

Sesión 6. Prueba T y Comparaciones de proporciones

Curso: POL278 - Estadística para el análisis político 1

2023-09-27



Recordando lo avanzado

En la sesión anterior de teoría, nos adentramos al análisis inferencial bivariado, teniendo como base del tema el cálculo del Intervalo de Confianza (IC) para una media (variables numéricas) y para una proporción (variables categóricas). Recordemos que gracias al IC podemos determinar si la estimación es representativa de la población. La idea era calcular los intervalos de confianza para cada grupo y ver si los intervalos se interceptan o no. La regla era que si los intervalos de ambos grupos no se interceptaban, podíamos extrapolar que la diferencia muestral existe en la población al 95% de confianza.

Para profundizar sobre estadística inferencial, evaluaremos las hipótesis mediante la introducción a la prueba t de diferencia de medias y desarrollaremos los cinco pasos de la inferencia estadística. Recordemos que el objetivo es corroborar que es posible extrapolar un resultado de la muestra a la población.

¿Qué es la prueba T de diferencia de medias?

Generalmente, cuando queremos comparar dos grupos centramos nuestra atención en el promedio de cada uno. Sin embargo, el hecho de que los promedios sean distintos no supone, necesariamente, que existe una diferencia estadísticamente significativa.

Para saber si la diferencia observada entre las medias de dos grupos es o no significativa se emplean métodos paramétricos como el de Z-scores o la distribución T-student. Estas técnicas calculan los intervalos de confianza de cada grupo y concluyen si existe una diferencia real entre las medias.

La prueba T compara la media de una variable numérica para dos grupos o categorías de una variable nominal u ordinal. Los grupos que forman la variable nominal/ordinal tienen que ser independientes. Es decir, cada observación debe pertenecer a un grupo o al otro, pero no a ambos.

Pruebas T para muestras independientes

Condiciones

1. Independencia: Las muestras deben ser independientes. El muestreo debe ser aleatorio.
2. Igualdad de varianza: La varianza de ambas poblaciones comparadas debe ser igual.

3. La variable numérica se distribuye de manera normal.

IMPORTANTE

La condición de normalidad también es considerada si es que la muestra fuera pequeña (Agresti y Finlay proponen que se aplica con un n pequeño menor a 30 observaciones). A medida que el n se hace más grande, el supuesto de normalidad es menos importante pues con grandes n confiamos en el teorema del límite central que nos indica que la distribución muestral será siempre normal.

Pasos para realizar la Prueba T

1. Establecer hipótesis
2. Calcular el estadístico (parámetro estimado) que se va a emplear
3. Determinar el nivel de significancia (α)
4. Calcular el p-value y realizar la prueba prop.test
5. Interpretar

Recuerda El H_0 de la prueba T es la siguiente:

H_0 : No existe diferencia estadísticamente significativa entre las medias de los dos grupos comparados. H_1 : Sí existe diferencia estadísticamente significativa entre las medias de los dos grupos comparados.

La H_0 es **generalmente** la hipótesis de no efecto, de no diferencias.

Indicador Proxy

También llamado indicador indirecto se usa ante la imposibilidad de medir lo que efectivamente es de importancia. El indicador mide una variable distinta a la que nos interesa de manera específica, pero presenta una relación lo más directa posible con el fenómeno en estudio. Un indicador proxy es una medición o señal indirecto que aproxima o representa un fenómeno en la ausencia de una medición o señal directo.

Indicador Aditivo

Pasos para construir un indicador:

1. Verificar que las variables que construyan el indicador correspondan al concepto que se desea medir.
Ejemplo: Si deseo medir Satisfacción del Usuario, las preguntas deben ser sobre ello.
2. Revisar el cuestionario e identificar el sentido de las categorías. *Ejemplo: El valor 5 es "Muy insatisfecho" y 1 "Muy satisfecho"*
3. Si las categorías de las variables están en el correcto sentido proceder a sumarlas, si no lo están, proceder a recodificarlas para luego sumar.
4. Una vez realizada la suma, identificar el mínimo y el máximo.
5. Restar a todos los valores el valor mínimo.
6. Al resultado de lo anterior, dividir por el nuevo máximo, con ello, se va a obtener valores entre 0 y 1.
7. Multiplicar por 100 si se desea el índice de 0 a 100, o por 10 si se desea el índice de 0 a 10.

