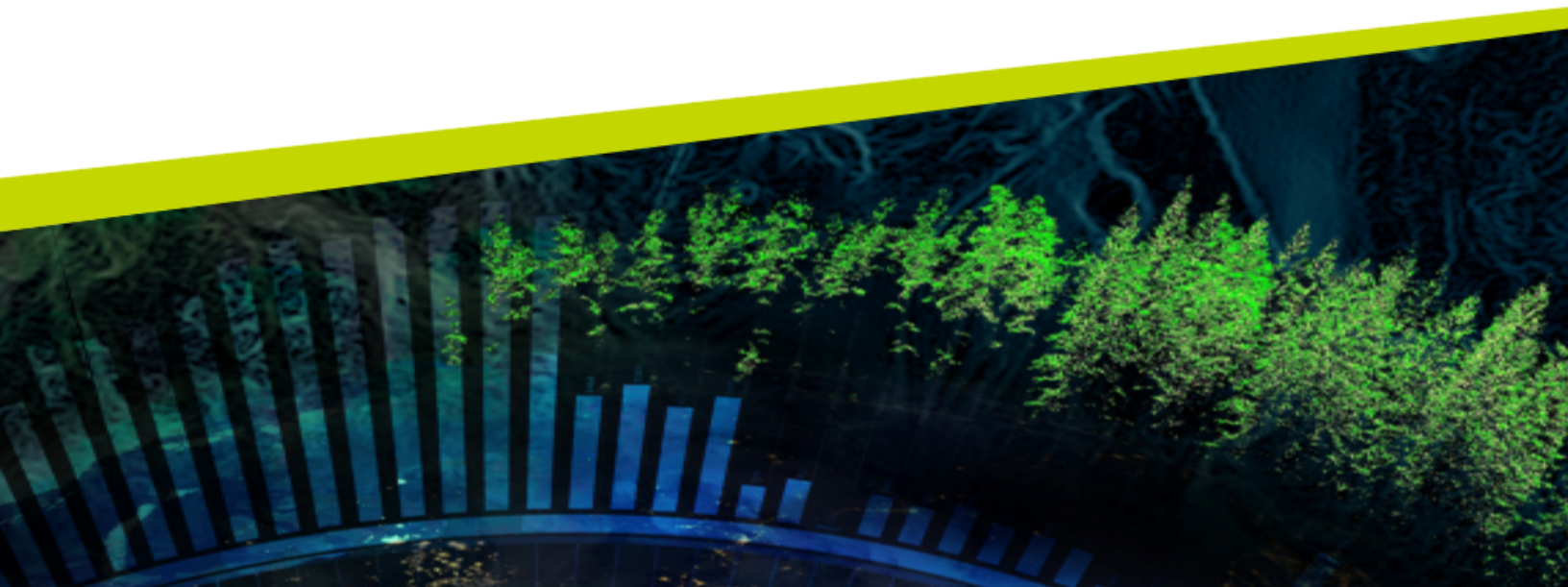
INFERENCIA Y MODELOS ESTADÍSTICOS

Jacqueline Köhler C. y José Luis Jara V.



CAPÍTULO 4. FUNDAMENTOS PARA LA INFERENCIA

En el capítulo 4 se definen los conceptos de población, entendido como todo el conjunto de interés, y muestra, que es un subconjunto de la población. También se introducen las nociones de parámetro, correspondiente a un valor que resume la población (por ejemplo la media de la población, μ), y de estadístico, como valor que resume una muestra (por ejemplo, la media muestral, $\overline{x}$). La **inferencia estadística** tiene por objeto entender cuán cerca está el estadístico del parámetro real de la población. En este capítulo conoceremos los principios necesarios para la inferencia estadística, con base en Diez y col. (2017, pp. 168-202) y Field y col. (2012, pp. 40-47).

4.1 ESTIMADORES PUNTUALES

Como ya dijimos, los parámetros y los estadísticos son valores que resumen, respectivamente, una población y una muestra. En consecuencia, podemos decir que un estadístico corresponde a un **estimador puntual** de un parámetro. El valor de un estimador puntual cambia dependiendo de la muestra que usemos para obtenerlo. Así, por más que su valor se acerque al parámetro de la población, difícilmente será igual a este último. Sin embargo, el estimador tiende a mejorar a medida que aumentamos el tamaño de la muestra, por efecto de la **ley de los grandes números**. Para ilustrar este fenómeno, consideremos la **media móvil**, que es una secuencia de medias muestrales en que cada una de ellas toma un elemento más de la población que su antecesora. La figura 4.1, elaborada con el script 4.1, ejemplifica este fenómeno.

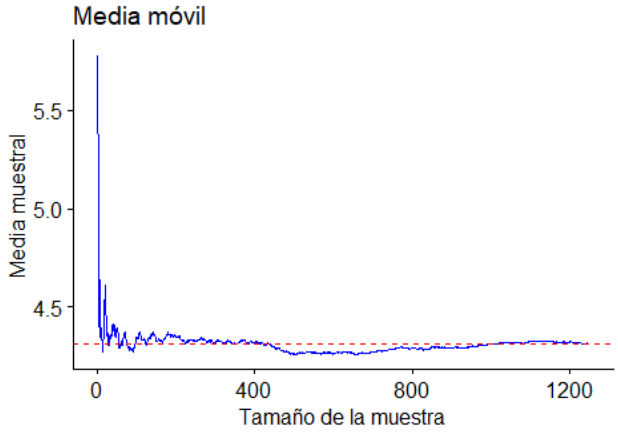


Figura 4.1: medias obtenidas al agregar a la muestra un elemento cada vez.

Script 4.1: representación gráfica de la media móvil.

```
library(ggpubr)

# Establecer la semilla para generar números aleatorios.
set.seed(9437)

# Generar aleatoriamente una población de tamaño 1500
# (en este caso, con una distribución cercana a la normal).
```

```r
poblacion <- rnorm(n = 1500, mean = 4.32, sd = 0.98)

# Calcular la media de la población.
media_poblacion <- mean(poblacion)
cat("Media de la población:", media_poblacion, "\n")

# Tomar una muestra de tamaño 1250.
tamano_muestra <- 1250
muestra <- sample(poblacion, tamano_muestra)

# Calcular las medias acumuladas (es decir, con muestras de
# 1, 2, 3, ... elementos).
n <- seq(along = muestra)
media <- cumsum(muestra) / n

# Crear una matriz de datos con los tamaños y las medias muestrales.
datos <- data.frame(n, media)

# Graficar las medias muestrales.
g <- ggline(data = datos,
            x = "n",
            y = "media",
            plot_type = "l",
            color = "blue",
            main = "Media móvil",
            xlab = "Tamaño de la muestra",
            ylab = "Media muestral")

# Añadir al gráfico una recta con la media de la población.
g <- g + geom_hline(aes(yintercept = media_poblacion),
                    color = "red", linetype = 2)

print(g)
```

Para determinar qué tan adecuado es un estimador, necesitamos saber cuánto cambia de una muestra a otra. Si esta variabilidad es pequeña, es muy probable que la estimación sea buena. Podemos estudiar la variabilidad de la muestra con ayuda de la **distribución muestral**, que representa la distribución de estimadores puntuales obtenidos con **todas** las diferentes muestras de igual tamaño de una misma población. La figura 4.2 (construida con el script 4.2) representa las medias para diferentes muestras de una población, aunque solo una selección aleatoria de todas las posibles muestras, incluyendo además una línea vertical roja que señala la media de la población. Podemos destacar que las medias muestrales tienden a aglutinarse en torno a la media poblacional, pues de acuerdo al **teorema del límite central**, la distribución de $\overline{x}$ se aproxima a la normalidad. Esta aproximación mejora a medida que aumenta el tamaño de la muestra.

Script 4.2: distribución de la media muestral.

```r
library(ggpubr)

# Establecer la semilla para generar números aleatorios.
set.seed(94)

# Generar aleatoriamente una población de tamaño 1500
# (en este caso, con una distribución cercana a la normal).
poblacion <- rnorm(n = 1500, mean = 4.32, sd = 0.98)

# Calcular la media de la población.
media_poblacion <- mean(poblacion)
cat("Media de la población:", media_poblacion, "\n")

```

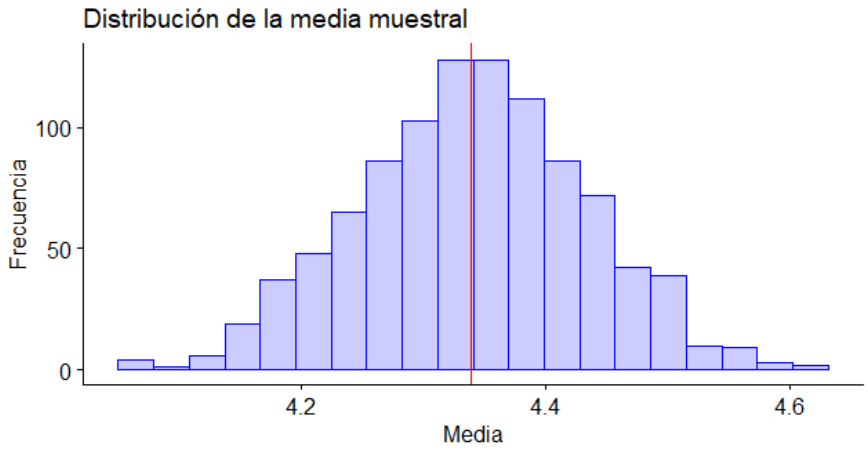


Figura 4.2: distribución muestral de la media para muestras con 100 observaciones.

```r
# Tomar 1000 muestras de tamaño 100. Quedan almacenadas
# como una matriz donde cada columna es una muestra.
tamano_muestra <- 100
repeticiones <- 1000

muestras <- replicate(repeticiones,
                      sample(poblacion, tamano_muestra))

# Calcular medias muestrales y almacenar los resultados
# en forma de data frame.
medias <- colMeans(muestras)
medias <- as.data.frame(medias)

# Construir un histograma de las medias muestrales.
g <- gghistogram(data = medias,
                 x = "medias",
                 bins = 20,
                 title = "Distribución de la media muestral",
                 xlab = "Media",
                 ylab = "Frecuencia",
                 color = "blue",
                 fill = "blue",
                 alpha = 0.2)

# Agregar línea vertical con la media de la población.
g <- g + geom_vline(aes(xintercept = media_poblacion),
                    color = "red", linetype = 1)

print(g)
```

4.2 MODELOS ESTADÍSTICOS

Ahora que hemos conocido más conceptos, podemos definir con precisión qué es un **modelo estadístico**. En el capítulo 4 dijimos que un modelo es simplemente una representación y que los modelos estadísticos pueden

emplearse para diversos propósitos:

- Describir o resumir datos.
- Clasificar objetos o predecir resultados.
- Anticipar los resultados de intervenciones (en ocasiones).

