Clase 2: Métodos Experimentales

Santiago Olivella

July 18, 2017

Profesor Asistente de la Universidad de Carolina de Norte, Chapel Hill

Clase 2

- Mecanismos de Asignación Aleatoria
- Evaluación de hipótesis
- Límites de Confianza
- Anílisis de poder

Mecanismos de asignación

- Un **mecanismo de asignación** es una distribución de probabilidad sobre el conjunto $\mathcal{D}$.
- Tres características:
    - No confundido: $\Pr(\mathbf{D}|\mathbf{X}, \mathbf{Y}(1), \mathbf{Y}(0)) = \Pr(\mathbf{D}|\mathbf{X})$
    - Individualístico: $\Pr(D_i|\mathbf{X}) = \Pr(D_i|\mathbf{x}_i)$
    - Probabilístico: $0 < \Pr(D_i|\mathbf{x}_i) < 1$
- Podemos clasificar los distintos tipos de experimentos por el tipo de mecanismo de asignación del tratamiento.

Tipos de Experimentos Aleatorizados

- Experimento de ensayo Bernoulli (¿probabilístico?):

$$\Pr(\mathbf{D}|\mathbf{X}) = \prod_{i=1}^{N} e(\mathsf{x}_i)^{D_i}\left(1 - e(\mathsf{x}_i)\right)^{1-W_i}$$

- Experimento completamente aleatorizado:

$$\Pr(\mathbf{D}|\mathbf{X}) = \begin{cases} \binom{N}{N_t}^{-1} & \text{si } \sum_i W_i = N_t \\ 0 & \text{de lo contrario} \end{cases}$$

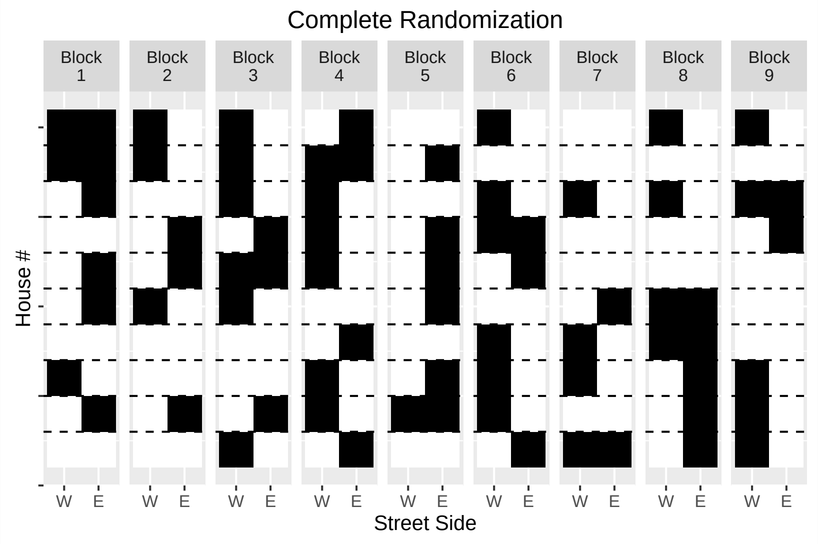
- Experimento estratificado (para bloques $B(i) \in \{1, \ldots, J\}$; pares igualados es caso especial):

$$\Pr(\mathbf{D}|\mathbf{X}) = \begin{cases} \binom{N(j)}{N_t(j)}^{-1} & \text{si } \sum_{i:B(i)=j}^{N} W_i = N_t(j) \ \forall j \\ 0 & \text{de lo contrario} \end{cases}$$