Apliquemos lo aprendido

Carguemos la base de datos

Descripción del Proyecto: Satisfacción de la ciudadanía con los servicios públicos transaccionales en regiones

Este estudio fue realizado por la Secretaría de Gestión Pública de la Presidencia del Consejo de Ministros en el año 2021. El propósito del estudio consistió en identificar los conductores de calidad (variables explicativas)

en la satisfacción de una persona con la realización de un servicio público transaccional (duplicado de DNI, pago en el Banco de la Nación, pasaporte, etc).

Se identificó que los factores que impactan en las regiones respecto a la satisfacción sobre los servicios públicos son: i. el tiempo de desplazamiento hacia la sede de la entidad, ii. la calidad del trato, iii. la rapidez del trabajador, iv. el procedimiento de atención, v. el resultado de la gestión, y, vi. la confianza.

Lo que buscaremos en este ejercicio es corroborar que los grupos de edad hasta 35 años y mayor a 35 años tienen diferencias estadísticas sobre el nivel de satisfacción.

Más sobre el proyecto accediendo al siguiente enlace: <https://www.gob.pe/institucion/pcm/informes-publicaciones/2244351-estudio-en-las-regiones-del-peru-que-factores-influyen-en-la-satisfaccion-de-las-personas-con-los-servicios-publicos-brindados>

```
#No olvides cambiar el directorio de trabajo
```

```
library(rio)
data <- import("data.sav")
```

Exploramos las variables que tiene la base de datos:

Utilizamos *glimpse* para ver la estructura de la data.

Utilizamos *colnames* para verificar los nombres de la data.

