**Preparado: José Tobias Agudelo Gutierrez, CC. 3015842**

**Grupo: AVDV-88**

**Modulo-2 Tarea-1**

# **Florence Nightingale y sus métodos estadísticos**

Florence Nightingale (1820-1910)

Nacionalidad inglesa de la era Victoriana impulsó la reforma hospitalaria y la profesión moderna de enfermería. pionera en el análisis estadístico de enfermedades, base de la epidemiología. En este campo, los modelos matemáticos se usan para seguir la evolución de enfermedades, tales como el COVID-19, donde se habla de la pendiente de las curvas, se siguen las tendencias y se hacen predicciones de escenarios futuros.

|  |
| --- |
|  |
| Este diagrama de "ala de murciélago" fue precursor de los "gráficos de rosa", y compara las causas de muerte de los soldados ingleses durante la guerra de Crimea entre abril 1854 y marzo 1855, por mes (derecha), y los fallecimientos entre abril 1855 a marzo 1856 (izquierda). El color verde representa las muertes debidas a enfermedades, rojo son las heridas y gris son todas las causas. El largo de cada línea radial es proporcional a la tasa de mortalidad para ese mes, con un pico en muertes en enero de 1855. ELMER BELT FLORENCE NIGHTINGALE COLLECTION/UNIV. OF CALIFORNIA LIBRARIES/ARCHIVE.ORG |

Nightingale fue innovadora en la recolección, tabulación, interpretación y presentación gráfica de las estadísticas descriptivas; mostró como la estadística proporciona un marco de organización para controlar y aprender, y puede llevar a mejoramientos en las prácticas quirúrgicas y médicas. También desarrolló una Fórmula Modelo de Estadística Hospitalaria para que los hospitales recolectaran y generaran datos y estadísticas consistentes. Compara la tasa de mortalidad durante el tiempo de paz de los civiles ingleses con los soldados ingleses en cuatro grupos etario.

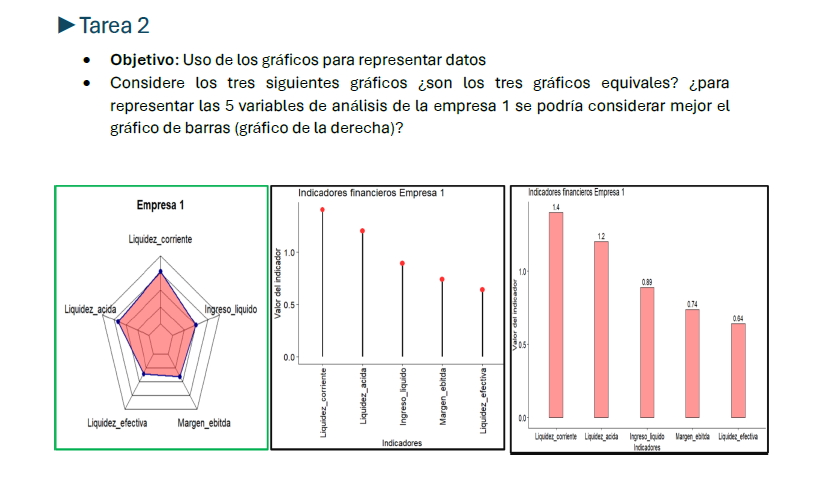
**Referencias**

* <https://www.cientec.or.cr/articulos/florence-nightingale-y-sus-metodos-estadisticos>
* Larousse, Dictionary of Women, New York, 1996
* Cohen, I.B. "Florence Nightingale", Scientific American, 250 (March 1984), 128-137
* Stinnett, S. "Women in Statistics: Sesquicentennial Activities," The American Statistician, May 1990, Vol 44, No. 2, 74-80.
* Kopf, E.W. "Florence Nightingale as a Statistician," J. Amer. Statist. Assoc., 15 (1916), 388-404.
* Nuttall, P. "The Passionate Statistician," Nursing Times, 28 (1983), 25- 27.

**Preparado: José Tobias Agudelo Gutierrez, CC. 3015842**

**Grupo: AVDV-88**

**Modulo-2 Tarea-2**



**El Objetivo:** Uso de los gráficos para representar datos

**Gráfico de radar (Gráfico de la izquierda):**

El gráfico de radar o gráfico de araña muestra cinco indicadores financieros para la Empresa 1: Liquidez corriente, Liquidez ácida, Ingreso líquido, Margen ebitda y Liquidez efectiva.

