

Proyecto Inferencia

Universidad Finis Terrae

Gianfranco Astorga, Jean Lucas Peñaloza, Diego Godoy

3 de noviembre de 2023

Indice

- Introducción
- Objetivos
- Objetivo General
- Objetivos Específicos
- Trabajo Previo
- Marco Teórico
- Regresión Lineal
- Correlación Lineal de Pearson
- Desarrollo
- Análisis Exploratorio de Datos
 - Carga de Datos
 - Visualización de Datos
 - Análisis de Datos
 - Estadística Descriptiva
 - Gráficos de caja y de dispersión
 - Correlación
- Investigación de modelo de regresión lineal
- Correlación lineal de Pearson
- Gráfico de dispersión
 - Calidad percibida vs Agilidad en la entrega del servicio
 - Calidad percibida vs Flexibilidad de pago
 - Calidad percibida vs Atención del personal
- Modelo de regresión lineal “Calidad percibida vs Agilidad en la entrega del servicio”
- Coeficientes del modelo
 - Interpretación de los coeficientes del modelo
 - Interpretación del coeficiente de determinación
 - Interpretación del valor p del coeficiente de regresión
- Investigación ANOVA modelo de regresión lineal
 - Interpretación de la tabla ANOVA
- Conclusiones
- Bibliografía

1. Introducción

El análisis de regresión lineal se ha utilizado ampliamente en diversos campos para estudiar la relación entre variables. El término “regresión” fue acuñado por Sir Francis Galton, quien lo empleó para describir los resultados de sus experimentos con arvejas. El contexto de este informe es el ámbito empresarial, específicamente en relación a la calidad del servicio al cliente. El problema a resolver es cómo mantener la preferencia de los clientes y lograr una opinión positiva a través de la calidad del servicio al cliente, considerando que esta puede ser una herramienta estratégica para ofrecer un valor añadido a los clientes y lograr una percepción de diferencias en la oferta global de la empresa. Además, se busca brindar recomendaciones para medir la percepción de los clientes sobre el servicio entregado y para mejorar la calidad del servicio al cliente considerando factores relevantes. En primer lugar, se realizará un análisis de estadística descriptiva utilizando los datos proporcionados en un archivo de texto mediante R Studio. Esto para saber los datos de los clientes y cómo está relacionado con el servicio entregado. Además, se investigará el concepto de regresión lineal simple a través de referencias confiables, como libros especializados. Esto permitirá comprender en detalle los fundamentos teóricos y las implicaciones del modelo a proponer.

2. Objetivos

2.1 Objetivo General

- Realizar un análisis de datos efectivo utilizando R Studio para obtener información relevante y valiosa a partir de conjuntos de datos empresariales.

2.2 Objetivos Específicos

- Realizar un análisis exploratorio de datos en R Studio, utilizando gráficos y estadísticas descriptivas para comprender la distribución y relaciones entre variables clave.
- Aplicar técnicas de modelado estadístico (regresión lineal, clustering, análisis de series temporales) según sea necesario, para obtener información predictiva o descriptiva.
- Interpretar resultados de análisis en R Studio y presentar hallazgos en informes claros y visuales para respaldar decisiones empresariales.
- Automatizar procesos de análisis de datos mediante scripts y programación en R Studio para futuros análisis eficientes.
- Colaborar con el equipo para aplicar conocimientos del análisis de datos en decisiones estratégicas de la empresa.

3. Trabajo Previo

Luego de investigaciones sobre la utilización de la regresión lineal en algunos ámbitos de la vida cotidiana se encontró un proyecto en Colombia que si bien no fue para el mismo propósito que en este proyecto, se utilizó el modelo de regresión lineal para realizar un análisis de los determinantes de la inversión extranjera en Colombia. Los datos utilizados para el análisis en los años 1989-2009, fueron obtenidos en su mayoría de la base de datos del Banco Mundial, el World Data Bank, a excepción de la variable del Gasto en Educación la cual fue obtenida de la base de datos de la CEPAL. Para ésta variable en específico, debido a la falta de información para los años 1989 y 2009 se obtuvo el promedio de los dos años siguientes para 1989 y de los dos años anteriores para 2009.

	N	Mínimo	Máximo	Media	Desv. típ.	Varianza	Asimetría		Curtosis	
	Estadístico	Estadístico	Estadístico	Estadístico	Estadístico	Estadístico	Estadístico	Error típico	Estadístico	Error típico
IED	21	6,552E8	1,025E10	3,70653E9	3,048793E9	9,295E18	1,019	,501	-,229	,972
PIB	21	-4,204015243	6,905843681	3,415311698	2,543249685	6,468	-1,224	,501	2,746	,972
GM	21	2,202499910	4,141208229	3,049886623	,5083466595	,258	,208	,501	-,481	,972
TC	21	78,07356180	121,4188159	97,33452278	13,71389810	188,071	,049	,501	-1,304	,972
GE	21	2,34	4,48	3,2367	,62457	,390	,466	,501	-,551	,972
NRR	21	,4322562630	2,160857430	1,055826439	,3902451690	,152	1,030	,501	2,185	,972
N válido (según lista)	21									

Figure 1: Análisis de los datos del Banco

4. Marco Teórico

Para el marco teórico, se detallarán los contenidos necesarios para la realización de los problemas dados, en el cual se podrán definir los conceptos básicos, matemáticos y físicos, para la ejecución del presente proyecto, en base a la asignatura de inferencia estadística:

4.1 Regresión Lineal

Su objetivo es estimar los coeficientes de regresión ($B_0, B_1, B_2, \dots, B_n$) que minimicen la suma de los cuadrados de los residuos, es decir, la diferencia entre los valores observados de Y y los valores predichos por el modelo. El proceso de ajuste del modelo de regresión lineal implica varios pasos, como la selección de variables, la comprobación de supuestos, la estimación de los coeficientes y la evaluación de la calidad del ajuste mediante medidas como el coeficiente de determinación (R^2) y el análisis de los residuos. Por otra parte, la regresión lineal tiene varias extensiones, como la regresión lineal múltiple (cuando hay más de una variable independiente), la regresión lineal ponderada (cuando se asignan pesos diferentes a los puntos de datos) y la regresión lineal robusta (que es menos sensible a valores atípicos). Por ende en la realización del proyecto, se utilizó la regresión lineal para modelar la relación entre variable dependiente y una o más variables independientes. El cual se expresa mediante la siguiente ecuación:

$$Y = B_0 + B_1X_1 + B_2X_2 + \dots + B_nX_n + E$$

Donde:

- Y es la variable dependiente que se intenta predecir.
- X_1, X_2, \dots, X_n son las variables independientes que se utilizan para predecir Y .
- $B_0, B_1, B_2, \dots, B_n$ son los coeficientes de regresión que representan el efecto de cada variable independiente sobre Y .
- E es el término de error, que captura la variabilidad no explicada por el modelo y se asume que sigue una distribución normal con media cero.

