

Práctica dirigida 5

Contents

Para la realización del ejercicio se toman las siguientes variables de la encuesta:	1
Ejercicio 1: Describe la variable	2
Graficando una variable	2
Ejercicio 2:	4
Ejercicio 3	8
Ejercicio 4	8
Ejercicio 5	8
Ejercicio 6	8

FACULTAD DE CIENCIAS SOCIALES - PUCP

Curso: POL 278 - Estadística para el análisis político 1 | Semestre 2023 - 2

Ejercicios:

Utilizamos la base de datos de la Encuesta Nacional de Programas Presupuestales (ENAPRES) 2021. Descargala en: <https://www.datosabiertos.gob.pe/dataset/encuesta-nacional-de-programas-presupuestales-enapres-2021-instituto-nacional-de-estad%C3%ADstica>

El cuestionario, base de datos, diccionario de variables se encuentran en la carpeta de la práctica dirigida.

Para todas las preguntas utilice un gráfico para evidenciar su respuesta y redacte un párrafo donde exprese su respuesta.

Abrimos la base de datos

Usamos el paquete *rio* para importar la base de datos.

```
library(rio)
data=import("data_ejercicio.csv")
```

Para la realización del ejercicio se toman las siguientes variables de la encuesta:

AREA 1 = urbano 2 = rural P800A_1_1 EN LOS ÚLTIMOS 12 MESES, ¿USTED VISITÓ ALGÚN: Monumento histórico (Iglesia, Catedral, fortaleza, muralla, casona, centro histórico, mausoleo, etc.)? P800A_2_1 EN LOS ÚLTIMOS 12 MESES, ¿USTED VISITÓ ALGÚN: Monumento Arqueológico (sitio arqueológico, huaca, ruina, sitios de fósiles)? P800A_3_1 EN LOS ÚLTIMOS 12 MESES, ¿USTED VISITÓ ALGÚN: Museo? P800A_8_1 ¿USTED ASISTIÓ A UN/UNA: Espectáculo de teatro? P800A_8_2 ¿USTED ASISTIÓ A UN/UNA: Espectáculo de danza? P800A_8_4 ¿USTED ASISTIÓ A UN/UNA: Espectáculo musical

(conciertos, festivales, etc.)? P800A_8_5 ¿USTED ASISTIÓ A UN/UNA: Función del Cine? P800A_8_8 ¿USTED ASISTIÓ A UN/UNA Biblioteca y/o sala de lectura? P800A_13_1 ¿USTED OBTUVO O ADQUIRIÓ: Libros? P207 SEXO 1 hombre 2 mujer P208_A Años cumplidos REGIONNATU Región Natural ESTRATO ESTRATO SOCIODEMOGRÁFICO indicador (indicador aditivo que mide el consumo de cultura)

Para mayor facilidad de entendimiento, se han renombrado las variables:

AREA, visita_monumentohistorico=P800A_1_1, visita_monumentoarqueologico=P800A_2_1, visita_museo=P800A_3_1, asistencia_teatro=P800A_8_1, asistencia_danza=P800A_8_2, asistencia_espectaculomusical=P800A_8_4, asistencia_cine=P800A_8_5, asistencia_biblioteca=P800A_8_8, adquirir_libro=P800A_13_1, sexo=P207, region_natural=REGIONNATU, estrato=ESTRATO indicador

Ejercicio 1: Describe la variable

Obtenga estadísticos, grafice y describa el comportamiento de la nueva variable: consumo de cultura (indicador)

