



PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública, Facultad de Medicina, UBA

# TEST DE HIPÓTESIS CORRELACIÓN Y REGRESIÓN

## CLASE 9





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

- Test de hipótesis
- Intervalos de confianza





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

# TEST DE HIPÓTESIS



- Hipótesis: afirmación acerca de algún PARÁMETRO
  - Parámetro: poblacional
  - Estimador: muestral
- 2 tipos de hipótesis (en la mayoría de los casos y según el test):
  - Nula ( $H_0$ ): no hay diferencia
  - Alternativa ( $H_1$ ): hay diferencia





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

# ¿Qué decisión tomar?

- 1º determinar nivel de significación ( $\alpha$ ): 0,05 o 0,01
- 2º decidir:
  - Si  $p\text{-valor} < \alpha$ : la diferencia es estadísticamente significativa. Rechazo  $H_0$
  - Si el  $p\text{-valor} \geq \alpha$ : la diferencia NO es estadísticamente significativa. No se rechaza  $H_0$





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

# INTERPRETACIÓN

- Rechazo  $H_0$ : la diferencia es significativa
- No rechazo  $H_0$ : no significa que aceptamos  $H_0$ , decimos que no hay evidencia suficiente para decir que son distintos





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública, Facultad de Medicina, UBA

# PROCEDIMIENTO INFERENCIAL



- 1) Plantear las hipótesis estadísticas
- 2) Elegir un valor de nivel de significación (en general  $< 0,05$ ).
- 3) Calcular un estadístico (fórmula)
- 4) Calcular el p-valor asociado al estadístico
- 5) Comparar el valor de p-valor con el valor del nivel de significación para rechazar o no rechazar la hipótesis nula





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

# EJEMPLO

- Se compararon dos muestras de 100 alumnos de un colegio primario, un grupo cuyos padres tenían diabetes y el otro sin ese antecedente. Se determinó el nivel de glucemia en sangre, siendo el promedio de 90 mg/dl en el primer grupo y de 80 en el segundo. Se fija  $\alpha=0,05$ .
- $H_0=$
- $H_1=$
- Test t – p-valor: 0.06
- Conclusión?





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública, Facultad de Medicina, UBA

# TIPOS DE ERRORES

	Ho Verdadera	Ho Falsa
No rechazo Ho	OK	Error tipo II
Rechazo Ho	Error tipo I	OK

Definiciones:

Error tipo I: Rechazar  $H_0$  cuando es verdadera.

Error tipo II: No rechazar  $H_0$  cuando es falsa.

$\alpha$ : Probabilidad de error tipo I

$\beta$ : Probabilidad de error tipo II

Potencia:  $1 - \beta$ : probabilidad de rechazar  $H_0$  cuando es  $H_0$  es falsa







PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

- Error tipo I: se controla con el nivel de significación del test
- Error tipo II: se controla con el tamaño muestral
- Error tipo I y II están inversamente relacionados: cuando  $\alpha$  aumenta,  $\beta$  disminuye y viceversa.
- Fijando  $\alpha$ , se puede disminuir  $\beta$  aumentando el tamaño muestral





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

	N pequeño	N grande
Rechazo $H_0$	CONCLUYENTE El efecto es tan grande, que aún con un n pequeño pudo ser detectado	CUIDADO!!! Ver si la magnitud del efecto es clínicamente importante
No rechazo $H_0$	CUIDADO!!! Es posible que el efecto exista y no pueda ser demostrado	CONCLUYENTE El efecto existente es despreciable





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

# INTERVALO DE CONFIANZA

- IC es un rango de valores dentro del cual se cree que se encuentra el PARÁMETRO, con una confianza del 95 % (o 99% o lo que definamos).
- Ejemplo de cálculo IC para media:

$$\text{Error Estándar de la media (ESM)} = \frac{\sigma}{\sqrt{n}}$$

$$\text{Error Estándar de la proporción (ESP)} = \sqrt{\frac{pq}{n}}$$

$$\text{IC}_{95\%} = 1,96 \pm \text{ESM} = 1,96 \pm \frac{s}{\sqrt{n}}$$





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA



# TEST DE HIPÓTESIS





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

## Muestras independientes

Variable dependiente	Variable independiente	Prueba estadística	Ejemplo
Cuantitativa	Cuali: 2 grupos	T de Student (P) U de Mann-Whitney (NP)	Nivel de glucemia entre niños con padres con DBT y sin DBT
	Cuali: más de 2 grupos	ANOVA (P) Kruskal-Wallis (NP)	Nivel de colesterol en 3 grupos con diferente dieta
	Cuanti	Correlación de Pearson (P) Spearman (NP)	Relación entre Tensión arterial y edad
Cualitativa	Cuali	Chi cuadrado Test exacto de Fisher	Asociación entre mate y gastritis

