Analítica de los Negocios

Pruebas de Hipótesis II

Carlos Cardona Andrade

¿Qué deberían saber para el parcial?

1. Cargar un archivo y un paquete en R
2. Los verbos del paquete `tidyverse`: filter, select, arrange, mutate, summarise
3. Describir una distribución (medidas de tendencia central, dispersión)
4. Interpretar una correlación
5. Calcular probabilidades bajo la curva normal
6. ¿Qué es un intervalo de confianza?¿Qué es el nivel de confianza?
7. ¿Qué es una prueba de hipótesis?
    - ¿Cómo se rechaza?
    - ¿Qué implicaciones tiene rechazar o no rechazar?

Plan para hoy

1. La Distribución t
2. Consideraciones al elaborar pruebas de hipótesis
3. Algunas ideas incorrectas sobre la prueba de hipótesis

La Distribución t

William Gosset

"Guinness is the best beer available, it does not need advertising as its quality will sell it, and those who do not drink it are to be sympathized with rather than advertised to." –W.S. Gosset (aka "Student")

Recordemos el TLC

$$\bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

Esto significa que:

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

Como habíamos mencionado antes, el problema radica en que σ usualmente es desconocida para nosotros...

¿Qué pasa si σ es desconocida?

En la práctica, nunca conocemos el valor verdadero de σ, por lo que lo estimamos a partir de nuestros datos con s

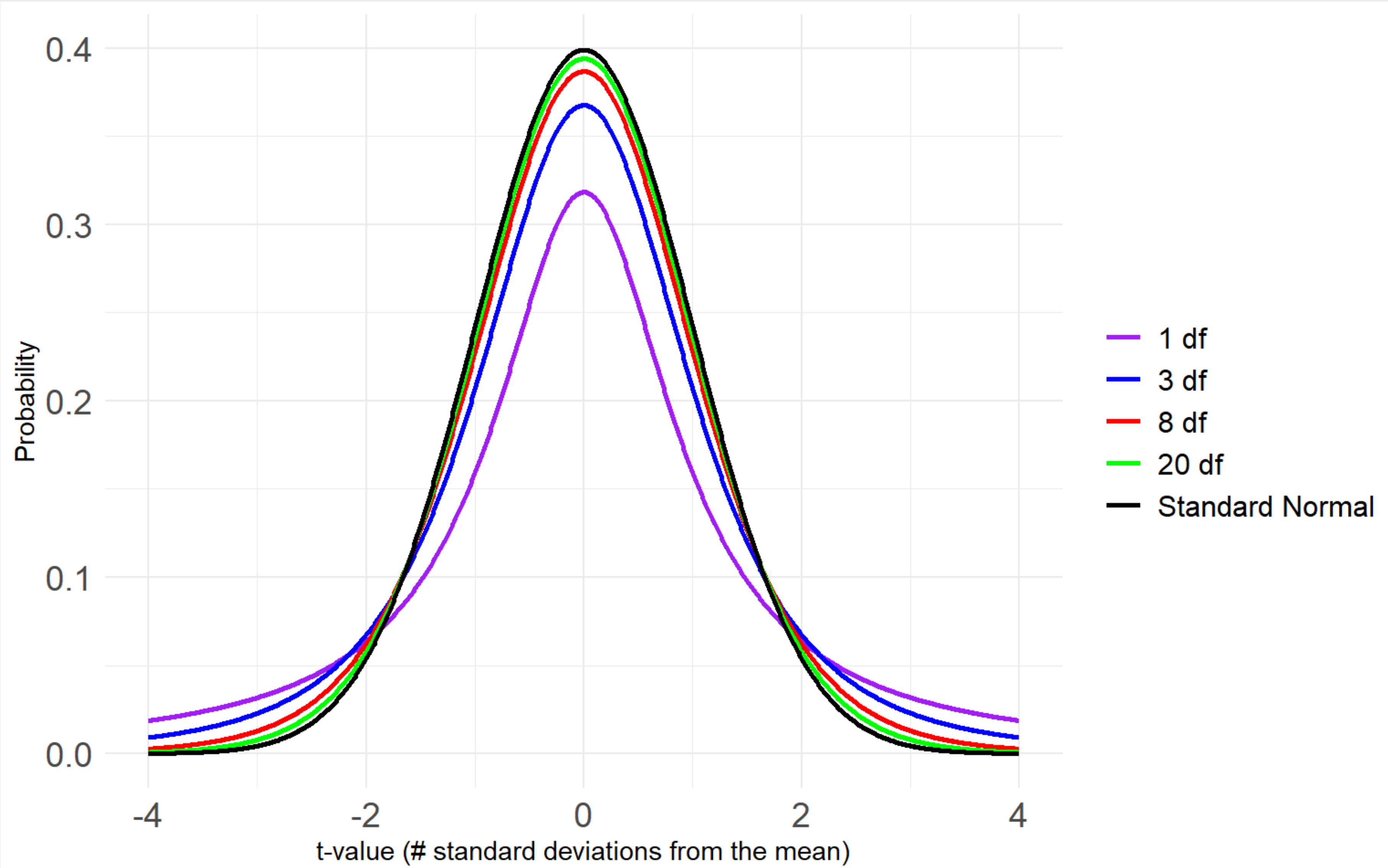
Podemos construir el siguiente estadístico de prueba para contrastar la media poblacional de una sola muestra, la cual sigue una distribución t con $n - 1$ grados de libertad:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$$

