Contrastes de hipótesis

Contenidos adaptados del libro "Probability and statistics for computer scientists, Second edition, M. Baron" (Capítulo 9.4)

Contenido

1. Objetivos

- Conocer el papel de los contrastes de hipótesis
- Determinar en qué situaciones hay que aplicar un contraste paramétrico o no paramétrico
- Realizar contrastes de hipótesis para la media poblacional y una proporción poblacional comprobando la región de aceptación
- Realizar contrastes de hipótesis para la media poblacional y una proporción poblacional calculando el p-valor
- Realizar contrastes de hipótesis para la diferencia de medias y de proporciones de dos poblaciones independientes comprobando la región de aceptación
- Realizar contrastes de hipótesis para la diferencia de medias y de proporciones de dos poblaciones independientes calculando el p-valor

2. Introducción

- Uno de los posibles resultados de la inferencia estadística es la comprobación o contraste de una hipótesis sobre los valores para los parámetros estimados, junto con una probabilidad de error cometido.
- Se suelen utilizar sinónimos de "contraste de hipótesis", como "prueba de hipótesis" o "test de hipótesis".

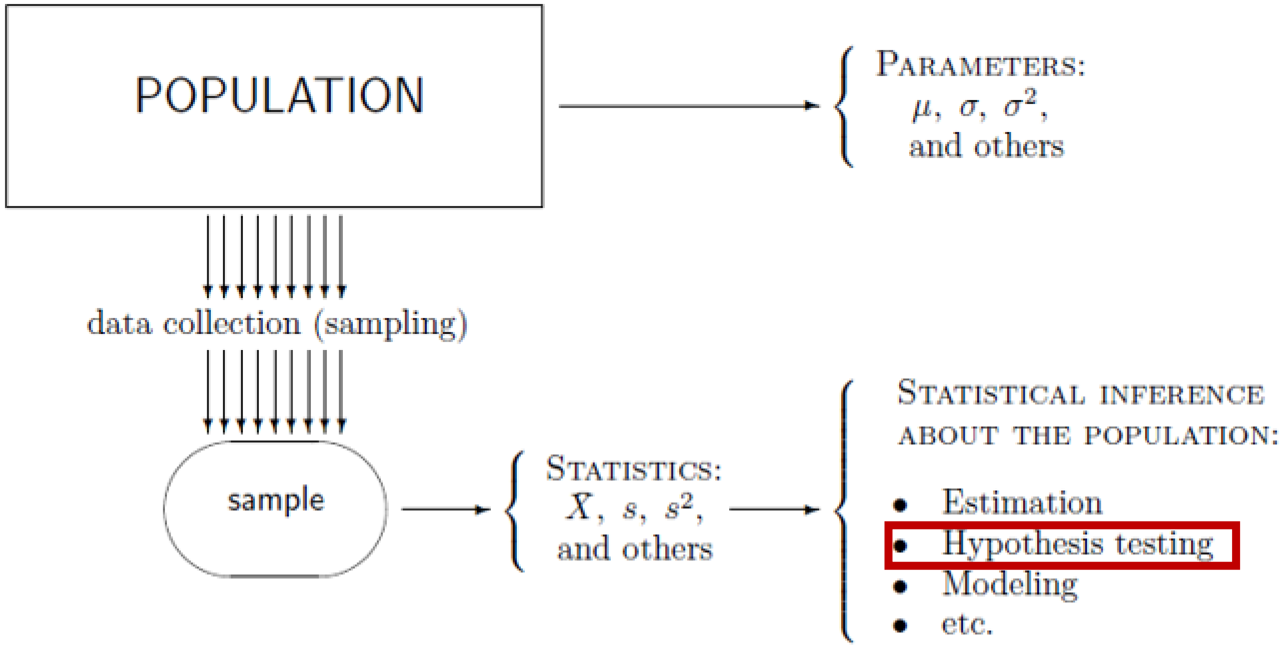


FIGURE 8.1: *Population parameters and sample statistics.*

2. Introducción
Contrastes paramétricos y no paramétricos

- Un contraste de hipótesis puede ser
  - Paramétrico: cuando los datos tienen una distribución de probabilidad Normal.
  - No Paramétrico: Cuando no hay evidencias de que los datos tengan una distribución Normal. Por ejemplo, porque la muestra sea pequeña o tenga datos atípicos (*outliers*)
- Comprobación de la normalidad
  - Hay diferentes <u>métodos para comprobar la normalidad de los datos</u>
    - Chi-Square, Kolmogorov-Smirnov, Lilliefors, Anderson-Darling, Shapiro-Wilk, Cramér-von-Mises
  - El software estadístico (R, SPSS, Statgraphics, Excel, …) ofrece funcionaldad para comprobar la normalidad
  - En los ejercicios no será necesario comprobar la normalidad, ya que indicará si los datos de cada ejercicio son normales o no, pero es recomendable utilizar un programa estadístico para confirmarlo.
  - Cuando una muestra es grande (>30), se puede considerar Normal, aplicando el Teorema Central del Límite.

2. Introducción
Planteamiento de un contraste de dos hipótesis

- En el contexto de la estadística inferencial, una hipótesis es una afirmación sobre el valor de un parámetro (θ) de una o varias poblaciones
- Cuando se plantea un contraste, se enfrentan dos hipótesis
  - H_0 versus (contra) H_A
  - H_0: Hipótesis, o Hipótesis nula (es la que se quiere comprobar)
  - H_A: Alternativa, o Hipótesis alternativa
- Ambas son declaraciones mutuamente excluyentes.
- El resultado de un contraste de hipótesis es la aceptación de H_0 o su rechazo en favor de H_A.
- Al realizar el contraste con una muestra de datos, no con la población completa, nunca se podrá decir que H_0 o H_A sean ciertas, sólo se puede decir
  - si hay suficientes evidencias como para rechazar H_0 en favor de H_A,
  - o si no hay suficientes evidencias como para rechazar H_0 en favor de H_A

2. Introducción Planteamiento de un contraste. Ejemplo 9.22

- Para verificar que la velocidad de conexión promedio en una población de hogares es de 54 Mbps, se puede plantear el siguiente contraste de hipótesis sobre el parámetro media poblacional (μ)
  - $H_0: \mu = 54Mbps$ vs $H_A: \mu \neq 54Mbps$
- Pero si lo que nos interesa es comprobar si la conexión es inferior a ese valor medio, se puede plantear un contraste
  - $H_0: \mu = 54Mbps$ vs $H_A: \mu < 54Mbps$
- En ambos casos,
  - si no encontramos evidencias que respalden H_0, entonces se rechazará H_0
  - si encontramos evidencias que respalden H_0, entonces no se rechazará H_0

2. Introducción
Tipos de contrastes según la hipótesis alternativa

- Contraste bilateral o de dos colas (*two-sided test*)
  - $H_0: \theta = valor$ vs $H_A: \theta \neq valor$
- Contraste uni-lateral o de una cola, con cola izquierda (*one-sided test, left-tail*)
  - $H_0: \theta = valor$ vs $H_A: \theta < valor$
- Contraste uni-lateral o de una cola, con cola derecha (*one-sided, right-tail*)
  - $H_0: \theta = valor$ vs $H_A: \theta > valor$

2. Introducción
Tipos de errores en un contraste de hipótesis

	Decisión: **Rechazar H_0**	**Decisión:** **No rechazar H_0**
Realidad: **Hipótesis H_0 CIERTA**	FALLO Error tipo I	**ACIERTO**
Realidad: **Hipótesis H_0 FALSA**	**ACIERTO**	FALLO Error tipo II

