Contrastes de Hipótesis

1) En un Hospital se determinó que el promedio de colesterol en sangre no se ajusta al estándar establecido de 180 mg/dl. Para comprobarlo, seleccionamos aleatoriamente a 15 individuos, obteniendo las determinaciones siguientes:

   **187 203 168 190 182 196 175 190 184 179 178 183 198 200 175**

   ¿Apoyan estos datos muestrales la sospecha del centro hospitalario? Considera un nivel de significación del 5%.

2) Se estudiaron 40 muestras de aceite crudo de determinado proveedor con el fin de detectar la presencia del níquel mediante una prueba que nunca da un resultado erróneo. Si en 5 de dichas muestras se observó la presencia de níquel, ¿podemos creer al proveedor cuando asegura que a lo sumo el 8% de las muestras contienen níquel? Considera un nivel de significación del 5%.

3) Un psicólogo educacional considera que el número de alumnos que abandonan los estudios es inferior al 15% que se establece por las autoridades educativas. Para comprobar la certeza de su creencia, selecciona al azar un grupo de 500 alumnos resultando que sólo 59 dejaron los estudios. ¿Podemos aceptar la hipótesis planteada por el psicólogo? Considera un nivel de significación del 5%.

4) Durante los últimos años se ha establecido en 1.1 la desviación típica de las calificaciones en las pruebas de acceso a la Universidad. Ante la sospecha de que en la actualidad ha aumentado la dispersión de las calificaciones, se seleccionan las notas correspondientes a 25 alumnos presentados en la última convocatoria, resultando como desviación típica 1.64. Considerando un nivel de significación del 1%, ¿podemos concluir que aumentó la dispersión de las calificaciones de las pruebas de acceso?

5) Las calificaciones de 18 alumnos en la asignatura de Psicología Matemática (PM) y en Psicología General (PG) están caracterizadas por las siguientes estimaciones:

Media muestral en PM: 5.333
Media muestral en PG: 6.117
Desviación típica muestral en PM: 1.9785
Desviación típica muestral en PG: 1.6507
Covarianza muestral: 2.4708
Coeficiente de correlación de Pearson: 0.7565

Con un nivel de significación del 5%, ¿podemos afirmar que las calificaciones en ambas asignaturas están igualmente dispersas?

6) Se ha llevado a cabo un estudio sobre la relación entre la actividad cerebral mientras se ven anuncios en televisión y la capacidad de la persona para recordar dichos anuncios. Se han mostrado anuncios de dos marcas para diez productos a las personas en la muestra. Para cada anuncio se ha medido la capacidad de cada persona para recordarlo pasadas 24 horas, y a cada anuncio de un producto se le han asignado las etiquetas “recuerdo fuerte” o “recuerdo débil”. La siguiente tabla muestra un índice de la actividad cerebral de las personas que han visto estos anuncios en el estudio.

| producto: $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| recuerdo fuerte: $x_i$ | 137 | 135 | 83 | 125 | 47 | 46 | 114 | 157 | 57 | 144 |
| recuerdo débil: $y_i$ | 53 | 114 | 81 | 86 | 34 | 66 | 89 | 113 | 88 | 111 |
| dif.: $d_i = x_i - y_i$ | 84 | 21 | 2 | 39 | 13 | $-20$ | 25 | 44 | $-31$ | 33 |

¿Se podría aceptar que en promedio la actividad cerebral es mayor para el grupo con recuerdo fuerte que para el grupo con recuerdo débil?

7) Se realiza un estudio entre auditores sobre la actividad de las mujeres en su profesión. A los encuestados se les pide que den su opinión con un valor entre 1 (muy en desacuerdo) y cinco (muy de acuerdo) sobre la afirmación “En auditoría se asignan los mismos trabajos a las mujeres y a los hombres”.

De una muestra de 186 auditores (varones) se obtuvo una respuesta promedio de 4.059 con una cuasi desviación típica de 0.839. Para una muestra independiente de 172 mujeres auditoras la respuesta promedio fue de 3.680 con una cuasi desviación típica de 0.966.

Contraste la hipótesis nula (para un nivel de significación igual a 0.0001) de que las medias de las dos poblaciones son iguales, frente a la alternativa de que la media de la población es mayor para los auditores varones.

# Resumen de teoría y fórmulas

Recordemos algunos conceptos anteriores:

Nivel de confianza: Usualmente del 95%, aunque también suelen usarse 90% y 99%. Se representa como $NC = 1 - \alpha$, donde $\alpha$ es el margen de error o nivel de significación.

Nivel de significación: $\alpha$. Usualmente 5%, 10% o 1%.

**Contraste de Hipótesis:** Procedimiento por el cual decidimos si una propuesta sobre la población puede aceptarse o no.

$H_0$: Hipótesis Nula
$H_1$: Hipótesis Alternativa

Usualmente ponemos en la alternativa lo que se pregunta en el enunciado, o lo que evidencian los datos muestrales.

Tipos de contraste según la hipótesis alternativa: Bilateral ($\neq$), Unilateral derecho ($>$), Unilateral izquierdo ($<$).

Nota: En la hipótesis nula NUNCA puede ir un $\neq$. En la hipótesis alternativa NUNCA puede ir un $=, \geq, \leq$.

Tipos de errores que se pueden cometer:

- Error de Tipo I: Rechazar $H_0$ cuando es cierta. La probabilidad del error de tipo I coincide con el nivel de significación $\alpha$.
- Error de Tipo II: No rechazar $H_0$ cuando es falsa. La probabilidad del error de tipo II se le llama $\beta$.

Potencia del contraste: $1 - \beta$

Región de Aceptación: Intervalo en el que no se puede rechazar la hipótesis nula.