Utilizamos *colSums(is.na())* para contar los nulos de la data.

```
str(data[,1:5]) #Visualice la estructura de la bbdd
```

```
## 'data.frame': 4142 obs. of 5 variables:
## $ SbjNum : num 1.52e+08 1.52e+08 1.52e+08 1.52e+08 1.52e+08 ...
## .. attr(*, "label")= chr "SbjNum"
## .. attr(*, "format.spss")= chr "F10.0"
## $ d3 : num 17 17 17 17 24 24 10 10 14 23 ...
## .. attr(*, "label")= chr "DC3d. ¿En qué departamento vives?"
## .. attr(*, "format.spss")= chr "F8.0"
## .. attr(*, "labels")= Named num [1:26] 1 2 3 4 5 6 7 8 9 10 ...
## .. ..- attr(*, "names")= chr [1:26] "Amazonas" "Áncash" "Apurímac" "Arequipa" ...
## $ PROVINCIA : num 1701 1701 1701 1701 2401 ...
## .. attr(*, "label")= chr "DC3p. ¿En qué provincia vives?"
## .. attr(*, "format.spss")= chr "F8.0"
## .. attr(*, "labels")= Named num [1:196] 101 102 103 104 105 106 107 201 202 203 ...
## .. ..- attr(*, "names")= chr [1:196] "Chachapoyas" "Bagua" "Bongará" "Condorcanqui" ...
## $ DISTRITO : num 170101 170101 170101 170101 240101 ...
## .. attr(*, "label")= chr "DC3dd. ¿En qué distrito vives?"
## .. attr(*, "format.spss")= chr "F8.0"
## .. attr(*, "labels")= Named num [1:1874] 10101 10102 10103 10104 10105 ...
## .. ..- attr(*, "names")= chr [1:1874] "Chachapoyas" "Asunción" "Balsas" "Cheto" ...
## $ ORGANIZACION: num 71 71 71 71 6 71 21 21 71 71 ...
## .. attr(*, "label")= chr "ORGANIZACION - ENTIDAD:"
## .. attr(*, "format.spss")= chr "F8.0"
## .. attr(*, "labels")= Named num [1:86] 1 2 3 4 5 6 7 8 9 10 ...
## .. ..- attr(*, "names")= chr [1:86] "Gobierno Regional de Ancash" "Gobierno Regional de La Libertad"
```

```
colnames(data[,1:5]) #Visualice los nombres de las variables de la bdd
```

```
## [1] "SbjNum"      "d3"          "PROVINCIA"   "DISTRITO"    "ORGANIZACION"
```

```
colSums(is.na(data))
```

```
##      SbjNum      d3      PROVINCIA      DISTRITO      ORGANIZACION      A
##      0      0      0      0      0      0
##      B      D      E      P1      P2      P3
##      0      0      0      0      0      0
##      P4      P5_01      P5_02      P5_03      P5_04      P6_1
##      0      1536      4040      4141      4142      2378
##      P6_2      P7      P8      P9      P10      P11
##      3400      0      0      0      0      0
##      P12      P13      P14      P15      P16A      P16B_01
##      0      0      0      0      2537      2538
##      P16B_02      P16B_03      P16B_04      P16B_05      P16B_06      P16B_07
##      4071      4136      4140      4142      4142      4142
##      P16B_08      P16B_09      P17      P18      P19      P20
##      4142      4142      2536      0      993      993
##      P21      P22      P23      P24_1      P24_2      P24_3
##      0      0      0      0      0      0
##      P24_4      P25      P26      P27      P28      P29
##      0      0      0      0      0      2259
##      P30      P31      P32      P33      P34      P35
##      0      0      0      0      3034      0
##      P36      P37      P38      P39      P40      P41
##      1176      3327      3368      3451      3451      3451
##      P42      P43      I_1_P44      I_2_P44      I_3_P44      I_4_P44
##      0      0      625      625      625      625
##      I_5_P44      P45      P46      P47      P48      P49
##      625      0      1      0      0      0
##      P50      P51      P52      P53      I_1_P54      I_2_P54
##      0      0      0      0      0      0
##      I_3_P54      I_4_P54      I_1_P55      I_2_P55      I_1_P56      I_2_P56
##      0      0      0      0      0      0
##      P57      P58      P59      NSE      REGX      ent
##      0      0      0      0      0      0
##      SEX0      edad
##      0      0
```

Limpieza de las variables previo al análisis: a. Seleccionar variables que conceptualmente generen un índice de satisfacción. Revisar preguntas en encuesta.

Según el cuestionario, 5 significa *muy insatisfecho* y 1 significa *muy satisfecho*.

- P10 = satisfacción con trabajador que lo atendió
- P25 = satisfacción con tiempo de espera desde que llegó a la entidad hasta ser atendido
- P30 = satisfacción con el proceso de gestión / trámite realizado (cantidad de documentos)
- P31 = satisfacción con la cantidad de pasos requeridos para completar gestión / trámite

Si deseamos crear un indicador de satisfacción, entonces el máximo valor debe ser la calificación más alta de satisfacción, y por tanto el valor mínimo de mostrar la insatisfacción:

- 1 'Muy satisfecho' a 5
- 2 a 4
- 3 a 3
- 4 a 2
- 5 'Muy insatisfecho' a 1

Ojo: Si bien estas variables deberían estar catalogadas como factor, para poder crear el índice necesitamos que se mantengan como numéricas para poder sumarlas.

```
summary(data$P10)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000   2.000   2.000   2.145   2.000   5.000
```

```
summary(data$P25)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000   2.000   2.000   2.437   3.000   5.000
```

```
summary(data$P30)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000   2.000   2.000   2.344   3.000   5.000
```

```
summary(data$P31)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000   2.000   2.000   2.415   3.000   5.000
```

b. Recodificar los valores de las variables

Ejemplo: La recodificación de una variable a la vez.