4.2 Correlación Lineal de Pearson

La correlación lineal de Pearson es una medida de la relación estadística entre dos variables cuantitativas. Es una herramienta utilizada en estadística para medir la dependencia lineal entre dos conjuntos de datos. El coeficiente de correlación de Pearson es una medida de la fuerza y dirección de la relación lineal entre dos variables. El coeficiente de correlación de Pearson oscila entre -1 y 1, donde un valor de -1 indica una correlación negativa perfecta, un valor de 0 indica que no hay correlación y un valor de 1 indica una

correlación positiva perfecta. El coeficiente de correlación de Pearson se calcula dividiendo la covarianza de las dos variables por el producto de sus desviaciones estándar. La correlación de Pearson es independiente de la escala de medida de las variables y se utiliza para medir la relación lineal entre dos variables continuas. En resumen, la correlación lineal de Pearson es una medida de la relación estadística entre dos variables cuantitativas y se utiliza para medir la dependencia lineal entre dos conjuntos de datos. La fórmula de la correlación de Pearson se representa de la siguiente manera:

$$r = \frac{(n \cdot S_{xy} - S_x S_y)}{\sqrt{(n S_x^2 - (S_x)^2)(n S_y^2 - (S_y)^2)}}$$

- Sumatoria = S

Donde:

- r es el coeficiente de correlación de Pearson.
- n es el número de observaciones.
- S_{xy} es la suma de los productos de las puntuaciones de X e Y.
- S_x es la suma de las puntuaciones de X.
- S_y es la suma de las puntuaciones de Y.
- S_x^2 es la suma de los cuadrados de las puntuaciones de X.
- S_y^2 es la suma de los cuadrados de las puntuaciones de Y.

5. Desarrollo

5.1 Análisis Exploratorio de Datos

Para el análisis exploratorio de datos se utilizó el software R Studio, el cual es un entorno de desarrollo integrado (IDE) para R, un lenguaje de programación para estadísticas y gráficos. Por otro lado se consideró caracterizar la opinión de los clientes con respecto a la calidad del servicio al cliente. A continuación, se presenta el análisis exploratorio de datos realizado en R Studio:

5.1.1 Carga de Datos

Para la carga de datos se utilizó el siguiente código:

```
# Cargar el paquete 'readr' para leer archivos de texto # nolint
library(readr)

df <- read_delim("C:/Users/choco/OneDrive/Documentos/GitHub/Proyecto_Inferencia/PROVLOGIS.csv", delim = ";")

## Rows: 100 Columns: 7
## -- Column specification -----
## Delimiter: ";"
## db1 (7): id, Calidad, agilidadentrega, flexpago, AtencionCH, imagen, cliente...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

View(df)
```

5.1.2 Visualización de Datos

Para la visualización de datos se utilizó el siguiente código:

```
# Visualizar las primeras filas del DataFrame para verificar la estructura de los datos # nolint
head(df)
```

```
## # A tibble: 6 x 7
##       id Calidad agilidadentrega flexpago AtencionCH imagen clientepref
##   <dbl>   <dbl>         <dbl>    <dbl>      <dbl>   <dbl>      <dbl>
## 1     1     4.2           4.1      3.9       4.4     4.7         1
## 2     2     4.3           1.8      3.3       4.5     2         2
## 3     3     5.2           3.4      2.7       6.3     3         2
## 4     4     3.9           6        4.1       3.8     5.9         2
## 5     5     6.1           5.2      6.3       5.7     2         1
## 6     6     4.4           1.9      4.9       4.6     4.8         1
```

Se comienza por visualizar las primeras phineas del Data Frame para tener una primera impresión de la estructura de los datos. Lo cual nos permitió verificar que los datos fueron cargados correctamente y tener una idea de como se presentan obteniendo los siguientes datos

5.1.3 Análisis de Datos

Para el análisis de datos se utilizó el siguiente código:

```
# Obtener información básica del DataFrame
str(df)
```

```
## spc_tbl_ [100 x 7] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ id      : num [1:100] 1 2 3 4 5 6 7 8 9 10 ...
## $ Calidad : num [1:100] 4.2 4.3 5.2 3.9 6.1 4.4 5.8 4.3 5.4 5.4 ...
## $ agilidadentrega: num [1:100] 4.1 1.8 3.4 6 5.2 1.9 4.6 1.3 5.5 4 ...
## $ flexpago  : num [1:100] 3.9 3.3 2.7 4.1 6.3 4.9 6.5 3.2 6.4 3.5 ...
## $ AtencionCH : num [1:100] 4.4 4.5 6.3 3.8 5.7 4.6 5.5 4.8 5.5 5.7 ...
## $ imagen    : num [1:100] 4.7 2 3 5.9 2 4.8 2 5.1 3 3 ...
## $ clientepref : num [1:100] 1 2 2 2 1 1 1 2 1 1 ...
## - attr(*, "spec")=
## .. cols(
## ..   id = col_double(),
## ..   Calidad = col_double(),
## ..   agilidadentrega = col_double(),
## ..   flexpago = col_double(),
## ..   AtencionCH = col_double(),
## ..   imagen = col_double(),
## ..   clientepref = col_double()
## .. )
## - attr(*, "problems")=<externalptr>
```

Se utilizó la función “str(df)” para obtener información básica sobre la estructura del Data Frame, obtener los nombres de las columnas y los tipos de datos en cada columna.

5.1.4 Estadística Descriptiva

Para la estadística descriptiva se utilizó el siguiente código:

```
# Calcular estadísticas descriptivas
```

```
summary(df)
```

```
##      id      Calidad  agilidadentrega  flexpago
## Min.   : 1.00   Min.   :2.000   Min.   :0.60   Min.   :2.000
## 1st Qu.: 25.75  1st Qu.:4.200   1st Qu.:2.80   1st Qu.:3.700
## Median : 50.50  Median :4.800   Median :3.65   Median :5.050
## Mean   : 50.50  Mean   :4.758   Mean   :3.62   Mean   :4.894
## 3rd Qu.: 75.25  3rd Qu.:5.400   3rd Qu.:4.70   3rd Qu.:6.100
## Max.   :100.00  Max.   :7.000   Max.   :6.00   Max.   :7.000
##  AtencionCH  imagen  clientepref
## Min.   :3.200  Min.   :1.500  Min.   :1.00
## 1st Qu.:4.400  1st Qu.:3.650  1st Qu.:1.00
## Median :5.000  Median :4.800  Median :1.00
## Mean   :4.926  Mean   :4.531  Mean   :1.33
## 3rd Qu.:5.500  3rd Qu.:5.700  3rd Qu.:2.00
## Max.   :6.600  Max.   :7.000  Max.   :2.00
```

Por otro lado se empleó la función “summary(df)” para calcular estadísticas descriptivas básicas de todas las columnas en el Data Frame. Esto proporcionó una visión general de la tendencia central y la dispersión de los datos en cada variable, lo que es crucial para comprender la distribución de la opinión de los clientes.

