



**MSIS545L 232-711**

**Prof. Gissella Bejarano**

## **FINAL PROJECT REPORT**

### **GENDER PAY ANALYSIS**

*Submitted by*

*Keerthija Vegi (20150979)*

*Vamsi Priyanka Vallamreddy (20148180)*

#### **Introduction:**

In our project, we're using RStudio to examine the pay gap between men and women over several decades. We obtained our data from the “United States Census Bureau”, which provides information on median earnings for both total workers and full-time workers from 1960 to 2022, categorized by gender. This dataset is split into two sections: one for total workers and another for full-time workers.

#### **Methodology:**

In our two-sample hypothesis testing, we aim to investigate the presence of a pay gap between men and women.

Null Hypothesis ( $H_0$ ): There is no pay gap between men and women.

Alternative Hypothesis ( $H_a$ ): There is a pay gap between men and women.

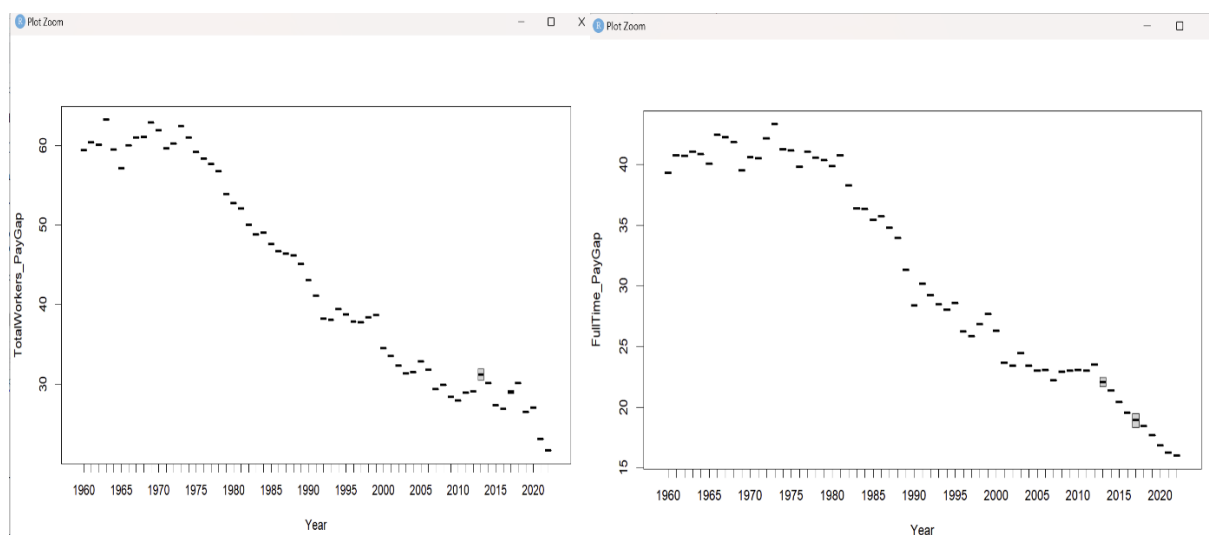
First, we fetched the data from an online source and imported it into RStudio. Then, we assigned variable names to each set of earnings data, distinguishing between total workers and full-time workers.

```
Total_Workers_data <- read_excel("C:\\Intro to DataAnalytics\\Final  
project\\Total_Workers.xlsx")  
Total_male_earnings <- Total_Workers_data$`Median earnings (dollars)/Male`  
Total_female_earnings <- Total_Workers_data$`Median Earnings(dollars)/Female`  
Year <- Total_Workers_data$Year
```

```
FullTime_Workers_data<-read_excel("C:\\Intro to DataAnalytics\\Final
project\\FullTime_Workers.xlsx")
FullTime_Male_Earnings<- FullTime_Workers_data$`Male Earnings(Dollars)`
FullTime_Female_earnings<- FullTime_Workers_data$`Female Earnings(dollars)`
```

Next, we computed the gender pay gap for both datasets using the following formula and visualized the differences using boxplots.

$$\text{Gender Pay Gap} = \frac{\text{Male Earnings} - \text{Female Earnings}}{\text{Male Earnings}} * 100$$



From the plots, it's evident that there exists a gender pay gap, and notably, it has experienced a significant decrease over the years, following an exponential trend.

