Prueba de hipótesis, efectos promedio de tratamiento, errores estándar y p-valores

Abrebocas: Pruebas de hipótesis

- ¿Qué es una hipótesis?
- ¿Cómo es una buena hipótesis?
- Aceptar o rechazar estadísticamente un hipótesis
- Pasos para realizar una prueba de hipótesis:
  - Asumir lo negativo (hipótesis nula)
  - Rechazarla o no rechazarla (con cierto nivel de confianza o p-valor)
- El **p**-valor
- Calcular **p**-valores

¿Qué es una hipótesis?

- Son probablemente **VERDADERAS** o **FALSAS**
- Son **falsificables**
- Son **afirmaciones** sobre el mundo, no su análisis
- Son **sencillas** (no son de doble cañón)
- Involucran **conceptos claros**
- Son pocas, y están **motivadas**
- Se **cuestionan**: Aprenderán algo si los datos las apoyan o si las rechazan.
- Más importante: **no está seguro** si son verdaderas o falsas
- Están **enumeradas**, e incluso **nombrada**

Algunas hipótesis

- Consideremos las siguientes:
  - La educación es muy importante
  - La educación aumenta sus ingresos
  - La educación aumenta, disminuye o no tiene efecto sobre sus ingresos
  - La educación es buena para usted porque fortalece su carácter de maneras muy fundamentales que nunca podría medir
- Ahora:
  - Sólo una de estas no es una hipótesis. ¿Cúal?
  - Sólo una de estas es una buena hipótesis. ¿Cúal?

Nulas: Un foco de confusión

- Debido a una convención inusual, los científicos sociales suelen describir hipótesis en términos de lo que esperan encontrar, pero luego prueban la hipótesis nula de ningún efecto
- Entonces:
  - H1: El conflicto reduce el crecimiento
  - H-nula: El conflicto no tiene ningún efecto sobre el crecimiento
  - Prueba: ¿Cuál es la probabilidad de observar los datos que observados dado la hipótesis nula de ningún efecto?

Pruebas: Las hipótesis son usualmente rechazadas, algunas veces se mantienen, pero raramente se aceptan

- En el enfoque clásico para probar una hipótesis nos preguntamos:
- ¿Qué tan probable es que veamos datos como estos si la hipótesis es verdadera?
  - Si la respuesta es "bastante probable" entonces tratamos la hipótesis como sospechosa.
  - Si la respuesta es "poco probable", entonces se mantiene la hipótesis.
- ¿Qué tan improbable es "poco probable"?

Preliminares: Efecto (Estimado) Promedio de Tratamiento o ATE

- Hay un "verdadero" efecto promedio de tratamiento (ATE) en el mundo
- Tratamos de estimarlo con un solo experimento
- ATE estimado = (Promedio de los resultados de las unidades de tratamiento) - (Promedio de los resultados de las unidades de control)
- Si repetimos el experimento muchas veces, para todas las asignaciones de tratamiento posibles, el promedio de todos estos ATE estimados convergería en la ATE verdadero. Pero sólo ejecutamos un experimento ¿Cómo sabemos si nuestro ATE estimado es estadísticamente significativo?

Preliminares: Variance & Desviación Estándar

- Ambos son **medidas de dispersión** de un estadístico
- Variación: desviación media cuadrada de la media de una variable:

$$Var(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

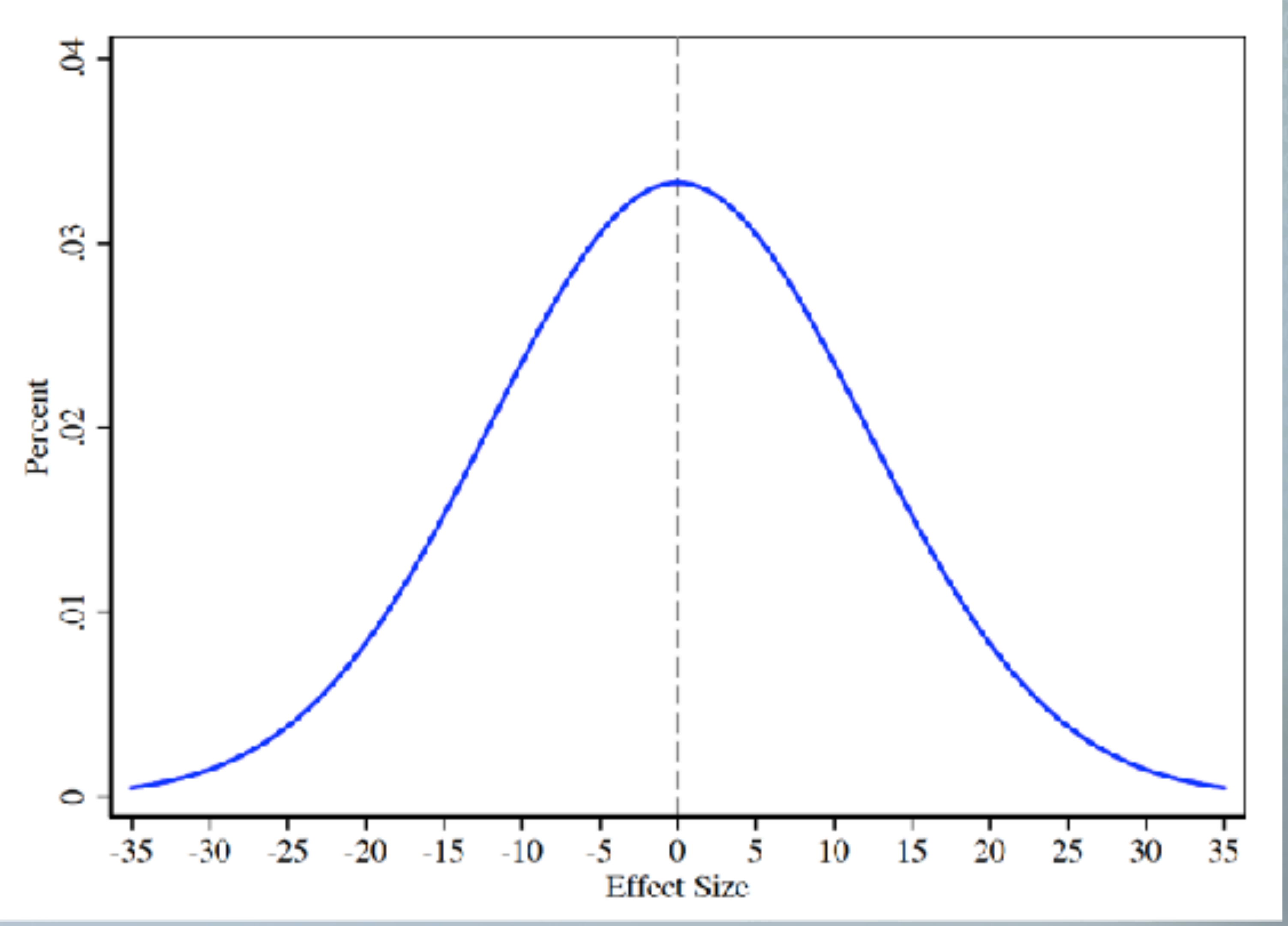
- Desviación Estándar: la **raíz cuadrada de la varianza**:

$$SD_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

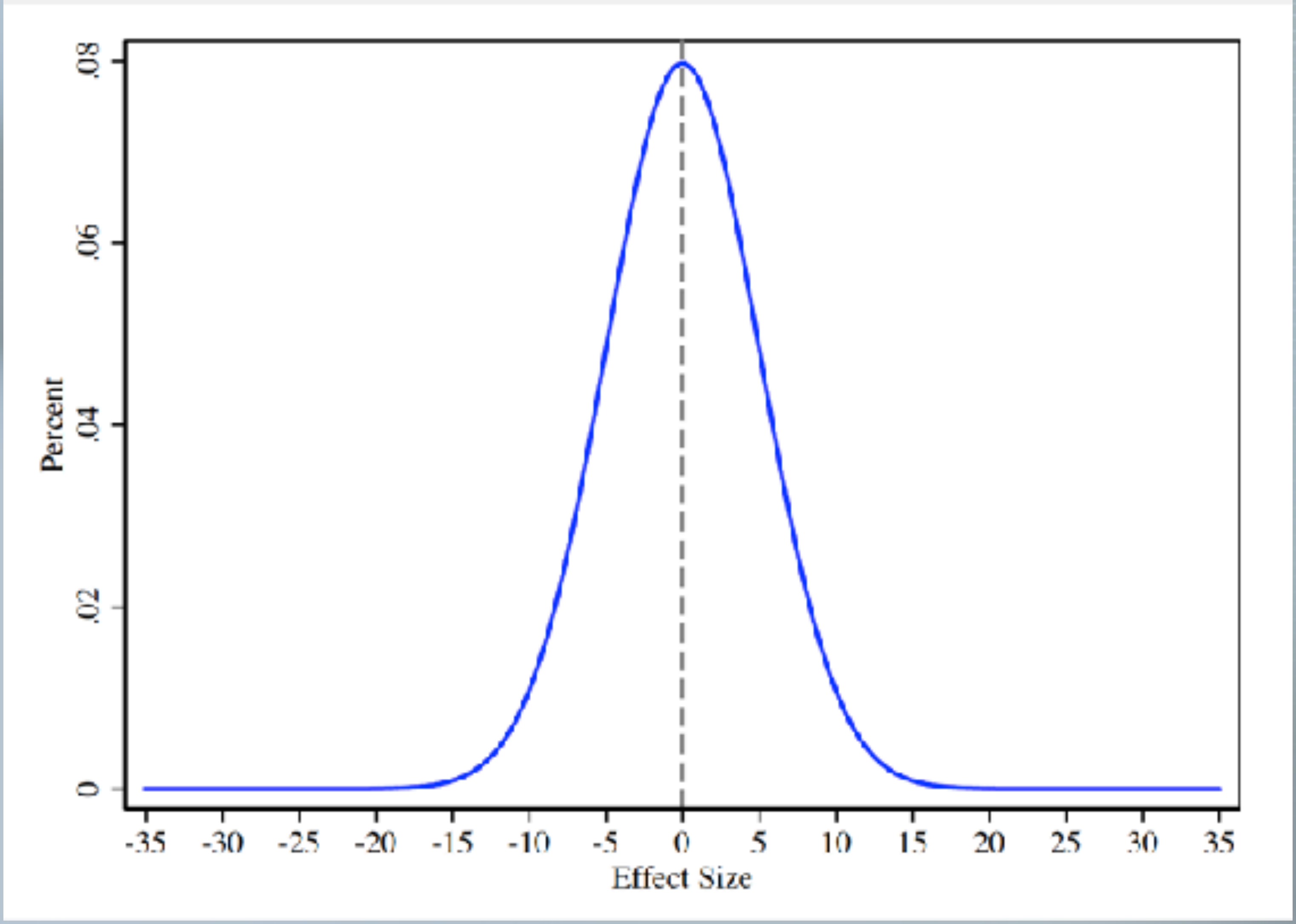
Preliminares: Error Estándar

- ¿Qué es un error estándar?
  - Error estándar = desviación estándar de la distribución muestral
  - Es una medida de la variabilidad de la muestra
- Un error estándar mayor significa que la estimación es más incierta
- Para estimaciones precisas, necesitamos que el error estándar sea pequeño en relación con los efectos