La distribución t

- La distribución t también es unimodal y simétrica, y está centrada en 0
- Tiene colas más gruesas que la distribución normal
- Esto compensa la variabilidad adicional introducida al usar s en lugar de σ en el cálculo del error estándar (SE)
- Está definida por los grados de libertad $n - 1$

Dist. Normal vs Dist. t



Distribución *t* en R

¿Cómo encontrar probabilidades?

```r
#P(t < -1.96)
pt(-1.96, df = 9)
```

```
[1] 0.0408222
```

```r
#P(t > -1.96)
pt(-1.96, df = 9,
   lower.tail = FALSE)
```

```
[1] 0.9591778
```

¿Cómo encontrar valores críticos?

```r
# Encontrando Q1
qt(0.25, df = 9)
```

```
[1] -0.7027221
```

```r
# Q3
qt(0.75, df = 9)
```

```
[1] 0.7027221
```

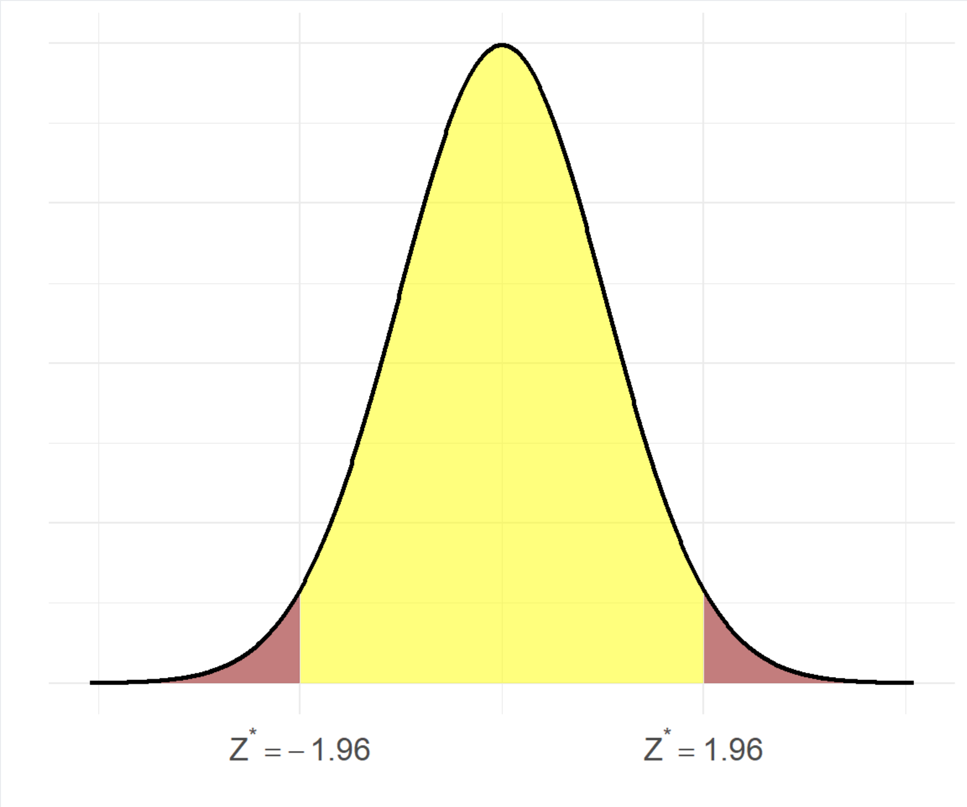
Dist. Normal vs Dist. t (valores críticos)

