Introducción a la prueba de hipótesis

Sesión 6

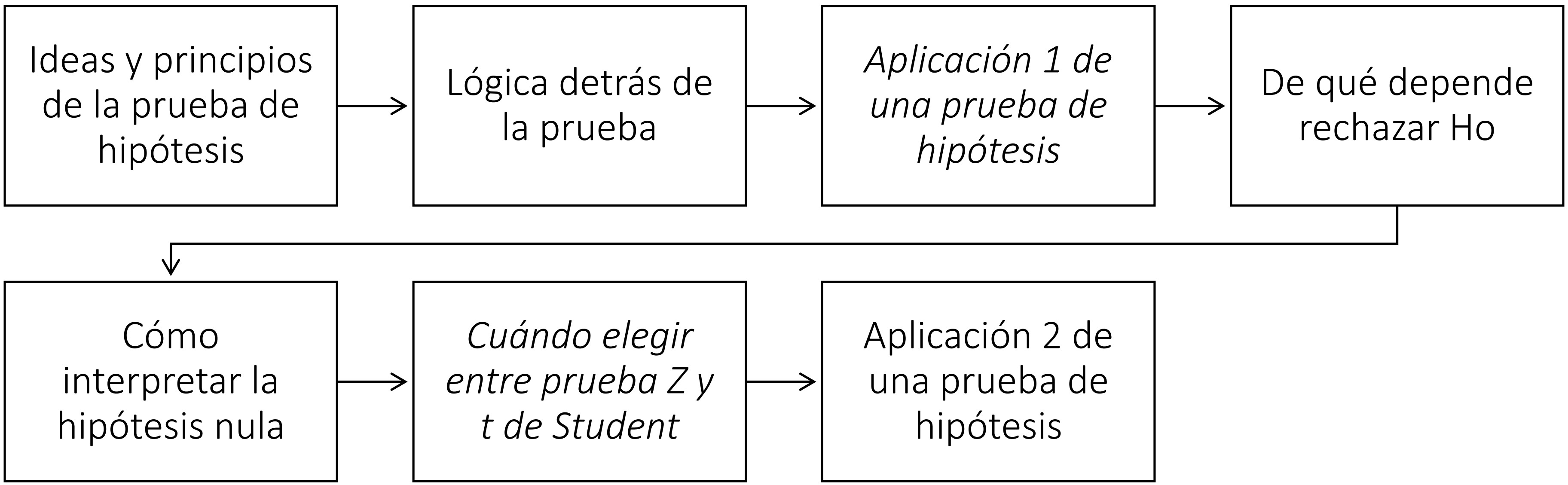
Carlos Vilalta, PhD
Contacto: carlos.vilalta@cide.edu

Lo que han aprendido hasta hoy

1. Tipos de variables y niveles de medición
2. *Aplicación 1 de la ED en las PP*
3. Notación
4. Métodos tabulares y gráficos
5. Medidas de tendencia central y localización
6. Paradoja de Simpson y tablas de contingencia
7. Diagramas de dispersión y outliers
8. *Aplicación 2 en PP*
9. Nociones de correlación y covarianza
10. Medidas de dispersión
11. *Aplicación 3 en PP*
12. Medidas de forma
13. *Aplicación 4 en PP*
14. *Aplicación 1 de la Probabilidad en las PP*
15. Resumen del procedimiento de la aplicación
16. Los métodos de las probabilidades
17. Combinaciones y permutaciones
18. *Aplicación 2 de Comb. y Perm. en las PP*
19. Lo que la probabilidad realmente es
20. *Aplicación 1 de valor esperado y varianza en las PP*
21. Conceptos básicos de distribuciones de probabilidad
22. Distribución Binomial
23. *Aplicación 2 de la distribución Binomial en las PP*
24. Distribución de Poisson
25. *Aplicación 3 de la distribución de Poisson en las PP*
26. Repaso de lo que hemos hecho hasta el momento
27. Distribución Normal Estándar
28. *Aplicación 4 de la distribución Normal Estándar en las PP*
29. Nociones de muestreo probabilístico
30. Estimación puntual
31. Error estándar de la media aritmética
32. *Aplicación 1 para una estimación puntual y de intervalos en las PP*
33. Error estándar de una proporción
34. *Aplicación 2 para una estimación puntual y de intervalos en las PP*
35. Requisitos de los estimadores

