

Tarea_02

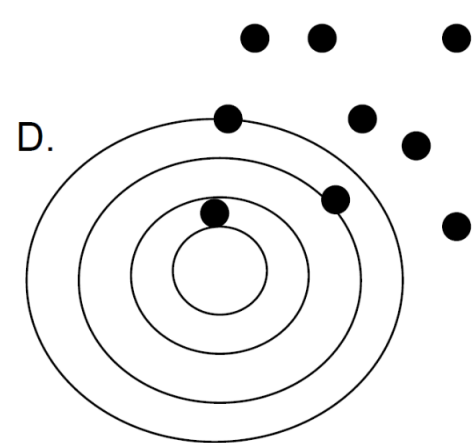
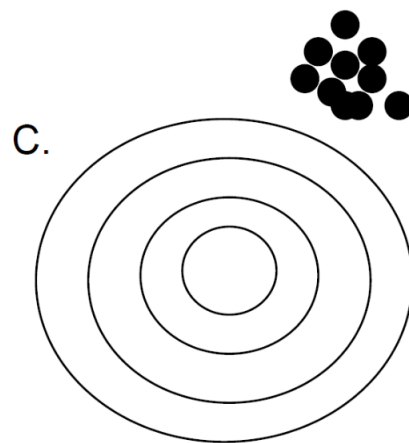
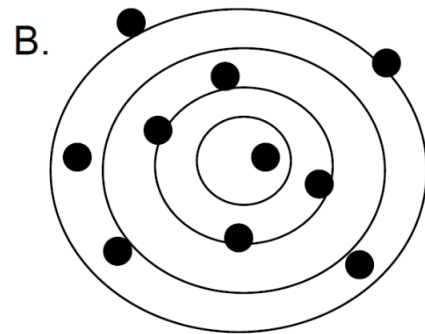
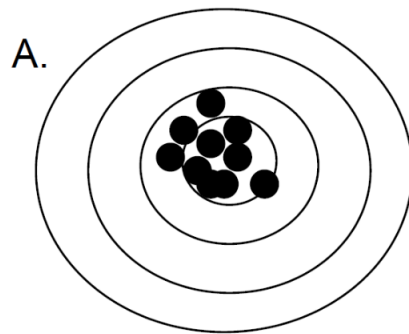
Nota: 9,0

```
library(tidyverse)
library(knitr)
library(rio)
library(visdat)
library(modelbased)
library(DescTools)
library(emmeans)
```

Nombre del estudiante: Alejandro Sebiani Calvo

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 preciso y exacto

B: no sesgado pero impreciso e inexacto

C: sesgado preciso e inexacto

D: sesgado impreciso e inexacto

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

```
Pinzones<- import("PinzonesAfricanos.csv")
Pinzones <- na.omit(Pinzones)
head(Pinzones)
```

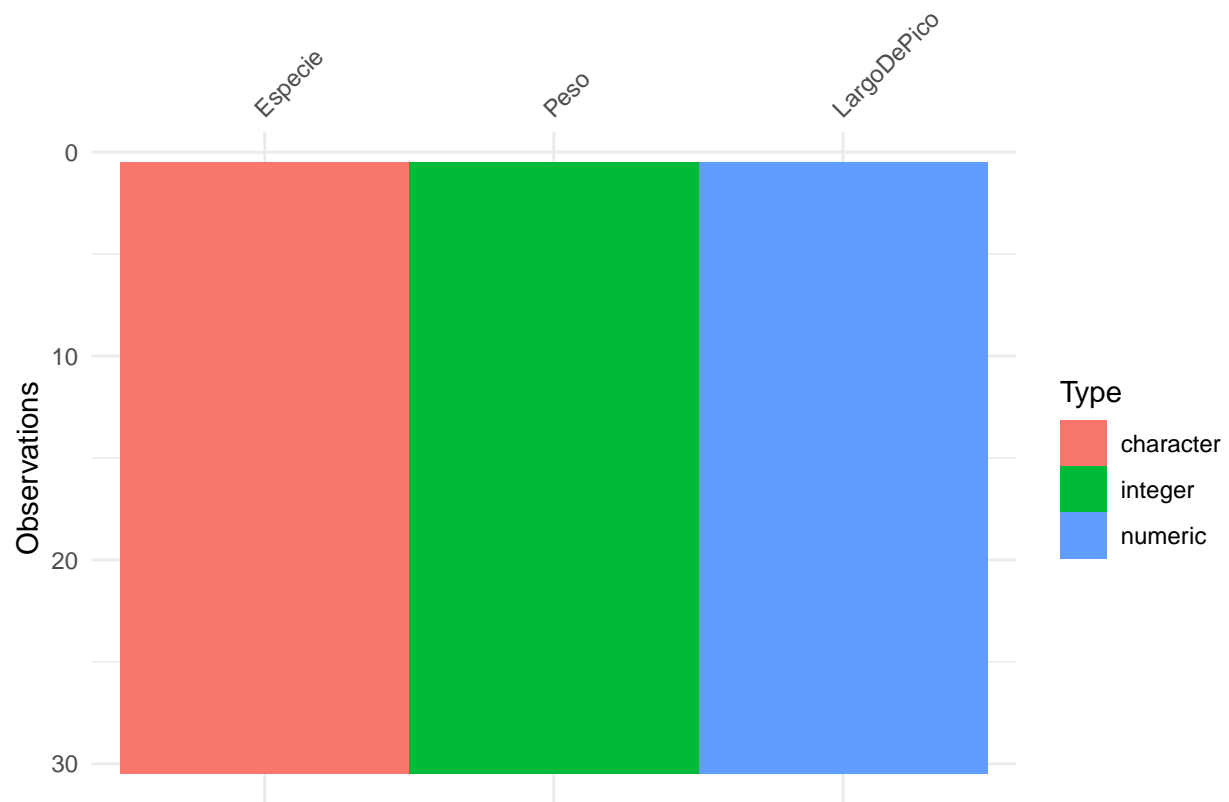
¿Cuál de las especies tiene el pico más largo?

