

Tarea_02

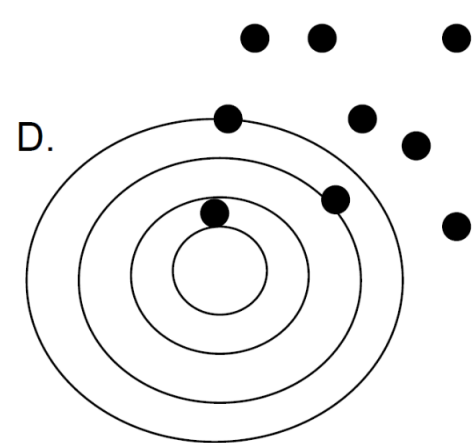
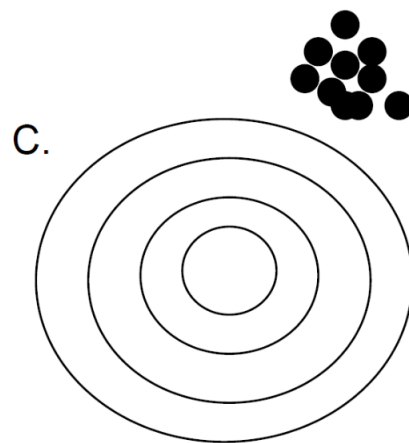
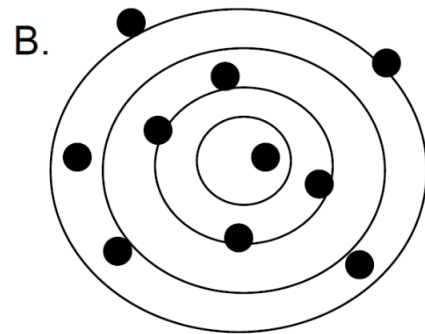
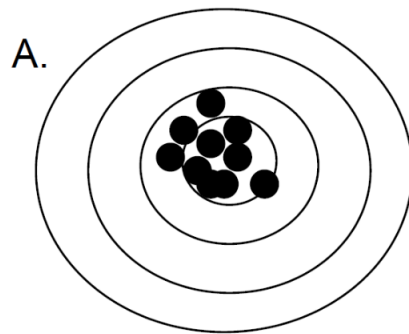
Nota: 9,5

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

Nombre del estudiante: Carlos Rafael Arita Portillo

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 sesgado pero preciso = exacto

B: no sesgado pero impreciso = inexacto

C: sesgado pero preciso = inexacto

D: sesgado pero 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

```
pinzones_africanos = import("PinzonesAfricanos.csv")
```

```
head(pinzones_africanos)
```

```
##      Especie Peso LargoDePico
## 1 WB.SPARG  40      10.6
## 2 WB.SPARG  43      10.8
## 3 WB.SPARG  37      10.9
## 4 WB.SPARG  38      11.3
## 5 WB.SPARG  43      10.9
## 6 WB.SPARG  33      10.1
```

¿Cuál de las especies tiene el pico más largo?

```
pico_largo <- pinzones_africanos %>%  
  group_by(Especie) %>%  
  summarize(media = mean(LargoDePico))
```

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

```
pico_largo
```

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

R//La especie WB.SPARW tiene en promedio el pico más largo

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

```
analisis_P = lm(LargoDePico ~ Especie, data = pinzones_africanos)
```

```
summary(analisis_P)
```

```
##  
## Call:  
## lm(formula = LargoDePico ~ Especie, data = pinzones_africanos)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -0.7375 -0.1786 -0.0375  0.2067  0.8214   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)      7.3786     0.1095   67.38  <2e-16 ***  
## EspecieWB.SPARW  3.3589     0.1500   22.40  <2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.4098 on 28 degrees of freedom  
## Multiple R-squared:  0.9471, Adjusted R-squared:  0.9453   
## F-statistic: 501.8 on 1 and 28 DF,  p-value: < 2.2e-16
```

```
contraste_p <- estimate_contrasts(analisis_P, "Especie")  
contraste_p
```

```
## Level1 | Level2 | Difference | SE |          95% CI |      t | df |      p | Difference (std.)  
## -----  
## CRU.WAXB | WB.SPARW |      -3.36 | 0.15 | [-3.67, -3.05] | -22.40 | 28 | < .001 |      -1.92
```

R//El analisis del valor de P se resume en la tabla anterior

Pero como se interpreta?

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

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

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

```
estad<- emmeans(analisis_P, "Especie")
```

```
estad
```

```
## Especie emmean    SE df lower.CL upper.CL  
## CRU.WAXB  7.38 0.110 28    7.15    7.6  
## WB.SPARG 10.74 0.102 28   10.53   10.9  
##  
## Confidence level used: 0.95
```

```
tamaño_efecto<-pairs(estad)
```