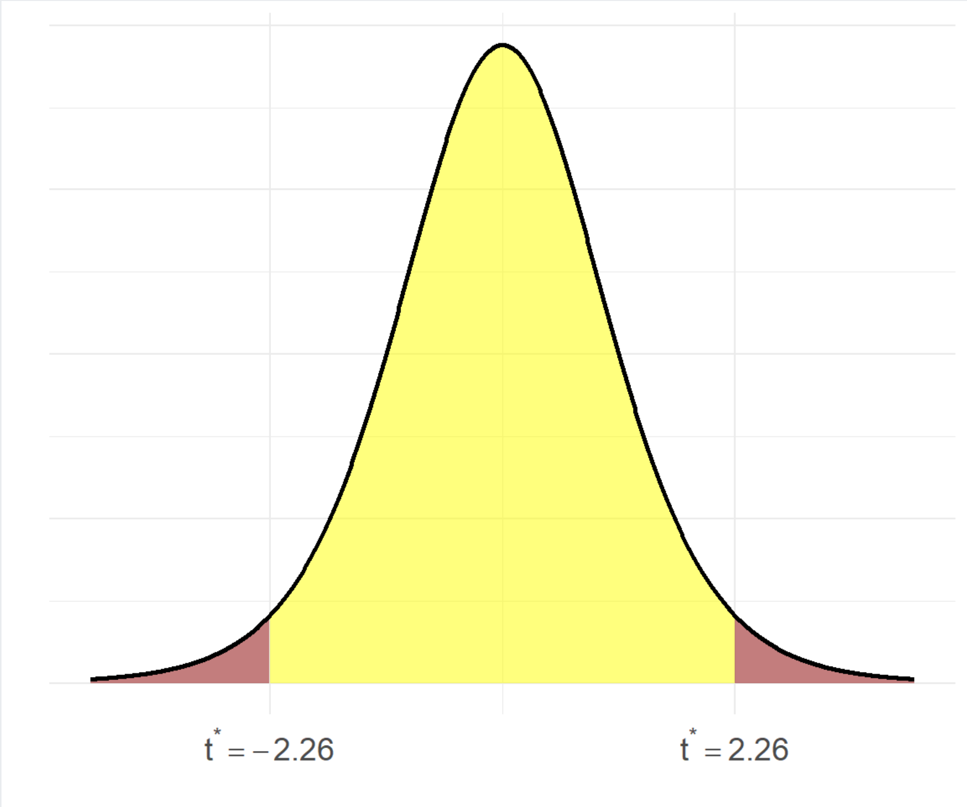
```r
qnorm(0.975)
```

```
[1] 1.959964
```

```r
qt(0.975, df = 9)
```

```
[1] 2.262157
```





Ejemplo: Rendimiento de un bono

Una empresa de gestión de activos recientemente cambió al gerente del fondo que administra un bono de alto rendimiento. Se espera que este cambio tenga un impacto positivo en los retornos del bono. Para evaluar si el cambio de manager ha mejorado significativamente los retornos del bono, se recogieron los retornos mensuales del bono durante 10 meses antes y 10 meses después del cambio de manager.

$$H_0 : \mu_D - \mu_A = 0$$

$$H_A : \mu_D - \mu_A \neq 0$$

Nuestra hipótesis nula es que no existe ningún cambio en el rendimiento promedio del bono con el cambio de gerente ($\mu_D = \mu_A$).

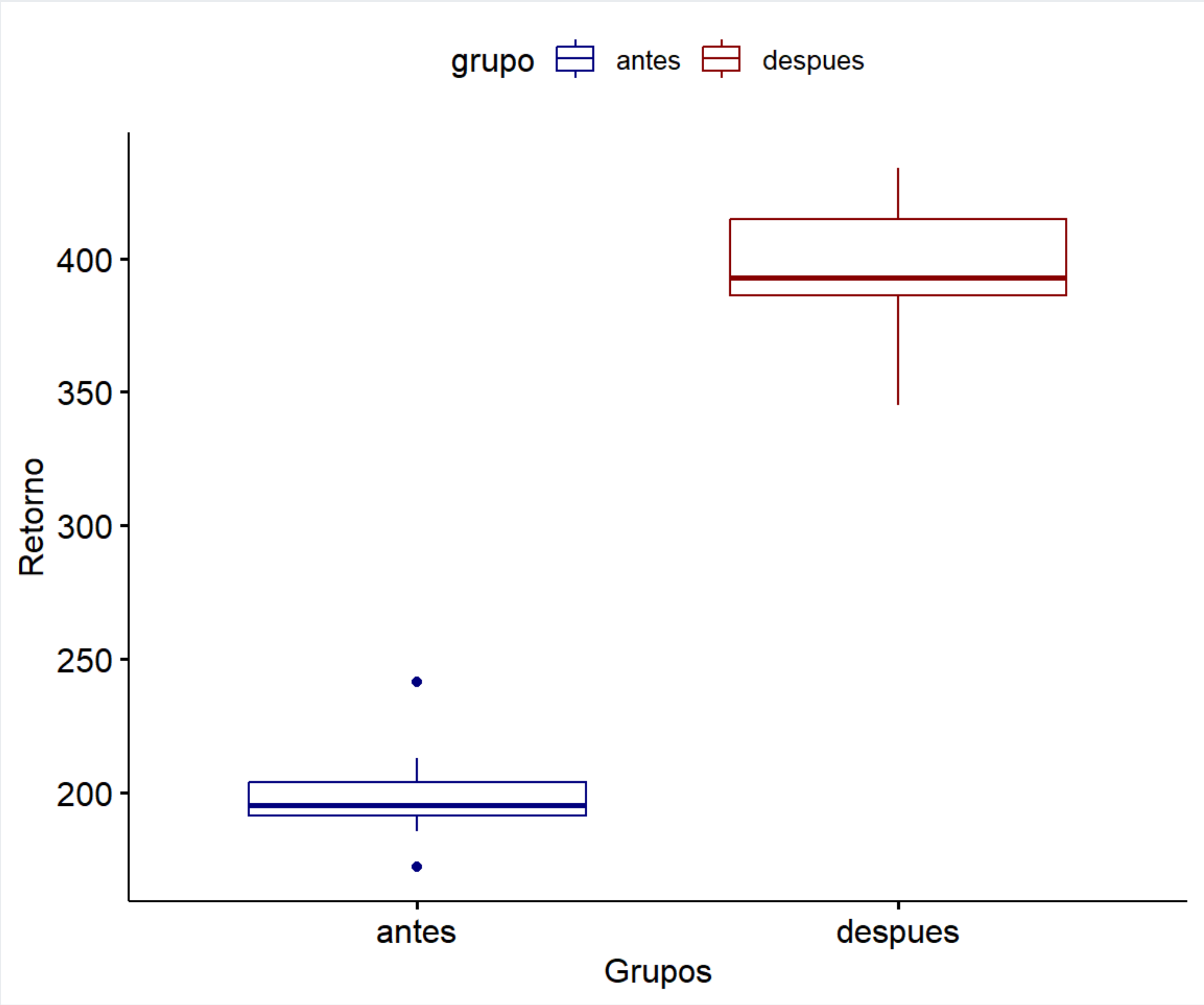
Ejemplo: Rendimiento de un bono

```r
# Retornos del bono antes del cambio de gerente
antes <-c(200.1, 190.9, 192.7, 213, 241.4, 196.9, 172.2, 185.5, 205.2, 193.
# Retornos del bono después del cambio de gerente
despues <-c(392.9, 393.2, 345.1, 393, 434, 427.9, 422, 383.9, 392.3, 352.2)
# Creamos el data frame
bono <- tibble(
                grupo = rep(c("antes", "despues"), each = 10),
                retorno = c(antes,  despues)
                )
bono
```

```
# A tibble: 20 × 2
   grupo   retorno
   <chr>     <dbl>
 1 antes      200.
 2 antes      191.
 3 antes      193.
 4 antes      213
 5 antes      241.
 6 antes      197.
 7 antes      172.
 8 antes      186.
 9 antes      205.
10 antes      194.
11 despues    393.
12 despues    393.
```

```
13 despues     345.
14 despues     393
15 despues     434
16 despues     428
```

