

Hola

Sobre tests de hipótesis con pocos datos

ENUNCIADO

La startup en la que trabajás tiene como objetivo aumentar el *ticket promedio* de las transacciones semanales de su aplicación a más de 500 USD.

ENUNCIADO

La startup en la que trabajás tiene como objetivo aumentar el *ticket promedio* de las transacciones semanales de su aplicación a más de 500 USD.

Tras implementar una serie de mejoras a lo largo de los últimos meses, han recopilado datos sobre el *ticket promedio* de las últimas 16 semanas, obteniendo un valor de 540 USD, y una desviación estándar muestral de 80 USD.

ENUNCIADO

La startup en la que trabajás tiene como objetivo aumentar el *ticket promedio* de las transacciones semanales de su aplicación a más de 500 USD .

Tras implementar una serie de mejoras a lo largo de los últimos meses, han recopilado datos sobre el *ticket promedio* de las últimas 16 semanas , obteniendo un valor de 540 USD , y una desviación estándar muestral de 80 USD .

Si bien este resultado es alentador, es crucial determinar si el incremento observado refleja un impacto real de las mejoras implementadas o si podría deberse a una fluctuación estadística en el comportamiento del mercado durante ese período.

ENUNCIADO

La startup en la que trabajás tiene como objetivo aumentar el *ticket promedio* de las transacciones semanales de su aplicación a más de 500 USD .

Tras implementar una serie de mejoras a lo largo de los últimos meses, han recopilado datos sobre el *ticket promedio* de las últimas 16 semanas , obteniendo un valor de 540 USD , y una desviación estándar muestral de 80 USD .

Si bien este resultado es alentador, es crucial determinar si el incremento observado refleja un impacto real de las mejoras implementadas o si podría deberse a una fluctuación estadística en el comportamiento del mercado durante ese período.

Resuelva el problema.

ENUNCIADO

La startup en la que trabajás tiene como objetivo aumentar el *ticket promedio* de las transacciones semanales de su aplicación a más de 500 USD.

Tras implementar una serie de mejoras a lo largo de los últimos meses, han recopilado datos sobre el *ticket promedio* de las últimas 16 semanas, obteniendo un valor de 540 USD, y una desviación estándar muestral de 80 USD.

Si bien este resultado es alentador, es crucial determinar si el incremento observado refleja un impacto real de las mejoras implementadas o si podría deberse a una fluctuación estadística en el comportamiento del mercado durante ese período.

Resuelva el problema.

ENUNCIADO

La startup en la que trabajás tiene como objetivo aumentar el *ticket promedio* de las transacciones semanales de su aplicación a más de 500 USD.

Tras implementar una serie de mejoras a lo largo de los últimos meses, han recopilado datos sobre el *ticket promedio* de las últimas 16 semanas, obteniendo un valor de 540 USD, y una desviación estándar muestral de 80 USD.

Si bien este resultado es alentador, es crucial determinar si el incremento observado refleja un impacto real de las mejoras implementadas o si podría deberse a una fluctuación estadística en el comportamiento del mercado durante ese período.

Resuelva el problema.

VARIABLES RELEVANTES PARA EL PROBLEMA

VARIABLES RELEVANTES PARA EL PROBLEMA

- Variable Aleatoria (X): *ticket promedio* de las transacciones semanales.
- Tamaño de la muestra (n): 16 semanas.
- Media muestral (\bar{X}): 540 USD .
- Desviación estándar muestral (s): 80 USD .
- Valor medio objetivo (μ_0): 500 USD .

