

Bioestadística

- MSc Ana Luisa Romero Pimentel

Instituto Nacional de Medicina Genómica
Laboratorio de Enfermedades Psiquiátricas

apimentel@inmegen.edu.mx

Dr. Raul Alejandro Mejía Pedroza

Instituto Nacional de Medicina Genómica
Departamento de Genómica Computacional

ralejandro@inmegen.edu.mx

- Página web
 - <https://github.com/raulmejia/>
- R
 - <https://cran.r-project.org/>

Formulas and Tables

Ch. 8: Test Statistics (one population)

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \quad \text{Proportion—one population}$$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad \text{Mean—one population} \quad (\sigma \text{ known})$$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad \text{Mean—one population} \quad (\sigma \text{ unknown})$$

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \quad \text{Standard deviation or variance—} \quad \text{one population}$$

Ch. 9: Test Statistics (two populations)

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}} \quad \text{Two proportions} \quad \bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad \text{df = smaller of } n_1 - 1, n_2 - 1$$

Two means—-independent; σ_1 and σ_2 unknown, and not assumed equal.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} \quad (\text{df} = n_1 + n_2 - 2) \quad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Two means—-independent; σ_1 and σ_2 unknown, but assumed equal.

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad \text{Two means—-independent; } \sigma_1, \sigma_2 \text{ known.}$$

$$t = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}} \quad \text{Two means—matched pairs} \quad (\text{df} = n - 1)$$

Ch. 11: Goodness-of-Fit and Contingency Tables

$$\chi^2 = \sum \frac{(O - E)^2}{E} \quad \text{Goodness-of-fit} \quad (\text{df} = k - 1)$$

$$\chi^2 = \sum \frac{(O - E)^2}{E} \quad \text{Contingency table} \quad [\text{df} = (r - 1)(c - 1)]$$

where $E = \frac{(\text{row total})(\text{column total})}{(\text{grand total})}$

Ch. 10: Linear Correlation/Regression

$$\text{Correlation } r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}}$$

where $z_x = z \text{ score for } x$
 $z_y = z \text{ score for } y$

$$\text{or } r = \frac{\sum (z_x z_y)}{n - 1}$$

Slope: $b_1 = \frac{n\sum xy - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$

or $b_1 = r \frac{s_y}{s_x}$

y-Intercept:

$$b_0 = \bar{y} - b_1 \bar{x} \quad \text{or} \quad b_0 = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

$$\hat{y} = b_0 + b_1 x \quad \text{Estimated eq. of regression line}$$

$$r^2 = \frac{\text{explained variation}}{\text{total variation}}$$

$$s_e = \sqrt{\frac{\sum (y - \hat{y})^2}{n - 2}} \quad \text{or} \quad \sqrt{\frac{\sum y^2 - b_0 \sum y - b_1 \sum xy}{n - 2}}$$

$$\hat{y} - E < y < \hat{y} + E \quad \text{Prediction interval}$$

$$\text{where } E = t_{\alpha/2} s_e \sqrt{1 + \frac{1}{n} + \frac{n(x_0 - \bar{x})^2}{n(\sum x^2) - (\sum x)^2}}$$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)} \quad \text{Rank correlation}$$

$$\left(\text{critical value for } n > 30: \frac{\pm z}{\sqrt{n - 1}} \right)$$

Ch. 11: One-Way Analysis of Variance

Procedure for testing $H_0: \mu_1 = \mu_2 = \mu_3 = \dots$

1. Use software or calculator to obtain results.
2. Identify the P -value.
3. Form conclusion:

If $P\text{-value} \leq \alpha$, reject the null hypothesis of equal means.