P: Paramétrico  
NP: No Paramétrico



PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública, Facultad de Medicina, UBA

## Muestras dependientes



Variable dependiente	Variable independiente	Prueba estadística	Ejemplo
Cuantitativa	Cuali: 2 grupos	Test de Wilcoxon (NP) Test del signo (NP)	Analizar la TAS antes y después de tratamiento
	Cuali: más de 2 grupos	ANOVA de medidas repetidas (P) Friedman (NP)	Comparar TAS al inicio, 3 meses y 6 meses de tto
Cualitativa	Cuali: 2 grupos	Test de McNemar	
	Cuali: más de 2 grupos	Test de Cochran	

P: Paramétrico  
NP: No Paramétrico





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

# CORRELACIÓN



 /labinnovacion

  /lab\_fmed



PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

- Evalúa si dos variables están asociadas y estima el grado de asociación
- Variables: numéricas y ordinales
- Evaluación por medio de:
  - Gráfico de dispersión (Scatter plot)
  - Coeficientes de correlación

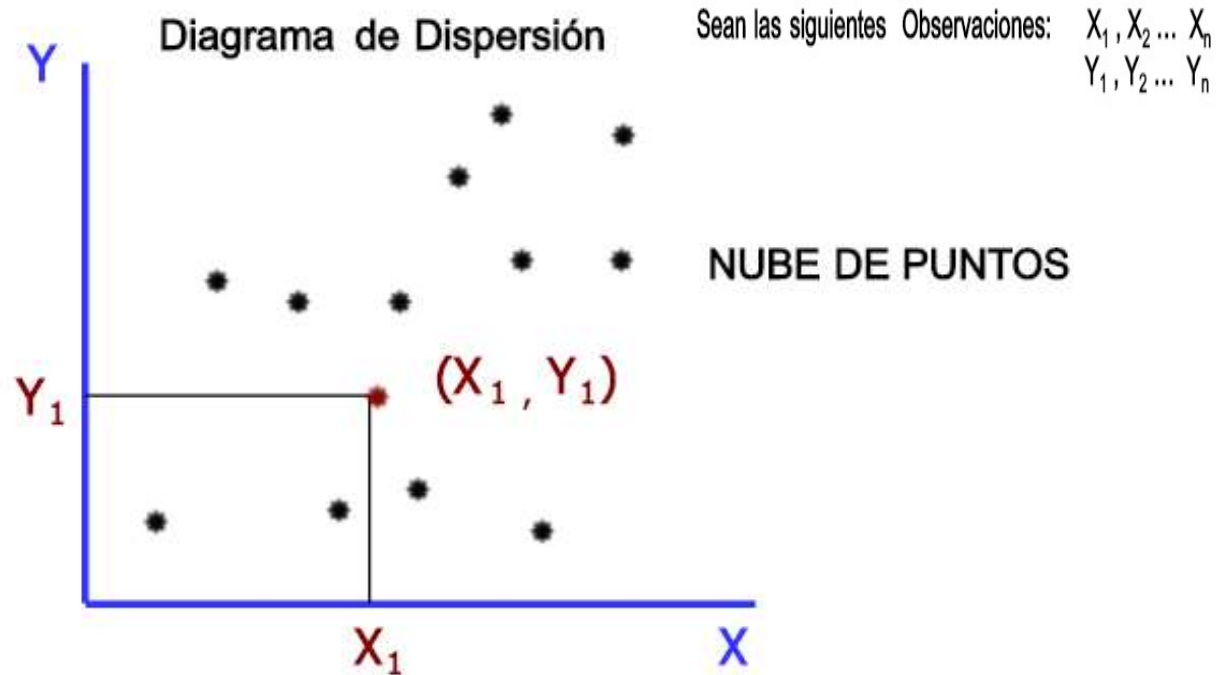






PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

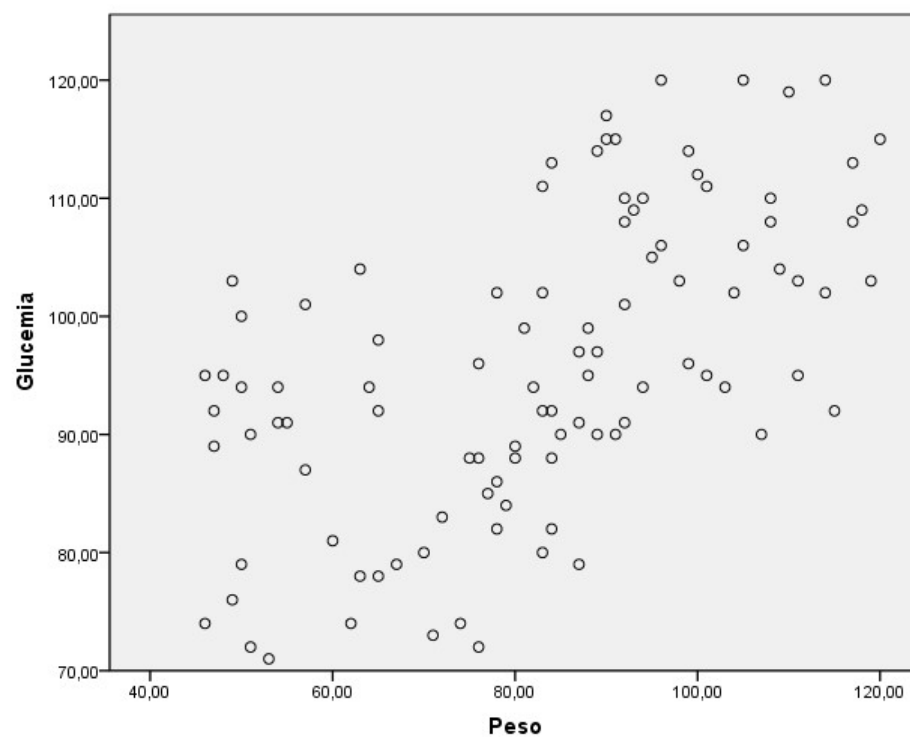
