Y sobre X debe ser lineal
- El error debe estar normal e independientemente distribuído con cero de promedio.

Las asunciones arriba indicadas están relacionadas con los términos de las ecuaciones de los modelos lineales aditivos del análisis de varianza de los diseños experimentales, a los que se agrega un término más que corresponde a la regresión entre la variable dependiente y concomitante. Así, para el diseño Completamente Randomizado, la descripción matemática del modelo aditivo lineal analizado con la covarianza está dado por la siguiente ecuación:

$$Y_{ij} = \mu + \tau i + \beta\,(X_{ij} - \bar{X}...) + E_{ij}$$

La ecuación anterior se puede escribir en la siguiente forma:

$$Y_{ij} - \beta(X_{ij} - \bar{X}..) = \mu + \tau i + E_{ij}$$

En esta forma de Y se elimina el efecto de la diferencia $(X_{ij} - \bar{X}..)$, lo que constituye un valor de Y ajustado por el efecto de la covariable X. La inclusión de la fuente de variación $\beta(X_{ij} - \bar{X}..)$ en el modelo, permite reducir el error experimental.

Es necesario anotar que dentro del diseño se puede incluir simultaneamente, más de una covariable.

3.2 Medición de covariables en ensayos con bovinos

En experimentos con vacas lecheras, la medición del nivel de producción de leche, antes de comenzar un ensayo es de primera importancia, Lucas (6). Esta medición resulta efectiva en la reducción del error experimental y podría ser hecha después de que todas las vacas alcanzaron el pico de producción y durante un período inmediatamente precedente a la colocación de las vacas dentro de los tratamientos a ser comparados. De acuerdo a la forma de manejo de aspectos nutricionales en un ensayo, propuesto por Lucas (6) denominado Alimentación Equitativa. Lucas recomienda que la medición de la covariable se realice durante el período de estandarización.

Otra covariable que puede medirse con ganado de leche, es el estado de gestación, en el caso que, dentro del ensayo se usen animales que hallan pasado la mitad del período de gestación. En este caso se usará un análisis de covarianza con dos covariables (covarianza múltiple).

En experimentos en que se evalúa ganancia de peso, se puede medir como covariable, el peso al nacimiento, el peso al destete o el peso al inicio del experimento. Los pesos iniciales producen ajustes significativos Ruiz (9).

En experimentos en los que el consumo es ad-libitum, se acostumbra a estimar la cantidad consumida para usarla como covariable, Steel y Torrie (11). En otros casos se mide como covariable el valor nutritivo de las raciones utilizadas. Algunas veces se mide tanto el consumo como el valor nutritivo. Estas covariables permiten comparar medias de ganancia de peso ajustadas a un consumo común.

3.3 Análisis de covarianza para el diseño irrestricto al azar en ensayos con bovinos

El diseño irrestricto al azar puede ser usado solamente cuando los animales que serán usados en el experimento son relativamente uniformes con respecto a aquellos factores (raza, edad, lactancia, peso inicial, estado de gestación), que influencia el rendimiento de las variables que se van a estudiar. Así, para un experimento con vacas de leche, se requiere:

- Que las vacas sean todas de la misma raza.
- Que no exista vacas ni extremadamente jóvenes, ni extremadamente viejas.
- Que exista un número suficiente de vacas al mismo tiempo disponibles para el ensayo.
- Que todas las vacas se encuentren antes de la mitad del período de gestación, o en caso de gestación avanzada, que todas sean uniformes al respecto.
- Que tengan niveles iniciales de producción uniforme.

En este caso, el requisito de uniformidad en producción inicial, podría no ser necesario ya que la variación debida a este factor podría ser controlada por covariancia, sin embargo, si se encuentra que el grupo puede ser dividido en vacas de alta producción y de baja producción, ya que se observa diferencias extremas entre los animales, es preferible usar técnicas de bloqueo para controlar esta fuente de variación.

Con respecto al estado de gestación, se podría usar covarianza para controlar esta fuente de variación, así, a aquellas vacas que no han alcanzado la mitad de la preñez, se les puede asignar el valor cero, a aquellas que han pasado en un mes la mitad de la gestación se les asigna valor 1, vacas con dos meses pasada la mitad de gestación se les asigna valor 2 y así sucesivamente. Estos valores constituyen los valores de la covariable.

Obviamente, el ajuste por covarianza funcionará, si las restantes condiciones que impone el modelo se cumplen, de lo contrario, es mejor usar criterios de bloqueo.

3.4 Análisis de covarianza en diseño en bloques al azar en ensayos con bovinos

El diseño en bloques al azar es aquel en el cual los animales son divididos en grupos o bloques de individuos similares o individuos bajo condiciones ambientales similares y dentro de cada grupo o bloque, los tratamientos se distribuyen al azar sobre los animales. Cada tratamiento debe estar presente en cada bloque. El bloqueo puede ser hecho con respecto a uno o varios factores simultaneamente. En este caso el agrupamiento tiene los siguientes objetivos:

- Reducir el error experimental
- Estimación y prueba de posibles interacciones entre tratamientos con factores tales como raza, edad, niveles de producción, etc.