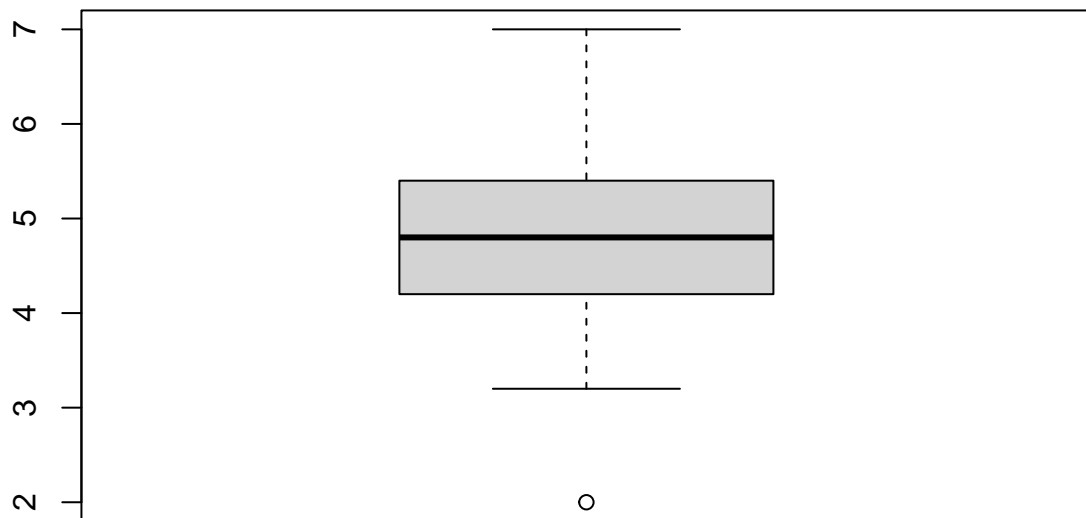
5.1.5 Gráficos de caja y de dispersión

Para los gráficos de caja el siguiente código:

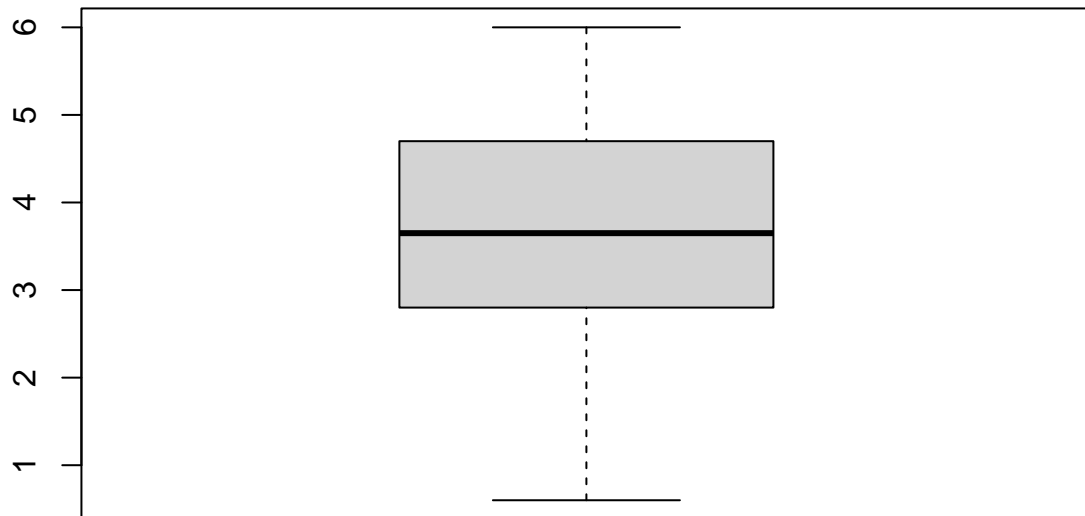
```
# Crear un gráfico de caja para identificar valores atípicos (por ejemplo, en la variable 'Calidad') #
```

```
names(df) <- c("ID", "CPS", "AES", "FP", "AP", "IP", "CP") # nolint
```

```
boxplot(df$CPS)
```



```
boxplot(df$AES)
```

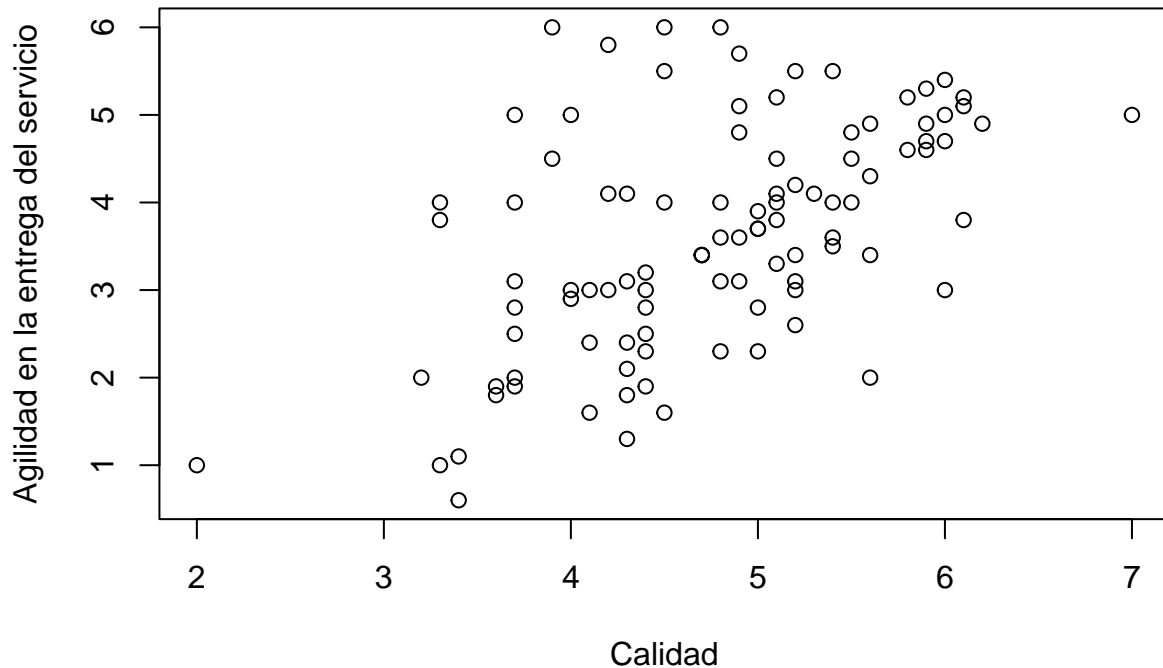



Se generaron dos gráficos de caja utilizando el comando “`boxplot(dfCalidad)`” y `boxplot(dfagilidad entrega)`”. El primer gráfico ayudó a visualizar la distribución de la variable “Calidad” e identificar posibles valores atípicos. Los valores atípicos pueden ser indicativos de experiencias extremas de los clientes y son importantes para el análisis. Por otro lado el gráfico de agilidad en la entrega entregó una línea central representando la mediana entre los valores 4 y 3, por lo que son unos valores para comprender la opinión de los clientes sobre la agilidad en la entrega del servicio y para identificar las posibles áreas de mejora o puntos fuertes en este aspecto.

Para el gráficos de caja el siguiente código:

```
# Crear un gráfico de dispersión para identificar relaciones entre variables (por ejemplo, 'Calidad' y
plot(df$CPS, df$AES, xlab = 'Calidad', ylab = 'Agilidad en la entrega del servicio', main = 'Calidad vs
```

