

Tarea_02

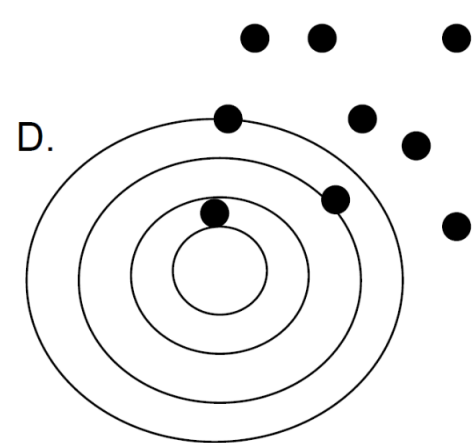
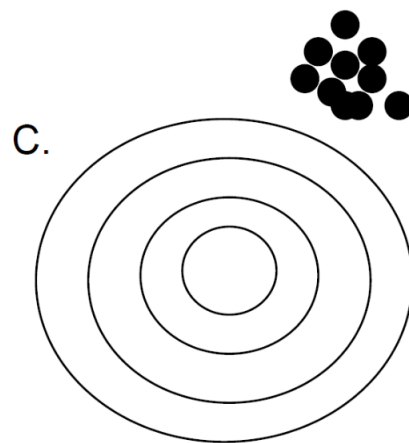
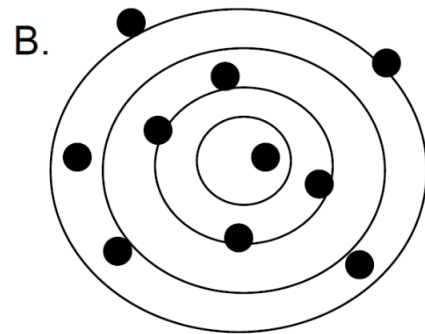
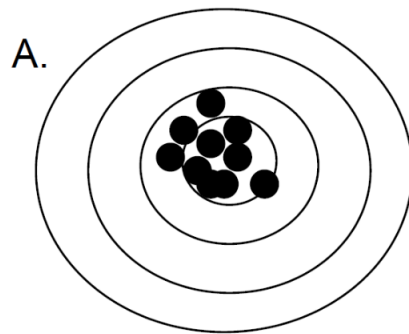
Nota: 9,0

```
library(tidyverse)
library(rio)
library(knitr)
library(DescTools)
library(data.table)
library(modelbased)
library(emmeans)
```

Nombre del estudiante: Marvin Jaroht Solano Lopez

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

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



A: No sesgada y precisa

B: no sesgado pero impreciso = inexacto

C: Precisa pero sesgada

D: Sesgado e impreciso

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?

- Analiza el resultado desde el punto de vista del valor de P.
- Analiza el resultado desde el punto de vista del tamaño del efecto.

```
Datospinzones <- import("PinzonesAfricanos.csv")
```