Cuando el agrupamiento tiene objeto de calcular interacciones, se podría complementar utilizando los conceptos de arreglos factoriales.

Con el objeto de facilitar la interpretación, es necesario que el N° de animales, en cada bloque sea el mismo múltiplo del N° de tratamientos.

De acuerdo a lo anotado anteriormente, con respecto al diseño irrestrico al azar se tiene que, para el caso de usar covarianza tomando el ejemplo en ganado de leche, si por ejemplo, los animales no son de la misma raza, formaríamos grupos por razas y consideraríamos las mismas covariables (producción incial y estado de gestación) y se realizaría el analisis de covarianza en un diseño en bloques al azar.

3.5 Procedimiento de análisis de covarianza usando SAS

El procedimiento de análisis de covarianza en forma manual, es explicado en detalle por Steel y Torrie (11) y por Calzada (2). En el presente documento, ilustraremos dos ejemplos de análisis de covarianza realizados en un computador IBM-370, serie 3341, utilizando del sistema S.A.S., el procedimiento General Lineal Model (PROG GLM). SAS USER'S GUIDE (10).

En el cuadro 4 se presenta la información tomada de Lucas (6), los datos simulan un experimento con 4 tratamientos y 3 bloques, asumiendo que X es la producción de leche corregida durante el período de estandarización de 3 semanas y Y es producción de leche corregida durante el período de comparación, 14 semanas.

En el cuadro 5 se muestra la información del cuadro, codificada adecuadamente para ser procesada por SAS. En donde la primera columna es el número de observación, la segunda codifica bloques, la tercera codifica tratamientos, la cuarta es la covariable X y la quinta es la variable de rendimiento de leche Y.

En el cuadro 6 se observa la primera página de salida SAS, la que es común a cualquier procedimiento de análisis de varianza y nos sirven para verificar si la codificación fue adecuada, así como para verificar si los grados de libertad en el análisis de varianza corresponden a las clases del modelo que quisimos construir para el análisis. Así, clases nos indica las fuentes de variación del modelo que no son error ni total. Cuando se usa covarianza, el procedimiento no considera la covariable como una clase, por eso no aparece en el cuadro. Seguidamente Levels indica el número de niveles de cada clase, así, tenemos 3 bloques, lo que equivale a tres niveles para bloques, tenemos 4 tratamientos, lo que equivale a 4 niveles para tratamientos. Por útlimo Values indica los valores que se usaron para codificar los niveles de cada clase dentro de la matriz de datos del procedimiento.

En el cuadro 7, que es la segunda hoja de salida de SAS, corresponde al análisis de covarianza, en este listado encontramos la información referente a la prueba de significación de cada uno de los componentes del modelo, en este caso bloque, tratamientos (trat) y la covariable (X), marcado con un 1, observamos los valores de Pr>f, el cual se interpreta considerando que cuanto más pequeño es el valor, tanto mas significativo es el efecto del componente en el modelo, en este caso tanto a bloque, como a trat. y a X le corresponde un valor de 0.0001, el cual es menor que 0.01, por lo tanto, los tres componentes o fuentes de variación son altamente significativos. Todos los componentes son probados contra el cuadrado medio del error 2.

En la siguiente página del listado SAS, cuadro 8 aparecen los resultados de la comparación de medias ajustadas de la variable Y por el efecto de la covariable X (LEAST SQUARES MEANS), en las primeras 6 líneas se encuentra la comparación de medias de bloques ajustadas, y en las siguientes la comparación de medias de tratamientos ajustadas. Las primeras 4 columnas dan la información referente a cada media ajustada, así en la primera columna para bloques, se indica a que bloque 1, 2, ó 3 se refiere cada promedio. En la segunda columna (Y Lsmean) se indica el valor de la media en cada bloque ajustada por la covariable, en la tercera columna (STD ERRLSMEAN) se indica la desviación estandard estimada de cada media ajustada. En la cuarta columna (PROB > ITI HO: LSMEAN = 0) se indica el resultado de probar la hipótesis de que la media ajustada es igual a cero, usando el estadístico "t" de Studend como comparador. La interpretación del valor numérico que aparece en las siguientes líneas, es similar a los valores de Pr > F explicado para la página anterior del listado, es decir, un valor que sea menor de 0.05 y mayor de 0.01, indicará que la media difiere significativamente de cero y un valor menor o igual que 0.01 indica que la diferencia de la media de cero es altamente significativa. Esto, dentro del contexto usual de presentar resultados con niveles significativos al 0.05 y altamente significativos al 0.01. En el ejemplo, todas las medias tienen diferencias de cero altamente significativas. En lo que resta de columnas, se presenta la comparación de las medias ajustadas, entre sí. (PROB > ITI HO: LSMEAN (i) = LSMEAN (J)) en una tabla de contingencia, el contenido de la tabla son valores de probabilidad cuya interpretación es similar al de Pr > F y Pr > ITI. Así para la comparación de la media ajustada del bloque 1 con respecto a la media ajustada del bloque 2 encontramos que el valor es de 0.0005, como el valor es menor que 0.01, indica que la medias ajustadas del bloque 1 con respecto al bloque 2 tienen diferencias altamente significativas. Para el caso de la media ajustada del bloque 1 con respecto al bloque 3, el valor es 0.0405, como este valor es mayor que 0.01 pero menor que 0.05, indica que la media del bloque 1 con respecto a la media del bloque 3 tiene diferencias significativas solamente.