- El error de tipo I ocurre cuando rechazamos una hipótesis nula que es cierta
  - Ejemplo: Decidimos que la altura media de los habitantes de una ciudad no es 1,6m a partir de una muestra de 100 personas, y posteriormente medimos a todos los habitantes y la media es 1,6m
- El error de tipo II ocurre cuando aceptamos una hipótesis nula que es falsa
  - Ejemplo: Decidimos que la altura media de los habitantes de una ciudad es 1,6m a partir de una muestra de 100 personas, y posteriormente medimos a todos los habitantes y la media es 1,7m

2. Introducción
Probabilidades en un contraste de hipótesis

- **Nivel de significación del contraste: α**. Es la probabilidad de rechazar la hipótesis cuando dicha hipótesis es cierta. Error tipo I.
  - $\boldsymbol{\alpha = P\{rechazar\ H_0 | H_0\ es\ cierta\} = Probabilidad\ de\ cometer\ error\ tipo\ I}$
- **Nivel de confianza del contraste: 1-α.** Es la probabilidad de no rechazar la hipótesis cuando dicha hipótesis es cierta.
- **Potencia del contraste**: Es la probabilidad de rechazar la hipótesis cuando dicha hipótesis es falsa.
  - $potencia = P\{rechazar\ H_0 | H_0\ es\ falsa\} = Probabilidad\ de\ no\ cometer\ error\ tipo\ II$
- **Debilidad del contraste**: $1 - potencia$. Es la probabilidad de no rechazar la hipótesis cuando dicha hipótesis es falsa. Error tipo II.

3. Métodos para realizar un contraste de hipótesis

- Existen dos métodos para realizar el contraste de una hipótesis y decidir si se acepta H_0 o, por el contrario, si se acepta H_A:
  - Comprobando la región de aceptación (no rechazo)
  - Calculando el p-valor

3. Métodos para realizar un contraste de hipótesis Comprobando la región de aceptación. Paso 1

Paso 1: Calcular un estadístico de contraste

- Un estadístico de contraste T, es una variable aleatoria que representa al estimador estandarizado (centrado en 0) del parámetro sobre el que se plantean las hipótesis
  - Si la distribución de probabilidad del estadístico es Normal, el estadístico se denomina Z → T=Z
  - Si la distribución de probabilidad del estadístico es T de Student, el estadístico se denomina t → T=t
  - En los contrastes no paramétricos se utilizan otros estadísticos llamados S, W, U
- El estadístico de contraste se utiliza para discriminar entre la hipótesis nula y la alternativa.
  - Primero se obtiene un valor "observado" del estadístico (T_o) a partir de los datos (observados) de la muestra y del valor del parámetro que se quiere comprobar en las hipótesis del contraste, considerando que H_0 es cierta.

3. Métodos para realizar un contraste de hipótesis Comprobando la región de aceptación. Paso 2

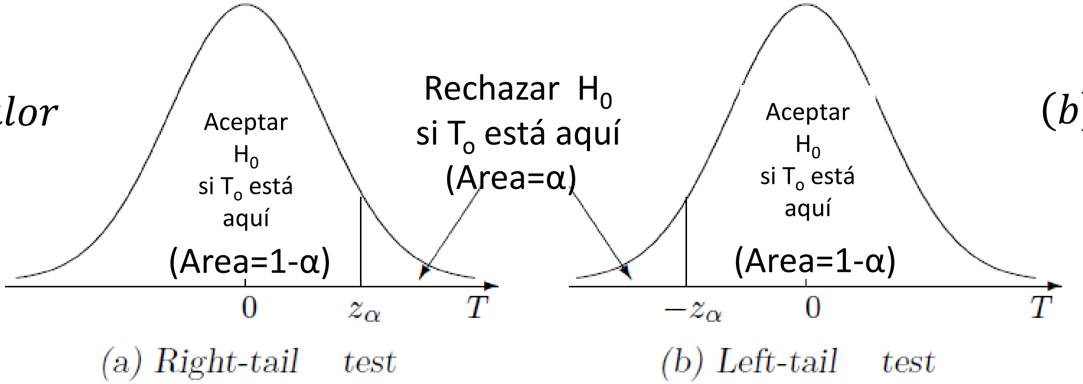
Paso 2: Calcular la región de aceptación

- El intervalo de valores del estadístico de contraste T, debajo del cual en su función de densidad de probabilidad (pdf) hay un área de 1-α, se llama **región de aceptación**.
- La parte restante se llama región de rechazo, con área α.
- Entonces
  - $P\{T \in Región\ de\ aceptación \mid H_0\ cierta\} = 1 - \alpha = Nivel\ de\ confianza$
  - $P\{T \in Región\ de\ rechazo \mid H_0\ cierta\} = \alpha = Nivel\ de\ significación$

3. Métodos para realizar un contraste de hipótesis Comprobando la región de aceptación. Gráficos

$(a)\ H_0: \theta = valor$ vs $H_A: \theta > valor$

$(b)\ H_0: \theta = valor$ vs $H_A: \theta < valor$



$(c)\ H_0: \theta = valor$ vs $H_A: \theta \neq valor$

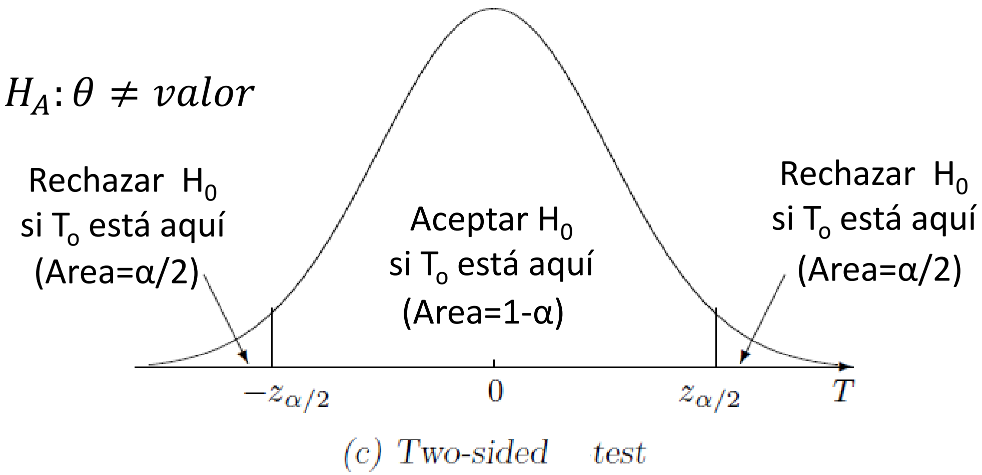


FIGURE 9.7: *Acceptance and rejection regions for a test with (a) a one-sided right-tail alternative; (b) a one-sided left-tail alternative; (c) a two-sided alternative.*

3. Métodos para realizar un contraste de hipótesis Comprobando la región de aceptación. Paso 3

Paso 3: Obtener e interpretar el resultado

- Resultado
  - Aceptar la hipótesis H_0 si el valor "observado" del estadístico de contraste T_o pertenece a la región de aceptación.
  - Rechazar H_0 en favor de la alternativa H_A si T_o no pertenece a la región de aceptación.
- Interpretación
  - Si se acepta la hipótesis, sólo significa que las evidencias obtenidas de los datos de la muestra no son suficientes para rechazarla.
  - Si se rechaza la hipótesis, todo lo que podemos afirmar es que los datos proporcionan evidencias suficientes contra H_0 y a favor de H_A.