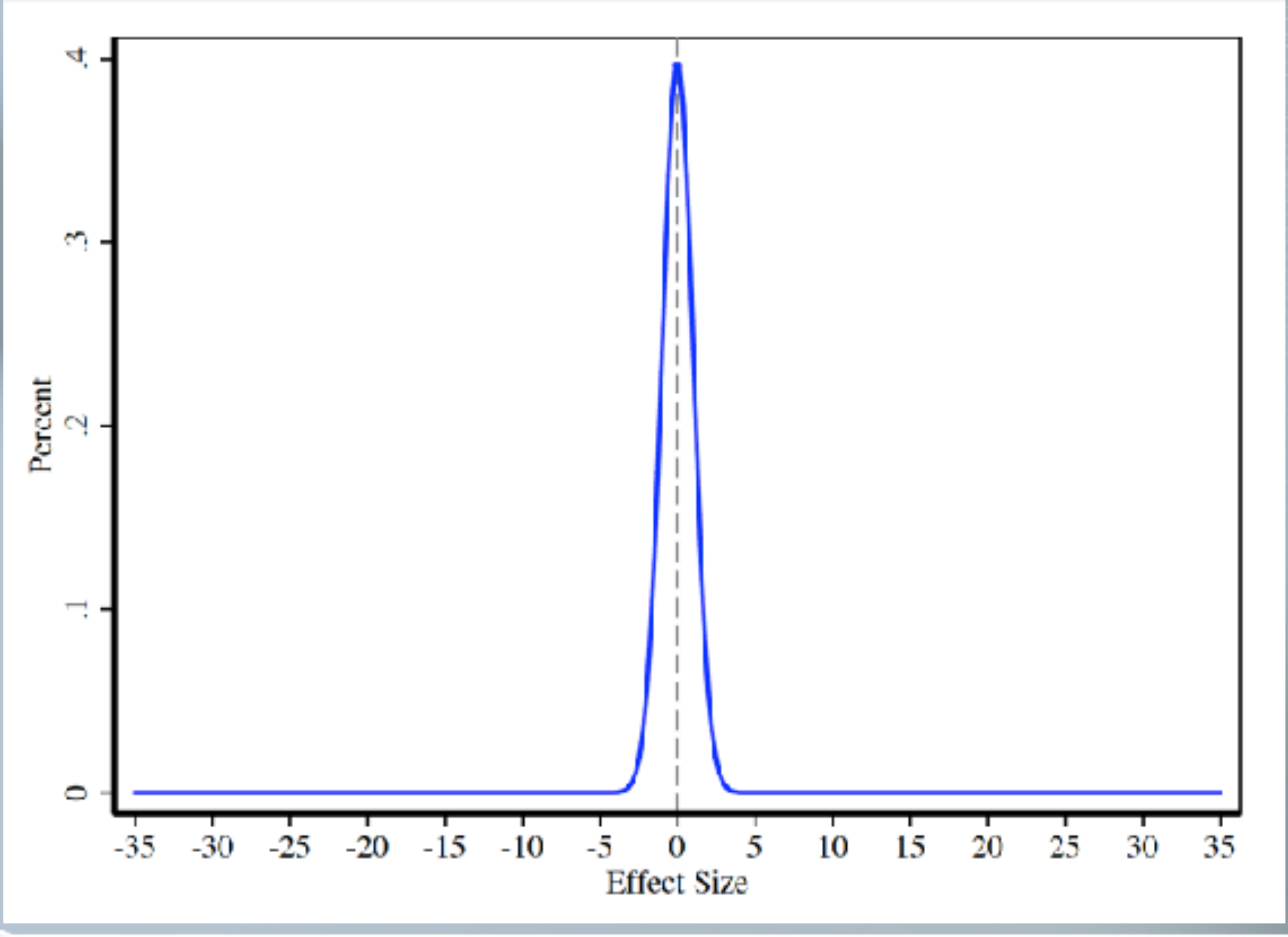
Distribución muestral: Muestra grande



Mayor o menor error estándar?