Ejemplo: Rendimiento de un bono - Gráfica



Ejemplo: Rendimiento de un bono - t.test()

```r
# Calculemos el test
test_resultado <- t.test(retorno ~ grupo, data = bono, conf.level=0.95)
test_resultado
```

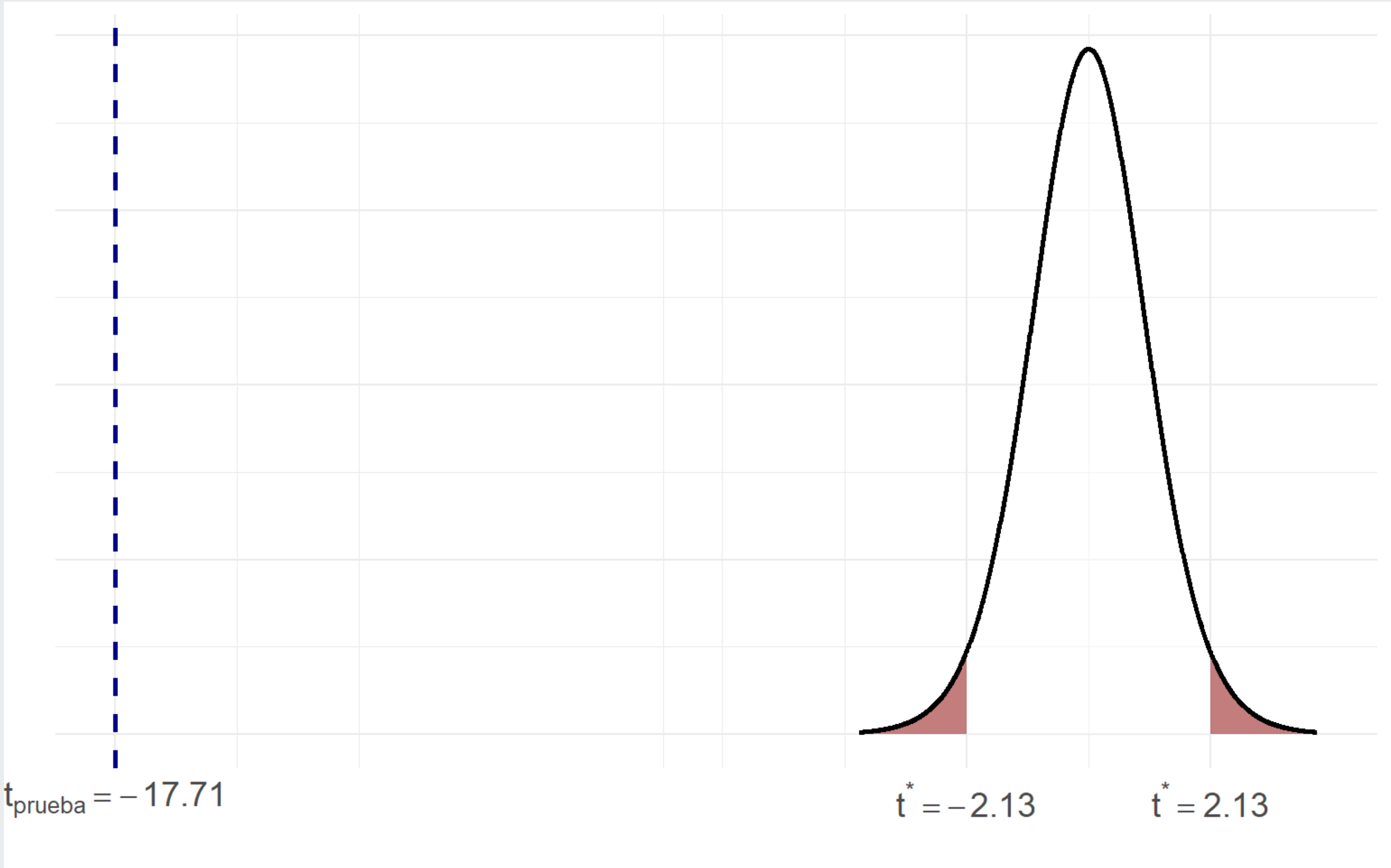
```

	Welch Two Sample t-test

data:  retorno by grupo
t = -17.714, df = 15.149, p-value = 1.544e-11
alternative hypothesis: true difference in means between group antes and group
despues is not equal to 0
95 percent confidence interval:
 -217.8724 -171.1076
sample estimates:
  mean in group antes mean in group despues 
               199.16                393.65 
```