VARIABLES RELEVANTES PARA EL PROBLEMA

- Variable Aleatoria (X): *ticket promedio* de las transacciones semanales.
- Tamaño de la muestra (n): 16 semanas.
- Media muestral (\bar{X}): 540 USD .
- Desviación estándar muestral (s): 80 USD .
- Valor medio objetivo (μ_0): 500 USD .
- Hipótesis nula (μ_0): $\mu \leq 500$ USD
(el *ticket promedio* semanal es menor a 500 USD)
- Hipótesis alternativa (μ_1): $\mu > 500$ USD
(el *ticket promedio* semanal es mayor o igual a 500 USD)

ELECCIÓN DE LA DISTRIBUCIÓN

ELECCIÓN DE LA DISTRIBUCIÓN

Como nuestra variable aleatoria, el *ticket promedio de las transacciones semanales*, es una variable aleatoria producto de muchísimas variables aleatorias independientes, es razonable asumir que ésta sigue una distribución normal: $X \sim N(\mu, \sigma)$.

ELECCIÓN DE LA DISTRIBUCIÓN

Como nuestra variable aleatoria, el *ticket promedio de las transacciones semanales*, es una variable aleatoria producto de muchísimas variables aleatorias independientes, es razonable asumir que ésta sigue una distribución normal: $X \sim N(\mu, \sigma)$.

Si tuviésemos la muestra de *infinitos* meses (más de 30), podríamos tener una sólida estimación de σ_0 y asumir que

$$z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0, 1).$$

ELECCIÓN DE LA DISTRIBUCIÓN

Como nuestra variable aleatoria, el *ticket promedio de las transacciones semanales*, es una variable aleatoria producto de muchísimas variables aleatorias independientes, es razonable asumir que ésta sigue una distribución normal: $X \sim N(\mu, \sigma)$.

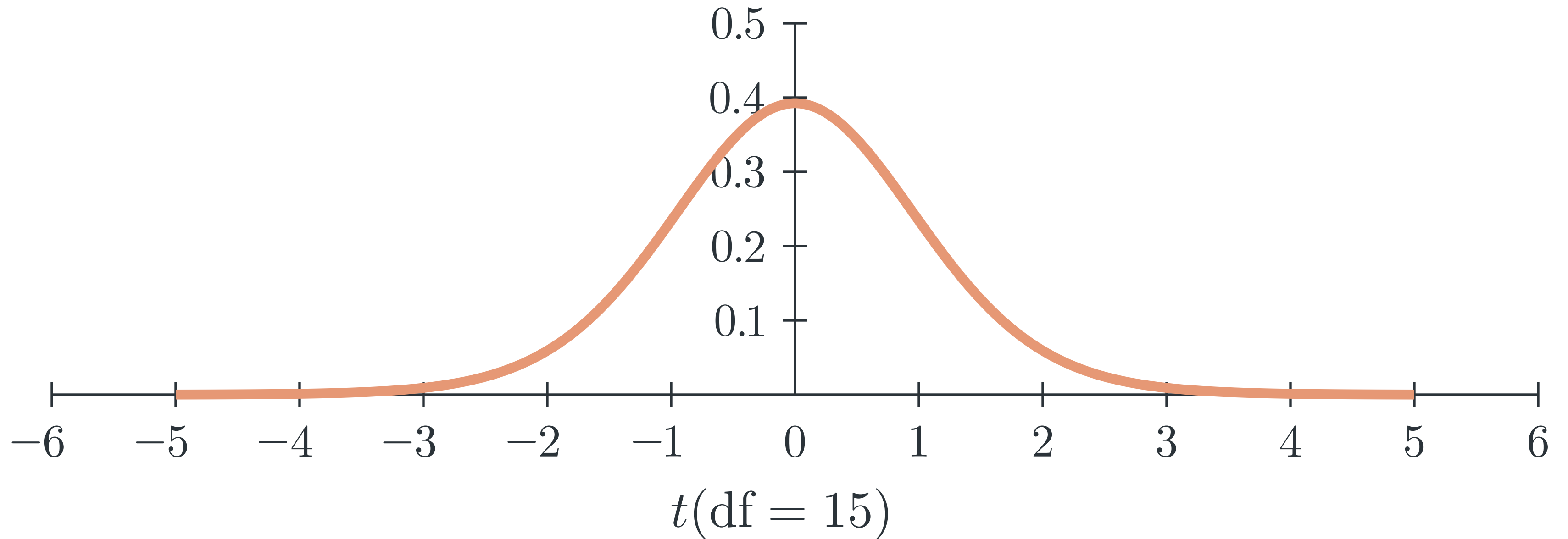
Si tuviésemos la muestra de *infinitos* meses (más de 30), podríamos tener una sólida estimación de σ_0 y asumir que