Más formalmente, un modelo estadístico es una descripción de un **proceso probabilístico** con **parámetros desconocidos** que deben ser **estimados** en base a **suposiciones** y un conjunto de datos **observados**. En general, tiene la forma dada en la ecuación 4.1:

$$y_i = (\text{modelo}) + \varepsilon_i \tag{4.1}$$

Donde:

- y_i es el i-ésimo valor observado de la variable respuesta Y (también llamada variable de salida o variable dependiente).
- modelo es el resultado de una función determinista basada en un conjunto de argumentos.
- ε_i es el error, correspondiente a la **variación natural**, y no a una equivocación, existente entre los valores observados y los valores pronosticados por el modelo. También recibe los nombres de variación no sistemática, variación aleatoria, residuos o incluso, residuales.

El error ε_i en la ecuación 4.1 se relaciona entonces con la calidad del modelo. Mientras menor sea el error, mejor será el modelo. Por el contrario, un error grande es señal de un modelo fallido, que no describe bien los datos, no ayuda a predecirlos bien, o no ayuda a su correcta clasificación.

La media y la proporción, y cualquier estadístico en general, son, en sí mismos, modelos estadísticos, aunque bastante simples.

4.3 ERROR ESTÁNDAR

En el capítulo 5 conocimos la desviación estándar como medida que estima la distancia de las observaciones respecto de la media. El **error estándar**, denotado usualmente por $SE_{\hat{\theta}}$ o $\sigma_{\hat{\theta}}$, corresponde a la desviación estándar de la distribución de un estimador muestral $\hat{\theta}$ de un parámetro θ. Por ejemplo, el error estándar de la media, es decir la desviación estándar de la distribución de las medias de todas las posibles muestras de n observaciones independientes, se calcula de acuerdo a la ecuación 4.2.

$$SE_{\overline{x}} = \frac{s}{\sqrt{n}} \tag{4.2}$$

En esta ecuación y los párrafos siguientes deberíamos hablar de σ y no s

Donde s es la desviación estándar de la muestra (ecuación 5.3) y n corresponde al tamaño de la muestra. En esta ecuación queda en evidencia que el error estándar de la media disminuye a medida que el tamaño de la muestra aumenta. Un método confiable que podemos usar para asegurar que las observaciones sean independientes es realizar un muestreo aleatorio simple[1] que abarque menos del 10 % de la población.

Volviendo a la ecuación para calcular el error estándar de la media muestral (equación 4.2), ¡debemos tener cuidado antes de usarla! Ya hemos mencionado antes que la distribución de las medias muestrales tiende a ser cercana a la normal, por lo que en dicho caso es posible usar el **modelo normal**, sustentado en el teorema del límite central. Las condiciones que deben cumplirse para usar este modelo y que, en consecuencia, el error estándar sea preciso, son:

[1] Es decir, una muestra en que todos los elementos de la población tengan igual probabilidad de ser escogidos. Las técnicas de muestreo se abordan con más detalle en capítulos posteriores.

1. Las observaciones de la muestra son independientes.
2. La muestra es grande (en general $n \geq 30$).
3. La distribución de la muestra no es significativamente asimétrica. Esto último suele además relacionarse con la presencia de valores atípicos. Mientras mayor sea el tamaño de la muestra, más se puede relajar esta condición.

Si no se cumplen las condiciones anteriores, debemos considerar otras opciones: para muestras pequeñas, se deben considerar métodos alternativos, y si la distribución de la muestra presenta una asimetría significativa, entonces tendremos que incrementar el tamaño de la muestra para compensar el efecto de la desviación.

4.4 INTERVALOS DE CONFIANZA

Hasta ahora sabemos que un estimador puntual es un único valor (obtenido a partir de una muestra) que, como su nombre indica, estima un parámetro de la población. Por ende, dicho valor rara vez es exacto. En consecuencia, lo lógico sería establecer un rango de valores plausibles para el parámetro estimado, que llamaremos **intervalo de confianza**, y que se construye en torno al estimador puntual. Dado que el error estándar representa la desviación estándar asociada al estimador, tiene sentido que lo usemos como guía en este proceso.

Recordemos que en el capítulo 6 vimos una regla empírica para la distribución normal (figura 6.5), la cual señala que (para distribuciones normales) alrededor de 95 % de las veces el estimador puntual se encontrará en un rango de 2 errores estándar del parámetro. Es decir, al considerar un intervalo de confianza de dos errores estándar (4.3), tendremos 95 % de **confianza** de haber capturado el parámetro real.

$$\overline{x} \pm 2 \cdot SE_{\overline{x}} \tag{4.3}$$

Podemos generalizar la ecuación 4.3 para calcular el intervalo de confianza para la media con cualquier **nivel de confianza** como muestra la ecuación 4.4.

$$\overline{x} \pm z^* \cdot SE_{\overline{x}} \tag{4.4}$$

El término z^* en la ecuación 4.4 corresponde, usualmente, al valor z tal que el área bajo la curva normal estándar comprendida entre $-z^*$ y z^* es igual al nivel de confianza deseado. La expresión $z^* \cdot SE$ recibe el nombre de **margen de error**.

Tomemos como ejemplo un **nivel de confianza** (que, por razones que veremos en la sección siguiente, denotaremos por $1 - \alpha$) de 90 % (es decir, $1 - \alpha = 0,9$). Eso significa, entonces, que nuestro intervalo de confianza excluye el 5 % del área correspondiente a la cola inferior (es decir, el percentil con valor 0,05) e igual porcentaje del área correspondiente a la cola superior (que, como la distribución Z es simétrica, es igual al área anterior). Puesto que conocemos el percentil, $(1 - \alpha)/2 = 0,05$, en R podemos usar la llamada `qnorm(0.05, mean = 0, sd = 1, lower.tail = FALSE)` y obtenemos $z^* = 1,64$. Es importante indicar que en esta llamada estamos en realidad trabajando con la cola superior para que z^* sea positivo. Si hacemos la llamada para la cola inferior, obtenemos $z^* = -1,64$.

Es importante destacar que, una vez más, debemos ser cuidadosos al interpretar un intervalo de confianza del x % (x $= 1 - \alpha$). Su significado es, sencillamente, "se tiene x % de certeza de que el parámetro de la población se encuentra entre..." (Diez y col., 2017, p. 180), es decir, que, en promedio, x % de los intervalos de confianza que se construyan en torno a un estadístico, con muestras de un tamaño fijo, capturarán el verdadero valor del parámetro. Esto **no es equivalente** a decir que el valor del parámetro tiene una "probabilidad de x %" de estar entre los valores del intervalo calculado, lo que sería incorrecto. Por otra parte, los intervalos de confianza no dicen nada acerca de observaciones individuales, sino que solo hablan del parámetro en cuestión.

4.5 PRUEBAS DE HIPÓTESIS

Supongamos que un banco ha desarrollado un nuevo sistema computacional para gestionar sus transacciones. El nuevo sistema (N) se ha puesto a prueba durante un mes, funcionando (con iguales condiciones de hardware) en paralelo con el sistema antiguo (A) y el banco ha llevado un registro del tiempo que tarda cada sistema en efectuar cada transacción. El gerente ha determinado que autorizará la migración al nuevo sistema únicamente si este es más rápido que el antiguo para procesar las transacciones. Se sabe que el sistema antiguo tarda en promedio $\mu_A = 530$ milisegundos en procesar una transacción. Para el sistema nuevo se han registrado $n = 1.600$ transacciones, realizadas en un tiempo promedio de $\overline{x}_N = 527,9$ [ms] con desviación estándar $s_N = 48$ [ms].

Una primera aproximación para tomar la decisión puede ser investigar si existe diferencia en los tiempos de ejecución de ambos sistemas, lo que puede expresarse en torno a dos **hipótesis** (palabra que la Real Academia Española (2014) define como "Suposición de algo posible o imposible para sacar de ello una consecuencia") que compiten entre sí:

H_0: El nuevo sistema, en promedio, tarda lo mismo que el antiguo en procesar las transacciones, es decir: $\mu_N = \mu_A$.
H_A: Los sistemas requieren, en promedio, cantidades de tiempo diferentes para procesar las transacciones, es decir: $\mu_N \neq \mu_A$

La primera hipótesis, H_0, recibe el nombre de **hipótesis nula** y suele representar una postura escéptica, es decir, que no hay cambios, por lo que **¡la hipótesis nula siempre se formula como una igualdad!**. La segunda (H_A), llamada **hipótesis alternativa**, representa en cambio una nueva perspectiva. Esta primera aproximación corresponde a una **prueba bilateral** o de dos colas, pues la diferencia puede ser en ambos sentidos: H_0 no parece correcta si $\mu_N < \mu_A$ o si $\mu_N > \mu_A$.

Como en este caso conocemos el valor de $\mu_A = 530$ [ms], también podríamos escribir la formulación matemática de las hipótesis de la siguiente manera:

H_0: $\mu_N = 530$
H_A: $\mu_N \neq 530$

En este planteamiento, "530" recibe el nombre de **valor nulo**, pues representa el valor del parámetro cuando se cumple la hipótesis nula.

Una aproximación más cercana al problema descrito puede ser investigar si el nuevo sistema es efectivamente **más rápido** que el antiguo. En este caso, se habla de una **prueba unilateral** o de una cola, pues solo interesa saber si el tiempo promedio empleado por el nuevo sistema es menor que el empleado por el sistema antiguo. Las hipótesis, en este caso, serían:

H_0: El nuevo sistema tarda, en promedio, lo mismo que el antiguo en procesar las transacciones, es decir: $\mu_N = \mu_A$.
H_A: El nuevo sistema tarda, en promedio, menos que el antiguo en procesar las transacciones, es decir: $\mu_N < \mu_A$

Obviamente en otros casos podría interesar solamente si valor alternativo es mayor que el valor nulo.

Teniendo las hipótesis planteadas, sigue decidir si la hipótesis nula parece o no plausible a través de una **prueba de hipótesis**. El marco para la prueba de hipótesis es **escéptico**: no se rechaza la hipótesis nula a menos que haya suficiente evidencia para rechazarla en favor de la hipótesis alternativa. Esta idea es muy parecida a la expresada en la expresión de uso común "se presume inocente mientras no se demuestre lo contrario". Sin embargo, el que no se logre rechazar H_0 **no significa aceptarla** como verdadera o como correcta sin más. Por eso se usa un lenguaje bastante peculiar, señalando que *se falla al rechazar* H_0 o bien que *se rechaza* H_0 *en favor de* H_A. Retomando la analogía con la expresión anterior, que no haya pruebas suficientes para la culpabilidad, no significa que una persona sea en verdad inocente.

Volvamos al escenario del ejemplo para la prueba de hipótesis bilateral (es decir, aquella en que solo queremos

ver si hay diferencias en el tiempo de procesamiento de transacciones entre ambos sistemas del banco). El valor de $\overline{x}_N = 527,9$ [ms] es, en efecto, distinto de $\mu_A = 530$ [ms]. No obstante, al ser una estimación puntual, como ya hemos estudiado, esta diferencia podría deberse simplemente a la muestra escogida, por lo que el parámetro real μ_N podría ser igual a μ_A [ms]. En consecuencia, resulta útil calcular el intervalo de confianza para $\overline{x}_N$.