3. Métodos para realizar un contraste de hipótesis
Calculando el p-valor

- Se define el **nivel crítico p del contraste** (o **p-valor**) como la probabilidad de obtener un valor del estadístico de contraste (T) tanto o más "extremo" que el valor "observado" del estadístico (T_o) que se ha calculado con los datos de la muestra, suponiendo que la hipótesis H_0 es cierta.
  - $\text{p}-\text{valor} = P\{|T| \geq |T_o|\}$
- Por tanto, p-valores pequeños nos llevarán a rechazar la hipótesis nula, concretamente en el caso que sea menor o igual que el nivel de significación (α) establecido de antemano.
  - Si p-valor > α, entonces se acepta la hipótesis nula H_0
  - Si p-valor < α, entonces se rechaza la hipótesis nula H_0

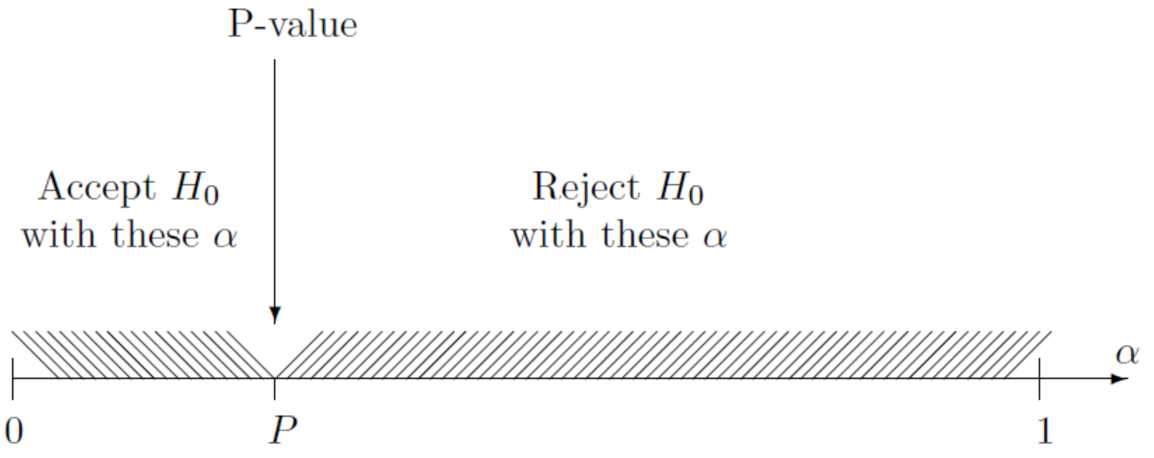
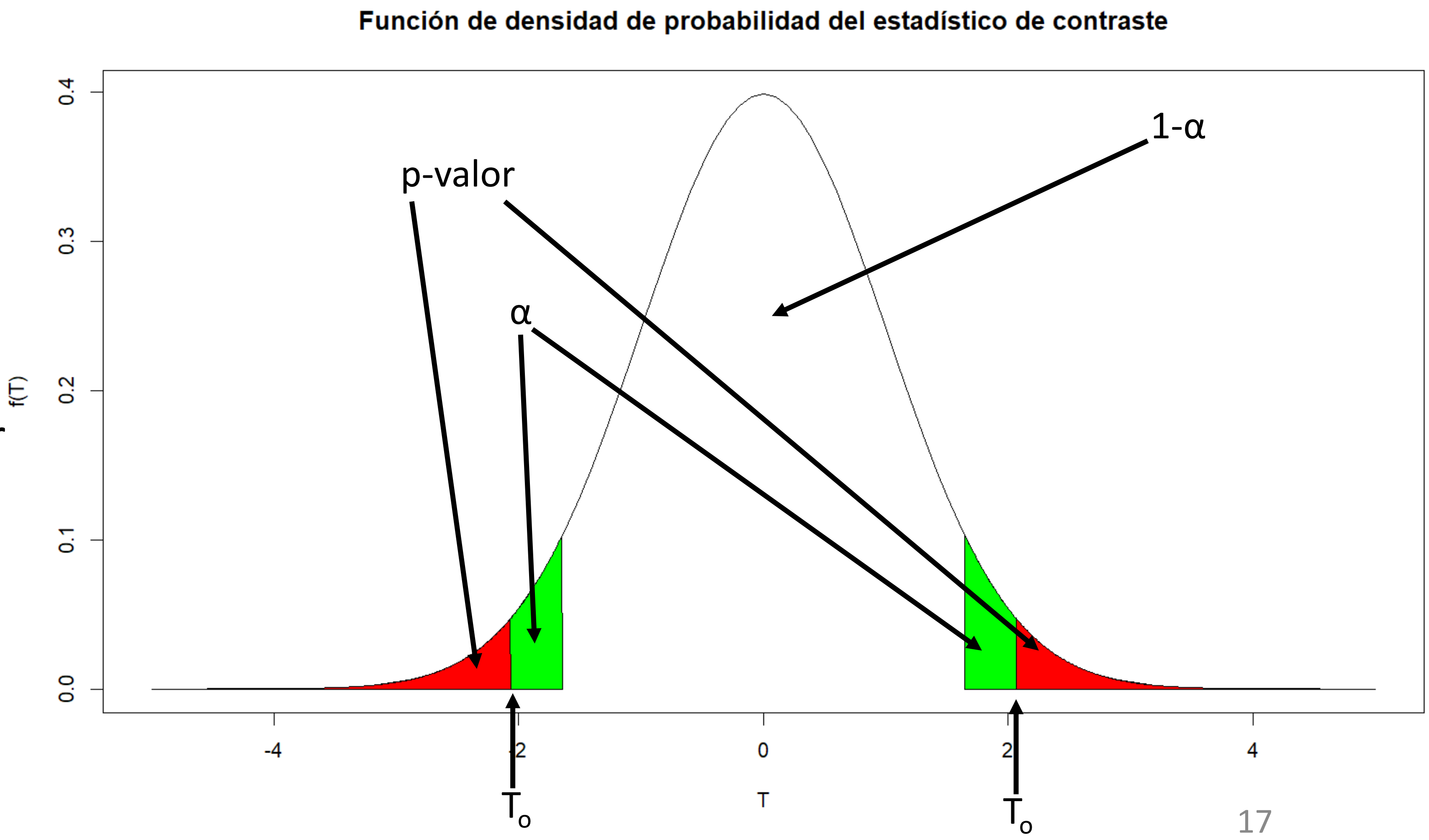


FIGURE 9.11: *P-value separates α-to-accept and α-to-reject.*

3. Métodos para realizar un contraste de hipótesis Calculando el p-valor. Ejemplo gráfico

- Ejemplo de contraste bilateral o de dos colas (two-sided test)
- H_0:θ=valor vs H_A:θ ≠valor
- El área verde total es α
- El área roja total es p-valor
- En este caso se observa que el área roja es menor que el área verde, es decir,
  p-valor< α , por lo que se rechaza la hipótesis H_0



4. Contraste de hipótesis para la media poblacional

- Casos posibles
  - Varianza poblacional conocida → Z-test
    - Se utiliza un estadístico de contraste T=Z con distribución Normal
  - Varianza poblacional desconocida
    - Muestra grande o muestra pequeña con distribución Normal → T-test
      - Se utiliza un estadístico de contraste T=t con distribución T de Student
    - Muestra pequeña sin distribución Normal o desconocida → No es posible, en su lugar se puede realizar un contraste no paramétrico sobre la mediana poblacional
- Nota
  - Una muestra se suele considerar grande cuando el tamaño es $n \geq 30$ o $n > 30$ (no hay unanimidad).