Calidad vs. Agilidad en la entrega del servicio



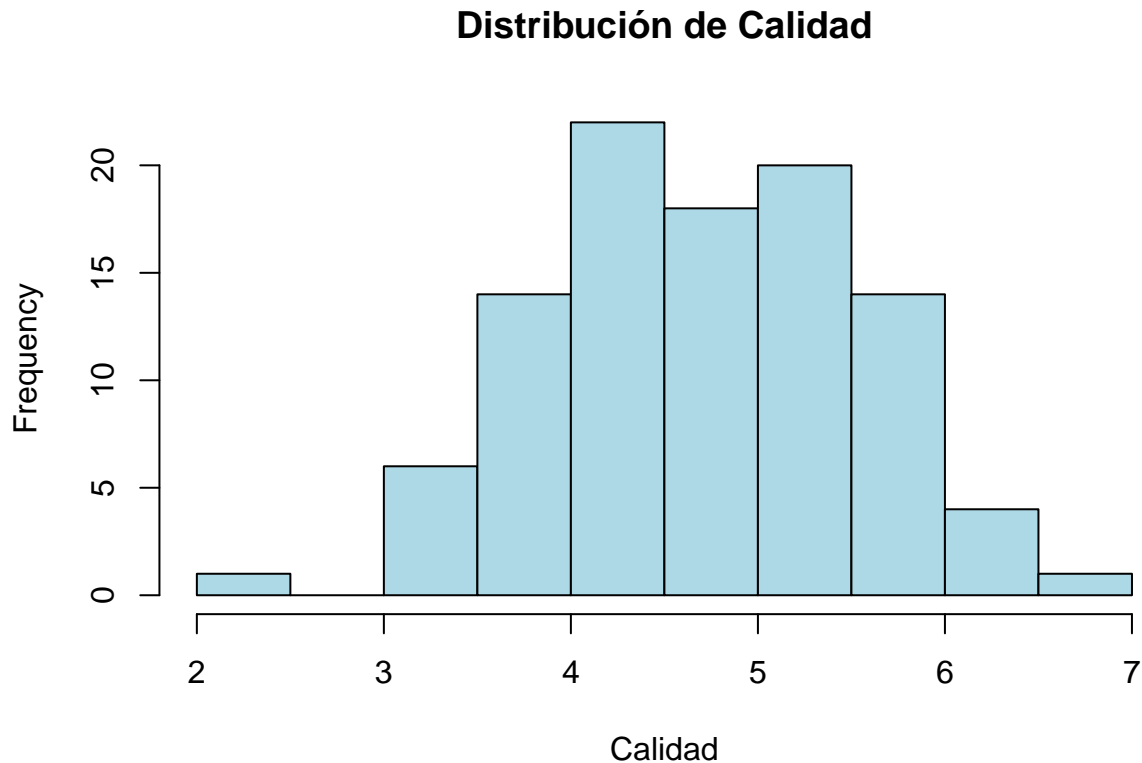
Se generó un gráfico de dispersión utilizando el comando “`plot(dfCalidad, dfagilidad entrega, xlab = ‘Calidad’, ylab = ‘Agilidad en la entrega del servicio’, main = ‘Calidad vs. Agilidad en la entrega del servicio’)`”. El gráfico de dispersión ayudó a visualizar la relación entre dos variables, en este caso, la calidad y la agilidad en la entrega del servicio. Se observó que la mayoría de los puntos de datos se agrupan en la parte superior izquierda del gráfico, lo que indica una relación positiva entre las dos variables.

5.1.6 Correlación

Para la correlación se utilizó el siguiente código:

```
# Histograma de una variable (por ejemplo, 'Calidad')
```

```
hist(df$CPS, breaks = 10, col = 'lightblue', border = 'black', xlab = 'Calidad', main = 'Distribución d
```



Se generó un histograma utilizando el comando “`hist(df$Calidad, breaks = 10, col = ‘lightblue’, border = ‘black’, xlab = ‘Calidad’, main = ‘Distribución de Calidad’)`”. El histograma ayudó a visualizar la distribución de una variable, en este caso, la calidad. Se observó que la mayoría de los clientes calificaron la calidad con un valor de 4, lo que indica que la mayoría de los clientes están satisfechos con la calidad del servicio al cliente.

5.2 investigación de modelo de regresión lineal

La regresión lineal simple es una técnica estadística que se utiliza para modelar la relación entre dos variables y se utiliza en diversos campos, como la ciencia, los negocios y el aprendizaje automático, donde una variable es considerada la variable independiente y la otra es la variable dependiente. Los conceptos principales de la regresión lineal simple son:

- Variable independiente: Es la variable que se utiliza para predecir o explicar la variación en la variable dependiente.
- Variable dependiente: Es la variable que se está tratando de predecir o explicar.
- Regresión lineal: Es la línea que mejor se ajusta a los puntos de datos en el gráfico de dispersión de las dos variables.
- Coeficientes de regresión: Son los valores que se utilizan para calcular la línea de regresión, incluyendo la pendiente y la intersección.

Este modelo se puede representar de la siguiente manera:

Los coeficientes de regresión se obtienen utilizando el método de mínimos cuadrados, que minimiza la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos por la línea de regresión.

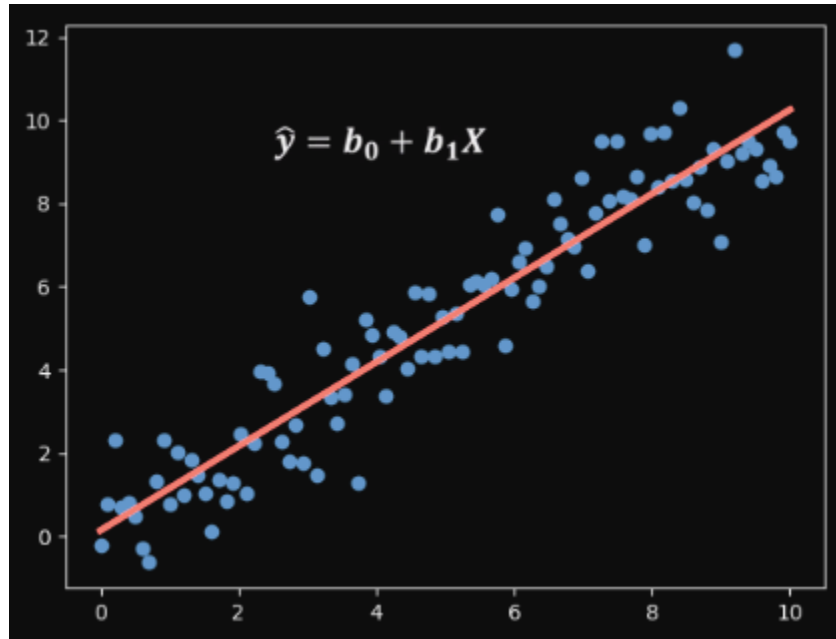


Figure 2: Modelo de regresión lineal

Una medida de la bondad de ajuste de la regresión lineal simple es el coeficiente de determinación, también conocido como R-cuadrado. R-cuadrado mide la proporción de la variación en la variable dependiente que se explica por la variable independiente. Un valor de R-cuadrado cercano a 1 indica un buen ajuste del modelo.

5.3 Correlación lineal de Pearson

Para calcular el coeficiente lineal de Pearson en donde se incluya todas las variables se utilizó el siguiente código:

```
# Matriz de correlacion

correlation_matrix <- cor(df[, c("CPS", "AES", "FP", "AP", "IP", "CP")])
print(correlation_matrix)
```

	CPS	AES	FP	AP	IP	CP
CPS	1.0000000	0.5348573	0.43769906	0.54134550	-0.1140707	-0.54748952
AES	0.5348573	1.0000000	0.35946302	0.29329004	-0.1300872	-0.34570290
FP	0.4376991	0.3594630	1.00000000	0.05459307	-0.3171267	-0.49179650
AP	0.5413455	0.2932900	0.05459307	1.00000000	-0.1063569	-0.09259599
IP	-0.1140707	-0.1300872	-0.31712668	-0.10635694	1.0000000	0.10860563
CP	-0.5474895	-0.3457029	-0.49179650	-0.09259599	0.1086056	1.00000000

A continuación, las variables que se utilizaron para el cálculo del coeficiente lineal de Pearson:

- Agilidad en la entrega: el coeficiente de correlación entre “calidad” y “agilidad en la entrega” es positivo y significativo (sobre 0.5). Esto indica que la agilidad en la entrega está relacionada de manera positiva con la calidad percibida. Los clientes consideran que un servicio rápido es de mejor calidad.