¿Cuál es el valor crítico t^*?

```r
qt(0.975, df = 15)
```

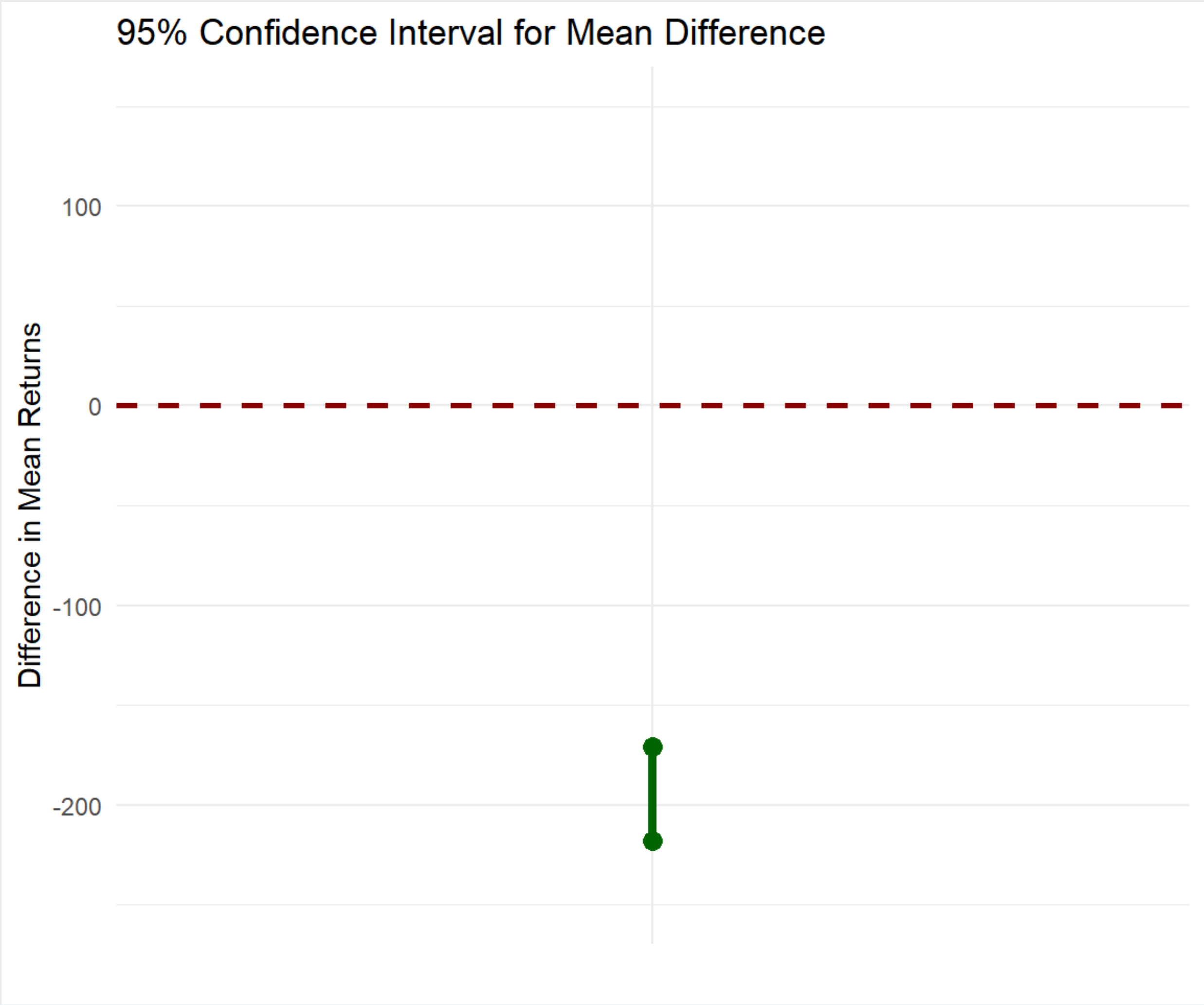
```
[1] 2.13145
```

Ejemplo - Decisión basada en el estadístico t



Rechazamos H_0 porque $t < t^*_{\alpha=0.05}$

Ejemplo - Decisión basada en el IC



Rechazamos la H_0 porque el IC no incluye el 0

Ejemplo - Decisión basada en el p-value

$$¿p < \alpha?$$

```
# p-value
test_resultado$p.value<0.05
```

```
[1] TRUE
```