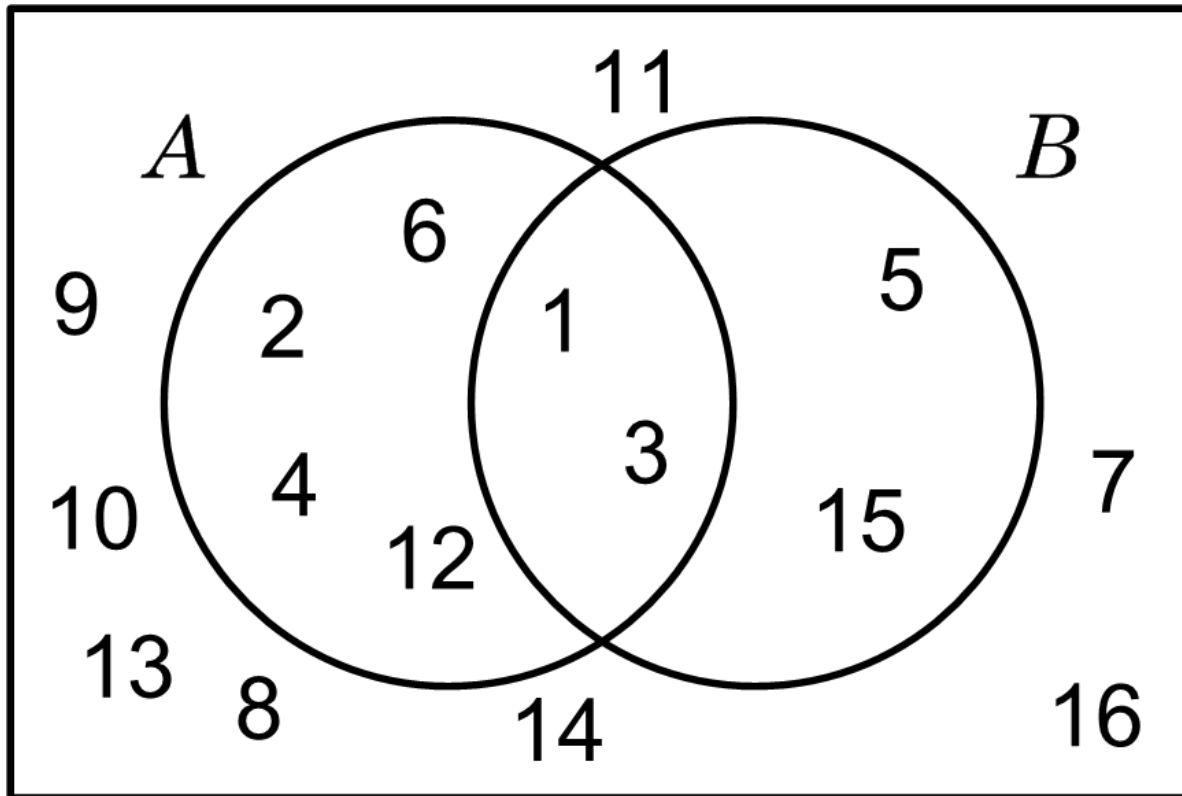
If $P\text{-value} > \alpha$, fail to reject the null hypothesis of equal means.