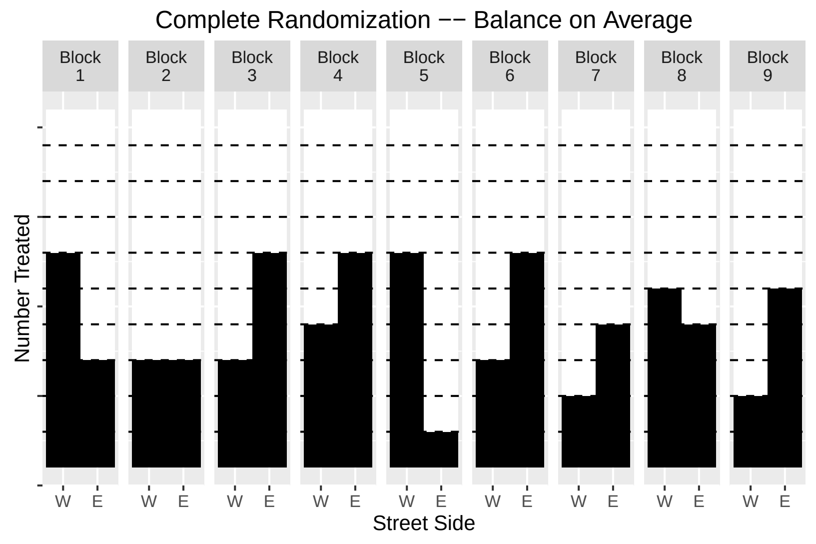
Tipos de Experimentos Aleatorizados

- La atribución causal en promedio es hecha posible por la aleatorización: los grupos de tratamiento son iguales, en esperanza, en todos los aspectos relevantes.
- Para una instancia: posibilidad del "bad draw" (falta de equilibrio en distribuciones de covariables para un experimento concreto).
- Bloques aseguran equilibrio en instancia observada, y mejoran precisión de estimación
- Especialmente útil cuando el número de uniades es pequeño.
- Mantra: **Estratifique lo que pueda, aleatorice lo demás** (Box et al. 1978).

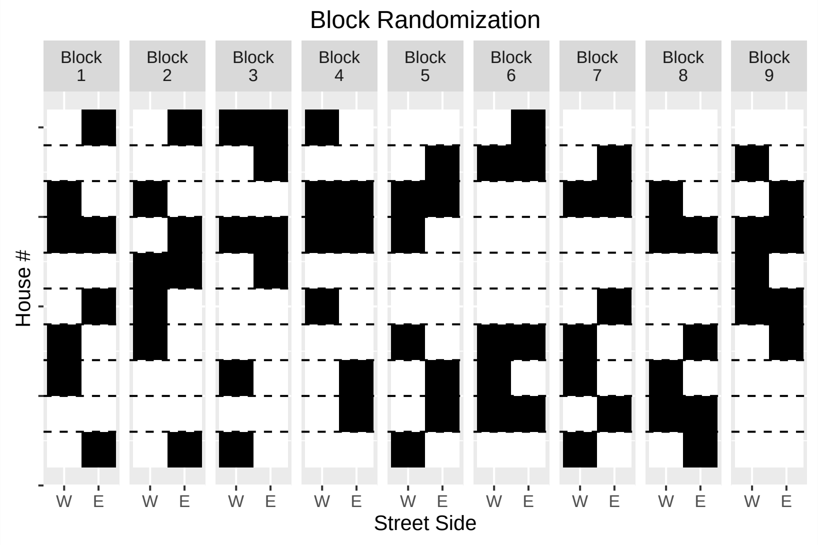
Ejemplo – Asignación GOTV (fuente: Erin Hartman)



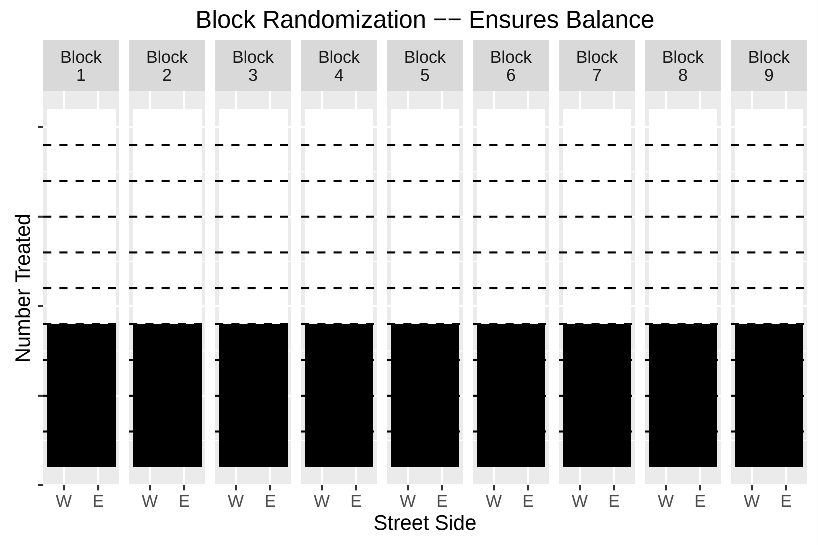
Ejemplo – Asignación GOTV (fuente: Erin Hartman)



Ejemplo – Asignación GOTV (fuente: Erin Hartman)



Ejemplo – Asignación GOTV (fuente: Erin Hartman)



Resultados Potenciales

Villa	Bloque	Y(I) Control	Y(I) Tratamiento
1	A	0	0
2	A	1	0
3	A	2	1
4	A	4	2
5	A	4	0
6	A	6	0
7	A	6	2
8	A	9	3
9	B	14	12
10	B	15	9
11	B	16	8
12	B	16	15
13	B	17	5
14	B	18	17

Table 3.3 Listado de resultados potenciales para proyectos de obras públicas cuando son auditados(Y(1)) y cuando no son auditados (Y(0))