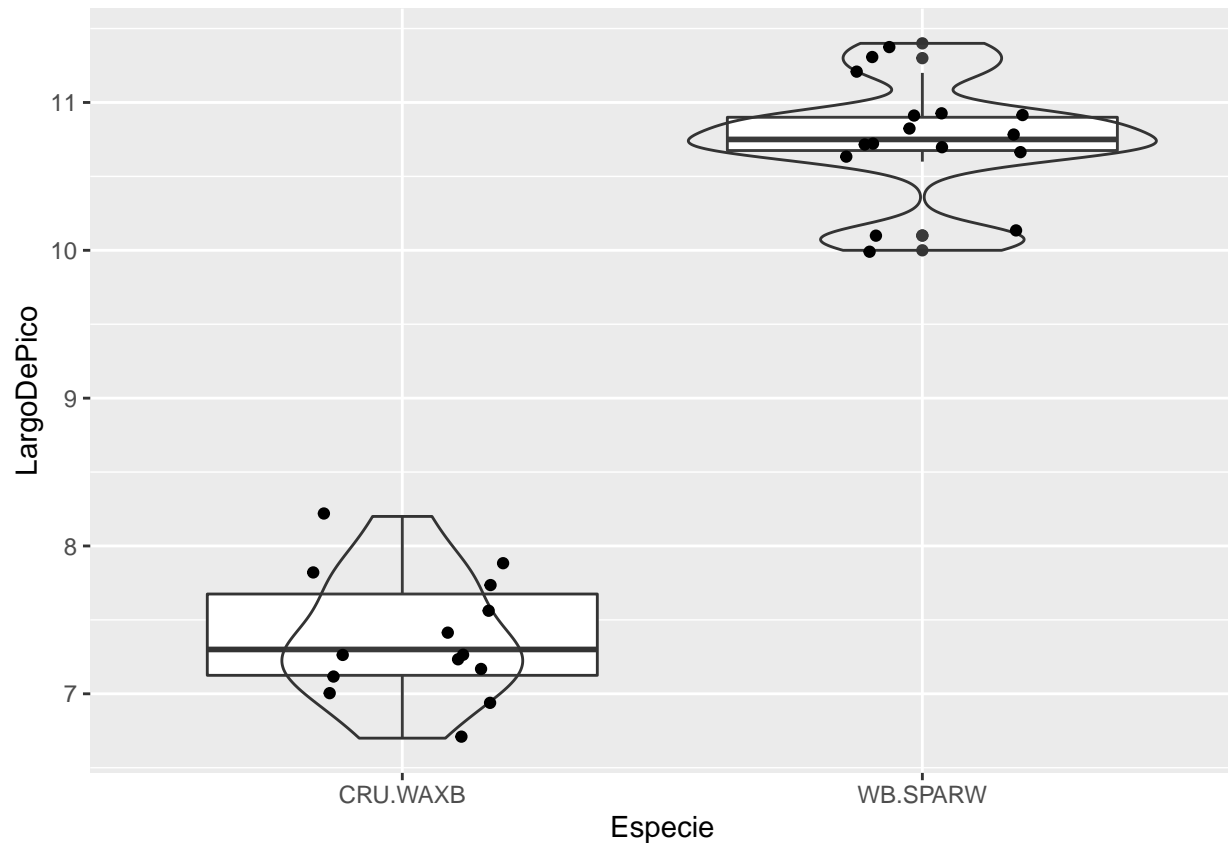
```
Datospinzones
```

```
##      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
## 7	WB.SPARW	35	10.7
## 8	WB.SPARW	37	10.7
## 9	WB.SPARW	36	10.9
## 10	WB.SPARW	42	11.4
## 11	WB.SPARW	36	10.8
## 12	WB.SPARW	36	11.2
## 13	WB.SPARW	39	10.7
## 14	WB.SPARW	37	10.0
## 15	WB.SPARW	34	10.1
## 16	WB.SPARW	41	10.7
## 17	CRU.WAXB	8	8.2
## 18	CRU.WAXB	8	7.2
## 19	CRU.WAXB	8	7.3
## 20	CRU.WAXB	8	7.2
## 21	CRU.WAXB	8	7.7
## 22	CRU.WAXB	8	7.4
## 23	CRU.WAXB	8	7.6
## 24	CRU.WAXB	6	6.9
## 25	CRU.WAXB	7	7.9
## 26	CRU.WAXB	7	6.7
## 27	CRU.WAXB	7	7.0
## 28	CRU.WAXB	8	7.3
## 29	CRU.WAXB	8	7.8
## 30	CRU.WAXB	8	7.1

```
ggplot(Datospinzones, aes(Especie, LargoDePico))+
  geom_boxplot()+
  geom_violin(alpha= 0.04) +
  geom_jitter(width = 0.2) +
  stat_summary(fun.data= "mean_cl_normal", colour = "red", size = 1)
```

```
## Warning: Computation failed in 'stat_summary()':
## object 'mean_cl_normal' of mode 'function' was not found
```



```
t.test(LargoDePico ~ Especie, data= Datospinzones, 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
```

Alfa = 0.05

p < 0.05

Pero como se interpreta?

```
medias <- Datospinzones %>%
  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.SPARG 10.7
```

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

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

La diferencia entre las especies CRU.WAXB Y WB.SPARG con respecto a la variable largo del pico es de 3.35 centímetros. Esta diferencia se encuentra entre 3.05 y 3.66 centímetros.

Así no se interpreta un intervalo de confianza

####3. Muchas personas creen que para lograr un estimado preciso de la media poblacional es necesario muestrear una fracción sustancial de la población. Esta pregunta está 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.

N =	100	1000	10000	1000000	10000000
n =	10	100	1000	10000	1000000
Error estándar =	5	1.58113883008419	0.5	0.05	0.05

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

Es interesante notar que al aumentar el número de muestras usando una misma desviación estándar para un tamaño de población mayor, el error se ve disminuido, considero que si la desviación estándar aumentara el error también aumentaría, a pesar de estar usando un 10% de muestras para las distintas poblaciones.