# Scatter plot o Gráfico de dispersión





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

# Gráfico de dispersión





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

# Coeficientes de correlación ( $\rho$ )

- Miden la fuerza de asociación **lineal** entre las variables
- Coeficiente de Pearson (r) es un test paramétrico y Coeficiente de Spearman (rho) es No paramétrico
- Va de -1 a 1
- Hipótesis:
  - $H_0$ : No hay asociación lineal entre x e y ( $\rho = 0$ )
  - $H_1$ : hay asociación lineal entre x e y ( $\rho \neq 0$ )





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública, Facultad de Medicina, UBA

## Interpretación:

- $r$  o  $\rho = 0$  ----> No hay asociación lineal
- $r$  o  $\rho = 1$  ----> Hay asociación perfecta positiva
- $r$  o  $\rho = -1$  ----> Hay asociación perfecta negativa

## Supuestos:

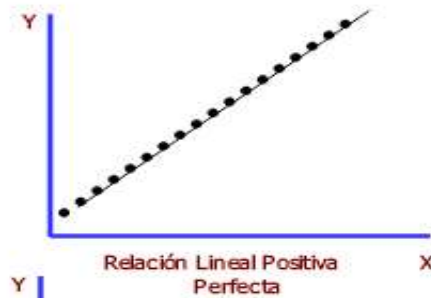
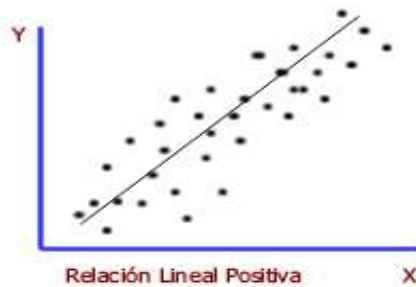
- Los pares de observaciones,  $(x_1, y_1) \dots (x_n, y_n)$ , son independientes entre sí
- Están igualmente distribuidos
- Cada muestra ( $x$  e  $y$ ) tienen distribución normal → Pearson, sino tiene distribución normal o son de escala ordinal → Spearman