Este tipo de gráfico es útil para mostrar datos multivariados de una manera que resalta las fortalezas y debilidades en múltiples categorías a la vez. Es particularmente bueno para ver patrones generales y comparar múltiples variables.

**Gráfico de puntos (Gráfico del medio):**

El gráfico de puntos muestra los mismos cinco indicadores financieros. Cada indicador está representado por un punto en un eje vertical.

Este gráfico proporciona una visualización clara del valor de cada indicador, facilitando la comparación de puntos individuales a través de las categorías.

**Gráfico de barras (Gráfico de la derecha):**

El gráfico de barras también representa los mismos cinco indicadores financieros. Cada barra corresponde a un indicador diferente, y la altura de la barra muestra el valor del indicador.

Los gráficos de barras son directos y efectivos para comparar la magnitud de diferentes categorías. Son fáciles de leer e interpretar, especialmente para comparar los valores de diferentes variables.

Los tres gráficos representan el mismo conjunto de datos, mostrando los cinco indicadores financieros de la Empresa 1.

Sin embargo, presentan los datos en formatos visuales diferentes.

El gráfico de barras generalmente es una mejor opción para representar estas cinco variables porque es directo y fácilmente interpretable. Permite una comparación rápida entre los valores de los diferentes indicadores.

**Recomendación:** El mejor grafico es el de la derecha, muestra Indicadores financieros de la Empresa-1 en las dos dimensiones (x,y) o sea (indicador, valor del indicador) y más fácil su comprensión para una audiencia general.

**Preparado: José Tobias Agudelo Gutierrez, CC. 3015842**

**Grupo: AVDV-88**

**Modulo-2 Tarea-3**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**►Tarea 3**

* **Objetivo**: Historia de la estadística
* ¿Es posible hacer inferencia a partir de gráficos estadísticos? (ejemplo de gráficos: diagramas de cajas y bigotes, histogramas, barras).

**Respuesta**. ¿Si es posible hacer inferencia a partir de gráficos estadísticos?

### **Inferencia a partir de gráficos estadísticos**

Los gráficos estadísticos son herramientas visuales poderosas que permiten hacer inferencias sobre los datos de manera más intuitiva y comprensible. A través de estos gráficos, podemos observar patrones, tendencias y anomalías que nos ayudan a tomar decisiones informadas. A continuación, se proporciona una explicación detallada sobre cómo se puede hacer inferencia a partir de gráficos estadísticos, con un ejemplo específico para una audiencia experta en estadística.

#### **Ejemplo: Comparación de Distribuciones con Diagramas de Cajas y Bigotes**

Supongamos que estamos analizando los salarios de empleados en dos departamentos diferentes de una empresa: Departamento A y Departamento B. Queremos inferir si hay una diferencia significativa en la distribución de los salarios entre estos dos departamentos.

#### **1. Diagramas de Cajas y Bigotes (Boxplots)**

Los diagramas de cajas y bigotes (boxplots) son útiles para resumir la distribución de un conjunto de datos a través de sus cuantiles. Nos permiten comparar rápidamente las distribuciones entre diferentes grupos.

* **Mediana (Q2):** La línea dentro de la caja representa la mediana del conjunto de datos.
* **Primer y Tercer Cuartil (Q1 y Q3):** Los bordes de la caja representan el primer y tercer cuartil.
* **Rango Intercuartílico (IQR):** La longitud de la caja (Q3 - Q1) muestra la dispersión de la mitad central de los datos.
* **Bigotes:** Extienden hasta el valor máximo y mínimo dentro de 1.5 veces el IQR desde los cuartiles.
* **Outliers:** Puntos individuales que se encuentran fuera de los bigotes.

