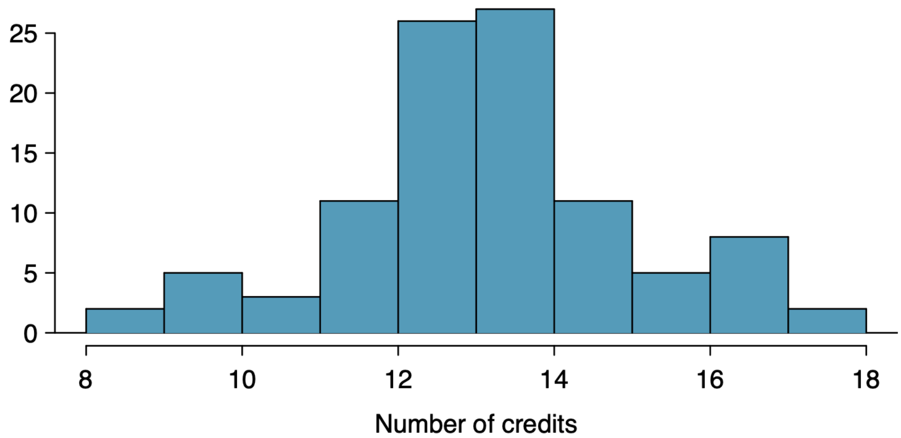
Unidad 4 - Actividad 5. Fundamentos para la inferencia.
Alumno: **Luis Fernando Izquierdo Berdugo**
Materia: **Estadística**
Fecha de Entrega: **26 de Septiembre de 2024**

Realiza los siguientes ejercicios del capítulo 4 Fundamentos para la inferencia.

Página	Ejercicios
204	4.3
207	4.11
211	4.24
217	4.41

4.3 College credits.
A college counselor is interested in estimating how many credits a student typically enrolls in each semester. The counselor decides to randomly sample 100 students by using the registrar's database of students. The histogram below shows the distribution of the number of credits taken by these students. Sample statistics for this distribution are also provided.



Min	8
Q1	13
Median	14
Mean	13.65
SD	1.91
Q3	15
Max	18

(a) What is the point estimate for the average number of credits taken per semester by students at this college? What about the median?

Estos datos de estimación puntual se pueden observar de manera sencilla en las muestras estadísticas que se proporcionan, siendo 13.65 para el promedio y 14 para la mediana.

(b) What is the point estimate for the standard deviation of the number of credits taken per semester by students at this college? What about the IQR?

La estimación puntual para la desviación estándar es de 1.91, de acuerdo con la información de la tabla.
El rango intercuartil (IQR) se puede calcular de la siguiente manera:

$$Q3 - Q1 = IQR$$

$$15 - 13 = 2$$

(c) Is a load of 16 credits unusually high for this college? What about 18 credits? Explain your reasoning. Hint: Observations farther than two standard deviations from the mean are usually considered to be unusual.

Para obtener la cantidad de desviaciones estándares de las que un valor se aleja del promedio, se utiliza la fórmula:

$$z = (x - \mu) / \sigma$$

donde:
z = desviaciones estándar alejadas de la media
x = dato a evaluar
μ = promedio
σ = desviación estándar

Entonces, con esta información se puede calcular z para 16 y 18

$$z = (16 - 13.65) / 1.91 = 1.2303665$$
$$z = (18 - 13.65) / 1.91 = 2.2774869$$

Teniendo los resultados podemos decir que 16 un carga de 16 créditos no es algo inusual en la escuela ya que no está alejado más de 2 desviaciones estándar de la media, sin embargo, una carga de 18 si sería algo inusual, ya que está alejado más de 2 desviaciones estándar.

(d) The college counselor takes another random sample of 100 students and this time finds a sample mean of 14.02 units. Should she be surprised that this sample statistic is slightly different than the one from the original sample? Explain your reasoning.

El consejero no debería estar sorprendido, ya que las estimaciones puntuales devuelven información acerca de la muestra analizada y no de la población completa, así que se esperan variaciones dependiendo de la muestra tomada, incluso si esta es dentro de la misma población de la muestra que se analizó.

(e) The sample means given above are point estimates for the mean number of credits taken by all students at that college. What measures do we use to quantify the variability of this estimate? Compute this quantity using the data from the original sample.

Usando la fórmula de error estándar para el promedio, siendo n el número de estudiantes:

$$SD_X = \frac{\sigma}{\sqrt{n}}$$
$$SD_X = \frac{1.91}{\sqrt{100}} = 0.191$$

Entonces se puede esperar una variación de 0.191 créditos en el promedio de la muestra.