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      intersect, setdiff, setequal, union
```

```
data=data %>%
mutate(satisfaccion_trabajador=case_when(
  P10 == 1 ~ "5",
  P10 == 2 ~ "4",
  P10 == 3~ "3",
  P10 == 4 ~ "2",
  P10 == 5 ~"1"),
satisfaccion_tiempo=case_when(
  P25 == 1 ~ "5",
  P25 == 2 ~ "4",
  P25 == 3~ "3",
  P25 == 4 ~ "2",
  P25 == 5 ~"1"),
satisfaccion_n_documentos=case_when(
  P30 == 1 ~ "5",
  P30 == 2 ~ "4",
  P30 == 3~ "3",
  P30 == 4 ~ "2",
  P30 == 5 ~"1"),
satisfaccion_n_pasos=case_when(
  P31 == 1 ~ "5",
  P31 == 2 ~ "4",
  P31 == 3~ "3",
  P31 == 4 ~ "2",
  P31 == 5 ~"1"))
```

Verifiquemos que nuestra recodificación se realizó de manera correcta.

```
table(data$P10)
```

```
##
##      1      2      3      4      5
## 619 2724  415  349   35
```

```
table(data$satisfaccion_trabajador)
```

```
##
##      1      2      3      4      5
##   35  349  415 2724  619
```

La recodificación de un conjunto de variables.

```
#library(dplyr)
#data=data %>%
#   mutate(across(c(P10,P25, P30,P31),
#               ~ case_when(. == 1 ~ "5",
#                           . == 2 ~ "4",
#                           . == 3~ "3",
#                           . == 4 ~ "2",
#                           . == 5 ~"1")))
```

c. Convertir las variables a numéricas

```
library(dplyr)
data=data %>% # objeto base de datos cargada
  mutate(satisfaccion_trabajador = as.numeric(satisfaccion_trabajador), #convertimos nuestras nuevas v
         satisfaccion_tiempo = as.numeric(satisfaccion_tiempo),
         satisfaccion_n_documentos = as.numeric(satisfaccion_n_documentos),
         satisfaccion_n_pasos = as.numeric(satisfaccion_n_pasos))
```

```
summary(data$satisfaccion_trabajador)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000   4.000   4.000   3.855   4.000   5.000
```

```
summary(data$satisfaccion_tiempo)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000   3.000   4.000   3.563   4.000   5.000
```

```
summary(data$satisfaccion_n_documentos)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000   3.000   4.000   3.656   4.000   5.000
```

```
summary(data$satisfaccion_n_pasos)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000   3.000   4.000   3.585   4.000   5.000
```

d. Recordemos los pasos para crear un índice aditivo:

$((\text{var_suma} - \text{mín_de_suma}) / \text{máx_de_suma}) \times \text{valor al que quiere que llegue el índice (si va del 0 al 10 será 10, del 0 al 50 será 50, etc.)}$.

Crearemos una variable nueva “índice_satisfacción”. Sumamos las variables

```
data=data %>%
  mutate(suma = satisfaccion_trabajador +
         satisfaccion_tiempo +
         satisfaccion_n_documentos +
         satisfaccion_n_pasos)
```

Revisamos mínimo y máximo

```
summary(data$suma)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      4.00   13.00   16.00   14.66   16.00   20.00
```

```
data=data %>%
  mutate(resta = ((suma - 4) # Menos el mínimo
                  / 20), # Cuarto paso: dividir entre el nuevo máximo
         indice_satisfaccion = resta * 100) # Queremos que el índice va del 1 al 100
```

Opción 2: Sin crear variable “resta”

```
#data = data %>%  
# mutate(indice_satisfaccion = ((suma-4)/20)*100)
```

```
summary(data$indice_satisfaccion)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##       0.0   45.0   60.0   53.3   60.0   80.0
```

e. Los grupos que compararemos serán dados por la variable P4. Damos formato a la variable categórica P4, si recibio orientación o no

