Unemployment Rate Gap Between Genders

Andre Gunawan – s3885488

Assignment 2

Introduction

- Around the world, finding a job is much tougher for women than it is for men. When women are employed, they tend to work in low-quality jobs in vulnerable conditions. Sadly, the forecast states that there will only be little improvement in the near future.

- Since my younger sister will graduate with her bachelor's degree this year and look for work in Australia, as a brother I have the curiosity to find out whether the unemployment rate gap between gender also occurs in developed country like Australia.

- The freedom to work by choice, in conditions of dignity, fairness, and safety is integral to a person's welfare. Hence, to answer the question of whether gender affects a person's probability of being unemployed. This report will analyze Australia's unemployment rate for men and women in the last 21 years.

Problem Statement

Question:

Does gender affect a person's probability of being unemployed in Australia?

Investigation method:

This report will use a two-sample t-test in hypothesis testing to compare the means of two groups and investigate whether the two groups are different from one another.

Data Pre-processing

- Both data is collected from **The World Bank** website.
- Resources:

- Male: http://databank.worldbank.org/data/reports.aspx?source=2&series=SL.UEM.TOTL.MA.NE.ZS&country=
- Female: https://databank.worldbank.org/reports.aspx?source=2&series=SL.UEM.TOTL.FE.ZS&country=

- Raw data is imported into RStudio using the **readr** package.

Data Pre-processing Cont.

- Showing raw data before pre-processing using head() function.

A tibble: 1 x 24

Gender <chr>	Country Name <chr>	Country Code <chr>	1999 [YR1999] <dbl>	2000 [YR2000] <dbl>	2001 [YR2001] <dbl>	2002 [YR2002] <dbl>	2003 [YR2003] <dbl>
Male	Australia	AUS	7.04	6.46	6.98	6.53	5.89

1 row | 1-8 of 24 columns

A tibble: 1 x 24

Gender <chr>	Country Name <chr>	Country Code <chr>	1999 [YR1999] <dbl>	2000 [YR2000] <dbl>	2001 [YR2001] <dbl>	2002 [YR2002] <dbl>	2003 [YR2003] <dbl>
Female	Australia	AUS	6.66	6.06	6.44	6.16	5.98

1 row | 1-8 of 24 columns

Data Pre-processing Cont.

- Use pivot_longer() function from tidyr package to combine variable 1999 to 2019 into one variable named “Year”.

A tibble: 6 x 5

Gender <chr>	Country Name <chr>	Country Code <chr>	Year <chr>	Unemployment rate <dbl>
Male	Australia	AUS	1999 [YR1999]	7.04
Male	Australia	AUS	2000 [YR2000]	6.46
Male	Australia	AUS	2001 [YR2001]	6.98
Male	Australia	AUS	2002 [YR2002]	6.53
Male	Australia	AUS	2003 [YR2003]	5.89
Male	Australia	AUS	2004 [YR2004]	5.28

6 rows

A tibble: 6 x 5

Gender <chr>	Country Name <chr>	Country Code <chr>	Year <chr>	Unemployment rate <dbl>
Female	Australia	AUS	1999 [YR1999]	6.66
Female	Australia	AUS	2000 [YR2000]	6.06
Female	Australia	AUS	2001 [YR2001]	6.44
Female	Australia	AUS	2002 [YR2002]	6.16
Female	Australia	AUS	2003 [YR2003]	5.98
Female	Australia	AUS	2004 [YR2004]	5.54

6 rows

Data Pre-processing Cont.

- Use bind_rows() function from dplyr package to merged both datas.
- Convert the data types into the proper type.
- Labelling the Year (factor variable).

```r
#merged both tables.
Unemployment_rate <- bind_rows(clean_male_UR, clean_female_UR)

#convert the gender variable as a factor.
Unemployment_rate$Gender <- as.factor(Unemployment_rate$Gender)

#convert the country code variable as a factor.
Unemployment_rate$`Country Code` <- as.character(Unemployment_rate$`Country Code`)

#convert the year variable as a factor.
Unemployment_rate$Year <- as.factor(Unemployment_rate$Year)

#convert the male unemployment rate variable as a numeric.
Unemployment_rate$`Unemployment rate` <- as.numeric(Unemployment_rate$`Unemployment rate`)

#labeling factor year
Unemployment_rate$Year <- factor(Unemployment_rate$Year, levels = c("1999 [YR1999]", "2000 [YR2000]", "2001 [YR2001]", "2002 [YR2002]", "2003
[YR2003]", "2004 [YR2004]", "2005 [YR2005]", "2006 [YR2006]", "2007 [YR2007]","2008 [YR2008]", "2009 [YR2009]", "2010 [YR2010]", "2011 [YR2011]",
"2012 [YR2012]", "2013 [YR2013]", "2014 [YR2014]", "2015 [YR2015]", "2016 [YR2016]","2017 [YR2017]", "2018 [YR2018]", "2019 [YR2019]"),
                              labels = c("1999", "2000", "2001", "2002", "2003", "2004", "2005", "2006", "2007", "2008", "2009", "2010",
"2011","2012", "2013", "2014", "2015", "2016", "2017", "2018", "2019"))
```

Data Pre-processing Cont.