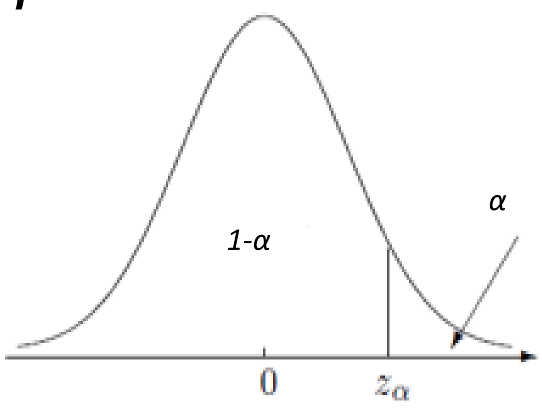
4. Contraste de hipótesis para la media poblacional. Fórmulas

Varianza conocida	Estadístico de contraste	Contraste de hipótesis	Región de aceptación de H_0	p-valor
Si	$Z_o = \dfrac{\bar{x} - \mu_o}{\sigma/\sqrt{n}}$	$H_0\colon \mu = \mu_o$ vs $H_A\colon \mu < \mu_o$	$(-z_{\alpha}, +\infty)$	$P\{Z \leq Z_o\}$
		$H_0\colon \mu = \mu_o$ vs $H_A\colon \mu > \mu_o$	$(-\infty, z_{\alpha})$	$P\{Z \geq Z_o\}$
		$H_0\colon \mu = \mu_o$ vs $H_A\colon \mu \neq \mu_o$	$\left(-z_{\alpha/2}, z_{\alpha/2}\right)$	$P\{\lvert Z\rvert \geq \lvert Z_o\rvert\} = 2P\{Z \geq \lvert Z_o\rvert\}$
No $n \geq 30$	$Z_o = \dfrac{\bar{x} - \mu_o}{s/\sqrt{n}}$	$H_0\colon \mu = \mu_o$ vs $H_A\colon \mu < \mu_o$	$(-z_{\alpha}, +\infty)$	$P\{Z \leq Z_o\}$
		$H_0\colon \mu = \mu_o$ vs $H_A\colon \mu > \mu_o$	$(-\infty, z_{\alpha})$	$P\{Z \geq Z_o\}$
		$H_0\colon \mu = \mu_o$ vs $H_A\colon \mu \neq \mu_o$	$\left(-z_{\alpha/2}, z_{\alpha/2}\right)$	$P\{\lvert Z\rvert \geq \lvert Z_o\rvert\} = 2P\{Z \geq \lvert Z_o\rvert\}$
No $n < 30$	$t_o = \dfrac{\bar{x} - \mu_o}{s/\sqrt{n}}$	$H_0\colon \mu = \mu_o$ vs $H_A\colon \mu < \mu_o$	$\left(-t_{\alpha,\, n-1}, +\infty\right)$	$P\{t \leq t_o\}$
		$H_0\colon \mu = \mu_o$ vs $H_A\colon \mu > \mu_o$	$\left(-\infty, t_{\alpha,\, n-1}\right)$	$P\{t \geq t_o\}$
		$H_0\colon \mu = \mu_o$ vs $H_A\colon \mu \neq \mu_o$	$\left(-t_{\alpha/2,\, n-1}, t_{\alpha/2,\, n-1}\right)$	$P\{\lvert t\rvert \geq \lvert t_o\rvert\} = 2P\{t \geq \lvert t_o\rvert\}$

4. Contraste de hipótesis para la media poblacional. Concepto de cuantil

Para calcular z_α, $z_{\alpha/2}$, t_α o $t_{\alpha/2}$ hay que recordar el concepto de cuantil.

- z_α=cuantil 1-α de una variable Z:Normal(0,1)
- Por tanto
  - $P\{Z \leq z_\alpha\} = 1 - \alpha$



4. Contraste de hipótesis para la media poblacional. Intervalo de confianza

- Existe una relación directa entre el contraste de hipótesis bilateral de cualquier parámetro y el intervalo de confianza para dicho parámetro.
- En un contraste bilateral con nivel α:
  - H_0: $\mu = \mu_o$ vs H_A: $\mu \neq \mu_o$
- Se rechaza H_0 con nivel de confianza $1-\alpha$, si el Intervalo de Confianza con nivel $1-\alpha$ para μ no contiene el valor μ_o

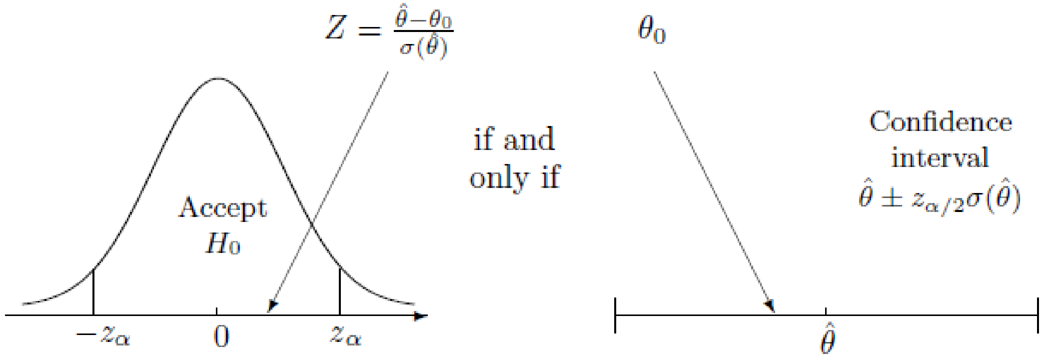


FIGURE 9.8: *Duality of tests and confidence intervals.*

4. Contraste de hipótesis para la media poblacional. Ejemplo 9.25 (Servicio de Internet)

- El número de usuarios simultáneos para algunos proveedores de servicios de Internet siempre ha sido de 5000 usuarios con una desviación estándar de 800 usuarios.
- Después de una actualización de equipos, el número promedio de usuarios en 100 momentos de tiempo seleccionados al azar es de 5200 usuarios.
- Suponer que la desviación estándar del número de usuarios simultáneos no ha cambiado.
- ¿Se puede afirmar con un nivel de significación del 5%, que la media de usuarios simultáneos ha aumentado?

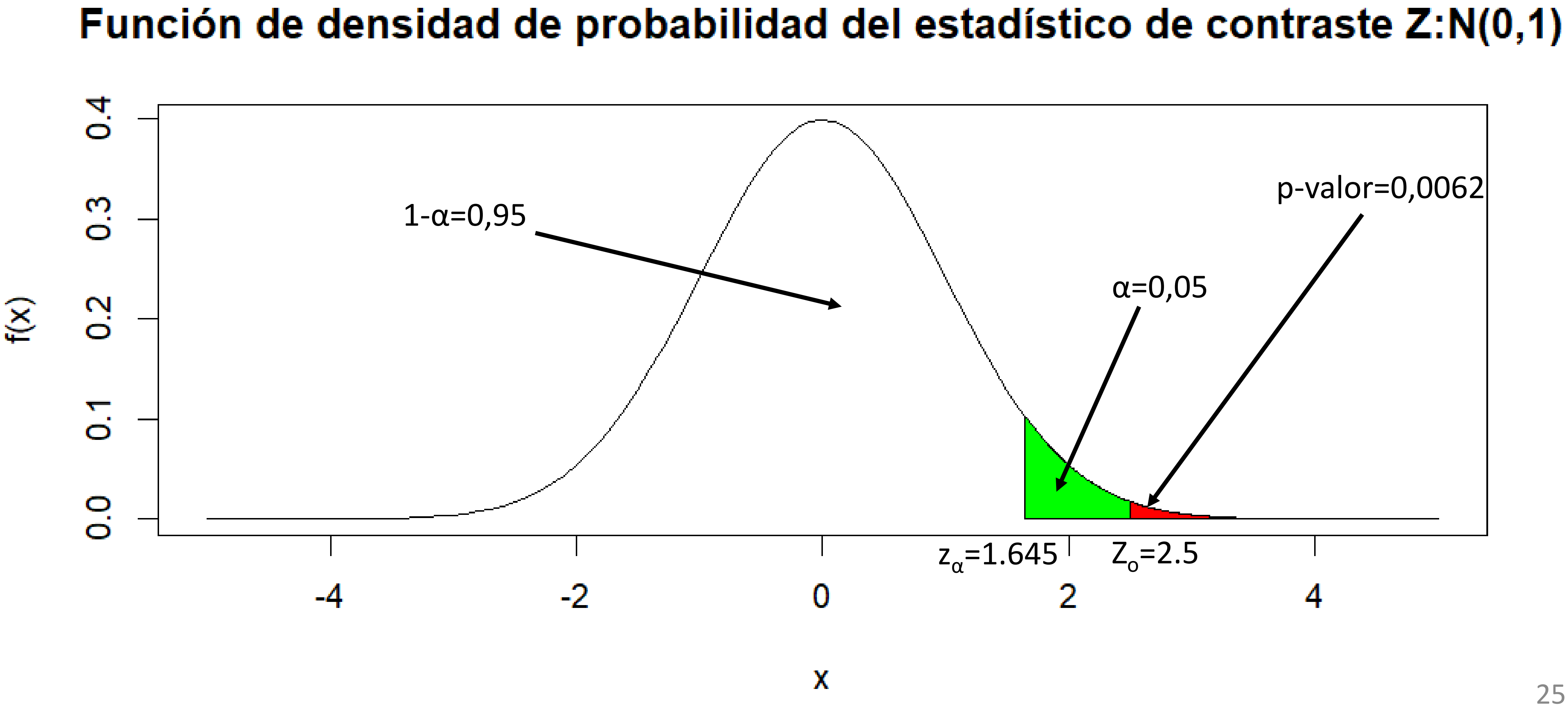
4. Contraste de hipótesis para la media poblacional. Ejemplo 9.25 (Solución) (1)