Para tratamientos, el listado es similar y la interpretación también. Con respecto a la nota (note), esta es válida cuando se usa pruebas de contrastes, en este caso, no debe considerarse.

En forma adicional al análisis de covarianza, se hizo que la máquina calcúle los promedios de los tratamientos sin ajustar, cuadro 9. Obsérvese que tanto el valor de la media ajustada como la media no ajustada, así como la desviación estandard de las medias difieren en magnitud, así para el tratamiento 1 tenemos que la media ajustada es 39.2534, con una desviación estandard de media igual a 0.0070, mientras que la media sin ajustar es 42.7666 con una desviación estandard de media igual a 3.7205. Las diferencias entre las medias ajustadas y sin ajustar, así como las diferencias entre su desviación estandard son una forma de entender la forma como el uso de una covariable adecuada puede permitir realizar comparaciones de medias más precisas, ya que se realiza un mejor control del error experimental.

En el cuadro 10, se presenta el análisis de regresión de la variable Y con su covariable X, se hizo con la finalidad de estimar el coeficiente β del modelo, en este caso es igual a 0.71895, este valor es el que permite ajustar cada valor de Y, usando la fórmula $Y - \beta(Xi - \bar{X}..)$. Con estos valores se podría calcular manualmente las medias ajustadas y su respectiva desviación estandard.

El cuadro 11 corresponde al análisis de covarianza de los datos del ejemplo anterior, pero considerando una fuente de variación más que es la interacción bloque X tratamiento, el mismo que resultó significativo ($Pr > F = 0.0302$) si se observa los valores de medias ajustadas ya no son los mismos que en el caso que no se consideró la interacción. Al respecto, el SAS es muy flexible en la construcción de modelos y la presencia de una o más covariables no dificulta en nada el análisis. A diferencia de un proceso manual en la que tanto el hecho de agregar una covariable; así como el hecho de cambiar de diseño, hace muy dificultoso el análisis de covarianza.

CUADRO 4. DATOS DE UN DISEÑO EN BLOQUES AL AZAR DONDE X PRODUCCION DE LECHE EN EL PERIODO DE COMPARACION y Y COVARIABLE PRODUCCION DE LECHE EN EL PERIODO DE ESTANDARIZACION.

Tratamiento / Variable	1		2		3		4		Suma de bloques	
	X	Y	X	Y	X	Y	X	Y	X	Y
Bloque 1	64.9	51.3	45.5	41.9	59.8	48.4	65.0	52.6		
	41.6	34.5	44.9	36.9	53.5	44.5	37.3	24.1		
	64.5	59.4	40.2	36.5	35.1	28.8	31.3	27.5		
Suma	171.0	145.2	130.6	115.3	148.4	121.7	133.6	104.7	553.6	486.9
Media	57.00	48.40	43.53	38.43	49.47	40.57	44.53	34.90	46.13	40.58
Bloque 2	56.1	43.6	37.6	19.6	74.3	55.8	67.2	47.4		
	89.9	57.7	62.2	47.6	33.3	27.1	45.0	33.7		
	57.7	41.3	38.1	27.0	33.3	23.3	42.9	33.6		
	50.0	36.7	40.8	32.7	38.7	24.5	53.3	40.8		
Suma	253.7	179.3	178.7	126.9	179.6	130.7	208.4	155.5	820.4	592.4
Media	63.42	44.82	44.68	31.72	44.90	32.68	52.10	38.88	51.28	37.02
Bloque 3	40.5	30.6	22.6	20.6	29.9	23.4	46.6	35.4		
	42.1	29.8	31.4	24.5	38.8	26.8	29.8	24.1		
Suma	82.6	60.4	54.0	45.1	68.7	50.2	76.4	60.5	281.7	216.2
Media	41.30	30.20	27.00	22.55	34.35	25.10	38.20	30.25	35.21	27.02
Tratamientos Suma	507.3	384.9	363.3	287.3	396.7	302.6	418.4	320.7	1685.7	1295.5
Media	56.37	42.77	40.37	31.92	44.08	33.62	46.49	35.63	46.82	35.99

Fuente: Lucas (6)