```
data$P4=as.factor(data$P4)  
data$P4=factor(data$P4,  
               levels = levels(data$P4),  
               labels = c("Si orientación", "No orientación"),  
               ordered = F)
```

Realizamos la Prueba T

Primer paso: Establecer la hipótesis.

La hipotesis de la prueba T queda establecida de la siguiente forma:

- Ho: No hay diferencia entre las medias del índice de satisfacción aditivo entre los grupos que sí recibieron orientación y los que no recibieron información. (**no diferencia de medias**)
- H1: Si existen diferencias entre las medias del índice de satisfacción aditivo entre los grupos que sí recibieron orientación y los que no recibieron información. (**sí diferencia de medias**)

Ambas hipótesis son acerca de los parámetros de la población.

Segundo paso: Calcular el estadístico a emplear

Para verificar la diferencia de medias se calcula el estadístico T, y uno de los primeros pasos es calcular las diferencias entre las medias muestrales, ya que es lo quiero extrapolar y por tanto saber si existe o no una diferencia significativa entre las medias poblacionales de ambos grupos:

Tercer paso: Determinar el nivel de significancia

Se trata de la probabilidad que define qué tan inusual debe ser la diferencia de medias muestrales para rechazar la H0 (que la diferencia de medias poblacionales sea 0). El valor más común es de $\alpha = 0.05$ a un nivel de confianza de 95%. De manera convencional establecemos la siguiente regla para nuestra prueba T:

- p-value ≥ 0.05 Rechazo la H0 y acepto H1
- p-value < 0.05 No rechazo la H0

Cuarto paso: Calcular el p-value y realizar la prueba t.test

El p-value mide la probabilidad de observar en una muestra una diferencia de medias como la observada, si la diferencia de medias poblacional fuera cero.

```
t.test(indice_satisfaccion ~ P4, data = data,
       alternative = "two.sided",
       conf.level = 0.95 #nivel de confianza (95%)
)

##
## Welch Two Sample t-test
##
## data: indice_satisfaccion by P4
## t = 12.521, df = 2706.3, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group Si orientación and group No orientación
## 95 percent confidence interval:
##  4.911234 6.735047
## sample estimates:
## mean in group Si orientación mean in group No orientación
##                55.45856                49.63542
```

Quinto paso: Interpretar

¿Cómo interpreto?

Recordando nuestras hipótesis:

- H_0 : No hay diferencia entre las medias del índice de satisfacción aditivo entre los grupos que sí recibieron orientación y los que no recibieron información.
- H_1 : Si existen diferencias entre las medias del índice de satisfacción aditivo entre los grupos que sí recibieron orientación y los que no recibieron información.

Asimismo, en el paso 4, determinamos el nivel de significancia de la siguiente manera:

- Si el p-value del t test es ≤ 0.05 Rechazo la H_0 y se afirma H_1 .
- Si el p-value del t test es > 0.05 No rechazo la H_0

Entonces, vemos que el p-value es $2.2e-16$, y es menor al α (0.05), entonces rechazo la H_0 , por tanto, existe una diferencia estadísticamente significativa entre las medias del índice de satisfacción aditivo entre los grupos que sí recibieron orientación y los que no recibieron información con un 95% de confianza.

Paso FINAL: Graficar

Otro método para evaluar la comparación entre grupos es realizar un gráfico de medias con intervalos de confianza de cada grupo.

Para calcular la diferencia de medias