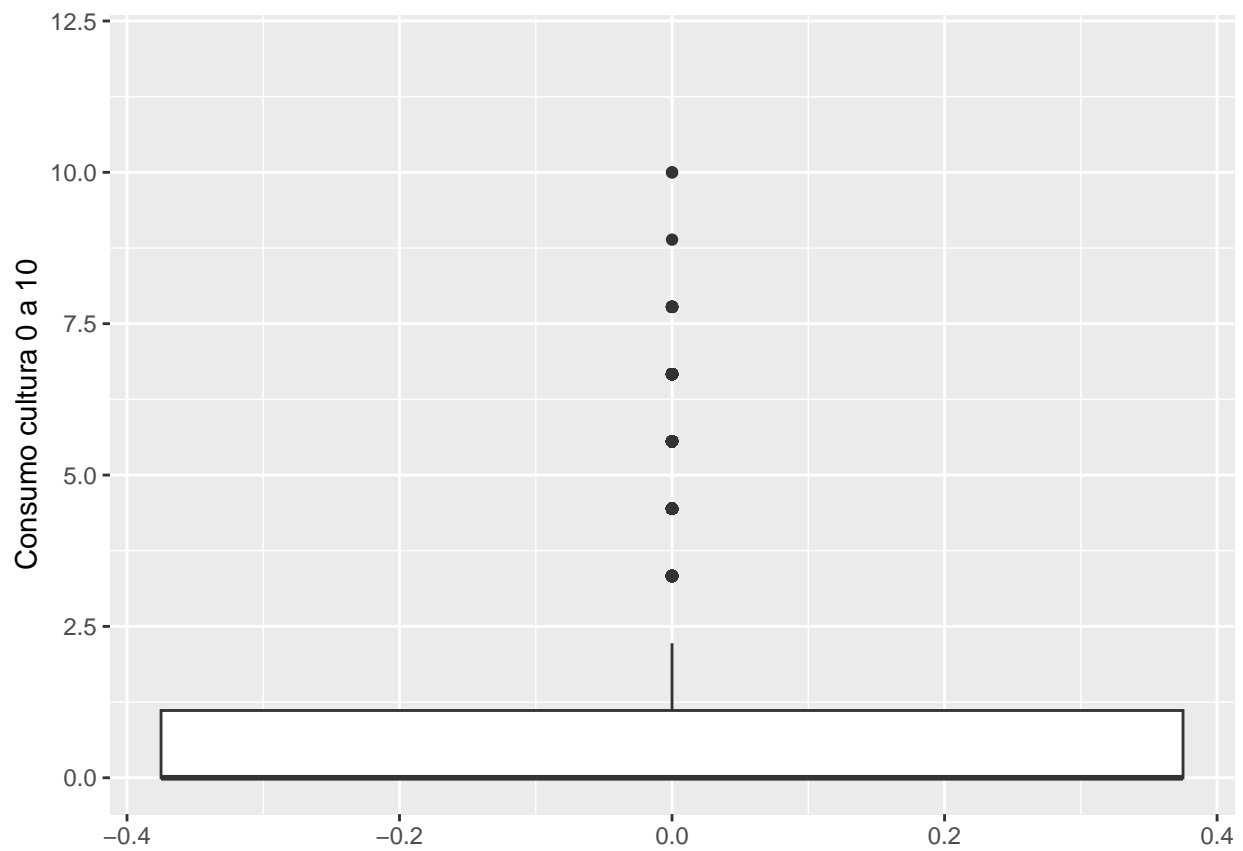
Usamos el comando *summarise* para solicitar una tabla resumen con mínimo (*min*), máximo (*max*), media (*mean*), mediana (*median*), desviación (*sd*), cuartiles (*quantile*).

```
library(dplyr)
resumen_consumo_cultura = data %>%
  dplyr::summarise(minimo = min(indicador, na.rm = TRUE),
    media = mean(indicador, na.rm = TRUE),
    mediana = median(indicador, na.rm = TRUE),
    maximo = max(indicador, na.rm = TRUE),
    desviacion = sd(indicador, na.rm = TRUE),
    iqr25= quantile(indicador, na.rm = TRUE, probs = 0.25),
    iqr75= quantile(indicador, na.rm = TRUE, probs = 0.75),
    iqr95= quantile(indicador, na.rm = TRUE, probs = 0.95))
resumen_consumo_cultura
```