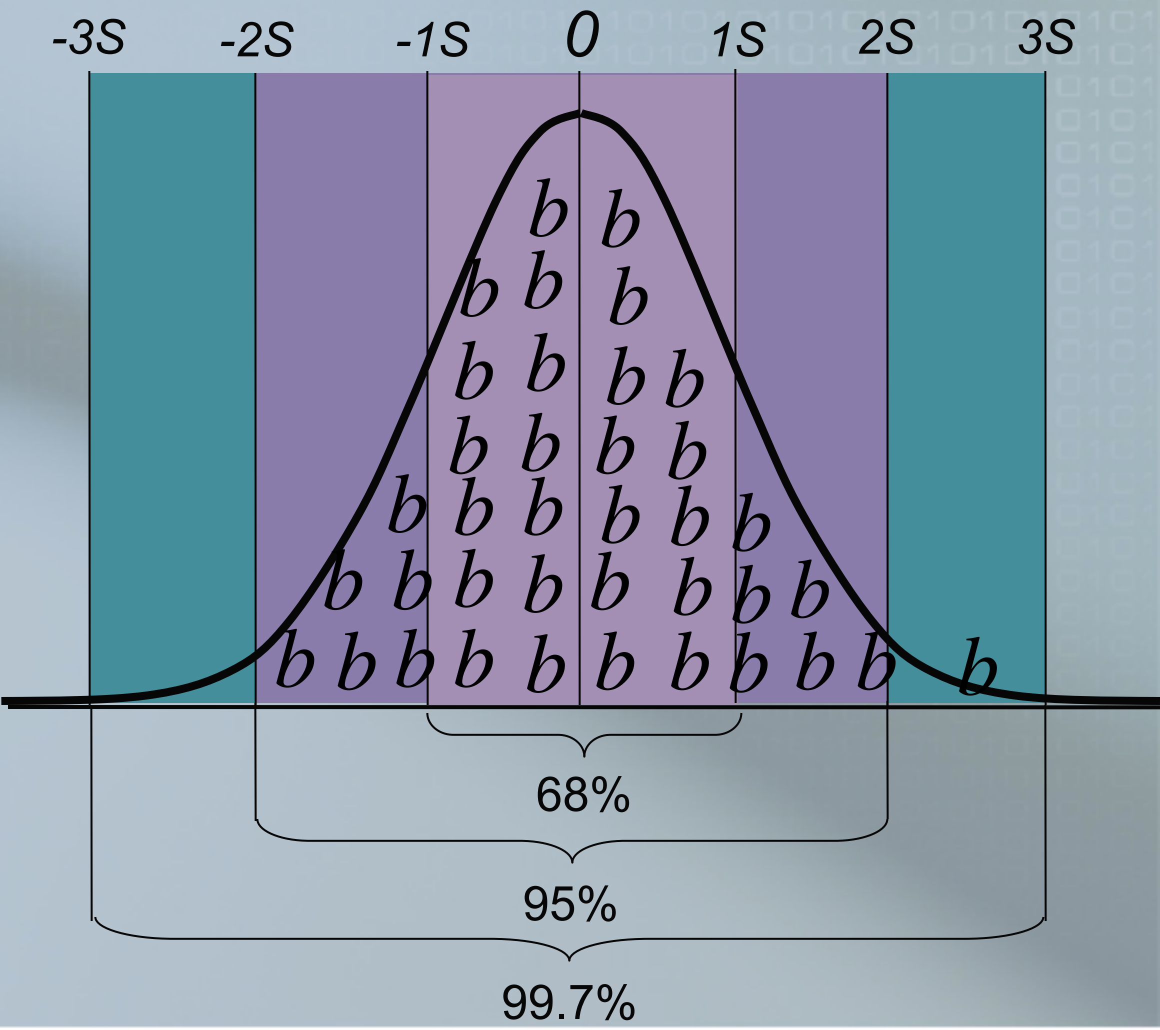


Mayor o menor error estándar?