CUADRO 5. INFORMACION CODIFICADA PARA ANALISIS DE COVARIANZA EN BLOQUES USANDO SAS

OBS	BLOQUE	TRAT	X	Y
1	1	1	54.9	51.3
2	1	1	41.6	34.5
3	1	1	64.5	59.4
4	2	1	56.1	43.6
5	2	1	39.9	57.7
6	2	1	57.7	41.3
7	2	1	50.0	35.7
8	3	1	40.5	30.6
9	3	1	42.1	29.8
10	1	2	45.5	41.9
11	1	2	44.9	36.9
12	1	2	43.2	35.5
13	2	2	37.6	13.6
14	2	2	62.2	47.6
15	2	2	35.1	27.0
16	2	2	40.8	32.7
17	3	2	22.6	20.6
18	3	2	31.4	24.5
19	1	3	59.8	48.4
20	1	3	53.5	44.5
21	1	3	35.1	28.3
22	2	3	74.3	55.8
23	2	3	33.3	27.1
24	2	3	33.3	23.3
25	2	3	38.7	24.5
26	3	3	29.9	23.4
27	3	3	38.8	26.8
28	1	4	65.0	52.8
29	1	4	37.3	24.1
30	1	4	31.1	27.8
31	2	4	67.2	47.4
32	2	4	45.0	33.7
33	2	4	42.9	33.6
34	2	4	53.3	40.8
35	3	4	46.6	36.4
36	3	4	29.8	24.1

CUADRO 6. PRIMER LISTADO SAS DE ANIMALES DE COVARIANZA. INFORMACION DE CLASES, NIVELES Y CODIFICACION

S T A T I S T I C A L   A N A L Y S I S   S Y S T E M   9:01 TUESDAY, AUGUST 9, 198[illegible]

GENERAL LINEAR MODELS PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
BLOQUE	3	1 2 3
TRAT	4	1 2 3 4

NUMBER OF OBSERVATIONS IN DATA SET = 36

CUADRO 7. SEGUNDO LISTADO SAS DE ANALISIS DE COVARIANZA. INFORMACION RESPECTO A LA SIGNIFICACION DE COMPONENTES DEL MODELO Y COVARIABLES

S T A T I S T I C A L   A N A L Y S I S   S Y S T E M   9:01 TUESDAY, AUGUST 9, [illegible]

GENERAL LINEAR MODELS PROCEDURE

[illegible] VARIABLE: Y

[illegible]	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V
[illegible]	6	[illegible]	674.97[illegible]591	57.04	0.0001	0.921896	[illegible]
[illegible]	29	[illegible]	11.[illegible]314[illegible]380		STD DEV		Y MEAN
[illegible] TOTAL	35	4392.[illegible]			3.43[illegible]385		3[illegible].9[illegible]1111

(1)

[illegible]	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
[illegible]	2	[illegible]	38.56	0.0001	2	207.9[illegible]804042	8.79	0.0010
[illegible]	3	613.83[illegible]4444	17.2[illegible]	0.0001	3	5.10641794	0.14	[illegible]
X	1	2523.04223546	213.25	0.0001	1	2523.04223546	213.25	0.000[illegible]

CUADRO 8. TERCER LISTADO SAS EN ANALISIS DE COVARIANZA. INFORMACION DE PRUEBAS DE DIFERENCIA DE MEDIAS AJUSTADAS

S T A T I S T I C A L   A N A L Y S I S   S Y S T E M   9:01 TUESDAY, AUGUST 9, 1983

GENERAL LINEAR MODELS PROCEDURE

LEAST SQUARES MEANS

LUGAR	Y LSMEAN	STD ERR LSMEAN	PROB > |T| HO:LSMEAN=0	PROB > |T| HO: LSMEAN(I)=LSMEAN(J) I/J	1	2	3
1	39.2534123	0.9970638	0.0001	1	.	0.0003	0.0366
2	33.7727996	0.8882944	0.0001	2	0.0003	.	0.3127
3	35.5117820	1.3474457	0.0001	3	0.0366	0.3127	.

NOTE: TO ENSURE OVERALL PROTECTION LEVEL, ONLY PROBABILITIES ASSOCIATED WITH PRE-PLANNED COMPARISONS SHOULD BE USED.

TRAT	Y LSMEAN	STD ERR LSMEAN	PROB > |T| HO:LSMEAN=0	PROB > |T| HO: LSMEAN(I)=LSMEAN(J) I/J	1	2	3	4
1	35.9865554	1.2217628	0.0001	1	.	0.6423	0.9256	0.9600
2	36.8353766	1.2297693	0.0001	2	0.6423	.	0.5400	0.6472
3	35.8231993	1.1804407	0.0001	3	0.9256	0.5400	.	0.6794
4	36.0721943	1.1632987	0.0001	4	0.9600	0.6472	0.6794	.

NOTE: TO ENSURE OVERALL PROTECTION LEVEL, ONLY PROBABILITIES ASSOCIATED WITH PRE-PLANNED COMPARISONS SHOULD BE USED.