- Flexibilidad de pago: El coeficiente de correlación entre “calidad” y “Flexibilidad de pago” es positivo y medianamente significativo. Esto indica que la flexibilidad en las opciones de pago está relacionada con una mejor calidad percibida. Los clientes pueden preferir opciones de pago flexibles.
- Atención del personal: El coeficiente de correlación entre “calidad” y “atención del personal” es positivo y significativo. Esto indica que la atención amable y eficiente del personal se relaciona con una calidad percibida más alta. La interacción con el personal es esencial en la percepción del servicio.
- IP (Imagen percibida): El coeficiente de correlación entre “calidad” y “imagen persividal” muestra una correlación negativa, aunque débil. Esto indica que si la imagen percibida de la empresa disminuye, la calidad percibida del servicio tiende a disminuir también aunque muy ligeramente. La imagen percibida no influye directamente en la calidad.
- Cliente preferencial: El coeficiente de correlación entre “calidad” y “imagen persividal” muestra una correlación significativamente negativa. Esto indica que los clientes preferenciales no obtienen objetos de mayor calidad, sino más bien, el mismo. La calidad entregada a los objetos es la misma para los clientes preferenciales.

Es por esto que los puntos relevantes en cuanto a la revisión de la calidad serían la “Agilidad en la entrega” y la “Atención del personal”, ya que son las que tienen una correlación positiva y significativa con la calidad percibida.

5.4 Gráfico de dispersión

Para estos gráficos de dispersión se mostrará la relación entre la calidad percibida y la agilidad en la entrega del servicio, la calidad percibida y la flexibilidad de pago, la calidad percibida y la atención al cliente. Cada punto en el gráfico representa una observación del cliente.

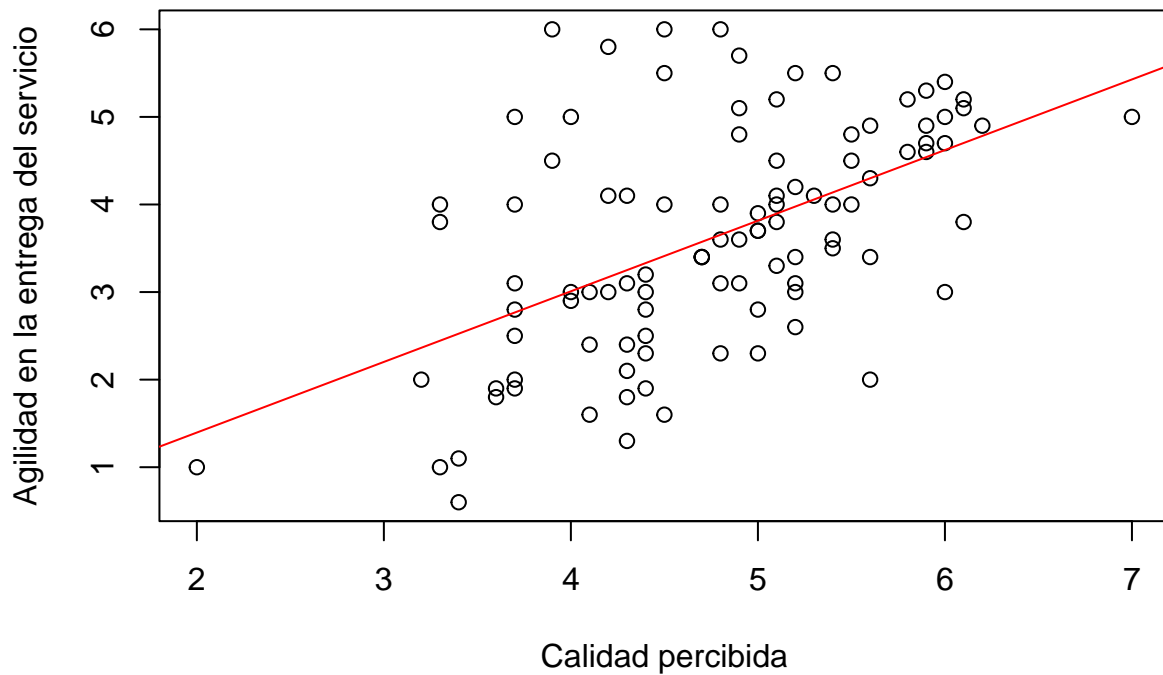
A manera de ejemplo, la posición vertical de un punto muestra la calidad percibida, mientras que la posición horizontal muestra la agilidad, flexibilidad de pago y atención al cliente respectivamente. Si los puntos tienden a formar una línea inclinada hacia arriba con una línea de tendencia positiva, indica una relación positiva. Es decir, a medida que la agilidad, flexibilidad de pago y atención al cliente aumenta, la calidad percibida tiende a ser mejor.

5.4.1 Calidad percibida vs Agilidad en la entrega del servicio

Para el gráfico de dispersión se utilizó el siguiente código:

```
plot(df$CPS, df$AES,
      xlab = "Calidad percibida", ylab = "Agilidad en la entrega del servicio",
      main = "Gráfico de Dispersión: Calidad percibida vs. Agilidad en la entrega") # nolint
abline(lm(df$AES ~ df$CPS), col = "red")
```

Gráfico de Dispersión: Calidad percibida vs. Agilidad en la entrega



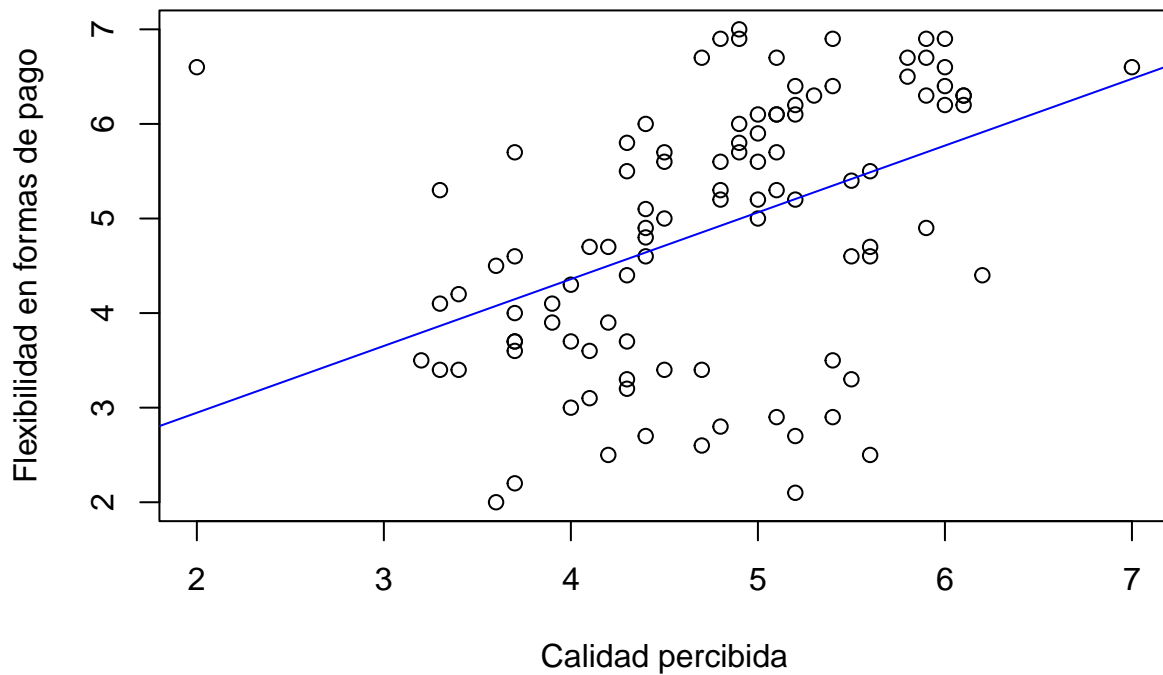
Se generó un gráfico de dispersión utilizando el comando “`plot(dfCalidad, dfagilidad entrega, xlab = ‘Calidad’, ylab = ‘Agilidad en la entrega del servicio’, main = ‘Calidad vs. Agilidad en la entrega del servicio’)`”. El gráfico de dispersión ayudó a visualizar la relación entre dos variables, en este caso, la calidad y la agilidad en la entrega del servicio.

