

Examen de Prueba

MPPP - Estadística

Juvenal Campos

10/4/2021

Sección Teórica.

- ¿Cuales diferencias puede recordar entre r-base y el tidyverse? ¿Se puede hacer lo mismo con cualquiera de las dos?
- ¿Que hace la librería haven? ¿Qué otras librerías utilizamos en lo que va del curso?
- ¿Qué es una media condicional?
- ¿Qué es la correlación? ¿Qué significa una correlación de 1, -1 y 0?
- ¿Qué es una tabla de contingencia?
- ¿Qué son las etiquetas en una base de datos?
- ¿Qué diferencia hay entre las etiquetas de variable y las etiquetas de valor?
- ¿Qué significa recodificar una variable?

Sección Práctica.

1. Abra la encuesta que se encuentra en el archivo “*Encuesta_covid.rds*” y guardala en un objeto llamado bd. ¿Qué variables contiene?

Ojo: Recuerde llamar a las librerías que va a utilizar desde el principio.

```
library(haven)
library(tidyverse)
library(sjmisc)
library(sjlabelled)

# Leo la información
bd <- readRDS("Encuesta_covid.rds")
```

2. ¿Qué información contiene la variable vacunado?

```
# Variable:
get_label(bd$estrato)
```

```
## [1] "Estrato"
```

```
get_labels(bd$estrato)
```

```
## [1] "Urbano" "Rural"
```

3. ¿Qué información contiene la variable `grado_educ`?

```
# Variable:  
get_label(bd$grado_educ)
```

```
## [1] "Último grado educativo terminado"
```

```
get_labels(bd$grado_educ)
```

```
## [1] "Primaria"      "Secundaria"    "Preparatoria" "Licenciatura" "Posgrado"
```

4. ¿Cual es el valor promedio del ingreso mensual? ¿Cual es su desviación estándar?

```
# Variable:  
mean(bd$ingreso_mensual)
```

```
## [1] 19093.98
```

```
sd(bd$ingreso_mensual)
```

```
## [1] 10838.79
```

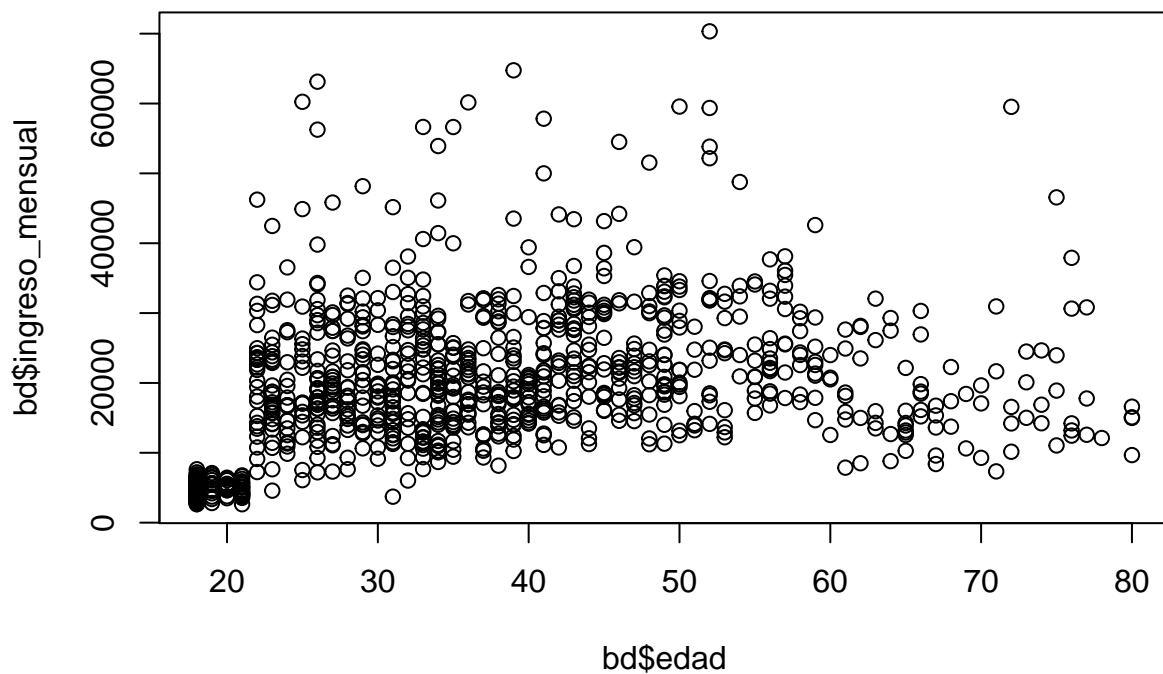
5. ¿Cual es la correlación entre ingreso y la edad? Utilice el método de Spearman. ¿Cómo podría interpretar el resultado?

```
# Variable:  
cor(bd$ingreso_mensual,  
    bd$edad,  
    method = "s")
```

```
## [1] 0.4964551
```

