UNIVERSIDAD NACIONAL DEL ALTIPLANO

Facultad de Ingeniería Estadística e Informática

METODOLOGÍA DE LA INVESTIGACIÓN

Vladimiro Ibañez Quispe, Dr.

C.U. agosto de 2024

TEMA:

ESTIMACION DE PARAMETROS Y PRUEBAS DE HIPOTESIS

Vladimiro Ibañez Quispe, Dr.

INFERENCIA ESTADÍSTICA

La inferencia estadística es un proceso que llega a inferencias respecto a una población, en base a resultados obtenidos de muestras tomadas de la población en referencia. La diferencia de la estimación con una prueba de hipótesis, es que una estimación en si, es una inferencia que se hace en los primeros de un experimento, donde no hay mucha información disponible, con los cuales se ofrece una aproximación del parámetro correspondiente a la población, de la cual se extrajo la muestra; en cambio una prueba de hipótesis se realiza cuando se dispone de mayor información de la situación experimental.

INFERENCIA ESTADÍSTICA:

Inferencia Estadística

→ **Estimación**: Posible en los primeros estadíos de un experimento.

→ **Prueba de Hipótesis**: Requiere de alguna previa información experimental.

I. TIPOS DE ESTIMACIÓN:

a. ESTIMACIÓN PUNTUAL.

De una población se puede obtener un sin número de muestras y de cada una de ellas un valor próximo al parámetro desconocido de la población. Consiste en dar un valor único para el parámetro que se desea estimar:

$$\hat{\theta} = \theta$$

PROPIEDADES DE UN BUEN ESTIMADOR.

CONSISTENCIA.- **En general, un estimador consistente es el que tiende a tener una mayor probabilidad de acercarse al parámetro de la población a medida que el tamaño de la muestra crece. Esto es:**

$$P(|\hat{\theta} - \theta| \leq \varepsilon) \rightarrow 1, \text{ cuando } n \rightarrow \infty$$

La media y varianza muestral son estimadores consistentes que tienden acercarse a los valores de la población a medida que aumenta el tamaño de la muestra.

INSESGAMIENTO.- **Un estimador es insesgado si su valor esperado es idéntico al parámetro de la población que se estima.**

$$E(\hat{\theta}) = \theta$$

La media y proporción muestral, la diferencia de medias y proporciones muestrales son estimadores insesgados. Sin embargo, la varianza muestral es un estimador sesgado de la varianza poblacional.

EFICACIA.- **Si se utilizan dos estadígrafos como estimadores de un parámetro, el más eficaz es aquel cuya distribución muestral tenga menor desviación estándar**

$$V(\hat{\theta}_1) < V(\hat{\theta}_2)$$

SUFICIENCIA.- **Un estimador es suficiente si aporta tanta información como sea posible acerca del parámetro contenido en la muestra, de modo que cualquier otro estimador proporcione escasa información adicional.**

b. ESTIMACIÓN POR INTERVALOS.

Un intervalo aleatorio, es un intervalo con límites finitos ó infinitos, donde por lo menos uno de los extremos es una variable aleatoria. En forma general, un intervalo de confianza para el parámetro θ. Es la estimación que consiste en dar un intervalo, de un grado aceptable de valores dentro de los cuales esperamos que se localice el parámetro que se desea estimar, llamado intervalo de confianza.

$$P(\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2) = 1 - \alpha$$

El número (1 - α) se llama el nivel de confianza, su elección depende del investigador y sus valores más utilizados son:

Si: $\alpha = 0.10$ **entonces** $1 - 0.10 = 0.90$

$\alpha = 0.05$ **entonces** $1 - 0.05 = 0.95$

$\alpha = 0.01$ **entonces** $1 - 0.01 = 0.99$

$\alpha = 0.02$ **entonces** $1 - 0.02 = 0.98$

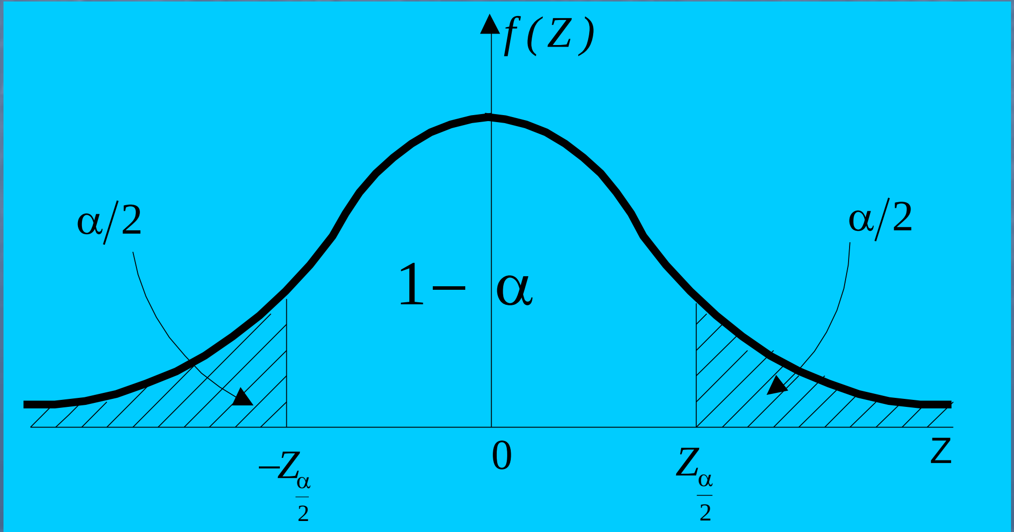
II. INTERVALOS DE CONFIANZA (IC)

1. Intervalo de Confianza para la Media con Varianza σ^2 conocida.

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

El Intervalo Confidencial (IC), estandarizado es:

$$P\left(\overline{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$



2. Intervalo de Confianza para la Media Poblacional (μ) con Varianza poblacional σ^2 desconocida.

i) Muestras grandes *n>30*.

Si la población no es normal pero el tamaño de la muestra es suficientemente grande (*n>30*), se utiliza la desviación estándar «*s*» de la muestra, como estimación puntual de la desviación estándar (σ) de la población.

$$Z = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

$$P(\underbrace{\overline{X} - Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}}_{L_I} \leq \mu \leq \underbrace{\overline{X} + Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}}_{L_S}) = 1 - \alpha$$

ii) Muestras grandes $n \leq 30$.

Cuando σ^2 es desconocida, la media y la varianza se debe estimar a partir de la muestra.

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}, \quad s^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1} = \frac{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}{n-1}$$

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

El Intervalo Confidencial (IC), es el siguiente:

$$P(\underbrace{\overline{X} - t_{\frac{\alpha}{2},(n-1)}\, S/\sqrt{n}}_{L_I} \leq \mu \leq \underbrace{\overline{X} + t_{\frac{\alpha}{2},(n-1)}\, S/\sqrt{n}}_{L_S}) = 1 - \alpha$$

EJEMPLO:

Las calificaciones de 40 estudiantes sobre una prueba de aptitud son las siguientes:

27.8	33.5	28.4	44.6	35.0	48.8	19.2	32.5
43.0	16.9	27.2	25.5	35.0	40.6	11.8	0.2
28.7	43.6	19.8	34.1	13.4	38.6	3.9	13.5
19.1	24.7	21.0	0.6	11.8	32.0	33.4	3.5
45.7	9.7	39.5	15.4	17.0	16.4	27.9	31.0

Construir un IC del 95% para el verdadero prueba de aptitud.

$$\overline{X} = \frac{\sum_{i=1}^{40} X_i}{n} = \frac{27.8 + 43.0 + \cdots + 3.53 + 31.0}{40} = \frac{1014.33}{40} = 25.358$$

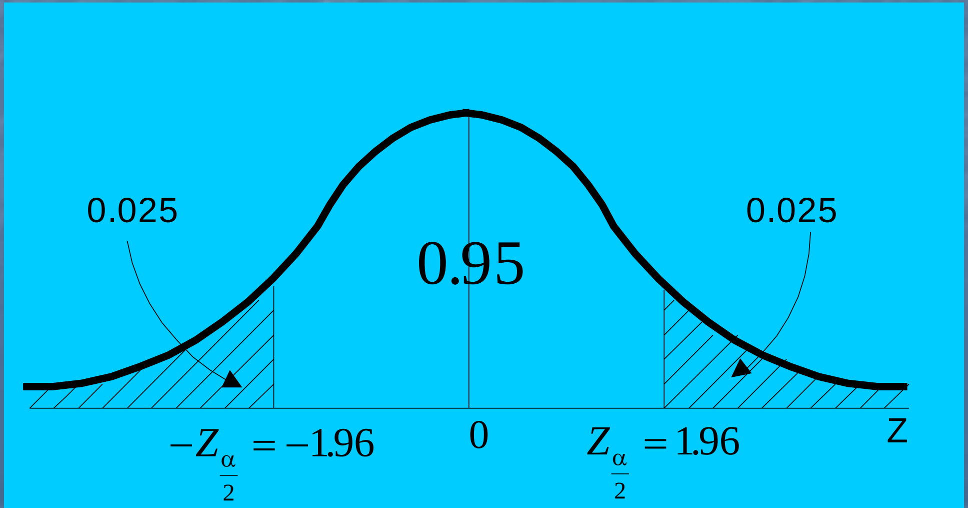
$$S^2 = \frac{\sum_{i=1}^{40} X_i^2 - \frac{\left(\sum_{i=1}^{40} X_i\right)^2}{n}}{n-1} = \frac{32390.8409 - \frac{(1014.33)^2}{40}}{39} = 171.0053123$$

S = 13.0769;   n = 40;   1 - α = 1 – 0.05 = 0.95

$$P\left(25.358 - (1.96)\frac{13.0769}{\sqrt{40}} \le \mu \le 25.358 + (1.96)\frac{13.0769}{\sqrt{40}} \right) = 0.95$$

$$P(21.305 \le \mu \le 29.41) = 0.95 \quad ó \quad (21.305, 29.41)$$

Hay un 95% de confianza de que la media de población de la prueba de aptitud, esté en un punto cualquiera comprendido entre 21.305 y 29.41 y un 5% estará fuera de este intervalo. También se puede decir, que de 100 intervalos aleatorios 95% de éstos contienen a μ, en tanto 5% no los contienen.



3. Intervalo de Confianza para la varianza de una población normal.

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de tamaño «n» y S^2 la varianza de muestra, se tiene:

Si $n \leq 30$

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

El intervalo de confianza de dos lados del 100x(1- α)% en σ^2 es:

$$P\left[\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},(n-1)}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},(n-1)}}\right] = 1-\alpha$$

Si *n > 30*

Si el grado de libertad «n» de una distribución es mayor que 30, se usará las siguientes aproximaciones para calcular el intervalo confidencial:

$$P\left[\frac{(n-1)S^2}{\frac{1}{2}\left(Z_{\alpha/2}+\sqrt{2n-1}\right)^2}\leq\sigma^2\leq\frac{(n-1)S^2}{\frac{1}{2}\left(-Z_{\alpha/2}+\sqrt{2n-1}\right)^2}\right]=1-\alpha$$

Donde:

$$\chi^2_{\alpha/2,(n-1)}=\frac{1}{2}\left(Z_{\alpha/2}+\sqrt{2n-1}\right)^2$$

$$\chi^2_{1-\alpha/2,(n-1)}=\frac{1}{2}\left(-Z_{\alpha/2}+\sqrt{2n-1}\right)^2$$

EJEMPLO:

Se registraron los valores de la capacidad vital de una muestra de 10 pacientes con obstrucción crónica severa de las vías respiratorias. La varianza de las 10 observaciones fue de 0.75. Calcular el Intervalo Confidencia (IC). Use $\alpha = 0.05$

Como $n \leq 30$

El intervalo de confianza de dos lados del 100x(1- α)% en σ^2 es:

$$P\left[\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},(n-1)}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},(n-1)}}\right] = 1-\alpha$$

$$P\left[\frac{9x0.75}{19} \leq \sigma^2 \leq \frac{9x0.75}{2.70}\right] = 0.95$$

$$P(0.3553 \leq \sigma^2 \leq 2.5) = 0.95$$

4. Intervalo de Confianza para la razón de dos varianzas de dos distribuciones normales.

Sean dos muestras aleatorias de tamaños n_1 y n_2 tomadas de X e Y, y sea $S^2{}_1$ y $S^2{}_2$ denotan las varianzas de las muestras. Para determinar el intervalo de confianza, notamos que la distribución de muestreo es:

$$F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F_{(n_1 - 1,\, n_2 - 1)},$$

$$\sigma_1^2 \text{ es estimado por } S_1^2 = \frac{\sum X_i^2 - \frac{(\sum X_i)^2}{n_1}}{n_1 - 1}$$

$$\sigma_2^2 \text{ es estimador por } S_2^2 = \frac{\sum Y_i^2 - \frac{(\sum Y_i)^2}{n_2}}{n_2 - 1}$$

El intervalo de confianza de 100(1- α)% para la razón es:

$$P\left[\frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{1-\frac{\alpha}{2},(n_1-1,n_2-1)}} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{S_1^2}{S_2^2} \frac{1}{F_{\frac{\alpha}{2},(n_1-1,n_2-1)}}\right] = 1 - \alpha$$

EJEMPLO:

Se dividió en dos grupos una clase de Estadística de 40 alumnos. Cada grupo utilizó durante un ciclo un método de enseñanza diferente. Al final del ciclo los alumnos se sometieron a una misma prueba de rendimiento obteniéndose los siguientes resultados: $n_1 = 19$, $S^2_1 = 280$, $n_2 = 21$, $S^2_2 = 200$. Calcule el IC para varianzas poblacionales. Use $\alpha = 0.10$.

Solución:

$n_1 - 1 = 19 - 1 = 18$, $S^2_1 = 280$, $n_2 - 1 = 21 - 1 = 20$, $S^2_2 = 200$.

Usando la tabla acumulativa de "F", se tiene:

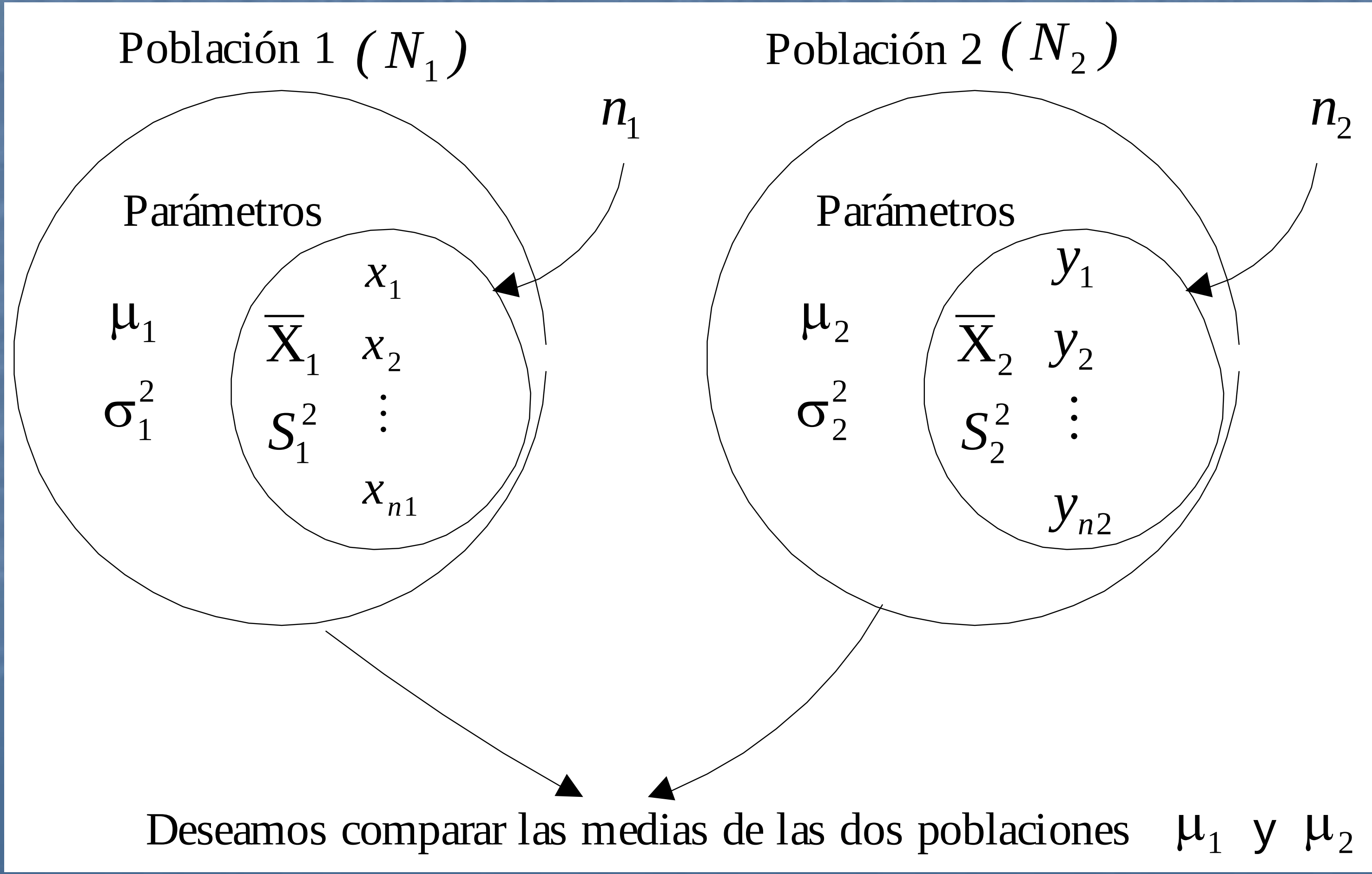
$$F_{0.90,(18,20)} = 1.811 \qquad F_{0.10,(20,18)} = 0.55208$$

Por tanto, el intervalo de confianza 100(1- α)% para σ_1^2/σ_2^2

$$P\left[\frac{280}{200}\frac{1}{1.811} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{280}{200}\frac{1}{0.552}\right] = 0.90 \qquad P\left[0.773 \le \frac{\sigma_1^2}{\sigma_2^2} \le 2.535\right] = 0.90$$

5. Intervalo de Confianza de dos medias poblacionales independientes (μ_1 - μ_2) y ambas varianzas conocidas (σ^2_1 y σ^2_2)

Sean $X_1, X_2, \ldots, X_{n1}$ de una muestra aleatoria extraída de una población $N(\mu_1, \sigma^2_1)$ y $Y_1, Y_2, \ldots, Y_{n2}$ otra muestra aleatoria extraída de una población $N(\mu_2, \sigma^2_2)$.



El estimador de μ_1 y μ_2 es:

$$\mu_1 \text{ es } \overline{X}_1 = \frac{\sum_{i=1}^{n_1} X_i}{n_1}$$

$$\mu_2 \text{ es } \overline{X}_2 = \frac{\sum_{i=1}^{n_2} Y_i}{n_2}$$

Si n_1 y n_2 son mayores que 30, se puede establecer un intervalo de confianza para μ_1 y μ_2 considerando la distribución muestral de promedios, la estadística es:

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

el intervalo de confianza del 100(1- α)% para (μ_1 - μ_2).

$$P\left(\underbrace{\overline{X}_1 - \overline{X}_2 - Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}_{L_I} \le \mu_1 - \mu_2 \le \underbrace{\overline{X}_1 - \overline{X}_2 + Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}_{L_S} \right) = 1 - \alpha$$

EJEMPLO:

Un grupo de 50 estudiantes de la UNA realizan un trabajo y obtuvieron los siguientes resultados: $\overline{X}_1$= 89 y varianza de $\sigma^2{}_1$=49 Otro grupo de 60 estudiantes de la UANCV realizan el mismo trabajo y obtuvieron un promedio de $\overline{X}_2$= 87 y varianza de $\sigma^2{}_2$= 25. Calcule el IC para la diferencia de medias entre mediciones obtenidas por alumnos de la UNA y UANCV. Use α = 0.05.

Solución:

$$Z_{0.05/2} = Z_{0.025} = 1.96 \quad \text{(Tabla Z)}$$

1- α = 1 – 0.05 = 0.95; $\sigma^2{}_1$=49; $\sigma^2{}_2$= 25; $\overline{X}_1$= 89; $\overline{X}_2$ = 87

Reemplazando en la fórmula, se tiene:

$$P\left(89 - 87 - (1.96)\sqrt{\frac{49}{50} + \frac{25}{60}} \le \mu_1 - \mu_2 \le 89 - 87 + (1.96)\sqrt{\frac{49}{50} + \frac{25}{60}}\right) = 0.95$$

$$P(-0.316340792 \le \mu_1 - \mu_2 \le 4.31634) = 0.95$$

Hay 95% de confianza que la diferencia verdadera de (μ_1 - μ_2), está en algún punto comprendido entre (-0.316, 4.316) mediciones.

6. Intervalo de Confianza de dos medias poblacionales normales independientes (μ_1 - μ_2) con varianzas desconocidas (σ^2_1 y σ^2_2)

Sean $X_1, X_2, \ldots\ldots, X_{n1}$ de una muestra aleatoria extraída de una población $N(\mu_1, \sigma^2_1)$ y $Y_1, Y_2, \ldots\ldots, Y_{n2}$ otra muestra aleatoria extraída de una población $N(\mu_2, \sigma^2_2)$. Donde μ_1 y μ_2, σ^2_1 y σ^2_2 desconocidas.

Caso I: Tamaños muestrales pequeños $n_1 + n_2 \leq 30$ pero $\sigma_1^2 = \sigma_2^2 = \sigma^2$

El Intervalo confidencial (IC) de dos lados del 100(1- α)% para (μ_1 - μ_2), está dada por:

$$t = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{S_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{(n_1 + n_2 - 2)\text{g.l.}}$$

$$S_P^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$P\left(\overline{X}_1 - \overline{X}_2 - t_{\frac{\alpha}{2}, n_1 + n_2 - 2}\, S_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq \overline{X}_1 - \overline{X}_2 + t_{\frac{\alpha}{2}, n_1 + n_2 - 2} S_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) = 1 - \alpha$$

EJEMPLO:

Se estudia el contenido de nicotina en los cigarrillos de dos marcas A y B, obteniéndose los siguientes resultados: A: 17, 20, 20, 23 y B: 18, 20, 21, 22, 24; con α = 0.05. Encontrar el intervalo confidencial al 95%.

Solución:

1- α = 1 – 0.05 = 0.95; como 4+5 ≤30, entonces usar t-Student.

Marca A:

$$\overline{X}_A = \frac{\sum_{i=1}^{4} X_i}{4} = \frac{80}{4} = 20$$

$$S_A^2 = \frac{\sum X_i^2 - n\overline{X}^2}{n-1} = \frac{1618 - 4(20)^2}{3} = \frac{18}{3} = 6$$

Marca B:

$$\overline{X}_B = \frac{\sum_{i=1}^{5} Y_i}{5} = \frac{105}{5} = 21$$

$$S_B^2 = \frac{\sum Y_i^2 - n\overline{Y}^2}{n-1} = \frac{2225 - 5(21)^2}{4} = \frac{20}{4} = 5.0$$

$$t_{\alpha/2,(n_A+n_B-2)} = t_{0.025,(4+5\text{-}2)} = t_{0.025,7} = 2.365$$

$$P\left((20-21)-(2.365)\sqrt{\frac{3(6)+(4)(5)}{7}\left(\frac{1}{4}+\frac{1}{5}\right)} \le \mu_A-\mu_B \le (20-21)+(2.365)\sqrt{\frac{3(6)+(4)(5)}{7}\left(\frac{1}{4}+\frac{1}{5}\right)}\right) = 0.95$$

$$P(-2.69641037 \le \mu_A - \mu_B \le 4.69641037) = 0.95$$

Comprobando si las varianzas son homogéneas o no:

$$F_{0.95,(3,4)} = 6.591$$

$$F_{0.05,(4,3)} = \frac{1}{9.117} = 0.1097$$

$$P\left[\frac{6}{5}.\frac{1}{6.591} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{6}{5}.\frac{1}{0.1097}\right] = 0.95$$

$$P\left[0.182 \le \frac{\sigma_1^2}{\sigma_2^2} \le 10.939\right] = 0.95$$

El intervalo cubre a 1, por consiguiente las varianzas son homogéneas, es decir:

$$\sigma_1^2 = \sigma_2^2 = \sigma^2$$

Caso II: Tamaños muestrales pequeños $n_1 + n_2 \le 30$, y varianzas poblacionales desconocidas y $\sigma_1^2 \ne \sigma_2^2$

En muchas situaciones no es razonable suponer que $\sigma^2_1=\sigma^2_2$. Cuando esta suposición es injustificada, aún sería posible encontrar un intervalo de confianza de 100(1- α)% en (μ_1 - μ_2) empleando el hecho de que la estadística es:

$$t^* = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_{(v)}$$

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1+1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2+1}} - 2$$

El intervalo confidencial está dada por:

$$P\left(\overline{X}_1 - \overline{X}_2 - t_{\frac{\alpha}{2},v}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \le \mu_1 - \mu_2 \le \overline{X}_1 - \overline{X}_2 + t_{\frac{\alpha}{2},v}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right) = 1 - \alpha$$

EJEMPLO:

Un investigador está interesado en saber si los niños nacidos prematuramente con acidosis metabólica tardía y los niños prematuros que no tienen dicha enfermedad difieren en lo que respecta a las concentraciones en la orina de cierta sustancia química. Las concentraciones medias, desviación estándar y el tamaño de la muestra para ambos grupos son las siguientes:

Muestra	n	$\bar{X}$	s
Con la condición	15	8.50	5.5
Sin la condición	14	4.80	2.9

Encuentre el intervalo confidencial al 95%.

