Laboratorio 2

Respuestas

Instrucciones

El laboratorio 2 deberá ser contestado en equipos de 2 o 3 personas (max.) y el archivo con las respuestas lo debes enviar por email a mi correo electrónico (`sergio.bejar@cide.edu`) a más tardar el viernes 15 de marzo, 2023 a las 12:00pm.

Ejercicio 1

Asumamos que la elección presidencial del próximo 2 de junio está empatada antes de que emitas tu voto. Además, supongamos que el 50% del electorado mexicano apoya a la candidata C, mientras que el otro 50% apoya a la candidata X. La probabilidad de votar (i.e. votar o no votar) es independiente de la preferencia electoral de cada persona. Dado este escenario,

a) ¿cuál es la probabilidad de que la elección termine en un empate exacto entre las dos candidatas si el número de votantes es 1,000, 10,000, y 100,000? (15 pts.) Hint: usa `dbinom` para simplificar el trabajo.

RESPUESTA:

```r
votantes <- c(1000, 10000, 100000)  ## definimos el # de votantes
dbinom(votantes / 2, size = votantes, prob = 0.5)  ## aplicamos fórmula de `dbinom`
```

```
[1] 0.025225018 0.007978646 0.002523126
```

b) ¿qué nos dicen estos resultados sobre la importancia de tu voto en una elección tan cerrada? (5 pts.)

RESPUESTA:

Hay dos cosas importantes/interesantes que podemos inferir de los resultado. Primero, la probabilidad de que una elección cerrada se decida por un voto es bastante pequeña, independientemente del número de votantes. Segundo, el resultado también nos indica que la probabilidad de que nuestro voto decida la elección disminuye a medida que se incrementa el número de votantes (De .02 cuando n = 1000 a .0025 cuando n = 100,000).

Ejercicio 2

Usa la función `rnorm()` para generar 1,000 valores aleatorios con $\mu = 100$ y $\sigma = 20$, y asigna los valores estimados a un objecto llamado `aleatorio`. A continuación, realiza los siguientes pasos y contesta a las preguntas.

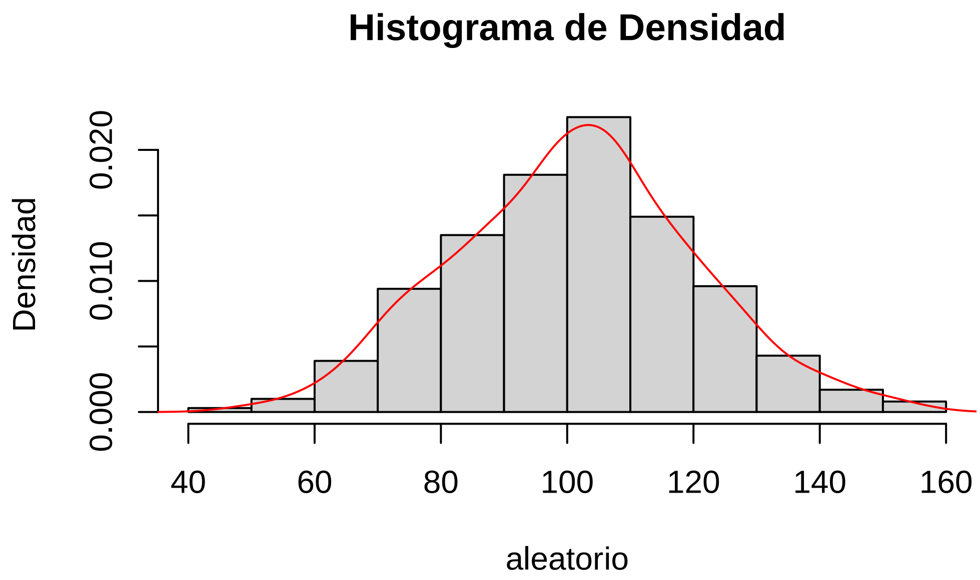
1. Presenta un histograma de densidad obtenido con los valores del objeto `aleatorio` y contesta ¿a qué distribución se asemeja el gráfico obtenido? (15 pts.)

RESPUESTA:

```
aleatorio <- rnorm(1000, mean = 100, sd = 20)  ## creamos objeto llamado `aleatorio`

## el gráfico lo podemos hacer al menos de dos formas diferentes:
## (i) con `hist` en `R` base,
## (ii) con ggplot2.
## Voy a hacerlo con `hist`.

hist(aleatorio,
     prob = TRUE,
     main = "Histograma de Densidad",
     ylab = "Densidad")
lines(density(aleatorio), col = "red")  ## dibuja una curva normal sobre histograma
```



1. ¿Es sorpresiva la forma del histograma? ¿Por qué? (5 pts.)

RESPUESTA:

Nada sorpresivo porque simulamos una distribución normal.