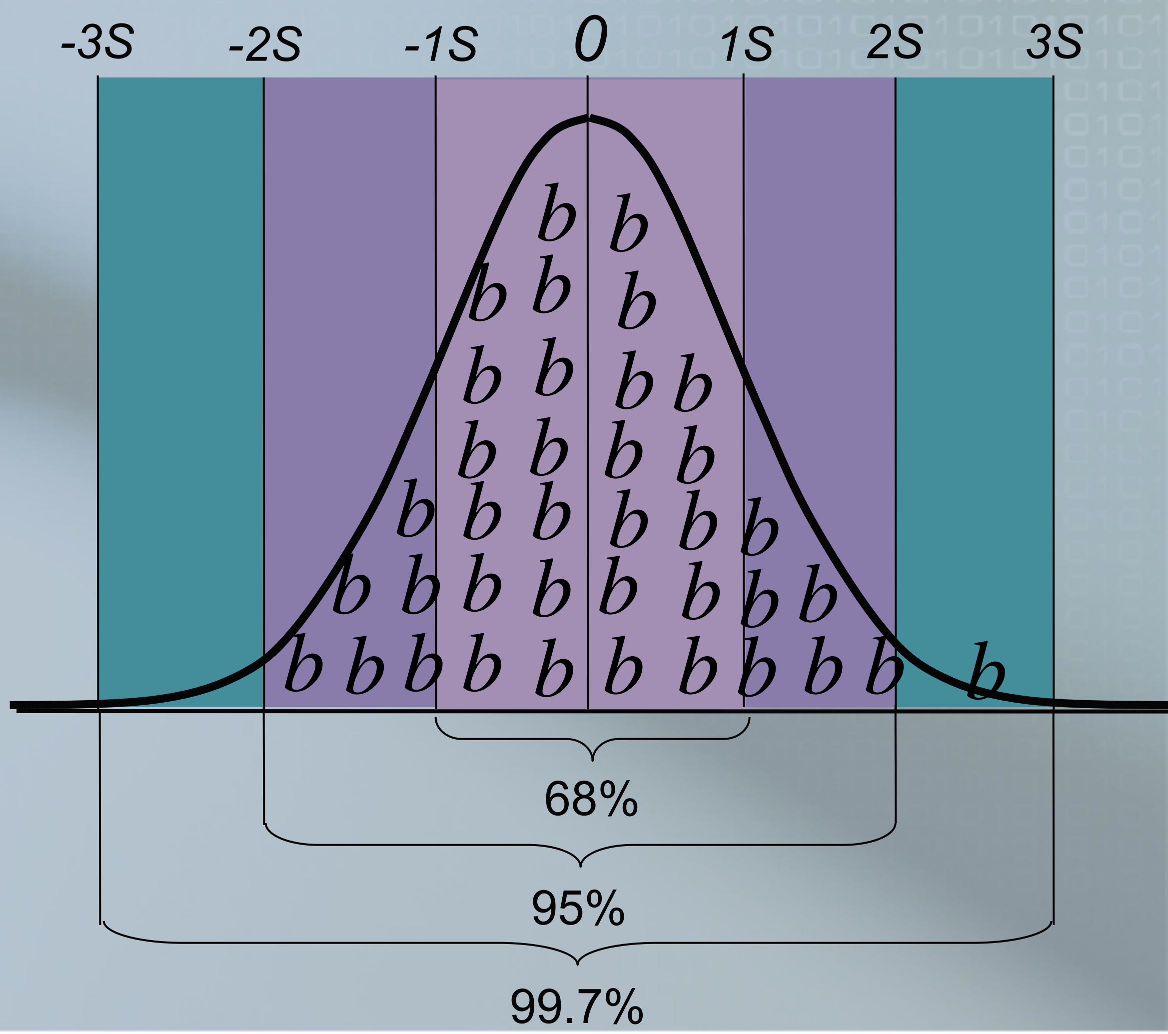
Distribución muestral

- Asuma que la hipótesis nula (diferencia entre tratamiento y control=0) es verdadera.
- Si se cuenta con muchas muestras, esperamos que las diferencias estén centradas alrededor de 0 (Teorema del límite central)



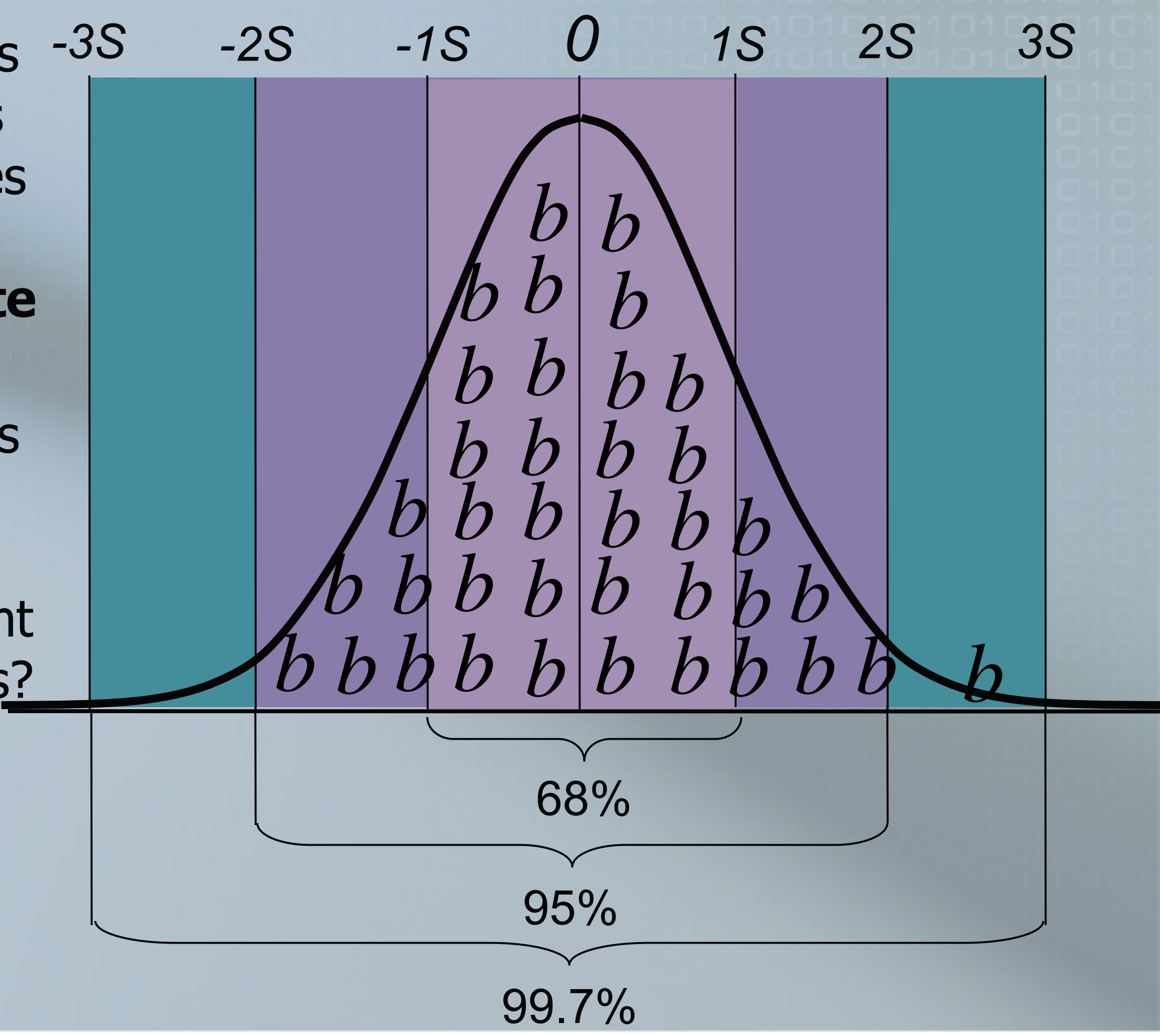
Distribución muestral

- Asuma que la hipótesis nula (diferencia entre tratamiento y control=0) es verdadera.
- Si se cuenta con muchas muestras, esperamos que las diferencias estén centradas alrededor de 0 (Teorema del límite central)