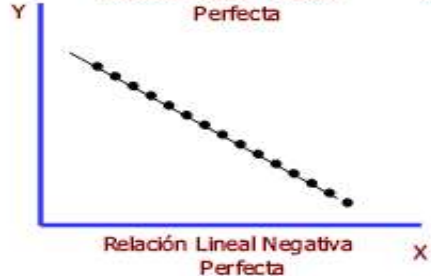
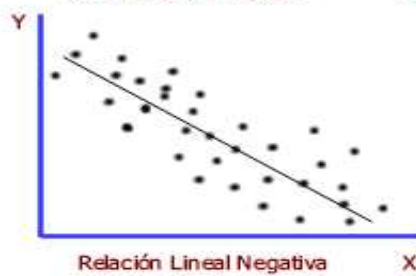
PROGRAMA DE INNOVACIÓN TECNOLÓGICA EN SALUD  
Instituto de Salud Pública Facultad de Medicina

$$0 > r < 1$$

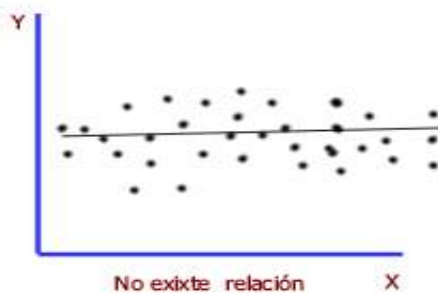


$$r = 1$$

$$-1 > r > 0$$



$$r = -1$$



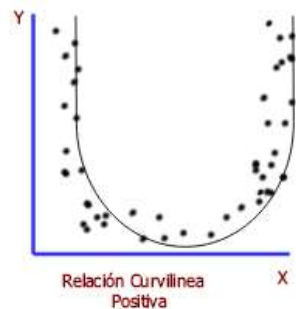
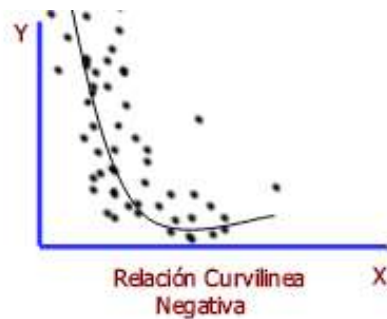
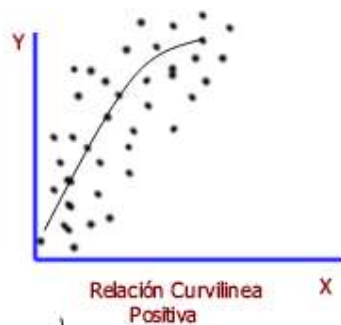
$$r \approx 0$$





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

# Otras asociaciones NO lineales





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

# REGRESIÓN LINEAL Y LOGÍSTICA



 /labinnovacion

  /lab\_fmed



PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

# Objetivo de la regresión



- **Explicar** la relación entre una variable dependiente y una o varias variables independientes.
- **Predecir** el resultado de una variable dependiente, con base en los valores de una o más variables independientes, explicativas o predictoras.







PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

# REGRESIÓN LINEAL



 /labinnovacion

  /lab\_fmed



PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

# MODELO GENERAL



- Modelo General:

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

Donde:

- $Y$  = variable dependiente o resultado
- $\beta_0$  = ordenada al origen o intercepto
- $\beta_1$  = pendiente de la recta
- $X_1$  = variable independiente o predictora
- $\varepsilon$  = error → variable desconocida





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública, Facultad de Medicina, UBA

# MODELO AJUSTADO

- Modelo ajustado:

$$\hat{Y} = b_0 + b_1 X$$

- Donde:
- $\hat{Y}$  = glucemia
- $b_0$  = nivel de glucemia cuando  $x = 0$
- $b_1$  = cambio de glucemia cuando  $x$  aumenta 1 unidad
- $X$  = peso





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

# SUPUESTOS ESTADÍSTICOS

- Los errores tienen una media de cero
- Los errores tienen todos la misma varianza (homocedasticidad)
- Los errores tienen distribución normal
- Los errores son independientes entre sí





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

# ANÁLISIS ESTADÍSTICO



- **1º) ANOVA (F)**
  - Permiten saber hay una regresión lineal significativa.
  - Regresión lineal significativa: existe una relación lineal entre una variable dependiente (Y) y una variable independiente (X)
  - $p < \alpha$ : hay regresión lineal significativa.
  - $p \geq \alpha$ : no hay regresión lineal





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

- **2º) Coeficiente de Determinación ( $R^2$ )**
  - Mide la proporción de la variación en Y que es explicada por la relación con X. En otras palabras, mide “cuán bueno es el ajuste de la ecuación de regresión”.
  - Va de 0 a 1
  - Un  $R^2$  cercano a 1 indica que la recta obtenida permite una buena predicción de valores de Y a partir de valores de X.
  - Un  $R^2$  cercano a 0 indica que la variable independiente (X) no predice los valores de la variable dependiente (Y) (mediante una ecuación lineal).