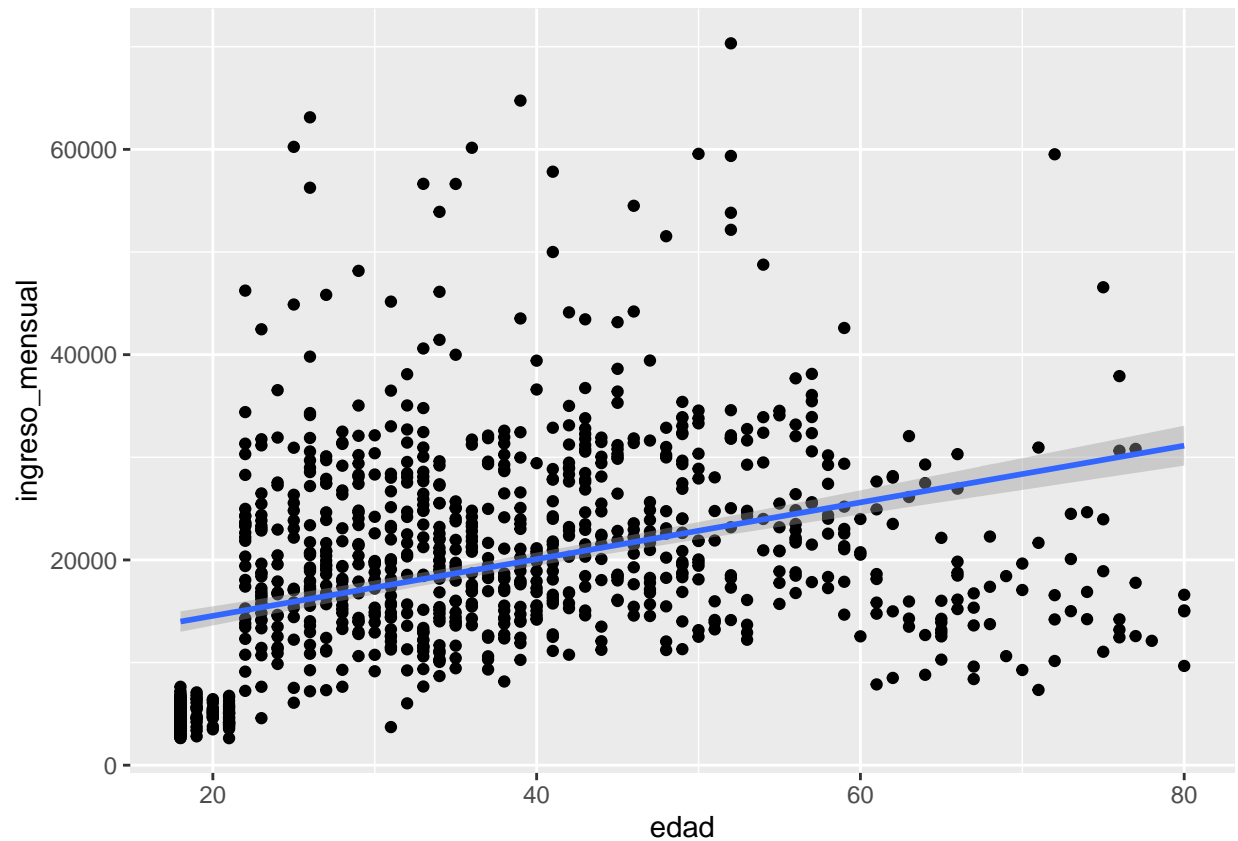
6. Grafique la relación entre ingreso y edad:

```
plot(bd$edad, bd$ingreso_mensual)
```



```
bd %>%  
  ggplot(aes(x = edad, y = ingreso_mensual)) +  
  geom_point() +  
  geom_smooth(method = "lm")
```

```
## 'geom_smooth()' using formula 'y ~ x'
```