Rechazamos porque $p < \alpha = 0.05$

Consideraciones al elaborar pruebas de hipótesis

Consecuencias de los errores Tipo I y II

Los errores de tipo I y tipo II son diferentes tipos de equivocaciones y tienen consecuencias distintas:

- H_0 representa el **status quo**, lo que generalmente creemos cierto
- No rechazar H_0 significa que el status quo se mantiene
- Rechazar H_0 indica que se refuta algo previamente creído, como un avance científico o una nueva estrategia de inversión
- Un error de tipo I puede llevar a conclusiones falsas y desperdicio de recursos hasta que se invalide el hallazgo

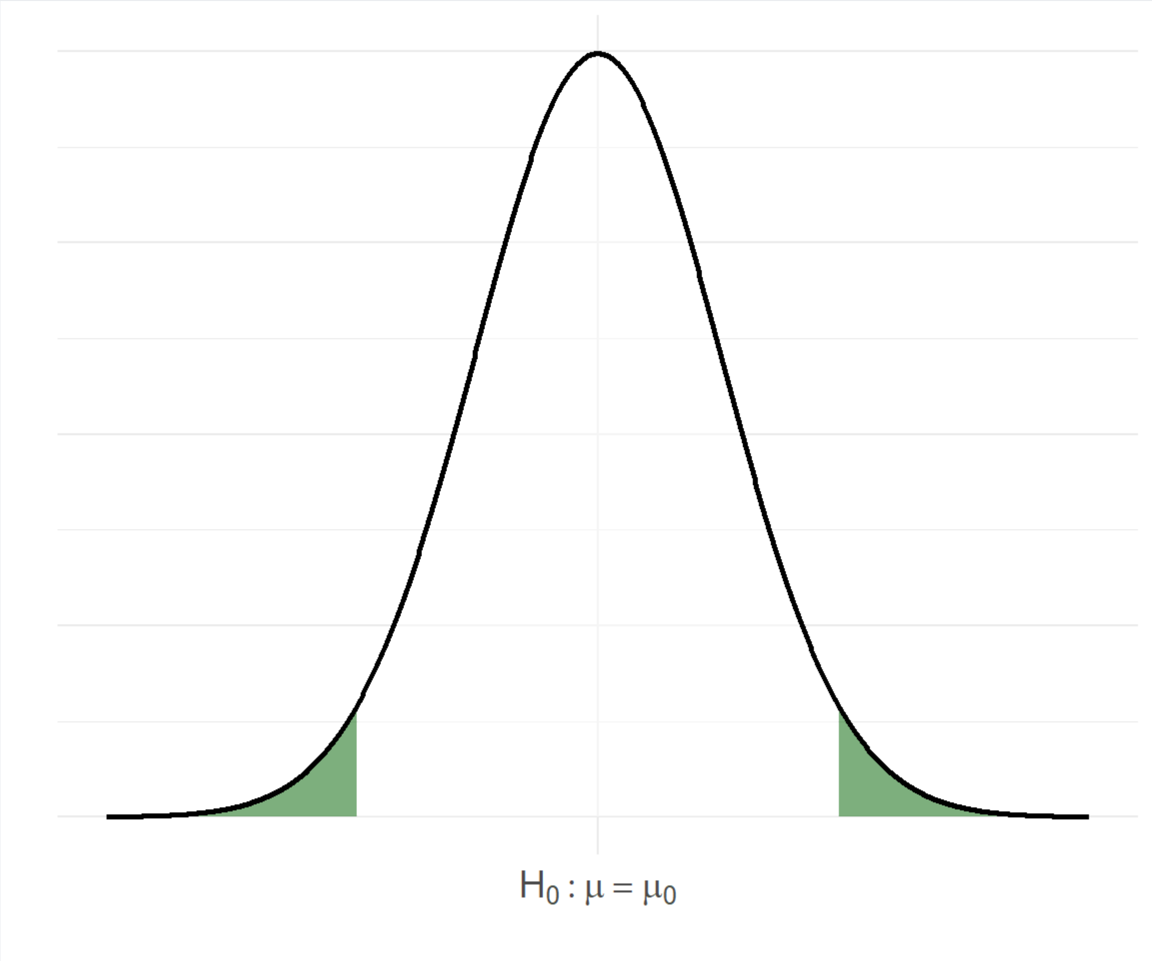
Consecuencias de los errores Tipo I y II

Un error de tipo II (no reconocer un avance científico o una nueva estrategia financiera) representa una oportunidad perdida para el progreso científico o para la empresa.

