

Tendencias en los clientes de casinos en Lima

Francesco Uccelli Meneses - 201810253, Alejandro Buitrago Zegarra - 201820079,
Carlos Mazuelos Selem - 201810195, Santiago Madariaga Collado - 201810677

Barranco, 2019

UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA

Profesor: Renom Andara, Jose Miguel

Curso: Estadística y Probabilidades

Introducción

Las estadísticas y probabilidades se pueden encontrar en casi todos los lugares y situaciones, sin embargo existen lugares donde estas te pueden ayudar a ganar mucho dinero o hacer que lo pierdas todo, los casinos. Los casinos, sin lugar a duda, son uno de lugares más tentativos en los que se puede gastar muchísimo dinero, a pesar de todo los clientes van siempre con esperanza de poder ganar. El poker, la ruleta, los dados, las máquinas tragamonedas y muchos juegos más del casino son en los cuales las personas pierden más dinero. Pero, ¿las personas realmente saben cuales son sus probabilidades de salir victoriosos? Debido a estas cuestiones decidimos realizar un trabajo de investigación de campo y análisis de los datos recolectados de este mismo. A continuación se podrá ver a más detalle los objetivos planteados en esta investigación, datos sobre la población objetivo y la muestra a tomar. Además se podrá ver específicamente el método de recolección de datos y las preguntas que se les hará a la muestra seleccionada.

Datos de la investigación a realizar

Los principales objetivos de esta investigación son dos. En primer lugar, conocer más sobre los perfiles de los clientes de los casinos y cómo son sus resultados normalmente. En segundo lugar, ver si realmente los clientes del casino saben cuales son sus probabilidades de salir victoriosos y con qué monto de dinero se conforman.

Para la realización de este estudio nos interesan los jugadores de los casinos, ya sea dentro de este o fuera, como en la universidad o calle. Debido a esto, el tipo de muestreo en esta investigación será de tipo conglomerado. De tal forma que se calcula que la unidad muestral sea de aproximadamente 15 personas por casino, siendo el número de casinos que se visitan, la unidad muestras será de 60 personas, esto irá de la mano con las personas que podamos entrevistar fuera del casino para llegar a un total de 100 muestras.

La recolección de data para la investigación será realizada por nosotros mismos. Durante los tres próximos fines de semana iremos a cuatro casinos distintos y entrevistamos a las personas que se encuentran jugando en el casino. Nosotros también nos acercaremos a personas aleatorias para que nos llenen la encuesta. Para la recolección de datos, hemos propuesto las siguientes preguntas, las cuales creemos que nos pueden ayudar a llegar a un análisis de la situación.

La planificación de la investigación a realizar durante el ciclo académico, puede ser observada en el siguiente diagrama de Gantt

Planificación sobre recolección de datos

Nuestro plan consta básicamente en dos muestras, la primera será de personas que se encuentren dentro de diferentes casinos y que estén realizando apuestas en diferentes juegos o máquinas. La siguiente sería de personas aleatorias tanto alumnos de la universidad como gente aleatoria en diversos lugares.

En la muestra, que es la de personas que se encuentran realizando apuestas dentro del casino, tenemos pensado generar un Google form con las preguntas. De esta manera, podremos entregarles un celular o dispositivo electrónico para que rápidamente llenen la encuesta sin hacerles perder mucho tiempo. Además esto nos garantiza una recolección de datos ordenada y fácil de analizar. Este método nos servirá también para entrevistar a las personas aleatorias.

Recolección de datos

La recolección se llevó a cabo mayoritariamente de forma virtual haciendo uso de redes sociales. Luego, se procedió a hacer la limpieza de datos muestra por muestra debido a que muchas respuestas estaban en diferentes unidades o respuestas incoherentes. Con todos los datos limpios, armamos un archivo csv para luego hacer unas gráficas que nos ayudarían a llegar a ciertas conclusiones relacionando variables. A continuación estas:

1. Tema delimitado

Características de los Jugadores de Casinos en Lima.

2. Problemática

Los casinos han sido por mucho tiempo pasatiempos destinados a un determinado nicho de la población, sin notorios cambios en su logística y atracciones, volviéndose poco atractivo a las nuevas generaciones (en comparación con las anteriores). Los casinos buscan tener un margen de ganancias óptimo, el cual no sea lo suficientemente grande como para que las personas no deseen jugar ya que nunca experimentaron la sensación de ganar. Sin embargo, varias personas sienten la necesidad de jugar y jugar (sin importar la cantidad de dinero que pierdan), siendo un

transtorno psicológico llamado ludopatía. Éste transtorno es muchas veces subestimado, pero es un problema para las personas que lo padecen. Con nuestro trabajo de investigación buscamos encontrar relaciones entre las diferentes características de los jugadores, para así detectar ciertas tendencias e identificarlas.

3. Justificación

Decidimos analizar el tema propuesto y abarcar dicha problemática debido a que consideramos que no existe información clara y reciente sobre este tema, con nuestra investigación lograremos identificar patrones que servirán de guía para futuras investigaciones y para ayudar a las empresas a estar al tanto de las personas asistentes a estos centros de entretenimiento.

4. Objetivos

Objetivos generales

- Describir las edades más comunes de personas asistentes al casino
- Describir el presupuesto que los jugadores encuestados están acostumbrados a tener
- Explorar las diferencias que existen entre un jugador de edad avanzada a uno joven
- Encontrar el juego preferido y describir porque es este
- Determinar la existencia del balance entre el presupuesto y el valor esperado de la ganancia

Objetivos específicos

- Descubrir si existe una relación notable entre la frecuencia de asistencia al casino con el valor esperado de la ganancia
- Encontrar semejanzas entre jugadores asistentes al mismo casino
- Descubrir si la mayoría de jugadores estan enterados de las probabilidades reales de ganar en los juegos que ellos frecuentan
- Validar si existe un patrón de ludopatía en nuestra muestra.

5. Definición de variables

Número	Relación con pregunta	Pregunta	Tipo de variable
1	7 , 17	Cuántos años tiene?	Numérica discreta
2	3	Sexo	Categórica Nominal
3	6	A qué edad vino por primera vez al casino?	Numérica discreta
4	11 ,12 ,16	En promedio, cuantas veces visita el casino al mes?	Numérica discreta
5	8, 14	En promedio, con cuánto dinero entra al casino?	Numérica discreta
6	5	La mayoría de veces gana o pierde o ninguno?	Categórica Nominal
7	4	Cual es el juego que suele jugar con mayor frecuencia?	Categórica Nominal
8	2,	Sabe cuales son las probabilidades de que gane en ese juego?	Numéricas continua
9	13, 18	Con cuánto dinero ganado se siente satisfecho?	Numérica discreta
10	7 , 10 , 15	Cuanto dinero es aceptable perder para usted?	Numérica discreta
11	9	Qué grado de satisfacción le da ir al casino? Fuera de su resultado	Categórica Ordinal
12	1 , 16	En que casino se encuentra?	Categórica Nominal

6. Preguntas de investigación:

Preguntas descriptivas:

- 1 ¿Cuál es el casino más visitado?
- 2 ¿Los integrantes de la muestra tienen idea de sus reales posibilidades de ganar?
- 3 ¿Quiénes tienen mayor interés en el casino (hombres o mujeres)?
- 4 ¿Cuál es el juego preferido?
- 5 ¿Cuál es el resultado más frecuente (ganar o perder)?
- 6 ¿A qué edad en promedio se empezó a ir al casino?

Preguntas exploratorias:

- 7 ¿Existe relación entre el presupuesto y la edad?
- 8 ¿La ganancia esperada tiene relación con el presupuesto?
- 9 ¿Existe relación entre la satisfacción obtenida al jugar y el resultado general (ganar o perder frecuentemente) ?
- 10 ¿Qué diferencias hay en pérdida aceptada y

Preguntas inferenciales:

11 ¿La frecuencia con la que una persona asiste al casino influye sobre su aceptación de pérdida?

12 ¿Las personas que asisten más de 5 veces al mes creen que ganan siempre?

Preguntas predictivas:

13 ¿Se puede conocer la satisfacción de un cliente sabiendo la frecuencia mensual?

14 ¿Es posible predecir la ganancia esperada de un jugador sabiendo su nivel de satisfacción al jugar?

Preguntas mecánicas

15 ¿Cómo afecta a la tolerancia a perder el conocimiento sobre las probabilidades de ganar?

16 ¿De qué manera determina la edad la regularidad de los clientes?

Preguntas casuales

17 ¿Qué factores influyen en el juego preferido de los clientes?

18 ¿Qué factores determinan la ganancia máxima?

```
library("graphics")

library("grid")
library("ggplot2") # beautiful graphs

library("ggthemes") # nice themes for ggplot2

library("ggbeswarm") # "beeswarm" plots

library("cowplot") # arrange graphs

##
## *****

## Note: As of version 1.0.0, cowplot does not change the
##   default ggplot2 theme anymore. To recover the previous
##   behavior, execute:
##   theme_set(theme_cowplot())

## *****

##
## Attaching package: 'cowplot'
```

```

## The following object is masked from 'package:ggthemes':
##
##     theme_map

library("pander") # nice tables

library("psych") # nice table of descriptive statistics

##
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha

library("fdth") # carga la extensión fdth (para el calculo de
distribución de frecuencias.

##
## Attaching package: 'fdth'

## The following objects are masked from 'package:stats':
##
##     sd, var

DF<-readr::read_csv("Casino.csv")

## Parsed with column specification:
## cols(
##   Sexo = col_character(),
##   Edad = col_double(),
##   `Frecuencia mensual` = col_double(),
##   `Edad de primera vez` = col_double(),
##   `Juego preferido` = col_character(),
##   Resultado = col_character(),
##   Presupuesto = col_double(),
##   `Probabilidad de Ganar` = col_character(),
##   `Ganancia maxima` = col_double(),
##   `Perdida Maxima` = col_double(),
##   Satisfaccion = col_double(),
##   Casino = col_character()
## )

DF$`Ganancia maxima`<-round(DF$`Ganancia maxima`)
DF$`Perdida Maxima`<-round(DF$`Perdida Maxima`)
DF$Presupuesto<-round(DF$Presupuesto)

```

Respuestas a las preguntas

Preguntas descriptivas:

- ¿Cuál es el casino más visitado?

```
###<
```

```
a<-table(DF$Casino)
```

- ¿Los integrantes de la muestra tienen idea de sus reales posibilidades de ganar?

```
na.omit(DF)
```

```
## # A tibble: 73 x 12
```

```
##   Sexo   Edad `Frecuencia men~` `Edad de primer~` `Juego preferid~`  
Resultado
```

```
##   <chr> <dbl>           <dbl>           <dbl> <chr>  
<chr>
```

```
## 1 Mujer    23                1                18 ruleta
```

```
Ninguna
```

```
## 2 Homb~    18                5                18 ruleta
```

```
Pierde
```

```
## 3 Homb~    18                5                18 blackjack
```

```
Pierde
```

```
## 4 Homb~    18                7                18 ruleta
```

```
Pierde
```

```
## 5 Mujer    18                1                18 ruleta          Gana
```

```
## 6 Mujer    18                3                18 pocker
```

```
Pierde
```

```
## 7 Homb~    18                3                18 maquina tragamo~
```

```
Pierde
```

```
## 8 Mujer    18                1                18 maquina tragamo~
```

```
Pierde
```

```
## 9 Homb~    18                0                18 blackjack
```

```
Pierde
```

```
## 10 Homb~   19                3                18 pocker          Gana
```