- H_0: $\mu = 5000$ vs H_A: $\mu > 5000$
  - Se plantea un contraste unilateral con cola derecha (one-sided, right-tail) porque nos interesa saber si el número medio de usuarios (μ) ha aumentado.
- Datos:
  - $\sigma = 800,\ n = 100,\ \alpha = 0.05,\ \mu_o = 5000,\ \bar{x} = 5200$
- Como se conoce la varianza poblacional (σ^2) se usa un estadístico de contraste Z
  - $Z_o = \frac{\bar{x}-\mu_o}{\sigma/\sqrt{n}} = \frac{5200-5000}{800/\sqrt{100}} = 2.5$

4. Contraste de hipótesis para la media poblacional. Ejemplo 9.25 (Solución) (2)

- Método de la región de aceptación:
  - $(-\infty, z_{\alpha}) = (-\infty, z_{0.05}) = (-\infty, 1.645)$ → Como $Z_o = 2.5$ está fuera de la región, se rechaza la hipótesis nula (H_0) y, en consecuencia, se acepta la hipótesis alternativa H_A
- Método del p-valor:
  - $\text{p}-\text{valor} = P\{Z \geq Z_o\} = P\{Z \geq 2.5\} = 1 - P\{Z \leq 2.5\} = 1 - 0.9938 = 0.0062$ → Como p-valor es menor que α=0.05, se rechaza la hipótesis nula (H_0) y, en consecuencia, se acepta la hipótesis alternativa H_A
- Respuesta = SI
  - Los datos (promedio de 5200 usuarios en 100 momentos de tiempo) proporcionan evidencias suficientes a favor de la hipótesis alternativa de que el número medio de usuarios ha aumentado.

4. Contraste de hipótesis para la media poblacional. Ejemplo 9.25 (Gráficos)



4. Contraste de hipótesis para la media poblacional. Ejemplo 9.28 (Acceso no autorizado)

- Recientemente se han propuesto varios métodos para detectar el uso no autorizado a una cuenta en un ordenador.
- Uno es medir el tiempo entre pulsaciones de teclas de un usuario al escribir su nombre de usuario y contraseña, y se compara con el tiempo habitual del propietario de la cuenta. Si hay diferencias significativas, se detecta un intruso.
- Se registraron los siguientes tiempos entre pulsaciones de teclas cuando un usuario escribía el nombre de usuario y contraseña:
  - 0.24, 0.22, 0.26, 0.34, 0.35, 0.32, 0.33, 0.29, 0.19, 0.36, 0.30, 0.15, 0.17, 0.28, 0.38, 0.40, 0.37, 0.27 segundos
- Si el usuario autorizado tiene habitualmente un tiempo medio entre pulsaciones de teclas de 0.2 segundos, ¿se puede afirmar con un nivel de significación del 5%, que se trata de un acceso no autorizado?

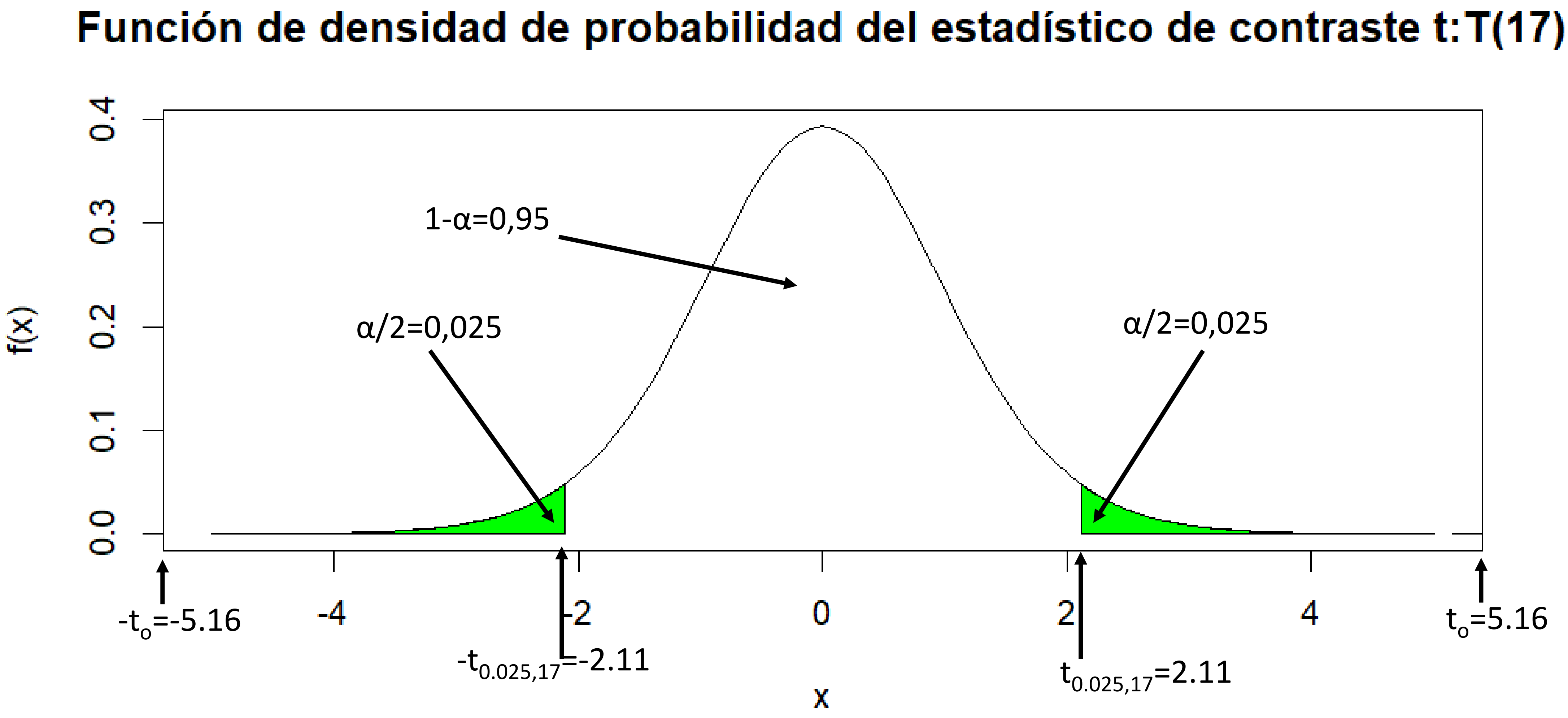
4. Contraste de hipótesis para la media poblacional. Ejemplo 9.28 (Solución) (1)