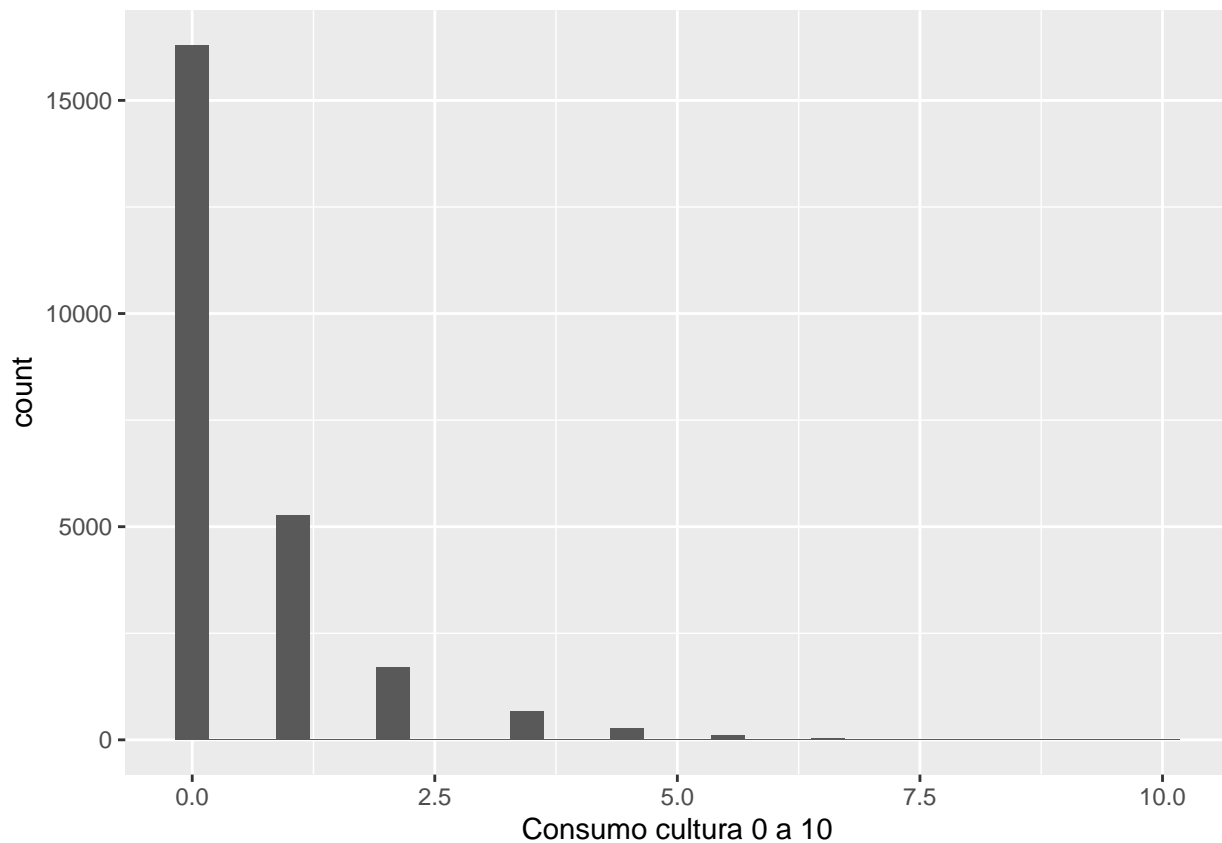
```
## minimo media mediana maximo desviacion iqr25 iqr75 iqr95
## 1 0 0.5745035 0 10 1.021341 0 1.111111 2.222222
```

Graficando una variable

```
library(ggplot2)
ggplot(data, aes(y=indicador))+
  geom_boxplot()+
  ylab("Consumo cultura 0 a 10") +
  scale_y_continuous(limits = c(0, 12))
```



```
ggplot(data, aes(x=indicador))+  
  geom_histogram() +  
  xlab("Consumo cultura 0 a 10")
```



Ejercicio 2:

Muchos investigadores afirman que existe una brecha entre el **consumo de cultura** según el estrato sociodemográfico **ESTRATO**. ¿Es cierto?