- Los errores de tipo II también pueden ser costosos, pero generalmente pasan desapercibidos
- Por eso, es más importante controlar la tasa de error de tipo I que la tasa de error de tipo II

¿Cuál es la probabilidad del error Tipo I?

¿Cuál es la probabilidad de rechazar H_0?



Es α!

Nivel de significancia = p(Error Tipo I)

Esto significa que, si H_0 es verdadera, solo hay un $(100 \times \alpha)$% de cometer el error Tipo I!

$$P(\text{Error de tipo I} \mid H_0 \text{ verdadera}) = \alpha$$

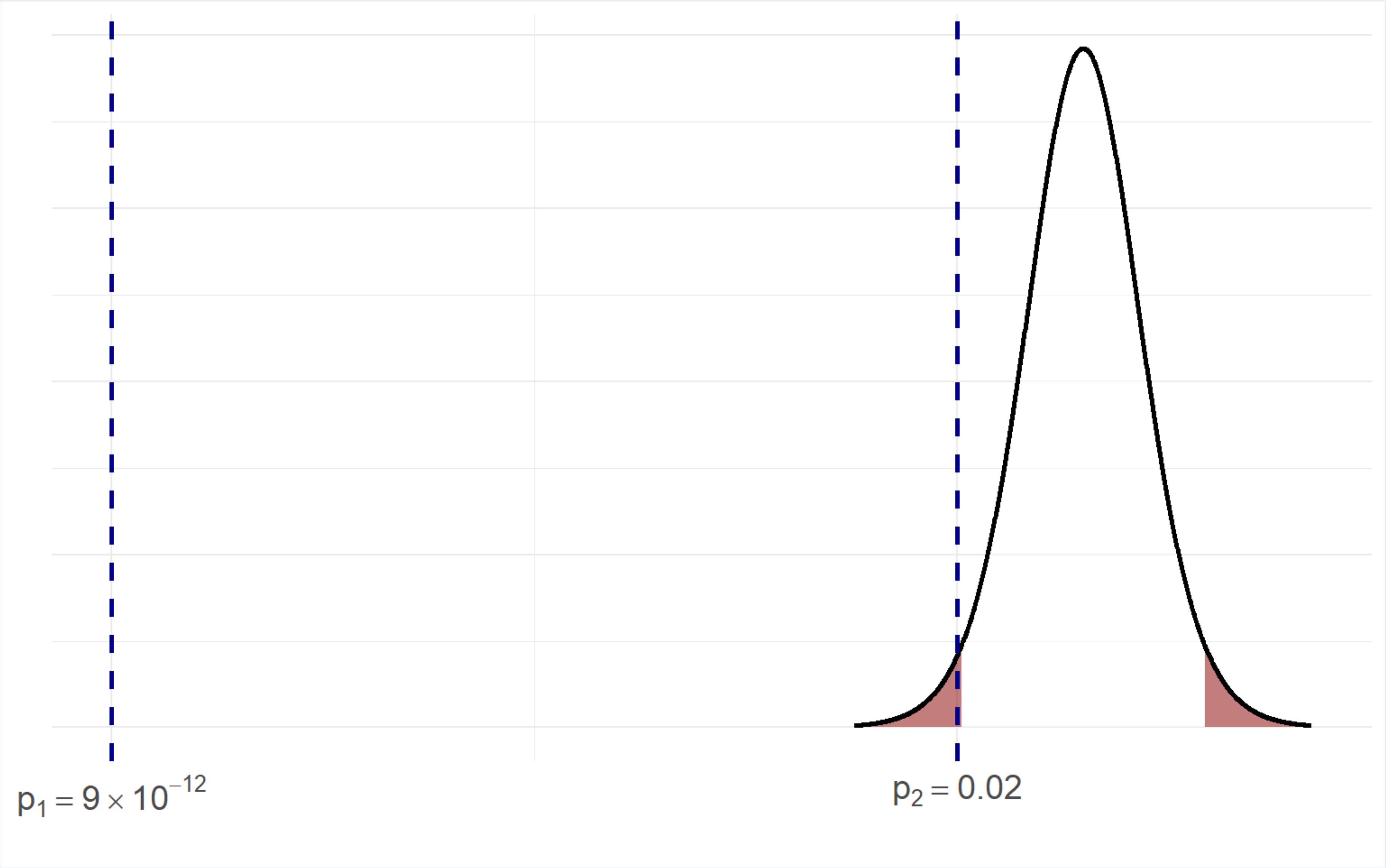
- Por eso preferimos valores pequeños de α — aumentar α incrementa la tasa de error de tipo I
- Sin embargo, el nivel de significancia no controla la tasa de error de tipo II

Reportando el p-value

No se limiten a reportar la conclusión de si se rechaza H_0. Muestren el p-value.