Definiciones

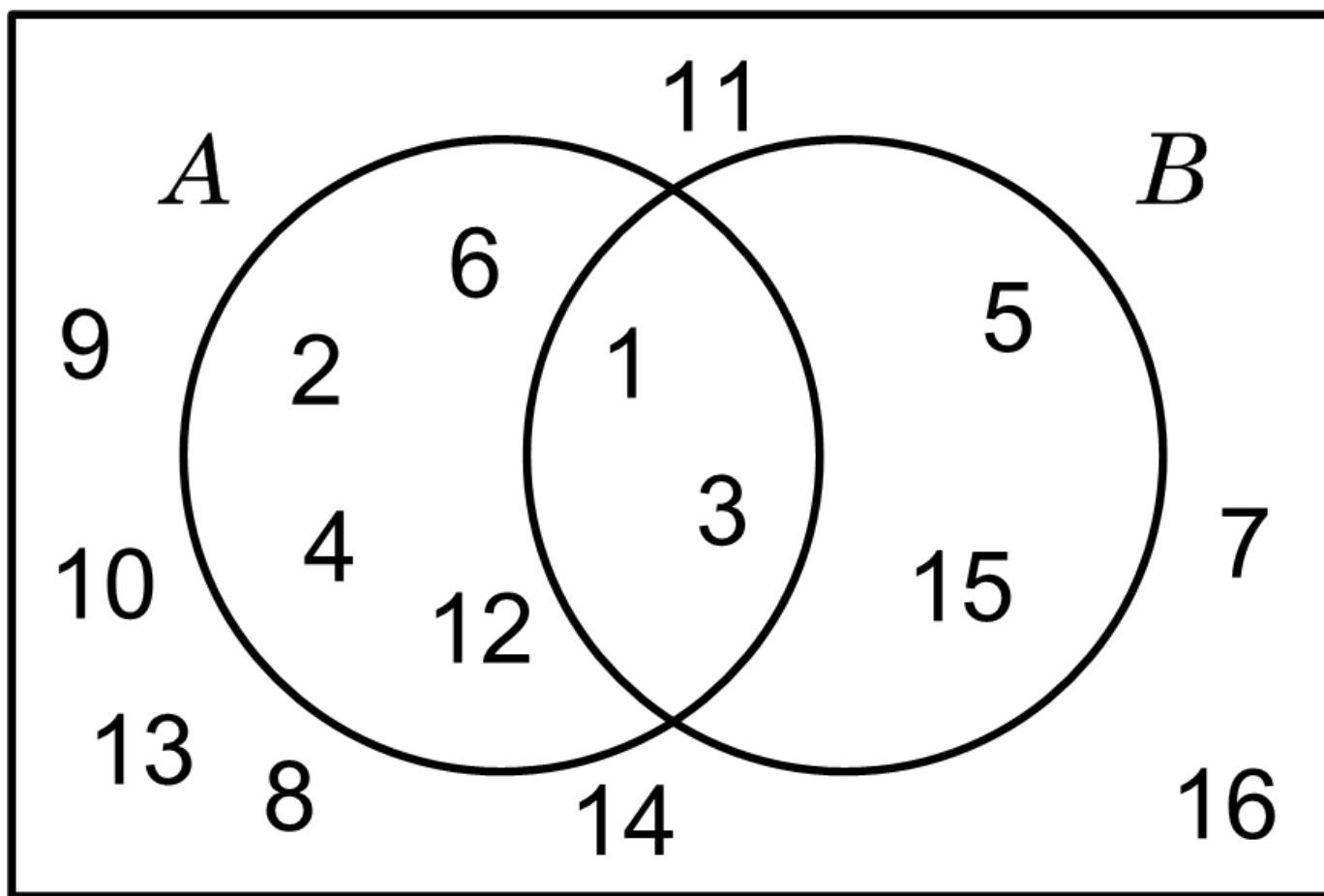
- “Rama de las matemáticas que estudia la recolección, análisis, interpretación y presentación de masas de información numérica.”
-Webster’s New Collegiate Dictionary -.
- “Es la rama del método científico que estudia los datos obtenidos por contar o medir propiedades de las poblaciones” -Stuart y Ord (1991)-.
- “Se ocupa esencialmente de procedimientos para analizar información, en especial aquella que en algún sentido vago tenga un carácter **aleatorio**” comentó: Rice (1995).
- “La ciencia de basar **inferencias** en datos observados y todo el problema de tomar decisiones frente a una incertidumbre” -Freund y Walpole (1987)-.