Comencemos por determinar el error estándar:

$$SE_{\overline{x}} = \frac{s_N}{\sqrt{n}} = \frac{48}{\sqrt{1600}} = 1,2$$

Ahora fijemos un nivel de confianza, por ejemplo 95 %, y usemos el valor z^* correspondiente para calcular el intervalo de confianza:

$$\overline{x}_N \pm z^* \cdot SE_{\overline{x}} = 527,9 \pm 1,96 \cdot 1,2 = [525,548; 530,252]$$

Como el parámetro del sistema antiguo ($\mu_A = 530$ [ms]) cae (a penas) dentro de este intervalo, se puede suponer que no existe una diferencia significativa entre los tiempos promedio requeridos por ambos sistemas, por lo que no se rechaza H_0. Así, tenemos un 95 % de confianza en que no existe una diferencia entre los tiempos que requieren ambos sistemas para procesar transacciones. Sin embargo, esta decisión es un tanto apresurada ya que el resultado está cerca del borde de rechazo y, en este caso, lo lógico sería investigar más (hacer crecer la muestra).

Revisemos ahora el caso planteado con hipótesis alternativa unilateral (es decir, queremos ver si el nuevo sistema es, en efecto, más rápido). Manteniendo nuestro nivel de confianza $1-\alpha = 0,95$, en este caso debemos considerar los valores menores a $\mu_A = 530$ [ms] para el cálculo de z^*. En otras palabras, el 5 % que descartamos corresponde únicamente a la cola superior. Así, nuestro valor para z^* está dado por la llamada `qnorm(0.05, mean = 0, sd = 1, lower.tail = FALSE)`, obteniéndose $z^* = 1,64$ (aprox.) por lo que se tiene que la cota superior es:

$$\overline{x}_N - z^* \cdot SE_{\overline{x}} = 527,9 - 1,64 \cdot 1,2 = 529.874$$

Luego, el intervalo de confianza va desde "cualquier valor" bajo la media observada en la muestra hasta el valor calculado arriba, por lo que el intervalo con 95 % confianza sería: $[-\infty; 529,874]$.

Ahora el valor $\mu_A = 530$ [ms] cae (apenas) fuera del intervalo y podemos decir que existe evidencia de que el nuevo sistema tarda en promedio menos tiempo que el antiguo en procesar las transacciones.

Ahora bien, siempre que se prueban hipótesis podemos cometer un error al momento de decidir si rechazar o no la hipótesis nula. Afortunadamente, la estadística ofrece herramientas para cuantificar cuán frecuentes son dichos errores. Existen cuatro posibles escenarios, los cuales se presentan en la tabla 4.1. El **error tipo I** corresponde a rechazar H_0 cuando en realidad es verdadera, mientras que el **error tipo II** corresponde a no rechazarla cuando en realidad H_A es verdadera.

		Conclusión de la prueba	
		No rechazar H_0	Rechazar H_0 en favor de H_A
Verdad	H_0 verdadera	Decisión correcta	Error tipo I
	H_A verdadera	Error tipo II	Decisión correcta

Tabla 4.1: posibles escenarios para una prueba de hipótesis.

Como ya hemos señalado, la prueba de hipótesis se basa en no rechazar H_0 a menos que se tenga evidencia contundente. Por regla general, no se desea cometer el error de rechazar incorrectamente la hipótesis nula (error tipo I) en más de 5 % de los casos. Esto corresponde a un **nivel de significación** de 0,05, denotado por $\alpha = 0,05$. Si usamos un intervalo de confianza de 95 % para evaluar una prueba de hipótesis en que la

hipótesis nula es verdadera, cometeremos un error tipo I cada vez que el estimador puntual esté a 1,96 o más errores estándar del parámetro de la población. Esto puede ocurrir un 5 % de las veces (2,5 % en cada cola de la distribución para el caso bilateral). Del mismo modo, un intervalo de confianza del 99 % es equivalente a un nivel de significación $\alpha = 0,01$.

El intervalo de confianza es de mucha ayuda para decidir si rechazar o no H_0. No obstante, no aporta información directa acerca de cuán fuerte es la evidencia para la decisión tomada.

4.5.1 Prueba formal de hipótesis con valores p

Antes de que la computación se hiciera masiva, las personas tenían dos procedimientos posibles para decidir una prueba de hipótesis. El primero es el realizado en la sección anterior, esto es, calcular el intervalo con $(1 - \alpha)$ % de confianza de acuerdo a los estadísticos observados en una muestra y revisar si el valor nulo cae o no dentro de este intervalo. El otro procedimiento clásico, que podemos encontrar en muchos libros y sitios en Internet, es estimar a qué valor z corresponde la media observada en la distribución normal estandarizada que define el valor nulo y el error estándar: si este estadístico z es mayor que z^*, entonces el estadístico cae en una "zona de rechazo" de H_0; en caso contrario ($|z| < z^*$), se falla en rechazar la hipótesis nula.

Si bien estos procedimientos siguen siendo útiles, su diseño respondía a la existencia de **tablas de probabilidad** en que se tabulaban probabilidades para algunos valores de percentiles de uso común, como 90 %, 95 %, 0,975 % o 0,99 %.

Con la llegada de los computadores, y en particular de entornos como R, es posible obtener probabilidades (casi) exactas para cualquier percentil. Esto hizo que un tercer método para decidir una prueba de hipótesis haya ido ganando popularidad: el uso del **valor p**, también llamado **p-valor**, que es definido por Diez y col. (2017, p. 186) como "la probabilidad de observar datos al menos tan favorables como la muestra actual para la hipótesis alternativa, si la hipótesis nula es verdadera". De esta forma, un p-valor permite cuantificar cuán fuerte es la evidencia en contra de la hipótesis nula (y en favor de la hipótesis alternativa).

Consideremos ahora el escenario de la hipótesis unilateral del ejemplo, con un nivel de significación $\alpha = 0,05$, bajo el supuesto de que H_0 es verdadera y que la muestra a su vez tiene una distribución cercana a la normal. Recordemos que $\overline{x}_N = 527,9$ [ms] y $s_N = 48$ [ms] en $n = 1600$ observaciones. Esta distribución se vería como muestra la figura 4.3, creada mediante el script 4.3.

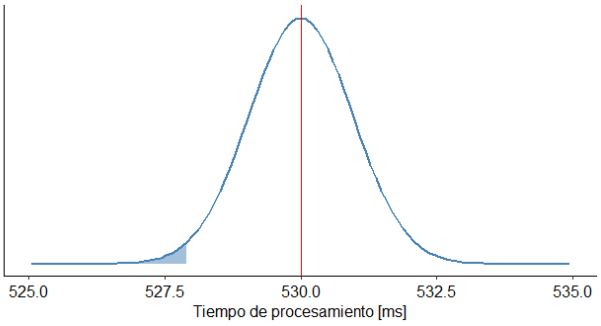


Figura 4.3: probabilidad de encontrar una media igual o menor que $\overline{x} = 527,9$ [ms] en la distribución muestral con $\mu_{\overline{x}} = 530$ y $\sigma_{\overline{x}} = 1,2$.

En este punto, resulta importante hacer una aclaración en relación al valor p. El área bajo la sección de la curva con valores menores o iguales a un estimador se calcula usando para ello el **valor z**, definido en la ecuación 4.5, como **estadístico de prueba**.

$$z = \frac{\text{estimador puntual} - \text{valor nulo}}{SE_{\text{estimador puntual}}} = \frac{\hat{\theta} - \theta_0}{SE_{\hat{\theta}}} \tag{4.5}$$

Un **estadístico de prueba** es un estadístico de resumen que resulta especialmente útil para evaluar hipótesis o calcular el valor p. El valor z se usa cuando el estimador puntual se acerca a la normalidad, aunque existen otros estadísticos de prueba adecuados para otros escenarios.

Script 4.3: cálculo del valor p para una prueba de una cola.

```
library(ggpubr)

# Generar una muestra donde la media cumpla con la hipótesis nula.
set.seed(872)

media_poblacion_antiguo <- 530
media_muestra_nuevo <- 527.9
desv_est <- 48
n <- 1600
error_est <- desv_est / sqrt(n)

x <- seq(media_poblacion_antiguo - 5.2 * error_est,
         media_poblacion_antiguo + 5.2 * error_est,
         0.01)

y <- dnorm(x, mean = media_poblacion_antiguo, sd = error_est)

datos <- data.frame(x, y)

# Graficar la muestra.
g <- ggplot(data = datos, aes(x))

g <- g + stat_function(fun = dnorm,
                       args = list(mean = media_poblacion_antiguo,
                                   sd = error_est),
                       colour = "steelblue", size = 1)

g <- g + ylab("")
g <- g + scale_y_continuous(breaks = NULL)
g <- g + scale_x_continuous(name = "Tiempo de procesamiento [ms]")
g <- g + theme_pubr()

# Colorear el área igual o menor que la media observada.
g <- g + geom_area(data = subset(datos,
                                 x < media_muestra_nuevo),
                   aes(y = y),
                   colour = "steelblue",
                   fill = "steelblue",
                   alpha = 0.5)

# Agregar una línea vertical para el valor nulo.
g <- g + geom_vline(aes(xintercept = media_poblacion_antiguo),
                    color = "red", linetype = 1)

print(g)

# Calcular el valor Z para la muestra.
Z <- (media_muestra_nuevo - media_poblacion_antiguo) / error_est

```

```r
# Calcular el valor p.
p_1 <- pnorm(Z, lower.tail = TRUE)

cat("Valor p: ", p_1, "\n")

# También se puede calcular el valor p directamente a partir de la
# distribución muestral definida por el valor nulo y el error
# estándar.
p_2 <- pnorm(media_muestra_nuevo, mean = media_poblacion_antiguo,
             sd = est_err)

cat("Valor p: ", p_2)
```

El valor p, en este caso $p = 0,040$, corresponde al área coloreada en la figura 4.3, y se calcula en la línea 51 del script 4.3. Esto nos indica, en este caso, que si H_0 fuera verdadera y el nuevo sistema tarda en promedio lo mismo que el antiguo en procesar las transacciones, la probabilidad de encontrar una media de a lo más 527,9 [ms] para una muestra de 1.600 transacciones es de 4 %, lo que sería bastante poco frecuente.

Cuanto menor sea el valor p, más fuerte será la evidencia en favor de H_A por sobre H_0. Y aquí la ventaja de usar este método para decidir: el valor p se puede **comparar directamente** con el nivel de significación α, y si p es menor que el nivel de significación se considera evidencia suficiente para rechazar la hipótesis nula en favor de la hipótesis alternativa. En este ejemplo, $p = 0,040 < \alpha = 0,05$, por lo que se falla al rechazar H_0 en favor de H_A. Pero como se dijo cuando usamos intervalos de confianza, el valor p está cerca del valor α y convendría ser menos tajante en la decisión y evaluar la posibilidad de ampliar la muestra para conseguir evidencia más definitiva.