Y con Excel…

Lo que vamos a ver a continuación



Ideas y principios

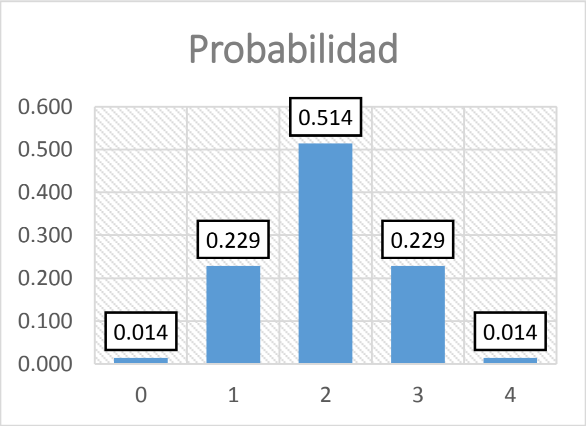
- Prueba de hipótesis y toma de decisiones son sinónimos
- Antes de entender cómo probamos hipótesis en PP hay que entender porqué probamos hipótesis (o tomamos decisiones en PP)
- Criterio principal en la prueba: maximizar la probabilidad de estar en lo correcto (reducir la probabilidad de estar en lo equivocado)
- La mayor parte de las pruebas/decisiones son sobre una igualdad en la Ho (prueba de dos colas)

$H_0\colon \mu \geq \mu_0$ $H_a\colon \mu < \mu_0$	$H_0\colon \mu \leq \mu_0$ $H_a\colon \mu > \mu_0$	$\boxed{\begin{array}{l} H_0\colon \mu = \mu_0 \\ H_a\colon \mu \neq \mu_0 \end{array}}$
$H_0\colon p \geq p_0$ $H_a\colon p < p_0$	$H_0\colon p \leq p_0$ $H_a\colon p > p_0$	$\boxed{\begin{array}{l} H_0\colon p = p_0 \\ H_a\colon p \neq p_0 \end{array}}$

Lógica de la prueba

- Fisher (1935): prueba exacta de Fisher
  - Una mujer argumentó que puede saber si en su té se sirvió leche antes o después de poner el té
  - Se realizó un experimento con 8 tazas y la mujer adivina en todos los casos la respuesta correcta; en 4 ocasiones se sirvió la leche antes del té
  - La pregunta simple era: ¿tiene esta persona una habilidad especial?
  - La pregunta científica era: ¿cuál es la probabilidad de obtener este resultado por simple azar?
    - Si es poco probable que esto se deba al simple azar, debemos creer su argumento
    - Si es algo o muy probable que esto se deba al simple azar, debemos rechazar su argumento
  - Esta es la pregunta y la decisión detrás de toda inferencia estadística
- ¿Cuál es la probabilidad de acertar en cada caso? $Combinaciones = \frac{8!}{4!(8-4)!} = 70$

Evento	Probabilidad	%
0	0.014	1.4%
1	0.229	22.9%
2	0.514	51.4%
3	0.229	22.9%
4	0.014	1.4%



Ya que la probabilidad de adivinar correctamente y por simple azar las 4 veces entre 8 es menor de 1.4%... creemos en su argumento (con un nivel de confianza del 98.6%)

Utilidad en Políticas Públicas

- Vilalta, C. 2016. Assessing the role of context on the relationship between adolescent marijuana use and property crimes in Mexico. *Substance Use & Misuse*. Status: Revise-Resubmit.