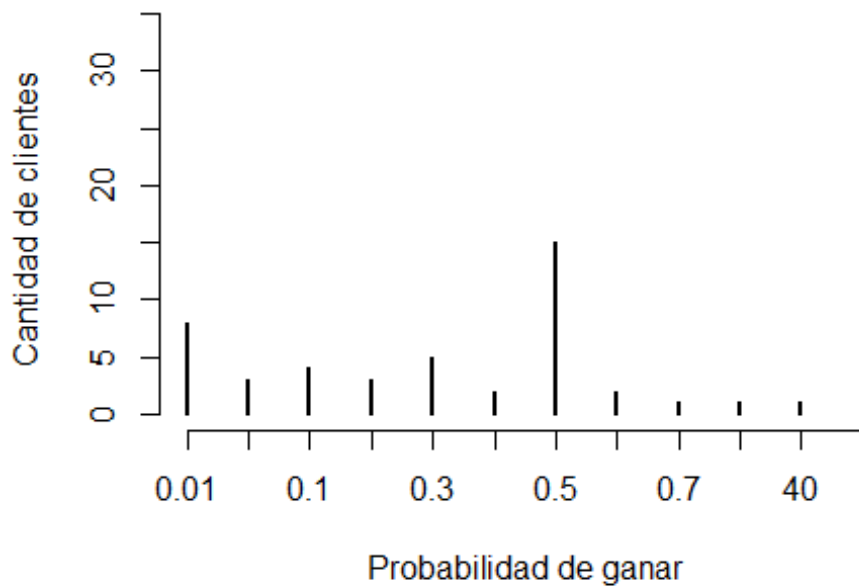
```
## # ... with 63 more rows, and 6 more variables: Presupuesto <dbl>,
```

```
## #   `Probabilidad de Ganar` <chr>, `Ganancia maxima` <dbl>, `Perdida
```

```
## #   Maxima` <dbl>, Satisfaccion <dbl>, Casino <chr>
```

```
al<-table(DF$`Probabilidad de Ganar`)
```

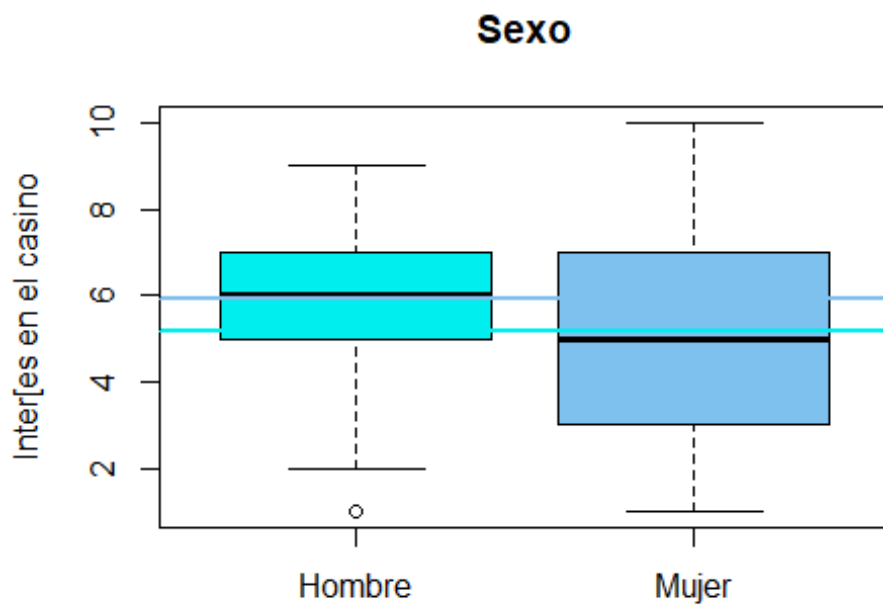
```
plot(al, xlab = "Probabilidad de ganar", ylab="Cantidad de clientes")
```



¿Quiénes tienen mayor interés en el casino (hombres o mujeres)?

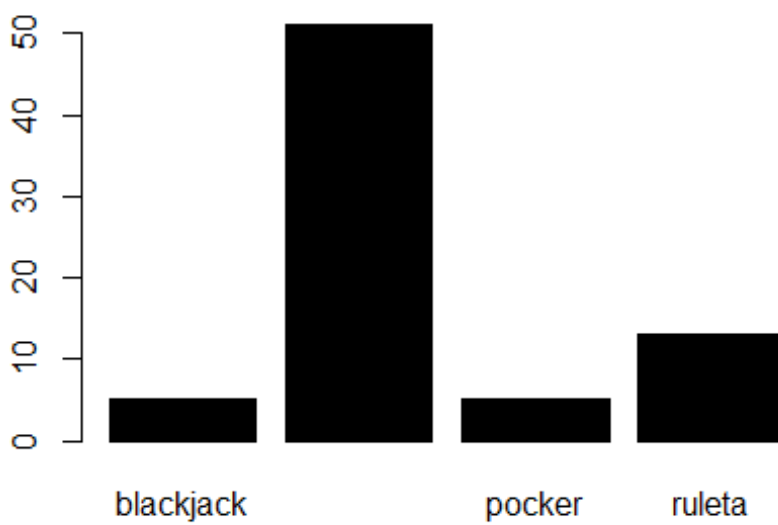
```
sas_hombres<-DF[which(DF$Sexo=="Hombre"),]$Satisfaccion
sas_mujeres<-DF[which(DF$Sexo=="Mujer"),]$Satisfaccion

boxplot(sas_hombres,sas_mujeres,col=c("cyan2","skyblue2"),main="Sexo",ylab="Inter[és en el casino", names=c("Hombre","Mujer"))
abline(h=mean(sas_hombres),col="skyblue2",lwd=2)
abline(h=mean(sas_mujeres),col="cyan2",lwd=2)
```

¿Cuál es el juego más jugado?

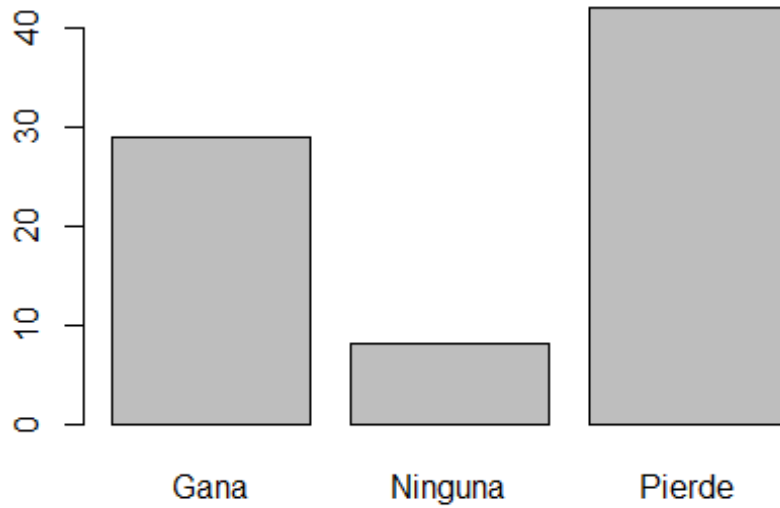
```
a<-table(DF$`Juego preferido`)  
barplot(a, col = "1")
```



```
##
```

¿Cuál es el resultado más frecuente (ganar o perder)?

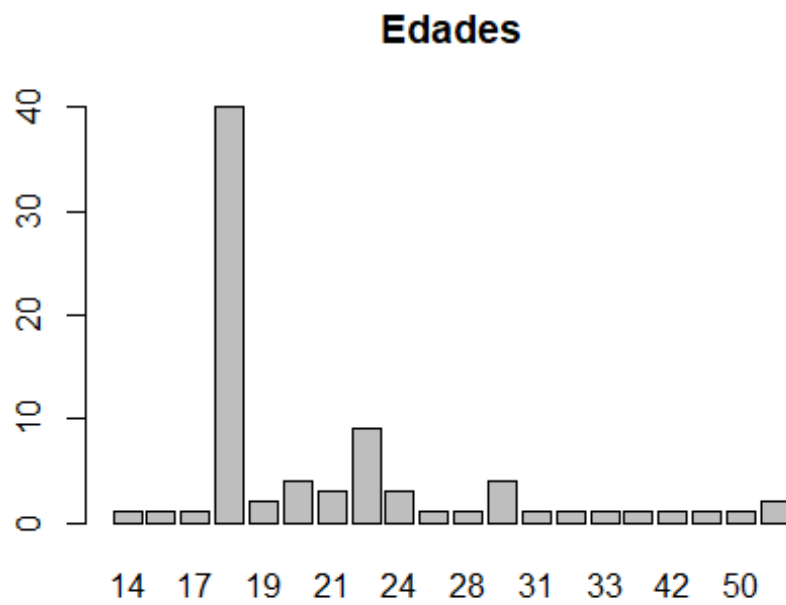
```
a<-table(DF$Resultado)  
barplot(a)
```



```
###
```

¿A qué edad en promedio se empezó a ir al casino?

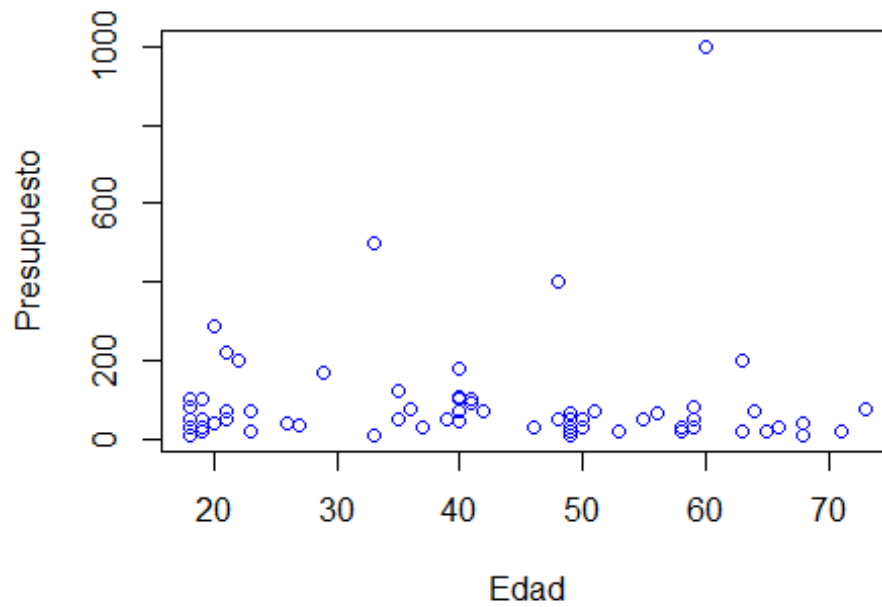
```
barplot(table(DF$`Edad de primera vez`), main = "Edades")
```



Respuesta a preguntas exploratorias:

¿Existe una relacion entre presupuesto y edad?

```
plot(DF$Edad,DF$Presupuesto,xlab="Edad",ylab="Presupuesto",col="blue")
```