Siempre es recomendable formular la conclusión de la prueba de hipótesis en lenguaje llano, para facilitar su comprensión. Así, en este caso concluimos que los datos sugieren que el nuevo sistema tarda menos que el antiguo en procesar transacciones, pero que es necesario hacer un estudio con más observaciones para tener un diagnóstico más definitivo.

Volvamos nuevamente al escenario de la prueba de hipótesis bilateral para el ejemplo, manteniendo el nivel de significación $\alpha = 0,05$. Puesto que en este caso nos interesa la diferencia en ambas direcciones, ya que la evidencia en ambas direcciones es favorable para H_A, debemos considerar el área bajo las dos colas de la curva normal, a diferencia del caso de la prueba de hipótesis unilateral en que solo se consideramos la cola correspondiente a la dirección de interés de la diferencia. Dado que el modelo normal es simétrico, el área bajo ambas colas es la misma (figura 4.4, script 4.4). El valor p, entonces, ahora es igual a dos veces el área de la cola inferior, es decir, $p = 0,080$. Puesto que $p > \alpha$, se falla en rechazar H_0. Es decir, no hay evidencia suficiente para concluir que existe una diferencia entre los tiempos promedio requeridos por ambos sistemas para procesar transacciones.

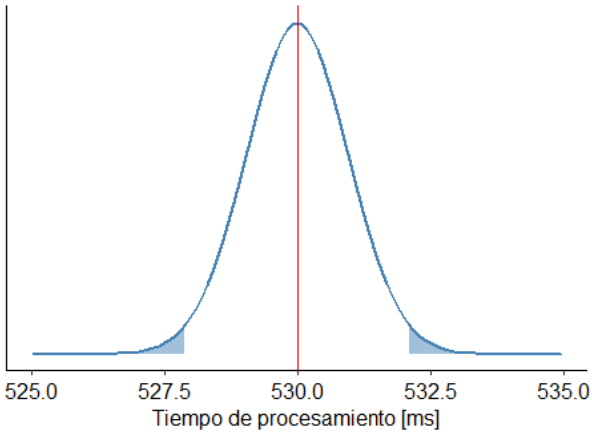


Figura 4.4: cuando la prueba de hipótesis es bilateral, se deben colorear ambas colas.

Script 4.4: cálculo del valor p para una prueba de dos colas.

```
library(ggpubr)

# Generar una muestra donde la media cumpla con la hipótesis nula.
set.seed(208)

media_poblacion_antiguo <- 530
media_muestra_nuevo <- 527.9
desv_est <- 48
n <- 1600
error_est <- desv_est / sqrt(n)

x <- seq(media_poblacion_antiguo - 5.2 * error_est,
         media_poblacion_antiguo + 5.2 * error_est,
         0.01)

y <- dnorm(x,
           mean = media_poblacion_antiguo,
           sd = error_est)

dataframe <- data.frame(x, y)

# Graficar la muestra.
g <- ggplot(data = dataframe, aes(x))

g <- g + stat_function(fun = dnorm,
                       args = list(mean = media_poblacion_antiguo,
                                   sd = error_est),
                       colour = "steelblue", size = 1)

g <- g + ylab("")
g <- g + scale_y_continuous(breaks = NULL)
g <- g + scale_x_continuous(name = "Tiempo de procesamiento [ms]")
g <- g + theme_pubr()

# Colorear el área igual o menor que la media observada.
g <- g + geom_area(data = subset(dataframe,
                                 x < media_muestra_nuevo),
                   aes(y = y),
                   colour = "steelblue",
                   fill = "steelblue",
                   alpha = 0.5)

# Calcular el área bajo la cola inferior.
area_inferior <- pnorm(media_muestra_nuevo,
                       mean = media_poblacion_antiguo,
                       sd = desv_est)


# Colorear igual área en la cola restante.
corte_x <- qnorm(1 - area_inferior,
                 mean = media_poblacion_antiguo,
                 sd = desv_est)

g <- g + geom_area(data = subset(dataframe,
                                 x > corte_x),
                   aes(y = y),
                   colour = "steelblue",
                   fill = "steelblue",
```

```
                    alpha = 0.5)

# Agregar una línea vertical para el valor nulo.
g <- g + geom_vline(aes(xintercept = media_poblacion_antiguo),
                    color = "red", linetype = 1)

print(g)

# Calcular el valor Z para la muestra.
Z <- (media_muestra_nuevo - media_poblacion_antiguo) / error_est

# Calcular el valor p (recordando ahora que la hipótesis es bilateral).
p <- 2 * pnorm(Z, lower.tail = TRUE)

cat("Valor p: ", p)
```

Un punto importante que debemos tener en cuenta es que **las pruebas unilaterales** se usan cuando se desea verificar un incremento o un decremento, pero no ambas. No obstante, esta decisión debe tomarse siempre **antes de examinar los datos**, pues de lo contrario se duplica la probabilidad de cometer errores de tipo I y se está cayendo en **prácticas poco éticas**.

4.5.2 El efecto del nivel de significación

Hemos visto que el nivel de significación (α) representa la proporción de veces en que se cometería un error de tipo I (es decir, rechazar H_0 en favor de H_A, cuando H_0 es en realidad verdadera). Si resulta costoso o peligroso cometer un error de este tipo, debemos requerir evidencia más fuerte para rechazar la hipótesis nula (es decir, reducir la probabilidad de que esto ocurra), lo que podemos lograr usando un valor más pequeño para el nivel de significación, por ejemplo, $\alpha = 0,01$. Sin embargo, esto necesariamente **aumentará** la probabilidad de cometer un error de tipo II.

Si, por el contrario, el costo o el peligro de cometer un error de tipo II (no rechazar H_0 cuando en realidad H_A es verdadera) es mayor, debemos escoger un nivel de significación más elevado (por ejemplo, $\alpha = 0,10$).

Así, **el nivel de significación seleccionado para una prueba siempre debe reflejar las consecuencias de cometer errores de tipo I o de tipo II**.

4.6 INFERENCIA PARA OTROS ESTIMADORES

Hasta ahora, solo hemos considerado la media como estimador para la inferencia. No obstante, muchos de los conceptos que hemos visto en este capítulo pueden aplicarse, con algunas ligeras modificaciones, usando otros estimadores.

4.6.1 Estimadores puntuales con distribución cercana a la normal

En realidad existen múltiples estimadores puntuales, además de la media, cuya distribución muestral es cercana a la normal si las muestras son lo suficientemente grandes, tales como las proporciones y la diferencia de medias. Si bien veremos con detalle la prueba de hipótesis con estos estimadores puntuales en capítulos posteriores, es importante contar con algunas orientaciones generales.

Un supuesto importante que debemos tener en cuenta es que el estimador puntual $\hat{\theta}$ debe ser **insesgado**. Esto significa que la distribución muestral de $\hat{\theta}$ tiene su centro en el valor del parámetro θ que estima. En otras palabras, un estimador insesgado (como la media) tiende a proveer una estimación cercana al parámetro real.

En términos generales, el intervalo de confianza para un estimador puntual insesgado cuya distribución es cercana a la normal (como la media, las proporciones o la diferencia de medias) está dado por la ecuación 4.6, donde z^* se escoge de manera tal que se condiga con el nivel de confianza seleccionado y y la lateralidad de la hipótesis alternativa. Como se dijo anteriormente, el valor $z^* \cdot SE_{\hat{\theta}}$ se denomina "margen de error". Debemos recordar que la ecuación 4.2 corresponde al error estándar de la media, pero los errores estándar para otros estimadores puntuales se estiman de manera diferente a partir de los datos.

$$\hat{\theta} \pm z^* \cdot SE_{\hat{\theta}} \tag{4.6}$$

El método de prueba de hipótesis usando valores p puede generalizarse para otros estimadores puntuales con distribución cercana a la normal. Para ello, Diez y col. (2017, p. 199) señalan que se debemos considerar los siguientes pasos:

Prueba de hipótesis usando el modelo normal:

1. Formular las hipótesis nula (H_0) y alternativa (H_A) en lenguaje llano y luego en notación matemática.
2. Identificar un estimador puntual (estadístico) adecuado e insesgado para el parámetro de interés.
3. Verificar las condiciones para garantizar que la estimación del error estándar sea razonable y que la distribución muestral del estimador puntual siga aproximadamente una distribución normal.
4. Calcular el error estándar. Luego, graficar la distribución muestral del estadístico bajo el supuesto de que H_0 es verdadera y sombrear las áreas que representan el valor p.
5. Usando el gráfico y el modelo normal, calcular el valor p para evaluar las hipótesis y escribir la conclusión en lenguaje llano.

4.6.2 Estimadores con otras distribuciones

Existen métodos de construcción de intervalos de confianza y prueba de hipótesis adecuados para aquellos casos en que el estimador puntual o el estadístico de prueba no son cercanos a la normal (por ejemplo, si la muestra es pequeña, se tiene una mala estimación del error estándar o el estimador puntual tiene una distribución distinta a la normal). No obstante, la selección de métodos alternativos debe hacerse siempre teniendo en cuenta la distribución muestral del estimador puntual o del estadístico de prueba.

Una consideración importante es que **siempre debemos verificar el cumplimiento de las condiciones requeridas por una herramienta estadística**, pues de lo contrario las conclusiones pueden ser erradas y carecerán de validez.