		Todos participantes		Bloque A participantes		Block B participantes	
Villa	Bloque	Y(0)	Y(1)	Y(0)	Y(1)	Y(0)	Y(1)
1	A	0	0	0	0		
2	A	1	0	1	0		
3	A	2	1	2	1		
4	A	4	2	4	2		
5	A	4	0	4	0		
6	A	6	0	6	0		
7	A	6	2	6	2		
8	A	9	3	9	3		
9	B	14	12			14	12
10	B	15	9			15	9
11	B	16	8			16	8
12	B	16	15			16	15
13	B	17	5			17	5
14	B	18	17			18	17
	Promedio	9.14	5.29	4.00	1.00	16.0	11.0
	Varianza	40.41	32.49	7.75	1.25	1.67	17.0
	Cov[Y(0),Y(1)]	31.03		2.13		1.00	

	Todos participantes		Bloque A		Bloque B	
	Y_i^c	Y_i^t	Y_i^c	Y_i^t	Y_i^c	Y_i^t
Promedio	9.14	5.29	4.00	1.00	16.00	11.00
Varianza	40.41	32.49	7.75	1.25	1.67	17.00
Covarianza	31.03		2.13		1.00	

Estimación del EPT con Asignación Aleatoria Estratificada

$$\text{SATE}_{strat} = \sum_{j=1}^{J} \frac{N(j)}{N} \text{SATE}_j$$

- Donde J es el numero de bloques y $\frac{N(j)}{N}$ es la proporción de todos los participantes en el bloque j
- Promedio ponderado del SATE en cada bloque.

Nivel de Aleatorización

- ¿Cómo decidir unidad experimental? i.e. ¿a qué nivel administramos tratamiento? Algunas consideraciones:
  - Unidad de medición de variable de respuesta: no aleatorizar a nivel menor de medición.
  - Derrame: puede ocurrir a niveles bajos de aleatorización (algunos diseños ayudan en este caso)
  - Cumplimiento y atrición
  - Poder estadístico: niveles bajos de agregación más poder (fuerza opuesta a las anteriores).

Prueba de Hipótesis según Fisher

- Podemos evaluar algunas conjeturas que nos entregan el listado completo de los resultados potenciales
- Una conjetura es la llamada *"hipótesis nula nítida"*: el efecto de tratamiento es constante para todas las observaciones (usualmente cero)
- Bajo esta hipótesis, e.g. $Y_i(1) = Y_i(0)$. En este caso podemos observar *ambos* resultados potenciales de cada observación
- Simulaciónes de las asignación aleatoria entregan un distribución muestral exacta del efecto promedio del tratamiento en caso que utilizamos la "hipótesis nula nítida"

Resultados Observables Presupuesto Local

	Porcentaje del presupuesto si el jefe de la villa es hombre	Porcentaje del presupuesto si el jefe de la villa es mujer
Villa 1	?	15
Villa 2	15	?
Villa 3	20	?
Villa 4	20	?
Villa 5	10	?
Villa 6	15	?
Villa 7	?	30

Ejemplo: Aleatorización

- De esta tabla se genera un valor estimado de EPT de 6.5
- ¿Cuán probable es que podamos obtener un estimado tanto o más grande que 6.5 si el valor verdadero fuera cero en todas las observaciones?
- La probabilidad asumiendo que la hipótesis nula sea correcta (o el *p*-value), de interés en este caso evalúa una *hipótesis unilateral*, es decir que asumimos que cuando el jefe de la villa es mujer hay un incremento en la asignación presupuestaria para instalaciones de agua potable

Ejemplo: Prueba bilateral - "1-tailed test"

- Basándonos en los resultados observados en la tabla, podemos calcular las 21 posible estimaciones del EPT que podrían haber sido generados si la hipótesis nula fuera verdad: {-7.5, -7.5, -7.5, -4.0, -4.0, -4.0, -4.0, -4.0, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, 3.0, 3.0, 6.5, 6.5, 6.5, 10.0, 10.0}
- Cinco de las estimaciones son mayores o iguales a 6.5. Por lo tanto, cuando evaluamos mediante un test unilateral en donde la jefe de la villa es mujer, asumimos que el presupuesto asignado a instalaciones de agua potable *aumenta*, y estamos concluyendo que hay una probabilidad de $5/21 = 0.24$ de obtener un estimado mayor que 6.5 si es que la hipótesis nula es verdad.

Ejemplo: Prueba bilateral - "2-tailed test"

- Si buscamos evaluar la *prueba de dos colas* - es decir si la jefe de la villa es mujer puede implicar un incremento o disminución en la asignación presupuestaria para instalaciones de agua potable
- Estariamos calculando el "p-value" de obtener un número que es mayor o igual a 6.5 o bien menor o igual a -6.5. Un test de hipótesis bilateral considera todas las instancias en que los estimados son menores a iguales a 6.5 *en su valor absoluto*. Ocho de los estimados cualifican este criterio, por lo que el valor del "p-value" de dos colas es $8/21 = 0.38$

Señora Degustando Té

- Ronald Fisher, *The Design of Experiments*
- Experimento de té aleatorio:
  - 8 Copas idénticas preparadas
  - 4 tazas preparadas al azar con leche primero
  - 4 tazas preparadas al azar con leche después
- ¡La mujer clasificó correctamente todas las tazas!