#### **Visualización:**

Imaginemos los siguientes boxplots para los salarios en los departamentos A y B:

less

Copiar código

Departamento A:

|--+--+--+--+--+--+--+--+--+--+--+--+--+--|

| 20 30 40 50 60 70 80 90 100 110 |

| [A] |

| [B] |

| [C] |

| [D] |

| |-----------| |

| |-----------| |

| |-----------| |

| 50 70 |

Departamento B:

|--+--+--+--+--+--+--+--+--+--+--+--+--+--|

| 20 30 40 50 60 70 80 90 100 110 |

| [E] |

| [F] |

| [G] |

| [H] |

| [I] |

| |-----------| |

| |-----------| |

| |-----------| |

| 40 60 |

#### Inferencia:

1. **Comparación de Medianas:**
   * La mediana de los salarios en el Departamento A es de aproximadamente 60, mientras que en el Departamento B es de aproximadamente 50. Esto sugiere que los salarios medianos son mayores en el Departamento A.
2. **Dispersión:**
   * La dispersión en los salarios del Departamento A es mayor (IQR más amplio) que en el Departamento B, indicando una mayor variabilidad en los salarios del Departamento A.
3. **Outliers:**
   * Ambos departamentos tienen outliers, pero el Departamento A parece tener más outliers hacia el extremo superior, lo que podría indicar la presencia de algunos salarios significativamente más altos.
4. **Superposición:**
   * Observamos que los rangos intercuartílicos (IQR) de los dos departamentos no se superponen. Esto puede ser un indicio de una diferencia significativa en las distribuciones de los salarios entre los dos departamentos.

#### **Conclusión:**

A partir de los boxplots, podemos inferir que el Departamento A tiene salarios medianos más altos y una mayor variabilidad en comparación con el Departamento B. La falta de superposición en los IQR también sugiere que estas diferencias podrían ser estadísticamente significativas. Para confirmar esta inferencia visual, podríamos realizar una prueba estadística formal, como una prueba de Mann-Whitney U, para evaluar si las diferencias observadas en las distribuciones son significativas.

Este ejemplo muestra cómo los gráficos estadísticos, como los diagramas de cajas y bigotes, no solo facilitan la visualización de datos complejos, sino que también proporcionan una base sólida para hacer inferencias estadísticas sobre las diferencias y características de las distribuciones de datos.

**Tarea-4**

* **Objetivo**: Cálculo de estadísticos descriptivos
* Complete la siguiente tabla. Indique si los valores estadísticos cambian en el tiempo. Los datos fueron entregados en clase.



**Preparado: José Tobias Agudelo Gutierrez, CC. 3015842**

**Grupo: AVDV-88**

**Modulo-2 Tarea-5**

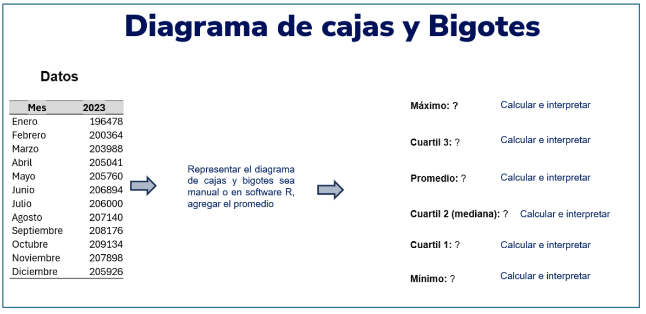
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**►Tarea 5**

**• Objetivo:** Uso del diagrama de cajas y bigotes y su interpretación

• Tome los datos del año 2023 para que analice el número de asociados mediante el Boxplot y las diferentes medidas estadísticas que le corresponden, sea en R o manualmente.

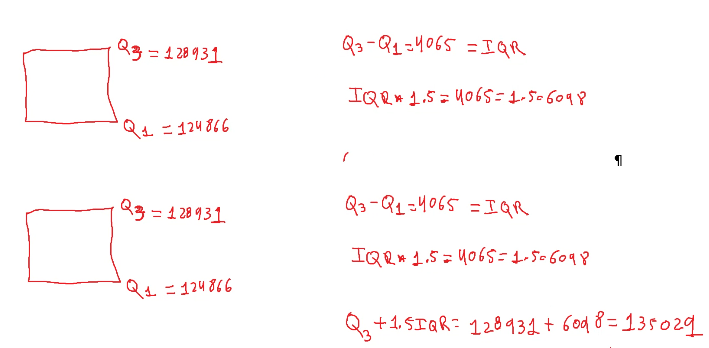
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_



