

Tarea_02

9,5

```
library(knitr)
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2    v purrr  0.3.4
## v tibble  3.0.4    v dplyr  1.0.2
## v tidyr   1.1.2    v stringr 1.4.0
## v readr   1.4.0    v forcats 0.5.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(rio)
library(flextable)
```

```
##
## Attaching package: 'flextable'

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

```
library(DescTools)
library(data.table)
```

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

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

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

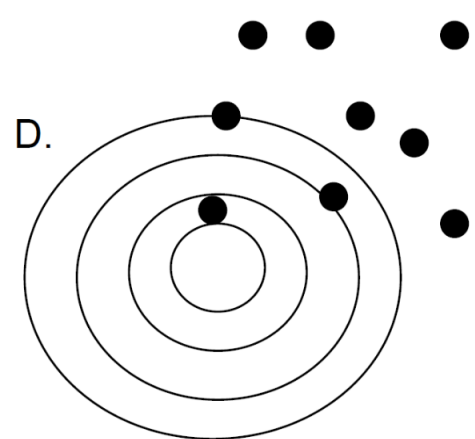
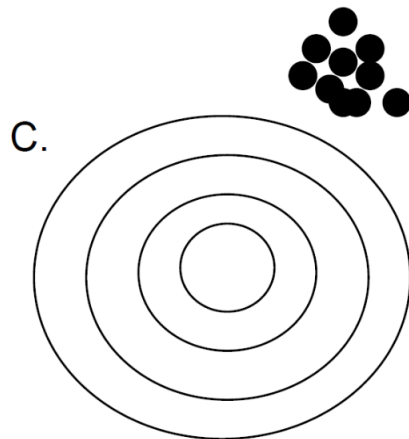
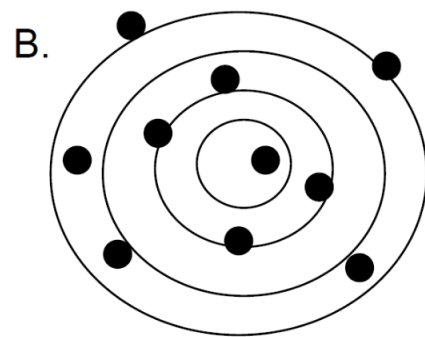
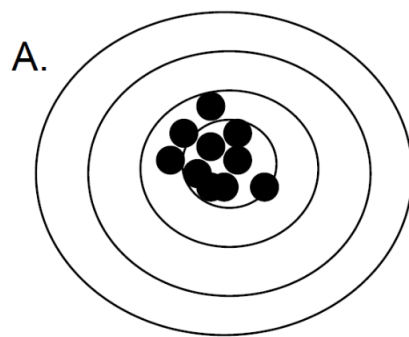
## The following object is masked from 'package:purrr':
##
##   transpose
```

```
library(modelbased)
```

Nombre del estudiante: Sergio Rodríguez Segura

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: sesgado, precisa = exacta.

B: no sesgado pero impreciso = inexacto

C: sesgado, pero preciso = inexacta

D: sesgado, impreciso = 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 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
```

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

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

```
picomas_largo
```

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

¿Cuál de las especies tiene el pico más largo?

La especie WB.SPARG es la que posee el pico más largo

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

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

```
summary(Puntovista_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.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
```

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

```
## 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
```

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
```

Si restamos (WB.SPARG - CRU.WAXB) = 3.4 (3,0 - 3,7) Se puede decir que con una confianza del 95% que la diferencia entre el largo de pico para ambas especies está representada por la función (3,0 - 3,7 cm)

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 desviación 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.

```
error_estandar_1 = (50/sqrt(10))*(sqrt(1-(10/100)))*(sqrt(100/(100-1)))
```

```
error_estandar_1
```

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

```
error_estandar_2 = (50/sqrt(100))*(sqrt(1-(100/1000)))*(sqrt(1000/(1000-1)))
```

```
error_estandar_2
```

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

```
error_estandar_3 = (50/sqrt(1000))*(sqrt(1-(1000/10000)))*(sqrt(10000/(10000-1)))
```

```
error_estandar_3
```

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

```
error_estandar_4 = (50/sqrt(10000))*(sqrt(1-(10000/1000000)))*(sqrt(1000000/(1000000-1)))
```

```
error_estandar_4
```

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

```
error_estandar_5 = (50/sqrt(1000000))*(sqrt(1-(1000000/10000000)))*(sqrt(10000000/(10000000-1)))
```

```
error_estandar_5
```

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

```
c <- data.table("N =" = c("n =", "Error estandar ="), "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 estandar =	15.07	4.75	1.50	0.4974	0.0474

b. Se te ocurre alguna explicaci3n de los resultados obtenidos.

Cuando aumentamos el tama±o y la muestra manteniendo el mismo desvio estandar, disminuira el error estandar. La variabilidad se expresa al aumentar la poblaci3n de la muestra esta disminuye