- H_0: $\mu = 0.2$ vs H_A: $\mu \neq 0.2$
  - Se plantea un contraste bilateral, porque nos interesa saber si la media del tiempo entre pulsaciones de teclas (μ) es significativamente diferente al habitual del usuario autorizado (0.2s).
- Datos:
  - $\alpha = 0.05, n = 18,$
  - $x = (0.24, 0.22, 0.26, 0.34, 0.35, 0.32, 0.33, 0.29, 0.19, 0.36, 0.30, 0.15, 0.17, 0.28, 0.38, 0.40, 0.37, 0.27)$
- Como no se conoce la varianza poblacional y n=18 es <30, se usa un estadístico de contraste con distribución T de Student con n-1=17 grados de libertad
  - $t_o = \frac{\bar{x}-\mu_o}{s/\sqrt{n}} = \frac{0.29-0.2}{0.074/\sqrt{18}} = 5.16$ Siendo $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ y $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

4. Contraste de hipótesis para la media poblacional. Ejemplo 9.28 (Solución) (2)

- Método de la región de aceptación:
  - $\left(-t_{\alpha/2,\, n-1}, t_{\alpha/2,\, n-1}\right) = (-t_{0.025,17}, t_{0.025,17}) = (-2.11, 2.11)$
  - En la Tabla A5 se puede encontrar que $t_{0.025,17} = 2.11$
  - Como $t_o = 5.16$ está fuera de la región de aceptación, se rechaza la hipótesis nula (H_0) y, en consecuencia, se acepta la hipótesis alternativa H_A
- Método del p-valor:
  - $\text{p}-\text{valor} = P\{|t| \geq |t_o|\} = 2P\{t \geq |t_o|\} = 2P\{t \geq 5.16\} = 2(1 - P\{t \leq 5.16\}) = 0.000078$
  - Se usa software estadístico porque no está en la tabla A5
  - Como p-valor es menor que α=0.05, se rechaza la hipótesis nula (H_0) y, en consecuencia, se acepta la hipótesis alternativa H_A
- Respuesta = SI
  - Los datos proporcionan evidencias suficientes a favor de la hipótesis alternativa, para afirmar con un nivel de significación del 5%, que se trata de un acceso no autorizado.

4. Contraste de hipótesis para la media poblacional. Ejemplo 9.28 (Gráficos)



4. Contraste de hipótesis para la media poblacional. Ejercicios propuestos

- Ejercicios 9.7(b), 9.9(b), 9.13 del libro
  - Las respuestas de 9.7 y 9.9 están disponibles en el libro

5. Contraste de hipótesis para una proporción poblacional. Fórmulas

Parámetro	Estadístico de contraste	Contraste de hipótesis	Región de aceptación de H_0	p-valor
p	$Z_o = \frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}}$	$H_0\!: p = p_o$ vs $H_A\!: p < p_o$	$(-z_\alpha, +\infty)$	$P\{Z \leq Z_o\}$
		$H_0\!: p = p_o$ vs $H_A\!: p > p_o$	$(-\infty, z_\alpha)$	$P\{Z \geq Z_o\}$
		$H_0\!: p = p_o$ vs $H_A\!: p \neq p_o$	$\left(-z_{\alpha/2}, z_{\alpha/2}\right)$	$P\{\lvert Z\rvert \geq \lvert Z_o\rvert\} = 2P\{Z \geq \lvert Z_o\rvert\}$

5. Contraste de hipótesis para una proporción poblacional. Ejemplo (Conductores)

- En una muestra de 100 habitantes de una ciudad, hay 60 personas que saben conducir.
- ¿Se puede afirmar con una confianza del 90% que más de la mitad de los habitantes de la ciudad saben conducir?
- NOTA: Considerar que para una variable aleatoria Z:Normal(0,1), se cumple que $P\{Z \leq 1.28\} = 0.9$

5. Contraste de hipótesis para una proporción poblacional. Ejemplo (Conductores) (Solución)

- Contraste de hipótesis
  - $H_0\colon p = p_o$ vs $H_A\colon p > p_o$ $\rightarrow$ $H_0\colon p = 0.5$ vs $H_A\colon p > 0.5$
- Datos: $n = 100, \alpha = 0.1, p_o = 0.5$, $\hat{p} = \frac{60}{100} = 0.6$, $P\{Z \leq 1.28\} = 0.9$
- $Z_o = \frac{\hat{p}-p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}} = \frac{0.6-0.5}{\sqrt{\frac{0.5(1-0.5)}{100}}} = 2$
- Método de la región de aceptación:
  - $(-\infty, z_\alpha) = (-\infty, z_{0.1}) = (-\infty, 1.28)$
  - Porque $z_{0.1} = cuantil(1-0.1)$ $\rightarrow P\{Z \leq z_{0.1}\} = 0.9$ (Es un dato, por lo que no hace falta usar la Tabla A4)
  - Como $Z_o = 2$ está fuera de la región de aceptación, se rechaza la hipótesis nula (H_0) y, en consecuencia, se acepta la hipótesis alternativa H_A
- Método del p-valor:
  - $\text{p}-\text{valor} = P\{Z \geq Z_o\} = P\{Z \geq 2\} = 1 - P\{Z \leq 2\} = 1 - 0.98 = 0.02$
  - Se usa la tabla A4
  - Como p-valor es menor que α=0.1, se rechaza la hipótesis nula (H_0) y, en consecuencia, se acepta la hipótesis alternativa H_A
- Respuesta = SI
  - Los datos proporcionan evidencias suficientes a favor de la hipótesis alternativa, para afirmar con un nivel de confianza del 95%, que más de la mitad de los habitantes de la ciudad saben conducir

5. Contraste de hipótesis para una proporción poblacional. Ejercicios propuestos

- Ejercicio 9.10(b) del libro
  - Las respuesta está disponible en el libro

6. Contraste de hipótesis para la diferencia de medias de dos poblaciones independientes (1)

Varianza conocida	Estadístico de contraste	Contraste de hipótesis	Región de aceptación de H_0	p-valor
Si	$Z_o = \dfrac{\bar{x}-\bar{y}-D}{\sqrt{\dfrac{\sigma_X^2}{n}+\dfrac{\sigma_Y^2}{m}}}$	H_0: $\mu_X-\mu_Y=D$ vs H_A: $\mu_X-\mu_Y<D$	$(-z_\alpha,+\infty)$	$P\{Z\le Z_o\}$
		H_0: $\mu_X-\mu_Y=D$ vs H_A: $\mu_X-\mu_Y>D$	$(-\infty,z_\alpha)$	$P\{Z\ge Z_o\}$
		H_0: $\mu_X-\mu_Y=D$ vs H_A: $\mu_X-\mu_Y\ne D$	$\left(-z_{\alpha/2},z_{\alpha/2}\right)$	$P\{\lvert Z\rvert\ge\lvert Z_o\rvert\}=2P\{Z\ge\lvert Z_o\rvert\}$
No $n,m\ge 30$ $\sigma_X\ne\sigma_Y$	$Z_o = \dfrac{\bar{x}-\bar{y}-D}{\sqrt{\dfrac{s_X^2}{n}+\dfrac{s_Y^2}{m}}}$	H_0: $\mu_X-\mu_Y=D$ vs H_A: $\mu_X-\mu_Y<D$	$(-z_\alpha,+\infty)$	$P\{Z\le Z_o\}$
		H_0: $\mu_X-\mu_Y=D$ vs H_A: $\mu_X-\mu_Y>D$	$(-\infty,z_\alpha)$	$P\{Z\ge Z_o\}$
		H_0: $\mu_X-\mu_Y=D$ vs H_A: $\mu_X-\mu_Y\ne D$	$\left(-z_{\alpha/2},z_{\alpha/2}\right)$	$P\{\lvert Z\rvert\ge\lvert Z_o\rvert\}=2P\{Z\ge\lvert Z_o\rvert\}$
No $n,m\ge 30$ $\sigma_X=\sigma_Y$	$Z_o = \dfrac{\bar{x}-\bar{y}-D}{s_p\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}}$	H_0: $\mu_X-\mu_Y=D$ vs H_A: $\mu_X-\mu_Y<D$	$(-z_\alpha,+\infty)$	$P\{Z\le Z_o\}$
		H_0: $\mu_X-\mu_Y=D$ vs H_A: $\mu_X-\mu_Y>D$	$(-\infty,z_\alpha)$	$P\{Z\ge Z_o\}$
		H_0: $\mu_X-\mu_Y=D$ vs H_A: $\mu_X-\mu_Y\ne D$	$\left(-z_{\alpha/2},z_{\alpha/2}\right)$	$P\{\lvert Z\rvert\ge\lvert Z_o\rvert\}=2P\{Z\ge\lvert Z_o\rvert\}$