- Use str() to show the data structure after pre-processing.
- Variable description:

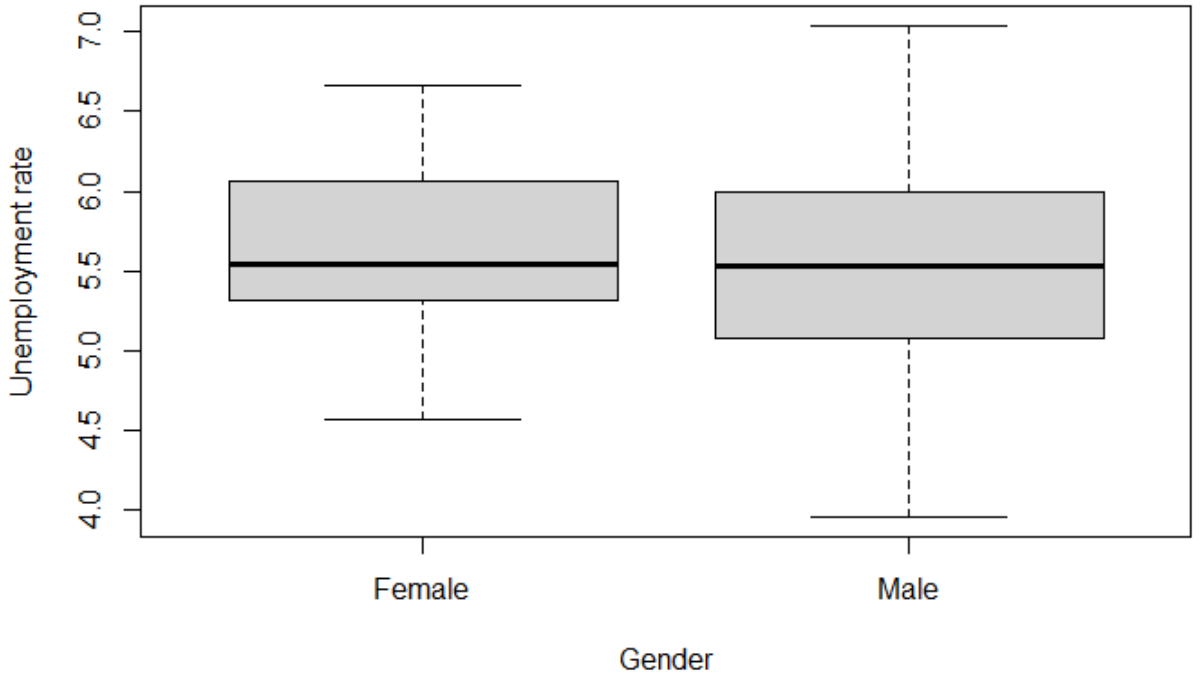
- Gender (Factor variable): Male and Female.
- Country name (Character): name of the country.
- Country code (Character): unique code of the country.
- Year (Factor variable): represent unemployment rate of different year. There are 21 levels on this factor from year 1999 – 2019 (21 years).
- Unemployment rate (Numeric variable): represent the rate/percentage of unemployment rate.

```
tibble [42 x 5] (S3: tbl_df/tbl/data.frame)
 $ Gender           : Factor w/ 2 levels "Female","Male": 2 2 2 2 2 2 2 2 2 2 ...
 $ Country Name     : chr [1:42] "Australia" "Australia" "Australia" "Australia" ...
 $ Country Code     : chr [1:42] "AUS" "AUS" "AUS" "AUS" ...
 $ Year             : Factor w/ 21 levels "1999","2000",..: 1 2 3 4 5 6 7 8 9 10 ...
 $ Unemployment rate: num [1:42] 7.04 6.46 6.98 6.53 5.89 ...
```