CUADRO 9. LISTADO SAS SOBRE PROMEDIOS DE TRATAMIENTOS SIN AJUSTAR. LISTADO ADICIONAL AL ANALISIS DE COVARIANZA

S T A T I S T I C A L   A N A L Y S I S   S Y S T E M   15:43 TUESDAY, AUGUST 9, 1983

VARIABLE	N	MEAN	STANDARD DEVIATION	MINIMUM VALUE	MAXIMUM VALUE	STD ERROR OF MEAN	SUM	VARIANCE	C.V.
TRAT=1									
Y	9	42.76666667	11.16154111	29.80000000	59.40000000	3.72051370	384.90000000	124.58000000	26.099
TRAT=2									
Y	9	31.92222222	9.68914054	19.50000000	47.50000000	3.22971351	287.30000000	93.87944444	30.352
TRAT=3									
Y	9	33.62222222	12.42555610	23.30000000	55.50000000	4.14185203	302.60000000	154.39444444	36.956
TRAT=4									
Y	9	35.63333333	9.97308879	27.10000000	52.80000000	3.32436293	320.70000000	99.46250000	27.988

CUADRO 10. LISTADO SAS SOBRE ANALISIS DE REGRESION ENTRE LA VARIABLE Y y X PARA ESTIMAR EL COEFICIENTE β

S T A T I S T I C A L   A N A L Y S I S   S Y S T E M   9:01 TUESDAY, AUGUST 9, 1983   3

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	3836.56128847	3836.56128847	234.69	0.0001	0.873462	11.2353
ERROR	34	555.80176708	16.34711080		STD DEV		Y MEAN
CORRECTED TOTAL	35	4392.36305556			4.04315694		35.98611111

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
X	1	3836.56128847	234.69	0.0001	1	3836.56128847	234.69	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	2.32123116	1.01	0.3197	2.29848627
X	0.71095094	15.32	0.0001	0.04692980

PAGINAS DE SALIDA EN 1

GENERAL LINEAR MODELS PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
BLOQUE	3	1 2 3
TRAT	4	1 2 3 4

NUMBER OF OBSERVATIONS IN DATA SET = 36

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	12	4108.47510267	342.37292522	27.74	0.0001	0.935368	9.7328
ERROR	23	283.88795289	12.34295447		STD DEV		Y MEAN
CORRECTED TOTAL	35	4392.36305556			3.51325411		35.98611111

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
BLOQUE	2	912.37555556	36.96	0.0001	2	207.71196513	8.41	0.0018
TRAT	3	613.83194444	16.58	0.0001	3	8.39466652	0.23	0.8768
BLOQUE*TRAT	6	213.76222222	7.89	0.0302	6	59.22536721	0.80	0.5801
[illegible]	1	2368.50538045	191.89	0.0001	1	2368.50538045	191.89	0.0001

GENERAL LINEAR MODELS PROCEDURE

LEAST SQUARES MEANS

TRAT	Y LSMEAN	STD ERR LSMEAN	PROB > |T| HO:LSMEAN=0	PROB > |T| HO: LSMEAN(I)=LSMEAN(J) I/J	1	2	3	4
1	35.7962634	1.2742432	0.0001	1	.	0.5961	0.8411	0.9802
2	37.0207631	1.2954502	0.0001	2	0.5961	.	0.4315	0.5826
3	35.6275797	1.2301105	0.0001	3	0.8411	0.4315	.	0.8130
4	35.0410541	1.2228853	0.0001	4	0.9802	0.5826	0.8130	.

3.6 Diseños de sobre cambio en Producción Animal

Siempre bajo la consideración de que en Producción Animal, sobre todo en especies mayores como bovinos, existe la limitante del número de animales, se pensó en la posibilidad de que un animal recibiera más de una tratamiento durante el período de comparación, esto permite obtener mayor número de repeticiones que las que se tendría en un ensayo contínuo como los expuestos anteriormente.

Los diseños en los que la unidad experimental recibe mas de un tratamiento durante el ensayo se denomina Diseño de Sobre Cambio o Diseño de Reversión (Change-Over Trials). En este caso el período experimental se subdivide en 2 o más períodos.

3.6.1 Uso de covarianza en diseños de sobre cambio

En ensayos contínuos, los tratamientos se distribuyen a los animales, después de que estos han pasado por un período de estandarización. La alta correlación existente entre tasas de producción a diferentes tiempos se puede tomar como una ventaja para reducir el error experimental, usando la técnica de covarianza, como ya habíamos visto. Con el procedimiento indicado se puede sustraer del error experimental la mayor parte de la variación entre individuos. En diseños de sobre cambio, desde que dos o más tratamientos son contrastados en el mismo animal, la varianza entre animales en lo que se refiere a niveles de producción no entra en el error experimental. De este modo, el uso de covarianza en diseños de sobre cambio no es necesario.

3.6.2 La tendencia de la curva de lactancia y los diseños de sobre cambio

Al igual que en los diseños contínuos, cuando se use diseños de sobre cambio, el período de comparaciones debe comenzar cuando los animales hallan alcanzado el pico de producción.

Como ya se anotó, la curva de lactancia tiene una tendencia característica, la cual puede tomar cursos irregulares debido a cambios intensos en factores ambientales o nutricionales. Cuando se usan diseños de sobre cambio en ganado de leche, se debe tener especial cuidado en que el efecto de los tratamientos sean independientes de la tendencia de la curva, dicho en otras palabras el diseño de sobre cambio debe proveer un control sobre la tendencia de la curva, el mismo que se denomina "Efecto de períodos". Separando este efecto, podemos comparar el efecto de tratamientos independientes de las variaciones en producción por efecto intriseco de la tendencia de la curva.