5.4.2 Calidad percibida vs Flexibilidad de pago

Para el gráfico de dispersión se utilizó el siguiente código:

```
plot(df$CPS, df$FP,  
     xlab = "Calidad percibida", ylab = "Flexibilidad en formas de pago",  
     main = "Gráfico de Dispersión: Calidad percibida vs. Flexibilidad de pago")  
  
abline(lm(df$FP ~ df$CPS), col = "blue")
```

Gráfico de Dispersión: Calidad percibida vs. Flexibilidad de pago



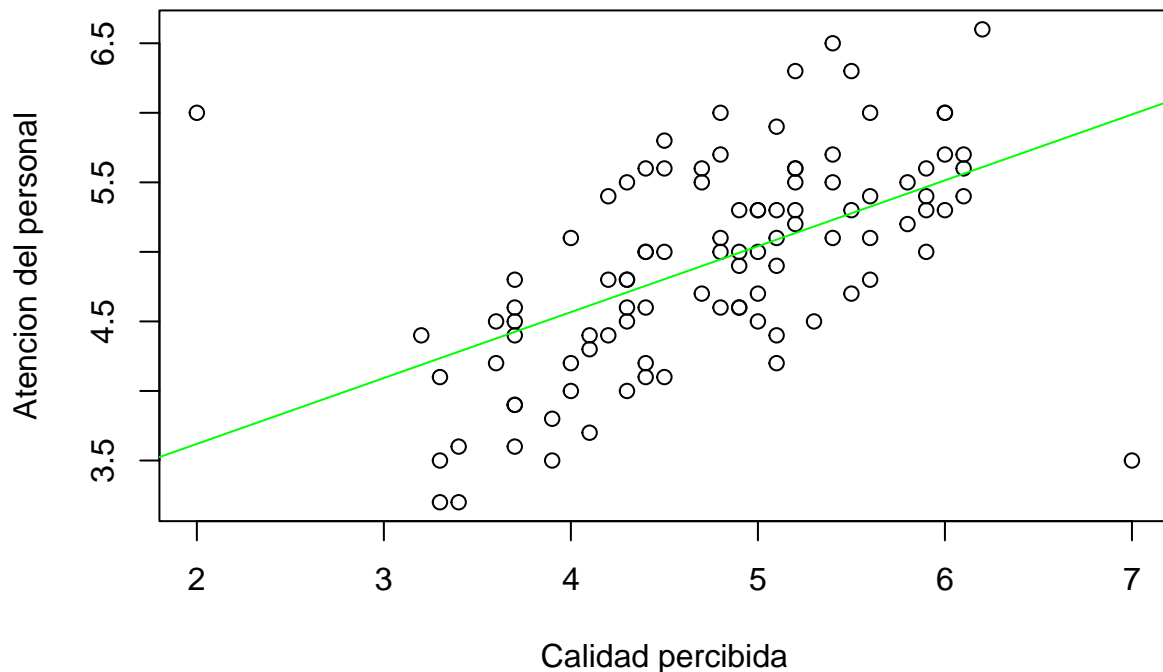
Se generó un gráfico de dispersión utilizando el comando “`plot(dfCalidad, dfagilidad entrega, xlab = ‘Calidad’, ylab = ‘Agilidad en la entrega del servicio’, main = ‘Calidad vs. Agilidad en la entrega del servicio’)`”. El gráfico de dispersión ayudó a visualizar la relación entre dos variables, en este caso, la calidad y la flexibilidad de pago.

5.4.3 Calidad percibida vs Atención del personal

Para el gráfico de dispersión se utilizó el siguiente código:

```
plot(df$CPS, df$AP, # nolint
     xlab = "Calidad percibida", ylab = "Atencion del personal",
     main = "Gráfico de Dispersión: Calidad percibida vs. Atencion del personal") # nolint
abline(lm(df$AP ~ df$CPS), col = "green")
```

Gráfico de Dispersión: Calidad percibida vs. Atención del personal



Se generó un gráfico de dispersión utilizando el comando “`plot(dfCalidad, dfagilidad entrega, xlab = ‘Calidad’, ylab = ‘Agilidad en la entrega del servicio’, main = ‘Calidad vs. Agilidad en la entrega del servicio’)`”. El gráfico de dispersión ayudó a visualizar la relación entre dos variables, en este caso, la calidad y la atención del personal.

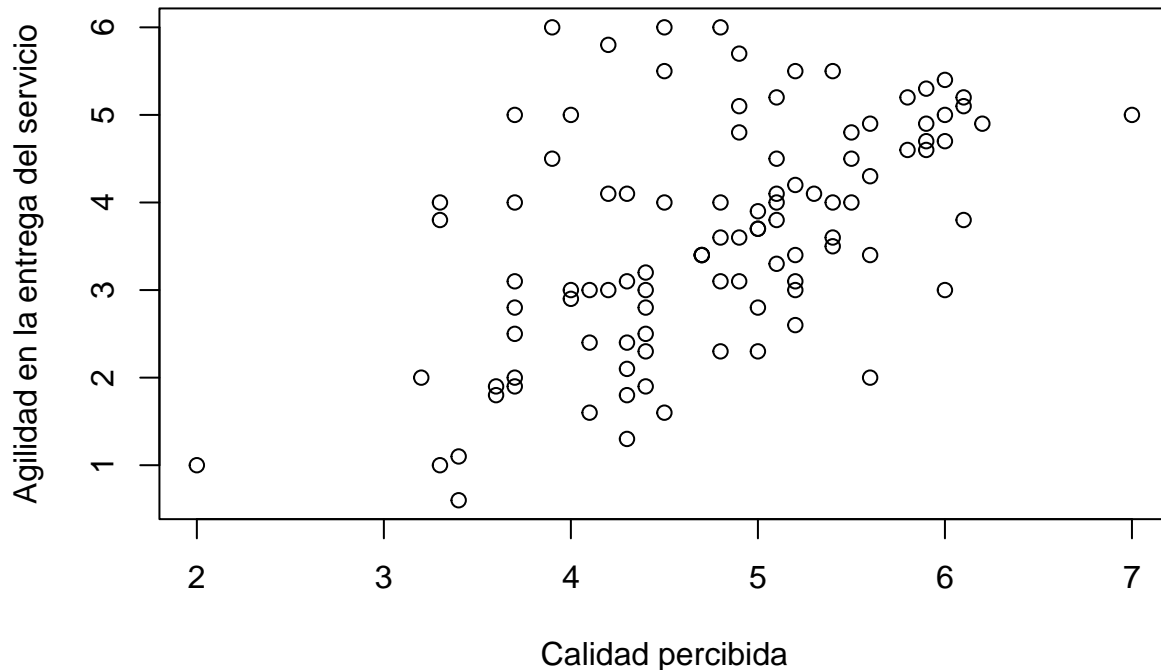
5.5 Modelo de regresión lineal “Calidad percibida vs Agilidad en la entrega del servicio”