Solución:

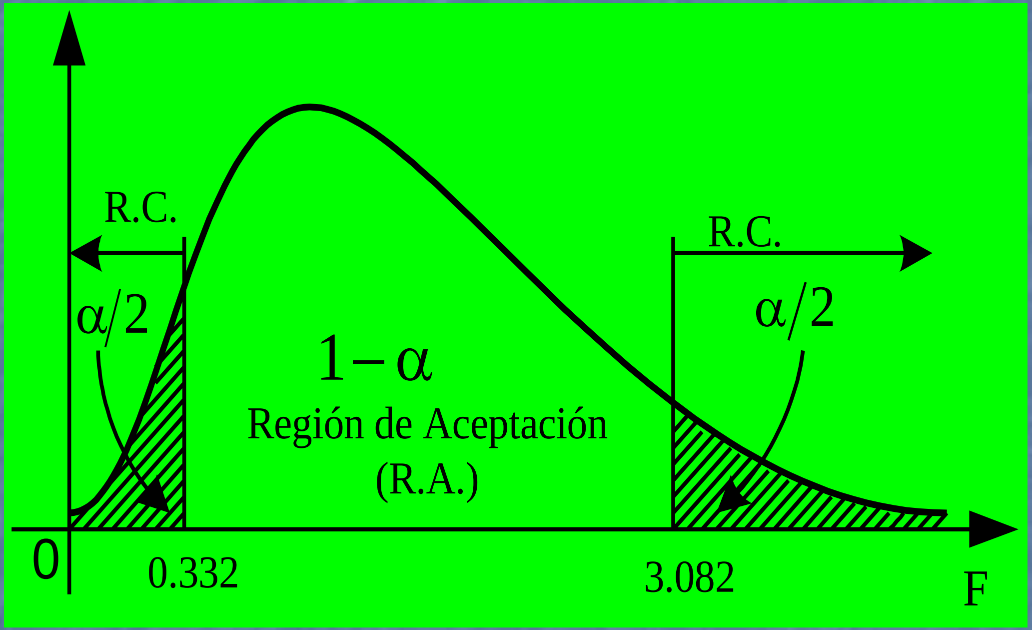
Primeramente se debe probar si la varianzas es:

$$\sigma_1^2 = \sigma_2^2 = \sigma^2 \quad ó \quad \sigma_1^2 \neq \sigma_2^2$$

$$F_c = \frac{(5.5)^2}{(2.9)^2} = \frac{30.25}{8.41} = 3.59 \sim 3.60$$

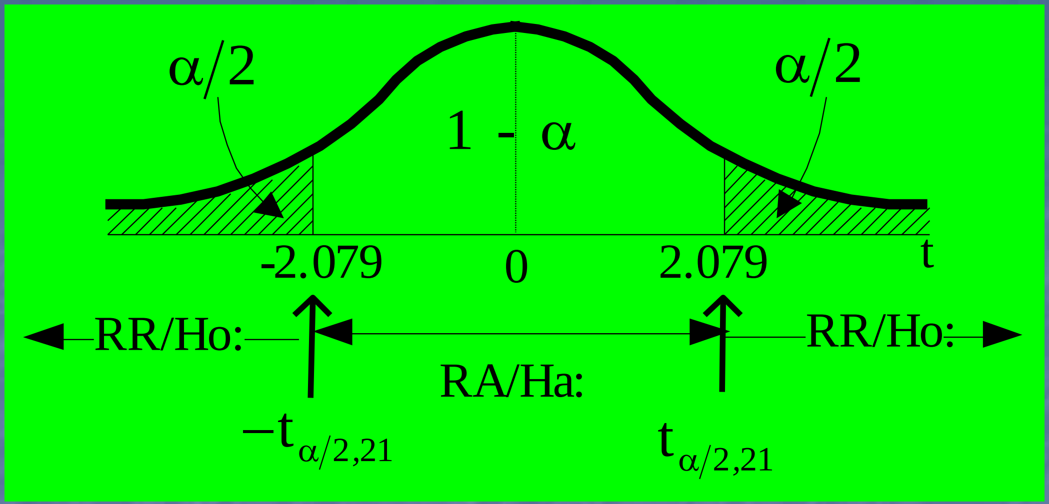
$$F_{0.975,14,13} = \frac{1}{F_{0.025,13,14}} = \frac{1}{3.012} = 0.332$$

$$\mathbf{F_{0.025,14,13} = 3.082}$$



Como F_c=3.60 cae en la región rechazo, entonces existe evidencia suficiente para afirmar que existe diferencia entre $\sigma^2_1 \neq \sigma^2_2$. Entonces, se prosigue a efectuar la prueba de la diferencia de promedios.

$$t = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)}} \sim t_{(v)}$$



***v* = Grados de libertad.**

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}} = \frac{\left(\frac{30.25}{15} + \frac{8.41}{14}\right)^2}{\frac{\left(\frac{30.25}{n_1}\right)^2}{14} + \frac{\left(\frac{8.41}{n_2}\right)^2}{13}} = \frac{6.85068}{0.31825} = 21.5 \sim 21$$

$$t_c = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)}} = \frac{8.5 - 4.8}{\sqrt{\left(\frac{30.25}{15} + \frac{8.41}{14}\right)}} = \frac{3.7}{1.6178} = 2.287$$

El intervalo confidencial es:

$$P\left(\overline{X}_1 - \overline{X}_2 - t_{\frac{\alpha}{2},v} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \le \mu_1 - \mu_2 \le \overline{X}_1 - \overline{X}_2 + t_{\frac{\alpha}{2},v} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right) = 1 - \alpha$$

$$P\left(8.5 - 4.8 - 2.079 \sqrt{\frac{30.25}{15} + \frac{8.41}{14}} \le \mu_1 - \mu_2 \le 8.5 - 4.8 + 2.079 \sqrt{\frac{30.25}{15} + \frac{8.41}{14}} \right) = 0.95$$

Por tanto, el intervalo de confianza del 95% para (μ_1 - μ_2) es:

$$P(0.3355 \le \mu_1 - \mu_2 \le 7.0644) = 0.95$$

De 100 muestras que se tome en el 95% se encontrará entre (0.3355, 7.064) las concentraciones en niños y el 5% no lo contendrá al intervalo.

Caso III: Tamaños muestrales grandes $n_1 + n_2 > 30$, y varianzas poblacionales desconocidas.

En este caso la estadística "t" tiene distribución aproximadamente normal:

$$\overline{X}_1 - \overline{X}_2 \;\sim\; N\left(\mu_1 - \mu_2, \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)$$

Entonces el estadístico es:

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \;\sim\; N(0,1)$$

El intervalo de confianza del 100(1- α)% para (μ_1 - μ_2).

$$P\left(\overline{X}_1 - \overline{X}_2 - Z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \le \mu_1 - \mu_2 \le \overline{X}_1 - \overline{X}_2 + Z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right) = 1 - \alpha$$

7. Intervalo de Confianza sobre (μ_1 - μ_2) con observaciones en pares (pareados, emparejados, relacionados).

Un procedimiento comúnmente utilizado para dos muestras no independientes es el denominado prueba «ANTES» y «DESPUÉS». Las mediciones se hacen sobre la muestra de sujetos tanto antes como después de la introducción de algún fenómeno, es decir hay dos mediciones en la misma unidad experimental, las dos mediciones dentro del mismo par pueden no ser independientes.

La muestra está formado por los pares: (X_1,Y_1), (X_2,Y_2), (X_3,Y_3), ... , (X_n,Y_n). Consideránse las «*n*» diferencias: $D_1 = X_1 - Y_1$, $D_2 = X_2 - Y_2$, ... , $D_n = X_n - Y_n$. En estas condiciones la media ($\mu_D = \mu_1 - \mu_2$) y varianza de las diferencias está dada por:

$$\overline{\mathrm{D}} = \frac{\sum_{i=1}^{n} \mathrm{D}_i}{n}, \quad S_D^2 = \frac{\sum_{i=1}^{n} \mathrm{D}_i^2 - \frac{\left(\sum D_i\right)^2}{n}}{n-1} = \frac{\sum_{i=1}^{n} \left(\mathrm{D}_i - \overline{\mathrm{D}}\right)^2}{n-1}$$

Caso I: Tamaños muestral pequeño (n≤30)

La estadística de prueba es:

$$t = \frac{\overline{D} - \mu_D}{S_D / \sqrt{n}} \sim t_{(n-1)}$$

$$P\left(-t_{\frac{\alpha}{2},(n-1)} \leq t \leq t_{\frac{\alpha}{2},(n-1)} \right) = 1 - \alpha$$

El intervalo de confianza del 100(1- α)% en ($\mu_D = \mu_1 - \mu_2$) es:

$$P\left[\overline{D} - t_{\frac{\alpha}{2},(n-1)} \frac{S_D}{\sqrt{n}} \leq \mu_D \leq \overline{D} + t_{\frac{\alpha}{2},(n-1)} \frac{S_D}{\sqrt{n}} \right] = 1 - \alpha$$

EJEMPLO:

Veinte estudiantes de Educación fueron divididas en 10 parejas, teniendo cada miembro de la pareja aproximadamente el mismo cociente de inteligencia. Uno de cada pareja se selecciona al azar y se asigna a una sección que utiliza videos. El otro miembro se asigna a una sección que cuenta con profesor. Al finalizar el ciclo ambos grupos se presentan al mismo examen, obteniéndose los resultados siguientes:

Pareja	Con video (V)	Con profesor (P)	$D = V - P$	D^2
1	15	16	-1	1
2	12	10	2	4
3	17	17	0	0
4	11	14	-3	9
5	18	17	1	1
6	15	16	-1	1
7	16	18	-2	4
8	13	12	1	1
9	14	15	-1	1
10	10	11	-1	1
Total	$\sum V =$ 141	$\sum P =$ 146	$\sum D_i =$ -5	$\sum D_i^2 =$ 23

Suponiendo que la característica en estudio es normal, obtener el intervalo de confianza del 95% para la diferencia real en el promedio de calificaciones de los dos procedimientos de enseñanza.

Solución:

$$\overline{X}_V = \frac{\sum_{i=1}^{10} V_i}{n} = \frac{141}{10} = 14.1$$

$$\overline{X}_P = \frac{\sum_{i=1}^{10} P_i}{n} = \frac{146}{10} = 14.6$$

$$\overline{D} = \overline{X}_V - \overline{X}_P = 14.1 - 14.6 = -0.5$$

$$S_D^2 = \frac{\sum_{i=1}^{n} D_i^2 - \frac{\left(\sum D_i\right)^2}{n}}{n-1} = \frac{23 - \frac{(-5)^2}{10}}{9} = \frac{20.5}{9} = 2.278$$

S_D = 1.509230856

$$t_{\alpha/2,(n-1)} = t_{0.05/2,(10-1)} = t_{0.025,9} = 2.262$$

$$P\left[-0.5 - (2.262)\frac{1.509230856}{\sqrt{10}} \le \mu_D \le -0.5 + (2.262)\frac{1.509230856}{\sqrt{10}}\right] = 0.95$$

$$P\ \ {-1.5795637} \le \mu_D \le 0.579563708 = 0.95$$

Caso II: Tamaño muestral grande (n>30)

Cuando n>30, la distribución de «t» de Student se apróxima a la distribución N(0,1). El intervalo confidencial para ($\mu_D = \mu_1 - \mu_2$) con coeficiente de confianza $1-\alpha$ esta dado por:

$$Z = \frac{\overline{D} - \mu_D}{S_D / \sqrt{n}} \sim N(0,1)$$

$$P\left(-Z_{\frac{\alpha}{2}} \leq Z \leq Z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

Estandarizando, se tiene:

$$P\left[\overline{D} - Z_{\frac{\alpha}{2}} \frac{S_D}{\sqrt{n}} \leq \mu_D \leq \overline{D} + Z_{\frac{\alpha}{2}} \frac{S_D}{\sqrt{n}}\right] = 1 - \alpha$$

8. Intervalo de Confianza para una Proporción Poblacional (P).

Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria extraída de una población Bernoullí de parámetro P.

Parámetro:

$$P = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{X}{N} = \frac{\text{Número de casos favorables}}{N}$$

Estimador:

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x}{n} = \frac{\text{Número de éxitos en la muestra}}{n}$$

El intervalo de confianza para P en el caso de muestras grandes (*n > 30*). Cuando «n» es suficientemente grande, la distribución es:

$$Z = \frac{\hat{p} - P}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim N(0,1)$$

Para construir el intervalo de confianza en P, se tiene:

$$P(-Z_{\frac{\alpha}{2}} \leq Z \leq Z_{\frac{\alpha}{2}}) = 1 - \alpha$$

El Intervalo de Confianza del 100(1- α)% para el parámetro P es aproximadamente dado por:

$$P\left[\underbrace{\hat{p} - Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}_{L_I} \leq P \leq \underbrace{\hat{p} + Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}_{L_S}\right] = 1 - \alpha$$

EJEMPLO:

El departamento de Oftalmología de la UNA, desea saber la proporción de estudiantes cuyo ojo derecho es el ojo dominante. Para ello se escogió una muestra de 400 estudiantes, habiendo encontrado que 280 estudiantes tienen el ojo derecho dominante. Estime la proporción de la población total, cuyo ojo derecho es el dominante por medio de un intervalo de confianza del 95%.

Solución:

$$\hat{p} = \frac{280}{400} = 0.7, \qquad \hat{q} = 1 - \hat{p} = 1 - 0.7 = 0.3$$

El valor de «Z» se encuentra en la Tabla de la siguiente manera:

$$Z_{\alpha/2} = Z_{0.05/2} = Z_{0.025} = 1.96$$

$$\text{como } P\left(Z < Z_{\alpha/2}\right) = 1 - \frac{\alpha}{2} = 1 - 0.025 = 0.975 \;\Rightarrow\; Z_{\alpha/2} = 1.96$$

Luego el intervalo de confianza del 95% se tiene:

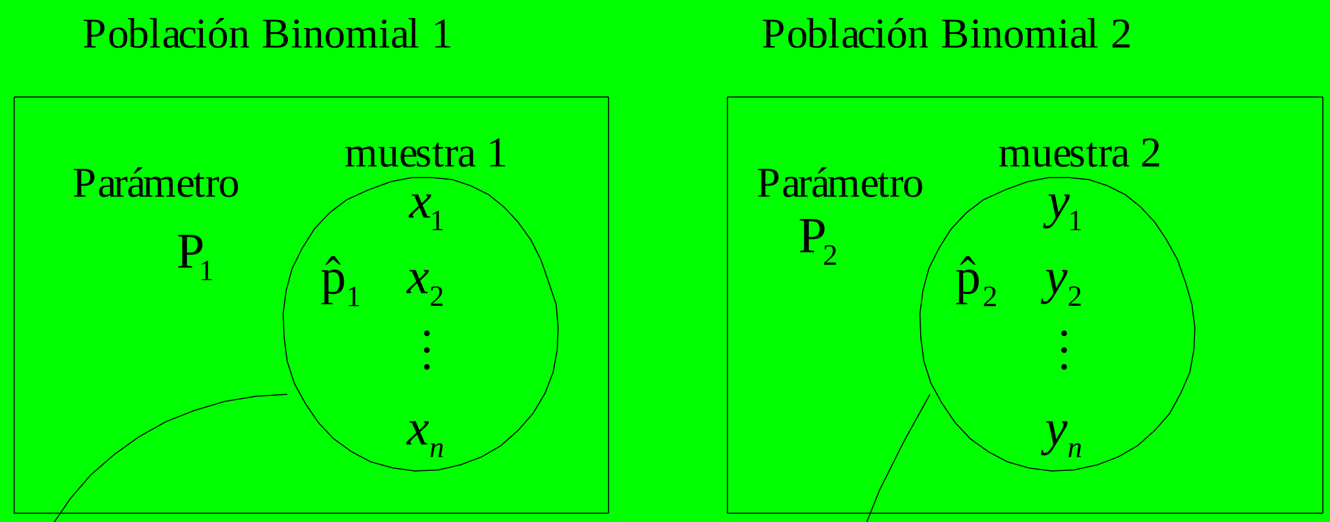
$$P\left[0.7 - (1.96)\sqrt{\frac{(0.7)(0.3)}{400}} \le P \le 0.7 + (1.96)\sqrt{\frac{(0.7)(0.3)}{400}}\right] = 0.95$$

$$0.7 \pm 0.044909241 \begin{cases} L_I = 0.7 - 0.044909241 = 0.6550 \\ L_S = 0.7 + 0.044909241 = 0.7449 \end{cases}$$

$$P\left(0.655 \le P \le 0.7449\right) = 0.95 \;\Rightarrow\; 0.655 \le P \le 0.7449$$

9. Intervalo de Confianza para la diferencia entre dos proporciones poblacionales $(P_1 - P_2)$

La estimación puntual de $(P_1 - P_2)$ se obtiene, tomando una muestra 1 de la población Binomial 1 y una muestra 2 de la población Binomial 2.



$$\hat{p}_1 = \frac{x_1 + x_2 + \cdots + x_n}{n_1} = \frac{\sum_{i=1}^{n_1} x_i}{n_1} = \frac{x}{n_1}$$

$$\hat{p}_2 = \frac{y_1 + y_2 + \cdots + y_n}{n_2} = \frac{\sum_{i=1}^{n_2} y_i}{n_2} = \frac{y}{n_2}$$

Sean $X_1, X_2, \ldots, X_{n1}$ una muestra aleatoria extraída de una población Binomial $B(n_1, p_1)$ y $Y_1, Y_2, \ldots, Y_{n2}$ una muestra aleatoria extraída de una población Binomial $B(n_2, p_2)$. Supomgamos que ambas poblaciones son independientes.

$$\hat{p}_1 = \frac{\sum_{i=1}^{n_1} x_i}{n_1} = \frac{x}{n_1} = \frac{\text{Número de éxitos de la muestra 1}}{n_1}$$

$$\hat{p}_2 = \frac{\sum_{j=1}^{n_1} y_j}{n_2} = \frac{y}{n_2} = \frac{\text{Número de éxitos de la muestra 2}}{n_2}$$

son estimadores independientes de P_1 y P_2.

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (P_1 - P_2)}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}} \sim N(0,1),$$

$$P(-Z_{\frac{\alpha}{2}} \leq Z \leq Z_{\frac{\alpha}{2}}) = 1 - \alpha$$

Por tanto, el intervalo de confianza del (1- α)100% para la diferencia de proporciones (P_1 - P_2) esta dado aproximadamente por:

$$P\left[\left(\hat{p}_1-\hat{p}_2\right)-Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}} \leq \left(P_1-P_2\right) \leq \left(\hat{p}_1-\hat{p}_2\right)+Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}\right] = 1 - \alpha$$

EJEMPLO:

En la Facultad de Ciencias de la Educación de la UNA, se desea determinar la opinión de los estudiantes y profesores en relación al calendario académico propuesto para el año siguiente. Una muestra aleatoria de 100 estudiantes y 50 profesores dan los siguientes resultados:

Opinión	Estudiantes	Profesores
Favorecen al Calendario Académ. Propuesto	63	30
Se oponen al calendario	37	20

¿Hay prueba de una diferencia en la actitud hacia el calendario académico propuesto entre los dos grupos?.

Solución:

$$\hat{p}_1 = \frac{63}{100} = 0.63 \quad y \quad \hat{p}_2 = \frac{30}{50} = 0.6$$

$$\hat{q}_1 = 1 - 0.63 = 0.37 \quad y \quad \hat{q}_2 = 1 - 0.6 = 0.4$$

$$\hat{q}_1 = \frac{37}{100} = 0.37 \quad y \quad \hat{q}_2 = \frac{20}{50} = 0.4$$

$$Z_{\alpha/2} = Z_{0.05/2} = Z_{0.025} = 1.96$$

(Se obtiene de la Tabla «Z»)

Por tanto, el intervalo de confianza de 95% para (P_1 - P_2) es:

$$P\left[(0.63-0.6)-(1.96)\sqrt{\frac{(0.63)(0.37)}{100}+\frac{(0.6)(0.4)}{50}} \le P_1 - P_2 \le (0.63-0.6)+(1.96)\sqrt{\frac{(0.63)(0.37)}{100}+\frac{(0.6)(0.4)}{50}}\right] = 0.95$$

$$P\left[0.03 - 0.165512687 \le P_1 - P_2 \le 0.03 + 0.165512687\right] = 0.95 \quad \Rightarrow \quad -0.1355 \le P_1 - P_2 \le 0.1955$$

El IC. cubre el cero (-0.1355, 0.1955), entonces se afirma que no hay evidencia de una diferencia en la actitud hacia el calendario académico propuesto entre los dos grupos en estudio.

III. PRUEBA DE HIPÓTESIS

El enunciado suele llamarse hipótesis, y el procedimiento de toma de decisiones en torno a la hipótesis recibe el nombre de Prueba de Hipótesis. Este es uno de los aspectos más útiles de la inferencia estadística, puesto que muchos tipos de problemas de decisión pueden formularse como problemas de prueba de hipótesis.

3.1. Hipótesis Estadística.

Es cualquier afirmación o conjetura que se hace acerca de la distribución de una o más poblaciones. La afirmación o conjetura puede referirse bien a la forma o tipo de distribución de probabilidad de la población o al valor o valores de uno o más parámetro de la distribución conocida.

EJEMPLO:

- Un agricultor de la sierra puede establecer la hipótesis de que la duración promedio de las lluvias es de 90 días.
- Un médico puede conjeturar que la incidencia de diabetes en pacientes femeninos es de 3%.
- La proporción de objetos defectuosos producidos por cierto proceso nunca es superior al 8%.
- Un investigador pretende estudiar en forma comparativa la eficacia de dos tratamientos (o procedimientos experimentales) para determinar cuál de los dos es mejor.
- Una enfermera puede hipotetizar de que un programa de educación pre-operatoria hace que el paciente entre en cierta confianza al quirófano.

a. Hipótesis Simple y Compuesta.

Se denomina HIPÓTESIS SIMPLE a cualquier hipótesis estadística que especifica completamente la distribución de la población, especifica la forma de la distribución y el valor de su parámetro.

EJEMPLO:

El ingreso mensual promedio de los empleados de cierta ciudad es μ=S/. 500, suponiendo que los ingresos mensuales se distribuyen según la normal con σ=S/. 30.0

Se denomina HIPÓTESIS COMPUESTA cuando no especifica completamente la distribución de la población.

EJEMPLO:

El ingreso promedio mensual es $\mu \neq 500$ ó $\mu < 500$ ó $\mu > 500$, con una desviación estándar de σ=S/. 30

b. Hipótesis Nula y Alternativa.

HIPÓTESIS NULA (H_0:). Se denomina hipótesis nula (H_0:) a la hipótesis que es aceptada provisionalmente como verdadera y cuya validez será sometida a comprobación experimental, algunos autores consideran como la hipótesis de ninguna diferencia.

EJEMPLO:

El rendimiento académico de estudiantes del grupo control (GC) y los rendimientos de estudiantes del grupo experimental (GE) son similares: H_0: $\mu_{GC} = \mu_{GE}$ (no hay diferencia en el rendimiento)

HIPÓTESIS ALTERNA (H_a: ó H_1:). Es la hipótesis que contradice a la H_0: a la hipótesis que se acepta en caso de que la hipótesis nula H_0: sea rechazada, a esta hipótesis se llama hipótesis de investigación.

EJEMPLO:

El rendimiento académico de estudiantes del grupo control (GC) es diferente al grupo experimental (GE). H_a: $\mu_{GC} \neq \mu_{GE}$ (Si hay diferencia en el rendimiento)

Si se asume que θ_0 es un valor del parámetro desconocido θ de una población cuya distribución se supone conocida, entonces son hipótesis nulas y alternativas, respectivamente tienen las siguientes afirmaciones:

1) $H_0:\ \theta = \theta_0 \quad \text{vs.}\ H_a:\ \theta \neq \theta_0$ **(Prueba bilateral de 2 colas)**

2) $H_0:\ \theta \leq \theta_0 \quad \text{vs.}\ H_a:\ \theta > \theta_0$ **(Prueba unilateral de cola a la derecha)**

3) $H_0:\ \theta \geq \theta_0 \quad \text{vs.}\ H_a:\ \theta < \theta_0$ **(Prueba unilateral de cola a la izquierda)**

PRUEBA DE UNA HIPÓTESIS. Es un proceso que nos conduce a tomar la decisión de aceptar o rechazar la hipótesis nula H_0, en contraposición de la hipótesis alternativa H_a. La aceptación de una hipótesis significa que los datos de la muestra no proporcionan evidencia suficiente para refutarla, en cambio el rechazo de la hipótesis significa que los datos de la muestra lo refutan.