- Un p-value de 0.04 y un p-value de 0.000001 no son lo mismo. Aunque H_0 se rechace en ambos casos, la fuerza de la evidencia es muy diferente
- Concluir simplemente si H_0 es rechazada sin el p-value es como concluir que la temperatura es “fría” o “caliente”
- Es mucho mejor reportar el p-value y permitir que la gente elija su propio nivel de significancia. Es similar a decirle a alguien la temperatura y dejar que decida cómo interpretarla

Reportando el p-value:¿Cuál gerente preferirían?



Algunas ideas incorrectas sobre la prueba de hipótesis

El método científico: prueba y refutación

- Hay una verdad sutil pero fundamental en el método científico, y es que nunca se puede realmente probar una hipótesis con él, solo `refutar` la hipótesis
- En palabras de Albert Einstein:

> “No amount of experimentation can ever prove me right; a single experiment can prove me wrong.”

- Por lo tanto, nunca decimos que la hipótesis nula es verdadera
- Cuando la evidencia no es lo suficientemente fuerte como para rechazar la nula, no decimos “aceptamos la hipótesis nula”, sino que decimos “no podemos rechazar la hipótesis nula”

Fallar al rechazar H_0 no prueba que H_0 sea cierta

Un error común es concluir a partir de un p-value alto que la H_0 es probablemente verdadera.

- **p-value bajo**: evidencia en contra de H_0
- **p-value alto**: no podemos concluir que H_0 es verdadera; podríamos cometer un error de tipo II
- La tasa de error de tipo II es mayor que la de tipo I, que se controla.
- Cuando no rechazamos H_0, a menudo significa que los datos no son capaces de distinguir entre H_0 y H_A (porque los datos son demasiado ruidosos, etc.).

Ejemplo: Dietas y cáncer

- Women's Health Initiative encontró que las dietas bajas en grasa reducen el riesgo de cáncer de seno con un p-value de **0.07**
- El titular del New York Times: "Estudio encuentra que las dietas bajas en grasa no detendrán el cáncer"
- El editorial principal afirmó que el estudio presentaba "evidencia sólida de que la guerra contra las grasas fue en vano" y añadió "este es el fin para la creencia de que reducir el porcentaje de grasa total en la dieta es importante para la salud"
- No encontrar evidencia del efecto no significa que las dietas bajas en grasa no tengan ningún efecto

No tomen la significancia al 0.05 demasiado enserio

- Un p-value de 0.049 y un p-value de 0.051 ofrecen casi la misma evidencia contra H_0
- Por ejemplo, un estudio famoso de 2009 sobre una vacuna que podría proteger contra el VIH reportó un p-value a dos colas de 0.08, mientras que el p-value a una cola fue 0.04
- Se desató mucho debate y controversia, en parte porque las dos formas de analizar los datos producen p-values a ambos lados de 0.05
- Gran parte de este debate y controversia es bastante inútil; ambos p-values te dicen esencialmente lo mismo: que la vacuna tiene potencial, pero que los resultados aún no son concluyentes

Las pruebas de hipótesis no pueden decirnos...

- si el diseño de un estudio está defectuoso
- si los datos se han recolectado adecuadamente

Por lo tanto, no podemos concluir a partir de un p-value pequeño si la conclusión se puede generalizar a una población más grande

$$\text{Garbage In} \rightarrow \text{Garbage Out}$$

Significancia estadística no significa importancia práctica

Otro error es leer demasiado en el término **estadísticamente significativo**.

- Decir que los resultados son estadísticamente significativos informa al lector que los hallazgos son poco probables de ser resultado del azar.
- Sin embargo, no dice nada sobre la importancia práctica del hallazgo.
- E.g., rechazar $H_0 : \mu_1 = \mu_2$ solo nos dice que $\mu_1 \neq \mu_2$, pero no qué tan grande o importante es $\mu_1 - \mu_2$. Puede ser que la diferencia no sea relevante por ser muy pequeña a pesar de ser significativa.
- Solución: `Reporten un intervalo de confianza` del parámetro para que la gente pueda decidir si la diferencia es lo suficientemente grande como para ser relevante.

Ejemplo: ¿Cuál es la diferencia de calidad entre el café colombiano y el resto?

$$H_0 : \mu_{\text{Colombia}} - \mu_{\text{No Colombia}} = 0$$

```
    Welch Two Sample t-test

data:  total_cup_points by colombia
t = -8.2894, df = 453.01, p-value = 1.304e-15
alternative hypothesis: true difference in means between group FALSE and group
TRUE is not equal to 0
95 percent confidence interval:
 -1.3659680 -0.8424154
sample estimates:
mean in group FALSE  mean in group TRUE
           82.00237            83.10656
```

En este caso, rechazamos la H_0 de que sean iguales pero, ¿es la diferencia de calidad considerable?

En resumen...

- Rechazar H_0 no significa que estamos 100% seguros que H_0 es falsa
- El p-value no es la probabilidad de que H_0 sea verdadera.
- No tomen el nivel de significancia de 0.05 demasiado en serio.
- Las pruebas de hipótesis no pueden decirnos si los datos se recolectaron adecuadamente o si el diseño de un estudio es malo.
- La significancia estadística no se traduce en importancia práctica