Descriptiva

```
table(data$ESTRATO)
```

```
##
##    1    2    3    4    5
## 2190 4077 5847 6302 5949
```

```
data = data %>%
  mutate(ESTRATO=factor(ESTRATO,
                        labels = c("A","B","C","D","E"),
                        ordered = T))
```

```
table(data$ESTRATO)
```

```
##
##    A    B    C    D    E
## 2190 4077 5847 6302 5949
```

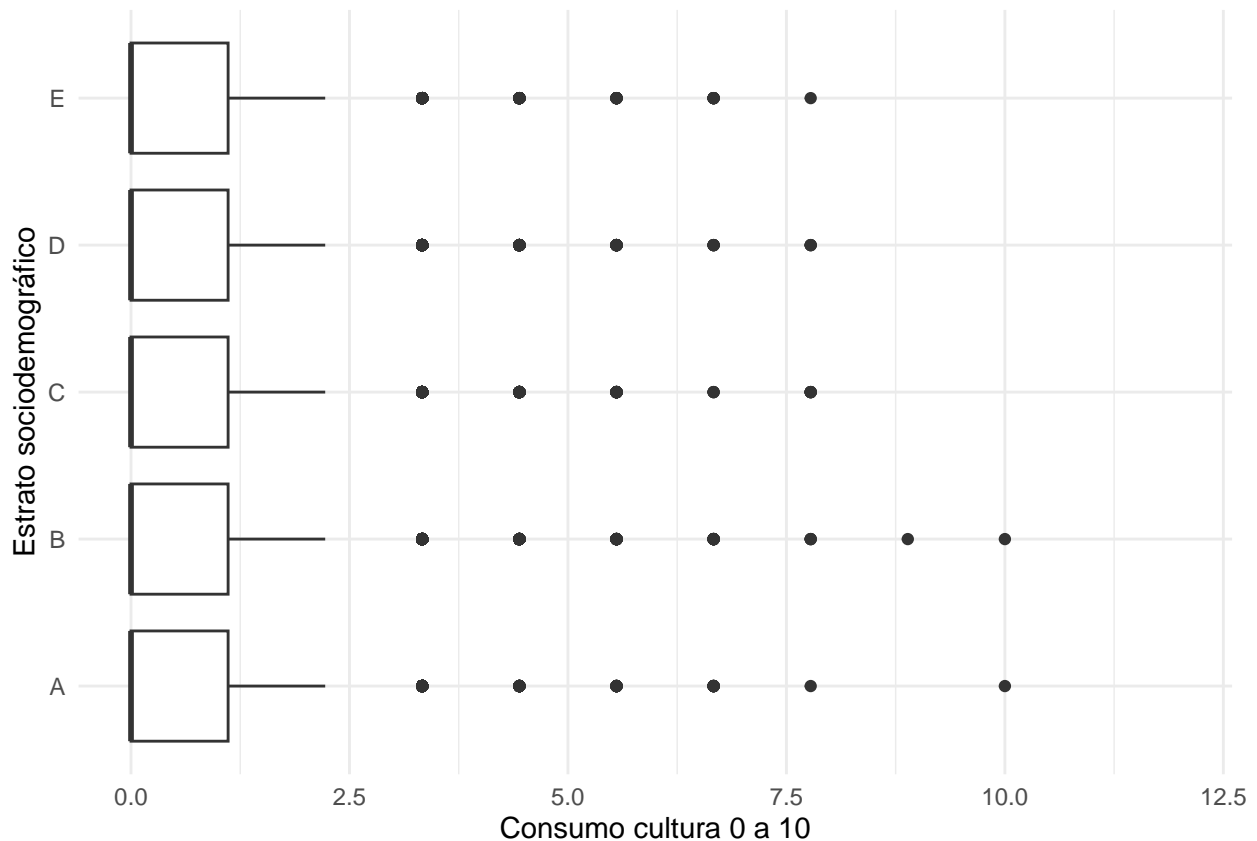
```
class(data$ESTRATO)
```

```
## [1] "ordered" "factor"
```

```
consumocultura_estrato = data %>%
  group_by(ESTRATO) %>%
  summarise(media = mean(indicador, na.rm = TRUE),
            desviacion = sd(indicador, na.rm = TRUE))
consumocultura_estrato
```

```
## # A tibble: 5 x 3
##   ESTRATO media desviacion
##   <ord>   <dbl>     <dbl>
## 1 A       0.833     1.21
## 2 B       0.744     1.17
## 3 C       0.555     1.01
## 4 D       0.516     0.959
## 5 E       0.444     0.875
```

```
ggplot(data, aes(y=indicador, x=ESTRATO))+
  geom_boxplot() +
  scale_y_continuous(limits = c(0, 12))+
  ylab("Consumo cultura 0 a 10")+
  xlab("Estrato sociodemográfico")+
  theme_minimal()+
  coord_flip()
```