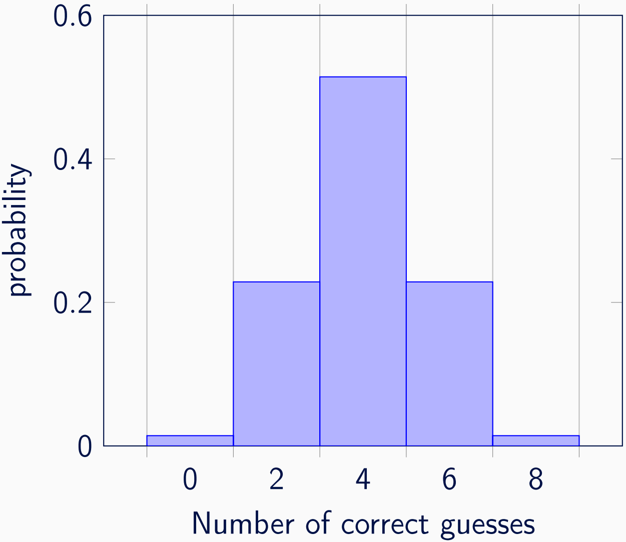
Señora Degustando Té

tazas	conjetura dama	orden real	escenarios			...
1	M	M	T	T	T	
2	T	T	T	T	M	
3	T	T	T	T	M	
4	M	M	T	M	M	
5	M	M	M	M	T	
6	T	T	M	M	T	
7	T	T	M	T	M	
8	M	M	M	M	T	
Número de conjeturas correctas		8	4	6	2	...

Señora Degustando Té

```
> ## truth: enumerate the number of assignment combinations

> true <- c(choose(4, 0) * choose(4, 4), choose(4, 1) *
    choose(4, 3), choose(4, 2) * choose(4, 2), choose(4, 3)
    * choose(4, 1), choose(4, 4) * choose(4, 0))

> ## compute probability: divide it by the total number of
    events

> true <- true/sum(true)

> ## number of correctly classified cups as labels

> names(true) <- c(0,2,4,6,8)

> true
         0          2          4          6          8
0.01428571 0.22857143 0.51428571 0.22857143 0.01428571
```

Señora Degustando Té



Señora Degustando Té Simulación

```
> ## Simulations

> sims <- 1000

> guess <- c("M", "T", "T", "M", "M", "T", "T", "M") # lady's guess

> correct <- rep(NA, sims) # place holder for number of correct guesses

> for (i in 1:sims) {
+   cups <- sample(c(rep("T", 4), rep("M", 4)), replace = FALSE)
+   correct[i] <- sum(guess == cups) # number of correct guesses
}

> ## comparison
> prop.table(table(correct)) - true
correct
          0           2           4           6           8
 0.001714286 0.004428571 -0.015285714 0.007428571 0.001714286
```

Fisher's Exact Test

```
> ## rows: actual assignments
> ## columns: reported guesses

> ## all correct
> x <- matrix(c(4, 0, 0, 4), byrow = TRUE, ncol = 2, nrow = 2)

> ## six correct
> y <- matrix(c(3, 1, 1, 3), byrow = TRUE, ncol = 2, nrow = 2)

> rownames(x) <- colnames(x) <- rownames(y) <- colnames(y) <- c("M", "T"
    )

> x
  M T
M 4 0
T 0 4
> y
  M T
M 3 1
T 1 3
```

Fisher's Exact Test

```
> fisher.test(x, alternative = "greater") # one-
    sided
```

Fisher's Exact Test for Count Data

data: x

p-value = 0.01429

hipótesis alternativa: verdadero valor del radio de posibilidades es mas

que 1 95 por ciento intervalo de confianza:

2.003768 Inf

estimaciones de la muestra:

radio de posibilidades

Inf

Fisher's Exact Test

```
> fisher.test(x, alternative = "greater") # one-
    sided
```

Fisher's Exact Test for Count Data

data: y

p-value = 0.2429

alternative hypothesis: true odds ratio is greater than 1 95 percent

confidence interval:

0.3135693 Inf

sample estimates:

odds ratio

6.408309

Fisher's Exact Test

```
> fisher.test(x) # two-sided
```

Fisher's Exact Test for Count Data

data: x

p-value = 0.02857

alternative hypothesis: true odds ratio is greater than 1 95 percent

confidence interval:

1.339059 Inf

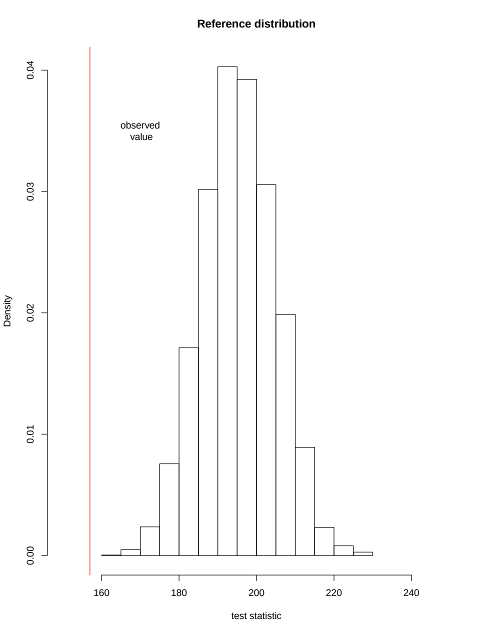
sample estimates:

odds ratio

Inf

Reanudar el experimento

```
> setwd("~/Dropbox/Experimental␣Methodology/DPIR␣2017/qss−master/
    CAUSALITY")
> resume <− read.csv("resume.csv")
> sims <− 5000
> z <− rep(NA, sims)
> for (i in 1:sims) {
+     ## shuffles treatment assignment
+     treat <− sample(resume$race, replace = FALSE)
+     ## test statistic
+     z[i] <− sum(resume$call[treat == "black"])
+ }
> ## observed z
> z.obs <− sum(resume$call[resume$race == "black"])
> z.obs
[1] 157
> ## one−sided p−value; proportion of simulated draws less than observed
     value
> pvalue <− mean(z[i] <= z.obs)
> pvalue
[1] 0
> ## histogram of reference distribution
> hist(z, freq = FALSE, xlim = c(150, 250),
+      xlab = "test␣statistic", main = "Reference␣distribution")
> abline(v = z.obs, col = "red")
> text(170, 0.035, "observed\n␣value")
```


Reference distribution
observed
value
Density
0.00
0.01
0.02
0.03
0.04
160
180
200
220
240
test statistic

Comentarios Sobre la Aleatorización

- Uno obtiene arbitrariamente valores precisos de p-values sin depender en los supuestos de distribución
- El mismo método se puede utilizar para una amplia gama de aplicaciones y estadísticas de prueba (por ejemplo, diferencias en medias, regresiones, diferencias en varianzas, etc.)
- Obliga al investigador a tomar un momento para pensar cuidadosamente acerca de cuál es la hipótesis nula y cómo debe ser testeada

Inferencia II: Intervalos de confianza de Neyman

	Porcentaje del presupuesto si el jefe de la villa es hombre	Porcentaje del presupuesto si el jefe de la villa es mujer	Efecto Tratamiento
Villa 1	10	15	5
Villa 2	15	15	0
Villa 3	20	30	10
Villa 4	20	15	-5
Villa 5	10	20	10
Villa 6	15	15	0
Villa 7	15	30	15
Promedio	15	30	15

Inferencia II: Intervalos de confianza de Neyman

Tabla 3.1 Distribución de muestra de EPTs estimados generados cuando dos de siete villas del listado Tabla 2.1 son asignadas para tratamiento

$\hat{\delta}_{fs}$	Frecuencia con la que un estimado ocurre
-1	2
0	2
0.5	1
1	2
1.5	2
2.5	1
6.5	1
7.5	3
8.5	3
9	1
9.5	1
10	1
16	1

Inferencia II: Intervalos de confianza de Neyman

- Basado en los números de la Tabla 3.1, calculamos el error estándar de la siguiente manera:

Suma del cuadrado de las desviaciones

$$
\begin{aligned}
&= (-1-5)^2 + (-1-5)^2 + (0-5)^2 + (0.5-5)^2 + (1-5)^2 + (1-5)^2 \\
&+ (1.5-5)^2 + (1.5-5)^2 + (2.5-5)^2 + (6.5-5)^2 + (7.5-5)^2 + (7.5-5)^2 \\
&+ (7.5-5)^2 + (8.5-5)^2 + (8.5-5)^2 + (8.5-5)^2 + (9-5)^2 + (9.5-5)^2 \\
&+ (10-5)^2 + (16-5)^2 = 445
\end{aligned}
$$

$$\textit{La raíz cuadrada del promedio de las desviaciones} = \sqrt{\frac{1}{21}(445)} = 4.60$$

Error estándar de $\hat{\delta}_{fs}$

$$ES(\hat{\delta}_{fs}) = \sqrt{\frac{1}{N-1}\left(\frac{N_t\ \mathbb{V}(Y_i(0))}{N_c} + \frac{(N_c)\ \mathbb{V}(Y_i(1))}{N_c} + 2\text{Cov}(Y_i(0), Y\right.}$$

$$ES(\hat{\delta}_{fs}) = \sqrt{\frac{1}{6}\left(\frac{(2)(14.29)}{5} + \frac{(5)(42.86)}{2} + (2)(7.14)\right)} = 4.60$$

