Tarea_02

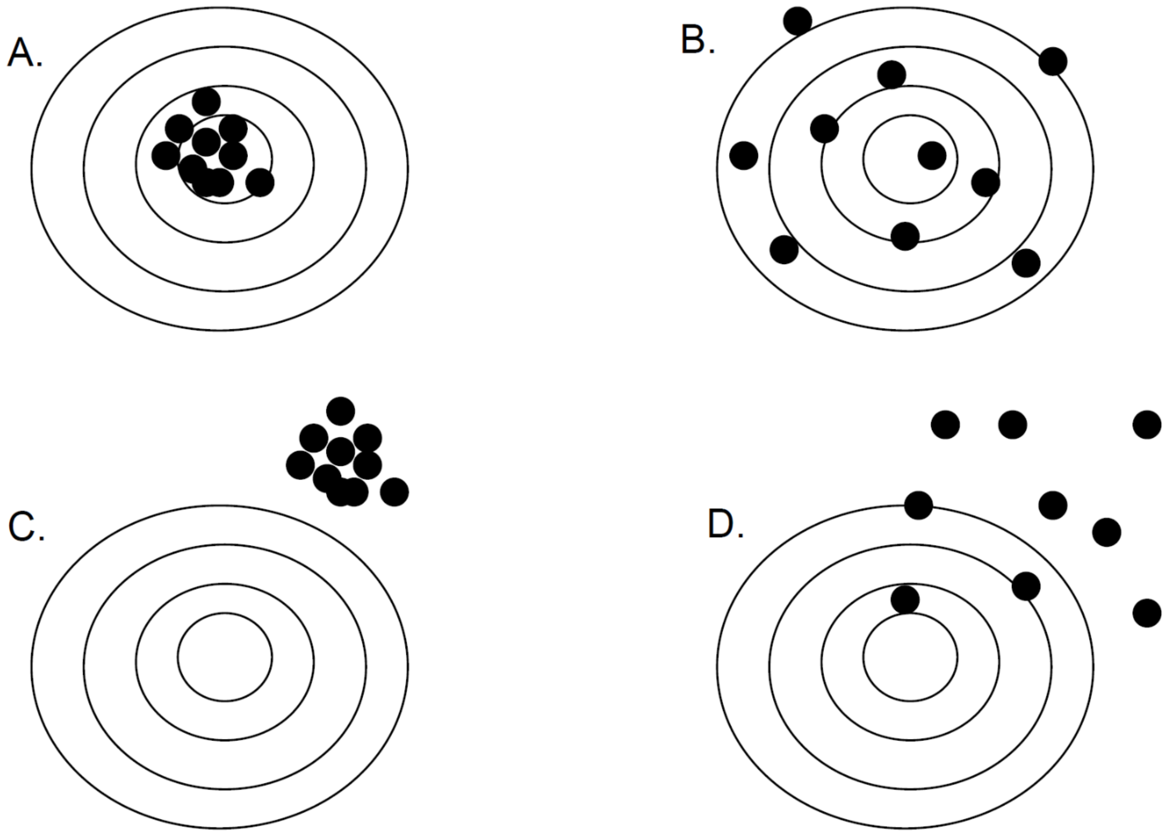
Nota: 10

```r
library(tidyverse)
library(tidyverse)
library(rio)
library(collapse)
library(DescTools)
library(emmeans)
library(modelbased)
library(see)
library(parameters)
library(knitr)
```

####Nombre del estudiante: Ana Carlota Reyes Ferrufino

####1. Identifica los conceptos de precisión, sesgo, y exactitud en los diagramas siguientes. Una pista, el diagrama B es no sesgado pero impreciso = inexacto.

```r
include_graphics("figura.png")
```



A: no sesgado y con alta precisión = exacto

B: no sesgado pero impreciso = inexacto

C: preciso pero con sesgo alto = inexacto

D: sesgo alto e impreciso = inexacto

####2. Un investigador esta investigando algunas características morfométricas, tales como el peso corporal (g) y la longitud del pico (mm), de 2 especies de de pinzones africanos.