Teoría de conjuntos

- Notación: pertenencia, inclusión, vacío.
- Diagrama de Venn



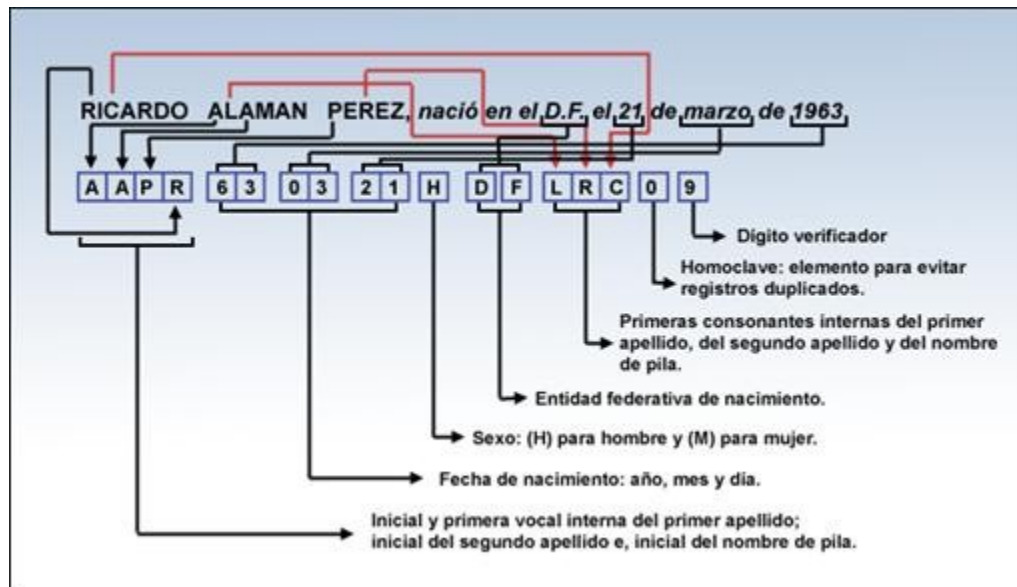
- Notación:
- Unión, intersección, diferencia, diferencia simétrica, complemento



Más definiciones

- Experimento: proceso por medio del cual se hace una observación.
- Evento simple: aquel que no se puede descomponer, cada evento simple corresponde a un sólo punto muestral.
- **Espacio muestral:** El conjunto formado por todos los posibles eventos muestrales.
- Espacio muestral discreto: Formado por un número finito o contable de puntos muestrales distintos.

- Ejercicio (experimento) M:
 - Llegan 5 pacientes a consulta.
 - Observemos su mes de nacimiento, registremos sólo las unidades y llámemosle “M” esta variable.
 - ¿Cuál es el espacio muestral?
 - Representar los eventos:
 - A= Observar un número impar.
 - $B = M < 3$
 - ¿A y B son mutuamente excluyentes? Dar un ejemplo de un evento
 - Dar un ejemplo de un evento simple



Probabilidad

- Sea “S” un espacio muestral asociado a un experimento. B es subconjunto de S. Asignamos un número a “B” llamado probabilidad de B, tal que:
 - Axioma 1: $P(A) \geq 0$
 - Axioma 2: $P(S) = 1$
 - Axioma 3: Si B1, B2, B3 son eventos mutuamente excluyentes por pares, entonces:
 - $P(B1 \cup B2 \cup B3) = P(B1) + P(B2) + P(B3)$
- Asignar probabilidades al ejercicio anterior
- Eventos equiprobables: dados, monedas, placas de coche.