TIPOS DE PRUEBA DE HIPÓTESIS. **El tipo de prueba depende básicamente de la hipótesis alternativa H_a:. Las hipótesis se plantearán de la siguiente forma: Prueba de una cola, se denomina a toda prueba de hipótesis donde la alternativa H_a,: es unilateral. Si la alternativa H_a: es bilateral, la prueba se denomina prueba de dos colas.**

c. Errores que se puede cometer al tomar una decisión.

La decisión para aceptar o rechazar la hipótesis nula, se basa en una estadística de prueba calculada a partir de los datos en una muestra aleatoria. Pueden producirse dos tipos de errores cuando se prueba la hipótesis.

a) **Error Tipo I:** **Es aquel que consiste en «rechazar la hipótesis nula H_0: siendo este verdadero». En términos de probabilidad es:**

$$\alpha = P\{\text{error tipo I}\} = P\{\text{rechazar } H_0 / H_0 \text{ es verdadera}\}$$

b) **Error Tipo II:** **Es aquel que consiste en «aceptar la hipótesis nula H_0: siendo este falso». En términos de probabilidad es:**

$$\beta = P\{\text{error tipo II}\} = P\{\text{aceptar } H_0 / H_0 \text{ es falsa}\}$$

Decisiones en la prueba de Hipótesis.

	H_0 es Verdadera	H_0 es falsa
Aceptación de H_0	Decisión Correcta Prob.: $1-\alpha$	Error del tipo II Prob.: β
Rechazo de H_0	Error del tipo I Prob: α	Decisión correcta Prob: $1-\beta$

La probabilidades de ocurrencia de los errores de tipo I y de tipo II tienen símbolos especiales. Algunas veces es más conveniente trabajar con la *potencia* de la prueba, donde:

$$\text{Potencia} = 1-\beta = P\{\text{rechazar } H_0 / H_0 \text{ es falsa}\}$$

EJEMPLO:

Un individuo y la naturaleza.

ACCIÓN	ESTADO DEL UNIVERSO	
	W1: Lluvioso	W2: Despejado
A1: Traer Paraguas	√	Error Tipo II
A2: No traer paraguas	Error Tipo I	√

Explicación:

Si traigo paraguas y llueve, entonces: Acerté.
Si traigo paraguas y no llueve, entonces: Error (II).
Si no traigo paraguas y llueve, entonces: Error (I).
Si no traigo paraguas y no llueve, entonces: Acerté.

d. Procedimiento de la Prueba de Hipótesis.

1. Formular la Hipótesis:

$H_0:\ \theta = \theta_0 \quad \text{vs.} \quad H_a:\ \theta \neq \theta_0$ **(Prueba bilateral de 2 colas)**

$H_0:\ \theta \leq \theta_0 \quad \text{vs.} \quad H_a:\ \theta > \theta_0$ **(Prueba unilateral de cola a la derecha)**

$H_0:\ \theta \geq \theta_0 \quad \text{vs.} \quad H_a:\ \theta < \theta_0$ **(Prueba unilateral de cola a la izquierda)**

2. Nivel de significación:

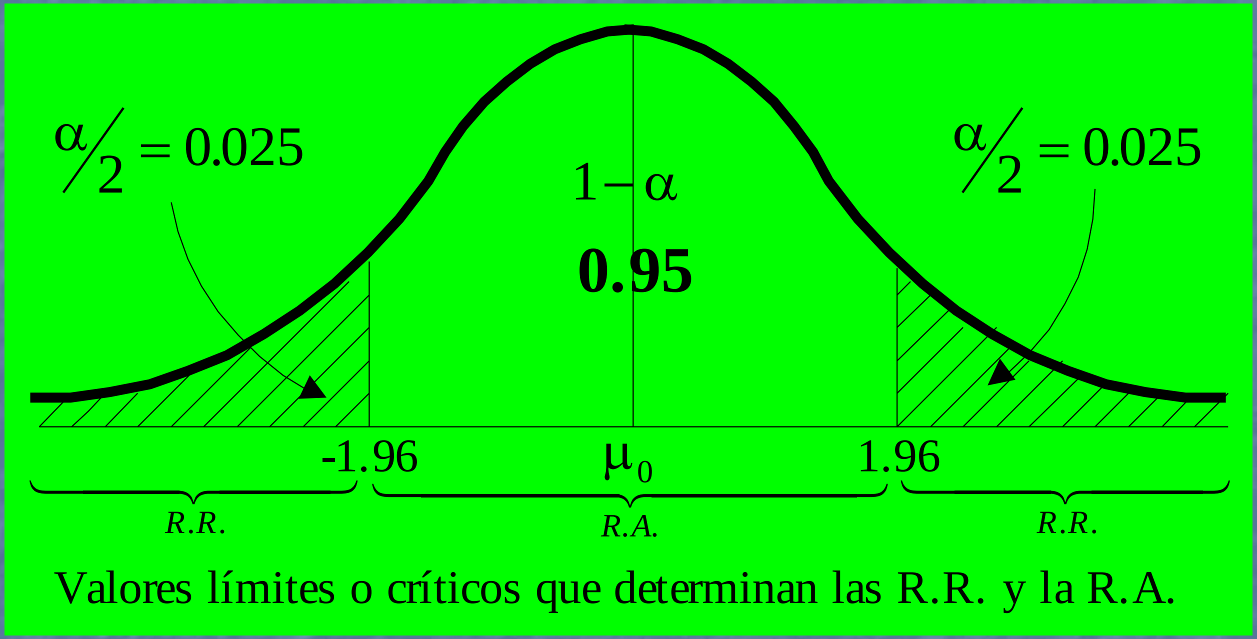
Se puede usar un valor pequeño para α, como 10%, 5%, u 1% (errores), por lo general es fijado a criterio del investigador.

3. Prueba Estadística.

Seleccionar la estadística apropiada a usar en la prueba que depende del parámetro sobre el que se realiza la hipótesis y depende de la distribución de la muestra.

4. Determinar la Región de Rechazo y Aceptación:

Establecer la regla de decisión, determinando la región crítica de la prueba.



5. Cálculo de la Prueba Estadística

Tabular y procesar estadísticamente los datos recopilados, obteniendo así un valor real calculado que se compara con el valor crítico de éste, llamado también valor tabular porque se encuentra en tablas pre-elaboradas.

6. Decisión Estadística.

Conlleva a rechazar o no rechazar la H_0 una vez comparada el valor calculado con el valor tabular al nivel de significación propuesto.

7. Conclusión:

Si se rechaza la H_0, se concluye que los datos son compatibles con la H_a, y sino, se rechaza la H_0, quiere decir que los datos son compatibles con la H_0,

3.2. Prueba de Hipótesis con una sola media de la población.

a) Cuando el valor de la Varianza σ^2 conocida.

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

EJEMPLO:

El promedio de calificaciones del examen de admisión a la universidad fue de 500 puntos y la desviación estándar de 40 puntos; hubo 16 estudiantes del Colegio A que obtuvieron un promedio de 485 puntos. ¿Podrán considerarse las calificaciones de los estudiantes del Colegio A como significativamente menores que la calificación promedio total ?. Usar $\alpha = 0.05$.

1. Hipótesis Estadística: H_0: $\mu = 500$ versus H_a: $\mu < 500$.

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

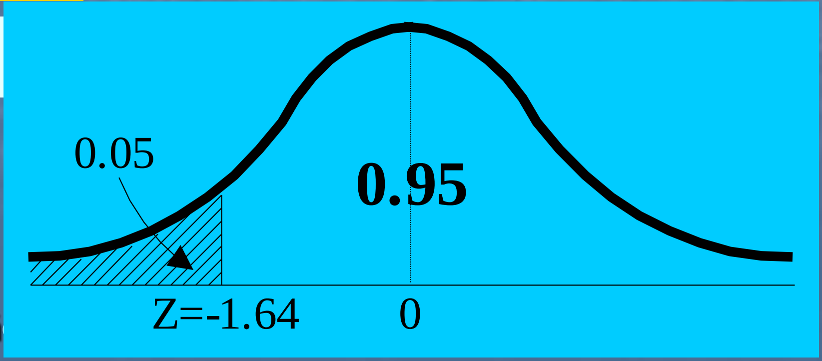
3. Distribución Muestral.

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

4. Región de Aceptación y Rechazo.

$$a)\ RA/H_0:\ Z_0 \geq -Z_\alpha = -1.64$$

$$b)\ RR/H_0:\ Z_0 < -Z_\alpha = -1.64$$



5. Cálculo de la prueba Estadística.

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{485 - 500}{40/\sqrt{16}} = -1.5$$

6. Decisión:

Como Z_0 = -1.5 pertenece a la región de aceptación, entonces se acepta la H_0: y se rechaza la H_a.

7. Conclusión.

Las calificaciones de los estudiantes del Colegio A, no pueden considerarse significativamente menores que la calificación del promedio total.

b) Cuando el valor de la Varianza σ^2 desconocida.

Si la población no tiene distribución normal y la varianza es desconocida, para probar la hipótesis acerca de la media de la población, se debe realizar mediante los siguientes estadísticos:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}, \qquad s^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1} = \frac{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}{n-1}$$

Para muestras grandes *n>30*, usar el estadístico:

$$Z = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

Para muestras pequeñas *n≤30*, usar el estadístico:

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}$$

EJEMPLO:

Se hizo un estudio de una muestra de 25 expedientes de enfermos crónicos atendidos como pacientes externos. El número medio de consultas por paciente fue de 4.8 y la desviación estándar de la muestra fue de 2. ¿Es posible concluir a partir de éstos datos que la media de la población es mayor que 4 visitas por paciente?. Suponer que la probabilidad de cometer un error de tipo I es de 0.05 ¿Cuáles son los supuestos que se deben cumplir?.

1. Hipótesis Estadística: $H_0: \mu \leq 4$ versus $H_a: \mu > 4$

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

3. Distribución Muestral.

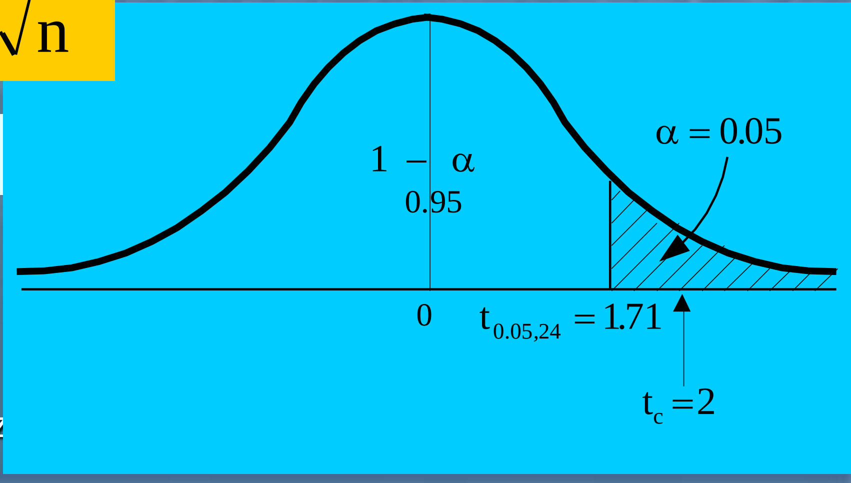
$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

$t_{0.05,24} = 1.71$ **(una cola)**

4. Región de Aceptación y Rechazo.

$$RA/H_0:\ t_c \leq t_{0.05,(24)} = 1.71$$

$$RR/H_0:\ t_c > t_{0.05,(24)} = 1.71$$



5. Cálculo de la prueba Estadística.

$$t_c = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{4.8 - 4}{2/\sqrt{25}} = \frac{5x0.8}{2} = 2$$

6. Decisión:

Decisión: Como t_c = 2 > $t_{0.05,\ 24}$ =1.71 cae en la región de rechazo, entonces se rechaza la H_0. y se acepta la H_a.

7. Conclusión.

Los expedientes de enfermos crónicos atendidos como pacientes externos de la población superan a 4 visitas por paciente.

Intervalo Confidencial (IC).

$$t_{\alpha/2,(n-1)} = t_{0.025,24} = 2.064$$

$$P\left(4.8 - (2.064)\frac{2}{\sqrt{25}} \le \mu \le 4.8 + (2.064)\frac{2}{\sqrt{25}}\right) = 0.95$$

$$P(3.97 \le \mu \le 5.63) = 0.95$$

3.3. Prueba de Hipótesis de dos medias poblacionales independientes de σ^2_1 y σ^2_2 conocidas.

Hipótesis Estadística:

a) $H_0: \mu_1 = \mu_2$

$\underbrace{H_a: \mu_1 \neq \mu_2}_{\text{Bilaterales}}$

b) $H_0: \mu_1 \geq \mu_2$

$\underbrace{H_a: \mu_1 < \mu_2}_{\text{Unilateral a la izquierda}}$

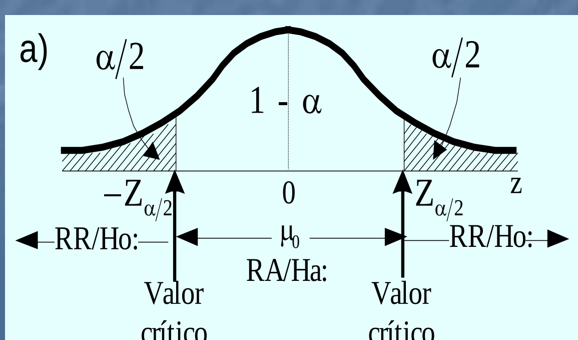
c) $H_0: \mu_1 \leq \mu_2$

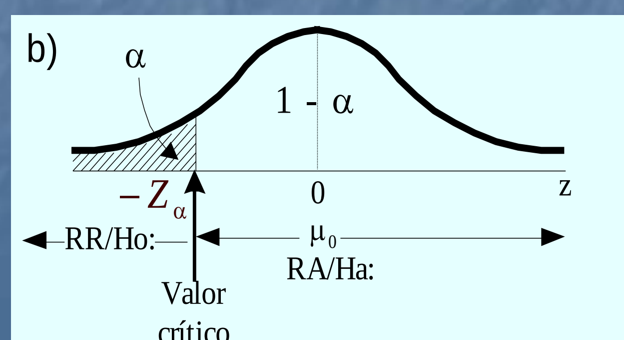
$\underbrace{H_a: \mu_1 > \mu_2}_{\text{Unilateral a la derecha}}$

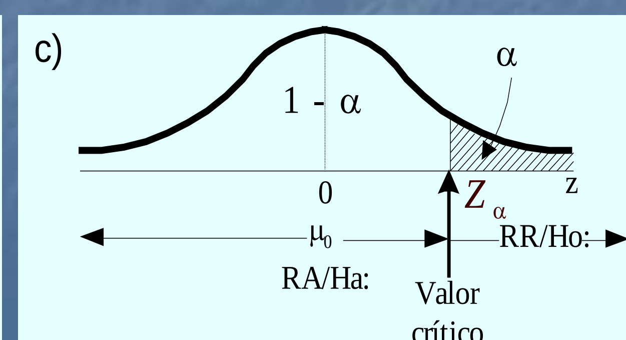
Prueba Estadística.

$$Z = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Región de rechazo y aceptación.







3.4. Prueba de Hipótesis de dos medias poblacionales normales independientes de varianzas σ^2_1 y σ^2_2 desconocidas.

Los parámetros μ_1, μ_2, σ^2_1 y σ^2_2 se estiman mediante los siguientes estadísticos:

$$\overline{X}_1 = \frac{\sum_{i=1}^{n_1} X_i}{n_1}, \qquad s_1^2 = \frac{\sum_{i=1}^{n_1} \left(X_i - \overline{X}\right)^2}{n_1 - 1} = \frac{\sum_{i=1}^{n_1} X_i^2 - \frac{(\sum_{i=1}^{n_1} X_i)^2}{n_1}}{n_1 - 1}$$

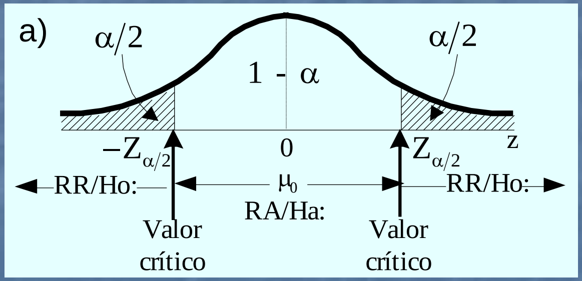
$$\overline{X}_2 = \frac{\sum_{j=1}^{n_2} Y_j}{n_2}, \qquad s_2^2 = \frac{\sum_{j=1}^{n_2} \left(Y_j - \overline{Y}\right)^2}{n_2 - 1} = \frac{\sum_{j=1}^{n_2} Y_j^2 - \frac{(\sum_{j=1}^{n_2} Y_i)^2}{n_2}}{n_2 - 1}$$

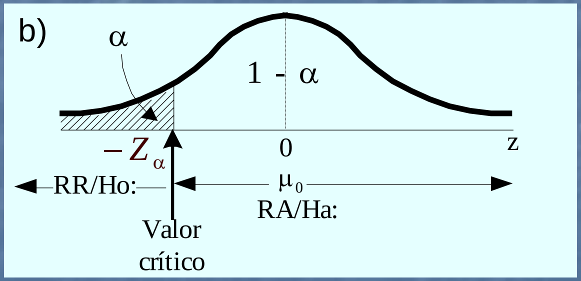
Caso I: Tamaño de muestra grandes $n_1+n_2>30$

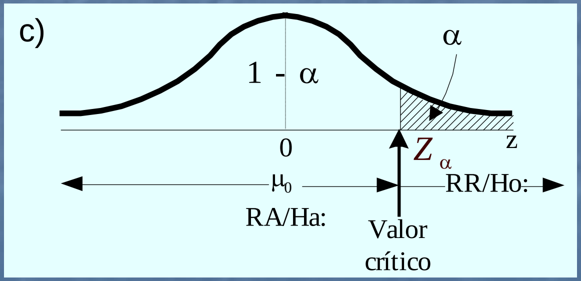
Prueba Estadística.

$$Z = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \sim N(0,1)$$

Región de Rechazo y Aceptación.







a) $RA/H_o: \ -Z_{\alpha/2} \leq Z_o \leq Z_{\alpha/2}$

$RR/H_o: \ |Z_o| < -Z_{\alpha/2} \text{ ó } |Z_o| > Z_{\alpha/2}$

b) $RA/H_0: \ Z_0 \geq -Z_\alpha$

$RR/H_0: \ Z_0 < -Z_\alpha$

c) $RA/H_0: \ Z_0 \leq Z_\alpha$

$RR/H_0: \ Z_0 > Z_\alpha$

EJEMPLO:

Se realizó el trabajo de investigación en el IES Aplicación "José Carlos Mariátegui" de la UNA Puno, considerando dos secciones de segundo grado «B» (Grupo Control) con 31 alumnos y «A» (Grupo Experimental) con 32 alumnos, a los cuales se les aplicó la prueba de salida, con la finalidad de identificar los aprendizajes significativos que poseen sobre el componente Historia y Sociedad, obteniéndose los siguientes resultados que se muestran a continuación: Sea α=0.05 y encuentre el intervalo confidencial al 95%.

1. Hipótesis Estadística:

$$H_0: \mu_A = \mu_B \text{ versus } H_a: \mu_A \neq \mu_B$$

APRENDIZAJE DE HISTORIA Y SOCIEDAD CATEGORIAS		GRUPO CONTROL					GRUPO EXPERIMENTAL				
Escala Cualitativa	Escala Cuantitativa	n_i	%	Yi	n_iYi	$n_iY^2_i$	n_i	%	Yi	n_iYi	$n_iY^2_i$
Logo destacado	[17 - 20]	2	6.45	18.5	37.0	684.5	6	18.75	18.5	111.0	2053.5
Logro	[13 - 16]	8	25.81	14.5	116.0	1682.0	23	71.88	14.5	333.5	4835.8
Proceso	[11 - 12]	18	58.06	11.5	207.0	2380.5	2	6.25	11.5	23.0	264.5
Inicio	[00 - 10]	3	9.68	5.0	15.0	75.0	1	3.13	5.0	5.0	25.0
TOTAL		31	100.00		**375.0**	4822.0	32	100.00		**472.5**	7178.8

5. Cálculo de la prueba Estadística.

$$\overline{Y}_C = \frac{\sum_{i=1}^{4} n_i Y_i}{n} = \frac{375}{31} = 12.097$$

$$S_C^2 = \frac{\sum_{i=1}^{4} n_i Y_i^2 - \frac{(\sum_{i=1}^{4} n_i Y_i)^2}{n}}{n} = \frac{4822 - \frac{(375)^2}{31}}{31} = 9.2164$$

$$\overline{Y}_E = \frac{\sum_{i=1}^{4} n_i Y_i}{n} = \frac{472.5}{32} = 14.766$$

$$S_E^2 = \frac{\sum_{i=1}^{4} n_i Y_i^2 - \frac{(\sum_{i=1}^{4} n_i Y_i)^2}{n}}{n} = \frac{7178.8 - \frac{(472.5)^2}{32}}{32} = 6.3138$$

$$Z_c = \frac{\left(\overline{X}_C - \overline{X}_E\right)}{\sqrt{\frac{S_C^2}{n_C} + \frac{S_E^2}{n_E}}} = \frac{12.097 - 14.766}{\sqrt{\frac{9.21644}{31} + \frac{6.3138}{32}}} = \frac{-2.669}{0.703286} = -3.795$$

Se puede concluir que existe evidencia suficiente para afirmar que el promedio aritmético de las notas de los alumnos del grupo experimental es diferente al promedio de las notas de los alumnos del grupo control, obtenidas en la prueba de salida.

Intervalo Confidencial (IC).

$$\Pr(-4.047 \leq \mu_C - \mu_E \leq -1.291) = 0.95$$

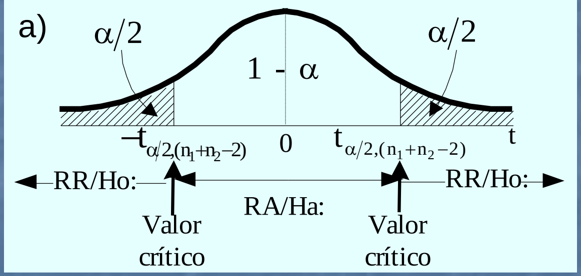
Caso II: a) Tamaño de muestra pequeña $n_1+n_2 \leq 30$

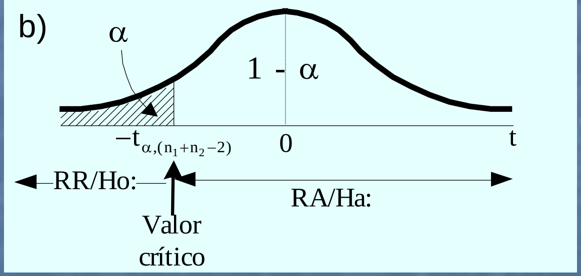
Cuando: $\sigma_1^2 = \sigma_2^2 = \sigma^2$

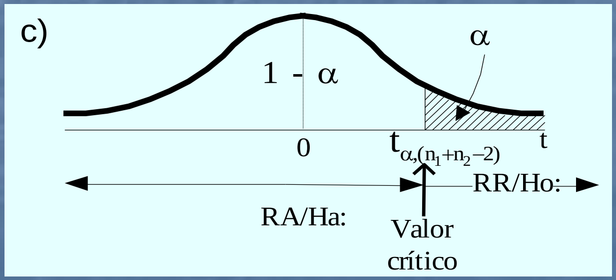
Prueba Estadística.

$$t = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}} \sim t_{(n_1+n_2-2)}$$

Región de Rechazo y Aceptación.







a) $RA/H_0:\ -t_{\alpha/2,(n_1+n_2-2)} \leq t_c \leq t_{\alpha/2,(n_1+n_2-2)}$

$RR/H_0:\ |t_c| < -t_{\alpha/2,(n_1+n_2-2)}$ ó $|t_c| > t_{\alpha/2,(n_1+n_2-2)}$

b) $RA/H_0:\ t_c \geq -t_{\alpha,(n_1+n_2-2)}$

$RR/H_0:\ t_c < -t_{\alpha,(n_1+n_2-2)}$

c) $RA/H_0:\ t_c \leq t_{\alpha,(n_1+n_2-2)}$

$RR/H_0:\ t_c > t_{\alpha,(n_1+n_2-2)}$

EJEMPLO:

El editor asegura que los estudiantes que reciben instrucción en matemáticas usando un nuevo texto, obtendrán mejor rendimiento que los que usan el texto antiguo. Se seleccionan aleatoriamente para Nuevo Texto 16 estudiantes con promedio de $X_1 = 83$, $S_1 = 7$ y el otro Texto Antiguo con 20 estudiantes con promedio $X_2 = 72$, $S_2 = 5$. Usando $\alpha = 0.01$, puede usted concluir que la aseveración del editor es correcta. Calcule el intervalo confidencial del 99% e interprete los resultados.