- ¿$\widehat{ES}(\hat{\delta}_j)$?
- Los dos primeros términos son fáciles de estimar: podemos usar las varianzas de muestra de $Y_i^{obs}|D_i = 1$ y $Y_i^{obs}|D_i = 0$, respectivamente
- La covariance entre valores potenciales no es posible estimarla

Error estándar de $\hat{\delta}_{fs}$

- La covarianza es negativa. Asumiendo que las varianzas son constantes, ¿cuándo es mínima la covarianza? ¿Cuándo es máxima?
- Cuando el efecto causal individual es constant y aditivo (i.e. $Y_i(1) - Y_i(0) = \delta$), la covarianza de valores potenciales es cero, y el estimador del error estándar se vuelve
$$\widehat{ES}(\delta)_{fs} = \frac{\widehat{\mathbb{V}}(Y_i(0))}{N_c} + \frac{\widehat{\mathbb{V}}(Y_1(1))}{N_t}$$
- Si el efecto es heterogéneo, este estimador tiene un sesgo positivo: es conservador.

Intervalos de Confianza

- Armados de con este estimador de la varianza de la distribución de muestreo de $\hat{\delta}_{fs}$, podemos construir intervalos de confianza de formas tradicionales:

$$\mathsf{CI}^{1-\alpha}(\delta_{fs}) = \hat{\delta}_{fs} \pm z_{\alpha/2}\sqrt{\mathbb{V}(\hat{\delta})}$$

- Tuvimos que asumir que la distribución de aleatorización (i.e. la distribución de muestreo) de $\hat{\delta}_i$ es Normal. No es muy atractivo, pero funciona en la práctica.

Analísis de Poder

Poder Estadístico

- ¿Cuál es el poder de una prueba estadística? H_0: hipótesis nula
- Aplique el estimador para probar una alternativa H_A
- Error tipo I: Falso positivo
    - Si la hipótesis nula es verdad, ¿Qué probabilidad tiene que el efecto estimado (o un valor mayor) ocurra por casualidad?
    - Nuestra tolerancia a estos errores esta dada por α
    - Cuando $\alpha = 0.05$, 95% del intervalo de confianza es construido en base a varias muestras que contendrá el valor real del parámetro

Poder Estadístico

- Error tipo II: Falso negativo
    - Si la hipótesis nula no es verdadera, ¿con qué frecuencia podemos rechazar la hipótenis nula con éxito?
    - Probabilidad o tasa de error tipo II, β
- Poder de un test: Probabilidad de que la prueba rechace $H_0, 1 - \beta$

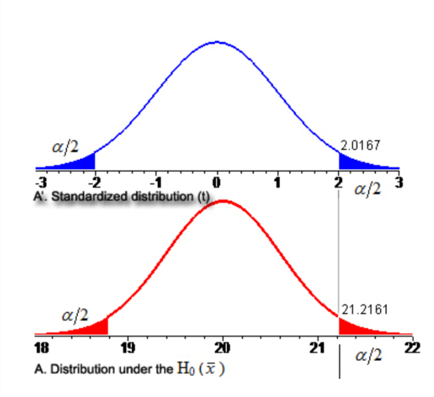
Inferencia Básica Revisada

- ¿Cuál es el efecto de perder Medicaid en la mortalidad infantil?
- $H_0 = 20$ muertes por cada 1,000 nacidos vivos (asumimos conocemos incertidumbre aquí)
- El verdadero efecto es un aumento de 2 muertes por cada 1,000 nacidos vivos
- La desviación estándar en la población es 4, tenemos N=44 observaciones; la distribución de muestreo produce un error estándar de 0.60

Inferencia Básica Revisada

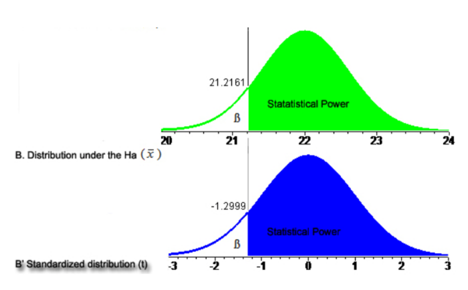
- $\hat{x}$ es nuestro valor estimado de la nueva tasa de mortalidad infantil
- Digamos que obtenemos un estimado $\hat{x} = 21.2161$
- ¿Qué tan improbable es que obtengamos este estimado, si la hipótesis nula es verdadera?