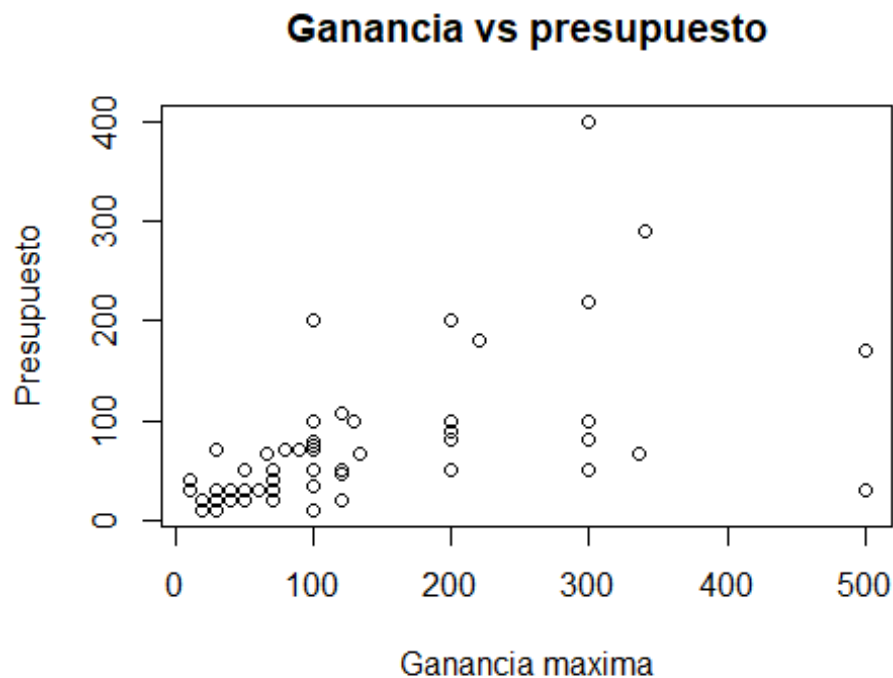
¿La ganancia esperada tiene relación con el presupuesto?

```
table(DF$Ganancia maxima, DF$Presupuesto)
```

```
##
##      10 20 30 35 40 46 50 67 70 75 78 80 90 100 108 125 170 180 200
220
## 10    0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0
## 20    1  2  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0
## 30    2  3  1  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0
0
## 40    0  1  3  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0
## 50    0  3  5  0  0  0  6  0  0  0  0  0  0  0  0  0  0  0  0
0
## 60    0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0
## 67    0  0  0  0  0  0  0  2  0  0  0  0  0  0  0  0  0  0  0
0
## 70    0  1  1  0  2  0  3  0  0  0  0  0  0  0  0  0  0  0  0
0
## 80    0  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0
0
## 90    0  0  0  0  0  0  0  0  2  0  0  0  0  0  0  0  0  0  0
0
```

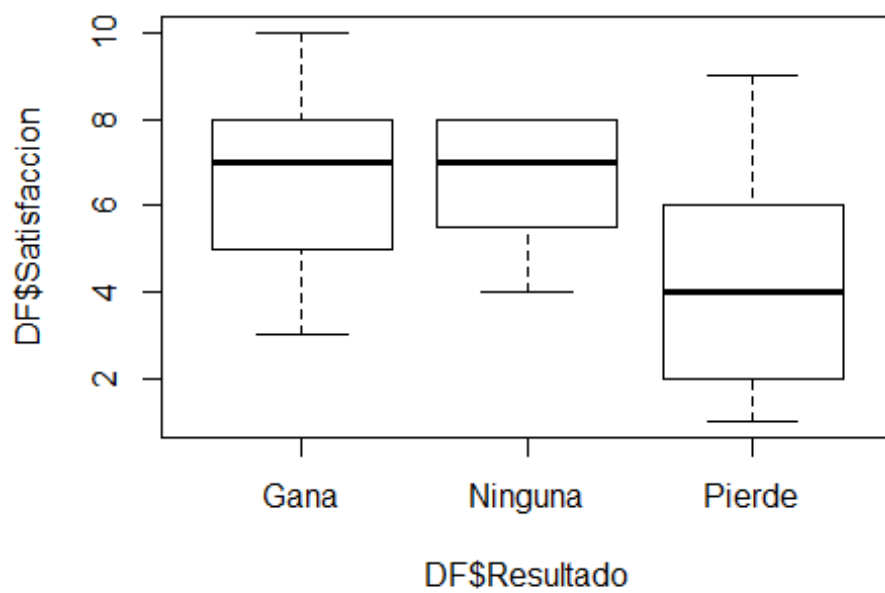


```
plot(DF[which(DF$`Ganancia maxima`<800),]$`Ganancia maxima`,
DF[which(DF$`Ganancia maxima`<800),]$Presupuesto,xlab="Ganancia
maxima",ylab="Presupuesto", main="Ganancia vs presupuesto")
```

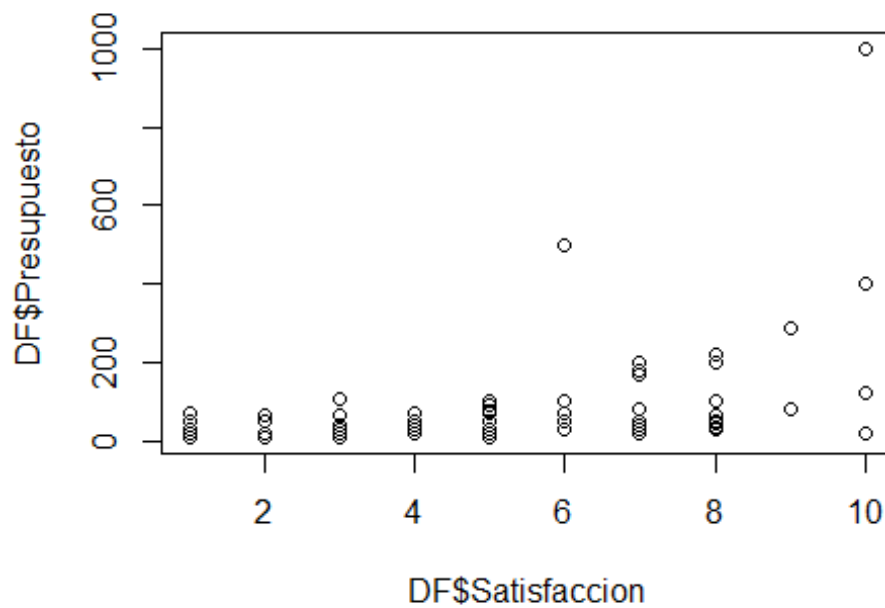


- ¿Existe relación entre la satisfacción obtenida al jugar y el resultado general (ganar o perder frecuentemente) ?

```
boxplot(DF$Satisfaccion~DF$Resultado)
```

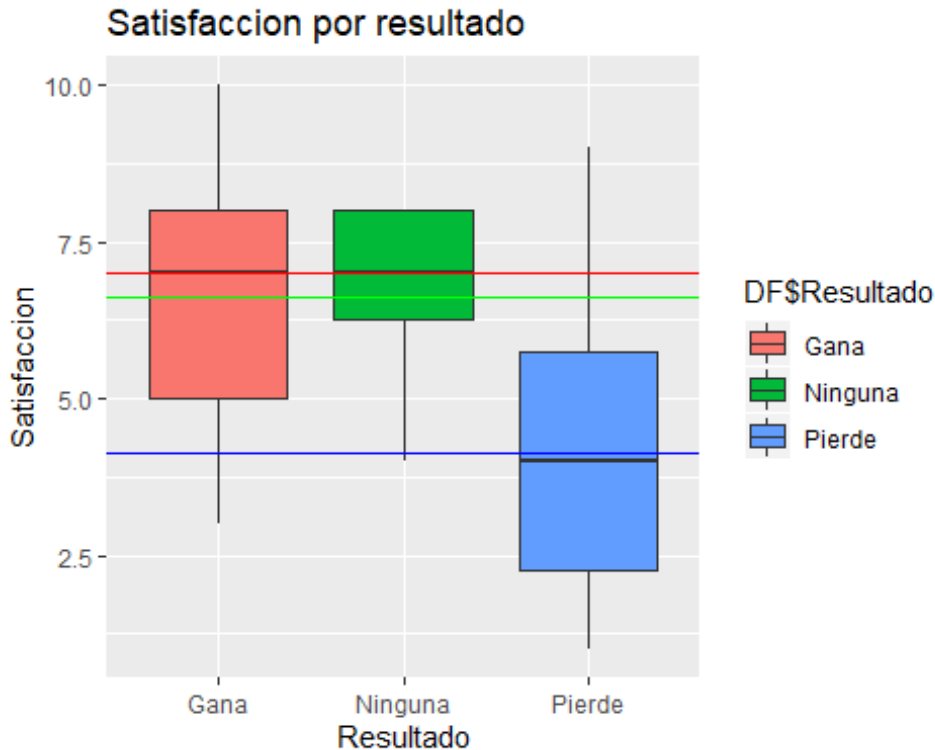


```
plot(DF$Presupuesto~DF$Satisfaccion)
```



```
ggplot(data = DF, aes(x=DF$Resultado, y=DF$Satisfaccion, fill=DF$Resultado))+geom_boxplot(
```

```
) + labs(title = "Satisfaccion por resultado", x = "Resultado", y = "Satisfaccion") + geom_hline(yintercept = 7, color = "red") + geom_hline(yintercept = 6.625, color = "green") + geom_hline(yintercept = 4.142857, color = "blue")
```



```
mean(DF[which(DF$Resultado=="Gana"),]$Satisfaccion)
## [1] 7

mean(DF[which(DF$Resultado=="Pierde"),]$Satisfaccion)
## [1] 4.142857

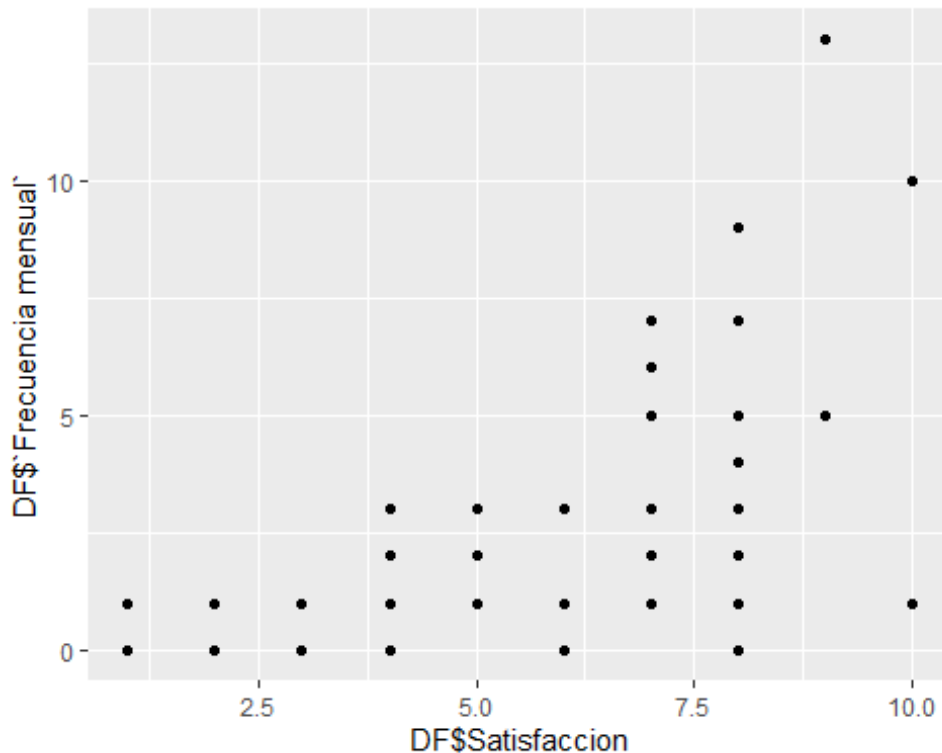
mean(DF[which(DF$Resultado=="Ninguna"),]$Satisfaccion)
## [1] 6.625
```

Preguntas inferenciales

¿Es coherente afirmar que un cliente con alta satisfacción asiste constantemente al casino y su perdida esperada es relativamente mayor?