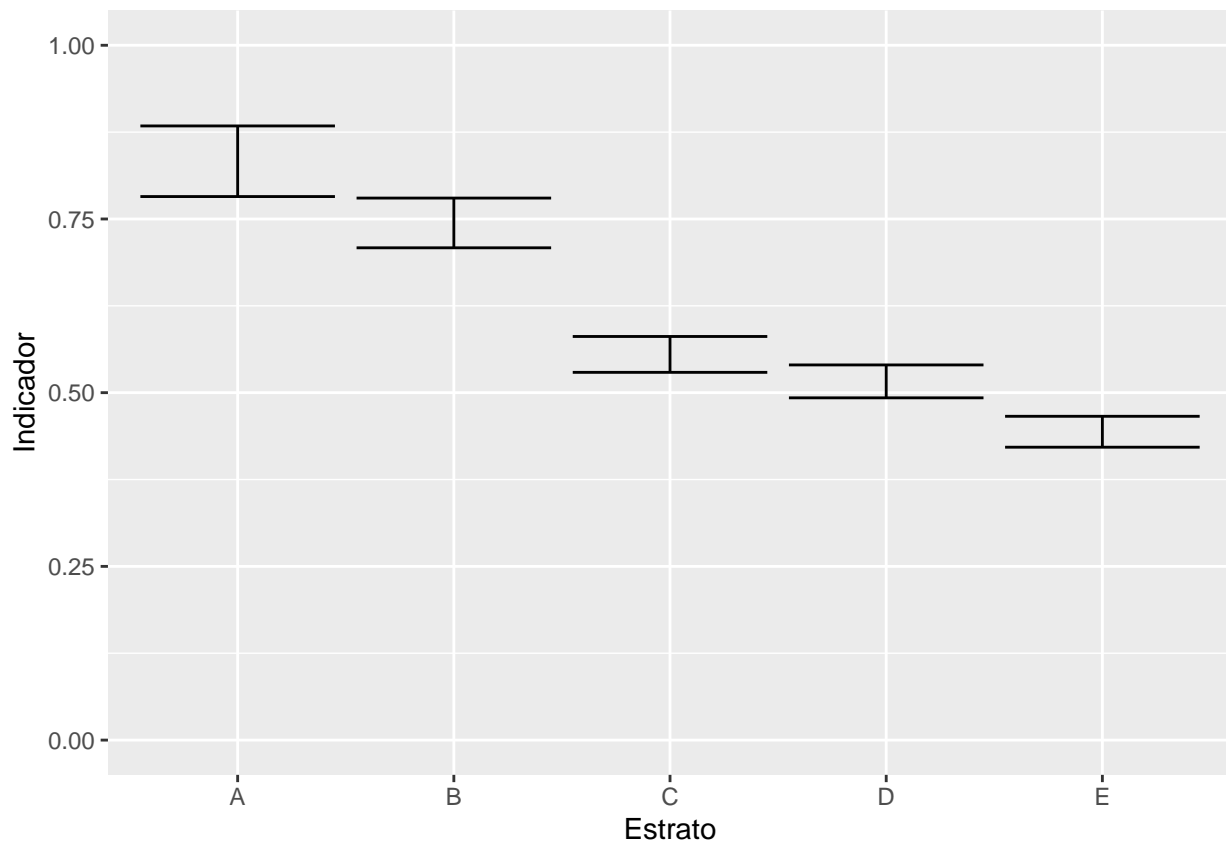
Inferencial

Ahora, realizaremos pruebas estadísticas para corroborar que efectivamente hay una diferencia por Estrato según el indicador de consumo cultural.

```
library(lsr)
tabla=data%>%
  group_by(ESTRATO) %>%
  summarise(Desviacion = sd(indicador, na.rm=T),
            Media = mean(indicador, na.rm=T),
            min = ciMean(indicador,conf = 0.95, na.rm=T)[1],
            max = ciMean(indicador,conf = 0.95, na.rm=T)[2],
            n=length(indicador)
            )
tabla
```

```
## # A tibble: 5 x 6
##   ESTRATO Desviacion Media   min   max     n
##   <ord>      <dbl> <dbl> <dbl> <dbl> <int>
## 1 A        1.21  0.833 0.782 0.884  2190
## 2 B        1.17  0.744 0.709 0.780  4077
## 3 C        1.01  0.555 0.529 0.581  5847
## 4 D        0.959 0.516 0.493 0.540  6302
## 5 E        0.875 0.444 0.422 0.466  5949
```

```
library(ggplot2)
ggplot(tabla,aes(x=ESTRATO,y=Media))+
  geom_errorbar(aes(ymin=min,ymax=max))+
  xlab("Estrato")+ylab("Indicador")+ylim(0,1)
```



```
anova = aov(indicador~ESTRATO, data =data)