Igualmente en los diseños de sobre cambio, podría ocurrir que un tratamiento tenga influencia sobre el tratamiento que se aplique en el siguiente período, a este efecto se le denomina "Efecto residual". Calzada (2) reporta que estos efectos pueden ser notables, de ligera o ninguna importancia. Al respecto Lucas (6) indica que la magnitud del efecto residual podría estar influenciada considerablemente por el método adoptado para controlar el alimento consumido. El método usual de ajustar el consumo individual de concentrado consumido de acuerdo a la producción propia del animal en un período precedente, resultaría en efecto residuales que podrían ser mayores que el efecto del tratamiento mismo. Sin embargo, el uso del método de alimentación equitativa, podría eliminar el efecto residual a un nivel por debajo del efecto de los tratamientos. Se recomienda también, como una forma de reducir el efecto residual, eliminar los datos de los primeros 7 a 14 días de cada secuencia.

Tipos de diseño de Sobre Cambio

Los diseños de sobre cambio más usados en ensayos con animales, son los diseños de sobre cambio simple y los diseños de sobre cambio doble, en el presente documento se describirá ambos diseños incluyendo las fórmulas de cálculo manual, ya que el volumen de información que se maneja en estos diseños es fácil de operar manualmente.

3.6.3 Diseño sobre cambio simple

Este diseño es definido como aquel en el que se prueban dos tratamientos y cada animal recibe el tratamiento una vez. Así, si se trata de seleccionar dos sistemas de racionamiento, se empieza por escoger animales representativos de la población y a la mitad de ellos, escogidos al azar, se le somete al racionamiento A y a la otra mitad al racionamiento B durante un primer período apropiado, una vez determinada las observaciones de las variables de respuesta, a cada animal se le cambia el racionamiento para comenzar el segundo período, así, el animal que reciba A recibe B y el que recibía B recibe A en este período, cuya duración es igual al anterior y en el que se toman las mismas observaciones. Si el número de animales es de 8, cada animal dará 2 datos. Según esto el análisis de varianza se ajusta al siguiente esquema:

	Animales							
	1	2	3	4	5	6	7	8
Período I	B	A	B	B	A	B	B	A
Período II	A	B	A	A	B	B	A	B

Las fuentes de variación y grados de libertad del análisis de varianza son:

Fuentes			G.L
Columnas	(r-1)	=	7
Períodos	(t-1)	=	1
Tratamientos	(t-1)	=	1
Error	(t-1) (r-1)	=	6
Total	tr-1	=	15

Como ejemplo numérico, usaremos los siguientes datos, tomados de Calcazada (2).

Períodos (Bloques)	UNIDADES EXPERIMENTALES (Columnas) 1	2	3	4	5	6	TOTALES
I	A 2	B 3	A 4	B 8	B 3	A 2	22
II	B 4	A 3	B 6	A 5	A 4	B 5	27
TOTALES	6	6	10	13	7	7	49

Total A = 20 Total B = 29

T.C $= 49^2 2/12 = 200.1$

S.C Columnas $= 1/2\ (6^2 + 6^2 + \ldots..7^2) - 200.1 = 19.4$

S.C Bloques $= 1/6\ (22^2 + 27^2) - 200.1 = 2.1$

S.C Tratamientos $= 1/6\ (20^2 + 29^2) - 200.1 = 6.7$

S.C Total $= (2^2 + 3^2 + \ldots..5^2) - 200.1 = 32.9$

ANALISIS DE VARIANZA

Fuentes	S.C	G.L	C.M	Significación
Columnas	19.4	5	3.88	N.S
Bloques	2.1	1	2.10	N.S
Tratamientos	6.7	1	6.70	N.S
Error	4.7	4	1.18	N.S
TOTAL	32.9	11	-	

Como se puede observar, no se detectó diferencias signficativas para tratamientos.

Si se quisiera estimar los límites de confianza dentro de los cuales se encuentra la verdadera diferencia ($\mu \bar{d}$) entre dos tratamientos, usaríamos la fórmula $\bar{d} \pm Sd\ (t_{\alpha})$, reemplazando los datos del ejercicio tendríamos:

$$\frac{(29-20)}{6} \pm (\sqrt{2(1.18)|6})\ (2.776)$$

donde $\bar{d} = \frac{21-20}{6}$ = promedio de diferencia de medias

$S\bar{d} = \frac{\sqrt{2C.M\ Error}}{r} = \sqrt{2(1.18)|6}$ = Desviación estandard de diferencia de medias, usando el cuadrado medio del error

$t(0.05) = 2.776$ Valor de T tabular con $\alpha = 0.05$ y 6 grados de libertad de error.

realizando los cálculos tenemos:

$$1.50 - 1.74 \leq \mu_{\bar{d}} \leq 1.50 + 1.74$$

3.6.4 Diseño sobre cambio doble

Este diseño se emplea principalmente cuando se tienen 2 tratamientos por estudiar y consiste en aplicar los tratamientos en dos secuencias diferentes, para lo cual es necesario usar tres o cuatro períodos, en el siguiente esquema, se muestra el caso en que cada animal recibe ambos tratamientos, uno de ellos dos veces en la sequencia 1, 2, 1 ó 2, 1, 2.