**Datos Afiliación del año 2023**

| **Mes** | **Número de personas** | **Concepto** | **Resultados 2023** | |
| --- | --- | --- | --- | --- |
| Enero | 196478 | **Medidas de posición o también de orden** | **Mínimo** | **196478** |
| Febrero | 200364 | **Máximo** | **209134** |
| Marzo | 203988 | Rango | 12656 |
| Abril | 205041 | Rango intercuartil | 2741 |
| Mayo | 205760 | **Medidas de tendencia central** | **promedio aritmético** | **205233** |
| Junio | 206894 | M.geometrica\_1 | 205204 |
| Julio | 206000 | M.armónica\_1 | 205175 |
| Agosto | 207140 | mediana | 205963 |
| Septiembre | 208176 | Promedio datos agrupados |  |
| Octubre | 209134 | Moda |  |
| Noviembre | 207898 | **Medidas de posición** | Percentile 10 | 200726 |
| Diciembre | 205926 | Percentile 15 | 202720 |
|  |  | Percentile 25 | 204778 |
|  |  | Percentile 75 | 207330 |
|  |  | **Cuartil 1** | **204778** |
|  |  | **Cuartil 2** | **205963** |
|  |  | **Cuartil 3** | **207519** |
|  |  | **Medidas de dispersión** | Varianza | 11722266 |
|  |  | Cuasi-Varianza | 12787926 |
|  |  | Desviación típica | 3424 |
|  |  | Cuasi-Desviación típica | 3576 |
|  |  | Coeficiente de variación | -1,6 |
|  |  | **Medidas de forma** | Sesgo | -1,4 |
|  |  | Curtosis | 2,5 |

**Diagramas de Cajas y Bigotes**



**Fuente: Referencia clase 2024-06-07 Modulo2 Estadística -Tech**

**Datos ejercicio de nuevos asociados de la cooperativa XX para el periodo 2023**

Q3 = 207519

Q1 = 204778

IQR = (Q3 – Q1) = (207519 – 204778) = 2741

IQR =2741

Ref1= IQR\*1.5 = 2741\*1.5 = 4112

BigoteS= Q3 + IQR\*1.5; 207519 + IQR\*1.5 = 207519 + 4112 =

**BigoteS= 211631**

BigoteI = Q1 - IQR\*1.5; 204778 – 4112 = 200666

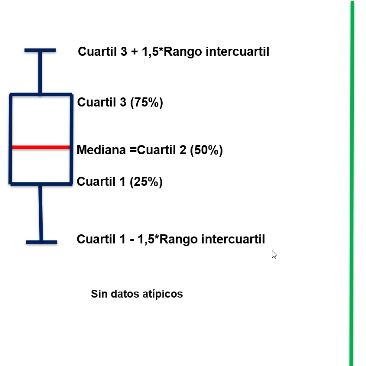
**BigoteI= 200666**

**Promedio** = **205233**

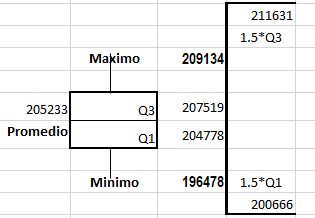
**Mediana: 205926**. Ordena los datos de menor a mayor o viceversa y elige justo el número que se encuentre en la mitad de todos.

|  |  |  |
| --- | --- | --- |
| **Item** | **Mes** | **Número de personas** |
| 1 | Enero | 196478 |
| 2 | Febrero | 200364 |
| 3 | Marzo | 203988 |
| 4 | Abril | 205041 |
| 5 | Mayo | 205760 |
| 6 | Diciembre | 205926 |
| 7 | Julio | 206000 |
| 8 | Junio | 206894 |
| 9 | Agosto | 207140 |
| 10 | Noviembre | 207898 |
| 11 | Septiembre | 208176 |
| 12 | Octubre | 209134 |

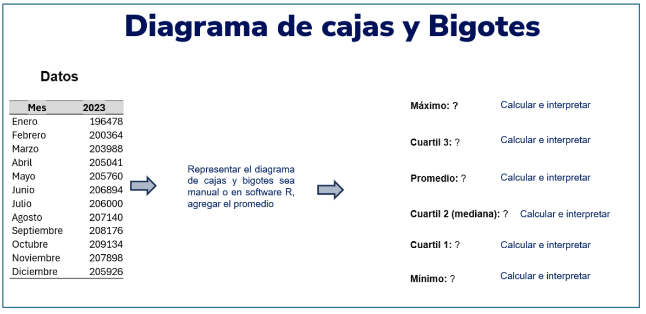
**Diagrama de Cajas y bigotes (boxplot)**



**Diagrama de Cajas y bigotes (boxplot)**



**Interpretación de los datos asociados Entidad cooperativa XX para el año 2023**



**Máximo**: **209134,** es el número máximo de asociados durante los doce meses fue 20914, que ocurrió en el mes de octubre del 2023.