1. Hipótesis Estadística: H_0: $\sigma^2_1 = \sigma^2_2$ versus H_a: $\sigma^2_1 \neq \sigma^2_2$

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

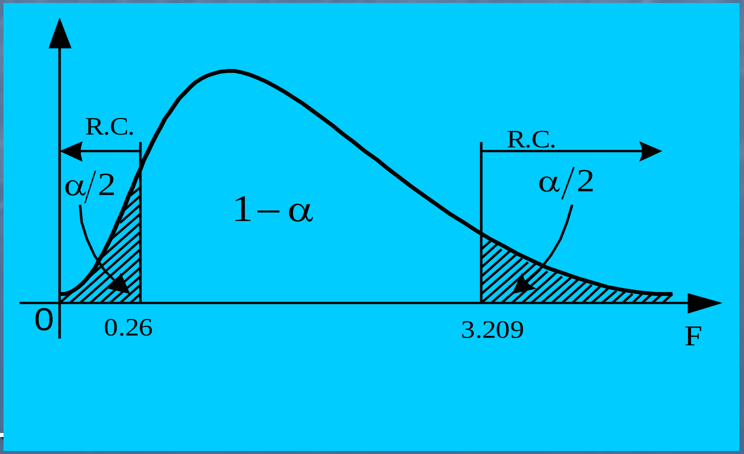
3. Distribución Muestral.

$$F = \frac{S_1^2}{S_2^2} \sim F_{(n_1-1),(n_2-1)}$$

4. Región de Aceptación y Rechazo.

$$F_{0.995,15,19} = \frac{1}{F_{0.005,19,15}} = \frac{1}{3.913} = 0.256$$

$$F_{0.005,15,19} = 3.58$$



Vladimiro Ibañez Q.

5. Cálculo de la prueba Estadística.

$$F_c = \frac{S_1^2}{S_2^2} = \frac{49}{25} = 1.96$$

6. Decisión:

Como F_c=1.96 cae en la región aceptación, entonces se acepta la Ho: y se rechaza la Ha:

7. Conclusión.

Estamos en condiciones de realizar el contraste de promedios, una vez comprobado que las varianzas son homogéneas ($\sigma^2_1 = \sigma^2_2 = \sigma^2$).

PROCEDIMIENTO:

1. Hipótesis Estadística: $H_0: \mu_1 = \mu_2$ versus $H_a: \mu_1 \neq \mu_2$

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

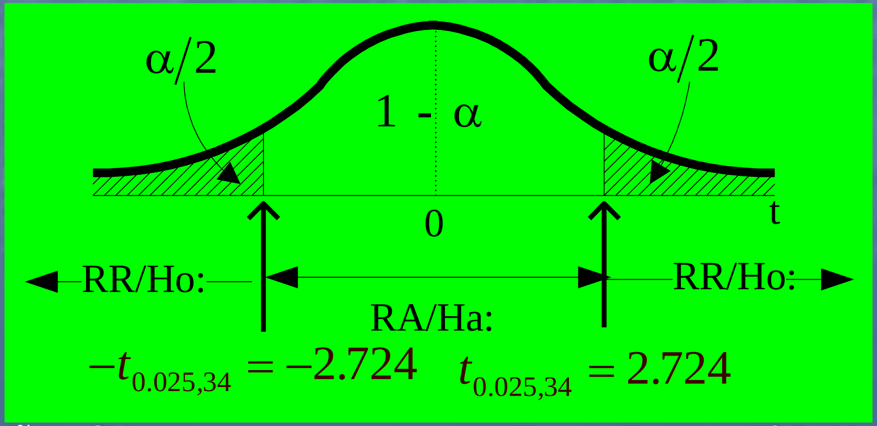
3. Distribución Muestral.

$$t = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}} \sim t_{(n_1+n_2-2)}$$

4. Región de Aceptación y Rechazo.

$$RA/H_o:\ -2.724 \leq t_c \leq 2.724$$

$$RR/H_o:\ |t_c| < -2.724 \ \ ó \ \ |t_c| > 2.724$$



5. Cálculo de la prueba Estadística.

$$t_c = \frac{83-79}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{4}{\sqrt{35.588\left(\frac{1}{16}+\frac{1}{20}\right)}} = \frac{4}{2.2098} = 1.999$$

$$S_c^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} = \frac{15\times 49+19\times 25}{34} = 35.588$$

6. Decisión:

Como la t_c = 1.99 cae en la región de aceptación, entonces se acepta la H_0: y se rechaza H_a:

7. Conclusión.

Existe evidencia suficiente para afirmar que los estudiantes de matemáticas con el texto nuevo y antiguo es indiferente en el rendimiento, esto implica que la aseveración del editor es incorrecta. sanos.

Intervalo Confidencial (IC).

$$\Pr\left(-1.4505 \le \mu_1 - \mu_1 \le 9.45\right) = 0.99$$

Caso II: b) Tamaño de muestra pequeña $n_1+n_2 \leq 30$

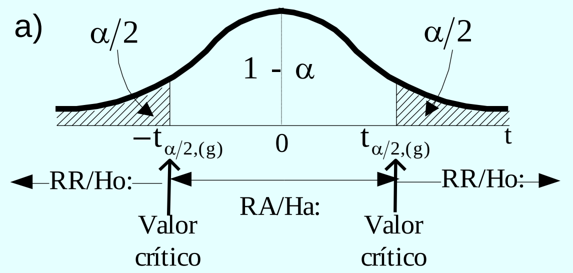
Cuando: $\sigma_1^2 \neq \sigma_2^2$

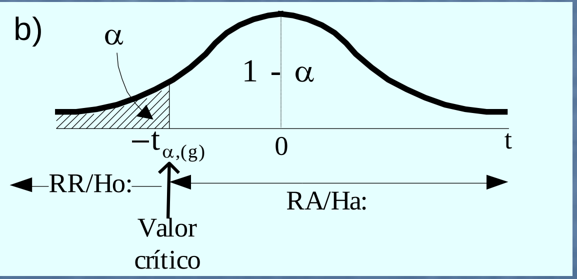
Prueba Estadística.

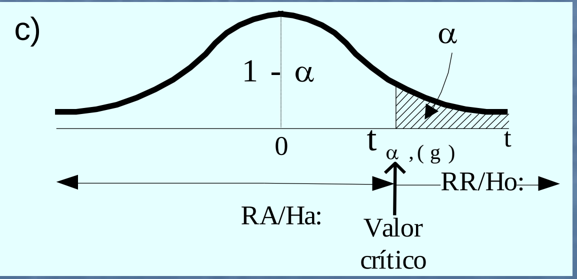
$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)}} \sim t_{(g)}$$

$$g = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Región de Rechazo y Aceptación.







a) $RA/H_0:\ -t_{\alpha/2,(g)} \leq t_c \leq t_{\alpha/2,(g)}$

$RR/H_0:\ |t_c| < -t_{\alpha/2,(g)}$ ó $|t_c| > t_{\alpha/2,(g)}$

b) $RA/H_0:\ t_c \geq -t_{\alpha,(g)}$

$RR/H_0:\ t_c < -t_{\alpha,(g)}$

c) $RA/H_0:\ t_c \leq t_{\alpha,(g)}$

$RR/H_0:\ t_c > t_{\alpha,(g)}$

EJEMPLO:

Un investigador está interesado en saber si los niños nacidos prematuramente con acidosis metabólica tardía y los niños prematuros que no tienen dicha enfermedad difieren en lo que respecta a las concentraciones en la orina de cierta sustancia química. Las concentraciones medias, desviación estándar y el tamaño de la muestra para ambos grupos son las siguientes:

¿Qué puede concluir el investigador con base en estos resultados?. Sea α =0.05. Se debe probar si las varianzas es: $\sigma^2_1=\sigma^2_2 = \sigma^2$ ó $\sigma^2_1 \neq \sigma^2_2$

Muestra	n	$\overline{X}$	s
Con la condición	15	8.50	5.5
Sin la condición	14	4.80	2.9

1. Hipótesis Estadística: H_0: $\sigma^2_1 = \sigma^2_2$ versus H_a: $\sigma^2_1 \neq \sigma^2_2$

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

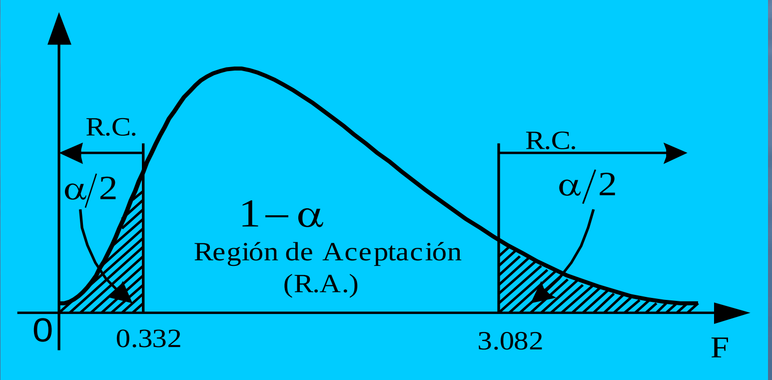
3. Distribución Muestral.

$$F = \frac{S_1^2}{S_2^2} \sim F_{(n_1-1),(n_2-1)}$$

4. Región de Aceptación y Rechazo.

$$F_{0.975,14,13} = \frac{1}{F_{0.025,13,14}} = \frac{1}{3.012} = 0.332$$

$$F_{0.025,14,13} = 3.082$$



5. Cálculo de la prueba Estadística.

$$F_c = \frac{(5.5)^2}{(2.9)^2} = \frac{30.25}{8.41} = 3.59 \sim 3.60$$

6. Decisión:

Como F_c=3.60 cae en la región rechazo, entonces se acepta la H_a: y se rechaza la H_0:

7. Conclusión.

Existe evidencia suficiente para afirmar que existe diferencia entre $\sigma^2_1 \neq \sigma^2_2$. Entonces, se prosigue a efectuar la prueba de la diferencia de promedios.

PROCEDIMIENTO:

1. Hipótesis Estadística: $H_0: \mu_1 = \mu_2$ versus $H_a: \mu_1 \neq \mu_2$

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

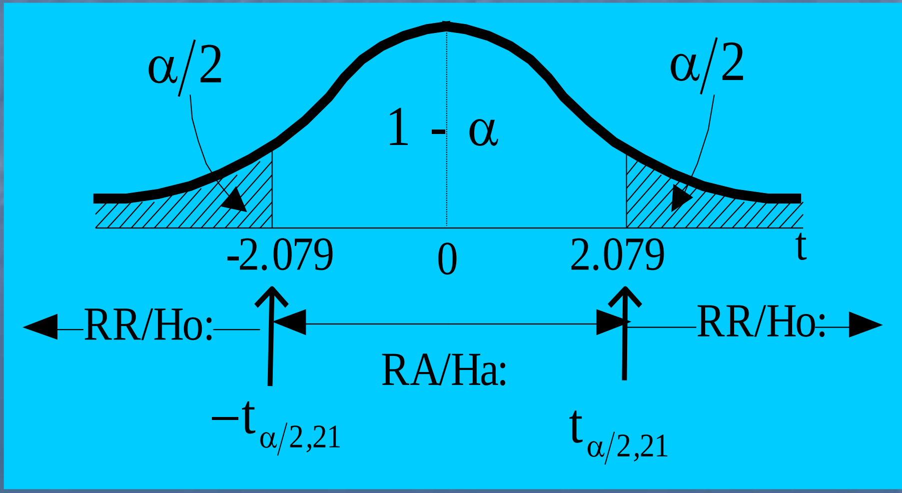
3. Distribución Muestral.

$$t = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)}} \sim t_{(g)}$$

4. Región de Aceptación y Rechazo.

$$RA/H_o: \ -t_{\alpha/2,(21)} = -2.079 \leq t_c \leq t_{\alpha/2,(21)} = 2.079$$

$$RR/H_o: \ |t_c| < -t_{\alpha/2,(21)} = -2.079 \ \text{ó} \ |t_c| > t_{\alpha/2,(21)} = 2.079$$



5. Cálculo de la prueba Estadística.

$$t_c = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)}} = \frac{8.5 - 4.8}{\sqrt{\left(\frac{30.25}{15} + \frac{8.41}{14}\right)}} = \frac{3.7}{1.6178} = 2.287$$

$$g = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}} = \frac{\left(\frac{30.25}{15} + \frac{8.41}{14}\right)^2}{\frac{\left(\frac{30.25}{n_1}\right)^2}{14} + \frac{\left(\frac{8.41}{n_2}\right)^2}{13}} = \frac{6.85068}{0.31825} = 21.5 \sim 21$$

6. Decisión:

Como $t_c = 2.29$ cae en la región de rechazo, entonces se rechaza la H_0: y se acepta la H_a:

7. Conclusión.

Existe evidencia suficiente para afirmar que hay diferencia entre los niños nacidos prematuramente tardía y los niños prematuros que no tienen dicha enfermedad.

Intervalo Confidencial (IC).

$$Pr\ 0.3355 \le \mu_1 - \mu_2 \le 7.06447\ = 0.95$$

3.5. Prueba de Hipótesis de dos medias con observaciones aparejadas o relacionadas.

La media y la varianza de la muestra de «n» diferencias $D_1, D_2, \ldots, D_n$ esta dada por:

$$\overline{D} = \frac{\sum_{i=1}^{n} D_i}{n}, \quad S_D^2 = \frac{\sum_{i=1}^{n} D_i^2 - \frac{\left(\sum D_i\right)^2}{n}}{n-1} = \frac{\sum_{i=1}^{n} \left(D_i - \overline{D}\right)^2}{n-1}$$

$$\therefore \ \overline{D} \sim N(\mu_D, \sigma_D^2/n)$$

Cuando hallamos un estimador de la diferencia entre dos medias poblacionales ($\mu_D = \mu_1 - \mu_2$), se pueden presentarse dos casos:

Caso I: Tamaño de muestra grande n>30

Cuando n>30, la distribución de «t» de Student se apróxima a la distribución N(0,1).

Hipótesis Estadística.

a) $$\underbrace{\begin{array}{l} H_0: \mu_D = 0 \\ H_a: \mu_D \neq 0 \end{array}}_{\text{Bilaterales}}$$

b) $$\underbrace{\begin{array}{l} H_0: \mu_D \geq 0 \\ H_a: \mu_D < 0 \end{array}}_{\text{Unilateral a la izquierda}}$$

c) $$\underbrace{\begin{array}{l} H_0: \mu_D \leq 0 \\ H_a: \mu_D > 0 \end{array}}_{\text{Unilateral a la derecha}}$$

Estadígrafo de contraste:

$$Z = \frac{\overline{D} - (\mu_1 - \mu_2)}{S_D / \sqrt{n}} \sim N(0,1)$$

$$Z_0 = \frac{\overline{D}}{S_D / \sqrt{n}}$$

Caso II: Tamaño de muestra pequeña n≤30

Cuando n≤30, se usa la distribución de «t» de Student con $t_{(n-1)}$.

Hipótesis Estadística.

a) $\underbrace{H_0: \mu_D = 0 \quad H_a: \mu_D \neq 0}_{\text{Bilaterales}}$

b) $\underbrace{H_0: \mu_D \geq 0 \quad H_a: \mu_D < 0}_{\text{Unilateral a la izquierda}}$

c) $\underbrace{H_0: \mu_D \leq 0 \quad H_a: \mu_D > 0}_{\text{Unilateral a la derecha}}$

Estadígrafo de contraste:

$$t = \frac{\overline{D} - (\mu_1 - \mu_2)}{S_D / \sqrt{n}} \sim t_{(n-1)}$$

$$t_c = \frac{\overline{D}}{S_D / \sqrt{n}}$$

EJEMPLO:

Veinte estudiantes de matemática I de la Facultad de Ingeniería Estadística e Informática de la UNA fueron divididas en 10 parejas, teniendo cada miembro de la pareja aproximadamente el mismo cociente de inteligencia. Uno de cada pareja se selecciona al azar y se asigna a una sección que utiliza videos. El otro miembro se asigna a una sección que cuenta con profesor. Al finalizar el ciclo ambos grupos se presentan al mismo examen, obteniéndose los resultados siguientes:

Pareja	1	2	3	4	5	6	7	8	9	10	Total
Con video (V)	15	12	17	11	18	15	16	13	14	10	141
Con profesor (P)	16	10	17	14	17	16	18	12	15	11	146

Suponiendo que la característica en estudio es normal, obtener el intervalo de confianza del 95% para la diferencia real en el promedio de calificaciones de los dos procedimientos de enseñanza.

Solución:

Pareja	Con video (V)	Con profesor (P)	D = V - P	D^2
1	15	16	-1	1
2	12	10	2	4
3	17	17	0	0
4	11	14	-3	9
5	18	17	1	1
6	15	16	-1	1
7	16	18	-2	4
8	13	12	1	1
9	14	15	-1	1
10	10	11	-1	1
Total	$\sum V = 141$	$\sum P = 146$	$\sum D_i = -5$	$\sum D_i^2 = 23$

PROCEDIMIENTO:

1. Hipótesis Estadística: H_0: $\mu_D = 0$ versus H_a: $\mu_D \neq 0$

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

3. Estadígrafo de contraste.

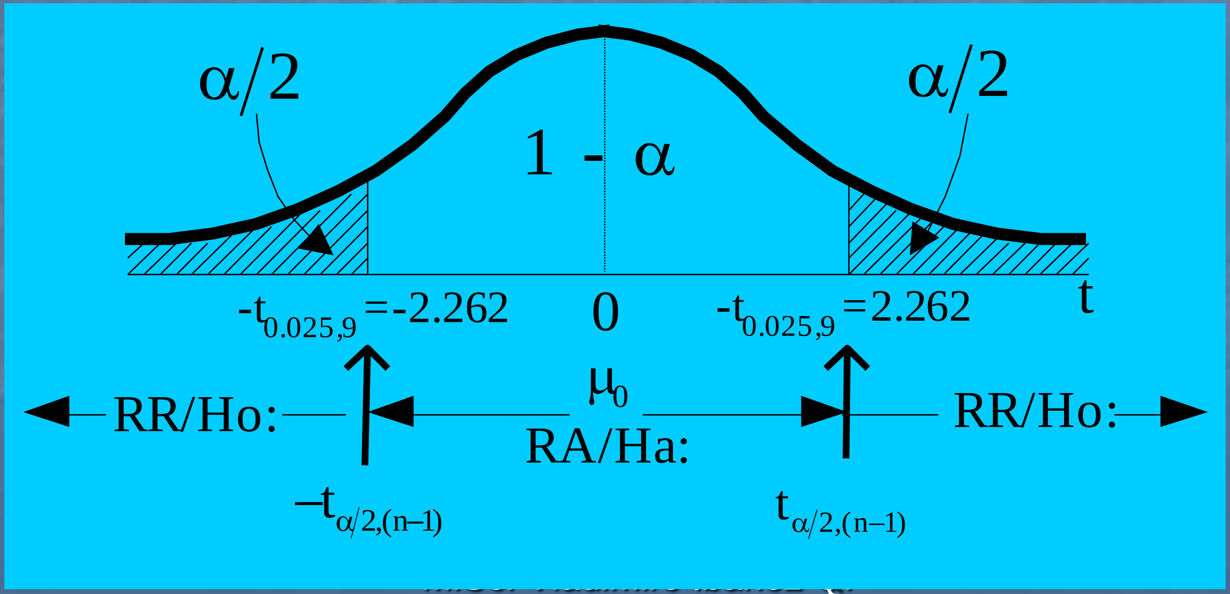
$$t = \frac{\overline{D} - (\mu_1 - \mu_2)}{S_D / \sqrt{n}}$$

4. Región de Aceptación y Rechazo.

$$RA/H_o: \; -t_{\alpha/2,(n-1)} \le t_c \le t_{\alpha/2,(n-1)}$$

$$RR/H_o: \; t_c < -t_{\alpha/2,(n-1)} \text{ ó } t_c > t_{\alpha/2,(n-1)}$$

$$t_{\alpha/2,(n-1)} = t_{0.05/2,(10-1)} = t_{0.025,9} = 2.262$$



5. Cálculo de la prueba Estadística.

$$\overline{X}_V = \frac{\sum_{i=1}^{10} V_i}{n} = \frac{141}{10} = 14.1 \qquad \overline{X}_P = \frac{\sum_{i=1}^{10} P_i}{n} = \frac{146}{10} = 14.6$$

$$\overline{D} = \overline{X}_V - \overline{X}_P = 14.1 - 14.6 = -0.5$$

$$S_D^2 = \frac{\sum_{i=1}^{n} D_i^2 - \frac{\left(\sum D_i\right)^2}{n}}{n-1} = \frac{23 - \frac{(-5)^2}{10}}{9} = \frac{20.5}{9} = 2.278$$

S_D = 1.509230856

$$|t_c| = \frac{\overline{D}}{S_D / \sqrt{n}} = \frac{-0.5}{1.50923085 / \sqrt{10}} = -1.047$$

6. Decisión:

Como t_c = 1.047 cae en la región de aceptación, entonces se acepta la H_0: y rechazamos la H_a:

7. Conclusión.

Se puede concluir que los estudiantes tienen el mismo cociente de inteligencia.

Intervalo Confidencial (IC).

$$P\ -1.5795637 \le \mu_D \le 0.579563708 = 0.95$$

3.6. Prueba de Hipótesis para proporción de una sola población (P)

Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria de tamaño «n», seleccionada de una población Bernoulli B(1,p), donde el parámetro desconocido «p» es la proporción de éxitos en la población:

Parámetro:

$$P = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{X}{N} = \frac{\text{Número de casos favorables}}{N}$$

Estimador:

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x}{n} = \frac{\text{Número de éxitos en la muestra}}{n}$$

Donde $\hat{p}$ es el estimador puntual de la proporción de la población que pertenece a esta clase,

$$\hat{p} \sim N\left[P, \frac{PQ}{n}\right]$$

Caso I: Tamaño de muestra grande n>30

$$\hat{p} \sim N\left[P, \frac{PQ}{n}\right]$$

Hipótesis Estadística.

a) $H_0: p = p_0$

$\underbrace{H_a: p \neq p_0}_{\text{Bilaterales}}$

b) $H_0: p \geq p_0$

$\underbrace{H_a: p < p_0}_{\text{Unilateral a la izquierda}}$

c) $H_0: p \leq p_0$

$\underbrace{H_a: p > p_0}_{\text{Unilateral a la derecha}}$

Estadígrafo de contraste:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} \sim N(0,1), \quad \hat{p} \sim \left(p_0, \frac{p_0(1-p_0)}{n}\right)$$

$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$$

Región de aceptación y rechazo:

Si Z_0 pertenece a RA/H_0, aceptamos H_0: y rechazamos la H_a:

Si Z_0 pertenece a RR/H_0, aceptamos H_0: y aceptamos la H_a:

EJEMPLO:

En una muestra de 150 pacientes admitidos a un hospital de urgencias con cierto diagnóstico, 128 de ellos presentaron vómitos. ¿Proporcionan estos datos la evidencia suficiente para indicar, al nivel 0.01 de significación, que la proporción de la población es menor que 0.92?.

1. Hipótesis Estadística:

$$H_0:\ P \geq \underbrace{0.92}_{p_0} \quad versus \quad H_a:\ P < \underbrace{0.92}_{p_0}$$

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

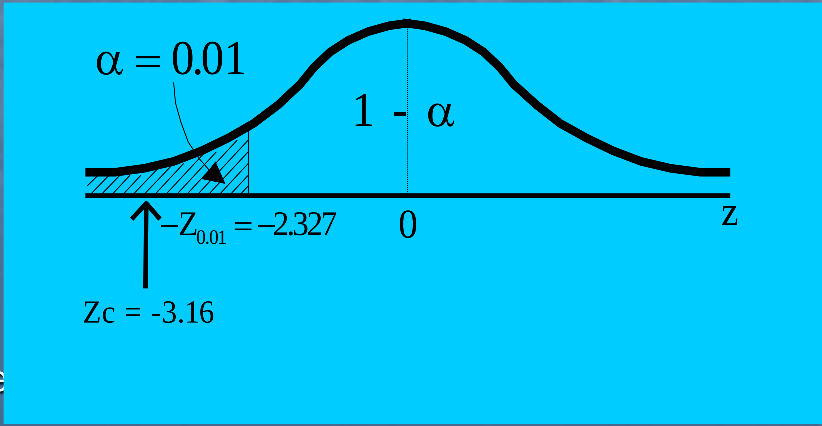
3. Estadígrafo de contraste: **n = 150, x = 128 presentaron vómitos.**

$$Z_C = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$$

4. Región de Aceptación y Rechazo.

$$RA/H_o:\ Z_C \geq -Z_{0.01} = -2.327$$

$$RR/H_o:\ Z_C < -Z_{0.01} = -2.327$$



5. Cálculo de la prueba Estadística.

$$Z_C = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.85-0.92}{\sqrt{\frac{(0.92)(1-0.92)}{150}}} = \frac{-0.07}{0.02215} = -3.16$$

$$\hat{p} = \frac{128}{150} = 0.85$$

$$\hat{q} = 1 - 0.92 = 0.08$$

6. Decisión:

Dado que $Z_c = -3.16 < -Z_{0.01} = -2.327$ cae en la región de rechazo, entonces rechazamos la H_0: y se acepta la Ha:

7. Conclusión.

Se puede afirmar que es evidente que los pacientes admitidos a un hospital de urgencias que presentaron vómitos es inferior al 92%.