7. Obtenga la media condicional del ingreso mensual por edad. ¿Cuál es la edad con mayor promedio de ingreso?

```
bd %>%
  group_by(edad) %>%
  summarise(media = mean(ingreso_mensual)) %>%
  print(n = Inf)
```

```
## # A tibble: 62 x 2
##   edad media
##   <dbl> <dbl>
## 1    18  4831.
## 2    19  5183.
## 3    20  5117.
## 4    21  4781.
## 5    22 21416
## 6    23 19166.
## 7    24 18591.
## 8    25 22480.
## 9    26 22886.
## 10   27 20856.
## 11   28 20028.
## 12   29 22996.
## 13   30 18965.
## 14   31 19459.
```

```

## 15    32 22570.
## 16    33 21402.
## 17    34 20998.
## 18    35 20602.
## 19    36 22018.
## 20    37 19078.
## 21    38 20768.
## 22    39 22555.
## 23    40 20445.
## 24    41 23694.
## 25    42 25056.
## 26    43 26672.
## 27    44 22833.
## 28    45 26797.
## 29    46 25318.
## 30    47 22219.
## 31    48 24866.
## 32    49 24387.
## 33    50 26342.
## 34    51 18852.
## 35    52 34294.
## 36    53 22232
## 37    54 31582.
## 38    55 23489.
## 39    56 24648
## 40    57 29699
## 41    58 24174.
## 42    59 24185.
## 43    60 19450.
## 44    61 18258.
## 45    62 20637
## 46    63 20385.
## 47    64 19566.
## 48    65 14409.
## 49    66 20809
## 50    67 12742.
## 51    68 17802.
## 52    69 14525
## 53    70 15330.
## 54    71 19983.
## 55    72 25112.
## 56    73 19862.
## 57    74 18587
## 58    75 25119.
## 59    76 21684.
## 60    77 20390
## 61    78 12117
## 62    80 14088.

```

8. Elabore una tabla de frecuencias de la variable **vacunado**. ¿Cual es el porcentaje de entrevistados vacunados?

```
frq(bd$vacunado)
```

```
##
## Ha sido vacunado contra el COVID-19 (x) <numeric>
## # total N=1000 valid N=1000 mean=1.16 sd=0.37
##
## Value | Label | N | Raw % | Valid % | Cum. %
## -----
## 1 | Vacunado | 838 | 83.80 | 83.80 | 83.80
## 2 | No vacunado | 162 | 16.20 | 16.20 | 100.00
## <NA> | <NA> | 0 | 0.00 | <NA> | <NA>
```

9. Elabore una tabla de contingencia de las variables `vacunado` y `opinion_vacuna`. ¿Cual es el porcentaje de personas vacunadas que no tienen una opinión favorable sobre la vacuna?

```
flat_table(data = bd, opinion_vacuna, vacunado, margin = "col")
```

```
##              vacunado Vacunado No vacunado
## opinion_vacuna
## Favorable              76.25          38.27
## Desfavorable           23.75          61.73
```

10. Obtenga los quintiles del ingreso.

```
quantile(x = bd$ingreso_mensual, seq(0,1,by = 0.2))
```

```
##      0%      20%      40%      60%      80%     100%
## 2635.0  9272.2 15957.0 20816.4 27598.0 70331.0
```

11. Genere una nueva variable llamada `grupo_edad`, en la cual va a meter en el grupo **Jóvenes** a los que tienen edades de entre 18 a 30 años, **Adultos** a los que tienen entre 31 y 60 años y **Adultos mayores** a los que tienen entre 61 y más. Después obtenga la tabla de contingencia entre `grupo_edad` y `opinion_vacuna`.

```
bd <- bd %>%
  mutate(grupo_edad = rec(edad,
                           rec = "min:30 = 1 [Jóvenes];
                                   31:60 = 2 [Adultos];
                                   61:max = 3 [Adultos mayores]"))

flat_table(bd, grupo_edad, opinion_vacuna, margin = "row")
```

```
##              opinion_vacuna Favorable Desfavorable
## grupo_edad
## Jóvenes              71.68          28.32
## Adultos              69.54          30.46
## Adultos mayores     65.82          34.18
```