```
library(Rmisc)
ic_grupo = group.CI(indice_satisfaccion~P4,data)
ic_grupo
```

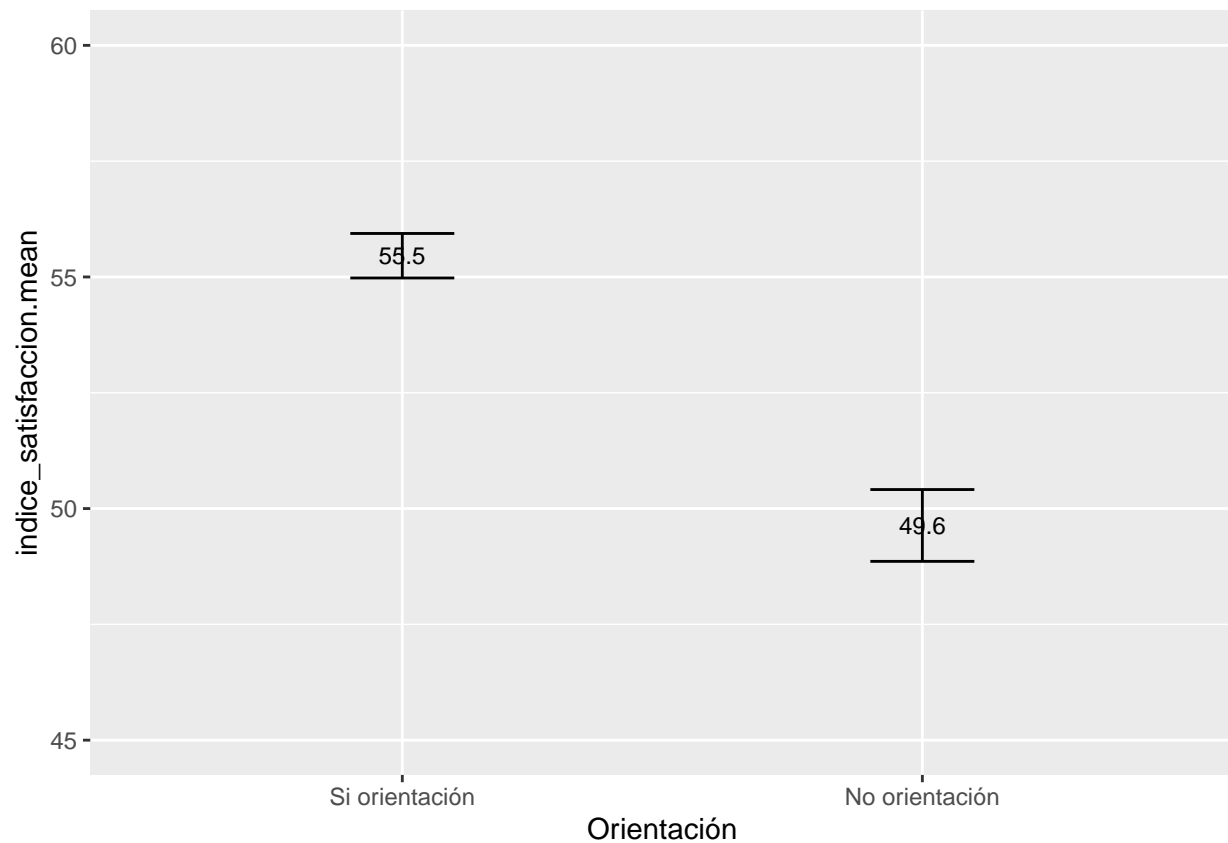
```
##           P4 indice_satisfaccion.upper indice_satisfaccion.mean
## 1 Si orientación          55.93908          55.45856
## 2 No orientación          50.41071          49.63542
##  indice_satisfaccion.lower
## 1          54.97803
## 2          48.86012
```

```
library(ggplot2)
```

Barras de Error

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
ggplot(ic_grupo, aes(x= P4, y =indice_satisfaccion.mean)) +
  geom_errorbar(aes(ymin=indice_satisfaccion.lower, ymax=indice_satisfaccion.upper), width = 0.2)+
  geom_text(aes(label=paste(round(indice_satisfaccion.mean,1))), size=3)+
  xlab("Orientación") +
  ylim(45, 60)
```



Interpretación: Tal como se observa ambos intervalos de confianza no se traslapan, por lo que se puede concluir gráficamente que existe una diferencia estadísticamente significativa entre los grupos. El grupo que si recibio orientación tiene mayor satisfacción que no recibio con un 95% de confianza en la población.