6. Contraste de hipótesis para la diferencia de medias de dos poblaciones independientes (2)

Varianza conocida	Estadístico de contraste	Contraste de hipótesis	Región de aceptación de H_0	p-valor
No $n, m < 30$ $\sigma_X \neq \sigma_Y$	$t_o = \dfrac{\bar{x} - \bar{y} - D}{\sqrt{\dfrac{s_X^2}{n} + \dfrac{s_Y^2}{m}}}$	$H_0\colon \mu_X - \mu_Y = D$ vs $H_A\colon \mu_X - \mu_Y > D$	$(-t_{\alpha,\nu}, +\infty)$	$P\{t \leq t_o\}$
		$H_0\colon \mu_X - \mu_Y = D$ vs $H_A\colon \mu_X - \mu_Y < D$	$(-\infty, t_{\alpha,\nu})$	$P\{t \geq t_o\}$
		$H_0\colon \mu_X - \mu_Y = D$ vs $H_A\colon \mu_X - \mu_Y \neq D$	$(-t_{\alpha/2,\nu}, t_{\alpha/2,\nu})$	$P\{\lvert t\rvert \geq \lvert t_o\rvert\} = 2P\{t \geq \lvert t_o\rvert\}$
No $n, m < 30$ $\sigma_X = \sigma_Y$	$t_o = \dfrac{\bar{x} - \bar{y} - D}{s_p\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}}$	$H_0\colon \mu_X - \mu_Y = D$ vs $H_A\colon \mu_X - \mu_Y > D$	$(-t_{\alpha,n+m-2}, +\infty)$	$P\{t \leq t_o\}$
		$H_0\colon \mu_X - \mu_Y = D$ vs $H_A\colon \mu_X - \mu_Y < D$	$(-\infty, t_{\alpha,n+m-2})$	$P\{t \geq t_o\}$
		$H_0\colon \mu_X - \mu_Y = D$ vs $H_A\colon \mu_X - \mu_Y \neq D$	$(-t_{\alpha/2,n+m-2}, t_{\alpha/2,n+m-2})$	$P\{\lvert t\rvert \geq \lvert t_o\rvert\} = 2P\{t \geq \lvert t_o\rvert\}$

$$s_p = \sqrt{\frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}}$$

$$\nu = \frac{\left(\dfrac{s_X^2}{n} + \dfrac{s_Y^2}{m}\right)^2}{\dfrac{s_X^4}{n^2(n-1)} + \dfrac{s_Y^4}{m^2(m-1)}}$$

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

$$s_Y^2 = \frac{1}{m-1}\sum_{i=1}^{m}(y_i - \bar{y})^2$$

6. Contraste de hipótesis para la diferencia de medias de dos poblaciones. Ejemplo 9.29

- Continuación del ejemplo 9.20
- Para estimar el efecto de la escritura en CD, se pide a 30 usuarios que trabajen en sus ordenadores portátiles hasta que aparezca el signo de "baja batería".
- Dieciocho usuarios sin CD trabajaron un promedio de 5,3 horas con una desviación estándar de 1,4 horas.
- Los otros doce, que utilizaron su grabador de CD, trabajaron un promedio de 4,8 horas con una desviación estándar de 1,6 horas.
- Suponer distribuciones Normales con las mismas varianzas poblacionales
- ¿La grabación de CD (Compact Disc) consume energía extra y, por tanto, afecta a la vida útil de la batería en los ordenadores portátiles?

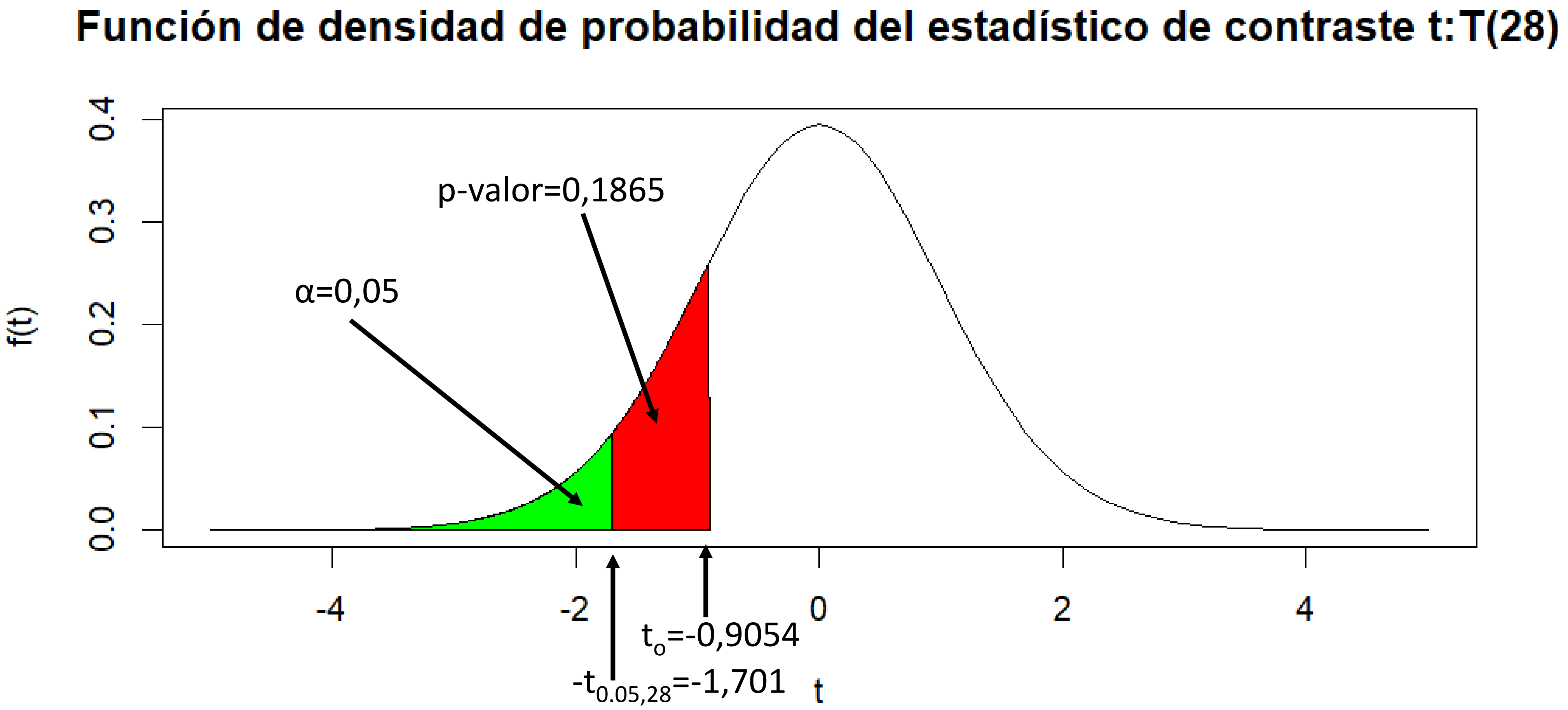
6. Contraste de hipótesis para la diferencia de medias de dos poblaciones. Ejemplo 9.29 (Solución)(1)