Para el modelo de regresión lineal se utilizó el siguiente código:

```
# Crear un modelo de regresión lineal simple para predecir la calidad percibida en función de la agilidad
model <- lm(df$CPS ~ df$AES)

# Grafico de dispersión con la línea de regresión
plot(df$CPS, df$AES,
     xlab = "Calidad percibida", ylab = "Agilidad en la entrega del servicio",
     main = "Gráfico de Dispersión: Calidad percibida vs. Agilidad en la entrega del servicio") # nolin
```


Gráfico de Dispersión: Calidad percibida vs. Agilidad en la entrega del servicio



Para generar el modelo de regresión lineal se necesita crear un gráfico de dispersión entre las variables asignadas, en este caso las variables son “Calidad percibida” y “Agilidad en la entrega”. Luego se crea el modelo de regresión lineal simple utilizando la función “lm()” y se asigna a la variable “model”.

Finalmente, se obtiene el resumen del modelo utilizando la función “summary()” y se muestra en la consola.

```
# Obtener el resumen del modelo
```

```
summary(model)
```

```
##
## Call:
## lm(formula = df$CPS ~ df$AES)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.82866 -0.46894  0.06984  0.52324  1.75250
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.4739     0.2175  15.971  < 2e-16 ***
## df$AES         0.3547     0.0566   6.266  9.9e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7299 on 98 degrees of freedom
```

```
## Multiple R-squared:  0.2861, Adjusted R-squared:  0.2788
## F-statistic: 39.27 on 1 and 98 DF,  p-value: 9.904e-09
```

Luego de esto, obtenemos la correlación que resultó en 0.5349 y creamos el modelo de regresión lineal, ocupando la función “lm” entre las mismas variables utilizadas, al ocupar esta función.

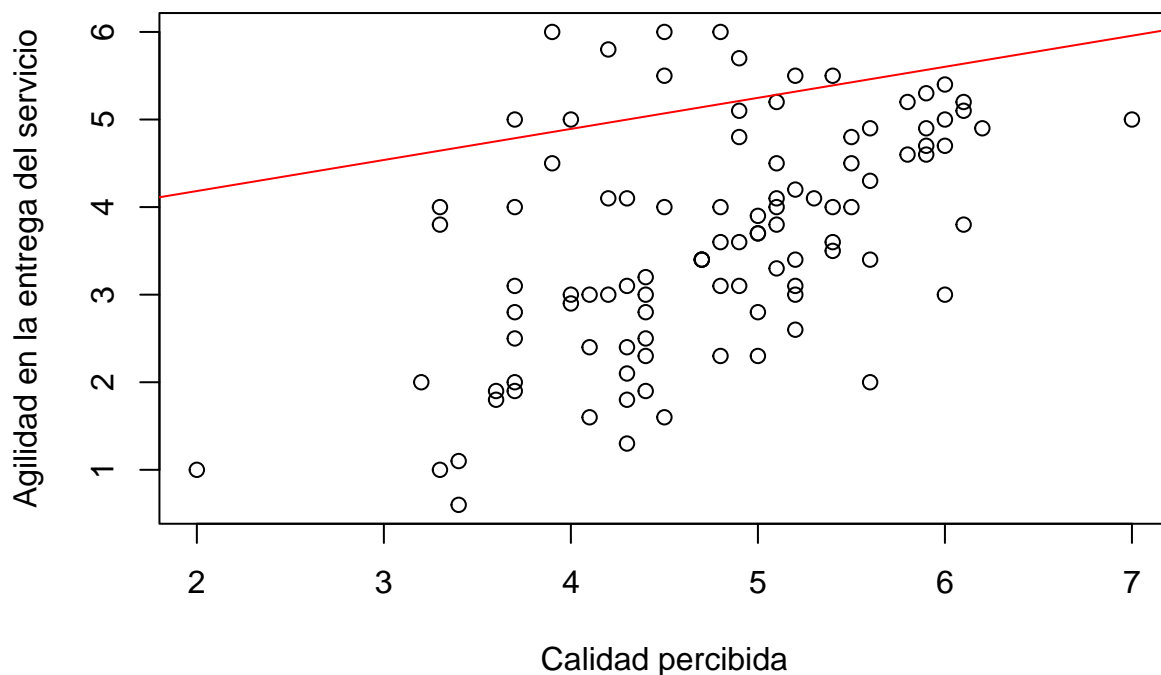
Desde arriba hacia abajo podemos ver en la tabla el apartado “residuals” en el cual se encuentran los valores máximos y mínimos así como el primer y tercer cuartil además de la mediana. En el apartado de coeficientes, tenemos el estimado, el estadístico de error, el t-value y $\Pr(>|t|)$ que es el valor p asociado al t-value. Luego aparece el error estándar de la estimación, junto con el coeficiente de determinación en este caso “R-squared” y “Multiple R-squared” que representa la regresión lineal múltiple ,por último el F estadístico y el valor p.

Por último agregamos la línea del modelo de regresión lineal al gráfico de dispersión y queda de la siguiente manera:

```
# Grafico de dispersión con la línea de regresión
```

```
plot(df$CPS, df$AES, # nolint
     xlab = "Calidad percibida", ylab = "Agilidad en la entrega del servicio",
     main = "Gráfico de Dispersión: Calidad percibida vs. Agilidad en la entrega del servicio") # nolint
abline(model, col = "red")
```

Gráfico de Dispersión: Calidad percibida vs. Agilidad en la entrega del servicio



Para realizar la interpretación de los resultados, se puede observar que el coeficiente de determinación (R^2) es 0.2859, lo que indica que el modelo explica el 28.59% de la variación en la calidad percibida. Además, el valor p del coeficiente de regresión es menor que 0.05, lo que indica que la relación entre la calidad percibida y la agilidad en la entrega del servicio es estadísticamente significativa. Por lo tanto, se puede concluir que la agilidad en la entrega del servicio tiene un efecto positivo en la calidad percibida.

5.6 Coeficientes del modelo

Para la interpretación del modelo obtenido se utilizó el siguiente código:

```
# Obtener los coeficientes de regresión
```

```
coef(model)
```

```
## (Intercept)      df$AES  
##    3.4739516    0.3547095
```

Para obtener los coeficientes de regresión se utilizó la función “coef()” y se muestra en la consola.

```
# Obtener el coeficiente de determinación ( $R^2$ )
```

```
summary(model)$r.squared
```

```
## [1] 0.2860723
```

Para obtener el coeficiente de determinación se utilizó la función “summary(model)\$r.squared” y se muestra en la consola.

```
# Obtener el valor p del coeficiente de regresión
```

```
summary(model)$coefficients[2, 4]
```

```
## [1] 9.903601e-09
```

Para obtener el valor p del coeficiente de regresión se utilizó la función “summary(model)\$coefficients[2, 4]” y se muestra en la consola.