EJERCICIO PRÁCTICO

Verifica si existe o no relación entre el número de veces que un ciudadano fue a la institución pública para realizar un trámite transaccional (grupo 1 sola vez vs grupo más de 1 vez) y el índice de satisfacción.

Comparación de proporciones

Para este ejercicio trabajaremos con dos variables:

- *P45: Confianza en la entidad pública*
- *P42: Preferencia realiza sus gestiones con entidades públicas*

Revisemos a nuestras variables

Variable V20:

```
class(data$P45) #Revisamos como está catalogada nuestra variable
```

```
## [1] "numeric"
```

```
summary(data$P45)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000   2.000   2.000   2.315   3.000   5.000
```

Recordemos que para comparar proporciones necesitamos que nuestra variable sea categórica. La recodificaremos para que tengamos dos grupos: Confianza alta y Confianza baja.

```
data = data %>%
  mutate(INDICE_AGRUPADO = case_when(indice_satisfaccion >= 60.00 ~ "Satisfecho",
                                     indice_satisfaccion < 60.00 ~ "No satisfecho"))
```

Realizamos el mismo ejercicio con nuestra variable P42:

```
class(data$P42) #Revisamos como está catalogada nuestra variable
```

```
## [1] "numeric"
```

```
summary(data$P42)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000   1.000   1.000   1.629   2.000   4.000
```

La recodificaremos para que tengamos dos grupos: Presencial y No presencial

```
data = data %>%
  mutate(P42_2 = case_when(P42 == 1 ~ "Presencial",
                           TRUE ~ "No presencial"))
```

```
#Realizamos una tabla de frecuencias
table(data$INDICE_AGRUPADO,data$P42_2)
```

```
##
##               No presencial Presencial
## No satisfecho      646      1165
## Satisfecho        677      1654
```

Identificamos lo que nos interesa: La frecuencia de los que tienen satisfacción alta y prefieren realizar gestiones presenciales es 1654; mientras que, los que tienen satisfacción alta y prefieren realizar gestiones no presenciales es de 677.

```
#Hallamos la proporción
prop.test(x=c(1654,677),n=c(1654+1165,677+646))
```