Datos: PinzonesAfricanos.csv

#####¿Cuál de las especies tiene el pico más largo?

a. Analiza el resultado desde el punto de vista del valor de P.

```r
pinzones <- import("PinzonesAfricanos.csv")
```

```r
head(pinzones)
```

```
##    Especie Peso LargoDePico
## 1 WB.SPARW   40        10.6
## 2 WB.SPARW   43        10.8
## 3 WB.SPARW   37        10.9
```

```
## 4 WB.SPARW   38        11.3
## 5 WB.SPARW   43        10.9
## 6 WB.SPARW   33        10.1
```

```
t.test(LargoDePico ~ Especie, data= pinzones, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  LargoDePico by Especie
## t = -22.4, df = 28, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.666094 -3.051763
## sample estimates:
## mean in group CRU.WAXB mean in group WB.SPARW
##               7.378571              10.737500
```

```
medias <- pinzones %>%
  group_by(Especie) %>%
  summarize (media= mean(LargoDePico))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
medias
```

```
## # A tibble: 2 x 2
##   Especie  media
##   <chr>    <dbl>
## 1 CRU.WAXB  7.38
## 2 WB.SPARW 10.7
```

Respuesta: Como el estadistico t supera el valor de 1.96, se rechaza la hipotesis nula. Esto indica que si existe una diferencia significativa entre el largo de los picos de las especies de pinzones, porque el valor de P es mucho menor que 0.05. Por lo tanto existe evidencia estadística que se rechaza la hipotesis nula, y que los largos de los picos entre especies son diferentes. Con el calculo de las medias del largo de pico para cada especie, mostrado también en la prueba de t-test, se puede concluir que la especie WB.SPARW tiene un pico mas largo que la espcie CRU.WAXB.

b. Analiza el resultado desde el punto de vista del tamaño del efecto.

```
MeanDiffCI(LargoDePico ~ Especie, data= pinzones)
```

```
##  meandiff    lwr.ci    upr.ci
## -3.358929 -3.667242 -3.050615
```

WB.SPARW - CRU.WAXB = 3.36 (IC= 3.05, 3.68)

Respuesta: En la investigación se encontró que la diferencia entre longitud de picos es de 3.36. Con el tamaño del efecto se puede concluir que con un 95% de confianza, la verdadera diferencia entre el largo de los picos entre especies se encuentra entre 3.05 y 3.68.

####3. Muchas personas creen que para logar un estimado preciso de la media poblacional es necesario muestrear una fracción sustancial de la población. Esta pregunta esta desarrollada para probar si tal aseveración es cierta o no.

a. Para una población con desvío estándar 50, encuentra el error estándar de los siguientes valores de N (tamaño de población) y n (tamaño de muestra). Coloca los errores estándares calculados en las celdas vacías de la tabla.

```
##
## Attaching package: 'data.table'

## The following object is masked from 'package:DescTools':
##
##     %like%

## The following objects are masked from 'package:dplyr':
##
##     between, first, last

## The following object is masked from 'package:purrr':
##
##     transpose
```

N =	100	1000	10000	1000000	10000000
n =	10	100	1000	10000	1000000
Error estándar =					

```
EE1 <- (50/sqrt(10))
EE1
```

```
## [1] 15.81139
```

```
EE2 <- (50/sqrt(100))
EE2
```

```
## [1] 5
```

```
EE3 <- (50/sqrt(1000))
EE3
```

```
## [1] 1.581139
```

```
EE4 <- (50/sqrt(10000))
EE4
```

```
## [1] 0.5
```

```r
EE5 <- (50/sqrt(1000000))
EE5
```

```
## [1] 0.05
```

N =	100	1000	10000	1000000	10000000
n =	10	100	1000	10000	1000000
Error estándar =	15.81	5.00	1.58	0.50	0.05

b. Se te ocurre alguna explicación de los resultados obtenidos.

Según el calculo de los errores estandar se puede concluir que este sera menor al aumentar considerablemente el tamaño de la muestra, y al tener la misma desviación estandar. En este caso, sucede porque la desviación estandar de la muestra es igual en todos los casos, y no es igual tener una distribución sd=50 en una muestra de n=10, a tener esa distribución sd= 50 en n=1000000. La desviación estandar de 50 es un valor pequeño para las muestras grandes, por lo tanto el error estandar es mucho menor, porque los datos estan menos dispersos de la media.