- Argumento: alguien dice que la proporción de estudiantes de bachilleratos públicos que en 2009 consumían marihuana era del 12.0%
- Pregunta: ¿Qué tan probable es esta estimación?
  - Población: estudiantes en bachilleratos públicos
  - Fuente: Encuesta de Exclusión, Intolerancia y Violencia de 2007 y 2009 (SEP)
  - Evidencia: conductas en encuesta auto-aplicada (auto-reporte)
    - Puede existir un sub-reporte de estas conductas
  - Resultados:
    - Que reportan uso de marihuana: 9.6%
    - n = 9,136

- Probemos la hipótesis de una diferencia de proporciones

Utilidad de políticas públicas

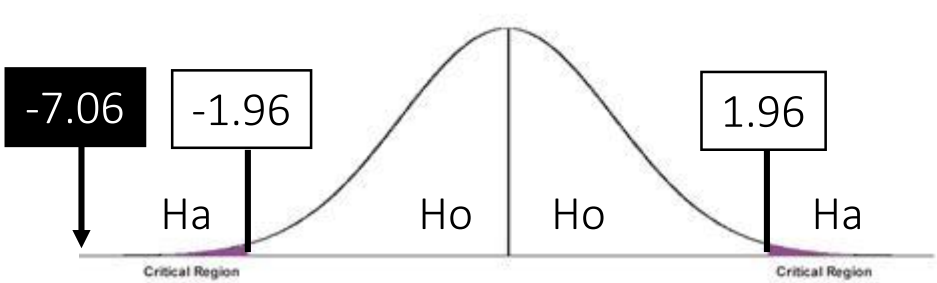
- Formulación de hipótesis:
  - Ho: $\bar{p} = p_0$; Ho: 0.096 = 0.120
  - Ha: $\bar{p} \neq p_0$; Ha: 0.096 ≠ 0.120

Aplicamos la prueba Z para diferencias de proporciones

$$\sigma_{\bar{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} \qquad z = \frac{\bar{p} - p_0}{\sigma_{\bar{p}}}$$

$$z = \frac{0.096 - 0.120}{0.003} = -7.059$$

- Con un nivel de confianza del 95%, el valor crítico de Z = 1.96 y -1.96
- Ya que -7.059 < 1.96 → rechazamos Ho



Nota: aquí usamos el procedimiento del valor crítico para rechazar Ho

https://www.medcalc.org/calc/test_one_proportion.php

Utilidad en Políticas Públicas

- Igualmente podemos probar la hipótesis y rechazar Ho sobre la base de los intervalos de confianza ($Z_{0.05/2}$=1.96)

$$\bar{p} - 1.96\sqrt{\frac{p(1-p)}{n}} \leq \pi \leq \bar{p} + 1.96\sqrt{\frac{p(1-p)}{n}}$$

$$0.096 - 1.96\sqrt{\frac{0.096(1-0.096)}{9{,}136}} \leq \pi \leq 0.096 + 1.96\sqrt{\frac{0.096(1-0.096)}{9{,}136}}$$

$$0.090 \leq \pi \leq 10.204$$

- Ya que 0.120 cae fuera del $IC_{0.95}$, rechazamos Ho con un nivel de confianza del 95%
    - O, dicho de otra manera, hay un 5% de posibilidades de que la proporción real sea menor de 9% o mayor de 10.2%, por lo que nos habríamos equivocado al rechazar Ho en ese caso

¿De qué depende rechazar Ho?

- Del nivel de confianza de la prueba y del margen de error (ME) del estimador puntual (p o M)
  - ME depende del tamaño de la muestra (n) y la variabilidad (s)
- Lo que sucede en este caso si modificamos el margen de error por conducto del tamaño de la muestra (n):
  - Con n = 913, el $IC_{0.95}$ estaría entre 7.69% y 11.51%; sigo rechazando Ho
  - Con n = 91, el $IC_{0.95}$ estaría entre 3.55% y 15.65%; no rechazo Ho
  - Regla empírica: a mayor n → más fácil rechazar Ho

La prueba de hipótesis nula

- Es un procedimiento estadístico de toma de decisiones (Neyman-Pearson, 1938)
- Se formula tanto una hipótesis nula como una alternativa
  - Ho: Hipótesis nula: lo que se busca rechazar (posición neutral)
  - Ha: Hipótesis alternativa: lo que se busca comprobar
- Hay 4 posibilidades en una toma de decisiones:

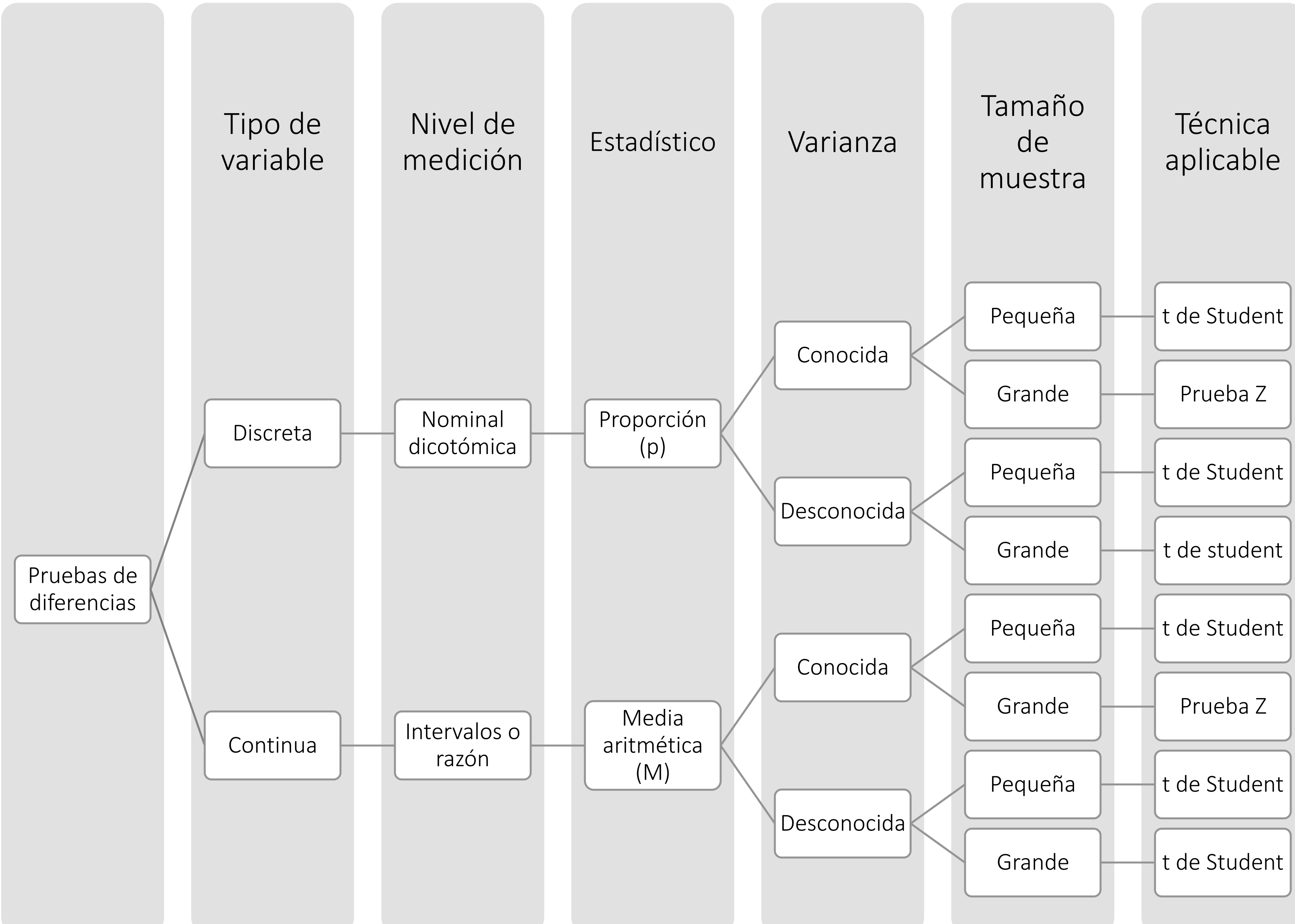
		Population Condition	
		H_0 True	**H_a True**
Conclusion	**Accept H_0**	Correct Conclusion	Type II Error
	Reject H_0	Type I Error	Correct Conclusion

- El error del tipo I = nivel de significancia estadística
  - Rechazar Ho cuando es verdadera (falso positivo)
  - Se busca que sea menor del 5% ($p < 0.05$: evidencia en contra de Ho)
  - Si $p < 0.05$ se rechaza Ho; si $p \geq 0.05$ no se rechaza Ho

¿Hay un tipo de error peor que otro?