Para responder la pregunta, usamos una grafica que represente la satisfaccion vs la frecuencia mensual de asistencia del jugador. Luego, por medio de la funcion `geom_smooth`, usamos el modelo de regresion lineal "lm" para intentar hallar una relación:


```
k<-ggplot(data = DF, aes(x=DF$Satisfaccion,y=DF`Frecuencia mensual`))
k+geom_point()
```



```
mm <- lm(DF`Frecuencia mensual`~DF$Satisfaccion)
summary(mm)
```

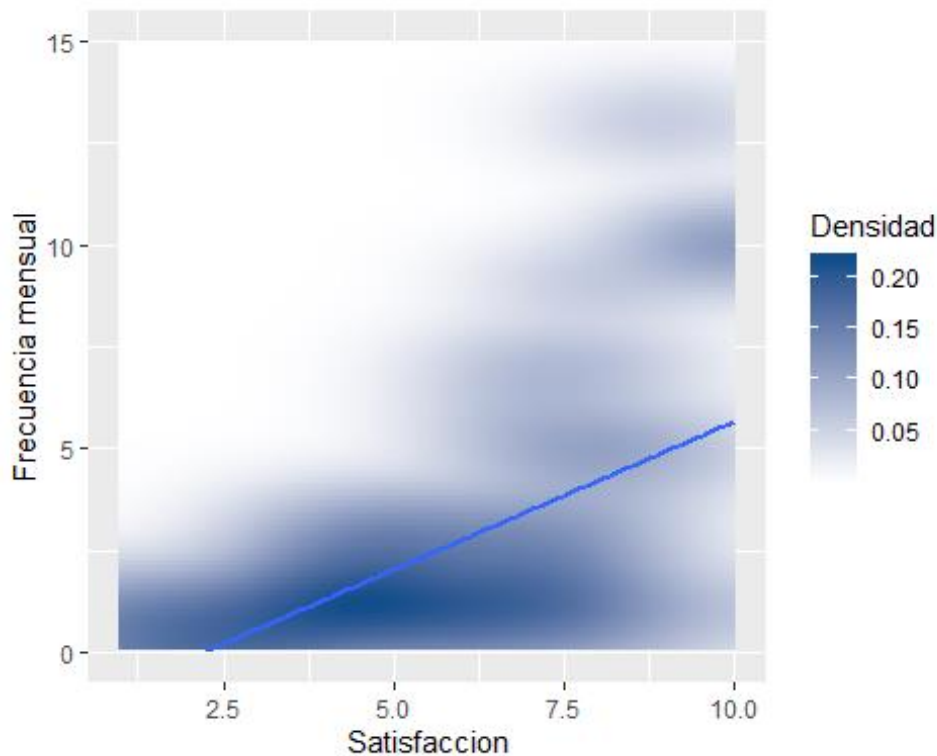
```
##
## Call:
## lm(formula = DF`Frecuencia mensual` ~ DF$Satisfaccion)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.6589 -1.3832 -0.2031  0.8924  8.0690
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1.62038    0.59299  -2.733  0.00779 **
## DF$Satisfaccion  0.72793    0.09925   7.334 1.92e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.173 on 77 degrees of freedom
## Multiple R-squared:  0.4113, Adjusted R-squared:  0.4036
## F-statistic: 53.79 on 1 and 77 DF, p-value: 1.924e-10

satis<-subset(DF, DF$Satisfaccion>=6)
y=0.025
```

```
d=0.025
```

```
gg<-k + stat_density2d(aes(fill = ..density..^0.5), geom = "tile",  
  contour = FALSE, n = 200) + scale_fill_continuous(low = "white", high =  
  "dodgerblue4")  
gg+geom_smooth(method = lm, se=FALSE)+ylim(0,15) + labs(x="Satisfaccion",  
  y="Frecuencia mensual", fill="Densidad")
```

```
## Warning: Removed 11 rows containing missing values (geom_smooth).
```



Con un r^2 de 0.4113, podemos señalar que existe una suficiente relación entre las variables.

Utilizaremos un nivel de confianza del 0.16 para hallar el estadístico de prueba:

```
mean(satis$`Frecuencia mensual`)  
## [1] 3.684211  
sd(satis$`Frecuencia mensual`)  
## [1] 3.495984  
t.test(satis$`Frecuencia mensual`,mu=3,alternative = "g",conf.level =  
  0.16)  
##  
## One Sample t-test
```

```
##
## data:  satis$`Frecuencia mensual`
## t = 1.2065, df = 37, p-value = 0.1176
## alternative hypothesis: true mean is greater than 3
## 16 percent confidence interval:
##  4.255873      Inf
## sample estimates:
## mean of x
##  3.684211
```

Descripción:

```
nsatis<-subset(DF,DF$Satisfaccion<6)
nsatis<-nsatis[complete.cases(nsatis),]
muh<-(mean(nsatis$`Perdida Maxima`))
t.test(satis$`Perdida Maxima`,nsatis$`Perdida Maxima`,conf.level = 0.9)

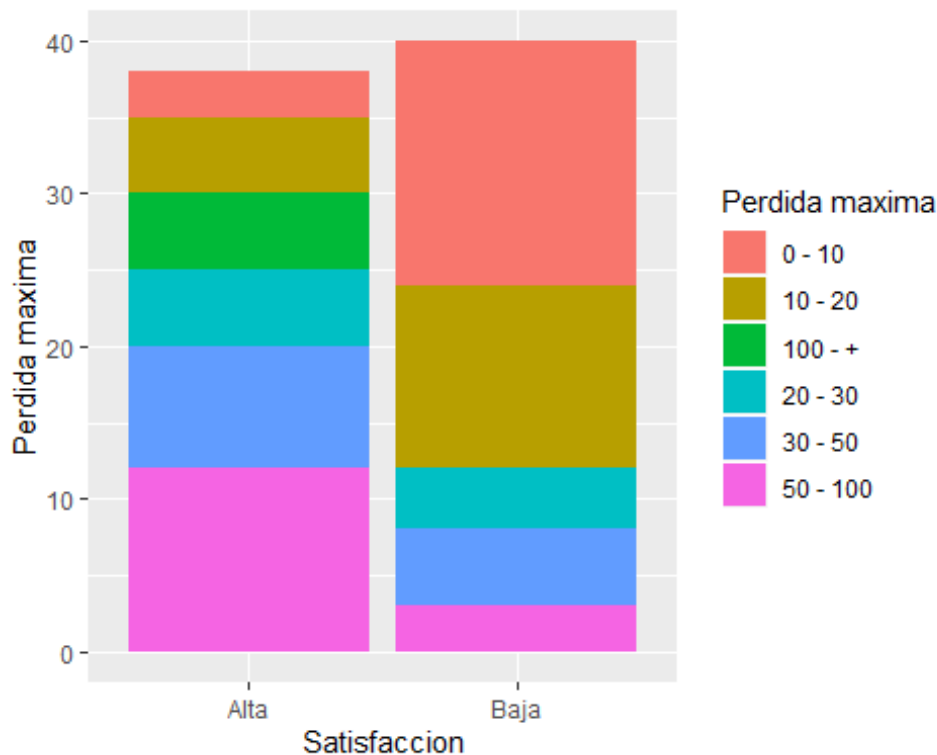
##
##  Welch Two Sample t-test
##
## data:  satis$`Perdida Maxima` and nsatis$`Perdida Maxima`
## t = 3.9161, df = 40.549, p-value = 0.0003371
## alternative hypothesis: true difference in means is not equal to 0
## 90 percent confidence interval:
##  31.15545 78.13200
## sample estimates:
## mean of x mean of y
##  76.10526  21.46154

DFC<-DF
DFC[which(DFC$Satisfaccion>=6),]$Satisfaccion<-"Alta"
DFC[which(DFC$Satisfaccion<6),]$Satisfaccion<-"Baja"
DFC[which(DFC$`Perdida Maxima`>100),]$`Perdida Maxima`<-175
DFC[which(DFC$`Perdida Maxima`<=100 & DFC$`Perdida Maxima`>50),]$`Perdida
Maxima`<-75
DFC[which(DFC$`Perdida Maxima`<=50 & DFC$`Perdida Maxima`>30),]$`Perdida
Maxima`<-35
DFC[which(DFC$`Perdida Maxima`<=30 & DFC$`Perdida Maxima`>20),]$`Perdida
Maxima`<-25
DFC[which(DFC$`Perdida Maxima`<=20 & DFC$`Perdida Maxima`>10),]$`Perdida
Maxima`<-20
DFC[which(DFC$`Perdida Maxima`<=10),]$`Perdida Maxima`<-5

DFC[which(DFC$`Perdida Maxima`==175),]$`Perdida Maxima`<-"100 - +"
DFC[which(DFC$`Perdida Maxima`==75),]$`Perdida Maxima`<-"50 - 100"
DFC[which(DFC$`Perdida Maxima`==35),]$`Perdida Maxima`<-"30 - 50"
DFC[which(DFC$`Perdida Maxima`==25),]$`Perdida Maxima`<-"20 - 30"
DFC[which(DFC$`Perdida Maxima`==20),]$`Perdida Maxima`<-"10 - 20"
DFC[which(DFC$`Perdida Maxima`==5),]$`Perdida Maxima`<-"0 - 10"

estos<-as.data.frame(table(DFC$Satisfaccion, DFC$`Perdida Maxima`))
```

```
ggplot(data = estos, mapping = aes(x = estos$Var1, y = estos$Freq, fill =
estos$Var2, order=estos$Var2)) + geom_bar(stat="identity",
position="stack")+labs(x="Satisfaccion", y="Perdida maxima",fill="Perdida
maxima")+theme_update()
```



¿Existe suficiente evidencia para señalar que la mayoría de personas que juegan a la maquina tragamonedas, son mayores que las que juegan los demas juegos?