##	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
```

```
#Revisión de clase de factores
```

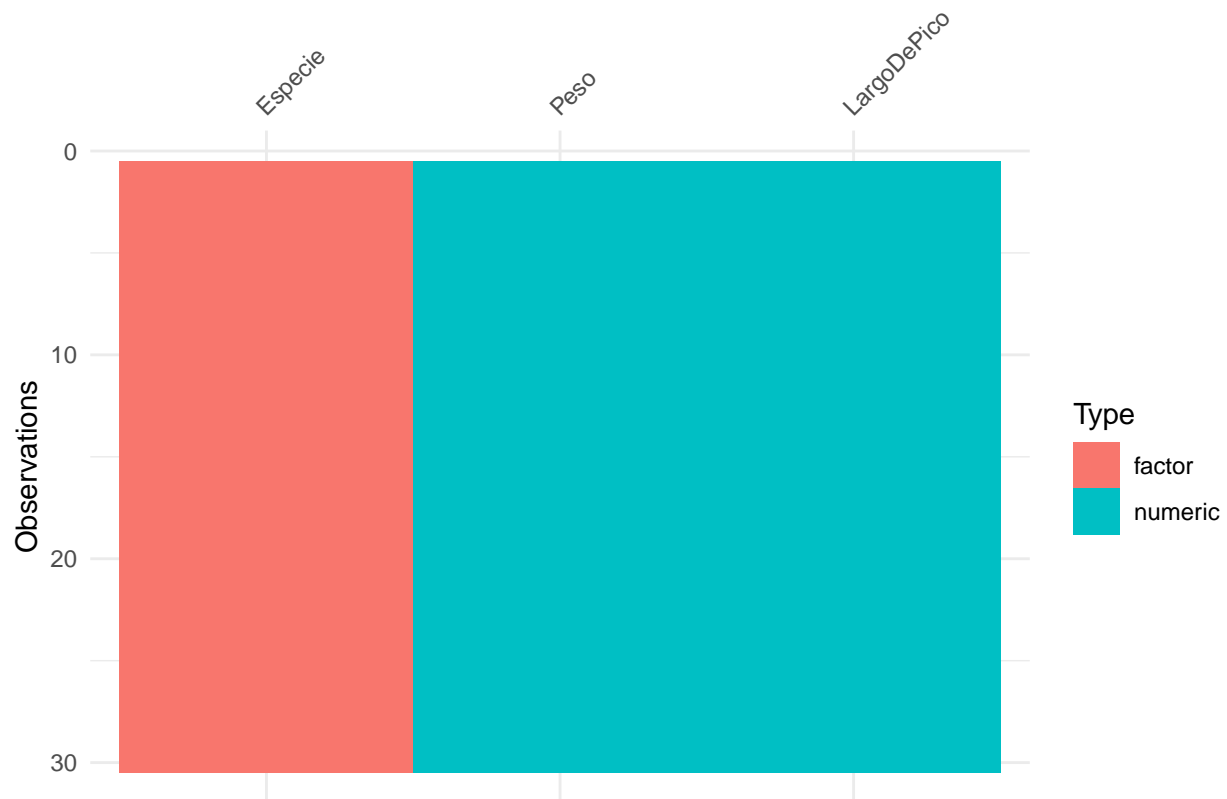
```
vis_dat(Pinzones)
```