- Los errores posibles son los siguientes:
  - Error tipo I: rechazar Ho cuando es verdadera
    - Concluir que $\bar{p} \neq p_0$ cuando la verdad es que $\bar{p} = p_0$
    - Decir que hay una correlación/diferencia cuando no la hay
    - Decir que alguien SI es culpable cuando NO lo es
    - El que afirma tiene la carga de la prueba… La presunción de inocencia es una Ho. La inocencia se presume, la culpabilidad hay que demostrarla, es decir, hay que producir evidencia para rechazar Ho
    - Falso positivo
    - La tasa de errores tipo I es la significancia estadística (p)
  - Error tipo II: no rechazar (aceptar) Ho cuando es falsa
    - Concluir que $\bar{p} = p_0$ cuando la verdad es que $\bar{p} \neq p_0$
    - Decir que NO hay un correlación/diferencia cuando SI la hay
    - Decir que alguien NO es culpable cuando SI lo es
    - Falso negativo
    - El poder de la prueba es la capacidad de rechazar Ho cuando es falsa o posibilidad de cometer el error tipo II

Cómo elegir la técnica para la prueba



Similitudes y diferencias entre Z y t

- La distribución de probabilidades es diferente en cada una
  - Z es una prueba paramétrica: Distribución normal estándar con M = 0 y s = 1
  - t es una prueba NO paramétrica: Distribución t de Student que se modifica con diferentes grados de libertad en cada prueba (g.l.= n - 1)
- Cuando n = 120, los valores críticos de Z y t son iguales
  - Z y t toman los valores de 1.64, 1.96, 2.58 en niveles de confianza del 90%, 95% y 99% respectivamente

Utilidad en Políticas Públicas

- Vilalta, C. y Fondevila, G. 2014. Perfiles Criminales II: Teorías, correlativos y políticas preventivas. México: Editorial CIDE.

- Se dice que los secuestradores tienen las edades mayores entre toda la población criminal. Pregunta: ¿son los reclusos por el delito de secuestro los que poseen una edad media (al momento de la detención) significativamente diferente a la población de la CDMX?
  - Fuente: Encuestas de Población en Reclusión (2002, 2005 y 2009). CIDE.
  - Reclusos por secuestro: M = 30.6 / s = 8.6 / n = 223
  - Población general en CDMX: μ = 33.0 años de edad
  - Prueba sobre una diferencia de medias aritméticas
- Formulación de hipótesis:
  - Ho: $\mathrm{M} = \mu_0$; Ho: 30.6 = 33.0
  - Ha: $\mathrm{M} \neq \mu_0$; Ha: 30.6 ≠ 33.0

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Varianza conocida y muestras grandes

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Varianza desconocida y/o muestras pequeñas

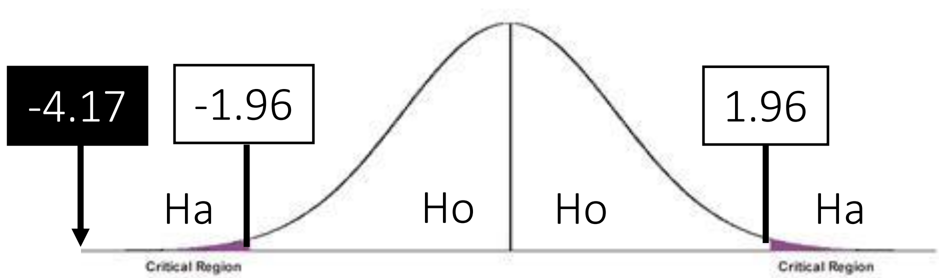
Utilidad en Políticas Públicas

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

- Resultado:

$$z = \frac{30.6 - 33.0}{8.6/\sqrt{223}} = \frac{30.6 - 33.0}{0.58} = -4.17$$

  - Conclusión: rechazo Ho ya que -4.17 < -1.96



- ¿Qué problemas habría con esta prueba?
- ¿Cómo refutarían esta conclusión o resultado?

Felicidades
Llegaste a la última clase
Pero…

… falta 1 tarea y 1 examen

Tarea

- Siguiente clase: repaso del curso
- Anderson et al.: Resolver ejercicios (entregar a mano y con procedimientos):
  - Pruebas de hipótesis (Cap. 9): ejercicios 11, 14, 24, 26 y 35
- Entrega: jueves en el laboratorio