## Analysis:

To further substantiate this analysis, t-test evaluations were performed on the earnings of both male and female workers for both total and full-time employment scenarios. Below are the outcomes obtained from these analyses.

R 4.3.2 · C:/Intro to DataAnalytics/Final project/ProjectGenderpay/ ↗

```
> t.test(Total_male_earnings, Total_female_earnings)

Welch Two Sample t-test

data: Total_male_earnings and Total_female_earnings
t = 17.933, df = 93.448, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 17598.45 21980.93
sample estimates:
mean of x mean of y
 46694.15  26904.46

> t.test(FullTime_Male_Earnings, FullTime_Female_earnings)

Welch Two Sample t-test

data: FullTime_Male_Earnings and FullTime_Female_earnings
t = 15.779, df = 108.23, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 15507.77 19963.61
sample estimates:
mean of x mean of y
 58613.69  40878.00
```

e

```
> TotalWorkers_MeanPayGap <- t.test(Total_male_earnings)$estimate - t.test(Total_female_earnings)$estimate
> TotalWorkers_MeanPayGap
mean of x
 19789.69
> FullTime_MeanPayGap <- t.test(FullTime_Male_Earnings)$estimate - t.test(FullTime_Female_earnings)$estimate
> FullTime_MeanPayGap
mean of x
 17735.69
```

The t-test analyses conducted on both total workers and full-time workers' earnings revealed statistically significant differences in mean earnings between males and females. For total workers, the mean earnings for males were approximately \$19,790 higher than for females, with a p-value well below the significance threshold. Similarly, for full-time workers, the mean earnings for males exceeded those of females by approximately \$17,736, again with a highly significant p-value. These results provide strong evidence supporting the presence of a gender pay gap in both total and full-time employment scenarios, indicating that males tend to earn significantly more than females on average.

Additionally, to strengthen the validity of our hypothesis, ANOVA tests were conducted on both total workers' earnings and full-time workers' earnings, as illustrated below.

```
> lm_TotalWorkers_data <- lm(Total_male_earnings~Total_female_earnings, data=Total_workers_data)
> TotalWorkers_anovaResult <- anova(lm_TotalWorkers_data)
> TotalWorkers_anovaResult
Analysis of Variance Table

Response: Total_male_earnings
              Df    Sum Sq  Mean Sq F value    Pr(>F)
Total_female_earnings  1 697657336 697657336  148.95 < 2.2e-16 ***
Residuals              63 295075043  4683731
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> lm_FullTime <- lm(FullTime_Male_Earnings ~ FullTime_Female_earnings, data = FullTime_workers_data)
> FullTime_anova_result <- anova(lm_FullTime)
> FullTime_anova_result
Analysis of Variance Table

Response: FullTime_Male_Earnings
              Df    Sum Sq  Mean Sq F value    Pr(>F)
FullTime_Female_earnings  1 1032407081 1032407081  137.73 < 2.2e-16 ***
Residuals              63  472238633  7495851
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The results from the ANOVA tests conducted on both total workers and full-time workers' earnings demonstrate statistically significant differences in mean earnings between males and females. For both scenarios, the p-values associated with the F-statistics are exceedingly small, indicating strong evidence against the null hypothesis of no difference in mean earnings between genders. These findings provide robust support for the existence of a gender pay gap, with males consistently earning significantly more than females across both total and full-time employment categories.

## Conclusion:

Based on our comprehensive analysis using two-sample hypothesis testing, t-tests, and ANOVA tests on earnings data obtained from the United States Census Bureau, it is evident that a significant gender pay gap exists between men and women across several decades. The results consistently revealed statistically significant differences in mean earnings between genders, with males consistently earning substantially more than females, whether considering total workers or full-time workers. Consequently, we reject the null hypothesis of no gender pay gap and conclude that there is indeed a substantial variation in earnings between men and women, highlighting the urgent need for continued efforts to address and eliminate this inequality.