$$z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0, 1).$$

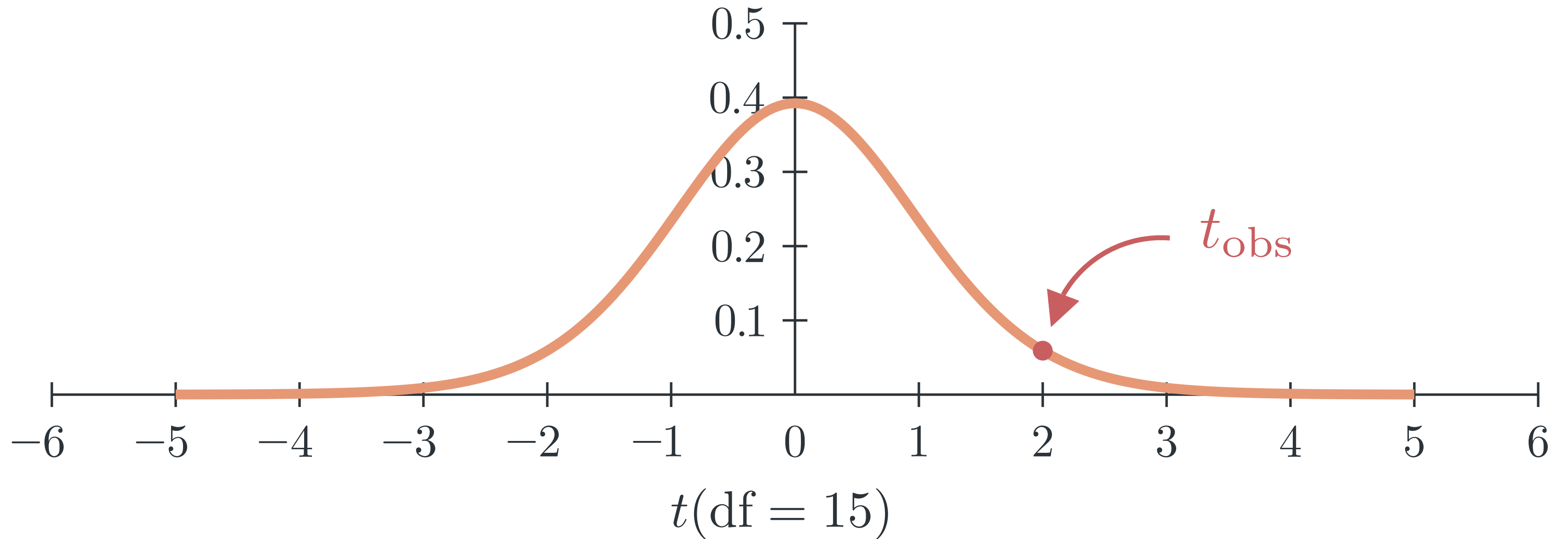
Pero como tenemos pocas muestras, tenemos que estimar σ , lo que se va a hacer que la distribución de nuestro estadístico sea otra; con colas más pesadas:

$$t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}} \sim t(n - 1).$$

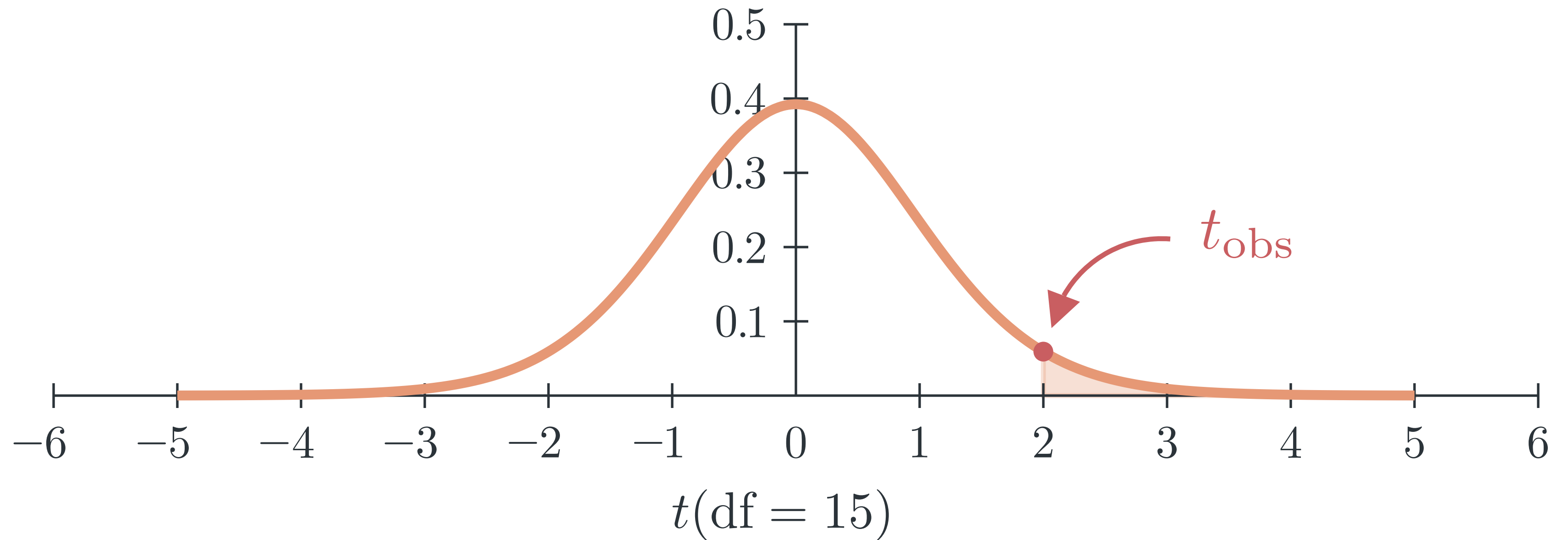
DISTRIBUCIÓN T-STUDENT



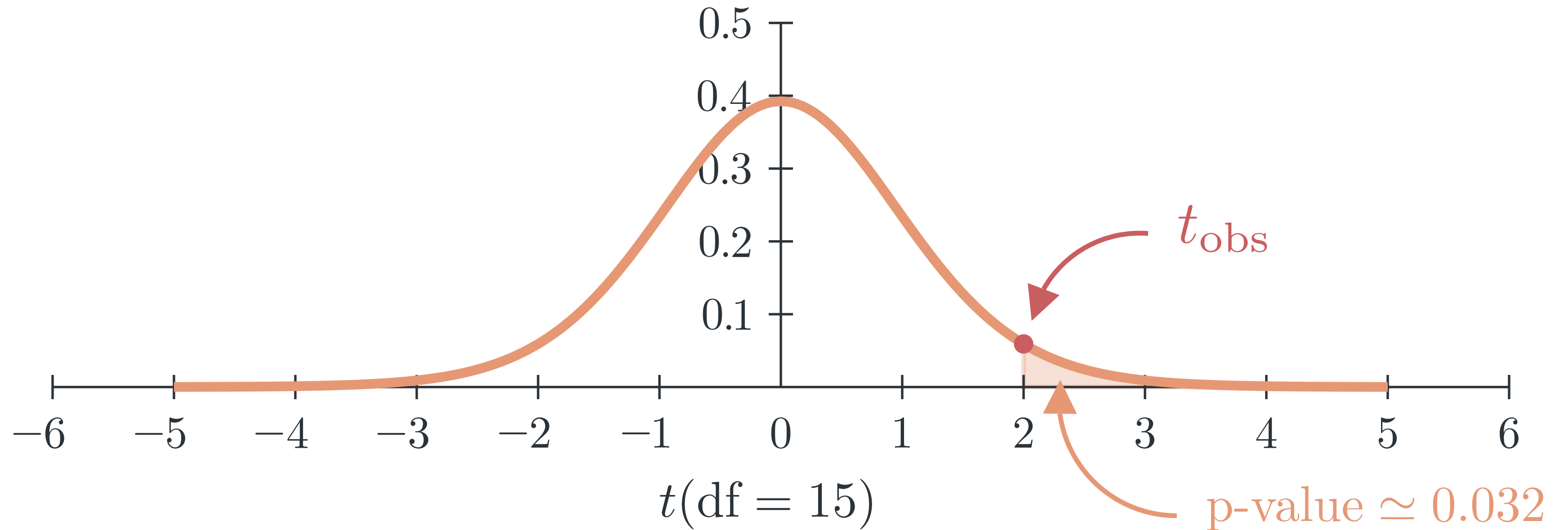
DISTRIBUCIÓN T-STUDENT



DISTRIBUCIÓN T-STUDENT



DISTRIBUCIÓN T-STUDENT



DISTRIBUCIÓN T-STUDENT EN R

```
# Graficar la t-student de 15 grados de libertad
x <- seq(-6, 6, length.out = 400)
y <- dt(x, df=15)
plot(x, y)

# Calcular e imprimir el p-value para 2 a cola derecha
p_val = pt(2, df=15, lower.tail=FALSE)
print(p_val)
```

UNIVERSOS ALTERNATIVOS

Si el valor real del *ticket promedio* fuese exactamente 500 USD y se repitiera este experimento en 10^4 universos alternativos en esta misma circunstancias; es decir, si se obtuviese el valor del ticket promedio para 10^4 conjuntos de 16 semanas independientes.

¿Cuántas de estas veces obtendría un valor de más de 540 USD ?

EL ERROR DE USAR UNA NORMAL

EL ERROR DE USAR UNA NORMAL

Tabla de diferencias normal/t-student (en %)

| Precio/ n | 3 | 10 | 16 | 30 | 50 |
|-------------|------|--------|---------|---------|----------|