4.7 EJERCICIOS PROPUESTOS

1. ¿Es correcto afirmar que, si se lanza un dado una y otra vez, la media móvil simple del número de puntos que aparecen en la cara superior crece monótonamente? Justifica tu respuesta.
2. ¿Es correcto afirmar que, si se lanza un dado una y otra vez, la proporción de veces que aparece un número impar de puntos (1, 3 o 5) en la cara superior es siempre 0,5? Justifica su respuesta.
3. Si se calcula la media de diez muestras distintas extraídas de la misma población, ¿se espera ver el mismo valor cada vez? ¿Cómo se llama a este fenómeno?
4. Completa las siguiente oraciones:
   a) Una estimación _____ _____ es un _____ calculado con datos de una muestra como aproximación del valor desconocido de un _____ de la población en estudio.
   b) $\overline{X}$ o $\overline{x}$ se usan para denotar la _____ _____, que es una estimación puntual de μ, la _____ _____.
5. Se sabe que una prueba para medir el coeficiente intelectual de jóvenes de 18 años produce puntuaciones que siguen una distribución $\mathcal{N}(\mu = 100, \sigma^2 = 100)$.
   a) Dibuja el histograma de la distribución muestral de medias para muestras de tamaño 25 de esta población.
   b) Una de las muestras anteriores presentó $\overline{x} = 95$ y $s = 15$. Determina el intervalo con 95 % de confianza para este caso.
   c) Con otra de las muestras se pudo determinar que su intervalo con 99 % confianza era $[90,26; 105,74]$. ¿Qué significa esto?
   d) El intervalo anterior, ¿es más grande o más pequeño que uno con 90 % de confianza?
6. Una empresa de tecnología quiere promocionar un software especializado para almacenar y recuperar imágenes médicas digitales. Con esta idea, está financiando un estudio para determinar el tiempo (en segundos) que necesita un grupo de médicos para recuperar imágenes desde sus propios registros en sus portátiles personales y desde la base de datos central con el software ofrecido y una conexión a la Web.
   a) Enuncia las hipótesis nula y alternativa (en castellano común).
   b) Identifica la variable aleatoria que se va a estudiar, el parámetro de interés y el correspondiente estadístico.
   c) Enuncia, más formalmente, las hipótesis nula y alternativa para este caso.
   d) Supón que el intervalo con 95 % confianza para el tiempo de recuperación promedio de una imagen digital desde la base de datos central resultó ser $[24; 36]$ [s]. ¿Qué decisión tomarías ante la hipótesis nula: la media del tiempo de recuperación de una imagen digital con el nuevo software es de 25 segundos? En este caso, ¿cuál podría ser la hipótesis alternativa?
   e) Para el intervalo de confianza anterior, ¿cuál sería un error de tipo I?
   f) Conociendo el intervalo de confianza anterior, ¿es posible cometer un error de tipo II? Explica.
7. Si una hipótesis nula es falsa, aumentar el nivel de significación para un tamaño de muestra dado, ¿reduce la probabilidad de rechazarla?
8. Qué significa que un estadístico tenga un valor p de 0,025?
9. Si una hipótesis nula es rechazada a un nivel de significación de 0,01, ¿será rechazada a un nivel de significación 0,05? Explica.
10. Si una hipótesis nula es rechazada por una prueba unilateral (una cola), ¿será también rechazada por una prueba bilateral (dos colas)? Explica.
11. Acabas de leer un artículo que hace la siguiente aseveración: "*a 95 % confidence interval for mean reaction time is from 0.25 to 0.29 seconds. Thus, about 95 % of individuals will have reaction times in this interval.*" Comenta.
12. Da el ejemplo de un estudio en que es más dañino cometer un error tipo II que un error tipo I.
13. Lista las condiciones que deben verificarse para asegurar que el TLC (teorema del límite central) está rigiendo y es posible hacer una prueba de hipótesis o calcular un intervalo de confianza.
14. Si para un estudio de una determinada variable aleatoria numérica es igualmente dañino cometer errores de tipo I como errores tipo II:
   a) Dibuja la distribución de una muestra de tamaño 16 (un diagrama de caja, por ejemplo) para la que el contraste de hipótesis con nivel de significación 0,05 sea confiable.

b) Dibuja la distribución de una muestra de tamaño 30 en que se requiera de un nivel de significación más exigente ($\alpha < 0,05$) para hacer el contraste de hipótesis más confiable.

c) Dibuja la distribución de una muestra en que es mejor no confiar en el contraste de hipótesis con métodos estudiados hasta ahora.

15. Si un estudio sobre el tiempo promedio de búsqueda y recuperación de imágenes médicas con dos tecnologías distintas reporta: "existe una diferencia significativa (p <0,02) entre el tiempo invertido con la tecnología A ($33 \pm 4[s]$) que con la tecnología B ($30 \pm 6[s]$)", ¿significa que se debe adoptar la tecnología B? ¿Por qué?
16. Explica por qué se incrementa la probabilidad de cometer errores tipo I al cambiar de una prueba de hipótesis bilateral a otra unilateral.

CAPÍTULO 5. INFERENCIA CON MEDIAS MUESTRALES

En el capítulo 4 conocimos los principios de la inferencia y definimos los principales conceptos involucrados. En dicho capítulo conocimos el modelo normal, es decir, que la distribución muestral de la media sigue aproximadamente una distribución normal, supuesto que en general se cumple si la muestra tiene a lo menos 30 observaciones.

Veremos que diversas pruebas estadísticas consideran el modelo normal, aunque otras consideran estadísticos (estimaciones puntuales) diferentes que siguen otras distribuciones que ya conocimos en el capítulo 6.

En este capítulo veremos nuestras primeras pruebas estadísticas, las cuales nos permitirán inferir acerca de una o dos medias muestrales. Para ello nos basaremos principalmente en las explicaciones que ofrecen Diez y col. (2017, pp. 219-239) y Meena (2020).

5.1 PRUEBA Z

Como ya adelantamos, la prueba Z es adecuada para inferir acerca de las medias con una o dos muestras, aunque aquí solo veremos el primer caso. Para poder usarla, debemos **verificar el cumplimiento** de algunas condiciones, muchas de las cuales están asociadas al modelo normal que conocimos en el capítulo anterior:

- La muestra debe tener al menos 30 observaciones. Si la muestra tiene menos de 30 observaciones, se debe conocer la varianza de la población.
- Las observaciones deben ser independientes, es decir que la elección de una observación para la muestra no influye en la selección de las otras.
- La población de donde se obtuvo la muestra sigue aproximadamente una distribución normal.

Esta prueba resulta adecuada si queremos **asegurar** o **descartar** que la media de la población tiene un cierto **valor hipotético**. Supongamos que queremos saber si, en promedio, las utilidades mensuales de una pequeña empresa son de 20 millones de pesos y que el gerente general, Esteban Quito, nos ha informado que la desviación estándar para las utilidades es de 2,32 millones de pesos. El Sr. Quito nos ha proporcionado una muestra, obtenida mediante muestreo aleatorio simple, con las utilidades (en millones de pesos) reportadas para 20 meses, que se muestra en la tabla 5.1.

Obs.	Utilidad [M$]	Obs.	Utilidad [M$]	Obs.	Utilidad [M$]	Obs.	Utilidad [M$]
1	19,33	6	22,22	11	22,55	16	29,68
2	29,37	7	31,26	12	20,69	17	29,27
3	29,14	8	26,92	13	24,68	18	26,72
4	32,10	9	31,40	14	28,74	19	27,08
5	25,04	10	17,66	15	26,85	20	20,62

Tabla 5.1: muestra para el ejemplo de prueba Z con una muestra.

El Sr. Quito nos ha dicho que debemos ser muy exigentes con respecto a nuestras conclusiones, por lo que se decide usar un nivel de significación $\alpha = 0,01$ (es decir, un nivel de confianza de 99 %).

Comencemos por formular nuestras hipótesis:

H_0: la media de las utilidades mensuales de la empresa (μ) es de 20 millones de pesos, es decir: $\mu = 20$ [M$].

H_A: las utilidades mensuales de la empresa son, en promedio, distintas de 20 millones de pesos, es decir: $\mu \neq 20$ [M$].

Ahora debemos verificar el cumplimiento de las condiciones para poder usar la prueba Z. En cuanto a la primera condición, el enunciado nos indica que, si bien la muestra tiene solo 20 observaciones, la desviación estándar de la población es conocida, por lo que se verifica su cumplimiento.

También podemos comprobar en el enunciado que las observaciones son independientes entre sí, pues fueron obtenidas mediante muestreo aleatorio simple. Si bien no estamos seguros que que esta muestra considere menos del 10 % de las observaciones, podemos suponerlo razonablemente.

En cuanto a la distribución de la muestra, el gráfico Q-Q de la figura 5.1 (obtenido mediante el script 5.1) nos muestra que no se observan valores atípicos. Otra forma de comprobar esta condición es mediante la prueba de Shapiro-Wilk (Parada, 2019), que podemos realizar en R mediante la función `shapiro.test(x)`, donde `x` es un vector con las observaciones de la muestra. La hipótesis nula de esta prueba es que la muestra fue extraída desde una distribución normal (por ende, la hipótesis alternativa es que la distribución detrás de la muestra es diferente a la normal). Al ejecutar el script, podemos ver que el valor p obtenido es $p = 0,244$, muy superior a nuestro nivel de significación, por lo que podemos suponer con relativa confianza que la población de donde proviene la muestra sigue una distribución muestral.

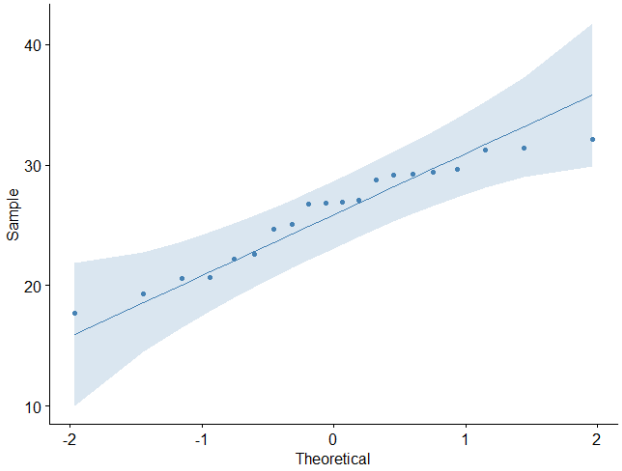


Figura 5.1: gráfico Q-Q para la muestra de la tabla 5.1.

Puesto que hemos comprobado que se cumplen todas las condiciones, podemos hacer una prueba Z para una muestra. Comencemos por calcular ahora el **estadístico de prueba** como ya hemos estudiado, usando para ello la ecuación 6.7:

$$Z = \frac{\overline{x} - \mu}{\sigma} = \frac{26,066 - 20}{2,32} = 2.6147$$

Con este resultado calculamos el valor p. Debemos recordar que las funciones de R (al igual que las antiguas tablas de probabilidades) nos entregan la probabilidad asociada al área correspondiente a una sola cola de la distribución, por lo que debemos multiplicar el resultado por 2 para considerar ambas colas si, como en este caso, se trata de una prueba bilateral. Al hacer la llamada `2 * pnorm(2.6147, lower.tail = FALSE)`, obtenemos que $p = 0,009 < 0.01$[1], con lo que se rechaza la hipótesis nula en favor de la hipótesis alternativa. Sin embargo, debemos ser cuidadosos puesto que el valor p es bastante cercano al nivel de significación establecido, por lo que sería prudente evaluar los resultados con una muestra más grande. Así, concluimos que los datos **sugieren** que, en promedio, las utilidades mensuales de la empresa difieren de los 20 millones de pesos establecidos.

[1]Más precisamente, $p = 0.008931758$, pero, por convención, p-valores suelen reportarse con tres decimales.

Desde luego, gracias a R podemos realizar esta prueba simplemente con una llamada a la función `z.test(x, mu, stdev, alternative, conf.level)`, disponible en el paquete `TeachingDemos`, donde:

- `x`: vector con las observaciones de la muestra.
- `mu`: valor nulo.
- `stdev`: desviación estándar de la población.
- `alternative`: tipo de hipótesis alternativa. Puede tomar los valores ‘‘`two.sided`’’ (hipótesis bilateral), ‘‘`less`’’ (hipótesis unilateral que la media de la población es menor que el valor nulo) o ‘‘`greater`’’ (hipótesis unilateral que la media de la población es mayor que el valor nulo).
- `conf.level`: nivel de confianza.

El script 5.1 muestra el desarrollo de este ejemplo en forma manual y luego, en la línea 42, usando la función `z.test()`. El resultado que se obtiene al usar esta función es el que se muestra en la figura 5.2, idéntico al obtenido en nuestro desarrollo previo.

```
     One Sample z-test

data:  media
z = 2.6147, n = 1.00, Std. Dev. = 2.32, Std. Dev. of the sample mean = 2.32, p-value = 0.008932
alternative hypothesis: true mean is not equal to 20
99 percent confidence interval:
 20.09008 32.04192
sample estimates:
mean of media
       26.066
```

Figura 5.2: resultado de la prueba Z para una muestra.