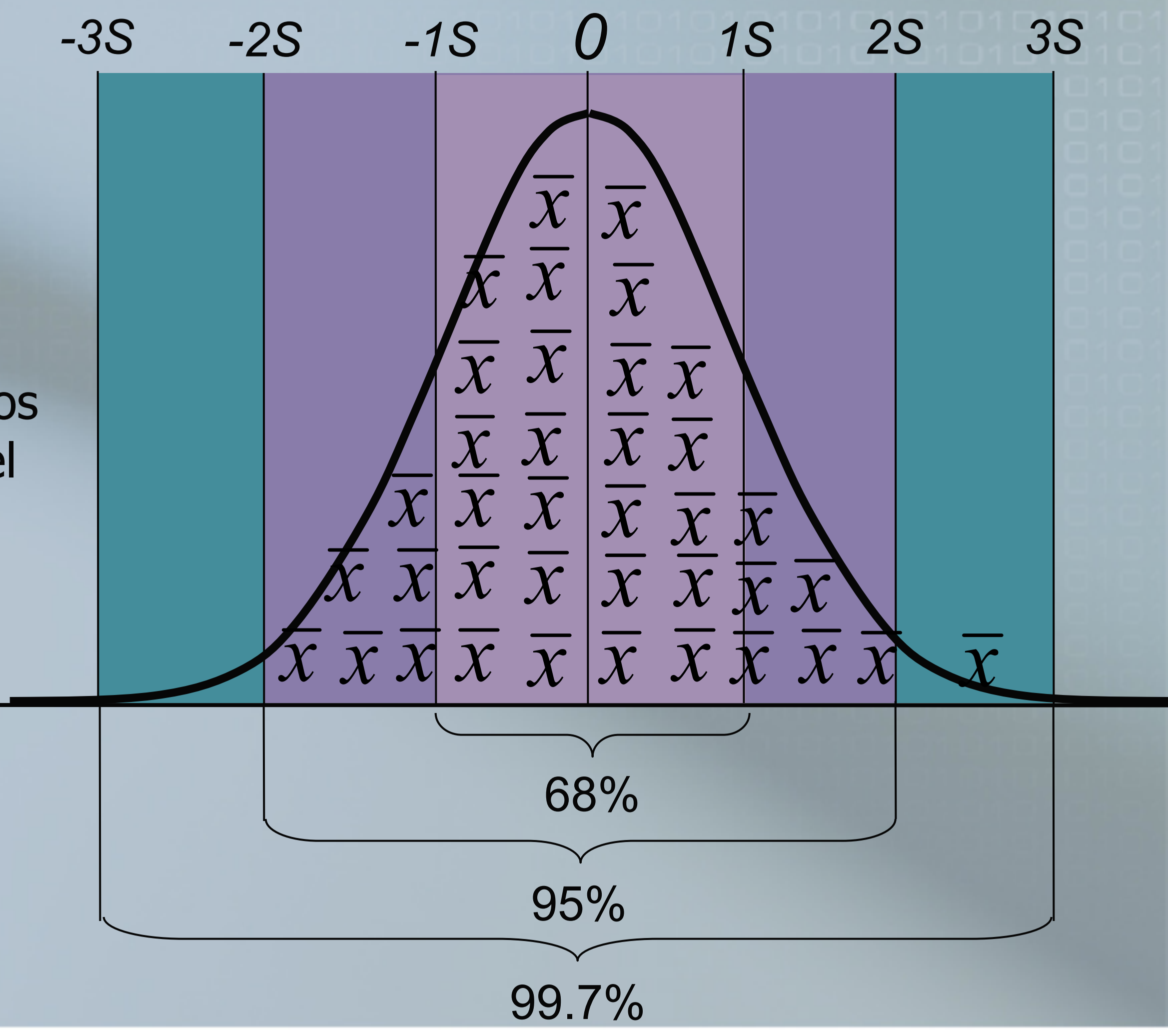
Distribución muestral

- Si la probabilidad de que la hipótesis nula sea verdades es < 5%, entonces el resultalado es **estadísticamente significativo**
  - ¿Qué muestras en el gráfico son estadísticamente significativas? Z-score> 2 es significativo
  - El puntaje z <-2 es significativo



Distribución muestral

- En lugar del puntaje-z, **estadístico t** (idéntico si n>30)
- Cada **estadístico t** tiene su **p-valor**
- **p-valor** = prob. de observar un estadístico t al menos tan extremo como el que obtuvimos
- p-valor de t=2 es? **0.05**
- p-valor de t=-2 es? **0.05**
- p-valor de t=3 es?
  - **0.003**



El Teorema del Límite Central

- Hace que las estadísticas funcionen
- CLT en esencia:
  - Con múltiple muestras, el promedio de las medias de la muestra se aproxima a la media de la población
  - La distribución de las media de la muestras es normal
  - Incluso cuando la distribución de la población no se distribuye normal
- El Teorema del Límite Central es tu amigo

El **p** valor

- El famoso p-valor indica la probabilidad de observar este tipo de evidencia: p.ej.: La probabilidad de obtener un efecto estimado tan grande como el que observamos.
- Considere esto: "Yo estimo que el tratamiento aumentó el ingreso en \$10, mi error estándar es 4 y mi p-valor es 0.05".
- ¿Cuál de estas afirmaciones es correcta? (¡sólo una!):
    1. La probabilidad de que el tratamiento aumentó los ingresos en \$10 es sólo del 5%
    2. La probabilidad de que el tratamiento aumentó los ingresos en \$10 es sólo del 95%
    3. La probabilidad de que el tratamiento aumentó el ingreso es del 95%
    4. La probabilidad de que el tratamiento no aumente el ingreso es del 5%
    5. La probabilidad de que el tratamiento incrementara el ingreso es del 95%
    6. La probabilidad de que estimáramos un efecto de \$10 si el efecto verdadero fuera 0 es del 5%
    7. La probabilidad de que estimáramos un efecto de \$10 si el efecto verdadero fuera positivo es del 95%