4.11 Relaxing after work.

The 2010 General Social Survey asked the question: “After an average work day, about how many hours do you have to relax or pursue activities that you enjoy?” to a *random sample of 1,155 Americans. A 95% confidence interval for the mean number of hours spent relaxing or pursuing activities they enjoy was (1.38, 1.92).*

(a) Interpret this interval in context of the data.

El promedio poblacional de horas para relajarse o actividades que se disfrutan es de entre 1.38 y 1.92 hrs, esto con un 95% de seguridad.

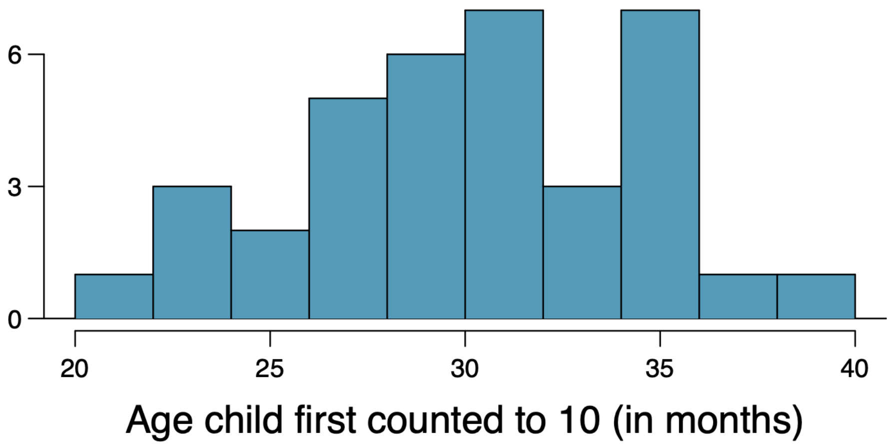
(b) Suppose another set of researchers reported a confidence interval with a larger margin of error based on the same sample of 1,155 Americans. How does their confidence level compare to the confidence level of the interval stated above?

Un margen de error más grande para la misma muestra indicaría que se utilizó un nivel de seguridad más bajo que el del 95%

(c) Suppose next year a new survey asking the same question is conducted, and this time the sample size is 2,500. Assuming that the population characteristics, with respect to how much time people spend relaxing after work, have not changed much within a year. How will the margin of error of the 95% confidence interval constructed based on data from the new survey compare to the margin of error of the interval stated above?

Con una muestra de tamaño mayor, se espera que el margen de error sea menor. La fórmula del margen de error involucra al error estándar, que es inversamente proporcional a la raíz cuadrada del tamaño de la mezcla, entonces mientras el tamaño de la muestra aumenta, el margen de error disminuye.

4.24 Gifted children, Part I.
Researchers investigating characteristics of gifted children collected data from schools in a large city on a random sample of thirty-six children who were identified as gifted children soon after they reached the age of four. The following histogram shows the distribution of the ages (in months) at which these children first counted to 10 successfully. Also provided are some sample statistics.



n	36
min	21
mean	30.69
sd	4.31
max	39

(a) Are conditions for inference satisfied?
Si, para la inferencia se cumplen las condiciones:

- Los datos provienen de una muestra aleatoria de niños prodigio
- Todas las observaciones (niños) son independientes

- El tamaño de la muestra es mayor a 30

(b) Suppose you read online that children first count to 10 successfully when they are 32 months old, on average. Perform a hypothesis test to evaluate if these data provide convincing evidence that the average age at which gifted children first count to 10 successfully is less than the general average of 32 months. Use a significance level of 0.10.

La hipótesis nula (Ho) es que el promedio de edad de los niños para contar hasta 10 es igual a 32 meses, mientras que la hipótesis alternativa (Ha) es que el promedio de edad de los niños para contar hasta 10 es menor a 32 meses.

Usando la función de distribución acumulada en R (pnorm) se puede calcular la probabilidad de que la edad sea 32 o mayor.

```
prom <- 30.69
sd <- 4.31
se <- sd/sqrt(36)
p <- pnorm(32, mean = prom, sd = SE)
1-p

## [1] 0.0341013
```

Esto significa que para un nivel de significancia de 0.10, siendo la probabilidad 0.0341013, se rechaza la hipótesis nula, y la hipótesis alternativa es posible.

(c) Interpret the p-value in context of the hypothesis test and the data.

El valor p de 0.0341013 indica que si el promedio de la población realmente es 32, hay un 3.41% de observar un promedio muestral como el que se obtuvo (30.69), esto indica que hay suficientes pruebas para decir que, en promedio, los niños prodigio cuentan hasta 10 antes de los 32 meses.

(d) Calculate a 90% confidence interval for the average age at which gifted children first count to 10 successfully.

Con la función qnorm de R se puede calcular el valor de una distribución t, con lo cual podemos calcular el intervalo de confianza de 90%.