Script 5.1: prueba Z para una muestra.

```
1  library(TeachingDemos)
2  library(ggpubr)
3
4  # Ingresar los datos.
5  muestra <- c(19.33, 29.37, 29.14, 32.10, 25.04, 22.22, 31.26, 26.92,
6               31.40, 17.66, 22.55, 20.69, 24.68, 28.74, 26.85, 29.68,
7               29.27, 26.72, 27.08, 20.62)
8
9  # Establecer los datos conocidos.
10 desv_est <- 2.32
11 n <- length(muestra)
12 valor_nulo <- 20
13
14 # Crear gráfico Q-Q para verificar la distribución de la muestra,
15 datos <- data.frame(muestra)
16
17 g <- ggqqplot(datos, x = "muestra", color = "SteelBlue")
18 print(g)
19
20 # Verificar distribución muestral usando la prueba de normalidad
21 # de Shapiro-Wilk.
22 normalidad <- shapiro.test(muestra)
23 print(normalidad)
24
25 # Fijar un nivel de significación.
26 alfa <- 0.01
27
28 # Calcular la media de la muestra.
```

```r
cat("\tPrueba Z para una muestra\n\n")
media <- mean(muestra)
cat("Media =", media, "M$\n")

# Calcular el estadístico de prueba.
Z <- (media - valor_nulo) / desv_est
cat("Z =", Z, "\n")

# Calcular el valor p.
p <- 2 * pnorm(Z, lower.tail = FALSE)
cat("p =", p, "\n")

# Hacer la prueba Z con R.
prueba <- z.test(media, mu = valor_nulo, alternative = "two.sided",
                 stdev = desv_est, conf.level = 1-alfa)

print(prueba)
```

5.2 PRUEBA T DE STUDENT

En la práctica, rara vez podemos conocer la desviación estándar de la población y a menudo nos encontraremos con muestras pequeñas, por lo que la prueba Z no es muy utilizada.

En el caso de la media, el teorema del límite central se cumple para datos normales, es decir, independientemente del tamaño de la muestra, la media muestral tendrá una distribución cercana a la normal siempre que las observaciones sean independientes y provengan de una distribución cercana a la normal. Sin embargo, cuando el conjunto de datos es pequeño, resulta muy difícil comprobar el cumplimiento de estas condiciones.

En el capítulo 6 conocimos la distribución t de Student, o simplemente distribución t. Vimos que un aspecto destacado de esta distribución, siempre centrada en 0 y definida únicamente por los grados de libertad (ν) como parámetro, es su semejanza con la distribución normal pese a que sus colas son algo más gruesas. Este grosor adicional de las colas tiene como consecuencia que, para la distribución t, es más probable que una observación esté a más de dos desviaciones estándares de la media que en el caso de la distribución normal. Este fenómeno permite que la estimación del error estándar sea más certera que al usar la distribución normal cuando el conjunto de datos es pequeño.

La prueba t de Student, basada en la distribución t, es en consecuencia la alternativa más ampliamente empleada para inferir acerca de una o dos medias muestrales.

5.2.1 Prueba t para una muestra

Aunque la prueba t no opera bajo el supuesto de normalidad, aún así requiere verificar algunas condiciones para poder usarla:

1. Las observaciones son independientes entre sí.
2. Las observaciones provienen de una distribución cercana a la normal.

Podemos ver que estas condiciones son casi las mismas que para la prueba Z, excepto por el hecho de que no limitan el tamaño de la muestra para que sea mayor a 30. La ventaja evidente de eliminar esta restricción es

que la distribución t permite su uso para muestras pequeñas, pero es igualmente adecuada cuando la muestra es grande. Esto se debe a que la forma de la distribución t es regulada por los grados de libertad y, a medida que aumentan, más se parece a una distribución normal. Este parámetro, al trabajar con medias de muestras de tamaño n, siempre estará dado por $\nu = n - 1$.

Tomemos el siguiente problema para ilustrar la prueba de hipótesis para la media de una muestra usando el modelo t: un ingeniero en Informática necesita determinar si el tiempo promedio que tarda una implementación dada de un algoritmo en resolver un problema, sabiendo que el algoritmo siempre se ejecuta en las mismas condiciones (misma máquina, igual disponibilidad de recursos de hardware y tamaño constante de las instancias), es inferior a 500 milisegundos. Para ello, ha seleccionado aleatoriamente 15 instancias del problema y registrado el tiempo de ejecución del algoritmo (en milisegundos) para cada una de ellas, como muestra la tabla 5.2.

Obs.	**t [ms]**	**Obs.**	**t [ms]**	**Obs.**	**t [ms]**
1	411,5538	6	388,6731	11	418,1169
2	393,2753	7	430,0382	12	408,4110
3	445,8905	8	469,4734	13	463,3733
4	411,4022	9	409,5844	14	407,0908
5	498,8969	10	442,0800	15	516,5222

Tabla 5.2: tiempo de ejecución para las instancias de la muestra.

El primer paso es formular las hipótesis:

H_0: el tiempo promedio que tarda el algoritmo en resolver una instancia del problema es igual a 500 milisegundos.
H_A: el tiempo promedio que tarda el algoritmo en resolver una instancia del problema es inferior a 500 milisegundos.

Recordemos que μ_0 es el valor nulo, por lo que en este caso $\mu_0 = 500$ [ms]. Matemáticamente, las hipótesis anteriores pueden formularse como:

Denotando como μ al tiempo medio que tarda la implementación del algoritmo en resolver una instancia cualquiera del problema:
H_0: $\mu = \mu_0$, esto es $\mu = 500$
H_A: $\mu < \mu_0$, es decir $\mu < 500$

Ahora debemos verificar que se cumplen las condiciones necesarias para usar la distribución t:

- Como las muestras fueron elegidas al azar, se puede asumir que son independientes.
- El gráfico de la figura 5.3 muestra que es válido suponer una distribución cercana a la normal. Si bien los puntos de la muestra no forman una recta, no se observan valores atípicos que se alejen de la región aceptable.

La media de la muestra es de $\overline{x} = 434,2921$, y la desviación estándar, $s = 38,0963$.

En este caso, el estadístico de prueba es el estadístico T, el cual sigue una distribución t con $\nu = n-1$ grados de libertad y está dado por la ecuación 5.1, donde la subexpresión $(s/\sqrt{n})$ corresponde al error estándar de la media (cuando no se conoce la desviación estándar de la población, σ).

$$T = \frac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}} \tag{5.1}$$

Así, para el ejemplo tenemos que:

$$T = \frac{434,2921 - 500}{\frac{38,0963}{\sqrt{15}}} = -6,6801$$

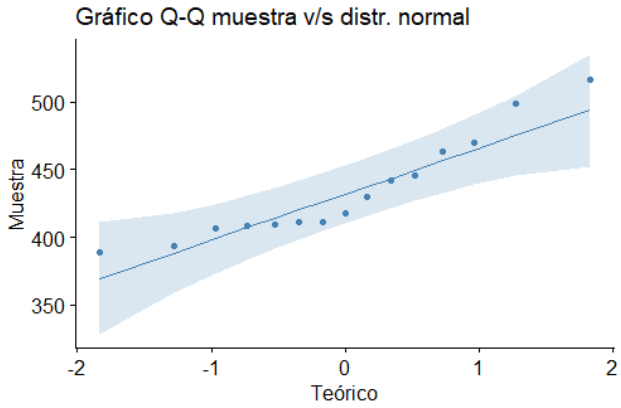


Figura 5.3: gráfico para comprobar el supuesto de normalidad.

A partir de este resultado, obtenemos el valor p con ayuda de la función `pt()`, obteniéndose $p = 5,219 \cdot 10^{-6}$, o simplemente, como dicta la convención, $p < 0,001$.

La fórmula para construir el intervalo de confianza usando la distribución t es ligeramente diferente al caso normal, como muestra la ecuación 5.2. Para este ejemplo consideraremos un nivel de confianza de 97,5 % (es decir, un nivel de significación $\alpha = 0,025$).

$$\overline{x} \pm t^*_\nu \cdot SE \tag{5.2}$$

Fijémonos en que en la ecuación 5.2 aparece el nuevo valor t^*_ν, el cual se obtiene a partir del nivel de confianza y la distribución t con ν grados de libertad (en este caso, $\nu = 14$), usando para ello una tabla de distribución t o la función `qt()` en R. Como puede verse al ejecutar el script 5.2, en este caso $t^*_\nu = 2,1448$.

Para el cálculo del error estándar, nuevamente se emplea la ecuación 4.2:

$$SE_{\overline{x}} = \frac{s}{\sqrt{n}} = \frac{38,0963}{\sqrt{15}} = 9,8364$$

Así, el intervalo de confianza está dado por:

$$(-\infty, t^*_{nu} * SE_{\overline{x}}] = (-\infty, 2,1448 * 9,8364] = (-\infty, 455,3892]$$

Una vez más, R permite realizar esta prueba de manera rápida y sencilla, gracias a la función `t.test(x, alternative, mu, conf.level)`, donde:

- `x`: vector no vacío de valores numéricos (la muestra).
- `alternative`: tipo de prueba de hipótesis. Los posibles valores son `"two.sided"` (prueba bilateral), `"greater"` (hipótesis unilateral que la media de la población es mayor que el valor nulo) o `"less"` (hipótesis unilateral que la media de la población es menor que el valor nulo).
- `mu`: valor nulo.
- `conf.level`: nivel de confianza.

El script 5.2 muestra el desarrollo en R para este ejemplo, incluyendo la construcción del gráfico de la figura 5.3, con iguales resultados al realizar la prueba paso a paso y con la función `t.test()`.

A partir de los resultados podemos observar que el valor p obtenido es muy pequeño, dando a entender que, si se cumple el supuesto de que la verdadera media es $\mu = 500$ [ms] (hipótesis nula), sería muy improbable obtener una media muestral de $\overline{x} = 434,2921$. Además, el valor p es muchísimo menor que el nivel de significación, por lo que la evidencia a favor de H_A es muy fuerte. En consecuencia, se rechaza H_0 en favor

de H_A. Se puede afirmar, con 97,5 % de confianza, que el tiempo promedio que tarda el algoritmo en resolver una instancia del problema es inferior a 500 milisegundos.