Intervalo Confidencial (IC).

$$P\left(0.868 \leq \hat{p} \leq 0.9715\right) = 0.99$$

Caso II: Tamaño de muestra pequeña n≤30

A: Para ambas colas o bilateral. **Si** $\hat{p} < p_0$ **entonces se calcula:**

$$P = P\ \ X \le x/p = p_0\ \ = \sum_{x=0}^{k} \binom{n}{x} p_o^x (1 - p_0)^{n-x}$$

Si $\hat{p} > p_0$ **entonces se calcula:**

$$P = P\ \ X \ge x/p = p_0\ \ = \sum_{x=0}^{k} \binom{n}{x} p_o^x (1 - p_0)^{n-x}$$

Se rechaza H_0, si $P \le \alpha$, en caso contrario no se rechaza.

B: Para la cola derecha (unilateral)

$$P = P\ \ X \le x/p = p_0\ \ = \sum_{k=x}^{n} \binom{n}{k} p_o^k (1 - p_0)^{n-k}$$

Se rechaza H_0, si $P \le \alpha$, en caso contrario no se rechaza.

C: Para la cola izquierda (unilateral)

$$P = P\ \ X \le x/p = p_0\ \ = \sum_{k=0}^{x} \binom{n}{k} p_o^k (1 - p_0)^{n-k}$$

EJEMPLO:

Se afirma que cierto medicamento que se prescribe para aliviar determinada enfermedad es efectivo en más del 80% de los casos. Al parecer esta afirmación es exagerada por lo que se suministra tal medicamento a una muestra aleatoria de 15 pacientes resultando que 13 de ellos han experimentado alivio. ¿Esta afirmación es suficiente para concluir que realmente el medicamento es efectivo en más del 80% de los casos al nivel de significación del 5%?.

SOLUCIÓN:

Sea X el número de pacientes que se sanan en n = 15 casos. Entonces X~B(15,p) donde «p» es el porcentaje de pacientes que se sanan en la población de todos los pacientes que sufren de enfermedad.

1. **Hipótesis Estadística:** **H_o: p = 0.80 Vs. H_a: p >0.80**

2. **Nivel de significación:** **α = 0.05**

3. **Estadígrafo de contraste:** **Si la H_o: es cierta, la variable X tiene distribución binomial con n = 15, p = 0.80**

4. **Región de Aceptación y Rechazo:** **Se rechazará H_o: a favor de H_a:, si el valor de:**

$$P = P\left(X \geq 13/\, p = 0.80\right) < \alpha = 0.05, \ \ (\mathrm{p} < \alpha)$$

5. **Cálculo de la prueba Estadística.**

$$P = P\ \ X \geq 13/\, p = 0.80\ \ = \sum_{k=13}^{15} \binom{15}{k} (0.80)^{k} (0.20)^{15-k} = 0.397$$

6. **Decisión:** **Dado que P = 0.3970 > = 0.05, entonces no se rechaza la H_o:**

7. **Conclusión.**

Se puede concluir que el medicamento permite aliviar determinada enfermedad en los pacientes.

3.7. Prueba de Hipótesis para diferencia de dos proporciones (P)

Sean $X_1, X_2, ..., X_{n1}$ una muestra aleatoria de una población Bernoullí $B(1,p_1)$ y $Y_1, Y_2, .., Y_{n2}$ una muestra aleatoria de una población Bernoullí $B(1,p_2)$, los parámetros desconocidos p_1 y p_2 son las proporciones de éxitos poblacionales respectivos. Supongamos que las poblaciones son independientes y sean las proporciones de éxito muestrales:

$$\hat{p}_1 = \frac{\sum_{i=1}^{n_1} X_i}{n_1} = \frac{X_1}{n_1} = \frac{\text{Número de éxitos de la muestra 1}}{n_1}$$

$$\hat{p}_2 = \frac{\sum_{j=1}^{n_2} y_j}{n_2} = \frac{X_2}{n_2} = \frac{\text{Número de éxitos de la muestra 2}}{n_2}$$

Hipótesis Estadística.

a) $H_0: p_1 = p_2 = p$

$\underbrace{H_a: p_1 \neq p_2}_{\text{Bilaterales}}$

b) $H_0: p_1 \geq p_2$

$\underbrace{H_a: p_1 < p_2}_{\text{Unilateral a la izquierda}}$

c) $H_0: p_1 \leq p_2$

$\underbrace{H_a: p_1 > p_2}_{\text{Unilateral a la derecha}}$

Estadígrafo de contraste:

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}} \sim N(0,1),$$

Si H_o: $p_1 = p_2$ es cierta, $p_1 = p_2 = p$

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1),$$

Para calcular :

$$Z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{\pi}(1-\hat{\pi})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

se debe estimar el valor de «p» (valor común de los parámetros p_1 y p_2). Cuya estimación insesgada de «p» es:

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}, \quad \text{ó} \quad \hat{\pi} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

Es un estimador combinado de la proporción de la población cuando H_o: es VERDADERA.

EJEMPLO:

Un alumno que realiza su tesis de grado en Veterinaria, cree que los cerdos criollos de la región de Puno están infestados con Cisticercosis en un 15% más que los cerdos de la región del Cusco. Para comprobar la suposición el alumno hace un seguimiento al azar de 98 cerdos sacrificados en la región de Puno y 118 en la región del Cusco, encontrando 64 cerdos con está enfermedad en Puno, y 60 en el Cusco. Con los datos disponibles se podrá corroborar la sospecha del alumno?

1. Hipótesis Estadística: $H_o: P_P - P_C \leq 0.15$ Vs. $H_a: P_P - P_C > 0.15$

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

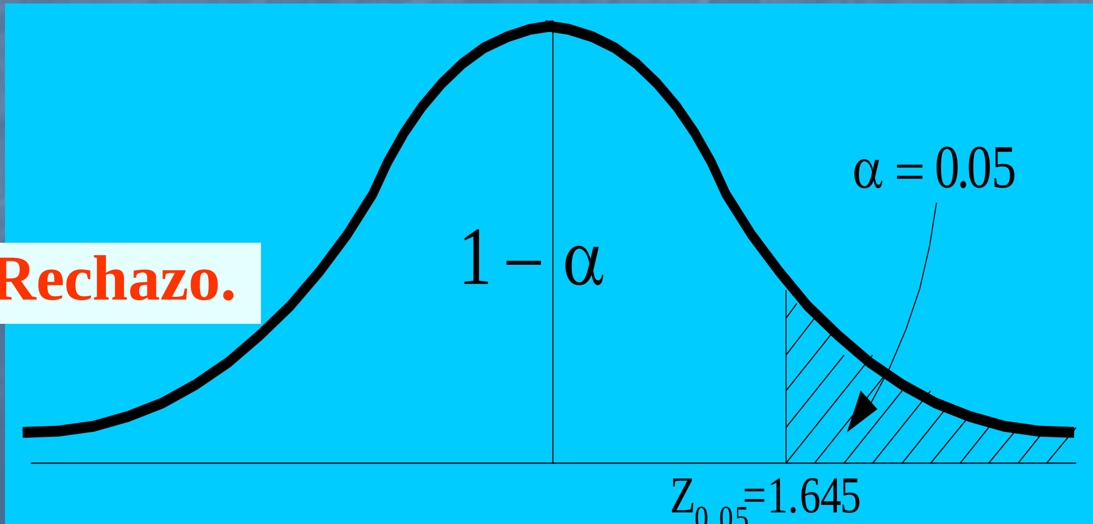
3. Estadígrafo de contraste:

$$Z_0 = \frac{(\hat{p}_1 - \hat{p}_2) - p_0}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}}$$

4. Región de Aceptación y Rechazo.

$$RA/H_0:\ Z_0 \leq Z_{\alpha/2}$$

$$RR/H_0:\ Z_0 > Z_{\alpha/2}$$



5. Cálculo de la prueba Estadística.

$$\hat{p}_P = \frac{64}{98} = 0.65306 \quad \Rightarrow \hat{q}_P = 1 - \frac{64}{98} = 0.34694$$

$$\hat{p}_C = \frac{60}{118} = 0.50847 \quad \Rightarrow \hat{q}_C = 1 - \frac{60}{118} = 0.49153$$

$$Z_0 = \frac{\left(\hat{p}_P - \hat{p}_C\right) - p_0}{\sqrt{\frac{\hat{p}_P \hat{q}_P}{n_P} + \frac{\hat{p}_C \hat{q}_C}{n_C}}} = \frac{(0.65306 - 0.50847) - 0.15}{\sqrt{0.0023119 + 0.002118}} = \frac{-0.00541}{0.066558} = -0.08128$$

6. Decisión:

No hay evidencia para rechazar la H_0, porque Z_C = -0.08Z_t=-1.65, entonces se acepta la H_0. y se rechaza la H_a.

7. Conclusión.

La proporción de cerdos con cisticercosis no es 15% más en Puno que en el Cusco, por lo que la suposición del alumno no es verdadera, siendo la proporción iguales en ambas regiones.

TEMA:

ANÁLISIS DE REGRESIÓN

Vladimiro Ibañez Quispe, Dr.

REGRESIÓN LINEAL SIMPLE

La regresión se utiliza para analizar datos que provienen de experimentos que no fueron diseñados, como estudio de fenómenos no controlados o de registros históricos.

La regresión se usa para determinar la «*mejor*» relación funcional entre las variables en estudio.

REGRESIÓN LINEAL SIMPLE

En pocas palabras, regresión es la *cantidad de cambio* de una variable asociado a un cambio único de otra variable.

La RLS permite determinar si existe relación entre las variables (X e Y), el cual utiliza el comportamiento de una variable (X = independiente), para predecir el comportamiento de otra variable (Y = dependiente). Las dos características por lo general deben ser cuantitativas.

USOS DE RLS:

- Para hacer predicciones futuras de *Y*, teniendo como base la *X*.
- Para ver si la variable (Y) depende de otra (X), estimando por consiguiente la medida de dicha relación o asociación.
- Para determinar la forma de la curva de la regresión.
- Para conocer el error real implicado en un experimento, después que haya sido descontado el efecto de una variable relacionada.
- Sirve de base para el análisis estadístico.

EJEMPLOS:

- ✓ Predecir el tiempo meteorológico basado en los datos del pasado.
- ✓ Predecir la producción de lana/año, basada en la información recogida en años anteriores.
- ✓ Etc.

MODELO DE REGRESIÓN LINEAL SIMPLE (MRLS)

Variable dependiente

Variable independiente

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \qquad i = 1,2,\ldots,n$$

Error de perturbación

Parámetro que indica la ordenada de la línea de regresión

Pendiente de la recta de regresión

PROPIEDADES DEL MODELO RLS.

a. **Toda perturbación aleatoria tiene de media cero:**

$$E(\varepsilon_i) = 0$$

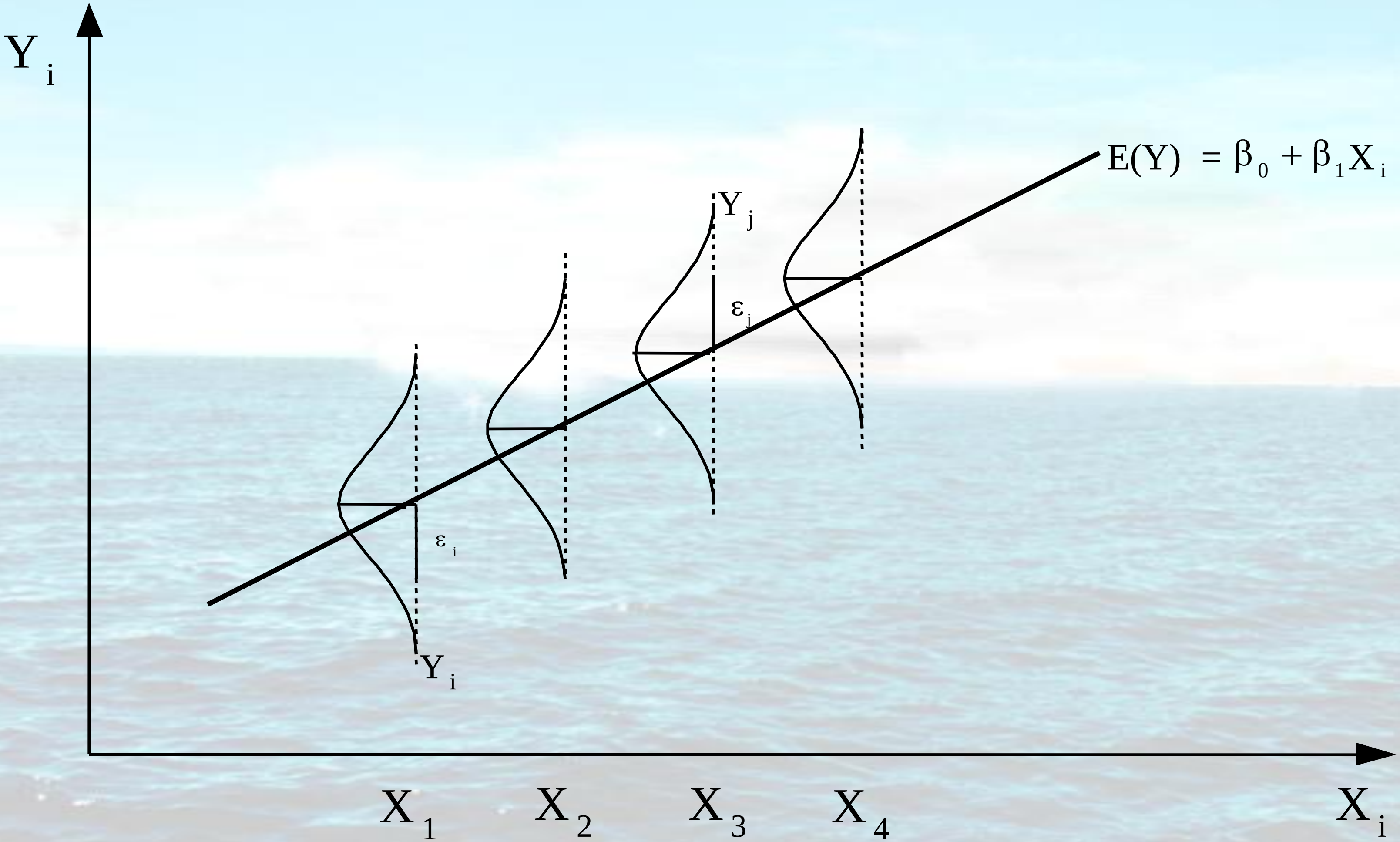
b. **Todas las perturbaciones aleatorias tienen la misma varianza.**

$$V(\varepsilon_i) = \sigma_i^2$$

c. **Las perturbaciones son independientes entre sí:**

$$\operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j$$

Para el EJEMPLO b)

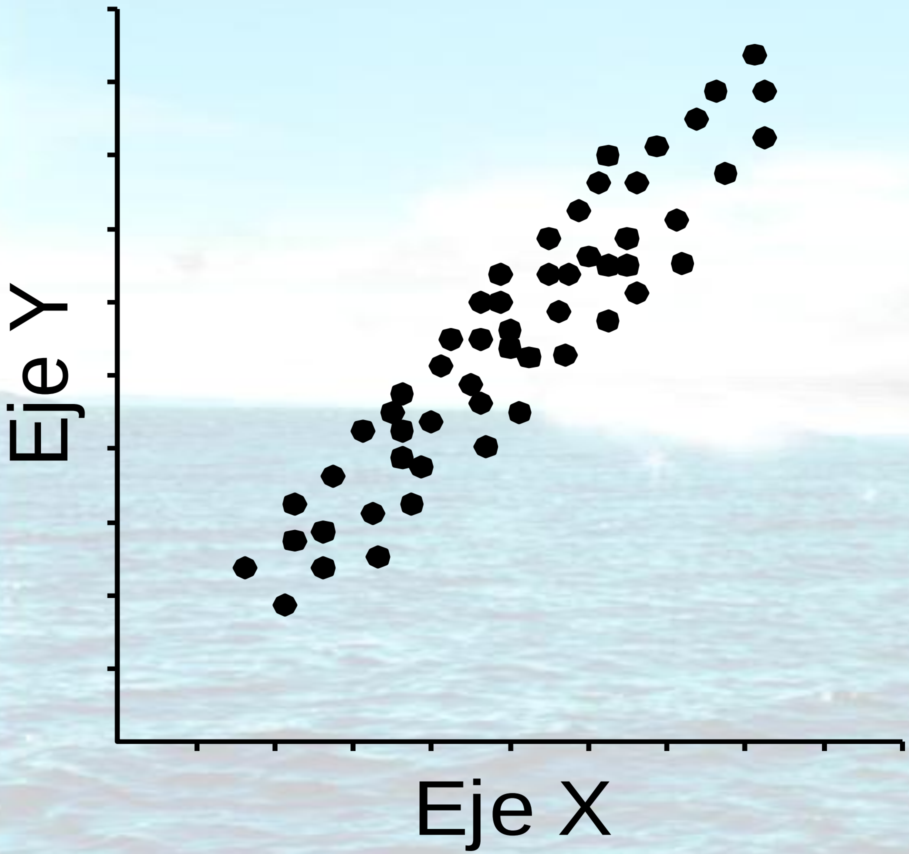


Suposiciones en regresión

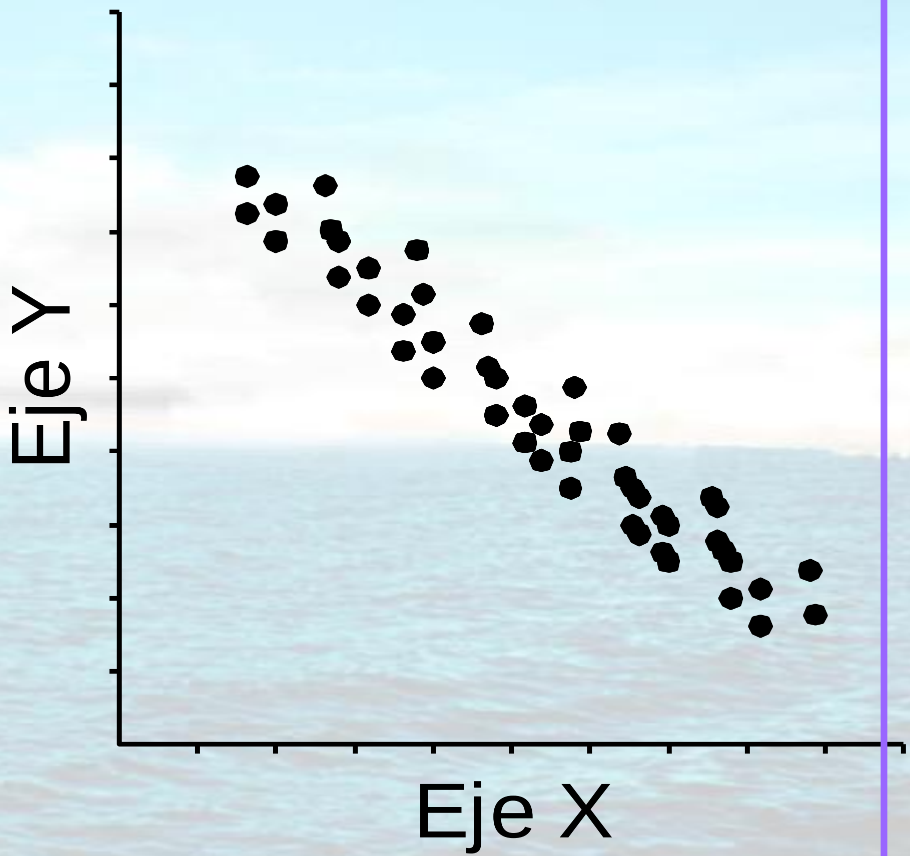
DIAGRAMA DE DISPERSIÓN:

La representación da origen a una nube de puntos que se denomina *diagrama de dispersión ó esparcimiento*; es la forma más usual para detectar si la función es lineal, exponencial, potencial, cuadrática, etc. al cual el experimentador ajustará su información recopilada, de tal forma que describa adecuadamente la relación entre las variables en estudio.

EJEMPLOS DE DIAGRAMAS DE DISPERSIÓN

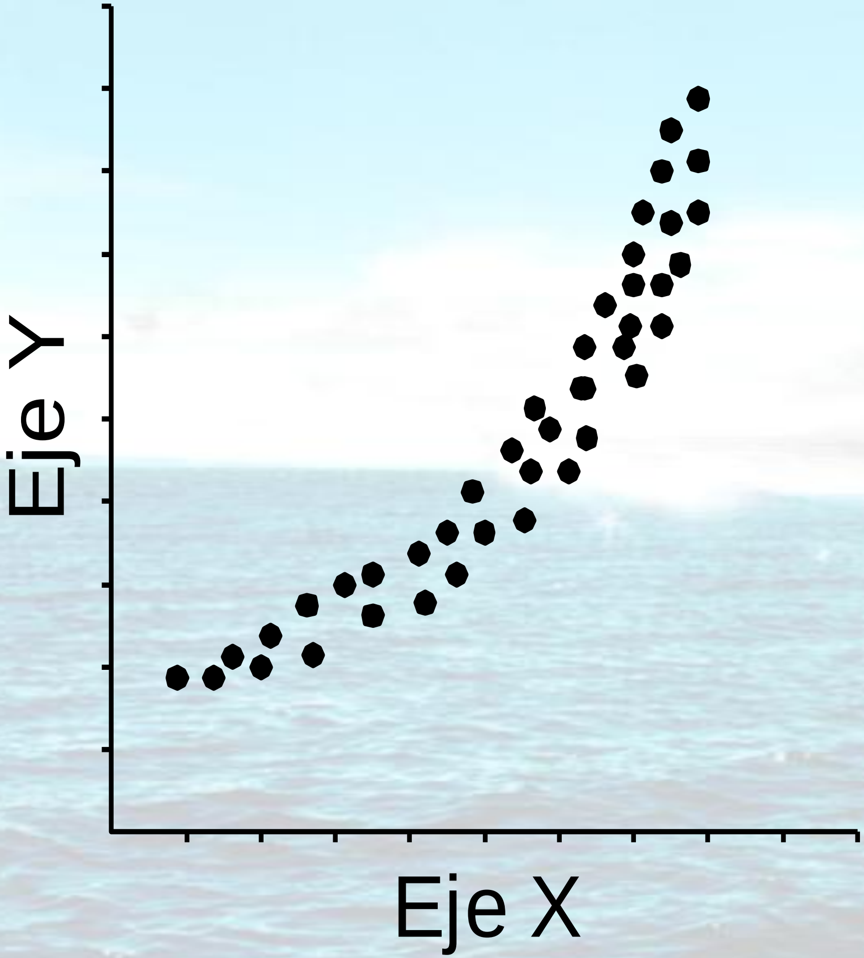


Lineal positiva

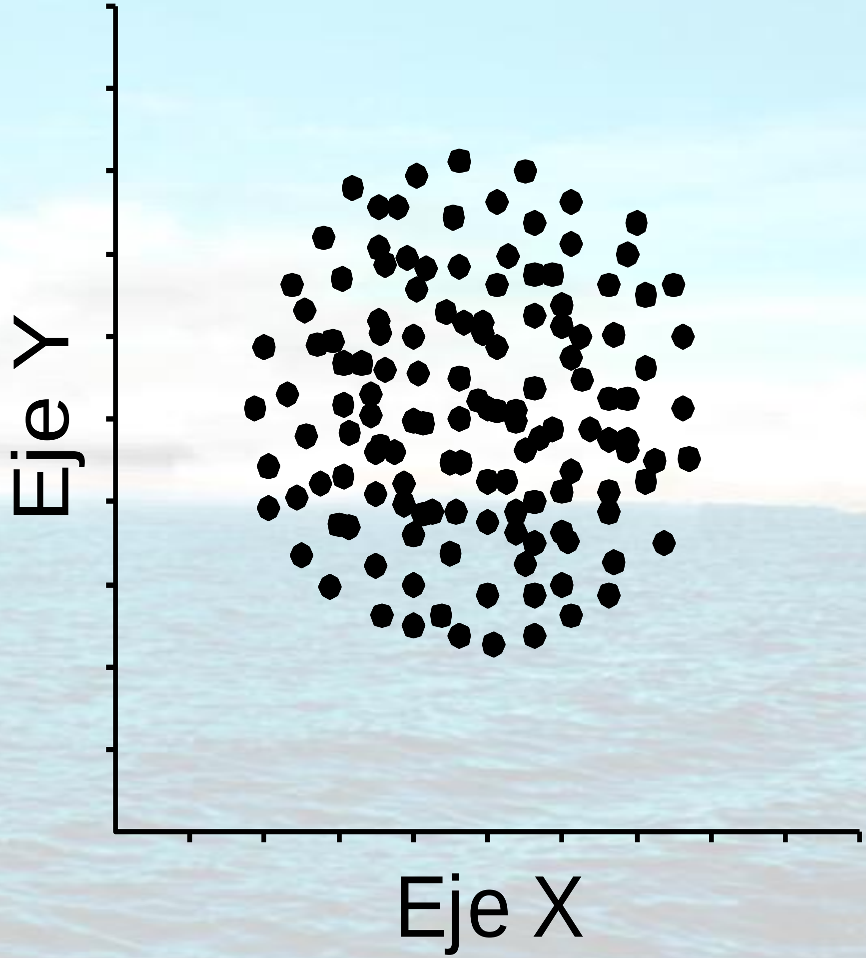


Lineal negativa

EJEMPLOS DE DIAGRAMAS DE DISPERSIÓN



No Lineal



Ninguna relación

ESTIMACIÓN DE LA ECUACIÓN DE REGRESIÓN.