| 510 USD | 2.5% | 0.6% | 0.36% | 0.19% | 0.11% |
| 540 USD | 6.9% | 1.6% | 0.92% | 0.47% | 0.28% |
| 600 USD | 1.9% | 0.037% | 0.0079% | 0.0012% | 0.00036% |

Cómo leer la tabla. Ej: Para 540 USD y $n = 16$ el p-value asociado a la distribución normal es 2.28% y el de la distribución t-student 3.2% , dando una diferencia de 0.92% entre ellas.

OTROS PROBLEMAS

que en realidad son el mismo problema

OTROS PROBLEMAS

que en realidad son el mismo problema

- ¿Bajó el tiempo promedio de respuesta de una aplicación móvil?
- ¿Aumentó el monto promedio de depósitos en una billetera digital?
- ¿Cambió el número promedio de usuarios activos diarios?
- ¿Aumentó el número de compras in-app?
- ¿Cambió el número promedio de transacciones fallidas en una pasarela de pago?
- ¿Aumentó el monto promedio de inversión en una plataforma de crowdfunding?
- ¿Se redujo el KLOC en el desarrollo de un software?
- ¿Mejóro la calificación promedio de satisfacción del cliente en soporte técnico?
- ¿Aumentó la tasa de aprobación promedio de Inferencia Estadística este cuatrimestre?

Eso es todo