```
#Conversión de clase de factores
```

```
Pinzones$Especie <-factor(Pinzones$Especie)
Pinzones$Peso <-as.numeric(Pinzones$Peso)
```

```
vis_dat(Pinzones)
```



¿Cuál de las especies tiene el pico más largo?

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

#Antes de analizar propiamente el valor de P quisiera revisar y graficar las estadísticas, principalmente la media, desviación estándar, error estándar e intervalos de confianza de los valores de las mediciones del largo de pico según especie

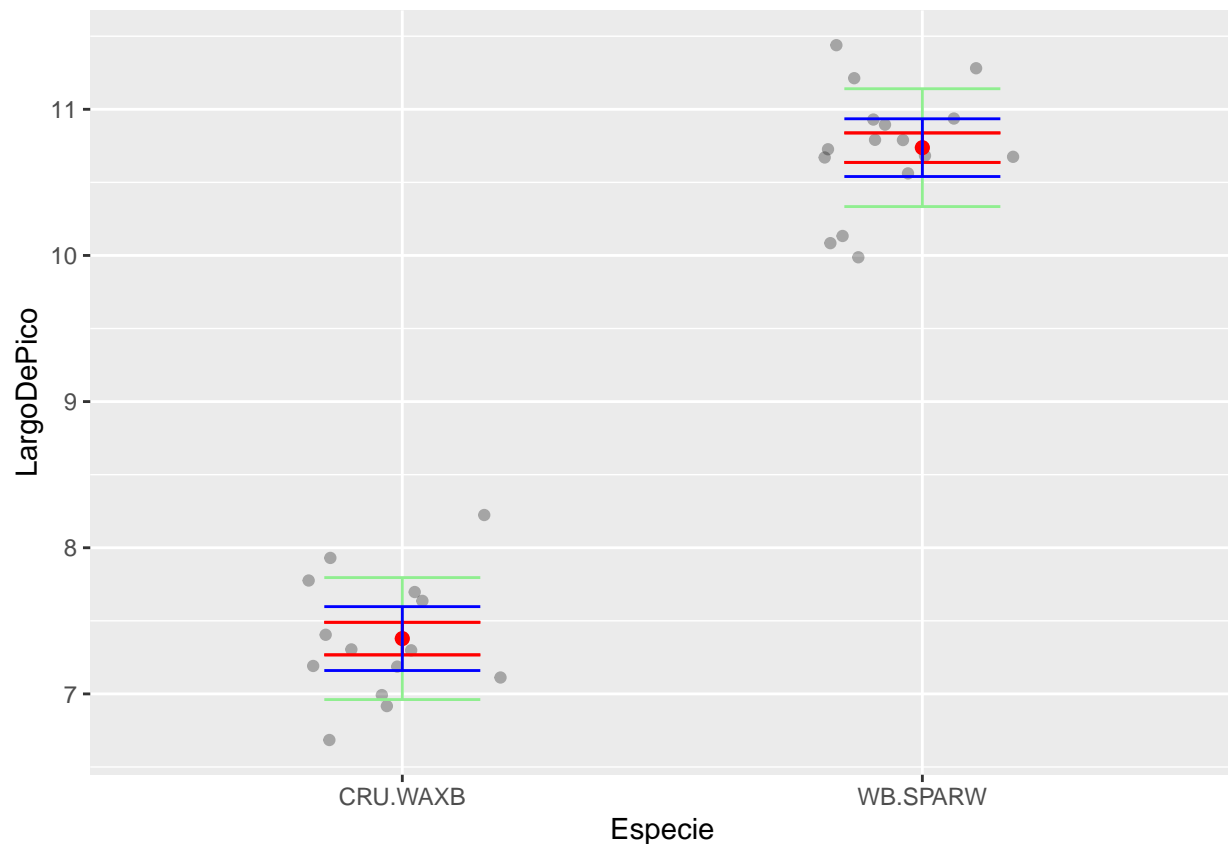
```
Estadisticas <- Pinzones %>%
  group_by(Especie) %>%
  summarise(media=mean(LargoDePico),
            n = n(),
            de = sd(LargoDePico),
            ee= de/sqrt(n()),
            li = media - 1.96*ee,
            ls = media + 1.96*ee
  )
```

Estadisticas