Para lograr la estimación de los parámetros desconocidos (β_0 y β_1), se utiliza uno de los métodos más conocido el «METODO DE MÍNIMOS CUADRADOS». Este método consiste en encontrar los valores β_0 y β_1 de la ecuación de regresión muestral.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i , \qquad i = 1,2,\ldots,n$$

$$\varepsilon_i = Y_i - \beta_0 - \beta_1 X_i$$

$$a)\ \frac{\partial \sum_{i=1}^{n} \varepsilon_i^2}{\partial \beta_0} = 0 \qquad y \qquad b)\ \frac{\partial \sum_{i=1}^{n} \varepsilon_i^2}{\partial \beta_1} = 0$$

ECUACIONES NORMALES (EN).

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i Y_i$$

PARÁMETROS ESTIMADOS:

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n} X_i^2 - \overline{X}\,^2} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\,\overline{Y}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}$$

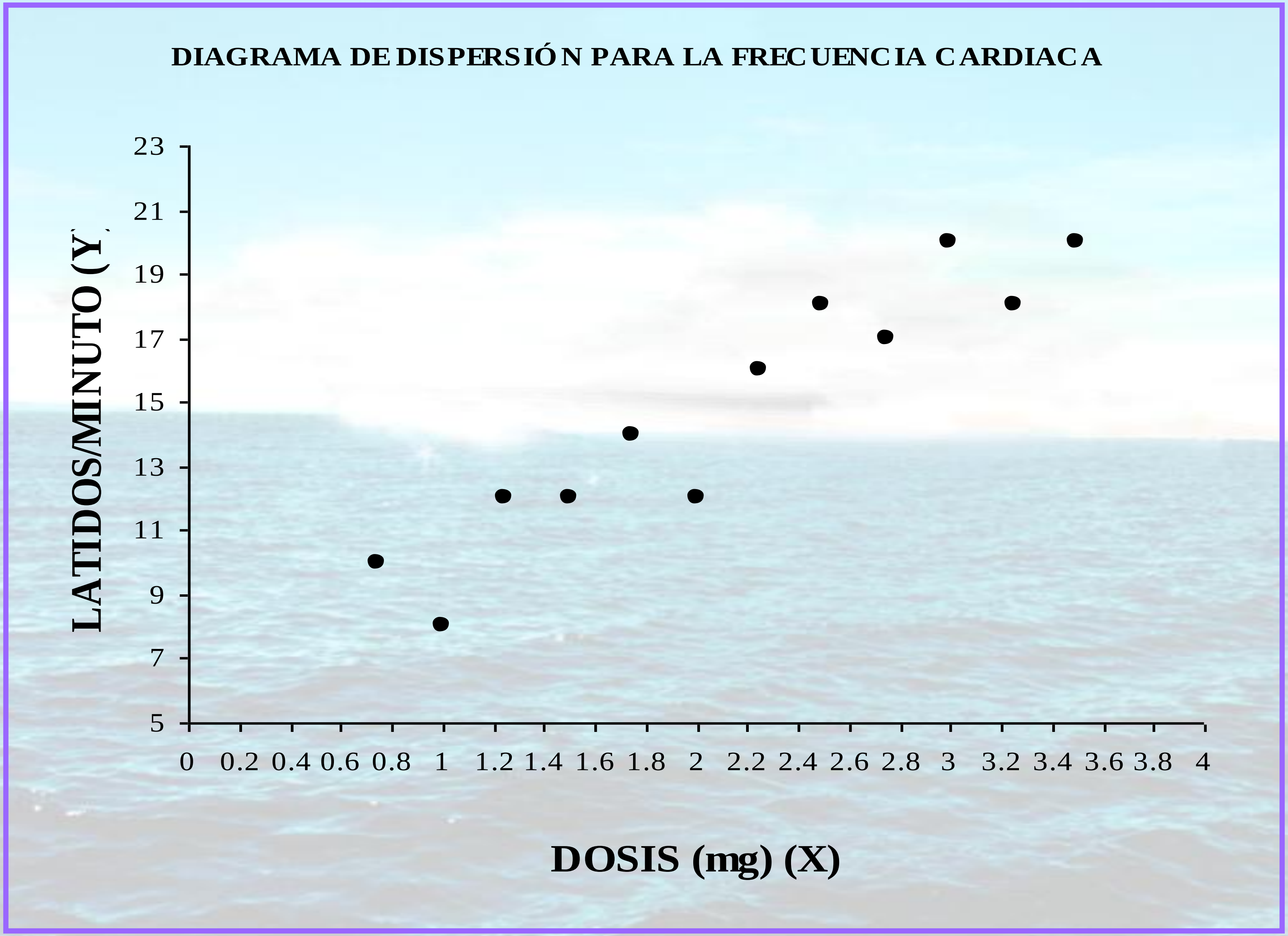
$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

EJEMPLO:

Se llevó a cabo un experimento para estudiar el efecto de cierto medicamento para disminuir la frecuencia cardiaca en adultos. La variable independiente es la dosis en miligramos del medicamento, y la variable dependiente es la diferencia entre la frecuencia cardiaca más baja después de la administración del medicamento y un control antes de administrarlo. Se reunieron los siguientes datos:

DOSIS (mg) (X)	REDUC. DEL RITMO CARD (latidos/min) (Y)
0.50	10
0.75	8
1.00	12
1.25	12
1.50	14
1.75	12
2.00	16
2.25	18
2.50	17
2.75	20
3.00	18
3.25	20
3.50	21

DIAGRAMA DE DISPERSIÓN



Los resultados son los siguientes:

$$\sum X_i = 26, \sum Y_i = 198, \sum X_i Y_i = 442.5, \sum X_i^2 = 63.375, \sum Y_i^2 = 3226$$

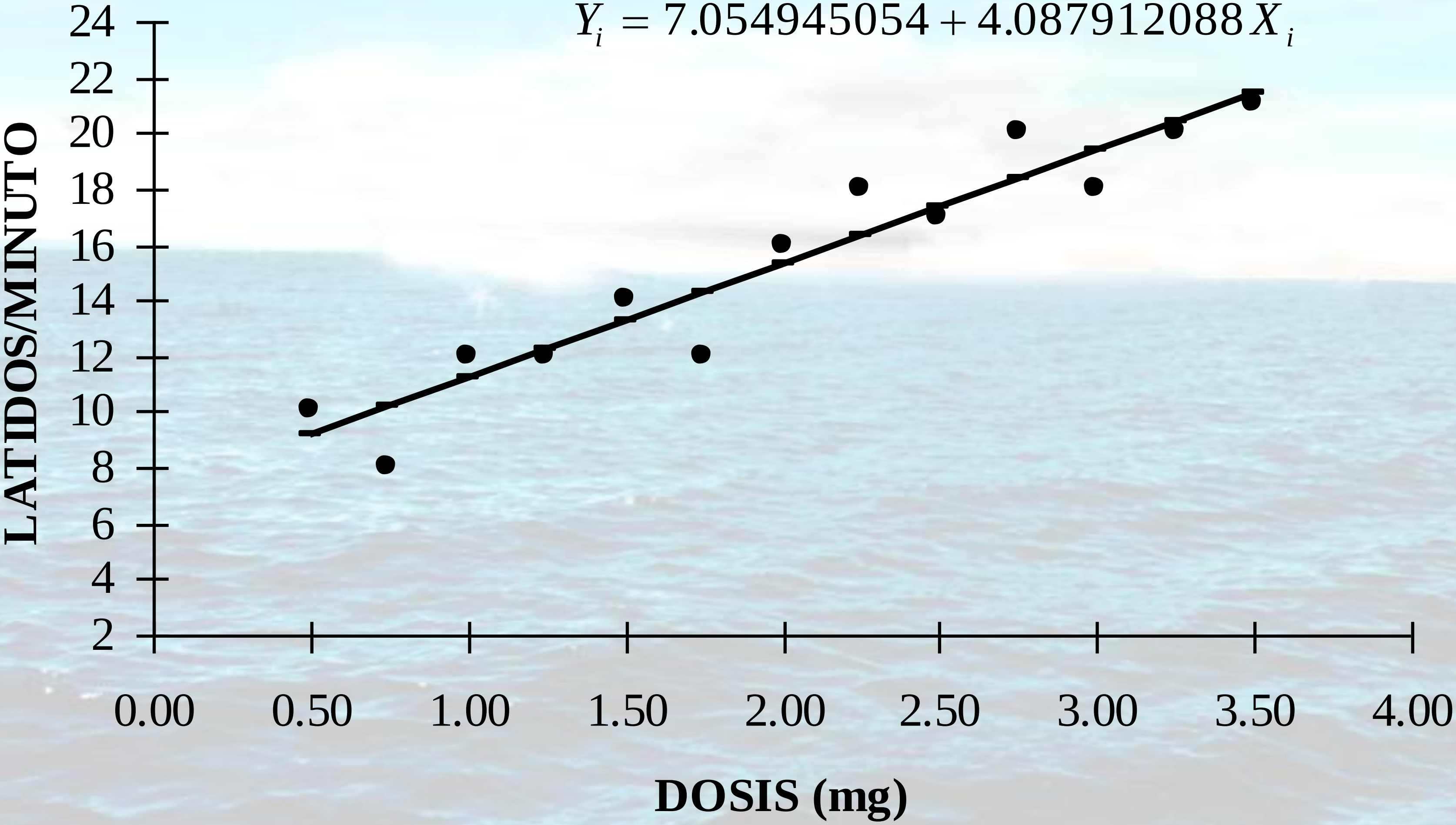
$$\hat{\beta}_1 = \frac{\sum_{i=1}^{13} X_i Y_i - n\overline{X}\overline{Y}}{\sum_{i=1}^{13} X_i^2 - n\overline{X}^2} = \frac{422.5 - 13(2)(15.23076923)}{63.375 - 13(4)} = 4.087912088$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} = 15.23 - 4.087912088x2 = 7.054945054$$

$$\hat{Y}_i = 7.054945054 + 4.087912088 X_i$$

Recta de regresión lineal ajustada para la Frecuencia Cardíaca

$$\hat{Y}_i = 7.054945054 + 4.087912088\, X_i$$



ANÁLISIS DE VARIANZA (ANOVA)

Tabla de Análisis de Varianza para la Frecuencia Cardíaca

F. de V.	G.L.	S.S.	M.S.	Fc.	Signif
Regresión	1	190.087912	190.08791	103.41	**
Error	11	20.219780	1.83816		
Total	12	210.307692	17.52564100		

B. CORRELACIÓN (*r*)

Coeficiente de correlación.

La correlación lineal permite medir si hay asociación entre dos variables ó mide la intensidad de dicha asociación. Para ello es importante que la muestra bivariado sea tomada al azar, tanto para la variable X como para la variable Y. La "*r*" de Pearson, se ajusta al tipo de información cuantitativa. Los valores del coeficiente de correlación oscilan entre +1 y –1.

Cálculo del coeficiente de Correlación (*r*).

$$r_{xy} = \frac{\sum X_i Y_i - (\sum X_i)(\sum Y_i) \big/ n}{\sqrt{\sum X_i^2 - (\sum X_i)^2 \big/ n \quad \sum Y_i^2 - (\sum Y_i)^2 \big/ n}}$$

Los rangos que del coeficiente de correlación, se pueden interpretarse haciendo uso de las siguientes expresiones:

***r* = 0.2 a *r* = 0.3**	**(coeficiente de correlación muy bajo).**
***r* = 0.4 a *r* = 0.5**	**(coeficiente de correlación bajo).**
***r* = 0.6 a *r* = 0.7**	**(coeficiente de correlación alto).**
***r* = 0.8 a *r* = 1.0**	**(coeficiente de correlación muy alto).**

El coeficiente de correlación:

$$r_{xy} = \frac{\sum X_i Y_i - (\sum X_i)(\sum Y_i)/n}{\sqrt{\sum X_i^2 - (\sum X_i)^2/n \quad \sum Y_i^2 - (\sum Y_i)^2/n}}$$

$$= \frac{442.5 - (26)(198)/13}{\sqrt{(63.375 - (26)^2/13)(3226 - (198)^2/13)}} = 0.9507$$

C. COEFICIENTE DE DETERMINACIÓN

El coeficiente de determinación, R^2, es el valor cuadrado del coeficiente *r* de Pearson, y representa la proporción de la varianza explicada por una variable respecto a la varianza total. Se considera como medida del grado de influencia de una variable frente a otra variable.

$$R^2 = r^2 = \frac{\text{Varianza Explicada}}{\text{Varianza Total}} x100 = \frac{SCR}{SCT} x100$$

COEFICIENTE DE DETERMINACIÓN AJUSTADO $\tilde{R}^2$

$$\tilde{R}^2 = 1 - \frac{CME}{CMT} \quad o \quad \tilde{R}^2 = 1 - \frac{MSE}{MST_m}$$

EJEMPLO:

$$R^2 = \frac{SSR_m}{SST_m} = \frac{190.08791}{210.30769} x100 = 90.39\%$$

$$\tilde{R}^2 = 1 - \frac{MSE}{MST_m} = 1 - \frac{1.83816}{17.525641} = 89.51\%$$

D. ANÁLISIS DE REGRESIÓN LINEAL MÚLTIPLE.

Muchos problemas de regresión involucran más de una variable regresiva. La regresión lineal múltiple es una de las técnicas estadísticas más ampliamente utilizadas en la actualidad. En esta oportunidad se presenta las técnicas básicas de la estimación de parámetros mediante el método de Mínimos Cuadrados.

El modelo es el siguiente:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i ; \quad ó \quad Y = X\,\beta + \varepsilon$$

Estimación de parámetros:

Se usa método de Mínimos Cuadrados para estimar los coeficientes de regresión. Supóngase que se dispone n>k observaciones, y X_{ij} denota la observación iésima o el nivel de la variable X_j.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i$$

$$= \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} + \varepsilon_i \qquad i = 1,2,\ldots,N$$

Usando las derivadas parciales para estimar a cada uno de los parámetros:

$$\left.\frac{\partial\left(\sum \varepsilon_i^2\right)}{\partial \beta_j}\right|_{\hat{\beta}_0,\hat{\beta}_1,\ldots,\hat{\beta}_k} = 2\sum_{i=1}^{N}\left(Y_i - \beta_0 - \sum_{j=1}^{k}\beta_j X_{ij}\right)\left(-X_{ij}\right) = 0,$$

$$j = 1,2,\ldots,k$$

Ecuaciones Normales:

$$\begin{array}{ccccccccccc}
N\hat{\beta}_0 & + & \hat{\beta}_1\sum_{i=1}^{N} X_{i1} & + & \hat{\beta}_2\sum_{i=1}^{N} X_{i2} & + & \cdots & + & \hat{\beta}_k\sum_{i=1}^{N} X_{ik} & = & \sum_{i=1}^{N} Y_i \\
\hat{\beta}_0\sum_{i=1}^{N} X_{i1} & + & \hat{\beta}_1\sum_{i=1}^{N} X_{i1}^2 & + & \hat{\beta}_2\sum_{i=1}^{N} X_{i1}X_{i2} & + & \cdots & + & \hat{\beta}_k\sum_{i=1}^{N} X_{i1}X_{ik} & = & \sum_{i=1}^{N} X_{i1}Y_i \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\hat{\beta}_0\sum_{i=1}^{N} X_{ik} & + & \hat{\beta}_1\sum_{i=1}^{N} X_{ik}X_{i1} & + & \hat{\beta}_2\sum_{i=1}^{N} X_{ik}X_{i2} & + & \cdots & + & \hat{\beta}_k\sum_{i=1}^{N} X_{ik}^2 & = & \sum_{i=1}^{N} X_{ik}Y_i
\end{array}$$

FORMA MATRICIAL: La forma matricial de las ecuaciones normales. El modelo en términos de las observaciones, se puede escribir en notación matricial como: $Y = X\beta + \varepsilon$

$$Y = \begin{pmatrix} Y_1 \\ Y_1 \\ \vdots \\ Y_N \end{pmatrix}_{Nx1} \qquad X = \begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{N1} & X_{N2} & \cdots & X_{Nk} \end{pmatrix}_{N(k+1)}$$

Vector de parámetros y errores residuales:

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}_{(k+1)x1} \quad y \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}_{Nx1}$$

PROPIEDADES.

a) $E(\varepsilon)=0, \quad E(Y)=X\beta, \quad Y=E(Y)+\varepsilon, \quad \Rightarrow \quad \varepsilon = Y-E(Y)$

b) $Var(\varepsilon)=E\;\;\varepsilon-E(\varepsilon)\;\;\varepsilon-E(\varepsilon)\,'=E(\varepsilon\varepsilon')=\sigma^2 I_N$

Realizando las derivadas parciales (Forma matricial), se encuentra la siguiente expresión:

$$(X'X)^{-1}(X'X)\hat{\beta} = (X'X)^{-1}X'Y, \;\Rightarrow\; \hat{\beta} = (X'X)^{-1}X'Y$$

Ecuaciones normales de forma matricial:

$$\begin{pmatrix} N & \sum_{i=1}^{N} X_{i1} & \sum_{i=1}^{N} X_{i2} & \cdots & \sum_{i=1}^{N} X_{ik} \\ \sum_{i=1}^{N} X_{i1} & \sum_{i=1}^{N} X_{i1}^2 & \sum_{i=1}^{N} X_{i1}X_{i2} & \cdots & \sum_{i=1}^{N} X_{i1}X_{ik} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{N} X_{ik} & \sum_{i=1}^{N} X_{ik}X_{i1} & \sum_{i=1}^{N} X_{ik}X_{i2} & \cdots & \sum_{i=1}^{N} X_{ik}^2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{N} Y_i \\ \sum_{i=1}^{N} X_{i1}Y_i \\ \vdots \\ \sum_{i=1}^{N} X_{ik}Y_i \end{pmatrix}$$

El modelo de regresión ajustado es: $\hat{Y} = X\hat{\beta}$ en notación escalar, el modelo ajustado es:

$$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j x_{ij}, \quad i = 1,2,..,N$$

EJEMPLO:

En un estudio que corresponde a los resultados encontrados por un grupo de 10 alumnos tomados al azar del archivo académico las siguientes variables: Y 0 Calificación (en puntos) en el examen de matemáticas, X1 = Calificación en el examen psicológico, y X2 = Calificación en el examen de cultura general. La información se presenta a continuación:

La información de 10 alumnos.

Examen de matemáticas (Y)	Examen psicológico (X1)	Examen de cultura general (X2)
77	76	84
83	71	85
64	96	95
59	96	95
53	55	73
77	73	89
37	16	65
15	15	47
28	51	71
35	51	59

$$X'X = \begin{bmatrix} n & \sum X_{i1} & \sum X_{i2} \\ \sum X_{i1} & \sum X_{i1}^2 & \sum X_{i1}X_{i2} \\ \sum X_{i2} & \sum X_{i1}X_{i2} & \sum X_{i2}^2 \end{bmatrix} = \begin{bmatrix} 10 & 600 & 763 \\ 16.8 & 43286 & 49546 \\ 763 & 49546 & 60537 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} \sum Y_i \\ \sum X_{i1}Y_i \\ \sum X_{i2}Y_i \end{bmatrix} = \begin{bmatrix} 528 \\ 36119 \\ 43093 \end{bmatrix}$$

$$\det(X'X) = \begin{vmatrix} 10 & 600 & 763 \\ 600 & 43286 & 49546 \\ 763 & 49546 & 60537 \\ 10 & 600 & 763 \\ 600 & 43286 & 49546 \end{vmatrix} = 27214926 = D$$

$$(X'X)^{-1} = \frac{Adj(X'X)}{\det(X'X)} = \begin{bmatrix} 6.084839841 & 0.054433291 & -0.12124295 \\ 0.054433291 & 0.00085251 & -0.00138399 \\ -0.121242953 & -0.001383799 & 0.00267721 \end{bmatrix}$$

Encontrando la inversa de la matriz y multiplicando X'Y, se tiene las betas estimadas:

$$\hat{\beta} = \begin{bmatrix} 6.084839841 & 0.054433291 & -0.12124295 \\ 0.054433291 & 0.00085251 & -0.00138399 \\ -0.121242953 & -0.001383799 & 0.00267721 \end{bmatrix} \times \begin{bmatrix} 528 \\ 36119 \\ 13093 \end{bmatrix} = \begin{bmatrix} -45.85109 \\ -0.099479 \\ 1.3711643 \end{bmatrix}$$

El modelo de regresión múltiple estimado es:

$$\hat{Y} = -45.85109 - 0.09948 X_{i1} + 1.37116 X_{i2}$$

TABLA ANOVA: FORMA MATRICIAL.