- X = tiempo de uso por usuarios con CD
- Y = tiempo de uso por usuarios sin CD
- H_0: $\mu_X = \mu_Y$ vs H_A: $\mu_X < \mu_Y$
  - Es decir: H_0: $\mu_X - \mu_Y = D$ vs H_A: $\mu_X - \mu_Y < D$ siendo D=0
  - Se plantea como un contraste unilateral con cola izquierda (one-sided, left-tail).
- Datos
  - $n = 12, \bar{x} = 4.8 \text{ horas}, s_X = 1.6 \text{ horas}, m = 18 \ \ \bar{y} = 5.3 \text{ horas}, s_Y = 1.4 \text{ horas}$
  - $1 - \alpha = 0.95 \rightarrow \alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025$
- Es un caso de muestras pequeñas (<30) con varianzas desconocidas pero iguales, por lo que se utiliza t como estadístico de contraste
  - $t_o = \frac{\bar{x} - \bar{y} - D}{s_p\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{4.8 - 5.3 - 0}{1.48\sqrt{\frac{1}{12} + \frac{1}{18}}} = -0.9054$
  - Donde $s_p = \sqrt{\frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}} = \sqrt{\frac{(11)(1.6)^2 + (17)(1.4)^2}{12+18-2}} = 1.48$

6. Contraste de hipótesis para la diferencia de medias de dos poblaciones. Ejemplo 9.29 (Solución)(2)

- Método de la región de aceptación:
  - $\left(-t_{\alpha,\, n+m-2}, +\infty\right) = \left(-t_{0.05,\, 28}, +\infty\right) = (-1.701, +\infty)$
  - Según la tabla A5: $t_{\alpha,n+m-2} = t_{0.05,\,28} = 1.701$
  - Como $t_o = -0.9054$ está dentro, se acepta la hipótesis nula (H_0) y, en consecuencia, se rechaza la hipótesis alternativa H_A
- Método del p-valor:
  - $\mathrm{p{-}valor} = P\{\mathrm{T} \leq t_o\} = P\{\mathrm{T} \leq -0.9054\} = 0.1865$
  - Calculado con software estadístico porque no está en la tabla A5
  - Como p-valor es mayor que α=0.05, se acepta la hipótesis nula (H_0) y, en consecuencia, se rechaza la hipótesis alternativa H_A
- Respuesta = NO
  - Los datos no proporcionan evidencias suficientes a favor de la hipótesis alternativa de que la grabación de CD (Compact Disc) consuma energía extra y afecte a la vida útil de la batería en los ordenadores portátiles.

6. Contraste de hipótesis para la diferencia de medias de dos poblaciones. Ejemplo 9.29 (Gráficos)



6. Contraste de hipótesis para la diferencia de medias de dos poblaciones. Ejercicios propuestos

- Ejercicios 9.14, 9.18 del libro
  - La respuesta de 9.18 está disponible en el libro

7. Contraste de hipótesis para la diferencia de proporciones de dos poblaciones independientes

Diferencia de parámetros	Estadístico de contraste	Contraste de hipótesis	Región de aceptación de H_0	p-valor
$p_1 - p_2$	$Z_o = \dfrac{\hat{p}_1 - \hat{p}_2 - D}{\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{m}}}$	$H_0\colon p_1 - p_2 = D$ vs $H_A\colon p_1 - p_2 < D$	$(-z_\alpha, +\infty)$	$P\{Z \leq Z_o\}$
		$H_0\colon p_1 - p_2 = D$ vs $H_A\colon p_1 - p_2 > D$	$(-\infty, z_\alpha)$	$P\{Z \geq Z_o\}$
		$H_0\colon p_1 - p_2 = D$ vs $H_A\colon p_1 - p_2 \neq D$	$\left(-z_{\alpha/2}, z_{\alpha/2}\right)$	$P\{\lvert Z\rvert \geq \lvert Z_o\rvert\} = 2P\{Z \geq \lvert Z_o\rvert\}$

7. Contraste de hipótesis para la diferencia de proporciones de dos poblaciones. Ejemplo 9.26

- Un inspector de calidad encuentra 10 piezas defectuosas en una muestra de 500 piezas recibidas del fabricante A.
- De 400 piezas del fabricante B, encuentra 12 piezas defectuosas.
- Una empresa de fabricación de computadores utiliza estas partes en sus computadores y afirma que la calidad de las piezas producidas por A y B es la misma.
- Con un nivel de significación del 5 %, ¿tenemos suficientes pruebas para refutar esta afirmación?

7. Contraste de hipótesis para la diferencia de proporciones de dos poblaciones. Ejemplo 9.26 (Solución) (1)

- $n = 500,\ m = 400,\ \alpha = 0.05$
- $\hat{p}_A = \frac{10}{500} = 0.02$
- $\hat{p}_B = \frac{12}{400} = 0.03$
- H_0: $p_A = p_B$ *vs* H_A: $p_A \neq p_B$
  - Es decir: $H_0: p_A - p_B = D$ *vs* $H_A: p_A - p_B \neq 0$ siendo D=0
  - Se plantea como un contraste bilateral.
- Estadístico de contraste
  - $Z_o = \frac{\hat{p}_A - \hat{p}_B - D}{\sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{n} + \frac{\hat{p}_B(1-\hat{p}_B)}{m}}} = \frac{0.02 - 0.03 - 0}{\sqrt{\frac{0.02(1-0.02)}{500} + \frac{0.03(1-0.03)}{400}}} = -0.95$

7. Contraste de hipótesis para la diferencia de proporciones de dos poblaciones. Ejemplo 9.26 (Solución) (2)

- Método de la región de aceptación:
  - $\left(-z_{\alpha/2}, z_{\alpha/2}\right) = (-z_{0.025}, z_{0.025}) = (-1.96, 1.96)$
  - Según la tabla A4: $P\{Z \leq z_{0.025}\} = 1 - 0.025 = 0.975 \ \rightarrow z_{0.025} = 1.96$
  - Como $Z_o = -0.945$ está dentro, no se rechaza la hipótesis nula (H_0) y, en consecuencia, se rechaza la hipótesis alternativa H_A
- Método del p-valor:
  - $\text{p}-\text{valor} = 2P\{Z \geq |Z_o|\} = 2(1 - P\{Z \leq |Z_o|\}) = 2(1 - P\{Z \leq 0.95\} = 2(1 - 0.83) = 0.34$
  - Usando la tabla A4
  - Como p-valor es mayor que α=0.05, se acepta la hipótesis nula (H_0) y, en consecuencia, se rechaza la hipótesis alternativa H_A
- Respuesta = NO
  - Los datos no proporcionan suficientes evidencias para refutar la afirmación de la empresa de que la calidad de las piezas es la misma.

7. Contraste de hipótesis para la diferencia de proporciones de dos poblaciones. Ejercicios propuestos

- Ejercicios 9.11, 9.15, 9.16(b), 9.17
    - Las respuestas de 9.11, y 9.17 están disponibles en el libro

8. Conclusiones

- En el contexto de la estadística inferencial, una hipótesis es una afirmación sobre el valor de un parámetro (θ) de una o varias poblaciones.
- El contraste de hipótesis consiste en formular una hipótesis nula (H_0) contra a una hipótesis alternativa (H_A) y decidir si se acepta la hipótesis nula rechazando la alternativa o viceversa.
- Para realizar un contraste se utiliza un estadístico de contraste, que es una variable aleatoria.
- En el caso de contrastes de medias, se debe utilizar un estadístico con distribución Normal o con distribución T de Student dependiendo de unas condiciones previas.
- En el caso de contrastes de proporciones se utiliza un estadístico con distribución Normal.
- El contraste se puede realizar aplicando el método de la región de aceptación o el método del p-valor.