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

- 3º) Test t para coeficientes  $b_0$  y  $b_1$

- Hipótesis:

- $H_0: b=0$

- $H_1: b \neq 0$

- $p < \alpha$ : rechazo  $H_0$

- $p \geq \alpha$ : no rechazo  $H_0$

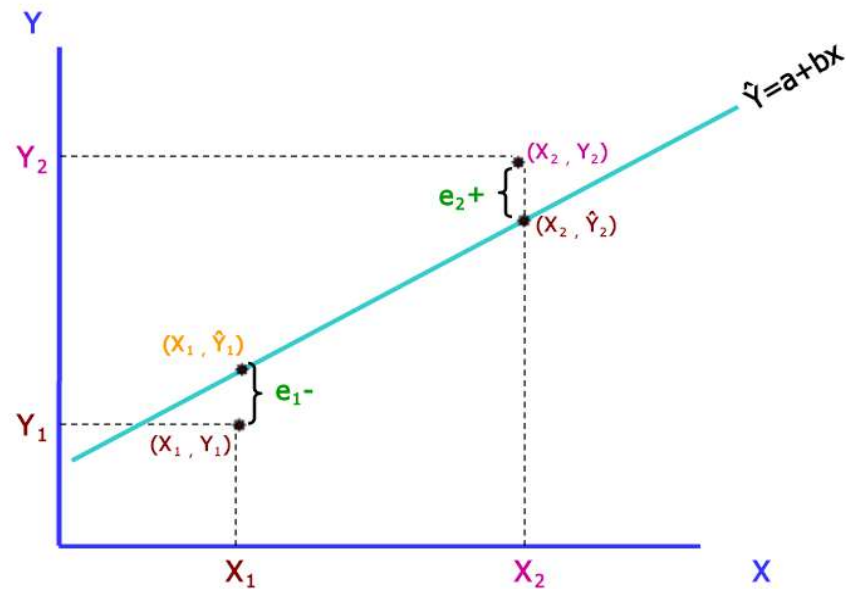




PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

# VERIFICACIÓN DE SUPUESTOS

- Como los errores son desconocidos, se estiman por los residuos







PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

# Verificación de supuestos con residuos

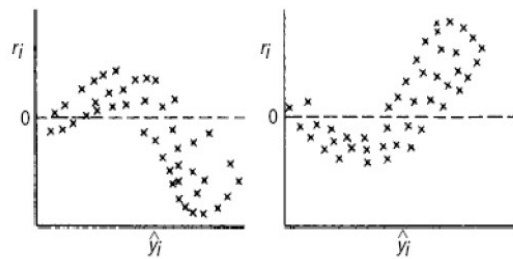
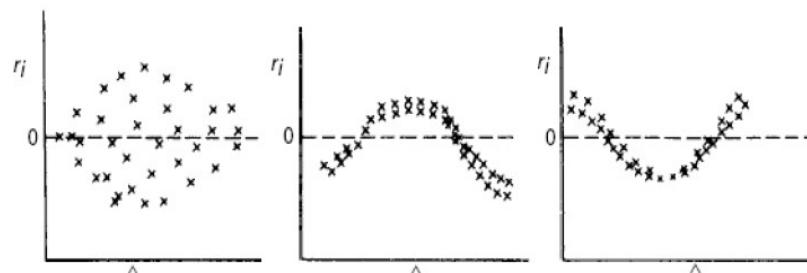
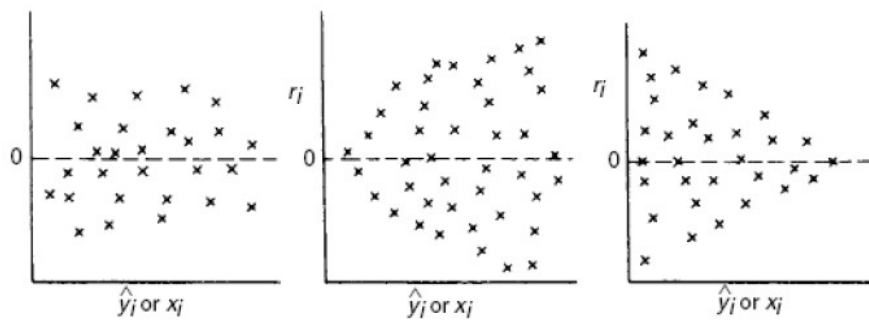


- Normalidad:
  - Gráficos (histograma, qq-plot, box plot)
  - Test de normalidad (Shapiro, Kolmogorov)
- Homocedasticidad y linealidad:
  - Gráficos de dispersión de residuos estandarizados (eje y) versus pronosticados (eje x)