El **p** valor

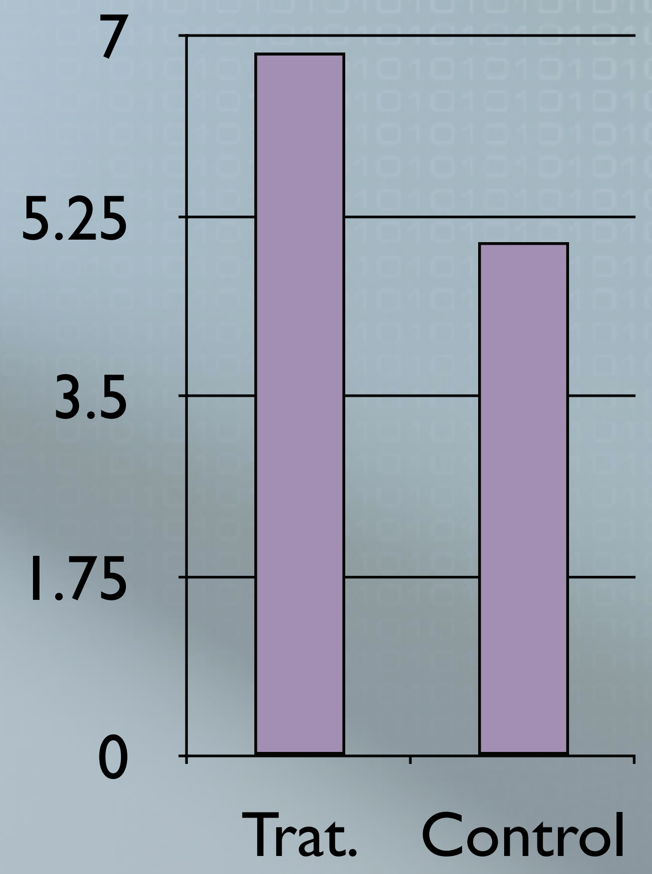
- El famoso p-valor indica la probabilidad de observar este tipo de evidencia: p.ej.: La probabilidad de obtener un efecto estimado tan grande como el que observamos.
- Considere esto: "Yo estimo que el tratamiento aumentó el ingreso en $10, mi error estándar es 4 y mi p-valor es 0.05".
- ¿Cuál de estas afirmaciones es correcta? (¡sólo una!):
  1. La probabilidad de que el tratamiento aumentó los ingresos en $10 es sólo del 5%
  2. La probabilidad de que el tratamiento aumentó los ingresos en $10 es sólo del 95%
  3. La probabilidad de que el tratamiento aumentó el ingreso es del 95%
  4. La probabilidad de que el tratamiento no aumente el ingreso es del 5%
  5. La probabilidad de que el tratamiento incrementara el ingreso es del 95%
  6. **La probabilidad de que estimáramos un efecto de $10 si el efecto verdadero fuera 0 es del 5%**
  7. La probabilidad de que estimáramos un efecto de $10 si el efecto verdadero fuera positivo es del 95%

Ejemplo: Sesgo Racial

- Pregunta central: ¿Podrían estas observaciones realmente ocurrir por casualidad?
- Ejemplo: Sesgo racial en la selección del jurado del Sur, 1960-1980
  1. El 50% de todos los jurados elegibles en la región eran afroamericanos
  2. De un panel de 80 personas de jurados potenciales, sólo 4 eran afroamericanos
  3. ¿Fue casual? ¿Pura casualidad?
  4. Probabilidad de 4 en 80 = 0.0000000000000000000014
- Igual que tres “royal flushes” en una fila en el póquer

¿Cómo sabría si hay una diferencia significativa?

Tratamiento	Control	
6	5	
8	0.7	
7	7	
5.7	6	
7.33	6	
5	7.7	
8	6.7	
6	2.7	
6.3	7.3	
6.7	6	
5.7	3	
8.7	3.3	
5.3	6.3	
10	2.3	
6.84	5	Media
	1.84	Est. ATE
2.04	4.78	Variance
14	14	N



Prueba clave de significancia estadística: prueba t

- Se utiliza para determinar si dos conjuntos de datos son estadísticamente diferentes entre sí
- Diferencia de medias
- Fórmula:

$$t_{(\bar{X}_1 - \bar{X}_2)} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$

Ejercicio: En grupos de a 2 calculen el estadístico “t”

Tratamiento	Control	
6	5	
8	0.7	
7	7	
5.7	6	
7.33	6	
5	7.7	
8	6.7	
6	2.7	
6.3	7.3	
6.7	6	
5.7	3	
8.7	3.3	
5.3	6.3	
10	2.3	
6.84	5	Media
	1.84	Est. ATE
2.04	4.78	Variance
14	14	N

$$t_{(\bar{X}_1 - \bar{X}_2)} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$

Donde Xbar = promedio
Sigma al cuadrado = Varianza
N = Número de observaciones

```
> t.test(treat, control)

	Welch Two Sample t-test

data:  treat and control
t = 2.6341, df = 22.379, p-value = 0.01503
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.3922873 3.2834269
sample estimates:
mean of x mean of y 
 6.837857  5.000000 
```

¿Cómo se ve esta prueba en R?