Script 5.2: prueba t para una muestra.

```r
library(ggpubr)

# Cargar los datos.
tiempo <- c(411.5538, 393.2753, 445.8905, 411.4022, 498.8969,
            388.6731, 430.0382, 469.4734, 409.5844, 442.0800,
            418.1169, 408.4110, 463.3733,407.0908, 516.5222)

# Establecer los datos conocidos.
n <- length(tiempo)
grados_libertad <- n - 1
valor_nulo <- 500


# Verificar si la distribución se acerca a la normal.
g <- ggqqplot(data = data.frame(tiempo),
              x = "tiempo",
              color = "steelblue",
              xlab = "Teórico",
              ylab = "Muestra",
              title = "Gráfico Q-Q muestra v/s distr. normal")

print(g)

# Fijar un nivel de significación.
alfa <- 0.025

# Calcular el estadístico de prueba.
cat("\tPrueba t para una muestra\n\n")
media <- mean(tiempo)
cat("Media =", media, "M$\n")
desv_est <- sd(tiempo)
error <- desv_est / sqrt(n)
t <- (media - valor_nulo) / error
cat("t =", t, "\n")

# Calcular el valor p.
p <- pt(t, df = grados_libertad, lower.tail = TRUE)
cat("p =", p, "\n")

# Construir el intervalo de confianza.
t_critico <- qt(alfa, df = grados_libertad, lower.tail = FALSE)
superior <- media + t_critico * error
cat("Intervalo de confianza = (-Inf, ", superior, "]\n", sep = "")

# Aplicar la prueba t de Student con la funcuón de R.
prueba <- t.test(tiempo,
                 alternative = "less",
                 mu = valor_nulo,
                 conf.level = 1 - alfa)

print(prueba)
```

5.2.2 Prueba t para dos muestras pareadas

Para esta prueba, supongamos ahora que el ingeniero en Informática del ejemplo anterior tiene dos algoritmos diferentes (A y B) que, en teoría, deberían tardar lo mismo en resolver un problema. Para ello, probó ambos algoritmos con 35 instancias del problema (elegidas al azar) de igual tamaño y registró los tiempos de ejecución (en milisegundos) de ambos algoritmos bajo iguales condiciones para cada una de ellas, además de calcular la diferencia en los tiempos de ejecución, como muestra la tabla 5.3. El ingeniero desea comprobar si efectivamente el rendimiento de ambos algoritmos es equivalente.

instancia	t_A **[ms]**	t_B **[ms]**	**dif [ms]**	**instancia**	t_A **[ms]**	t_B **[ms]**	**dif [ms]**
1	436,5736	408,5142	28,0594	19	438,5959	458,2536	-19,6577
2	470,7937	450,1075	20,6862	20	439,7409	474,9863	-35,2454
3	445,8354	490,2311	-44,3957	21	464,5916	496,0153	-31,4237
4	470,9810	513,6910	-42,7100	22	467,9926	485,8112	-17,8186
5	485,9394	467,6467	18,2927	23	415,3252	457,4253	-42,1001
6	464,6145	484,1897	-19,5752	24	495,4094	483,3700	12,0394
7	466,2139	465,9334	0,2805	25	493,7082	510,7131	-17,0049
8	468,9065	502,6670	-33,7605	26	433,1082	467,5739	-34,4657
9	473,8778	444,9693	28,9085	27	445,7433	482,5621	-36,8188
10	413,0639	456,3341	-43,2702	28	515,2049	453,5986	61,6063
11	496,8705	501,1443	-4,2738	29	441,9420	385,9391	56,0029
12	450,6578	471,7833	-21,1255	30	472,1396	548,7884	-76,6488
13	502,9759	441,1206	61,8553	31	451,2234	467,2533	-16,0299
14	465,6358	544,1575	-78,5217	32	476,5149	494,7049	-18,1900
15	437,6397	447,8844	-10,2447	33	440,7918	451,9716	-11,1798
16	458,8806	432,4108	26,4698	34	460,1070	522,3699	-62,2629
17	503,1435	477,1712	25,9723	35	450,1008	444,1270	5,9738
18	430,0524	482,4828	-52,4304				

Tabla 5.3: tiempos de ejecución de cada algoritmo para las instancias de la muestra.

Para este ejemplo, tenemos dos tiempos de ejecución diferentes para cada instancia del problema: uno con cada algoritmo. En consecuencia, los datos están **pareados**. Es decir, cada observación de un conjunto tiene una correspondencia o conexión especial con exactamente una observación del otro. Una forma de uso común para examinar datos pareados es usar la diferencia entre cada par de observaciones, para lo cual podemos usar la técnica de la distribución t (también llamada prueba t de Student) vista en la sección anterior.

La media de las diferencias es $\overline{x}_{dif} = -12,08591$ y la desviación estándar es $s_{dif} = 36,08183$.

Una vez más, comenzamos por formular las hipótesis:

H_0: la media de las diferencias en los tiempos de ejecución es igual a 0.
H_A: la media de las diferencias en los tiempos de ejecución es distinta de 0.

Que matemáticamente se expresan como:

Denotando la media de las diferencias en los tiempos de ejecución necesitados por ambos algoritmos para cualquier instancia del problema como μ_{dif}:
H_0: $\mu_{dif} = 0$
H_A: $\mu_{dif} \neq 0$

Como siguiente paso, verificamos el cumplimiento de las condiciones. Como las instancias fueron escogidas al azar, se puede suponer razonablemente que las observaciones son independientes, pues además el conjunto de instancias posibles es muy grande (o infinito) y las 35 seleccionadas no superan el 10 % de la población. Además, al aplicar una prueba de normalidad de Shapiro-Wilk (ver script 5.3, línea 23) se obtiene $p = 0,357$, con lo que podemos concluir que la diferencia en los tiempos de ejecución se acerca razonablemente a una

distribución normal. En consecuencia, podemos proceder con la prueba t de Student. El ingeniero no necesita ser especialmente riguroso, por lo que usaremos un nivel de confianza del 95 %.

En este caso, la función `t.test()` de R permite efectuar la prueba de dos maneras diferentes (con idéntico resultado), como muestra el script 5.3. La primera de ellas (línea 30) es aplicar la prueba t directamente a las diferencias, tal como en la sección anterior (es decir, una prueba t para una muestra). La segunda (línea 39) consiste en entregar a la función ambas muestras por separado e indicarle que están pareadas. En este caso, la llamada tiene la forma `t.test(x, y, paired, alternative, mu, conf.level)`, donde los argumentos son:

- `x`: vector de valores numéricos para la primera muestra).
- `y`: vector de valores numéricos para la segunda muestra).
- `paired`: booleano (por defecto falso) que, cuando es verdadero, indica que ambas muestras están pareadas.
- `alternative`: tipo de prueba de hipótesis.
- `mu`: valor nulo.
- `conf.level`: nivel de confianza.

Script 5.3: inferencia con la media de las diferencias entre dos muestras pareadas usando la distribución t.

```
1  # Cargar los datos.
2  instancia <- seq(1, 35, 1)
3
4  t_A <- c(436.5736, 470.7937, 445.8354, 470.9810, 485.9394,
5           464.6145, 466.2139, 468.9065, 473.8778, 413.0639,
6           496.8705, 450.6578, 502.9759, 465.6358, 437.6397,
7           458.8806, 503.1435, 430.0524, 438.5959, 439.7409,
8           464.5916, 467.9926, 415.3252, 495.4094, 493.7082,
9           433.1082, 445.7433, 515.2049, 441.9420, 472.1396,
10          451.2234, 476.5149, 440.7918, 460.1070, 450.1008)
11
12 t_B <- c(408.5142, 450.1075, 490.2311, 513.6910, 467.6467,
13          484.1897, 465.9334, 502.6670, 444.9693, 456.3341,
14          501.1443, 471.7833, 441.1206, 544.1575, 447.8844,
15          432.4108, 477.1712, 482.4828, 458.2536, 474.9863,
16          496.0153, 485.8112, 457.4253, 483.3700, 510.7131,
17          467.5739, 482.5621, 453.5986, 385.9391, 548.7884,
18          467.2533, 494.7049, 451.9716, 522.3699, 444.1270)
19
20 diferencia <- t_A - t_B
21
22 # Verificar si la distribución se acerca a la normal.
23 normalidad <- shapiro.test(diferencia)
24 print(normalidad)
25
26 # Fijar un nivel de significación.
27 alfa <- 0.05
28
29 # Aplicar la prueba t de Student a la diferencia de medias.
30 prueba_1 <- t.test(diferencia,
31                    alternative = "two.sided",
32                    mu = valor_nulo,
33                    conf.level = 1 - alfa)
34
35 print(prueba_1)
36
37 # Otra alternativa puede ser aplicar la prueba t de Student
38 # para dos muestras pareadas.
39 prueba_2 <- t.test(x = t_A,
40                    y = t_B,
```

```
                    paired = TRUE,
                    alternative = "two.sided",
                    mu = valor_nulo,
                    conf.level = 1 - alfa)

print(prueba_2)
```

Los resultados para esta prueba son:

- El valor para el estadístico de prueba T es $t = -1,9816$.
- Se consideran $df = 34$ grados de libertad para la distribución t.
- El valor p obtenido es $p = 0,05565$.
- El intervalo de confianza obtenido es $[-24,4804542; 0,3086313]$.
- La media de la muestra es $\overline{x} = -12,08591$.

En este caso, la media de las diferencias está dentro del intervalo de confianza, y además el valor p es mayor que el nivel de significación, por lo que se falla al rechazar la hipótesis nula. Pero, nuevamente, el resultado está cerca del borde de significación. En consecuencia, se puede afirmar con 95 % de confianza que pareciera no haber diferencia entre los tiempos de ejecución de ambos algoritmos, aunque sería necesario conseguir una muestra más grande para tener mayor certeza.

5.2.3 Prueba t para dos muestras independientes

En este caso, la prueba t se usa para comparar las medias de dos poblaciones en que las observaciones con que se cuenta no tienen relación con ninguna de las otras observaciones, ni influyen en su selección, ni en la misma ni en la otra muestra. En este caso la inferencia se hace sobre la diferencia de las medias: $\mu_1 - \mu_2 = d_0$, donde d_0 es un valor hipotético fijo para la diferencia. Usualmente se usa $d_0 = 0$, en cuyo caso las muestras podrían provenir de dos poblaciones distintas con igual media, o desde la misma población. Para ello, la prueba usa como estimador puntual la diferencia de las medias muestrales $(\overline{x}_1 - \overline{x}_2)$. Así, el estadístico T en este caso toma la forma de la ecuación 5.3.

$$T = \frac{(\overline{x}_1 - \overline{x}_2) - d_0}{SE_{(\overline{x}_1 - \overline{x}_2)}} \tag{5.3}$$

Al usar la distribución t de Student para la diferencia de medias, se deben cumplir los siguientes requisitos:

1. Cada muestra cumple las condiciones para usar la distribución t.
2. Las muestras son independientes entre sí.

Veamos el funcionamiento de esta prueba con un ejemplo. El doctor E. L. Matta Sanno desea determinar si una nueva vacuna A es más efectiva que otra vacuna B, a fin de inmunizar a la población mundial contra una terrible enfermedad. Para ello, ha reclutado a un grupo de 28 voluntarios en diferentes países, 15 de los cuales (seleccionados al azar) recibieron la vacuna A y los 13 restantes, la vacuna B. La tabla 5.4 muestra, para cada voluntario, la concentración de anticuerpos (en microgramos por cada mililitro de sangre) al cabo de un mes de recibir la vacuna.

Las hipótesis a formular en este caso son:

H_0: no hay diferencia entre la efectividad promedio de ambas vacunas.
H_A: la vacuna A es, en promedio, más efectiva que la B.