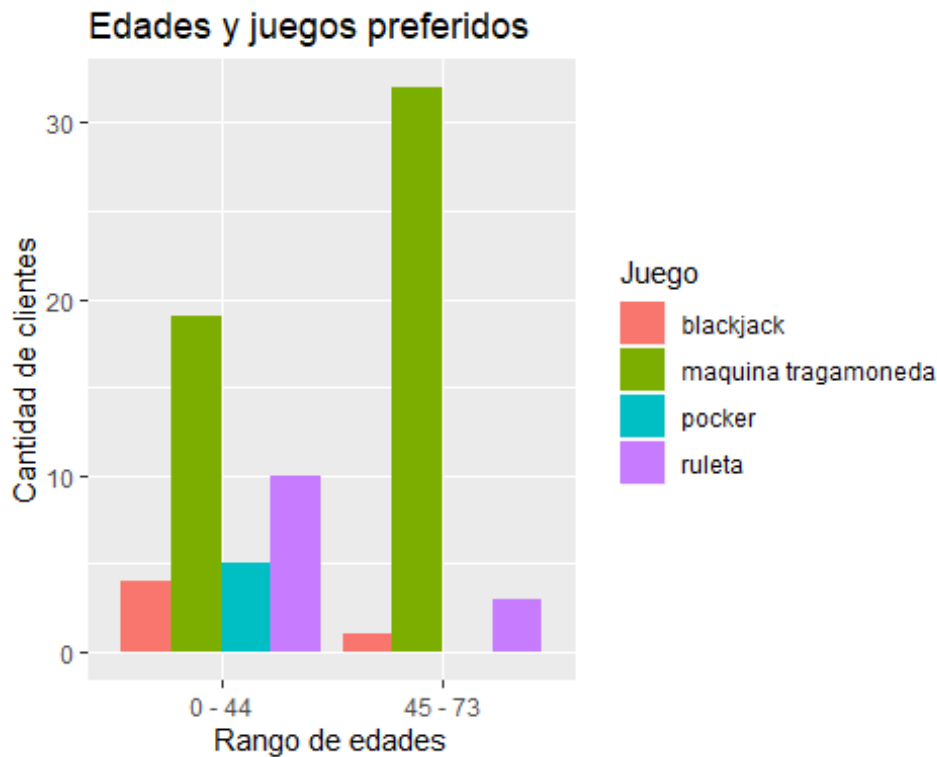
Para responder esta pregunta, se preparará un data frame, con las edades como variable categórica:

```
dfp<-DF
max(DF$Edad)

## [1] 73

dfp[which(dfp$Edad>44),]$Edad<-45
dfp[which(dfp$Edad<=44),]$Edad<-40

dfp[which(dfp$Edad==45),]$Edad<-"45 - 73"
dfp[which(dfp$Edad==40),]$Edad<-"0 - 44"
pref<-as.data.frame(table(dfp$Edad,dfp$`Juego preferido`))
ggplot(data = pref, mapping = aes(x = pref$Var1, y = pref$Freq, fill =
pref$Var2))+geom_bar(stat="identity", position="dodge") + ylab("Cantidad
de clientes") + xlab("Rango de edades") + ggtitle("Edades y juegos
preferidos") + labs(fill="Juego")
```



Sacamos las medias de cada conjunto, de los que prefieren la maquina tragamonedas Y de los que prefieren los demás juegos

```
nmaq<-subset(DF$Edad,DF$`Juego preferido`=="maquina tragamonedas")
maq<-subset(DF$Edad,DF$`Juego preferido`!="maquina tragamonedas")
```

Prueba de hipotesis $H[0]$: edad de no tragamonedas = edad de tragamonedas $H[1]$: edad de no tragamonedas \neq edad de tragamonedas

```
t.test(nmaq,maq,conf.level = 0.9,alternative = "two.sided")

##
##  Welch Two Sample t-test
##
## data:  nmaq and maq
## t = 4.9913, df = 43.432, p-value = 1.02e-05
## alternative hypothesis: true difference in means is not equal to 0
## 90 percent confidence interval:
##  11.89427 23.97103
## sample estimates:
## mean of x mean of y
## 46.01961 28.08696
```

Como demuestra la prueba de hipotesis, aceptamos la hipótesis nula, concluyendo que la edad promedio de las personas que no les gusta la maquina tragamonedas, 1.5 veces menor en el mayor de los casos, con una confianza del 75%.

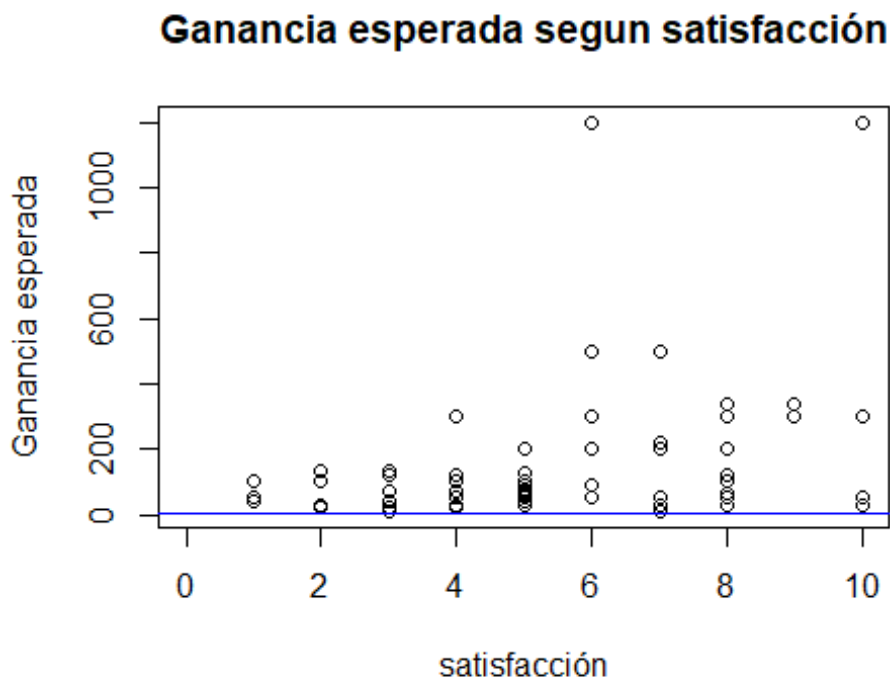
Preguntas predictivas

¿Es posible predecir la ganancia esperada de un jugador sabiendo su nivel de satisfaccion al jugar?

Grafica Ganancia Máxima vs Nivel de satisfacción al jugar

```
satisfaccionaljugar<-DF$Satisfaccion[which(DF$`Ganancia maxima`<1500)]
gananciaesperada<-(DF$`Ganancia maxima`[which(DF$`Ganancia
maxima`<1500)])
plot(satisfaccionaljugar, gananciaesperada
     ,main="Ganancia esperada segun satisfacción"
     ,ylab="Ganancia esperada"
     ,xlab="satisfacción"
     ,xlim=c(0,10))

abline(lm(satisfaccionaljugar~gananciaesperada), col="blue")
coeficiente_es<-c("r:",round(cor(satisfaccionaljugar,
gananciaesperada),2))
legend(25, 70,legend=c(coeficiente_es),cex=0.8)
```



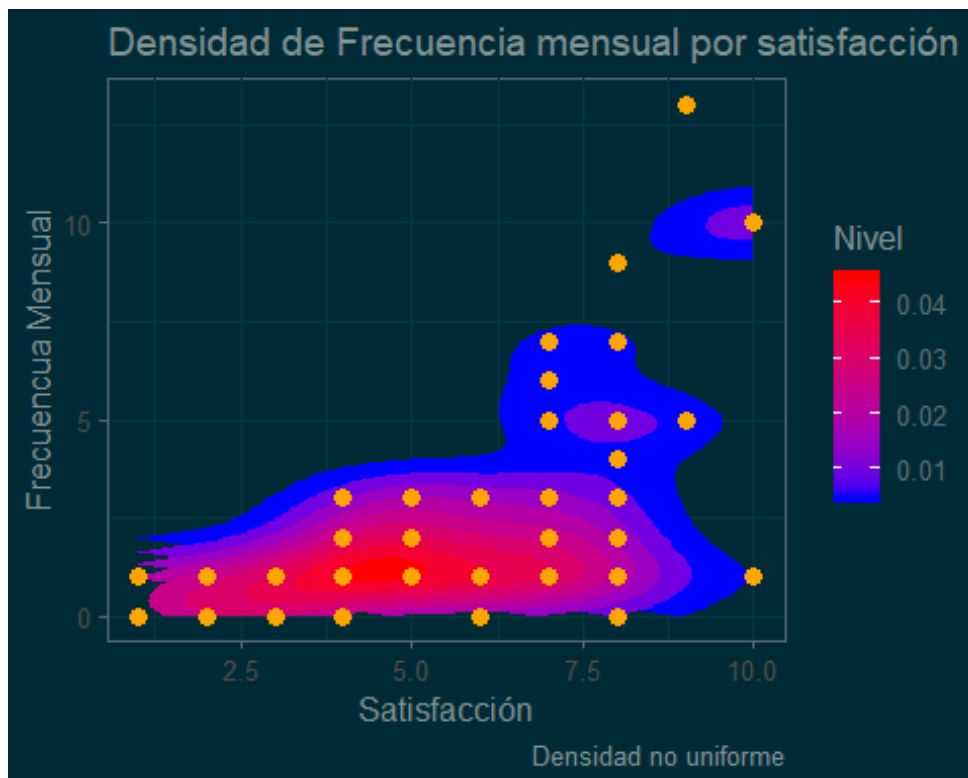
```
summary(lm(gananciaesperada~satisfaccionaljugar))
```

```
##
## Call:
## lm(formula = gananciaesperada ~ satisfaccionaljugar)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -230.13  -98.16  -33.87   39.29 1043.45
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.185      53.184   0.022  0.98229
## satisfaccionaljugar 25.894       9.006   2.875  0.00524 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 192.8 on 76 degrees of freedom
## Multiple R-squared:  0.0981, Adjusted R-squared:  0.08623
## F-statistic: 8.266 on 1 and 76 DF,  p-value: 0.005236
```