Ejercicio 3

Para este ejercicio tienes que descargar el archivo `lab2.csv` que se encuentra en la carpeta Files del repositorio de nuestra clase en Github. La variable que vamos a utilizar es *age*.

```
## cargamos los datos

lab2 <- rio::import("https://raw.githubusercontent.com/Sergio-Bejar/MCA_CIDE/main/Files/la
```

1. Calcula la media de *age*. (10 pts.)

RESPUESTA:

```
media_age <- mean(lab2$age) ## calcula media de variable de interés

media_age  ## muestra resultado
```

```
[1] 60.36749
```

2. Calcula el error estándar de *age*

RESPUESTA:

```
d_e <- sd(lab2$age)  ## calcula d.e. de variable de interés
e_e <- d_e/sqrt(length(lab2$age))  ## calcula error estándar


e_e ## muestra resultado
```

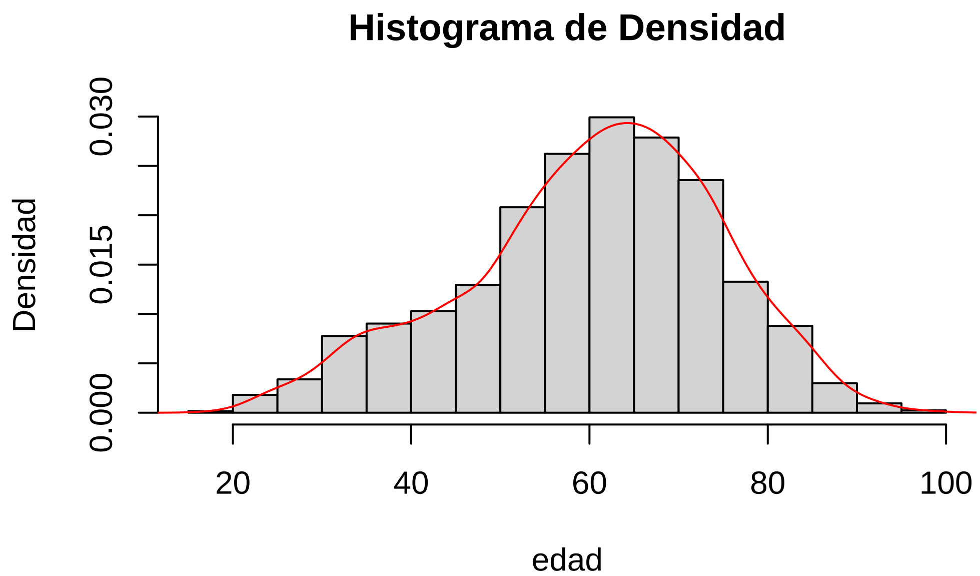
```
[1] 0.2815446
```

3. Calcula el intervalo de confianza para la variable *age* asumiendo un 95% de confianza. (10 pts.)

RESPUESTA:

Vamos primero a gtraficar un histograma de la variable `age` para ver su distribución.

```
hist(lab2$age,
     prob = TRUE,
     main = "Histograma de Densidad",
     xlab = "edad",
     ylab = "Densidad")
lines(density(lab2$age), col = "red")  ## dibuja una curva normal sobre histograma
```



Podemos ver en el gráfico que no tenemos una distribución normal perfecta. Los datos están sesgados hacia la derecha. Aunque nuestras conclusiones serán las mismas -porque n es suficientemente grande- lo mas adecuado es asumir que tenemos una distribución *t-student*. También hay que notar que no tenemos información suficiente sobre los parámetros poblacionales (i.e. de nuestros datos solo podemos calcular parámetros muestrales).

Y hay dos formas para calcular el intervalo de confianza.

a) Podemos usar la siguiente fórmula:

```
int_conf <- c(media_age - (1.96 * e_e), media_age + (1.96 * e_e)) ## ver notas para fórmul
int_conf  ## imprime resultado
```

```
[1] 59.81566 60.91932
```

b) Nos simplificamos la vida usando `t.test`

```
t.test(lab2$age)
```

```

	One Sample t-test
```

```
data:  lab2$age
t = 214.42, df = 2546, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 59.81541 60.91957
sample estimates:
mean of x
 60.36749
```

Podemos ver que los intervalos de confianza en (a) y (b) son idénticos.

4. Interpreta el resultado del inciso anterior (i.e. ¿qué quiere decir?). (5 pts.)

RESPUESTA:

El resultado significa que μ se encuentra entre 59.8 y 60.9 años con un 95% de probabilidad.

Ejercicio 4

Las hipótesis nulas y alternativa para determinar si la media muestral es igual a la media poblacional obtenidas en el ejercicio 3 son iguales (o no) son las siguientes.

- $H_0 : \bar{x} = \mu$
- $H_1 : \bar{x} \neq \mu$

1. ¿Cuál es el p-value cuando $\bar{x} = 60$? (15 pts.)

RESPUESTA:

```r
t.test(lab2$age, mu = 60)
```

```

    One Sample t-test

data:  lab2$age
t = 1.3053, df = 2546, p-value = 0.1919
alternative hypothesis: true mean is not equal to 60
95 percent confidence interval:
 59.81541 60.91957
sample estimates:
mean of x
 60.36749
```

$p = 0.19$

2. El resultado anterior, ¿confirma la hipótesis nula o la hipótesis alternativa? ¿por qué) (10 pts.)

RESPUESTA:

En base al resultado anterior, $p > .05$, aceptamos la hipótesis nula.