**Mínimo:196478,** es el número mínimo de asociados durante los doce meses fue 196478, que ocurrió en el mes de enero del 2023.

**Cuartil\_3**: **207519,** es el 75% de los meses (o en 9 de los meses) el número de asociados fue menor a 207519.

**Promedio:** **205233**. En promedio el número de asociados fue de 205233 para el año 2023, también conocido como media.

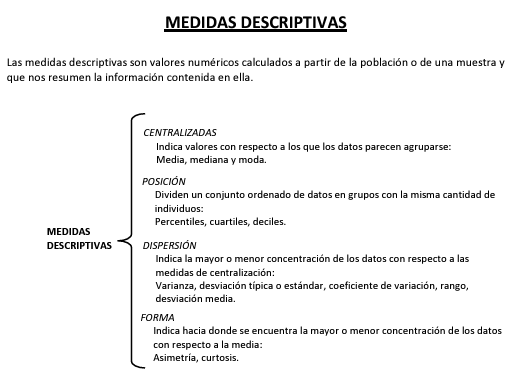
**Cuartil\_2 (Mediana):** Es el 50% de los meses (o en 6 de los meses) el numero de asociados fue menor a 205926.

Corresponde a la mitad de los datos, conocido como mediana, es el 50% de la concentración de los datos.

**Mediana: 205926**. Ordena los datos de menor a mayor o viceversa y elige justo el número que se encuentre en la mitad de todos.

**Cuartil\_1**: **204778,** En el 25% de los meses (o en 3 de los meses) el número de asociados fue menor a 204778.

**Gráfico de Caja y Bigotes (Bigotes)** representa la mediana, los cuartiles el promedio y los valores atípicos de las edades.



**Preparado: José Tobias Agudelo Gutierrez, CC. 3015842**

**Grupo: AVDV-88**

**Modulo-2 Tarea-6**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

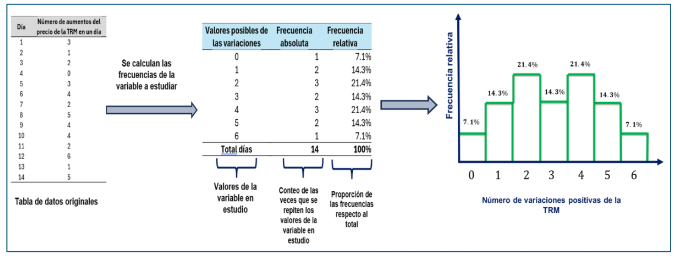
**Tarea 6**

• Objetivo: Inferencia a partir de la distribución de una variable (uso del histograma)

• De acuerdo con los datos de la siguiente tabla, se estudia la variable “Número de variaciones positivas durante un día de la TRM (Tasa Representativa del Mercado)”.

Respecto a estos datos suponga que, Juan compra y vende dólares todo el día, todos los días, su ganancia está cuando vende a un precio alto la TRM. ¿Es aconsejable que Juan continue en este negocio por los próximos meses? Justifique su respuesta.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_



**Recomendación**

Los valores positivos de las variaciones en la TRM, no son suficientes para aconsejar a JUAN en continuar con el negocio de compra y venta de dólares USD porque existen otros factores o variables de mercado, políticos y económicos que no se están validando por lo tanto NO recomiendo esta negociación es muy riesgoso.