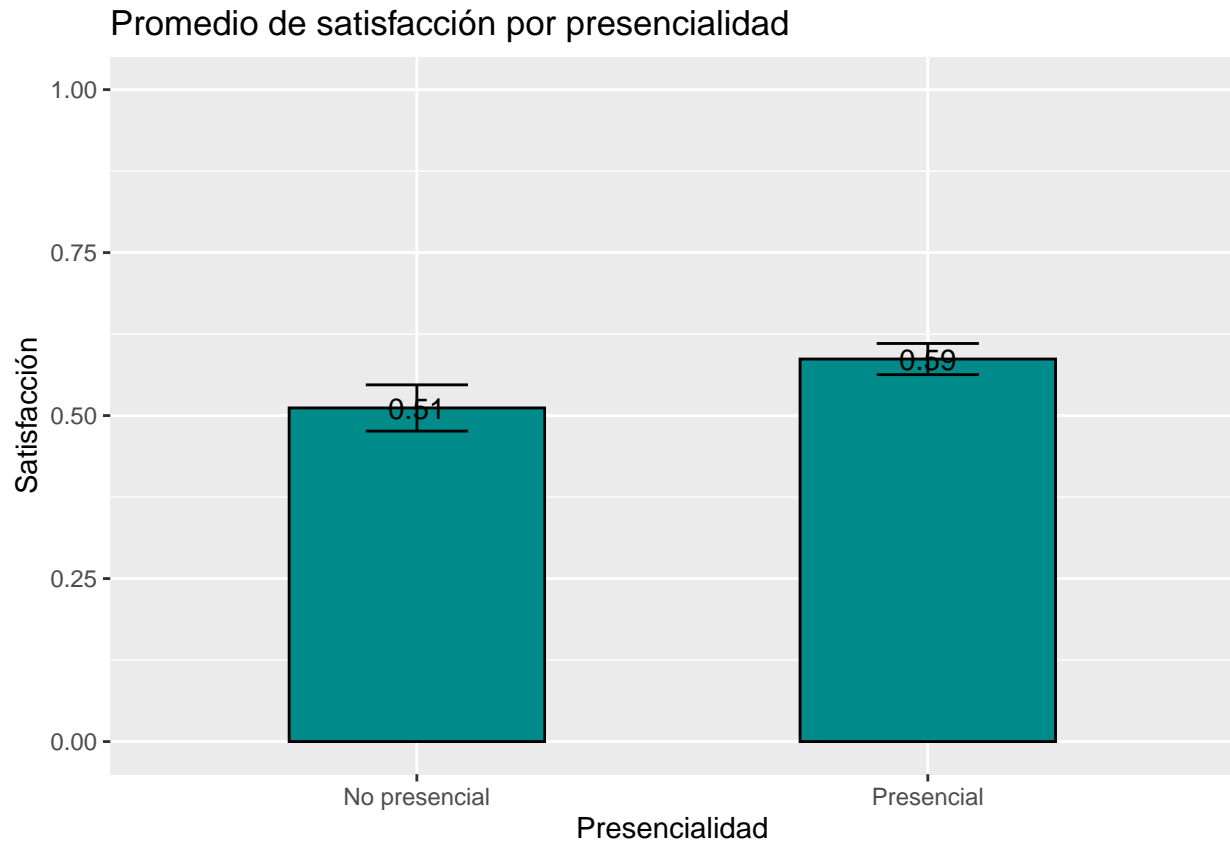
```
##
## 2-sample test for equality of proportions with continuity correction
##
## data:  c(1654, 677) out of c(1654 + 1165, 677 + 646)
## X-squared = 20.289, df = 1, p-value = 6.657e-06
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.04196679 0.10806739
## sample estimates:
##   prop 1    prop 2
## 0.5867329 0.5117158
```

Interpretación: la diferencia entre aquellas personas que tienen satisfacción alta y preferencia por gestión presencial vs personas que tienen satisfacción alta y preferencia por gestión no presencial, se encuentra entre 4% y 10%, a un 95% de confianza.

```
# Grafiquemos la comparación de proporciones
library(fastDummies)
data_grafico =dummy_cols(data,select_columns=c("INDICE_AGRUPADO"))
library(lsr)
tabla2=data_grafico%>%
  dplyr::group_by(P42_2)%>%
  dplyr::summarise(Media = mean(INDICE_AGRUPADO_Satisfecho, na.rm=T),
                   limiteinferior = ciMean(INDICE_AGRUPADO_Satisfecho,conf = 0.99, na.rm=T)[1],
                   limitesuperior = ciMean(INDICE_AGRUPADO_Satisfecho,conf = 0.99, na.rm=T)[2]
  )
tabla2
```

```
## # A tibble: 2 x 4
##   P42_2      Media limiteinferior limitesuperior
##   <chr>      <dbl>          <dbl>          <dbl>
## 1 No presencial 0.512            0.476            0.547
## 2 Presencial   0.587            0.563            0.611
```

```
# Grafiquemos la comparación de proporciones
tabla2=as.data.frame(tabla2)
library(ggplot2)
ggplot(tabla2, aes(x=P42_2, y = Media))+
  geom_bar(width=0.5, fill="darkcyan", colour="black", stat="identity")+
  geom_errorbar(aes(ymin=limiteinferior, ymax=limitesuperior), width=0.2)+
  geom_text(aes(label=paste(round(Media, 2))), size = 4)+
  xlab("Presencialidad") + ylab("Satisfacción")+
  ggtitle("Promedio de satisfacción por presencialidad")+ylim(0,1)
```



Ejercicios

- Recodifica variables en 2 categorías y compara sus proporciones con sentido de investigación.