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA



# REGRESIÓN LOGÍSTICA



 /labinnovacion

  /lab\_fmed



PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

# Repaso de algunos conceptos

- Odds: cociente que expresa cuanto más probable es que se produzca un evento frente a que no se produzca el evento en cuestión
- La razón de odds (OR) de una condición en los expuestos comparados con el odds de la condición en los no expuestos

$$OR = \frac{\text{Odds de enfermedad en expuestos}}{\text{Odds de enfermedad en No expuestos}}$$





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública, Facultad de Medicina, UBA

# ODDS RATIO

	Enfermo	No Enfermo	Total
Expuesto	a	b	a + b
No Expuesto	c	d	c + d
Total	a + c	b + d	a + b + c + d

Odds en expuestos =  $a/b$

Odds en NO expuestos =  $c/d$

$$OR = \frac{a/b}{c/d} = \frac{a \times d}{c \times b}$$





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

- Odds =  $\frac{\text{proporción}}{1 - \text{proporción}}$
- Proporción =  $\frac{\text{odds}}{1 + \text{odds}}$

Odds y proporción → son medidas de frecuencias

OR o Razón de odds → medida de asociación





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

# MODELO GENERAL

3 formas de expresar el Modelo General:

$$\text{logit} ( p(x) ) = \alpha + \beta x$$

$$\ln (\text{odds}(x)) = \ln \left( \frac{p(x)}{1-p(x)} \right) = \alpha + \beta \cdot x$$

$$E(Y / X = x) = p(x) = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}$$

Donde:

- $\text{Logit} (p(x)) \rightarrow Y \rightarrow$  variable dependiente dicotómica
- $X \rightarrow$  variable predictora
- $\beta \rightarrow$  OR para  $Y=1$  (tiene el evento) dado  $X=x$
- $\alpha \rightarrow$  medida de frecuencia para  $Y=1$  dado  $X=0$





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA



- El modelo establece que la transformación logit de la proporción de la variable dependiente (Y) se relaciona linealmente con la variable regresora o predictora o independiente (X).
- En este modelo la variable Y condicional al valor de la variable  $X = x$  tiene distribución binomial
- El parámetro de mayor interés es la pendiente  $\beta$







PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

# ANÁLISIS ESTADÍSTICO

- Tabla de codificación interna de variable dependiente: ver si la codificación interna corresponde con nuestra codificación
- $-2 \log$  verosimilitud: sirve para comparar modelos, mientras menor sea mayor es la verosimilitud, por lo tanto el agregar la variable mejora el modelo





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

- Test de máxima verosimilitud (test omnibus):
  - $H_0$ : todos los  $\beta = 0$  vs
  - $H_1$ : algún  $\beta \neq 0$
- Test de wald:
  - $H_0$ : cada coeficiente aislado es igual a 0
  - $H_1$ : cada coeficiente aislado es distinto de 0





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

- Tabla de clasificación:
  - Valora que tan bien clasifica el modelo a los valores observados según los pronosticados por el modelo.
- Test de Hosmer-Lemeshow:
  - Es un test de bondad de ajuste que indica que tan bien ajusta el modelo
  - Compara frecuencias observadas y esperadas
  - $H_0$ : observadas = esperadas
  - Se busca **NO** rechazar  $H_0$





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

# Interpretación de coeficientes



- Variable independiente continua:
  - Es el cambio en el  $\text{logit}(p)$  cuando  $x$  aumenta 1 unidad. O sea es el aumento o disminución de las chances de  $Y=1$  (tener el evento) cuando  $x$  aumenta 1 unidad.
- Variable independiente categórica:
  - Es el cambio en el  $\text{logit}(p)$  entre las categorías de  $x=1$  y  $x=0$  (tener y no tener el factor de exposición).





PROGRAMA DE INNOVACIÓN  
TECNOLÓGICA EN SALUD PÚBLICA  
Instituto de Salud Pública Facultad de Medicina, UBA

# Bibliografía

- Material docente de la Unidad de Bioestadística Clínica. Hospital Universitario Ramón y Cajal.  
Disponible en: <http://www.hrc.es/bioest/Mdocente.html>
- Correlación y modelos de regresión lineal.  
Disponible en: <http://www.hrc.es/bioest/Reglin1.html>
- Modelos de regresión logística.  
Disponible en: <http://www.hrc.es/bioest/Reglog1.html>