```
## # A tibble: 2 x 7
##   Especie media    n    de    ee    li    ls
##   <fct>    <dbl> <int> <dbl> <dbl> <dbl> <dbl>
## 1 CRU.WAXB  7.38     14 0.417 0.112  7.16  7.60
## 2 WB.SPARG 10.7     16 0.403 0.101 10.5  10.9
```

```
ggplot(Pinzones, aes(x=Especie, y= LargoDePico)) +
  geom_jitter (width=0.2, alpha=0.3)+
  geom_point (data = Estadisticas, aes( x=Especie, y=media), color = "red", size=2)+

  geom_errorbar(data= Estadisticas,aes(x=Especie, y = media, ymin=media - de, ymax= media + de), width=0.1, color="green"),
  geom_errorbar(data= Estadisticas,aes(x=Especie, y = media, ymin=media - ee, ymax= media + ee), width=0.1, color="blue"),
  geom_errorbar(data= Estadisticas,aes(x=Especie, y = media, ymin=media - ee, ymax= media + ee), width=0.1, color="red"),
  geom_errorbar(data= Estadisticas,aes(x=Especie, y = media, ymin= li, ymax= ls), width=0.3, color = "black")
```



#Se visualiza gráficamente que la media de las mediciones del largo de pico de la especie WB.SPARW es mayor a la de la especie CRU.WAXB. Además se nota que la desviación estándar entre ambos no se sobrepone, con una prueba de T student conoceremos con certeza si existe diferencia estadística significativa para el largo de pico entre ambas especies

```
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.SPARG
##          7.378571          10.737500
```

#El valor de P es de 2.2e-16, mucho menor al valor establecido como referente de 0.05 por lo que se puede inferir con un 95% de confianza que existe diferencia estadística significativa entre las mediciones del largo de pico por especie de ave. #Con el mismo 95% de confianza se puede inferir que el largo del pico de la especie WB.SPARG es un valor entre 10.540 mm y 10.935 mm mientras que el largo del pico de la especie CRU.WAXB es un valor de entre 7.160 mm y 7.60 mm. #####¿Cuál de las especies tiene el pico más largo?

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

Pero lo que respondes no es lo que se pregunta?

#Conociendo que existe diferencia estadística significativa entre las medias del largo de pico por especie, se quiere evaluar el tamaño de esa diferencia entre las medias

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

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

```
mean(Pinzones$LargoDePico)
```

```
## [1] 9.17
```

```
3.36/9.17
```

```
## [1] 0.3664122
```

Conociendo la diferencia entre las medias de ambas especies y dividiendola entre la media de la variable, se puede inferir que existe una diferencia del 36,6% en el largo del pico de ambas especies, teniendo entonces la especie WB.SPARG un pico 36.3% más largo que la especie CRU.WAXB

#Otra manera de evaluar el tamaño del efecto es analizando el contraste entre las medias a partir de un modelo lineal

```
modelo <- lm(LargoDePico ~ Especie, data = Pinzones )
```

```
summary(modelo)
```

```
##
## Call:
## lm(formula = LargoDePico ~ Especie, data = Pinzones)
##
```

```
## 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.SPARG  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
```

#En el resumen del modelo lineal se observa la diferencia entre las medias que es 3.3589