¿Se puede conocer la satisfacción de un cliente sabiendo la frecuencia mensual?

```
p1<-ggplot(data = DF, aes(x = DF$Satisfaccion, y =DF$`Frecuencia
mensual`))

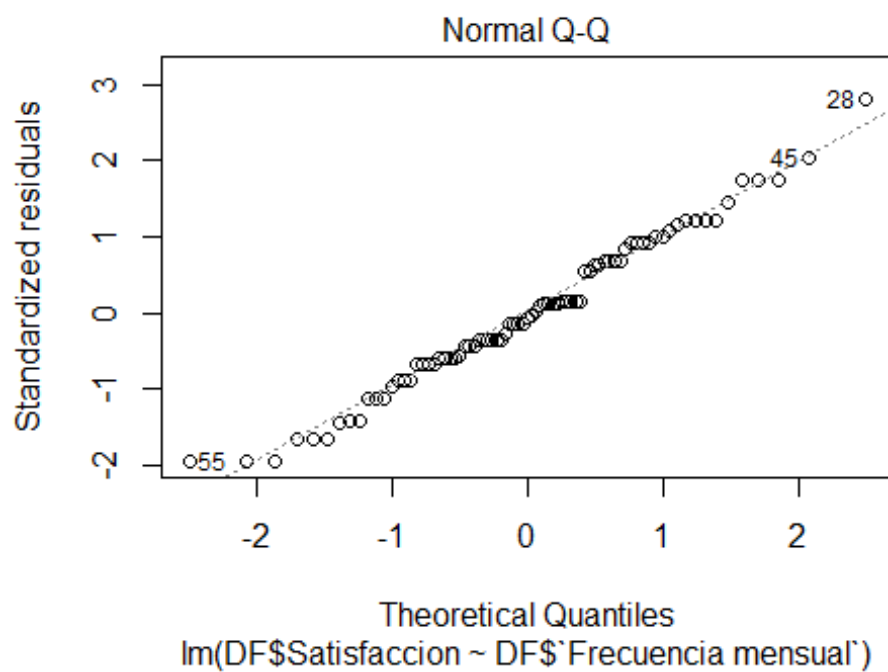
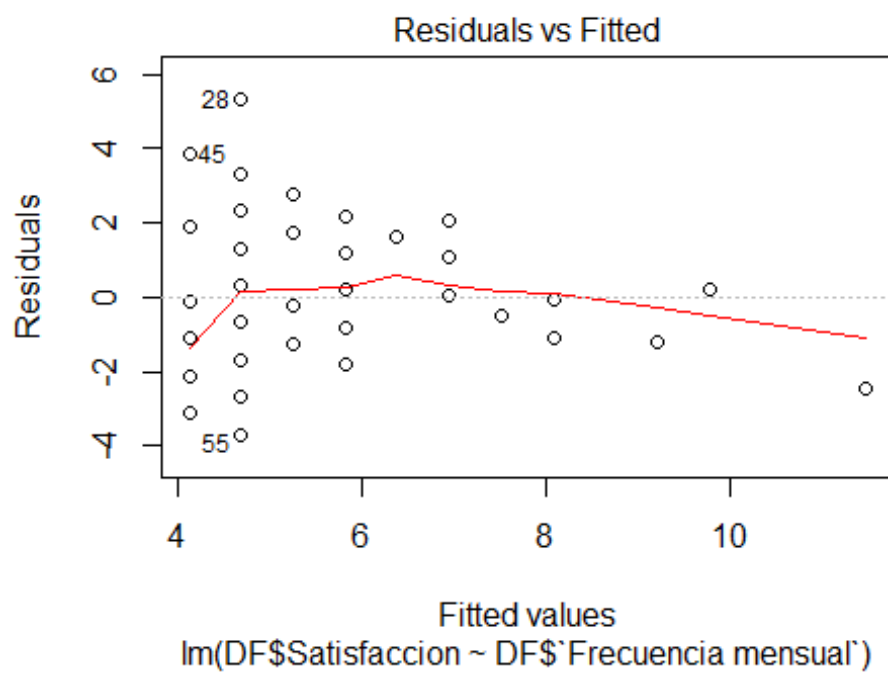
p1 +
  stat_density_2d(aes(fill = ..level..), geom = "polygon") +
  scale_fill_gradient(low = "blue", high = "red") + labs(title = "Density
Plot of Outcome by Predictor") + geom_point(color = "orange", size =3) +
  theme_solarized(light = FALSE) + scale_color_solarized("blue") +
  labs(x="Satisfacción", y="Frecuencia Mensual", title = "Densidad de
Frecuencia mensual por satisfacción", caption = "Densidad no uniforme",
fill="Nivel")
```

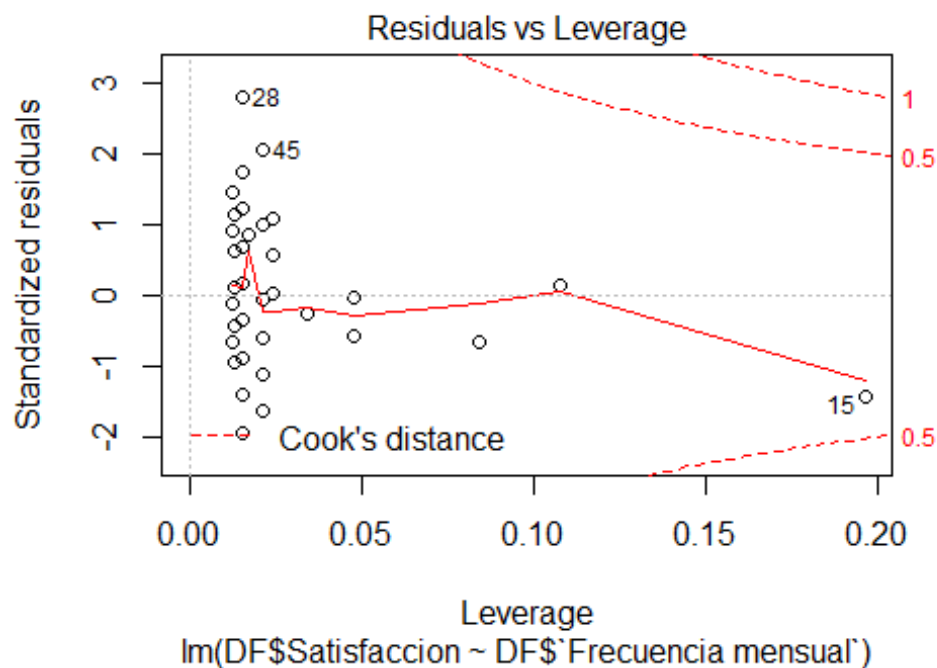
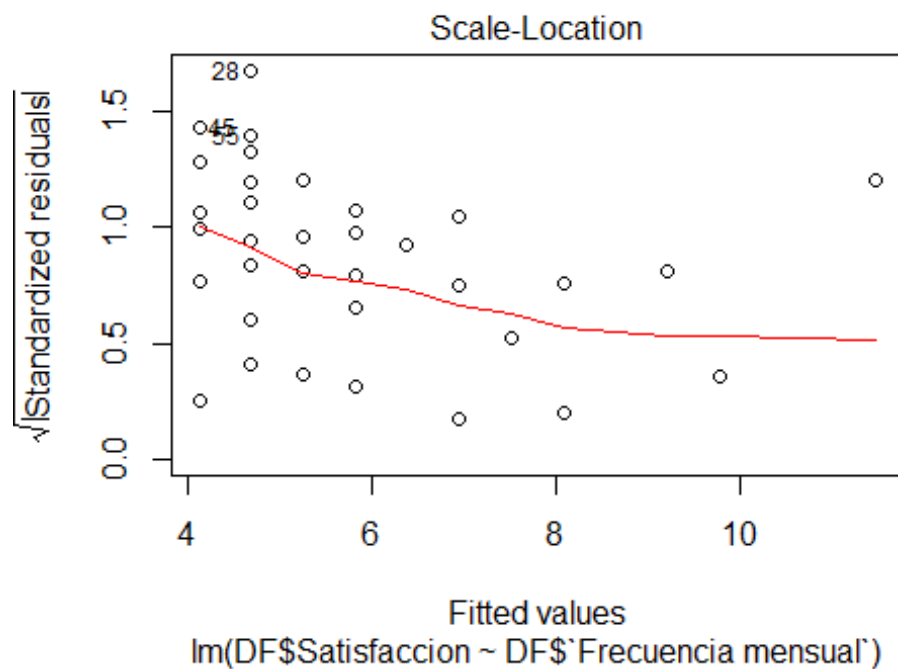


```
ax<-lm(DF$Satisfaccion~DF$`Frecuencia mensual`)
summary(ax)

##
## Call:
## lm(formula = DF$Satisfaccion ~ DF$`Frecuencia mensual`)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6849 -1.1624 -0.1199  1.3151  5.3151
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.11995    0.28098   14.663 < 2e-16 ***
## DF$`Frecuencia mensual` 0.56499    0.07703    7.334 1.92e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.915 on 77 degrees of freedom
## Multiple R-squared:  0.4113, Adjusted R-squared:  0.4036
## F-statistic: 53.79 on 1 and 77 DF, p-value: 1.924e-10

plot(ax)
```

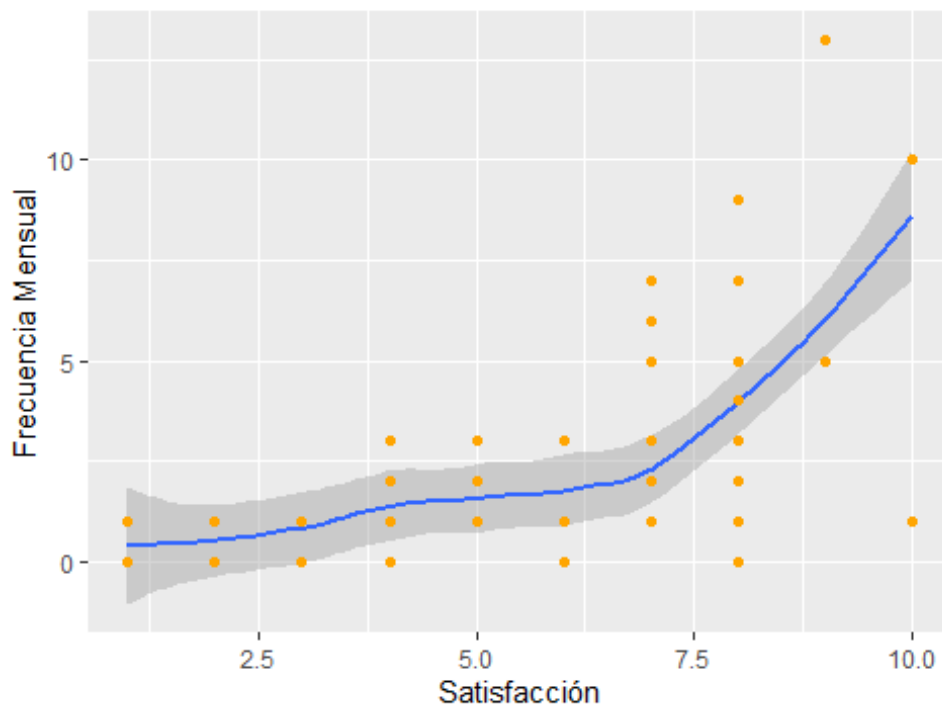





```
p1 + geom_smooth() + geom_point(color = "orange") + labs(title="Regresión suavizada",x="Satisfacción",y="Frecuencia Mensual") + xlim(1,10)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Regresión suavizada



```
p1a<-loess(DF$`Frecuencia mensual`~DF$Satisfaccion, span = 0.1)

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 0.955

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 1.045

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 0

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 1.092

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : zero-width neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : zero-width neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : zero-width neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
```

```

## parametric, : zero-width neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : zero-width neighborhood. make span bigger

summary(p1a)