Distribución de Muestreo Bajo la Hipótesis Nula



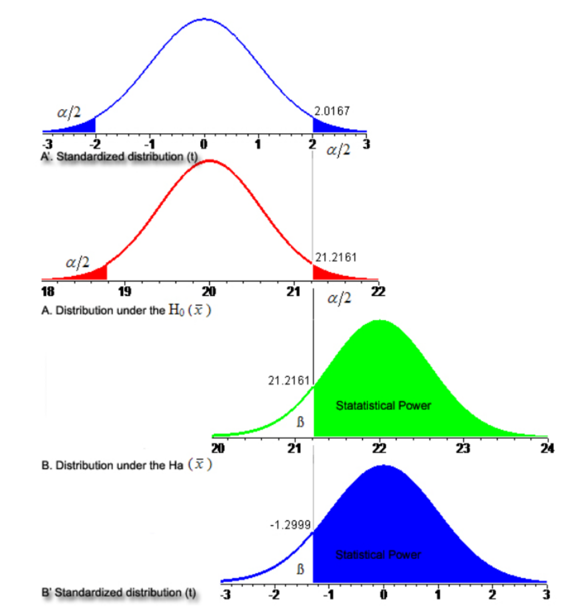
- Digamos que nuestro test es $\alpha = 0.05$
- Puedes cambiar la escala mediante una transformación Z
- ¿Qué significa este gráfico?
- Para $\hat{x} = 21.2161$,
- t-stat=2.0167, $p = 0.011$

Distribución de muestreo de $\hat{x}$



- Interpretemos este gráfico
- $1-\beta$ es la fracción de los estimados que rechazan la hipótesis nula
- Poder del test: probabilidad de identificar un efecto de un tamaño dado como significativo.
- ¿Qué parámetros se necesitan?

La Relación entre α y β



El Tamaño de la Muestra Aumenta el Poder Estadístico

- De interés primordial porque puede ser manipulada
- Ley de los grandes números: Para datos independientes, la precisión estadística de las estimaciones aumenta con la raíz cuadrada del tamaño de la muestra, $\sqrt{n}$
- Las pruebas estadísticas generalmente tienen la forma $T = \hat{\theta}/\sqrt{\hat{V}(\hat{\theta})}$
- Por ejemplo: Promedio de la distribución normal θ, data $y = (y_1, ..., y_n)$, iid

$$\hat{\theta} = n^{-1}\sum_{i=1}^{n} y_i = \bar{y}$$

$$\hat{V}(\hat{\theta}) = V(y)/n \text{ and } \sqrt{\hat{V}(\hat{\theta})} = s_y/\sqrt{n}$$

$$T = \bar{y}/(s_y/\sqrt{n})$$

El Tamaño de la Muestra Aumenta el Poder Estadístico

- Esta lógica se extiende para los casos con dos muestras (por ejemplo, tratados vs control en un experimento), regresiones, regresión logística, etc.

"Reverse Engineer T" para Determinar el Tamaño de la Muestra

- ¿Cuánta muestras necesito para darme una oportunidad "razonable" de rechazar H_0, dadas las expectativas respecto a la magnitud del "efecto"
- Ejemplo:

  Una proporción $\theta \in [0, 1]$ estimada como $\hat{\theta}$

  La varianza es $\theta(1 - \theta)/n$, máximo hasta 0.5

  Un Intervalo de Confianza de 95% $\theta = 0.5$ es $0.5 \pm 2\sqrt{0.25/n}$

  El ancho de ese intervalo es $W = 4\sqrt{0.25/n} \rightarrow n = 4/W^2$

“Reverse Engineer T” para Determinar el Tamaño de la Muestra

- Uso típico: ¿cuán grande debe ser una encuesta para obtener un margen de error razonable?
- Para los investigadores, ¿cuán grande debe ser una encuesta para detectar el efecto de un campaña?
  - La respuesta depende de la creencia que tenemos acerca de la magnitud probable de los efectos de la campaña

Ejemplo 2: efecto campaña

- En R, power.prop.test()
- El investigador piensa que es probable que los efectos se mueven en una proporción entre 50% y 52% (por ejemplo, apoyo a votar)
- Desea poder detectar efectos de este tamaño en niveles convencionales de significancia estadística
- ($p = 0.05; 95\%$ Intervalo de confianza para el efecto excluye cero), con poder $(1 - \beta)$ igual a 0.50
- $H_0 : \delta = \theta_1 - \theta_2 = 0; H_A : \delta \neq 0$ (Un test de dos colas es una alternativa)

Estimación de Poder Para 2 Puntos de Efectos

Test dos colas con un nivel convencional de significancia

```
>power.prop.test(p1 = 0.5, p2 = 0.52, power
    = 0.5)
```