En lenguaje matemático:

Anticuerpos [mg/ml]	
Vacuna A	Vacuna B
6,04	5,32
19,84	3,31
8,62	5,68
13,02	5,73
12,20	4,86
14,78	5,68
4,53	2,93
26,67	5,48
3,14	6,10
19,14	2,56
10,86	7,52
13,13	7,41
6,34	4,02
11,16	
7,62	

Tabla 5.4: Concentración de anticuerpos de los pacientes vacunados.

Si μ_A y μ_B son la concentraciones medias de anticuerpos presentes en personas luego de un mes de recibir la vacuna A y B, respectivamente, entonces:
H_0: $\mu_A = \mu_B$
H_A: $\mu_A > \mu_B$

Como es habitual, debemos ahora verificar el cumplimiento de las condiciones. Ambas muestras son independientes entre sí, pues son diferentes voluntarios y fueron designados aleatoriamente a cada grupo. Además, se puede asumir que las observaciones son independientes, pues cada muestra es significativamente menor a la población total a vacunar. En cuanto al supuesto de normalidad para cada muestra, al aplicar a cada una la prueba de Shapiro-Wilk (script 5.4, líneas 13 y 15) se obtiene, respectivamente, $p = 0,428$ y $p = 0,445$. En ambos casos el valor p es bastante alto, por lo que podemos concluir que ambas muestras provienen de poblaciones que se distribuyen de forma aproximadamente normal. Puesto que hemos verificado las condiciones, podemos llevar a cabo la prueba t para dos muestras independientes.

Ahora bien, como las muestras son algo pequeñas, sería prudente proceder con algo más de cautela. Además, en este escenario, un error tipo I (rechazar H_0 cuando es verdadera) implicaría reducir innecesariamente la cantidad de vacunas disponibles y retrasar el proceso de vacunación, poniendo en riesgo a todos los habitantes del planeta. Un error tipo II, en cambio, podría causar que se continúe el uso indistinto de ambas vacunas retrasando ligeramente el efecto inmune en la población. En consecuencia, el error tipo I es más grave, por lo que el nivel de significación debiese ser aún más exigente. En consecuencia, optaremos por $\alpha = 0,01$.

Al aplicar la prueba t (script 5.4), obtenemos que la diferencia entre las medias es $6,683$ [mg/ml] y que el intervalo de confianza es $[2,2739; \infty)$. Además, el valor p es $p < 0.001$, muy inferior al nivel de significación $\alpha = 0,01$. Esto significa que la evidencia en favor de H_A es muy fuerte, por lo rechazamos la hipótesis nula. En consecuencia, podemos concluir con 99 % de confianza que la vacuna A es, en promedio, mejor que la vacuna B (produce una mayor concentración media de anticuerpos en las personas vacunadas con ella que la producida por la vacuna B).

Script 5.4: prueba t para dos muestras independientes.

```
1  library(ggpubr)
2
3  # Cargar los datos.
4  vacuna_A <- c(6.04, 19.84, 8.62, 13.02, 12.20, 14.78, 4.53, 26.67,
5                3.14, 19.14, 10.86, 13.13, 6.34, 11.16, 7.62)
6
7  vacuna_B <- c(5.32, 3.31, 5.68, 5.73, 4.86, 5.68, 2.93, 5.48, 6.10,
```

```
              2.56, 7.52, 7.41, 4.02)

# Verificar si las muestras se distribuyen de manera cercana
# a la normal.
normalidad_A <- shapiro.test(vacuna_A)
print(normalidad_A)
normalidad_B <- shapiro.test(vacuna_B)
print(normalidad_B)

# Fijar un nivel de significación.
alfa <- 0.01

# Aplicar la prueba t para dos muestras independientes.
prueba <- t.test(x = vacuna_A,
                 y = vacuna_B,
                 paired = FALSE,
                 alternative = "greater",
                 mu = 0,
                 conf.level = 1 - alfa)

print(prueba)

# Calcular la diferencia entre las medias.
media_A <- mean(vacuna_A)
media_B <- mean(vacuna_B)
diferencia <- media_A - media_B
cat("Diferencia de las medias =", diferencia, "[mg/ml]\n")
```

Si estás leyendo atentamente, te habrás dado cuenta que ¡no hemos definido el error estándar para cuando tenemos dos muestras! En este caso, SE se construye a partir del error estándar de cada muestra, como se aprecia en la ecuación 5.4. En este escenario, la determinación de los grados de libertad es más compleja, por lo que se recomienda usar programas estadísticos o, en su defecto, escoger el menor valor entre $n_1 - 1$ y $n_2 - 1$.

$$SE_{\overline{x}_1 - \overline{x}_2} = \sqrt{SE_{\overline{x}_1}^2 + SE_{\overline{x}_2}^2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \tag{5.4}$$

Se puede lograr un mejor ajuste de la distribución t si se sabe con certeza que las desviaciones estándares de ambas poblaciones son casi iguales. En este caso, se puede usar una **varianza agrupada** (s_p^2, del inglés *pooled variance*) que reemplaza tanto a s_1^2 como a s_2^2 en la ecuación 5.4. Esta varianza agrupada se calcula como muestra la ecuación 5.5 y, en este caso, se consideran $n_1 + n_2 - 2$ grados de libertad.

$$s_p^2 = \frac{s_1^2 \cdot (n_1 - 1) + s_2^2 \cdot (n_2 - 1)}{n_1 + n_2 - 2} \tag{5.5}$$

Por defecto, R utiliza la corrección de Welch para la prueba t de Student de la diferencia de dos medias, variante considerada más segura, que en general entrega resultados muy similares a la versión original de la prueba cuando las muestras tienen varianzas similares. No obstante, los resultados son bastante mejores cuando los tamaños de las muestras y sus desviaciones estándares son muy diferentes (Kassambara, 2019). La corrección de Welch calcula el error estándar como muestra la ecuación 5.4, pero ajusta los grados de libertad de acuerdo a la ecuación 5.6.

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)}} \tag{5.6}$$

5.3 EJERCICIOS PROPUESTOS

1. Investiga acerca de la prueba de Kolmogorov-Smirnov y explica cómo puede usarse para verificar si una distribución se asemeja a la normal. Compara esta prueba con la de Shapiro-Wilk.
2. Para confirmar que el tiempo que requieren los estudiantes de ingeniería para desarrollar una guía de ejercicios de Cálculo I es de dos horas, se eligió aleatoriamente a 16 estudiantes de esta asignatura y se les pidió anotar el tiempo [min.] invertido en la tarea. Los resultados fueron los siguientes: 140,6; 133,3; 142,4; 86,4; 129,9; 110,8; 133,2; 129,1; 142,5; 150,2; 141,6; 111,0; 127,2; 137,9; 131,9; 121,9.
   a) Enuncia las hipótesis nula y alternativa a contrastar.
   b) Analiza si es razonable en este caso considerar que los datos cumplen las condiciones para usar una prueba t de Student.
   c) Independientemente del resultado anterior, aplica la prueba propuesta y obtenga un intervalo de confianza y un valor p.
   d) Usando un nivel de significación adecuado, entrega una conclusión para la cuestión planteada.
3. El departamento de control de calidad de un importante laboratorio requiere analizar la concentración de ingredientes activos presente en una muestra de 10 botellas diferentes de detergente líquido que ellos seleccionaron aleatoriamente en el último mes. Como se sospecha que esta concentración depende del catalizador que se use, la mitad del contenido de cada botella fue sometida a un catalizador, y la otra mitad a otro catalizador. En orden por botella seleccionada, los resultados fueron:
   - Catalizador 1: 62,9; 67,2; 67,4; 67,4; 67,2; 64,6; 69,6; 65,7; 68,2; 72,0.
   - Catalizador 2: 66,8; 69,3; 69,6; 67,3; 68,8; 68,4; 68,6; 70,3; 69,6; 71,7.

   a) Como primer paso, el departamento de control de calidad necesita saber si la concentración media de concentraciones de ingredientes activos depende del catalizador elegido.
   b) Propón las hipótesis nula y alternativa que permitan responder el problema planteado con una prueba t de Student.
   c) Muestra que es razonable considerar que estos datos cumplen las condiciones para usar la prueba propuesta y fija un nivel de significación apropiado.
   d) Aplica la prueba propuesta y obtenga un intervalo de confianza y un valor p.
   e) ¿Cuál sería tu respuesta al departamento de control mencionado?
4. Una fábrica de detectores de radón recibió consultas de sus clientes sobre si era conveniente comprar su nuevo modelo de detectores Radolmes+®para reemplazar los antiguos aparatos Radolmes®en su poder. Si bien los técnicos están seguros que la inversión es conveniente, la gerencia decidió hacer un estudio previo a la recomendación. Para esto, se introdujeron en una tómbola oscura los números de serie de los aparatos producidos en los últimos meses de ambos modelos y se seleccionaron 26 números sin mirar y girando la tómbola cinco veces entre cada selección, resultando escogidos 12 aparatos Radolmes y 14 aparatos Radolmes+. Luego, cada detector seleccionado se expuso a 100 pCi/l de radón. Las lecturas resultantes fueron las siguientes:
   - Radolmes: 105,6; 100,1; 90,9; 105,0; 91,2; 99,6; 96,9; 107,7; 96,5; 103,3; 91,3; 92,4.
   - Radolmes+: 98,9; 94,3; 95,9; 107,7; 102,0; 94,2; 100,6; 98,5; 99,1; 101,3; 94,4; 103,6; 95,3; 106,7.

   a) ¿Qué hipótesis nula y alternativa se deberían docimar[2] con una prueba t de Student para responder a la inquietud planteada?
   b) ¿Cumplen los datos obtenidos las condiciones para usar esta prueba t de Student?
   c) Aplicando la prueba t de Student para este caso, obtén un intervalo de confianza y un valor p.
   d) ¿Qué aconsejarías a los directivos de la fábrica?

[2]Término que suele ocuparse en estadística como sinónimo de "probar".

REFERENCIAS

Diez, D., Barr, C. D. & Çetinkaya-Rundel, M. (2017). *OpenIntro Statistics* (3.ª ed.). https://www.openintro.org/book/os/.

Field, A., Miles, J. & Field, Z. (2012). *Discovering statistics using R*. SAGE Publications Ltd.

Kassambara, A. (2019). *Practical Statistics in R II - Comparing Goups: Numerical Variables*. Datanovia.

Meena, S. (2020). *Statistics for Analytics and Data Science: Hypothesis Testing and Z-Test vs. T-Test*. Consultado el 22 de septiembre de 2021, desde https://www.analyticsvidhya.com/blog/2020/06/statistics-analytics-hypothesis-testing-z-test-t-test/#h2_1

Parada, L. F. (2019). *Prueba de normalidad de Shapiro-Wilk*. Consultado el 22 de septiembre de 2021, desde https://rpubs.com/F3rnando/507482

Real Academia Española. (2014). *Diccionario de la lengua española* (23.ª ed.). Consultado el 30 de marzo de 2021, desde https://dle.rae.es