Región de Rechazo o Región crítica: Intervalo en el que se rechaza la hipótesis nula.

Función de decisión o estadístico de contraste: Es el estadístico que permite resolver el contraste.

Nivel crítico de decisión: Extremo (o extremos) del intervalo o región de aceptación.

# Contrastes de hipótesis para los diferentes parámetros

## Notación

<u>Población</u>
$\mu$: media poblacional
$\sigma$: varianza poblacional
$\pi$: proporción poblacional

<u>Muestra</u>
$n$: tamaño muestral
$\bar{x}$: media muestral
$s^2$: varianza muestral
$s'^2$: cuasivarianza muestral
$\hat{p}$: proporción muestral
$r$: coeficiente de correlación muestral

## Funciones de decisión y su distribución

### 1. Contraste para una MEDIA

Ejemplo:
$H_0: \mu = \mu_o$
$H_0: \mu \neq \mu_o$

a. Varianza de la población conocida:

$$Z = \frac{\bar{x} - \mu_o}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

<u>Niveles críticos:</u> $z_{\alpha/2}$ si el contraste es bilateral y $z_\alpha$ o $-z_\alpha$ si el contraste es unilateral (derecha e izquierda respectivamente)

b. Varianza de la población desconocida y muestras grandes (n>30):

$$Z = \frac{\bar{x} - \mu_o}{\frac{s}{\sqrt{n-1}}} \sim N(0,1)$$

Con la cuasi-desviación típica:

$$Z = \frac{\bar{x} - \mu_o}{\frac{s'}{\sqrt{n}}} \sim N(0,1)$$

c. Varianza de la población desconocida y muestras pequeñas (n<=30):

$$T = \frac{\bar{x} - \mu_o}{\frac{s'}{\sqrt{n}}} \sim t_{n-1;\alpha}$$

## 2. Contraste para una VARIANZA

Ejemplo:
$H_0: \sigma^2 = \sigma_0^2$
$H_0: \sigma^2 \neq \sigma_0^2$

$$\hat{\chi} = \frac{ns^2}{\sigma_0^2} \sim \chi^2_{n-1;\alpha}$$

## 3. Contraste de comparación de 2 MEDIAS

a. Muestras independientes y varianzas de las poblaciones conocidas.

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

b. Muestras independientes y varianzas de las poblaciones desconocidas pero iguales.

i. Muestras grandes ($n_1 + n_2 - 2 > 50$)

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1)$$

ii. Muestras pequeñas $(n_1 + n_2 - 2 \leq 50)$

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

c. Muestras independientes y varianzas de las poblaciones desconocidas y distintas.

i. Muestras pequeñas pero de igual tamaño $n_1 = n_2 = n$

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2 + s_2^2}{n - 1}}} \sim t_{2n-2}$$

ii. En otro caso, calcularemos la expresión "gl=grados de libertad":

$$gl = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 + 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 + 1}} - 2$$

- Muestras grandes $(gl > 50)$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_1 - \mu_2}{\sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}} \sim N(0,1)$$

- Muestras pequeñas $(gl \leq 50)$

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_1 - \mu_2}{\sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}} \sim t_{gl}$$

d. Muestras relacionadas (pareadas) y varianza de la diferencia conocida $\sigma_D^2$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\frac{\sigma_D}{\sqrt{n}}} \sim N(0,1)$$

e. Muestras relacionadas y varianza de la diferencia desconocida (muestras grandes)

- Primero calcular la varianza muestral de la diferencia:

$$s_D^2 = \frac{\sum(D_i - \overline{D})^2}{n}$$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\frac{s_D}{\sqrt{n-1}}} \sim N(0,1)$$

f. Muestras relacionadas y varianza de la diferencia desconocida (muestras pequeñas)

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\frac{s_D}{\sqrt{n}}} \sim t_{n-1}$$

$T = \frac{\frac{\overline{D} - D_0}{s_D}}{\sqrt{n}} \sim t_{n-1}$, donde $D$ es la diferencia y $D_0$ es 0

**4. Contraste de comparación de 2 VARIANZAS**

Ejemplo:
$H_0 : \sigma_1^2 = \sigma_2^2$
$H_1 : \sigma_1^2 \neq \sigma_2^2$

a. Observaciones relacionadas (pareadas)
Con varianzas:

$$T = \frac{(s_1^2 - s_2^2)\sqrt{n-2}}{2\, s_1\, s_2 \sqrt{1 - r^2}} \sim t_{n-2}$$

Con cuasi-varianzas:

$$T = \frac{(s_1'^2 - s_2'^2)\sqrt{n-2}}{2\, s_1'\, s_2' \sqrt{1 - r^2}} \sim t_{n-2}$$

b. Observaciones independientes

$$\hat{F} = \left(\frac{s_1'^2}{s_2'^2}\right)\left(\frac{\sigma_2^2}{\sigma_1^2}\right) = \frac{\frac{n_1}{n_1 - 1} s_1^2}{\frac{n_2}{n_2 - 1} s_2^2}\left(\frac{\sigma_2^2}{\sigma_1^2}\right) \sim F_{n_1 - 1, n_2 - 2}$$

5. **Contraste para una PROPORCIÓN**

Ejemplo:

$H_0: p = p_0$

$H_1: p > p_0$

a. Muestras grandes n>30

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

b. Muestras pequeñas

$$T = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim t_{n-1}$$

6. **Contraste de comparación de 2 PROPORCIONES**

a. Muestras independientes con proporciones poblacionales iguales

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{D_{est}} \sim N(0,1)$$

donde $D_{est} = \sqrt{\left(\frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1+n_2}\right)\left(1 - \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1+n_2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

b. Muestras independientes con proporciones poblacionales desiguales

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (\pi_1 - \pi_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \sim N(0,1)$$