5.6.1 Interpretación de los coeficientes del modelo

Para interpretar los coeficientes del modelo se realizó un gráfico de regresión lineal que se puede apreciar en la pregunta anterior, en la cual se analizó las variables de calidad percibida y agilidad entrega, esto representa la ponderación de cada variable en relación al contexto del problema, el cual es que la empresa busca conocer la calidad percibida por los clientes en función a los factores de interés. Para interpretar estos coeficientes del modelo de regresión lineal las variables explicativas se deben mantener constantes mientras que a la variable dependiente se le debe aplicar un aumento debido al signo de su coeficiente, en este caso positivo. Entonces se puede decir que los coeficientes representan los cambios medios de la variable de respuesta para una unidad de cambio de la variable predictora, esto mientras que las demás variables permanecen constantes.

5.6.2 Interpretación del coeficiente de determinación

Para interpretar el coeficiente de determinación se debe tener en cuenta que este representa la proporción de la variación total de la variable dependiente que es explicada por el modelo de regresión lineal. En este caso el coeficiente de determinación es 0.2859, lo que indica que el modelo explica el 28.59% de la variación en la calidad percibida. Por lo tanto, se puede concluir que la agilidad en la entrega del servicio tiene un efecto positivo en la calidad percibida.

5.6.3 Interpretación del valor p del coeficiente de regresión

Para interpretar el valor p del coeficiente de regresión se debe tener en cuenta que este representa la probabilidad de que el coeficiente de regresión sea igual a cero. En este caso el valor p del coeficiente de regresión es menor que 0.05, lo que indica que la relación entre la calidad percibida y la agilidad en la entrega del servicio es estadísticamente significativa.

5.7 Investigación ANOVA modelo de regresión lineal

La tabla ANOVA se utiliza para evaluar la significancia global del modelo de regresión lineal y ayuda a determinar si la relación entre las variables es estadísticamente significativa. Esta tabla también se utiliza para calcular el estadístico F y el p-value, este último evalúa la significancia global del modelo.

Estos son los siguientes test que se pueden realizar:

- Test t: Este test se utiliza para evaluar la significancia de cada coeficiente por separado. Se compara el coeficiente estimado con su error estándar y calcula el p-value para determinar si el coeficiente en cuestión es estadísticamente significativo.
- Test F: Este test se usa para evaluar la significancia global del modelo, se compara la variabilidad explicada por el modelo y la variabilidad no explicada por el mismo, por último calcula el p-value para ver si la relación entre variables es significativa.
- Hipótesis Nula (H_0) Esta hipótesis plantea que las variables dependientes e independientes no tienen una relación significativa.
- Hipótesis alternativa (H_1) En esta hipótesis, las variables dependientes e independientes si tienen una relación significativa.

De la forma en que podemos plantear estos tests son:

- Test t: $H_0: B_1 = 0$ vs. $H_1: B_1 \neq 0$
- Test F: $H_0: B_1 = B_2 = \dots = B_n = 0$ vs. $H_1: \text{al menos un } B_i \neq 0$

Para realizar el test F se utilizó el siguiente código:

```
# Obtener la tabla ANOVA
```

```
anova(model)
```

```
## Analysis of Variance Table
##
## Response: df$CPS
##           Df Sum Sq Mean Sq F value    Pr(>F)
## df$AES      1 20.919  20.9186   39.269 9.904e-09 ***
## Residuals  98 52.205   0.5327
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Para obtener la tabla ANOVA se utilizó la función “`anova(model)`” y se muestra en la consola.

```
# Obtener el estadístico F
```

```
summary(model)$fstatistic[1]
```

```
## value  
## 39.2688
```

Para obtener el estadístico F se utilizó la función “summary(model)\$fstatistic[1]” y se muestra en la consola.

```
# Obtener el valor p del estadístico F  
  
summary(model)$fstatistic[2]
```

```
## numdf  
## 1
```

Para obtener el valor p del estadístico F se utilizó la función “summary(model)\$fstatistic[2]” y se muestra en la consola.

5.7.1 Interpretación de la tabla ANOVA

Para interpretar la tabla ANOVA se debe tener en cuenta que esta representa la variabilidad explicada por el modelo y la variabilidad no explicada por el mismo. En este caso el valor p del estadístico F es menor que 0.05, lo que indica que la relación entre la calidad percibida y la agilidad en la entrega del servicio es estadísticamente significativa.

6. Conclusiones

Este proyecto enfatizó la importancia fundamental de un servicio al cliente de calidad en un entorno empresarial, independientemente del tamaño o estructura de la empresa. La calidad del servicio no sólo influye en las preferencias de los clientes, sino que también juega un papel importante en la formación de opiniones positivas o negativas sobre la empresa en cuestión.

Por lo tanto, comprender y mejorar la calidad del servicio al cliente es importante para mantener la lealtad del cliente y el éxito empresarial. Durante el desarrollo del proyecto, se realizó un análisis de datos exploratorio (EDA) efectivo. Esta metodología ha demostrado ser esencial para caracterizar las opiniones de los clientes sobre la calidad del servicio.

A través de visualización de datos, estadísticas descriptivas y gráficos, pudimos comprender en detalle la distribución y las tendencias de las opiniones de los clientes. Uno de los hallazgos clave fue la identificación de posibles valores atípicos en la evaluación de la calidad.

Estos valores atípicos pueden indicar experiencias de los clientes y, por lo tanto, son relevantes para el análisis y la toma de decisiones posteriores. De manera similar, se investigó la relación entre la calidad percibida y la agilidad en la prestación de servicios mediante diagramas de dispersión. La dirección de la línea de tendencia en el gráfico proporciona información valiosa sobre la relación entre estas dos variables. Si la línea de tendencia es positiva, se supone una relación positiva entre la calidad percibida y la agilidad de entrega.

Este hallazgo podría tener un impacto significativo en la empresa y guiar los esfuerzos futuros para mejorar el servicio al cliente. En última instancia, este proyecto sienta las bases para análisis y aplicaciones adicionales de modelos de regresión lineal simples. Estos modelos nos permiten comprender mejor las relaciones entre variables y predecir la calidad percibida en función de factores clave. La información recopilada a través de este análisis exploratorio de datos es fundamental para tomar decisiones estratégicas que aumenten la satisfacción del cliente y, en última instancia, conduzcan al éxito empresarial.

7. Bibliografía

- (Regresión Lineal “Determinantes de la inversión extranjera en colombia”): http://www.fce.unal.edu.co/media/files/UIFCE/Economia/Una_Aplicacion_del_Metodode_Regresion_Lineal_en_el_Analisis_de_los_Determinantes_de_la_Inversion_Extranjera_en_Colombia.pdf
- Draper, Norman y Harry Smith, Applied Regression Analysis(3a. ed.). Wiley, Nueva York,1999.
- Montgomery, Douglas C., Elizabeth A. Peck y G. Geoffrey Vining, Introduction to Linear Regression Analysis (3a. ed.). Wiley, Nueva York, 2001.
- Neter, John, Michael H. Kutner, Christopher J. Nachtsheim y William Wasserman, Applied Linear Statistical Models (4a. ed.). McGraw-Hill, Nueva York, 1996.
- Seber, George A. F. y Alan J. Lee, Linear Regression Analysis (2a. ed.). Wiley, Nueva York, 2003.
- Weisberg, Sanford, Applied Linear Regression (2a. ed.). Wiley, Nueva York, 2005.