summary(anova)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## ESTRATO      4     389    97.31   94.72 <2e-16 ***
## Residuals 24360  25026     1.03
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

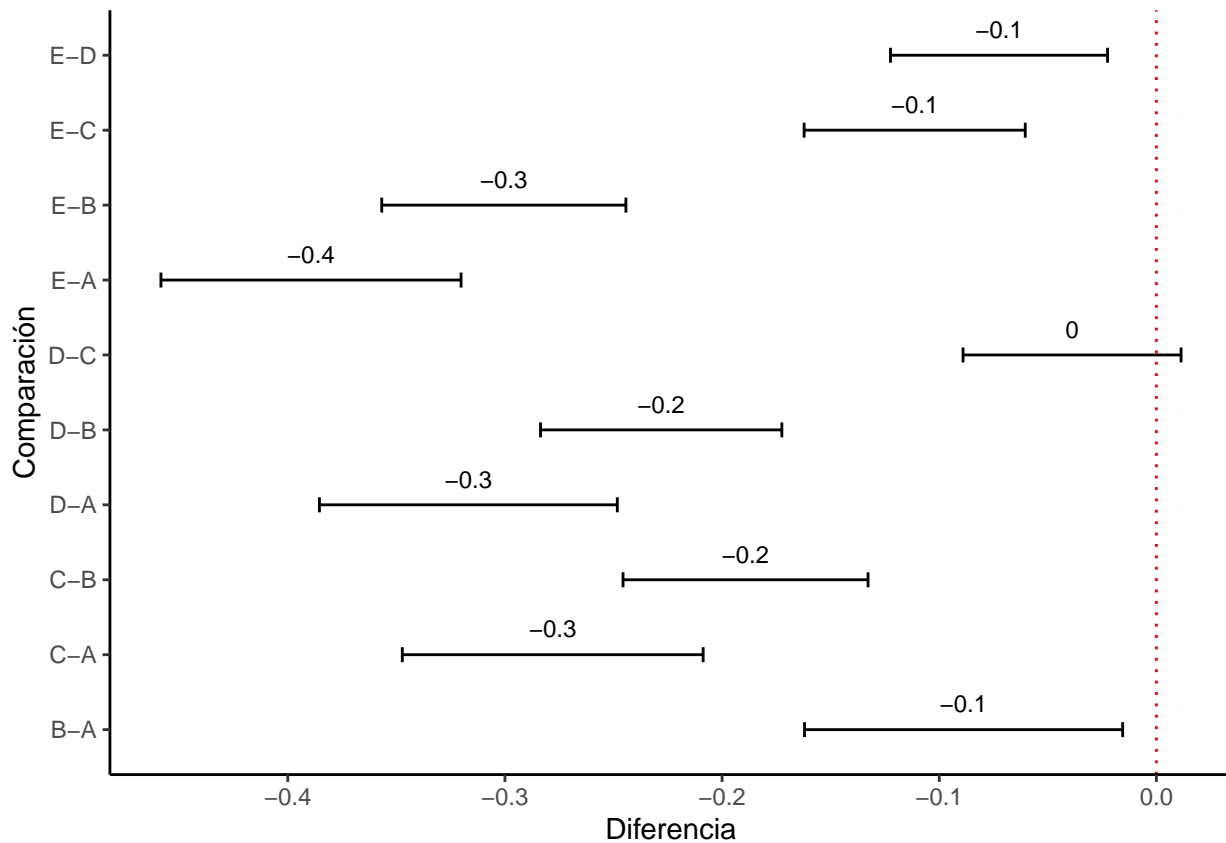
```
compara=TukeyHSD(anova)
compara
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = indicador ~ ESTRATO, data = data)
##
## $ESTRATO
##      diff      lwr      upr      p adj
## B-A -0.08879600 -0.16205033 -0.01554168 0.0083949
## C-A -0.27799918 -0.34727061 -0.20872774 0.0000000
## D-A -0.31684142 -0.38542807 -0.24825478 0.0000000
## E-A -0.38930759 -0.45841702 -0.32019817 0.0000000
## C-B -0.18920317 -0.24561917 -0.13278717 0.0000000
## D-B -0.22804542 -0.28361844 -0.17247240 0.0000000
## E-B -0.30051159 -0.35672854 -0.24429463 0.0000000
## D-C -0.03884225 -0.08904879  0.01136430 0.2154858
## E-C -0.11130842 -0.16222681 -0.06039003 0.0000000
## E-D -0.07246617 -0.12244895 -0.02248339 0.0007310
```

_ Interpretacion: al 95% de diferencia de medias, si existe una diferencia significativa entre ellas.