**Preparado: José Tobias Agudelo Gutierrez, CC. 3015842**

**Grupo: AVDV-88**

**Modulo-2 Tarea-7**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Tarea 7**

**• Objetivo:** Inferencia a partir de la regresión lineal

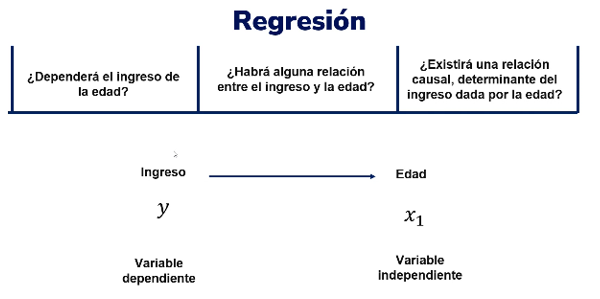
**•** En la sesión de clase número 7 se hizo un ejercicio de brecha de genero con regresión lineal, se estimó el modelo y se observó que las mujeres ganaban en promedio menos que los hombres. Ahora bien, se pidió que teniendo en cuenta la edad analice si por la edad hay brecha salarial de género. No realice una nueva regresión, solo diligencie la siguiente tabla y realice el análisis respectivo. Deberá calcula la brecha salarial en dinero y en porcentaje. En las diapositivas de la clase 7 están las fórmulas y en los archivos los datos de las 540 personas del estudio.

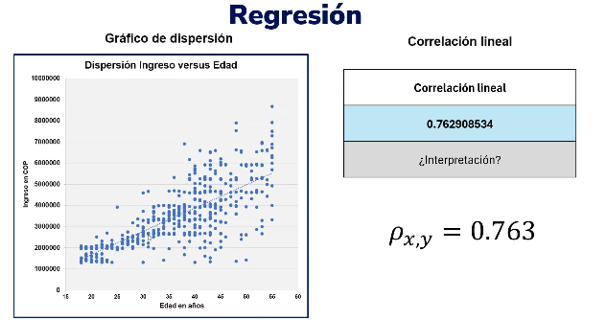
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

| **Frecuencias por edad** | | | **Salario Promedio** | | **Brecha salarial entre hombres y mujeres** | |
| --- | --- | --- | --- | --- | --- | --- |
| **Frecuencia Hombres** | **Frecuencia Mujeres** | **Edad** | **Hombre** | **Mujer** | **Brecha en $** | **Brecha en %** |
| 8 | 8 | 18 | 1767500 | 1641250 | 126250 | 7,14% |
| 10 | 5 | 19 | 1773000 | 1654400 | 118600 | 6,69% |
| 7 | 8 | 20 | 1787143 | 1680000 | 107143 | 6,00% |
| 4 | 7 | 21 | 1832500 | 1793143 | 39357 | 2,15% |
| 8 | 5 | 22 | 1841250 | 1818000 | 23250 | 1,26% |
| 4 | 6 | 23 | 1847500 | 1835000 | 12500 | 0,68% |
| 8 | 4 | 24 | 2099228 | 1915000 | 184228 | 8,78% |
| 3 | 5 | 25 | 2640000 | 1956410 | 683590 | 25,89% |
| 7 | 4 | 26 | 2225885 | 2196250 | 29635 | 1,33% |
| 7 | 8 | 27 | 2342916 | 2305568 | 37348 | 1,59% |
| 4 | 5 | 28 | 2482617 | 2370000 | 112617 | 4,54% |
| 5 | 5 | 29 | 2770000 | 2384679 | 385321 | 13,91% |
| 8 | 10 | 30 | 2978325 | 2690000 | 288325 | 9,68% |
| 3 | 7 | 31 | 3113333 | 2878571 | 234762 | 7,54% |
| 8 | 7 | 32 | 3416250 | 2881429 | 534821 | 15,66% |
| 8 | 6 | 33 | 3197500 | 2775000 | 422500 | 13,21% |
| 6 | 11 | 34 | 3214667 | 2974545 | 240121 | 7,47% |
| 7 | 10 | 35 | 3517026 | 3182000 | 335026 | 9,53% |
| 13 | 13 | 36 | 3631829 | 3365385 | 266444 | 7,34% |
| 17 | 4 | 37 | 3664843 | 3267500 | 397343 | 10,84% |
| 18 | 10 | 38 | 3717521 | 3481000 | 236521 | 6,36% |
| 10 | 7 | 39 | 4045000 | 3751429 | 293571 | 7,26% |
| 16 | 8 | 40 | 4104922 | 3778750 | 326172 | 7,95% |
| 7 | 9 | 41 | 4184286 | 3832222 | 352063 | 8,41% |
| 11 | 9 | 42 | 4207273 | 3907778 | 299495 | 7,12% |
| 16 | 12 | 43 | 4263750 | 4069076 | 194674 | 4,57% |
| 7 | 7 | 44 | 4304286 | 4087143 | 217143 | 5,04% |
| 10 | 11 | 45 | 4350000 | 4104545 | 245455 | 5,64% |
| 5 | 2 | 46 | 4409200 | 4225000 | 184200 | 4,18% |
| 6 | 1 | 47 | 4713333 | 4410000 | 303333 | 6,44% |
| 8 | 5 | 48 | 5130000 | 4424000 | 706000 | 13,76% |
| 2 | 5 | 49 | 5130000 | 4556000 | 574000 | 11,19% |
| 2 | 6 | 50 | 5240000 | 4278333 | 961667 | 18,35% |
| 4 | 3 | 51 | 5275000 | 5036667 | 238333 | 4,52% |
| 7 | 4 | 52 | 5311429 | 4460000 | 851429 | 16,03% |
| 5 | 6 | 53 | 5622000 | 4728333 | 893667 | 15,90% |
| 2 | 3 | 54 | 6615000 | 4876667 | 1738333 | 26,28% |
| 9 | 4 | 55 | 6747778 | 5202500 | 1545278 | 22,90% |
| 290 | 250 |  |  |  |  |  |