Resumen:

- Paso 1. Formular hipótesis
  - H_0 es la **hipótesis nula** – Por lo general, que las observaciones son el resultado puramente del azar (o error de muestreo) y que no existe una relación estadística entre las variables
  - H_1 (o H_a) es la **hipótesis alternativa** (la de ustedes) – Las observaciones son el resultado de un efecto real
- Paso 2. Identificar el **test estadístico** – Usualmente la prueba **t**
- Paso 3. **P-valor** – Probabilidad de observar un test-estadístico si H_0 es verdadera
- Paso 4. Comparar el p-valor con un nivel de significancia fijo (por ej. el umbral 0.05)
- Paso 5. **Rechazar** o **no rechazar** hipótesis nula

Inferencia Aleatorizada

- Una manera diferente de pensar en la significancia estadística (Fisher,1935)
  - También llamada "prueba de permutación" o "prueba exacta"
- Calcula el p-valor usando la información sobre la aleatorización
- Prueba la hipótesis nula estricta: no hay efecto del tratamiento para todas las unidades
- Revuelve la asignación del tratamiento para todas las unidades
  - Si N es pequeño, examina todas las permutaciones
  - Si N es grande, muchas permutaciones posibles (10.000 son comunes)
  - Si la hipótesis nula es verdadera, entonces la asignación del tratamiento no debe importar.
  - El efecto del tratamiento es 0 si la nula es verdadero, después de todo
- Así, RI compara el efecto del tratamiento en el experimento real con los posibles "efectos del tratamiento" en los datos codificados
- El p-valor en RI muestra la proporción de las observaciones revueltas cuyos efectos en el tratamiento son menos extremos que los datos experimentales reales
  - RI es a menudo superior a la regresión para el análisis de experimentos
- Pequeñas muestras, valores atípicos

Ejemplo de RI

Unidad	1	2	3	4	5	6	7	8	9	10
Exp.: Trat.	0	0	0	0	0	0	0	1	0	0
Exp.: Paz	3	2	4	6	7	2	4	9	8	2

Ejemplo de RI

Unidad	1	2	3	4	5	6	7	8	9	10
Exp.: Trat.	0	0	0	0	0	0	0	0	0	1
Exp.: Paz	2	2	2	3	4	4	6	7	8	9

Ejemplo de RI

Unidad	1	2	3	4	5	6	7	8	9	10
Exp.: Trat.	0	0	0	0	0	0	0	0	0	**1**
Exp.: Paz	2	2	2	3	4	4	6	7	8	**9**
P1: Trat.	**1**	0	0	0	0	0	0	0	0	0
P1: Paz	**2**	2	2	3	4	4	6	7	8	9

Ejemplo de RI

Unidad	1	2	3	4	5	6	7	8	9	10
Exp.: Trat.	0	0	0	0	0	0	0	0	0	**1**
Exp.: Paz	2	2	2	3	4	4	6	7	8	**9**
P1: Trat.	**1**	0	0	0	0	0	0	0	0	0
P1: Paz	**2**	2	2	3	4	4	6	7	8	9
P2: Trat.	0	**1**	0	0	0	0	0	0	0	0
P2: Paz	2	**2**	2	3	4	4	6	7	8	9

Ejemplo de RI

Unidad	1	2	3	4	5	6	7	8	9	10
Exp.: Trat.	0	0	0	0	0	0	0	0	0	**1**
Exp.: Paz	2	2	2	3	4	4	6	7	8	**9**
P1: Trat.	**1**	0	0	0	0	0	0	0	0	0
P1: Paz	**2**	2	2	3	4	4	6	7	8	9
P2: Trat.	0	**1**	0	0	0	0	0	0	0	0
P2: Paz	2	**2**	2	3	4	4	6	7	8	9
P3: Trat.	0	0	**1**	0	0	0	0	0	0	0
P3: Paz	2	2	**2**	3	4	4	6	7	8	9

Ejemplo de RI

Unidad	1	2	3	4	5	6	7	8	9	10
Exp.: Trat.	0	0	0	0	0	0	0	0	0	**1**
Exp.: Paz	2	2	2	3	4	4	6	7	8	**9**
P1: Trat.	**1**	0	0	0	0	0	0	0	0	0
P1: Paz	**2**	2	2	3	4	4	6	7	8	9
P2: Trat.	0	**1**	0	0	0	0	0	0	0	0
P3: Paz	2	**2**	2	3	4	4	6	7	8	9
P4: Trat.	0	0	**1**	0	0	0	0	0	0	0
P4: Paz	2	2	**2**	3	4	4	6	7	8	9
P5: Trat.	0	0	0	**1**	0	0	0	0	0	0
P5: Paz	2	2	2	**3**	4	4	6	7	8	9
P6: Trat.	0	0	0	0	**1**	0	0	0	0	0
P6: Paz	2	2	2	3	**4**	4	6	7	8	9
P7: Trat.	0	0	0	0	0	**1**	0	0	0	0
P7: Paz	2	2	2	3	4	**4**	6	7	8	9
P8: Trat.	0	0	0	0	0	0	**1**	0	0	0
P8: Paz	2	2	2	3	4	4	**6**	7	8	9
P9: Trat.	0	0	0	0	0	0	0	**1**	0	0
P9: Paz	2	2	2	3	4	4	6	**7**	8	9
P10: Trat.	0	0	0	0	0	0	0	0	**1**	0
P10: Paz	2	2	2	3	4	4	6	7	**8**	9