```
compara.df=as.data.frame(compara[1])
compara.df$compara=rownames(compara.df)
```

```
graf = ggplot(compara.df, aes(x=compara, y=ESTRATO.diff))+
  geom_errorbar(aes(ymin=ESTRATO.lwr, ymax=ESTRATO.upr), width=0.2)+
  geom_text(aes(label=paste(round(ESTRATO.diff, 1))), vjust=-1, size=3)+
  xlab("Comparación") + ylab("Diferencia")+
  coord_flip() +
  geom_hline(yintercept = 0, color = "red", linetype="dotted") +
  theme_classic()
graf
```



EJERCICIOS PARA CASA:

Ejercicio 3

¿Es cierto que existe una brecha entre el **consumo de cultura** según el sexo **sexo**? Realiza el análisis descriptivo y luego, realiza la prueba inferencial correspondiente.

Ejercicio 4

¿Es cierto que existe una brecha entre el **consumo de cultura** según la región natural **region_natural**? Realiza el análisis descriptivo y luego, realiza la prueba inferencial correspondiente.

Ejercicio 5

¿Es cierto que existe una brecha entre el **consumo de cultura** según área urbano vs rural **AREA**? Realiza el análisis descriptivo y luego, realiza la prueba inferencial correspondiente.

Ejercicio 6

Utiliza las variables que dispones en la base de datos y construye tu propio indicador de consumo cultural de 0 a 100.