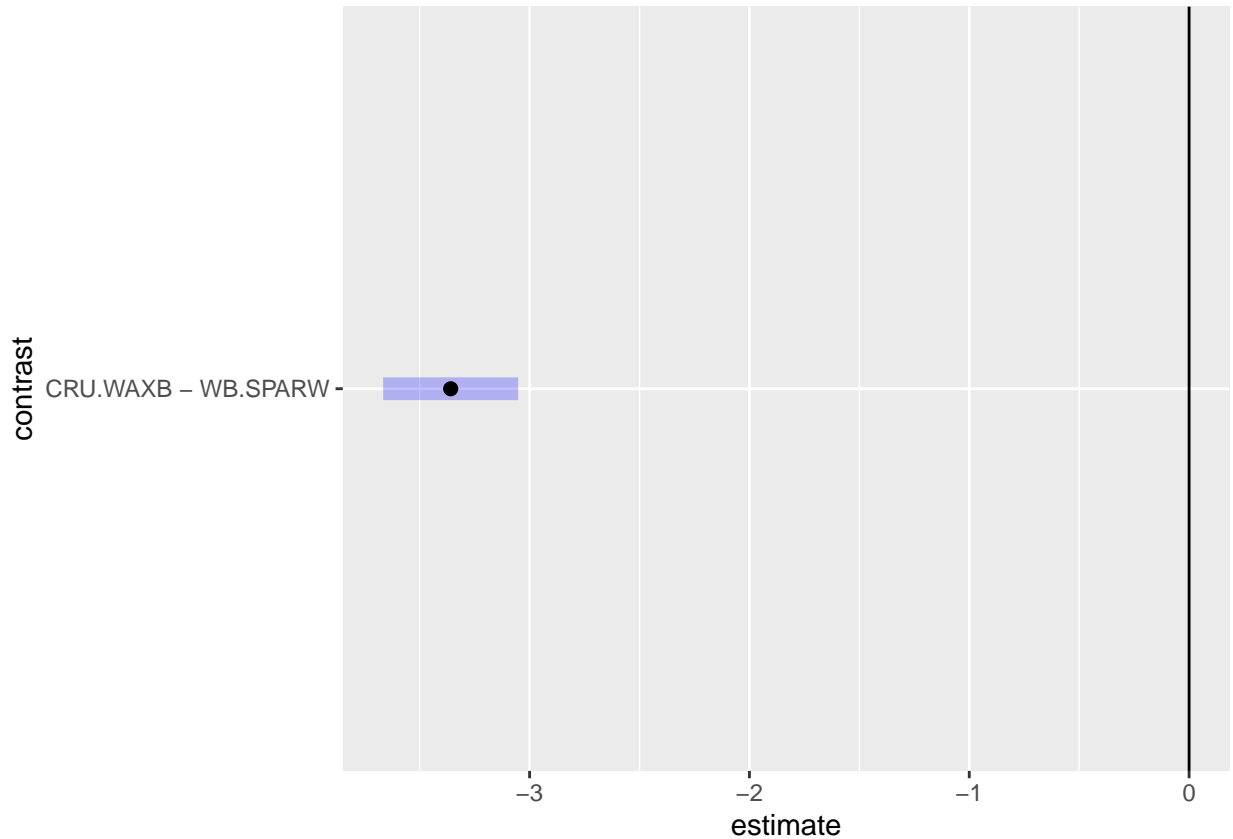
```
tamaño_efecto
```

```
## contrast          estimate    SE df t.ratio p.value  
## CRU.WAXB - WB.SPARG   -3.36 0.15 28 -22.400 <.0001
```

```
confint(tamaño_efecto)
```

```
## contrast          estimate    SE df lower.CL upper.CL  
## CRU.WAXB - WB.SPARG   -3.36 0.15 28   -3.67   -3.05  
##  
## Confidence level used: 0.95
```

```
plot(tamaño_efecto) + geom_vline(xintercept = 0)
```



R//La especie WB.SPARG tiene en promedio 3.36 mm mas largo el pico que la especie CRU.WAXB con un intervalo de confianza que va desde 3.67 a 3.05 mm

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 esta desarrollada para probar si tal aseveración es cierta o no.

- 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.

```
ee_1 = (50/sqrt(10))*(sqrt(1-(10/100)))*(sqrt(100/(100-1)))
```

```
ee_1
```

```
## [1] 15.07557
```

```
ee_2 = (50/sqrt(100))*(sqrt(1-(100/1000)))*(sqrt(1000/(1000-1)))
```

```
ee_2
```

```
## [1] 4.74579
```

```
ee_3 = (50/sqrt(1000))*(sqrt(1-(1000/10000)))*(sqrt(10000/(10000-1)))
```

```
ee_3
```

```
## [1] 1.500075
```

```
ee_4 = (50/sqrt(10000))*(sqrt(1-(10000/1000000)))*(sqrt(1000000/(1000000-1)))
```

```
ee_4
```

```
## [1] 0.497494
```

```
ee_5 = (50/sqrt(1000000))*(sqrt(1-(1000000/10000000)))*(sqrt(10000000/(10000000-1)))
```

```
ee_5
```

```
## [1] 0.04743417
```

```
c <- data.table("N =" = c("n =", "Error estándar ="), "100" = c(10, "15.07"), "1000" = c("100", "4.75")
kable(c)
```

N =	100	1000	10000	1000000	10000000
n =	10	100	1000	10000	1000000
Error estándar =	15.07	4.75	1.50	0.49	0.04

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

R// Se puede observar que a medida que se incrementa la población y la muestra y si el desvío estandar es el mismo, el error estandar disminuye. El error estandar mide la variabilidad del estadístico con respecto a la población, por lo que si se incrementa la población la variabilidad disminuye.