Períodos de comparación	Secuencia de grupo 1	2
1	1	2
2	2	1
3	1	2

Para el caso de experimentos con vacas lecheras, por este método se elimina del error la variación en niveles de producción entre vacas, así como la variación en persistencia entre vacas. Ya que los tratamientos forman un balance en cada secuencia y en cada período.

A continuación se tomó un ejemplo de Calzada (2) de un experimento que consiste en probar dos raciones cuyos efectos se han medido en kilos de leche producida.

Cuatro vacas correspondieron a la secuencia 1 y tres vacas a la secuencia 2. Los datos en clave están dados en el cuadro 11.

CUADRO 11. DATOS DE 4 VACAS SOMETIDAS A LA SECUENCIA 1 y 3 VACAS A LA SECUENCIA 2.

Secuencia N°1
Vacas

Períodos	X_1	X_2	X_3	X_4	Suma de Períodos (P)
a	A = 13	A = 10	A = 5	A = 8	$p_1 = 36$
b	B = 6	B = 5	B = 4	B = 7	$p_2 = 22$
c	A = 10	A = 8	A = 4	A = 7	$p_3 = 29$
Diferencias (D)	$d_1 = 11$	$d_2 = 8$	$d_3 = 1$	$d_4 = 1$	$P_1 = 21$

Secuencia N°2
Vacas

Períodos	X_1	X_2	X_3	X_4	Suma de Períodos (P)
a	B = 19	B = 7	B = 6	-	$p_4 = 22$
b	A = 14	A = 6	A = 8	-	$p_5 = 28$
c	B = 5	B = 3	B = 3	-	$p_6 = 11$
Diferencias (D)	$d_5 = -14$	$d_6 = -2$	$d_7 = -7$		$P_2 = -23$

A continuación se presenta el análisis estadístico tanto en forma indicada como en forma númerica.

1° Paso.

Hallar los valores de las p's (última columna del cuadro 11 de las secuencias 1 y 2, las que se obtienen sumando los resultados de cada período, así, para $p_1 = 13 + 10 + 5 + 8 = 36$

2° Paso.

Hallar para los resultados de las unidades experimentales de cada secuencia las d's según la siguiente ecuación:

$$1(a) - 2(b) + 1(c) = d$$

así, aplicando esta ecuación a las vacas X_1 y X_5 tenemos:

para $X_1 : 1(13) - 2(6) + 1(10) = 11 = d_1$

para $X_5 : 1(19) - 2(14) + 1(5) = -14 = d_5$

La misma ecuación se aplica a los valores de p_1, p_2 y p_5; así como a p_4, p_5 y p_6, para obtener los valores de P_1 y P_2, tal como se ve a continuación:

para la secuencia 1: $1(36) - 2(22) + 1(29) = 21 = P_1$

para la secuencia 2: $1(22) - 2(28) + 1(11) = -23 = P_2$

Para la comprobación de los cálculos se obtienen estos mismos resultados como sigue:

$$d_1 + d_2 + d_3 + d_4 = P_1 = 11 + 8 + 1 + 1 = 21$$

$$d_5 + d_6 + d_7 = P_2 = (-14) + (-2) + (-7) = -23$$

3° Paso.

La suma de cuadrados (S.C) de tratamientos, está dada por la siguiente ecuación:

$$\text{S.C tratamientos} = \frac{(r_2 P_1 - r_1 P_2)^2}{6\, r_1\, r_2\, (r_1 + r_2)}$$

siendo r_1 el número de repeticiones de la secuencia 1, y r_2 el número de repeticiones de la secuencia 2. Aplicando la ecuación al ejemplo tenemos:

$$\text{S.C tratamientos} = \frac{[3(21) - 4(-23)]^2}{(6)(4)(3)(4 + 3)} = 47.7$$

4° Paso.

La suma de cuadrados del error experimental está dado por la siguiente ecuación:

$$\text{S.C Error experimental} = \frac{1}{16} \left[\Sigma d^2 - \frac{P_1^{\,2}}{r_1} - \frac{P_2^{\,2}}{r_2} \right]$$

donde los grados de libertad del error experimental es igual a $r_1 + r_2 - 2$

Aplicando esta ecuación a los datos del ejemplo tenemos:

$$\text{S.C Error Experimental} = \frac{1}{16} \left[11^2+8^2+1^2+1^2+(-14)^2+(-2)^2+(-7)^2-\frac{21^2}{4}-\frac{23^2}{3}\right]$$

G.L. del error experimental = 4 + 3 - 2 = 5

5° Paso.

Con los resultados obtenidos se hace el análisis de varianza, tal como sigue:

Fuentes	S.C	G.L	C.M	F
Tratamientos	47.7	1	47.7	9.54 *
Error	25.0	5	5.0	

6° Paso.