Hipótesis a Probarse: $H_0: \beta = 0 \quad \text{vs.} \quad H_a: \beta \neq 0$

F. de V.	G.L.	S.S.	M.S.	Fobs.
Regresión	**r - 1**	$SSR_m = \hat{b}'X'Y - N\bar{Y}^2$	$\frac{SSR_m}{r-1} = MSR_m$	$\frac{MSR_m}{MSE}$
Error Residual	**N - r**	$SSE = Y'Y - \hat{b}'X'Y$	$\frac{SSE}{r-N} = MSE$	
Total	**N - 1**	$SST_m = Y'Y - N\bar{Y}^2$		

ANOVA para el ejemplo anterior:

Tabla de Análisis de Varianza para las calificaciones de estudiantes

F. de V.	G.L.	S.S.	M.S.	Fc.	Signif
Debido a Regresión	2	3406.722	1703.3610	8.22	*
Error residual	7	1450.878	207.2683		
Total	9	4857.600	539.7333		

Coeficiente de Determinación: R^2

$$R^2 = \frac{SSR_m}{SST_m} \times 100 = \frac{3406.722}{4857.6} \times 100 = 70.13\%$$

Coeficiente de Determinación ajustado:

$$\tilde{R}^2 = 1 - \frac{MSE}{MST_m} = (1 - \frac{207.268}{539.733}) \times 100 = 61.60\%$$

Matriz de Varianzas-Covarianzas:

$$V(\hat{\beta}) = (X'X)^{-1}\sigma^2$$

$$V(\hat{\beta}) = \begin{bmatrix} 6.084839841 & 0.054433291 & -0.12124295 \\ 0.054433291 & 0.00085251 & -0.00138399 \\ -0.121242953 & -0.001383799 & 0.00267721 \end{bmatrix} \times (207.268) = \begin{bmatrix} 1261.1943227 & 11.282295014 & -25.12981907 \\ 11.282295014 & 0.1766983124 & -0.286817743 \\ -25.12981907 & -0.286817743 & 0.5549001785 \end{bmatrix}$$

Vector de errores estimados:

$$\hat{e} = \begin{bmatrix} 77 \\ 83 \\ 64 \\ 59 \\ \vdots \\ 28 \\ 35 \end{bmatrix} - \begin{bmatrix} 1 & 76 & 84 \\ 1 & 71 & 85 \\ 1 & 96 & 95 \\ 1 & 96 & 95 \\ \vdots & \vdots & \vdots \\ 1 & 51 & 71 \\ 1 & 51 & 59 \end{bmatrix} \times \begin{bmatrix} -45.85109157 \\ -0.099479124 \\ 1.371164338 \end{bmatrix} = \begin{bmatrix} 15.234 \\ 19.365 \\ -10.860 \\ \vdots \\ \vdots \\ -18.428 \\ 5.026 \end{bmatrix}$$

Estimar los Intervalos Confidenciales (IC) para betas:

$$Pr\left[\hat{\beta}_i - t_{\frac{\alpha}{2},(n-r)}\sqrt{V(\hat{\beta}_i)} \le \beta_i \le \hat{\beta}_i + t_{\frac{\alpha}{2},(n-r)}\sqrt{V(\hat{\beta}_i)}\right] = 1-\alpha$$

Estimar el Intervalo Confidencial (IC) para β_0

$$\Pr\ -45.85109-(2.365)\sqrt{1261.194323} \le \beta_0 \le -45.85109+(2.365)\sqrt{1261.194323}\ = 0.95$$

$$\Pr\ -129.8267 \le \beta_0 \le 38.12451\ = 0.95$$

Estimar el Intervalo Confidencial (IC) para β_1

$$\Pr\ -0.099479124-(2.365)\sqrt{0.176698} \le \beta_1 \le -0.099479124+(2.365)\sqrt{0.176698}\ = 0.95$$

$$\Pr\ -1.09346 \le \beta_1 \le 0.89466\ = 0.95$$

Estimar el Intervalo Confidencial (IC) para β_2

$$\Pr\left[1.3711643-(2.365)\sqrt{0.55490}\le\beta_2\le1.3711643+(2.365)\sqrt{0.55490}\right]=0.95$$

$$\Pr\left[-0.39028\le\beta_2\le3.13261\right]=0.95$$

Estimar la respuesta Media $E(Y_h)$ y el IC:

$$X_h' = \begin{bmatrix}1 & 55 & 73\end{bmatrix}$$

$$\hat{Y}_h = X_h'\hat{\beta} = \begin{bmatrix}1 & 55 & 73\end{bmatrix}\times\begin{bmatrix}-45.85109\\ -0.099479\\ 1.3711643\end{bmatrix} = 48.77255$$

$$V(X_h'\hat{\beta}) = X_h'(X'X)^{-1}\sigma^2X_h = X_h'V(\hat{\beta})X_h \text{ ó } S^2(\hat{Y}_h) = \mathrm{MSE}[X_h'(X'X)^{-1}X_h]$$

$$V(X_h'\hat{\beta}) = \begin{bmatrix}1 & 55 & 73\end{bmatrix}\times\begin{bmatrix}1261.194323 & 11.2823 & -25.129819\\ 11.28229501 & 0.176698 & -0.2868177\\ -25.12981907 & -0.28682 & 0.55490018\end{bmatrix}\times\begin{bmatrix}1\\ 55\\ 73\end{bmatrix}$$

$$S(\hat{Y}_h) = 4.660704219$$

El Intervalo Confidencial (IC) para la respuesta media:

$$\Pr\left[48.77255 - (2.365)(4.660704) \le E(Y_h) \le 48.77255 + (2.365)(4.660704)\right] = 0.95$$

$$\Pr\left[37.7499878 \le E(Y_h) \le 59.79511878\right] = 0.95$$

Predicción para las observaciones futuras: $Y_{h(new)}$

$$X_h^{'} = \begin{bmatrix} 1 & 100 & 100 \end{bmatrix}$$

$$\hat{Y}_h = X_h^{'}\hat{\beta} = \begin{bmatrix} 1 & 100 & 100 \end{bmatrix} \times \begin{bmatrix} -45.85109 \\ -0.099479 \\ 1.3711643 \end{bmatrix} = 81.31743$$

$$S^2(\hat{Y}_{h(New)}) = MSE[1 + \mathrm{X}_\mathrm{h}^{'}(\mathrm{X'X})^{-1}\mathrm{X}_\mathrm{h}] = \mathrm{MSE} + \mathrm{X}_\mathrm{h}^{'}(\mathrm{X'X})^{-1} MSE\, \mathrm{X}_\mathrm{h} = MSE + \mathrm{X}_\mathrm{h}^{'}V(\hat{\beta})\mathrm{X}_\mathrm{h}$$

$$S^2(\hat{Y}_{h(New)}) = 207.268 + [1 \quad 100 \quad 100 \times \begin{bmatrix} 1261.19432 & 11.282295 & -25.129819 \\ 11.282295 & 0.1766983 & -0.2868177 \\ -25.129819 & -0.2868177 & 0.554900178 \end{bmatrix} \times \begin{bmatrix} 1 \\ 100 \\ 100 \end{bmatrix}$$

= 278.5878555

$$S(\hat{Y}_{h(New)}) = 16.6909513$$

El Intervalo Confidencial (IC) para la obs. Futuras es:

$$Pr\left[\hat{Y}_h - t_{\frac{\alpha}{2},(n-r)} S(\hat{Y}_{h(new)}) \le Y_{h(new)} \le \hat{Y}_h + t_{\frac{\alpha}{2},(n-r)} S(\hat{Y}_{h(new)})\right] = 1-\alpha$$

$$Pr\left[81.31743 - (2.365)(16.6909513) \le Y_{h(New)} \le 81.31743 + (2.365)(16.6909513)\right] = 0.95$$

$$Pr\left[41.84333 \le Y_{h(New)} \le 120.7915297\right] = 0.95$$

TEMA:

DISEÑOS EXPERIMENTALES

Vladimiro Ibañez Quispe, Dr.

1. DISEÑO COMPLETAMENTE AL AZAR (DCA, DIA)

Es aquel diseño en el cual los tratamientos son asignados en forma aleatoria a las unidades experimentales, o viceversa, sin ninguna restricción; por lo tanto se considera que es un diseño eficiente cuando las unidades experimentales de los que se dispone son muy homogéneas. Sus características son:

Características:

a) Los tratamientos se distribuyen en forma aleatoria en todas las unidades experimentales, y el número de repeticiones o unidades por tratamiento puede ser igual o diferente.

b) El diseño es útil cuando las unidades experimentales tienen una variabilidad uniformemente repartida.

c) El DIA, proporciona el máximo número de grados de libertad para la estimación del error experimental; además, no requiere estimar datos faltantes, es decir el diseño puede analizarse con diferente número de repeticiones por tratamiento.

MODELO ESTADISTICO LINEAL.

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \qquad i = 1,2,\ldots,t$$

$$j = 1,2,\ldots,r$$

Y_{ij} = Es una observación en la j-ésima unidad experimental, sujeto al i-ésimo tratamiento.

μ = Es el efecto de la media general o constante común.

τ_i = Es el efecto del i-ésimo tratamiento.

ε_{ij} = Efecto verdadero de la j-ésima unidad experimental, sujeta al i-ésimo tratamiento (error experimental).

ESQUEMA DEL DISEÑO

Repeticiones (j)	Tratamientos (i)					Total
	1	2	3	....	t	
1	Y_{11}	Y_{21}	Y_{31}	....	Y_{t1}	$Y_{.1}$
2	Y_{12}	Y_{22}	Y_{32}	....	Y_{t2}	$Y_{.2}$
3	Y_{13}	Y_{23}	Y_{33}	....	Y_{t3}	$Y_{.3}$
.	....	....	....	....	....	....
.	....	....	....	....	....	....
.	....	....	....	....	....	....
r	Y_{1r}	Y_{2r}	Y_{3r}	....	Y_{tr}	$Y_{.r}$
Total (t)	$Y_{1.}$	$Y_{2.}$	$Y_{3.}$	....	$Y_{t.}$	$Y_{..}$
Total (r)	$n_{1.}$	$n_{2.}$	$n_{3.}$	....	$n_{t.}$	$n_{..}$

HIPÓTESIS A PROBAR:

$H_0: \tau_i = 0$ **(Todo los tratamientos son iguales)**

$H_a: \tau_i \neq 0$ **(al menos uno de los tratamientos es diferente de los otros)**

TABLA ANVA: Fórmulas

F. de V.	G.L.	S.C.	C.M.	Fc	C.M.E.
Tratam.	t - 1	$\sum_{i=1}^{t}\frac{Y_{i.}^2}{r_i} - \frac{Y_{..}^2}{N}$	$\frac{SC_{trat.}}{t-1}$	$\frac{CM_{trat.}}{CM_{Error}}$	$\sigma^2 + r\sum_{i=1}^{t}\frac{\tau_i^2}{t-1}$
Error Exptal	N – t	$\sum_{i=1}^{t}\sum_{j=1}^{r_i} Y_{ij}^2 - \sum_{i=1}^{t}\frac{Y_{i.}^2}{r_i}$	$\frac{SC_{error}}{N-t}$		σ^2
TOTAL	N - 1	$\sum_{i=1}^{t}\sum_{j=1}^{r_i} Y_{ij}^2 - \frac{Y_{..}^2}{N}$			

EJEMPLO:

Se sometió a cuatro grupos de estudiantes a diferentes técnicas de enseñanza y se les examinó al final de un periodo específico. Debido a las deserciones de los grupos experimentales (por enfermedad, transferencias, etc.) el número de estudiantes varió de grupo a grupo. ¿Presentan los datos mostrados en la siguiente tabla la evidencia suficiente para indicar una diferencia en el rendimiento medio para las cuatro técnicas de enseñanza?.

Grupos	**1**	65	87	73	79	81	69		**454**
	2	75	69	83	81	72	79	90	**549**
	3	59	78	67	62	83	76		**425**
	4	94	89	80	88				**351**

¿Sugieren estos datos una diferencia en el rendimiento promedio de las 4 técnicas de enseñanza?. Sea alfa = 0.05. Realizar el ANOVA e interprete los resultados.

PROCEDIMIENTO:

a) Grados de Libertad:

GLtrat. = t -1 = 4 - 1 = 3

GLerror = N - t = 23 - 4 = 19

GLtotal = N - 1 = 23 - 1=22

b) Suma de Cuadrados (SS=S.C.) $TC = \frac{Y_{..}^2}{N} = \frac{1779^2}{23} = 137601.7826$

$$SC_{Trat.} = \sum_{i=1}^{4} \frac{Y_{i.}^2}{r_i} - TC = \frac{454^2}{6} + \frac{549^2}{7} + \frac{425^2}{6} + \frac{351^2}{4} - TC = 138314.369 - TC = 712.58644$$

$$SC_{Total} = \sum_{i=1}^{4} \sum_{j=1}^{r_i} Y_{ij}^2 - TC = 65^2 + 87^2 + \cdots + 76^2 - TC = 139511 - TC = 1909.2174$$

$SC_{Error\ Exptal}$ = SC_{total} - SC_{trat} = 1909.2174-712.586= 1196.63096

Tabla de ANOVA:

F. de V.	G.L.	S.C.	C.M.	Fc	Pr > F
Entre grupos	3	712.58644	237.52881	3.77	0.028
Dentro de grupos	19	1196.63095	62.98058		
Total	22	1909.21739			

$$CV = \frac{\sqrt{CM_{Error}}}{\bar{Y}..}(100) = \frac{\sqrt{62.980576}}{77.3478}(100) = 10.26\%$$

La prueba F de la tabla de ANOVA nos muestra que la Fc = 3.77>$F_{3,19,0.05}$=3.13, entonces se acepta la hipótesis alterna (Ha:), es decir existe diferencia estadística significativa en el rendimiento promedio para las cuatro técnicas de enseñanza. Por lo tanto se sugiere realizar la prueba múltiple de significación.

2. DISEÑO BLOQUE COMPLETO AL AZAR (DBCA)

Este diseño es uno de los más ampliamente conocidos y difundidos de los diseños experimentales, también se conoce como Diseño Bloque Completamente Aleatorizado, y se caracteriza porque los tratamientos se distribuyen en forma aleatoria, a un grupo de unidades experimentales denominado bloque, la finalidad es que las unidades experimentales dentro de un bloque sean lo más homogéneas posibles, es decir el número de unidades experimentales en cada bloque debe ser igual al número de tratamientos que se quiere estudiar.

MODELO ADITIVO LINEAL

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}; \qquad i = 1,2,\cdots,t \ (t = tratamientos)$$

$$j = 1,2,\cdots,r \ (r = bloques)$$

Y_{ij} = Variable de respuesta observada en la unidad experimental ubicada en el j-ésimo bloque que recibe el tratamiento "i".

μ = Constante para toda observación, es la media de la población.

τ_i = Es el efecto del tratamiento «i», el cual es igual a ($\mu_i - \mu$), es decir, a la Diferenci a entre el promedio poblacional del tratamiento y la media poblacional μ.

β_j = Es el efecto del bloque «j», el cual es igual a ($\mu_j - \mu$), es decir a la diferencia entre el promedio poblacional del bloque y la media poblacional μ.

ε_{ij} = Término que representa el error de su respectiva Y_{ij} se considera variable aleatoria distribuida en forma normal e indep. con media 0 y variancia cte.

HIPÓTESIS:

Respecto a tratamientos:

Modelo I

$$H_0 : \tau_i = 0$$

$$H_a : \tau_i \neq 0$$

Modelo II

$$H_0 : \sigma_\tau^2 = 0$$

$$H_a : \sigma_\tau^2 \neq 0$$

Respecto a los bloques:

Modelo I

$$H_0 : \beta_j = 0$$

$$H_a : \beta_j \neq 0$$

Modelo II

$$H_0 : \sigma_\beta^2 = 0$$

$$H_a : \sigma_\beta^2 \neq 0$$

TABLA ANVA: Fórmulas

F. de V.	G.L.	S.C.	C.M.	Fc	C.M.E.
Bloque	r - 1	$\sum_{j=1}^{r}\frac{Y_{.j}^2}{t}-\frac{Y_{..}^2}{tr}$	$\frac{SC_{bloque}}{r-1}$	$\frac{CM_{bloque}}{CM_{error}}$	$\sigma_e^2+t\sigma_\beta^2$
Tratam.	t - 1	$\sum_{i=1}^{t}\frac{Y_{i.}^2}{r}-\frac{Y_{..}^2}{tr}$	$\frac{SC_{trat.}}{t-1}$	$\frac{CM_{trat.}}{CM_{Error}}$	$\sigma_e^2+r\sigma_\tau^2$
Error Exptal	**(t-1)(r – 1)**	**Por diferencia**	$\frac{SC_{error}}{(t-1)(r-1)}$		σ_e^2
TOTAL	tr - 1	$\sum_{i=1}^{t}\sum_{j=1}^{r}Y_{ij}^2-\frac{Y_{..}^2}{tr}$			

EJEMPLO:

Supongamos que, en el ámbito de la psicología educativa, realizamos una investigación para examinar la influencia que ejerce la metodología de trabajo empleada en el aula (tratamientos) sobre el rendimiento presentado por un grupo de niños en la asignatura de Ciencias Naturales. No obstante, se considera que una posible variable extraña, capaz de contaminar los resultados del estudio, es el nivel de motivación de los sujetos (grupos o bloques). Con el objeto de controlar dicha variable se utiliza un diseño bloque completo al azar.

Continuación:

Se dividen los niños en cuatro bloques, de tres sujetos cada uno, en función de las puntuaciones que obtienen en una prueba destinada a medir su nivel de motivación con respecto al aprendizaje, a saber: I (nivel de motivación bajo, II (nivel de motivación medio-bajo), III (nivel de motivación medio-alto y IV (nivel de motivación alto). A continuación, a cada uno de los sujetos que configuran cada bloque se le somete a una de las tres siguientes metodologías de trabajo A: metodología basada en medios audiovisuales, B: metodología basada en el contacto directo con la naturaleza y C: metodología tradicional. La información se presenta a continuación.

Nivel de Motivación	Metodología de trabajo			
	A = 1	B = 2	C = 3	Total
I	4	14	2	20
II	7	12	4	23
III	8	15	5	28
IV	12	19	5	36
Total	31	60	16	**107**

PROCEDIMIENTO:

a) Grados de Libertad:

GLbloque = r -1 = 4 - 1 = 3

GLtrat. = t -1 = 3 - 1 = 2

GLerror = (r-1)(t –1)t = 2x3 = 6

GLtotal = tr - 1 = 3x4 – 1 = 11

b) Suma de Cuadrados (SS=S.C.)

$$TC = \frac{Y_{..}^2}{tr} = \frac{107^2}{12} = 954.08333$$

$$SC_{Bloque} = \sum_{i=1}^{4} \frac{Y_{.j}^2}{t} - TC = \frac{20^2 + 23^2 + 28^2 + 36^2}{3} - TC = 1003 - TC = 48.91667$$

$$SC_{Trat.} = \sum_{i=1}^{3} \frac{Y_{i.}^2}{r} - TC = \frac{31^2 + 60^2 + 16^2}{4} - TC = 1204.25 - TC = 250.16667$$

$$SC_{Total} = \sum_{i=1}^{3} \sum_{j=1}^{4} Y_{ij}^2 - TC = 4^2 + 7^2 + \cdots + 5^2 - TC = 1269 - TC = 314.916667$$

$$\mathbf{SC_{Error\ Exptal} = SC_{total} - SC_{trat} - SC_{bloque}}$$

$$\mathbf{= 314.916667 - 250.16667 - 48.91667 = 15.8333}$$

Tabla de ANOVA:

F. de V.	G.L.	S.C.	C.M.	Fc	Sign.
Entre bloques	3	48.91667	16.30556	6.18	*
Entre metodologías	2	250.16667	125.08334	47.40	**
Error Exptal.	6	15.83333	2.63889		
Total	11	314.91667			

$$CV = \frac{\sqrt{CM_{Error}}}{\bar{Y}_{..}}(100) = \frac{\sqrt{2.63889}}{8.92}(100) = 18.21\%$$

La prueba F de la tabla de ANOVA nos muestra que la Fc = 6.18>$F_{3,6,0.05}$=4.76, entonces se acepta la hipótesis alterna (Ha:), es decir existe diferencia entre los grupos en el nivel de motivación. Además se encontró que existe diferencia estadística altamente significativa para la metodología empleada en el aula en el rendimiento promedio de los niños en la asignatura de Ciencias Naturales (Fc = 47.40>$F_{2,6,0.01}$=10.93).

EXPERIMENTO FACTORIAL DE LA SERIE 2^n

Vladimiro Ibañez Quispe, Dr.

EL MODELO LINEAL ADITIVO
(dos factores)

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk};$$

$$i = 1,2,\cdots,p \; (Niveles\ de\ factor\ A)$$

$$j = 1,2,\cdots,q \; (Niveles\ de\ factor\ B)$$

$$k = 1,2,\cdots,r \; (repeticiones)$$

Donde:

Y_{ijk} = Es la variable respuesta del k-ésimo observación bajo el j-ésimo nivel de factor B, sujeto al i-ésimo nivel de tratamiento A.

μ = Constante común, media de la población a la cual pertenecen las observaciones.

α_i = Efecto del i-ésimo nivel del factor A.

β_j = Efecto del j-ésimo nivel del factor B.

$(\alpha\beta)_{ij}$ = Efecto de la interacción del i-ésimo nivel del factor A, en el j-ésimo nivel del factor B.

ε_{ijk} = Efecto del error experimental, que está distribuído como

$$\varepsilon_{ijk} \sim DNI(0, \sigma_e^2)$$

TABLA DE ANÁLISIS DE VARIANZA (ANOVA)

Fuentes de Variación	G. L.	S.C.	C.M.	Fc	E(CM) (Modelo II)
Tratamientos	***t - 1***	$\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{Y_{ij\bullet}^2}{r} - \frac{Y_{\bullet\bullet\bullet}^2}{abr}$	$\frac{SC_{tratam.}}{t-1}$	$\frac{CM_{tratam}}{CM_{error}}$	
Factor A	***a - 1***	$\sum_{i=1}^{t}\frac{Y_{i\bullet\bullet}^2}{br} - \frac{Y_{\bullet\bullet\bullet}^2}{abr}$	$\frac{SC_{(A)}}{a-1}$	$\frac{CM_{(A)}}{CM_{error}}$	$\sigma_e^2 + r\sigma_{\alpha\beta}^2 + rb\sigma_{\alpha}^2$
Factor B	***b - 1***	$\sum_{j=1}^{b}\frac{Y_{\bullet j\bullet}^2}{ar} - \frac{Y_{\bullet\bullet\bullet}^2}{abr}$	$\frac{SC_{(B)}}{b-1}$	$\frac{CM_{(B)}}{CM_{error}}$	$\sigma_e^2 + r\sigma_{\alpha\beta}^2 + ra\sigma_{\beta}^2$
Interacción AxB	***(a - 1)(b - 1)***	$\underbrace{\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{Y_{ij\bullet}^2}{r} - \frac{Y_{\bullet\bullet\bullet}^2}{abr}}_{SC_{comb(AxB)}} - (SC_A + SC_B)$	$\frac{SC_{(AB)}}{(a-1)(b-1)}$	$\frac{CM_{(AB)}}{CM_{error}}$	$\sigma_e^2 + r\sigma_{\alpha\beta}^2$
Error Exptal.	***ab(r -1)***	Por diferencia	$\frac{SC_{error}}{(ab-1)(r-1)}$		σ_e^2
Total	***abr - 1***	$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}Y_{ijk}^2 - \frac{Y_{\bullet\bullet\bullet}^2}{abr}$			

HIPÓTESIS A PROBARSE

Respecto al Factor A

$$H_0: \alpha_i = 0, \qquad i = 1,2,\ldots\ldots a$$

$$H_a: \alpha_i \neq 0, \quad para\ cualquier\ i$$

Respecto al Factor B

$$H_0: \beta_j = 0, \qquad j = 1,2,\ldots\ldots b$$

$$H_a: \beta_j \neq 0, \quad para\ cualquier\ j$$

Respecto a la interacción AxB

$$H_o: \alpha\beta_{ij} = 0, \qquad i = 1,2,\ldots\ldots a;\ \mathrm{j} = 1,2,3,\ldots,\mathrm{b}$$

$$H_a: \alpha\beta_{ij} \neq 0, \quad para\ cualquier\ ij$$

EJEMPLO.

De una población de 2500, se escogieron al azar 48 clases de psicología de cuarto de secundaria para un experimento en el cual se buscaba determinar la eficiencia de dos métodos y dos tipos de ayudas diferentes, así como de sus interacciones. Estas clases se dividieron aleatoriamente en cuatro combinaciones según el método y el tipo de ayudas. Transcurrido un semestre, a todas las clases se les administró la misma prueba de rendimiento en psicología. Siendo la unidad de muestreo «la clase», como unidad de análisis se toma la media de la clase, según el criterio de la puntuación obtenida en la prueba de rendimiento. Los resultados se presentan a continuación:

A: Método	Tradicional (1)		Moderno (2)		
B: Ayudas	Conferencia (1)	Instrucción (2)	Conferencia (1)	Instrucción (2)	
1	2	9	10	21	
2	5	12	13	25	
3	6	14	14	31	
4	7	15	16	33	
5	4	10	10	22	
6	6	13	13	26	
7	7	14	14	32	
8	8	16	17	34	
9	4	10	11	22	
10	6	13	13	30	
11	7	14	15	32	
12	10	17	17	35	
$Y_{ij.}$	72	157	163	343	$Y_{...}$= 735
$Y_{i..}$	$Y_{1..}$= 229		$Y_{2..}$= 506		
$Y_{.j.}$	$Y_{.1.}$= 235		$Y_{.2.}$= 500		

SOLUCIÓN

a) GRADOS DE LIBERTAD:

$GL_{MÉTODO} = a - 1 = 2 - 1 = 1$

$GL_{AYUDA} = b - 1 = 2 - 1 = 1$

$GL_{MxA} = (a-1)(b-1) = (2-1)(2-1) = 1$

$GL_{ERROR\ EXP} = ab(r-1) = 2x2x(12-1) = 44$

$GL_{TOTAL} = abr - 1 = 2x2x12 - 1 = 47$

b) SUMA DE CUADRADOS:

$$TC = \frac{Y^2_{\bullet\bullet\bullet}}{abr} = \frac{(735)^2}{2x2x12} = 11254.6875$$

$$SC_{Método} = \sum_{i=1}^{a} \frac{Y^2_{i\bullet\bullet}}{br} - TC = \frac{229^2 + 506^2}{2x12} - TC = 1598.520833$$

$$SC_{Ayuda} = \sum_{j=1}^{b=2} \frac{Y^2_{\bullet j\bullet}}{ar} - TC = \frac{235^2 + 500^2}{2x12} - TC = 1463.020833$$

$$SC_{comb(MxA)} = \sum_{i=1}^{a=2}\sum_{j=1}^{b=2}\frac{Y_{ij\bullet}^{2}}{r} - TC = \frac{72^2+157^2+163^2+343^2}{12} - TC = 3249.5625$$

$$SC_{Int(MxA)} = SC_{Comb(MxA)} - (SC_{Método} + SC_{Ayuda}) =$$

$$= 3249.5625 - (1598.520833 + 1463.020833) = 188.020834$$

$$SC_{Total} = \sum_{i=1}^{a=2}\sum_{j=1}^{b=2}\sum_{k=1}^{r=12} Y_{ijk}^{2} - TC$$

$$= 2^2 + 5^2 + \cdots + 32^2 + 35^2 - TC$$

$$= 118758.73 - TC = 14969 - 11254.6875$$

$$= 3714.3125$$

$$SC_{ERROR} = SC_{TOTAL} - SC_{MÉTODO} - SC_{AYUDA} - SC_{MxA}$$

$$= 3714.3125 - 1598.520833 - 1463.020833 - 188.020834$$

$$= 464.75$$

ANÁLISIS DE VARIANZA PARA LA PUNTUACIÓN OBTENIDA EN LA PRUEBA DE RENDIMIENTO

F. De V.	G.L.	S.C.	C.M.	Fc.	Signif.
Métodos (M)	1	1598.520833	1598.520833	151.34	**
Ayudas (A)	1	1463.020833	1463.020833	138.51	**
Interacción (MxA)	1	188.020834	188.020834	17.80	**
Error Experimental	44	464.750000	10.562500		
Total	47	3714.312500			

≤

$$CV. = \frac{\sqrt{CM_{error}}}{\bar{Y}_{...}} \text{x}100 = \frac{\sqrt{10.56250}}{735/48} \text{x}100 = 21.22\%$$

Al análisis de varianza (ANVA) del cuadro precedente, para la interacción Métodos por Ayudas, existe diferencia estadística altamente significativa (P<=0.01), lo que nos demuestra que estos factores no son independientes, anulando los efectos prinicipales entre Métodos y Ayudas, adquiere importancia los efectos simples para nuestras conclusiones.