Comparación de dos muestras de cálculo de potencia de proporciones

n = 4799.903

p1 = 0.5

p2 = 0.52

sig.level = 0.05

power = 0.5

alternative = two.sided

NOTE: n is number in *each* group

Estimación de Poder Para 2 Puntos de Efectos

Test unilateral con un nivel de convencional de significancia

```
> power.prop.test(p1 = 0.5, p2 = 0.52,
   power = 0.5,
alternative="one.sided")
```

Comparación de dos muestras de cálculo de potencia de proporciones

n = 3380.577

p1 = 0.5

p2 = 0.52

sig.level = 0.05

power = 0.5

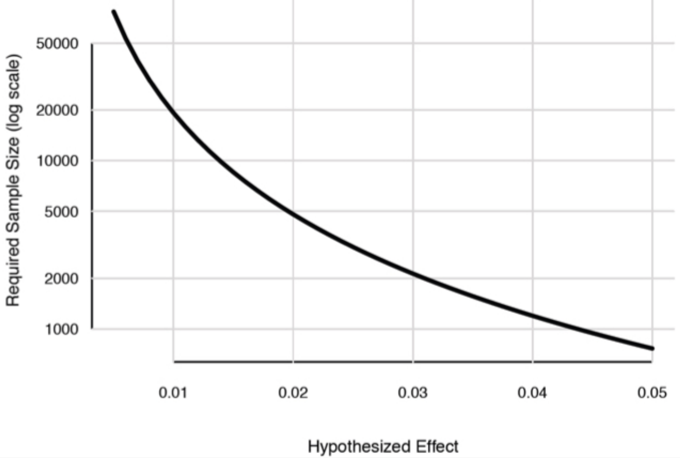
alternative = one.sided

NOTE: n is number in *each* group

Curvas de Poder

```
> p> effects <- seq(0.005, 0.05, by =
    0.001)

> base <- 0.5
> m <- length(effects)
> n <- rep(NA, m)
> for (i in 1:m) {
n[i] <- power.prop.test(p1 = base, p2 =
    base + effects[i],
+ power = 0.5)$n
+})
```

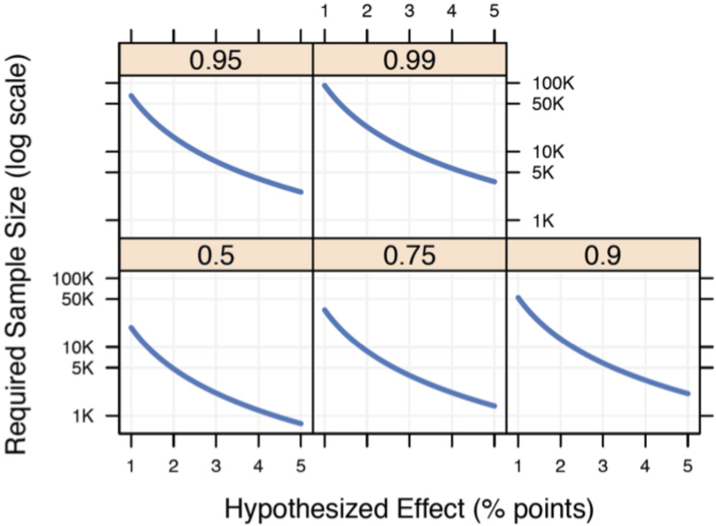
Curvas de Poder



Curvatura de las Curvas de Poder

```
> power <- c(0.5, 0.75, 0.9, 0.95, 0.99)
> effects <- seq(0.01, 0.05, by = 0.001)
> base <- 0.5
> m <- c(length(power), length(effects))
> n <- matrix(NA, m[1], m[2])
> for (i in 1:(m[1])) {
+ for (j in 1:(m[2])) {
+ n[i, j] <- power.prop.test(p1 = base, p2
    = base + effects[j],
+ power = power[i])$n
+ }
+ }
```

Curvas de Poder: Diferentes Niveles de Poder



Consejos Prácticos Sobre el Análisis de Poder

- Cuál es el tamaño "típico" de los efectos, y cómo podríamos adivinarlo?
    - Algunas reflexiones sobre un ejemplo posterior
- Generalmente, los estudios experimentales requieren $1 - \beta > 0.8$ cuando se solicitan fondos a agencias
- Máxima de Zaller: "Haga su análisis de poder, determina el tamaño de su muestra, luego dobla ese numero"

Consejos Prácticos sobre el Analísis de Poder

- Consideraciones de costos: Gerber y Green experimento en New Haven GOVT
  - Uno de los componentes involucró la prospección
  - $40 por hora con un par de estudiantes, 6,000 tratados
  - Son son 6 casas por hora, necesitas 1000 horas, por lo que son $40k
  - Implicaciones basadas en diapositivas de curva de potencia
- En particular, los costos son altos para los experimentos que buscan evaluar a la población en general
- ¿Alguien tiene una idea de cuánto cuestan las encuestas?
- ¿Cuánto es su valor?