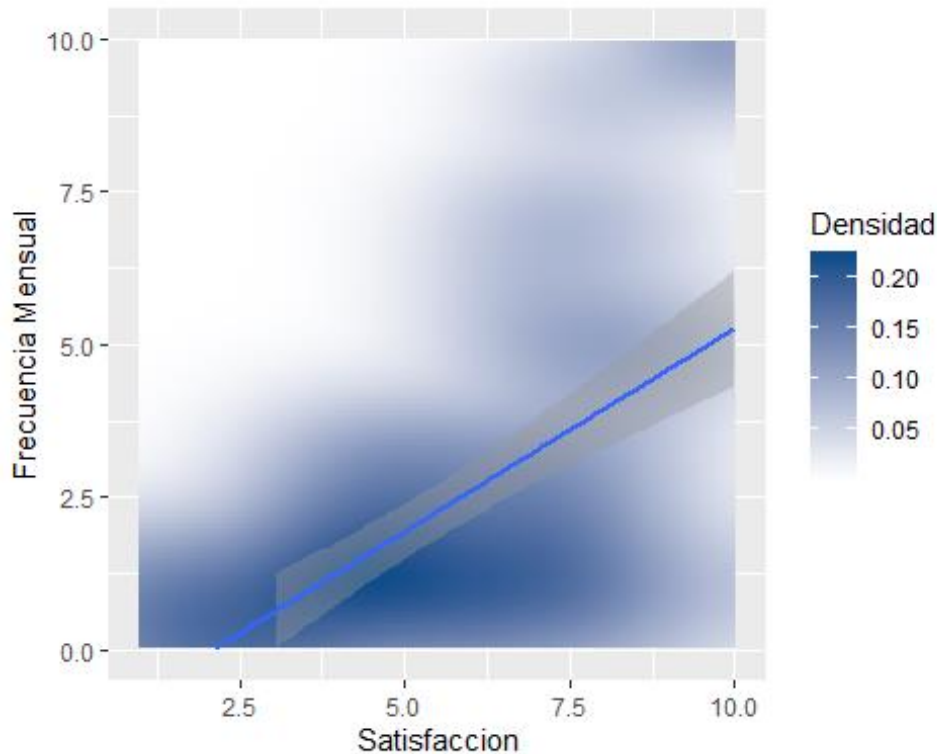
## Call:
## loess(formula = DF$`Frecuencia mensual` ~ DF$Satisfaccion, span = 0.1)
##
## Number of Observations: 79
## Equivalent Number of Parameters: NaN
## Residual Standard Error: NaN
## Trace of smoother matrix: NaN (exact)
##
## Control settings:
##   span      : 0.1
##   degree    : 2
##   family    : gaussian
##   surface   : interpolate      cell = 0.2
##   normalize: TRUE
##   parametric: FALSE
##   drop.square: FALSE

smp1a<-predict(p1a)

library(ggplot2)
satisfaccionaljugar<-DF$Satisfaccion
frec<-DF$`Frecuencia mensual`
ggplot(data =
DF,aes(x=satisfaccionaljugar,y=frec))+stat_density2d(aes(fill =
..density..^0.5), geom = "tile", contour = FALSE, n = 200) +
scale_fill_continuous(low = "white", high = "dodgerblue4") +
geom_smooth(method = lm)+ylim(0,10)+labs(x="Satisfaccion",y="Frecuencia
Mensual",fill="Densidad")

## Warning: Removed 1 rows containing non-finite values (stat_density2d).
## Warning: Removed 1 rows containing non-finite values (stat_smooth).
## Warning: Removed 10 rows containing missing values (geom_smooth).

```



```
summary(lm(DF$`Frecuencia mensual`~DF$Satisfaccion))

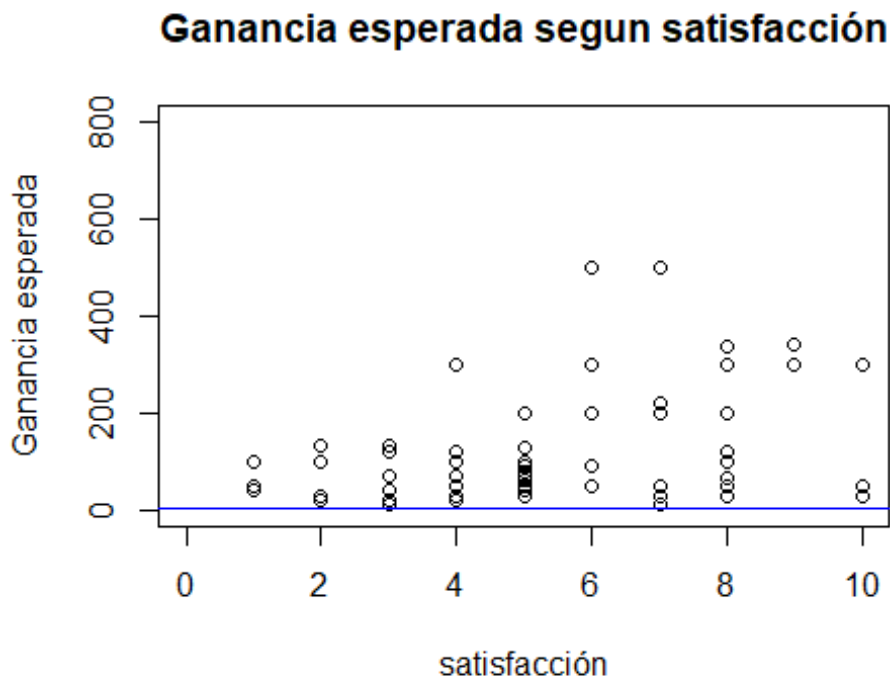
##
## Call:
## lm(formula = DF$`Frecuencia mensual` ~ DF$Satisfaccion)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.6589 -1.3832 -0.2031  0.8924  8.0690
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1.62038    0.59299  -2.733   0.00779 **
## DF$Satisfaccion  0.72793    0.09925   7.334 1.92e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.173 on 77 degrees of freedom
## Multiple R-squared:  0.4113, Adjusted R-squared:  0.4036
## F-statistic: 53.79 on 1 and 77 DF,  p-value: 1.924e-10
```

¿Es posible predecir la ganancia esperada de un jugador sabiendo su nivel de satisfaccion al jugar? Grafica Ganancia Esperada vs Nivel de satisfacción al jugar

```
satisfaccionaljugar<-DF$Satisfaccion[which(DF$`Ganancia maxima`<1500)]
gananciaesperada<-(DF$`Ganancia maxima`[which(DF$`Ganancia
maxima`<1500)])
```

```
plot(satisfaccionaljugar, gananciaesperada
     , main="Ganancia esperada segun satisfacción"
     , ylab="Ganancia esperada"
     , xlab="satisfacción"
     , xlim=c(0,10)
     , ylim=c(0,800))

abline(lm(satisfaccionaljugar~gananciaesperada), col="blue")
coeficiente_es<-c("r:", round(cor(satisfaccionaljugar,
gananciaesperada),2))
legend(25, 70, legend=c(coeficiente_es), cex=0.8)
```



```
summary(lm(gananciaesperada~satisfaccionaljugar))

##
## Call:
## lm(formula = gananciaesperada ~ satisfaccionaljugar)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -230.13  -98.16  -33.87   39.29 1043.45
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.185      53.184   0.022  0.98229
```

```
## satisfaccionaljugar 25.894 9.006 2.875 0.00524 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 192.8 on 76 degrees of freedom
## Multiple R-squared: 0.0981, Adjusted R-squared: 0.08623
## F-statistic: 8.266 on 1 and 76 DF, p-value: 0.005236
```

Preguntas macanísticas

15 ¿Cómo afecta a la tolerancia a perder el conocimiento sobre las probabilidades de ganar?

En esta pregunta se va a analizar si hay alguna relación entre la tolerancia a perder dinero en el casino y el conocimiento de las probabilidades de ganar. Lógicamente, si una persona sabe que sus probabilidades de ganar son bajas, debería de tolerar más la pérdida.

```
probabilidad<- (DF$`Probabilidad de Ganar`)
tolerancias <- (DF$`Perdida Maxima`)
probabilidad

## [1] "No" "No" "No" "No" "No" "0.3" "0.5" "0.5" "No"
"0.6"
## [11] "No" "0.5" "0.3" "0.5" "No" "0.02" "No" "0.4" "0.5"
"0.6"
## [21] "No" "0.5" "No" "40" "0.2" "0.3" "No" "0.5" "0.5"
"0.6"
## [31] "0.1" "0.5" "No" "No" "No" "No" "0.5" "0.1" "0.01"
"0.5"
## [41] "0.6" "No" "0.4" "0.3" "0.5" "0.5" "0.5" "0.2" "0.5"
"0.6"
## [51] "0.01" "0.01" "No" "No" "No" "0.3" "0.01" "0.01" "No"
"0.6"
## [61] "0.02" "0.7" "No" "No" "20" "No" "0.1" "0.01" "0.01"
"0.6"
## [71] "No" "No" "No" "0.2" "No" "0.1" "0.02" "0.01" "No"
"0.6"

tolerancias

## [1] 20 80 30 30 100 20 20 0 10 50 30 50 40 20 100 10
20
## [18] 50 100 100 50 10 30 20 120 200 10 100 50 20 10 20 20
100
## [35] 50 70 10 50 50 30 0 100 50 168 67 50 10 100 NA 30
67
## [52] 20 0 0 50 50 10 10 70 20 0 20 30 67 0 20 30
70
## [69] 20 500 20 150 20 30 10 20 0 10 10
```

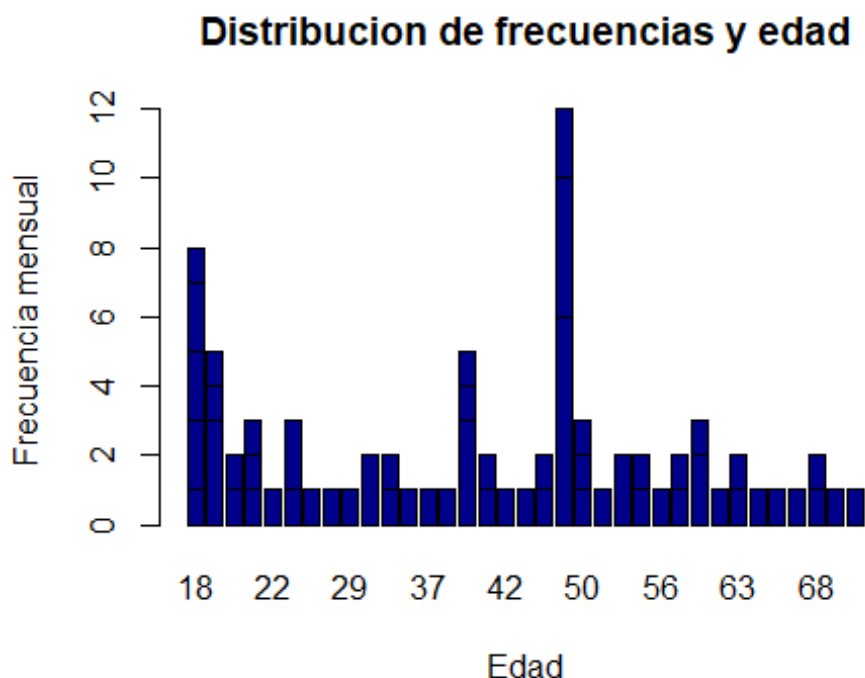
Como podemos ver, si hay una relación directa entre las tolerancias y la idea sobre probabilidades de ganar

16 ¿De qué manera determina el casino la regularidad de los clientes?

En esta pregunta vamos a analizar si la edad de los encuestados afecta la regularidad con la que van al casino. Esto se podría deber a diferentes niveles de estabilidad económica, tiempo libre, etc.

Por otro lado, en el capítulo 5 de la tesis “ESTUDIO SOBRE LA REPUTACIÓN PERCIBIDA POR LOS CLIENTES POTENCIALES, ENFOCADO EN LAS DIMENSIONES DE LA ÉTICA Y RESPONSABILIDAD CORPORATIVA, Y CALIDAD DEL SERVICIO EN LAS SALAS MÁS EXCLUSIVAS DEL SECTOR DE JUEGOS DE CASINOS Y MÁQUINAS TRAGAMONEDAS EN LOS DISTRITOS DE INDEPENDENCIA, SAN MIGUEL, MIRAFLORES, SAN ISIDRO, SANTIAGO DE SURCO, MAGDALENA Y LINCE” (CABRERA LEIVA, Will Jhino) se menciona el análisis de una pregunta relativamente parecida, la diferencia es que se divide la frecuencia y las edades de una manera diferente. Resaltando una gran mayoría de frecuencia en relación a los clientes entre 35 y 56 años.