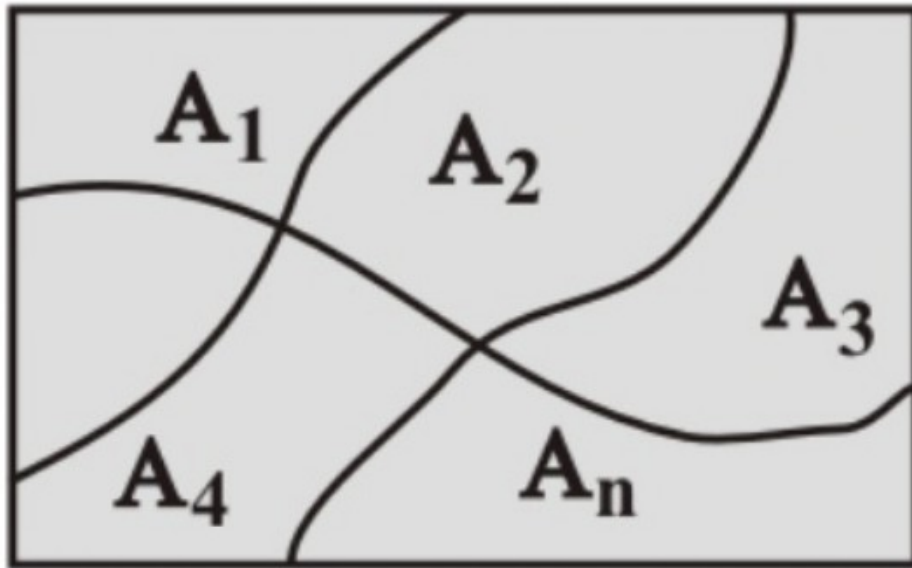
Probabilidad condicional

- La probabilidad condicional de A dado que B a ocurrido es dada por:

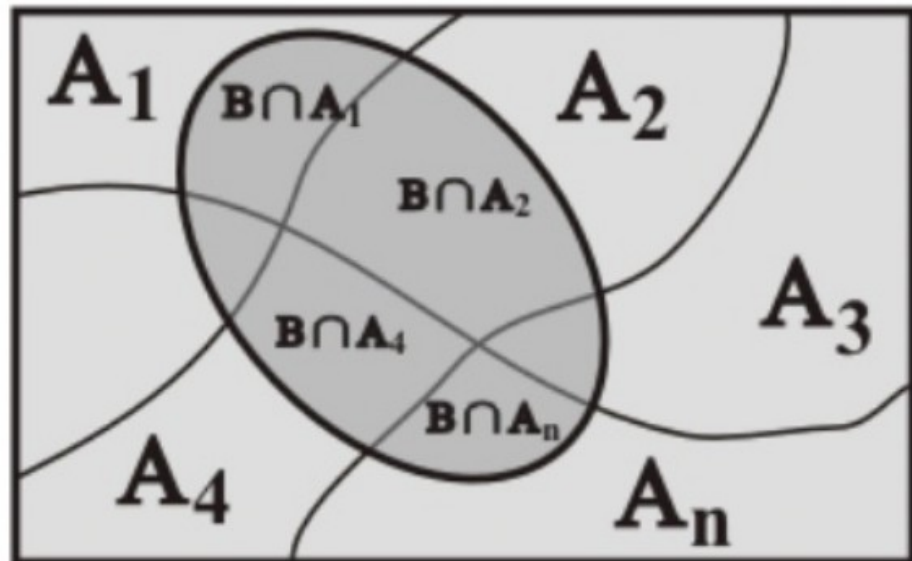
$$P(A | B) = \frac{P(A \cap B)}{P(B)}.$$

- A = M es > 6
- Considerar el experimento M , la información adicional (Los eventos):
 - “El paciente es Leo”.
 - Nació en Verano.
 - M es impar
 - “El paciente tiene menos de 10 años”

Regla de Bayes



$$B = (B \cap A_1) \cup (B \cap A_2) \cup \dots \cup (B \cap A_n)$$



$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \dots + P(B \cap A_n)$$
$$P(B) = P(B/A_1) \cdot P(A_1) + P(B/A_2) \cdot P(A_2) + \dots + P(B/A_n) \cdot P(A_n)$$

$$P(B) = \sum_{i=1}^n P(B/A_n) \cdot P(A_n)$$

Independencia (probabilidad)

Dos eventos son independientes si la probabilidad de que ocurran simultáneamente es igual al producto de sus probabilidades correspondientes.

$$P(A \cap B) = P(A)P(B)$$

- Condicional, independencia de eventos
- Dos leyes de probabilidad
- Regla de Bayes
- Variables aleatorias
- Muestreo aleatorio