Para establecer intervalo de confianza para la diferencia de medias de tratamiento se cálcula la media ajustada usando las siguientes fórmulas:

Cálculo de la media:

$$\bar{X} = \frac{1}{3(r_1 + r_2)} \left[\Sigma_p - \frac{(r_1 - r_2)\,(r_2 P_1 - r_1 P_2)}{8 r_1 r_2} \right]$$

Cálculo del factor de ajuste:

$$t_i = \pm \frac{r_2 P_1 - r_1 P_2}{8 v_1 v_2}$$

Luego se calcula $\bar{X} + t_i$, el resultado de t_i puede ser positivo o negativo, si es positivo se suma y si es negativo se resta a $\bar{X}$ para dar $\bar{X}_A$, realizándonse la operación inversa para obtener $\bar{X}B$. Aplicando el procedimiento a los datos del ejemplo, tenemos:

$$\bar{X} = \frac{1}{3(4+3)} \left[36+22+\ldots..11 - \frac{(4-3)((3)(21)-(4)(-23))}{(8)(4)(3)} \right] = 6.97$$

$$t_i = \pm \frac{3(21) - 4(-23)}{(8)(4)(3)} = 1.61$$

$\bar{X}_A$ 6.97 + 1.61 = 8.58 $\bar{X}_B$ = 6.97 - 1.61 = 5.36

Luego se calcula la varianza de la diferencia de medias, según la siguiente fórmula:

$$V(\bar{X}_D) = 2\left[\frac{3(r_1 + r_2)S^2}{16\ r_1\ r_2}\right]$$

donde S^2 es el cuadrado medio del error (C.M) tomado del análisis de varianza.

Para los datos del ejemplo tendremos:

$$V(\bar{X}_D) = 2\left[\frac{3(4 + 3)}{16x4x3}\ (5)\right] = 1.0937$$

Luego los límites para la diferencia de medias se calculan según la siguiente ecuación:

$$L(\bar{X}_D) = \bar{X}_A - \bar{X}_B \pm t_{(\alpha)}\sqrt{V\bar{X}_D}$$

Para el ejemplo tenemos, con "t" con 5 G.L. (G.L del error experimental) y $\alpha = 0.05$.

$$L(\bar{X}_D) = (8.58 - 5.36) \pm 2.571 \times (1.04)$$

Como se puede observar, los diseños de sobre cambio son fáciles de analizar manualmente. En diseños con ganado de leche, Lucas (6) reporta que el diseño es recomendable en dos casos:

a) En ensayos exploratorios donde se van a comparar tratamientos de efectos extremos y por alguna razón solo se pueden usar pocos animales.

b) En trabajos finales donde se quieren probar pequeñas diferencias entre dos tratamientos y se cuenta con un número sustancial de animales.

BIBLIOGRAFIA

1) BAJARANO, G. E. Producción de leche de cuatro grupos raciales en el trópico húmedo. Tesis Mg. Sc. Turrialba, Costa Rica, UCR/CATIE, 1979. 72 p.

2) CALZADA, B. J. Métodos estadísticos para la investigación. Lima, Perú, Ed. Jurídica, 1970. 640 p.

3) CERDAS, R. R. Banano de desecho (Musa acuminata) como suplemento a vacas lecheras en pastoreo,en diferentes estados de lactancia. Tesis Mg. Sc. Turrialba, Costa Rica, UCR/CATIE. 1981. 52 p.

4) LAZARTE, V.M. Efecto de la suplementación con yuca (Manihot scurenta grantz) como fuente de almidón sobre la producción de leche en vacas en pastoreo. Tesis Mg. Sc. Turrialba, Costa Rica, UCR/CATIE. 1978. 52 p.

5) LINDSTROM, M. B. El registro en los países en desarrollo. In Revista Mundial de Zootecnia, Roma. FAO 1977. 38-47 pp.

6) LUCAS, H. L. Designs for Animal Experiments, North Carolina Exp. Sta., Raleigh. 1948.

7) MOLINA, V. J. Análisis de la producción de leche en un hato Holstein puro. Tesis Ing. Agrónomo. San José, Costa Rica. Universidad de Costa Rica. 1978. 45 p.

8) PAEZ, B. G. Métodos de Investigación en Producción Animal. IICA, Turrialba, Costa Rica. 1964. 267 p.

9) RUIZ, V. A. Utilización de la gallinaza en la alimentación de bovinos. Tesis Mg. Sc. Turrialba, Costa Rica. UCR/CATIE. 1976. 81 p

10) SAS USER'S GUIDE: 1979 Edition. Cary, N.C., SAS Institute, 1979. 429 p.

11) STEEL, R.G. y TORRIE, J.H. Principles and Procedures of Statistics. New York, McGraw-Hill. 1960. 481 p.

12) VILLEGAS, Z.L.A. Suplementación con banano verde a vacas lecheras en pastoreo. Tesis Mg. Sc. Turrialba, Costa Rica, UCR/CATIE. 1979. 58 p.