```
mec1 <- table(DF$`Frecuencia mensual`,DF$Edad)
barplot(mec1, main="Distribucion de frecuencias y edad",
        xlab="Edad", col=c("darkblue"),ylab = "Frecuencia mensual")
```



Para esto vamos a comparar las medias de frecuencia mensual según la edad.


```
fiesta<-na.omit(DF$`Frecuencia mensual`[which(DF$Casino=="Fiesta")])
mean(fiesta)

## [1] 2.974359

atlantic<-na.omit(DF$`Frecuencia mensual`[which(DF$Casino=="Atlantic")])
atlantic

## [1] 1 3 1 2 1 3 1 9 1 5 2 1 1 1 0 2 1 1 10 2

mean(atlantic)

## [1] 2.4

ninguno<-na.omit(DF$`Frecuencia mensual`[which(DF$Casino=="Ninguno")])
mean(ninguno)

## [1] 0.6666667
```

Como se puede ver, la media de frecuencia en los clientes del casino Fiesta es mucho mas alta. Esto nos dice, por lo menos para nuestra muestra, que existen factores que hacen que los clientes sean mas fieles en ese casino.

Preguntas casuales:

17 ¿Influye la edad en el juego preferido de los clientes?

Si bien anteriormente analizamos la relación entre la edad y la frecuencia de asistencia al casino, en esta pregunta veremos si hay una relación entre la edad y el juego preferido de los clientes:

```
edad<-na.omit(DF$Edad)
juego<-na.omit(DF$Juego)

## Warning: Unknown or uninitialised column: 'Juego'.
```

```
plot(juego,edad,xlab="Juego favorito",ylab="Edad", main="Juego favorito vs Edad")
```

Como podemos ver en la gráfica, en realidad no hay ninguna particularidad entre los diferentes juegos favoritos. Esta falta de relación probablemente se da a que el juego favorito es muy subjetivo. Si por ejemplo, evaluáramos el juego más jugado, se podría encontrar una relación. Pero, para este caso, se podría concluir que, al menos en nuestra muestra, no hay una aparente relación.

18 ¿Qué factores determinan la ganancia máxima?

Dado, que nosotros conocemos teoría de probabilidades, si evaluamos las posibilidades de en este ejemplo:

Posibilidades de Duplicar tu dinero en porcentaje

Ruleta

```
(17/36)*100
```

```
## [1] 47.22222
```

En máquina tragamonedas:

La mayoría de máquina de tragamonedas tienen un retorno de entre 80 y 95 céntimos de retorno por sol invertido. Sin embargo, dado que hay muchos premios distintos que ocurren con frecuencias muy diferentes, no se puede determinar la probabilidad de duplicar tu dinero.

En blackjack:

La probabilidad de ganar en este juego esta severmente influenciada por las reglas con las que se juega, cada casino tiene diferentes reglas que beneficia más al jugador o a la casa.

En una situación irrealistamente positiva para el jugador: Si consideramos que se juega con un solo deck, el dealer pide hasta 17 y el pago de blackjack

Probabilidad de sacar blackjack, pago 3/2

```
result <- 4*16/choose(52,2)
```

```
result
```

```
## [1] 0.04826546
```

Probabilidad de ganar la mano Dada la gran cantidad de variables que entran en juego al calcular una sola mano lo calcularemos con lo que se sabe que tiene de ventaja casa en un juego general de blackjack que es 0.05 entonces :

Aproximados: Probabilidad de ganar:44 Probabilidad de de perder:48 Probabilidad de empatar: 8

Preguntas macanísticas

15 ¿Cómo afecta a la tolerancia a perder el conocimiento sobre las probabilidades de ganar?

En esta pregunta se va a analizar si hay alguna relación entre la tolerancia a perder dinero en el casino y el conocimiento de las probabilidades de ganar. Lógicamente, si una persona sabe que sus probabilidades de ganar son bajas, debería de tolerar más la pérdida.

```
probabilidad<- (DF$Probabilidad.de.Ganar)
```

```
## Warning: Unknown or uninitialised column: 'Probabilidad.de.Ganar'.
```

```
tolerancias <- (DF$Pérdida.Maxima)
```

```
## Warning: Unknown or uninitialised column: 'Pérdida.Maxima'.
```

```
probabilidad
```

```
## NULL
```

```
tolerancias
```

```
## NULL
```

Como podemos ver, si hay una relación directa entre las tolerancias y la idea sobre probabilidades de ganar

16 ¿De qué manera determina la la regularidad de los clientes?

En esta pregunta vamos a analizar si la edad de los encuestados afecta la regularidad con la que van al casino. Esto se podría deber a diferentes niveles de estabilidad económica, tiempo libre, etc.

Por otro lado, en el capítulo 5 de la tesis “ESTUDIO SOBRE LA REPUTACIÓN PERCIBIDA POR LOS CLIENTES POTENCIALES, ENFOCADO EN LAS DIMENSIONES DE LA ÉTICA Y RESPONSABILIDAD CORPORATIVA, Y CALIDAD DEL SERVICIO EN LAS SALAS MÁS EXCLUSIVAS DEL SECTOR DE JUEGOS DE CASINOS Y MÁQUINAS TRAGAMONEDAS EN LOS DISTRITOS DE INDEPENDENCIA, SAN MIGUEL, MIRAFLORES, SAN ISIDRO, SANTIAGO DE SURCO, MAGDALENA Y LINCE”(CABRERA LEIVA, Will Jhino) se menciona el análisis de una pregunta relativamente parecida, la diferencia es que se divide la frecuencia y las edades de una manera diferente. Resaltando una gran mayoría de frecuencia en relación a los clientes entre 35 y 56 años.

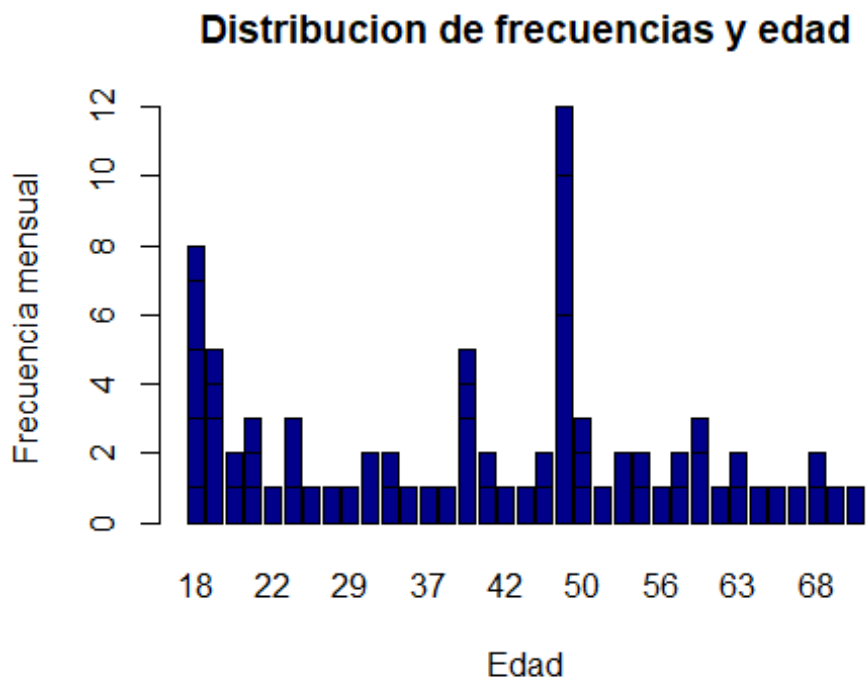
```
mec1 <- table(DF$`Frecuencia mensual`,DF$Edad)
length(DF$`Frecuencia mensual`)

## [1] 79

length(DF$Edad)

## [1] 79

barplot(mec1, main="Distribucion de frecuencias y edad",
        xlab="Edad", col=c("darkblue"),ylab = "Frecuencia mensual")
```



Para esto vamos a comparar las medias de frecuencia mensual según la edad.

```
joven<-na.omit(DF$Frecuencia.mensual[which(DF$Edad<35)])
## Warning: Unknown or uninitialised column: 'Frecuencia.mensual'.
mean(joven)
## Warning in mean.default(joven): argument is not numeric or logical:
## returning NA
## [1] NA

adulto<-na.omit(DF$Frecuencia.mensual[which(DF$Edad<56,DF$Edad>35)])
## Warning: Unknown or uninitialised column: 'Frecuencia.mensual'.
mean(adulto)
## Warning in mean.default(adulto): argument is not numeric or logical:
## returning NA
## [1] NA

adultomayor<-na.omit(DF$Frecuencia.mensual[which(DF$Edad>56)])
## Warning: Unknown or uninitialised column: 'Frecuencia.mensual'.
mean(adultomayor)
```

```
## Warning in mean.default(adultomayor): argument is not numeric or
logical:
## returning NA

## [1] NA
```

Como se puede ver, la media de frecuencia en los clientes del casino Flesta es mucho mas alta. Esto nos dice, por lo menos para nuestra muestra, que existen factores que hacen que los clientes sean mas fieles en ese casino.

Conclusión

En conclusión, pudimos analizar respuestas obtenidas mediante encuestas para luego de separarlas por variables y posteriormente correlacionarlas para poder ir formulando hipótesis y corroborarlas con nuestros datos. Mediante muchos tipos de preguntas planteadas, pudimos hallar relaciones. Dado que durante el ciclo no hemos todas las herramientas necesarias para contestar bien las preguntas causales y mecanísticas. Por esta razón, hemos simplemente hecho nuestro mejor intento para ganar información que podría ayudarnos a responder la pregunta en el futuro, nos hemos basado en fuentes académicas, cálculos probabilísticos y análisis de nuestros datos. Respecto a las mecanísticas : Como podemos ver en las gráficas, las personas mayores parecen tener una mayor frecuencia de asistencia y las personas que no conocen sus reales probabilidades de ganar suelen tolerar más las pérdidas. Respecto a las preguntas causales, parece que hay una relación entre el juego favorito y la edad. También se puede observar como varía la ganancia según el juego preferido. Con las respuestas de las preguntas predictivas, podemos afirmar que la ganancia esperada no es explicada por el nivel de satisfacción y que la satisfacción del cliente podría estar relacionada a la frecuencia de asistencia mensual. Esto y más nos ayuda a sacar conclusiones y validar o no hipótesis varias. Se podría usar el estudio para hacer inferencias sobre rasgos de ludopatía en personas.