ANÁLISIS DE VARIANZA PARA LOS EFECTOS SIMPLES

a) Suma de cuadrados de M en a_j

$$SC_{(\text{M dentro } a_0)} = \sum_{i=1}^{2}\sum_{j=}^{2}\frac{Y_{ij\bullet}^2}{r} - \frac{Y_{\bullet j\bullet}^2}{ar} = \frac{72^2 + 163^2}{12} - \frac{235^2}{2x12} = 345.0416667$$

$$SC_{(\text{M dentro } a_1)} = \sum_{i=1}^{2}\sum_{j=}^{2}\frac{Y_{ij\bullet}^2}{r} - \frac{Y_{\bullet j\bullet}^2}{ar} = \frac{157^2 + 343^2}{12} - \frac{500^2}{2x12} = 1441.5$$

	Métodos				
		m_0	m_1	$Y_{.j.}$	
Ayudas	a_0	72	163	235	
	a_1	157	343	500	
	$Y_{i..}$	229	506	735	**= Y...**

b) Suma de cuadrados de A en m_i

$$SC_{(A \text{ dentro } m_0)} = \sum_{i=1}^{2}\sum_{j=}^{2}\frac{Y_{ij\bullet}^2}{r} - \frac{Y_{i\bullet\bullet}^2}{br} = \frac{72^2 + 157^2}{12} - \frac{229^2}{2x12} = 301.0416667$$

$$SC_{(A \text{ dentro } m_1)} = \sum_{i=1}^{2}\sum_{j=}^{2}\frac{Y_{ij\bullet..}^2}{r} - \frac{Y_{i\bullet\bullet}^2}{br} = \frac{163^2 + 343^2}{12} - \frac{506^2}{2x12} = 1350.0$$

ANÁLISIS DE VARIANZA DE EFECTOS SIMPLES

F. De V.	G.L.	S.C.	C.M.	Fc.	Signif.
M dentro de a_0	1	345.041667	345.04167	32.67	**
M dentro de a_1	1	1441.500000	1441.50000	136.47	**
A dentro de a_0	1	301.041667	301.04167	28.50	**
A dentro de a_1	1	1350.000000	1350.00000	127.81	**
Error Experimental	44	464.750000	10.56250		

CONCLUSIONES:

M dentro a_0: Existe diferencia significativa entre los niveles de m_0 y m_1 bajo los niveles de a_0, es decir hay diferencia significativa en los métodos de enseñanza con la ayuda de conferencia de clase.

M dentro a_1: Existe diferencia significativa entre los niveles de m_0 y m_1 bajo los niveles de a_1, es decir hay diferencia significativa en los métodos de enseñanza con la ayuda de instrucción programada.

CONCLUSIONES:

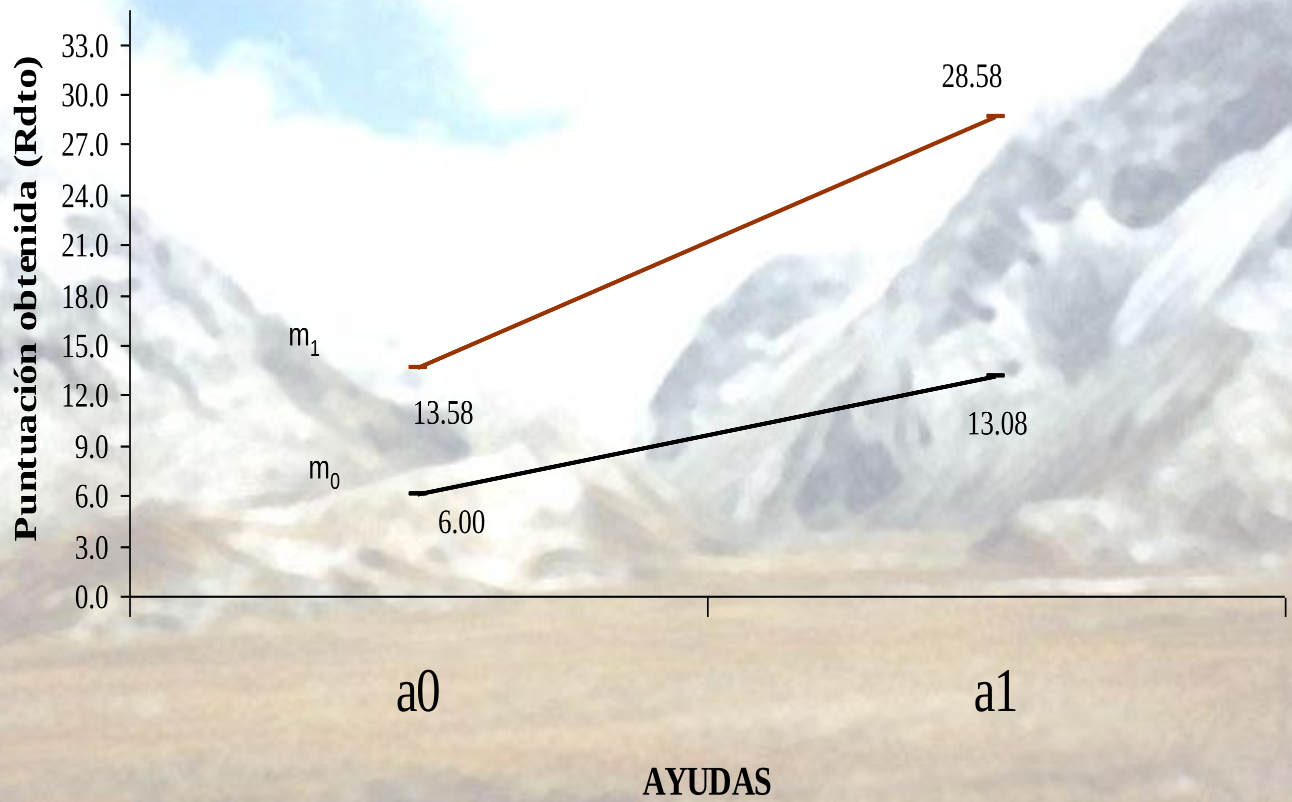
A dentro m_0: Existe diferencia significativa entre los niveles de a_0 y a_1 bajo los niveles de m_0, es decir hay diferencia significativa con las Ayudas de clase con el método tradicional.

A dentro m_1: Existe diferencia significativa entre los niveles de a_0 y a_1 bajo los niveles de m_1, es decir hay diferencia significativa con las Ayudas de clase con los métodos modernos.

Otra forma de observar la interacción de los dos factores (Método con Ayudas), es graficando con los promedios de la siguiente forma:

Métodos

Ayudas		m_0	m_1
	a_0	6.00	13.58
	a_1	13.08	28.58



EXPERIMENTO MULTIFACTORIAL DE LA SERIE 3^n

Existen experimentos con tres factores, se presenta interacción de primer orden e interacción de segundo orden; el análisis tiende a complicarse a medida que se detecte significación en las interacciones, por ejemplo, si existe interacción de 3 factores se analiza estos resultados por efectos simples-simples; en cambio, sino existe interacción de tres factores, pero se presenta interacciones de dos factores, entonces se analizará mediante efectos simples.

EL MODELO LINEAL ADITIVO
(tres factores)

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + B_l + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl}$$

$$i = 1,2,\cdots,a \;\; (\,Niveles\ de\ factor\ A\,)$$

$$j = 1,2,\cdots,b \;\; (\,Niveles\ de\ factor\ B\,)$$

$$k = 1,2,\cdots,c \;\; (\,Niveles\ de\ factor\ C\,)$$

$$l = 1,2,\cdots,r \;\; (\,Bloque\,)$$

Y_{ijkl} = Es la variable respuesta de la l-ésima bloque al que se le aplicó el i-ésimo nivel del factor A, j-ésimo nivel del factor B, y k-ésimo nivel del factor C.

μ **= Efecto de la media general.**

B_l **= Efecto del del l-ésimo bloque.**

α_i **= Efecto del i-ésimo nivel del factor A.**

β_j **= Efecto del j-ésimo nivel del factor B.**

γ_k **= Efecto del k-ésimo nivel el factor C**

$(\alpha\beta)_{ij}$ **= Efecto de la interacción del i-ésimo nivel del factor A y del j-ésimo nivel factor B.**

$(\alpha\gamma)_{ik}$ **= Efecto de la interacción del i-ésimo nivel del factor A y del k-ésimo nivel del factor C.**

$(\beta\gamma)_{jk}$ **= Efecto de la interacción del j-ésimo nivel del factor B y del k-ésimo nivel del factor C.**

$(\alpha\beta\gamma)_{ijk}$ **= Efecto de la interacción del i-ésimo nivel del factor A y del j-ésimo nivel del factor B y del k-ésimo nivel del factor C.**

ε_{ijkl} **= Efecto aleatorio o error experimental en la obtención de Yijkl**

$$\varepsilon_{ijkl} \sim DNI(0, \sigma_e^2)$$

ANÁLISIS DE VARIANZA (ANOVA) DE TRES FACTORES

Fuentes de Variación	G. L.	S.C.	C.M.	Fc	E(CM) (Modelo II)
Bloques	***r - 1***	$\sum_{l=1}^{r}\frac{Y_{\bullet\bullet\bullet l}^2}{br} - \frac{Y_{\bullet\bullet\bullet\bullet}^2}{abcr}$	$\frac{SC_{bloques}}{r-1}$	$\frac{CM_{bloq.}}{CM_{error}}$	$\sigma_e^2 + abc\,\sigma_\rho^2$
Tratamientos	***abc - 1***	$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\frac{Y_{ijk\bullet}^2}{r} - \frac{Y_{\bullet\bullet\bullet\bullet}^2}{abcr}$	$\frac{SC_{tratam.}}{abc-1}$	$\frac{CM_{tratam.}}{CM_{error}}$	$\sigma_e^2 + r\,\sigma_\alpha^2$
Factor A	***a - 1***	$\sum_{i=1}^{t}\frac{Y_{i\bullet\bullet\bullet}^2}{bcr} - \frac{Y_{\bullet\bullet\bullet\bullet}^2}{abcr}$	$\frac{SC_{(A)}}{a-1}$	$\frac{CM_{(A)}}{CM_{error}}$	$\sigma_e^2 + r\sigma_{\alpha\beta\gamma}^2 + rc\sigma_{\alpha\beta}^2 + rb\sigma_{\alpha\gamma}^2 + rbc\sigma_\alpha^2$
Factor B	***b - 1***	$\sum_{j=1}^{b}\frac{Y_{\bullet j\bullet\bullet}^2}{acr} - \frac{Y_{\bullet\bullet\bullet\bullet}^2}{abcr}$	$\frac{SC_{(B)}}{b-1}$	$\frac{CM_{(B)}}{CM_{error}}$	$\sigma_e^2 + r\sigma_{\alpha\beta\gamma}^2 + rc\sigma_{\alpha\beta}^2 + ra\sigma_{\beta\gamma}^2 + rac\sigma_\beta^2$
Factor C	***c - 1***	$\sum_{k=1}^{c}\frac{Y_{\bullet\bullet k\bullet}^2}{abr} - \frac{Y_{\bullet\bullet\bullet\bullet}^2}{abcr}$	$\frac{SC_{(C)}}{c-1}$	$\frac{CM_{(C)}}{CM_{error}}$	$\sigma_e^2 + r\sigma_{\alpha\beta\gamma}^2 + rb\sigma_{\alpha\gamma}^2 + ra\sigma_{\beta\gamma}^2 + rab\sigma_\gamma^2$
Interacción AxB	***(a - 1)(b - 1)***	$\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{Y_{ij\bullet\bullet}^2}{cr} - \frac{Y_{\bullet\bullet\bullet\bullet}^2}{abcr} - SC_A - SC_B$	$\frac{SC_{(AB)}}{(a-1)(b-1)}$	$\frac{CM_{(AB)}}{CM_{error}}$	$\sigma_e^2 + r\sigma_{\alpha\beta\gamma}^2 + rc\sigma_{\alpha\beta}^2$
Interacción AxC	***(a - 1)(c - 1)***	$\sum_{i=1}^{a}\sum_{k=1}^{c}\frac{Y_{i\bullet k\bullet}^2}{br} - \frac{Y_{\bullet\bullet\bullet\bullet}^2}{abcr} - SC_A - SC_C$	$\frac{SC_{(AC)}}{(a-1)(c-1)}$	$\frac{CM_{(AC)}}{CM_{error}}$	$\sigma_e^2 + r\sigma_{\alpha\beta\gamma}^2 + rb\sigma_{\alpha\gamma}^2$
Interacción BxC	***(b - 1)(c - 1)***	$\sum_{j=1}^{b}\sum_{k=1}^{c}\frac{Y_{\bullet jk\bullet}^2}{ar} - \frac{Y_{\bullet\bullet\bullet\bullet}^2}{abcr} - SC_B - SC_C$	$\frac{SC_{(BC)}}{(b-1)(c-1)}$	$\frac{CM_{(BC)}}{CM_{error}}$	$\sigma_e^2 + r\sigma_{\alpha\beta\gamma}^2 + ra\sigma_{\beta\gamma}^2$
Interacción AxBxC	***(a - 1)(b - 1)(c - 1)***	$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\frac{Y_{ijk\bullet}^2}{r} - \frac{Y_{\bullet\bullet\bullet\bullet}^2}{abcr} - SC_A - SC_B - SC_C - SC_{AB} - SC_{AC} - SC_{BC}$	$\frac{SC_{(ABC)}}{(a-1)(b-1)(c-1)}$	$\frac{CM_{(ABC)}}{CM_{error}}$	$\sigma_e^2 + r\,\sigma_{\alpha\beta\gamma}^2$
Error Experimental	***(abc - 1)(r -1)***	SC $_{Total}$ - SC $_{tratamientos}$ - SC $_{bloques}$	$\frac{SC_{error}}{(abc-1)(r-1)}$		σ_e^2
Total	***abcr - 1***	$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{r}Y_{ijkl}^2 - \frac{Y_{\bullet\bullet\bullet\bullet}^2}{abcr}$			

EJEMPLO.

Se pretende realizar un experimento con objeto de estudiar el efecto de la variable sexo (Factor A), cantidad de dimensiones irrelevantes (Factor B) y clase de concepto (Factor C), sobre una tarea de identificación de conceptos. El factor A actúa a dos niveles (a_1, varones y a_2, mujeres), el factor B a tres niveles (b_1, una dimensión irrelevante al concepto tal como el número; b_2, dos dimensiones irrelevantes, número y sombreado y b_3, tres dimensiones irrelevantes, número sombreado y barra). El factor C actúa a dos niveles (c_1, concepto conjuntivo, y c_2, concepto disyuntivo). Las dimensiones relevantes que fueron consideradas a fin de que el sujeto resolviera el problema son forma y color. Se trata, por tanto, de averiguar si el concepto «triángulo rojo» necesitaba menos cantidad de ensayos previos para ser identificado que el concepto «o triángulo rojo, o ambas cosas a la vez». Evidentemente, la identificación del concepto está en función de la cantidad de dimensiones irrelevantes que intervienen y en función de la variable sexo. Una vez realizado el experimento, se obtuvieron los datos que a continuación se presentan en la correspondiente tabla (El experimento se realizó mediante un diseño bloque al azar, siendo la variable de bloqueo «la capacidad de resolución de problemas medido a través de una prueba»).

Factor A	a_1						a_2						
Factor B	b_1		b_2		b_3		b_1		b_2		b_3		
Factor C	c_1	c_2	c_1	c_2	c_1	c_2	c_1	c_2	c_1	c_2	c_1	c_2	
I	10	10	5	6	40	32	15	12	8	27	27	25	
II	16	13	7	10	41	43	15	21	13	28	41	41	
III	20	15	7	13	46	47	15	26	22	38	47	57	
IV	20	19	7	17	51	52	17	31	27	38	52	60	
V	25	25	15	21	60	70	18	32	30	43	62	71	
$Y_{ijk.}$	91	82	41	67	238	244	80	122	100	174	229	254	$Y_{....}$ 1722
$Y_{ij..}$	173		108		482		202		274		483		
$Y_{i.k.}$	a_1c_1= 370		a_1c_2= 393				a_2c_1= 409		a_2c_2= 550				
$Y_{.jk.}$	b_1c_1= 171		b_1c_2= 204		b_2c_1= 141		b_2c_2= 241		b_3c_1= 467		b_3c_2= 498		
$Y_{i...}$	a_1 = 763						a_2 = 959						
$Y_{.j..}$	b_1 = 375			b_2 = 382			b_3 = 965						
$Y_{..k.}$	c_1 = 779						c_2 = 943						

SOLUCIÓN

a) GRADOS DE LIBERTAD:

GL_{BLOQUE} = r - 1 = 5 - 1 = 4

$GL_{tratamientos}$: abc - 1 = 2x3x2-1 = 11

GL_{A} = a - 1 = 2 - 1 = 1

GL_{B} = b - 1 = 3 - 1 = 2

GL_{C} = b - 1 = 2 - 1 = 1

GL_{AxB} = (a-1)(b-1) = (2-1)(3-1) = 2

GL_{AxC} = (a-1)(c -1) = (2-1)(2-1) = 1

GL_{BxC} = (b -1)(c - 1) = (3-1)(2-1) = 2

GL_{AxBxC} = (a-1)(b-1)(c-1) = (2 - 1)(3 - 1)(2 - 1) = 2

$GL_{ERROR\ EXP}$ = (r-1)(abc-1) = (4)(2x3x2-1) = 44

GL_{TOTAL} = abcr - 1 = 2x3x2x5 - 1 = 59

b) SUMA DE CUADRADOS:

$$TC = \frac{Y^2_{\bullet\bullet\bullet\bullet}}{abcr} = \frac{(1722)^2}{2x3x2x5} = 49421.40$$

$$SC_{Bloque} = \sum_{l=1}^{r} \frac{Y^2_{\bullet\bullet\bullet l}}{abc} - TC = \frac{217^2 + 289^2 + 353^2 + 391^2 + 472^2}{2x3x2} - TC = 3152.266666$$

$$SC_{tratam.} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} \frac{Y^2_{ijk\bullet}}{r} - TC$$

$$= \frac{91^2 + 82^2 + 41^2 + 67^2 + 238^2 + 244^2 + 80^2 + 122^2 + 100^2 + 174^2 + 229^2 + 254^2}{5} - TC$$

$$= 13753$$

S.C. De tratamientos se parten en: Factores principales

$$SC_A = \sum_{i=1}^{a} \frac{Y_{i\bullet\bullet\bullet}^2}{bcr} - TC = \frac{763^2 + 959^2}{3x2x5} - TC = 640.266666$$

$$SC_B = \sum_{j=1}^{b} \frac{Y_{\bullet j\bullet\bullet}^2}{acr} - TC = \frac{375^2 + 382^2 + 965^2}{2x2x5} - TC = 11467.30$$

$$SC_C = \sum_{k=1}^{c} \frac{Y_{\bullet\bullet k\bullet}^2}{abr} - TC = \frac{779^2 + 943^2}{2x3x5} - TC = 448.266666$$

SUMA DE CUADRADOS DE INTERACCIONES

$$SC_{Comb.(\text{AxB})} = \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{Y_{ij..}^{2}}{cr} - TC = \frac{173^2+108^2+482^2+202^2+274^2+483^2}{2x5} - TC = 12887.20$$

$$SC_{Int(\text{AxB})} = SC_{Comb(\text{AxB})} - (SC_A + SC_B)$$

$$= 12887.20 - (640.266666 + 11467.30) = 779.633334$$

$$SC_{Comb.(\text{AxC})} = \sum_{i=1}^{a}\sum_{k=1}^{c}\frac{Y_{i.k.}^{2}}{br} - TC = \frac{370^2+393^2+409^2+550^2}{3x5} - TC = 1320.60$$

$$SC_{Int(\text{AxC})} = SC_{Comb(\text{AxC})} - (SC_A + SC_C)$$

$$= 1320.60 - (640.266666 + 448.266666) = 232.066668$$

$$SC_{Comb.(\mathrm{BxC})} = \sum_{j=1}^{b}\sum_{l=1}^{r}\frac{Y_{.jk.}^{2}}{ar} - TC = \frac{171^2+204^2+141^2+241^2+467^2+498^2}{2\mathrm{x}5} - TC = 12069.80$$

$$SC_{Int(\mathrm{BxC})} = SC_{Comb(\mathrm{BxC})} - (SC_B + SC_C)$$

$$= 12069.80 - (11467.30 + 448.266666) = 154.233334$$

$$SC_{Comb.(\mathrm{AxBxC})} = \sum_{i=1}^{a=2}\sum_{j=1}^{b=3}\sum_{k=1}^{c=2}\frac{Y_{ijk.}^{2}}{r} - TC = SC_{tratam.} = 13753$$

$$SC_{Int(\mathrm{AxBxC})} = SC_{Comb(\mathrm{AxBxC})} - (SC_A + SC_B + SC_C + SC_{AxB} + SC_{AxC} + SC_{BxC})$$

La suma de cuadrados combinado corregido equivale a SC$_{\text{tratamiento corregido}}$

$$SC_{Inter(AxBxC)} = 13753 - (640.266666 + 11467.30 + 448.26666 +$$

$$779.633334 + 232.066668 + 154.233334)$$

$$= 31.233332$$

$$SC_{total} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{r} Y_{ijkl}^{2} - TC = 10^2 + 16^2 + \cdots + 60^2 + 71^2 - TC$$

$$= 67470 - TC = 18048.60$$

$$SC_{Error\,exptal} = SC_{total} - SC_{bloque} - SC_{Tratam.}$$

$$= 18048.60 - 3152.266666 - 13753$$

$$= 1143.333334$$

ANÁLISIS DE VARIANZA

F. De V.	G.L.	S.C.	C.M.	Fc.	Signific.
Bloques	4	3152.266666	788.066667	30.33	**
Tratamientos	11	13753.000000	1250.272727	48.12	**
Factor (A)	1	640.266666	640.266666	24.64	**
Factor (B)	2	11467.300000	5733.650000	220.65	**
Factor (C)	1	448.266666	448.266666	17.25	**
AxB	2	779.633334	389.816667	15.00	**
AxC	1	232.066668	232.066668	8.93	*
BxC	2	154.233334	77.116667	2.97	n.s.
AxBxC	2	31.233332	15.616666	0.60	n.s.
Error Experimental	44	1143.333334	25.984849		
Total	59	18048.600000			

$$CV. = \frac{\sqrt{CM_{error}}}{\overline{Y}_{....}} \text{ x}100 = \frac{\sqrt{25.984849}}{1722/60} X100 = 17.76\%$$

CONCLUSIÓN

Se encontró diferencia estadística altamente significativa para los factores: (A, B, y C) (P<=0.01), así mismo para las interacciones de primer orden se encontró diferencia estadística altamente significativa para la interacción de AxB, y significativa para la interacción AxC, en cambio no se encontró diferencia alguna para las interacciones BxC, tampoco para la interacción de segundo orden (AxBxC), esto quiere decir que los factores actúan en forma independiente o se deben al azar, por consiguiente las conclusiones se sacan de los efectos principales significativos, y las interacciones que han resultado ser significativas, para tal efecto se debe realizar los efectos simples para obtener las conclusiones en el presente trabajo.

BIBLIOGRAFÍA CONSULTADA

- **Córdova, M. (1999). Estadística Inferencial. 1ra. Edic. Edit. Moshera S.R.L. Lima – Perú.**
- **IBAÑEZ, V. (2000). “Estadística Inferencial Aplicada en Ganadería”. Edición Primera. FINESI - UNA-Puno.**
- **IBAÑEZ, V. (2006). “Estadística Inferencial Aplicada a la Salud”. Edición Primera. FINESI - UNA-Puno.**
- **IBAÑEZ, V. (2006). “Métodos Estadísticos para la Investigación”. Edición Primera. Fac. de Enfermería - UNA-Puno.**
- **MITACC, M.M.(1990). «Tópicos de Inferencia Estadística». Primera Edición. Editorial San Marcos. Lima - Perú.**
- **POZO, M. y ORIHUELA, P. (1992). «Estadística Aplicada a la Educación». Primera Edición. Editorial. Talleres Gráficos P.L. Villanueva S.A. Lima - Perú.**
- **VILLALTA, P. e IBAÑEZ, V. (1998). "Estadística Pecuaria". Primera Edición. Editorial Universitaria. Fac. Medicina Veterinaria y Zootecnia. UNA - Puno.**
- **ZEA, F.W. e IBAÑEZ, Q.V. (2003). "Tablas Estadísticas". Segunda Edición. Edit. Universitaria. Fac. Cs. Agrarias - FINESI. UNA - Puno.**