**Conclusión**

Al revisar la tabla, se encuentra que las brechas salariales por edad entre hombre y mujeres existe para todos los rangos de edad (18 a 55 años) entre los hombres a las mujeres y en los porcentajes entregados por la tabla.





Es una relación lineal fuerte, porque el factor de relación se encuentra cerca 1. (0.763)

Interpretación del Coeficiente de Correlación de 0.7629

Un coeficiente de correlación de

𝑟 = 0.7629

r=0.7629 se interpreta de la siguiente manera:

Magnitud de la Correlación:

0.0 a 0.1: Correlación despreciable.

0.1 a 0.3: Correlación pequeña.

0.3 a 0.5: Correlación moderada.

0.5 a 0.7: Correlación considerable.

0.7 a 0.9: Correlación fuerte.

0.9 a 1.0: Correlación muy fuerte

# **Tarea 8**

* + **Objetivo**: Inferencia a partir de la regresión logística
  + En la sesión de clase número 8 se hizo un ejercicio de regresión logística. Se habló de los odd ratios, frente a ellos se pide:

1. Se HP el evento que indica que una empresa tiene la Habilidad de Pagar sus deudas, la probabilidad de que este evento ocurra es 0.679. Se pide que calcule el 𝑂𝑑𝑑(𝐻𝑃) y lo interprete.

Habilidad de Pagos y su probabilidad

Esto quiere decir que la probabilidad de que la empresa realice pagos es 2.115 veces mayor que la probabilidad de que no los realice. Un odd ratio mayor a 1 indica una mayor probabilidad de que ocurra el evento de realizar pagos.

1. Se DF el evento que indica que una empresa puede quebrar y cerrar, la probabilidad de que este evento ocurra es 0.2312. Se pide que calcule el

𝑂𝑑𝑑(𝐷𝐹) y lo interprete.

Lo anterior indica que la probabilidad de que la empresa pueda quebrar debido al evento DF es aproximadamente 0.3007 veces la probabilidad de que la empresa pueda quebrar sin estar expuestos a ese evento DF.

* + En las notas de clase se indicó una base de datos que tiene la información de la liquidez de 60 empresas colombianas. Existe una columna de datos que indica si la empresa tiene **buen o mal** estado de la liquidez. Se pide que analice si las variables “Liquidez corriente”, “Liquidez inmediata” y el “Sector” pueden ser determinantes para explicar la variable de clasificación “Estado Liquidez”, justifique su respuesta.

Analizando las variables y haciéndonos la pregunta si, ¿las variables Liquidez inmediata y Liquidez corriente, están relacionadas? Decidimos hacer una gráfica de dispersión y calcular su regresión lineal.

Nos damos cuenta de que tienen una tendencia y el valor de R^2 nos indica que podemos relacionarlas mediante una ecuación de tipo lineal.

Así mismo cuando la liquidez inmediata es mayor a 0.5 es posible concluir que el Estado de liquidez pasa a determinarse como Buena.

Con respecto al Sector no podemos encontrar una relación que determine el mencionado Estado de Liquidez.