```
library(modelbased)
```

#tamaño del efecto o contraste entre medias

```
contraste <-estimate_contrasts(modelo, "Especie")
contraste
```

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

#De esta manera se puede analizar la diferencia entre las medias con respaldo estadístico , podemos observar que el intervalo de confianza del 95% es bastante estrecho y que la diferencia entre las medias es considerable (dependiendo de la pregunta biológica)

#El tamaño del efecto se puede graficar para así analizar la significancia biológica o práctica del mismo con mayor facilidad a los ojos del investigador/experto de la siguiente manera:

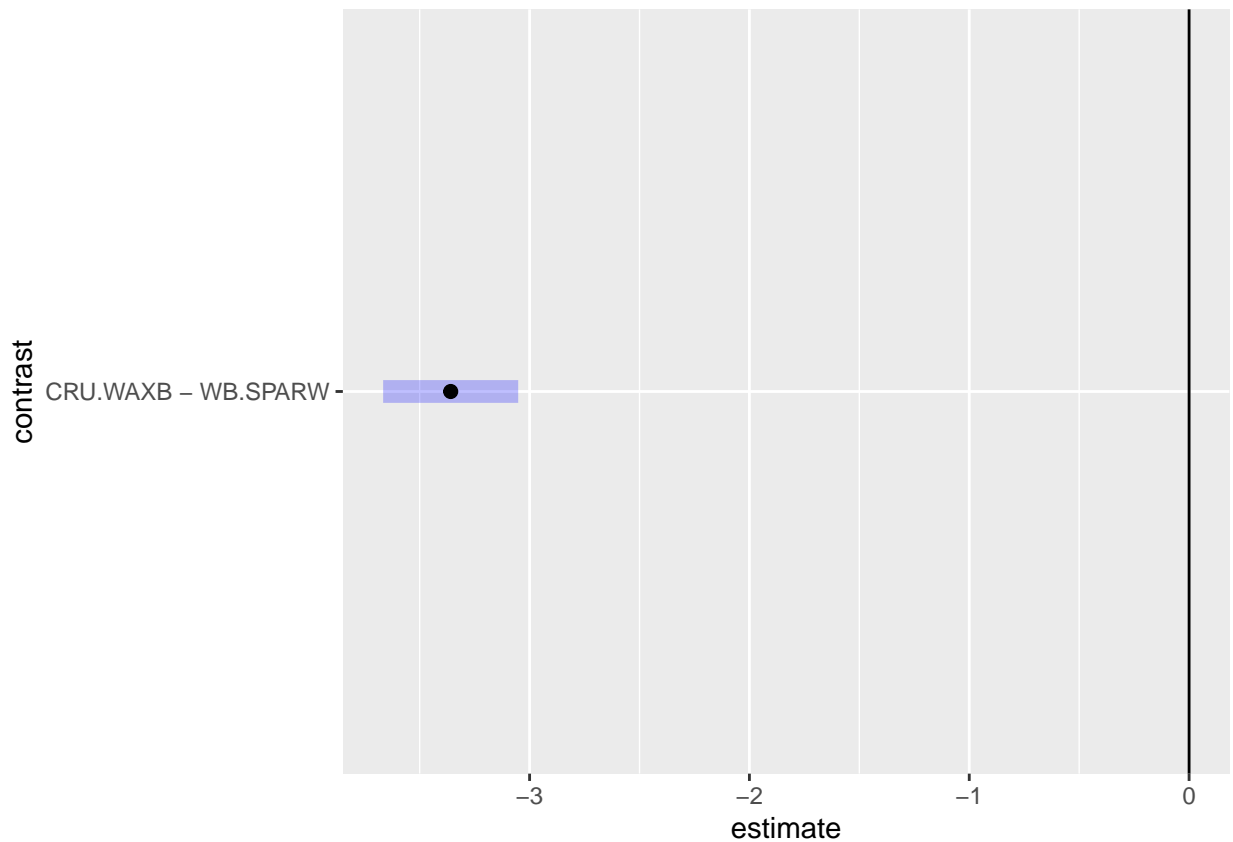
```
e <- emmeans(modelo,"Especie")
e
```

```
## 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
```

```
p<-pairs(e)
p
```

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

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



#El gráfico representa la diferencia de medias, se logra observar que nunca cruza o traslapa con el valor de 0 que indicaría que no hay diferencia entre las medias de una especie con respecto a otra.

Esta interpretación no es correcta

#Una interpretación que me atrevo a hacer sin ser experto, solo con fines de análisis del tamaño del efecto, es que el largo del pico de la especie WB.SPARW es de al menos 2 mm mayor que el largo del pico de la especie CRU.WAXB. según los datos analizados y visualizados aquí, por lo que podría tener una significancia en identificación taxonómica de las aves.

Por qué 2 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.

- 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.

#Utilicé la siguiente fórmula con la corrección de poblaciones finitas para obtener los datos tabulados :

```
(50/sqrt(n))*(sqrt((N-n)/(N-1)))
```

#Ejemplo de uso de la fórmula


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

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

```
##
```

```
## 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 =	15,08	4,75	1,50	0,50	0,05

```
##          N    100 1000 10000 1000000 10000000
```

```
## 1:      n =     10   100   1000    10000   1000000
```

```
## 2: Error estándar = 15,08 4,75  1,50    0,50    0,05
```

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

#El tamaño de la muestra tiene una relación inversamente proporcional al error estándar. #Entre más grande sea la muestra, las mediciones y datos obtenidos a partir de ella, serán mucho más representativos (siempre y cuando sean al azar) de la población o sistema.

A mayor tamaño de muestra menor error estándar y más precisa la estimación