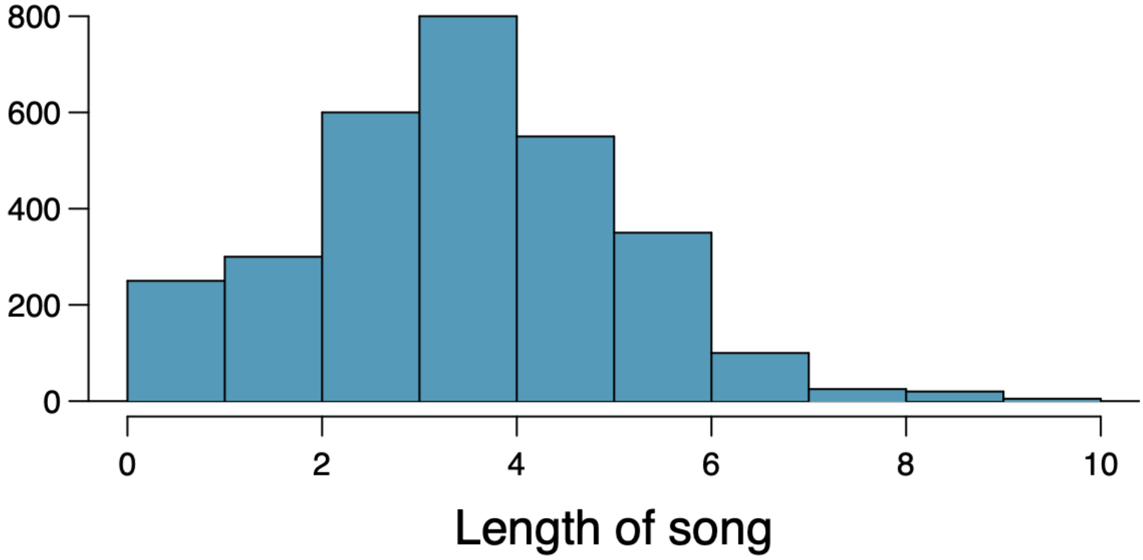
```
z <- qnorm(0.95)
mean + z*se

## [1] 31.87155

mean - z*se

## [1] 29.50845
```

Siendo el nuevo intervalo de entre 29.50845 y 31.87155.

(e) Do your results from the hypothesis test and the confidence interval agree? Explain

Si son congruentes, ya que el nuevo intervalo no incluye los 32 meses con seguridad de 90%, en palabras sencillas, esto indica que se está 90% seguro de que el promedio real de la población es menor a 32, rechazando la hipótesis nula.

4.41 Songs on an iPod.
Suppose an iPod has 3,000 songs. The histogram below shows the distribution of the lengths of these songs. We also know that, for this iPod, the mean length is 3.45 minutes and the standard deviation is 1.63 minutes.



(a) Calculate the probability that a randomly selected song lasts more than 5 minutes.

Usando R para desarrollar esta probabilidad, se usaría la función pnorm.

```
mean <- 3.45
sd <- 1.63
z <- (5-mean)/sd
p <- 1-pnorm(z)
p
## [1] 0.1708224
```

La probabilidad de que una canción dure más de 5 minutos es de 17.08224%

(b) You are about to go for an hour run and you make a random playlist of 15 songs. What is the probability that your playlist lasts for the entire duration of your run? Hint: If you want the playlist to last 60 minutes, what should be the minimum average length of a song?

Si se quiere que la lista de 15 canciones dure 60 minutos, se necesita que las canciones duren 4 minutos o más, entonces se calculará la probabilidad de que una canción tenga una duración mayor a 4 minutos.

```
p_4 <- 1 - pnorm((4 - mean) / sd)
p_4
## [1] 0.3678989
```

Con esto se puede sacar la probabilidad de que la lista de 15 canciones aleatorias dure 60 minutos.

```
p_60 <- p_4^15
p_60
## [1] 3.061452e-07
```

Se puede observar que la probabilidad de que una lista de reproducción de 15 canciones dure 60 minutos exactos es de 0.00003061452%, haciéndolo casi imposible.

(c) You are about to take a trip to visit your parents and the drive is 6 hours. You make a random playlist of 100 songs. What is the probability that your playlist lasts the entire drive?

Primero se obtendrá la probabilidad de que una canción dure 3.6 minutos.

```
p_36 <- 1 - pnorm((3.6 - mean) / sd)
p_36
## [1] 0.4633393
```

Lo siguiente será obtener la probabilidad de que las 100 canciones duren 3.6 minutos.

```
p_100 <- p_36^100
p_100
## [1] 3.889655e-34
```

Solamente viendo el número expresado en notación científica, se puede observar que la probabilidad de que 100 canciones aleatorias duren exactamente 360 minutos es casi imposible. La probabilidad sería de 0.00000000000000000000000000000003889655%.