Descriptive Statistics and Visualisation

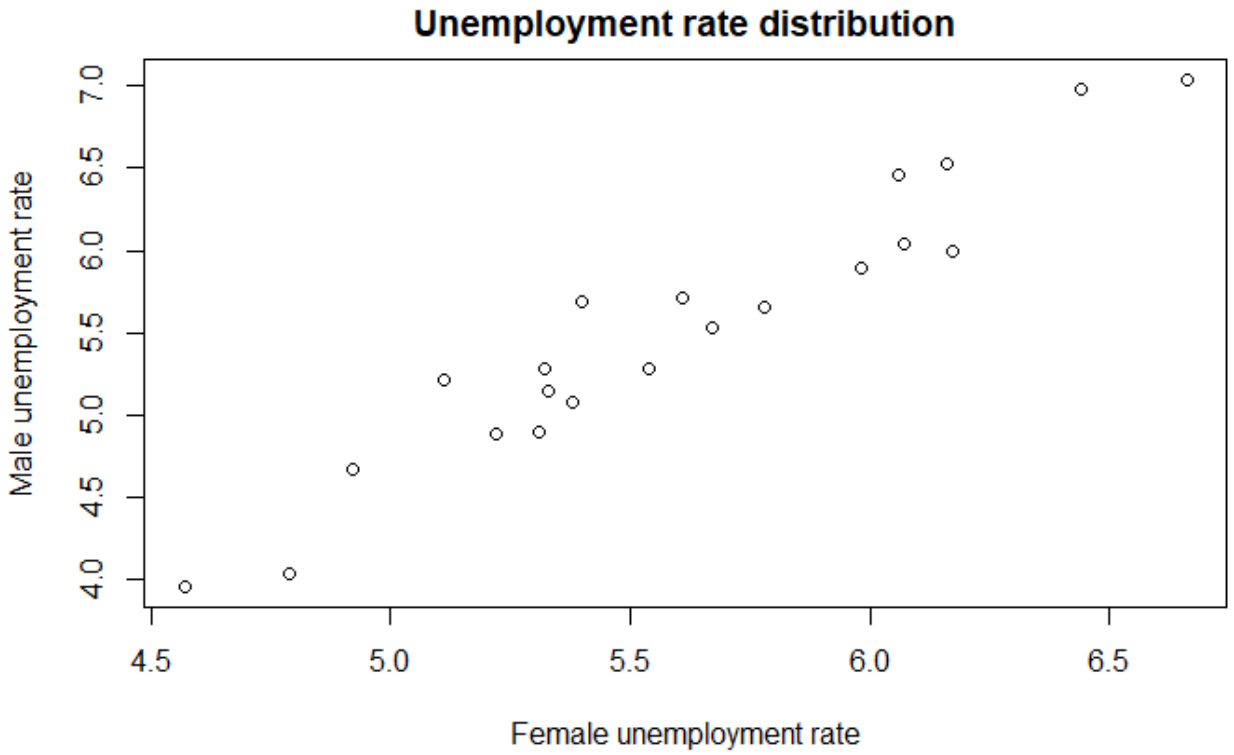
- Median and mean between both genders is approximately the same.
- Standard deviation for male is bigger than female showing a larger range of deviation for male unemployment rate.
- No missing value was found.
- No outliers shown on the boxplot.

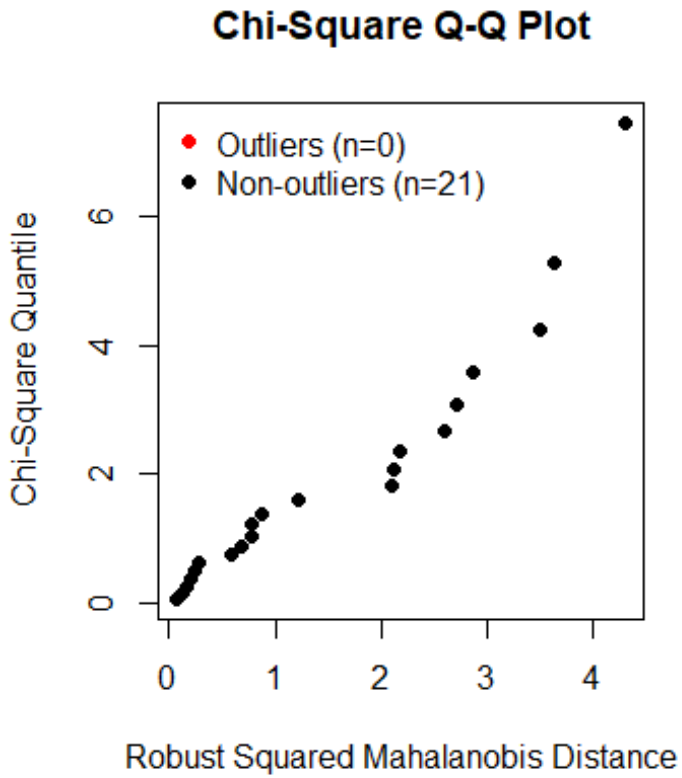
Gender	Min	Q1	Median	Q3	Max	Mean	SD	n	Missing
Female	4.57	5.31	5.54	6.06	6.66	5.594762	0.5464761	21	0
Male	3.96	5.08	5.53	6.00	7.04	5.521905	0.8293107	21	0



Checking for Outliers

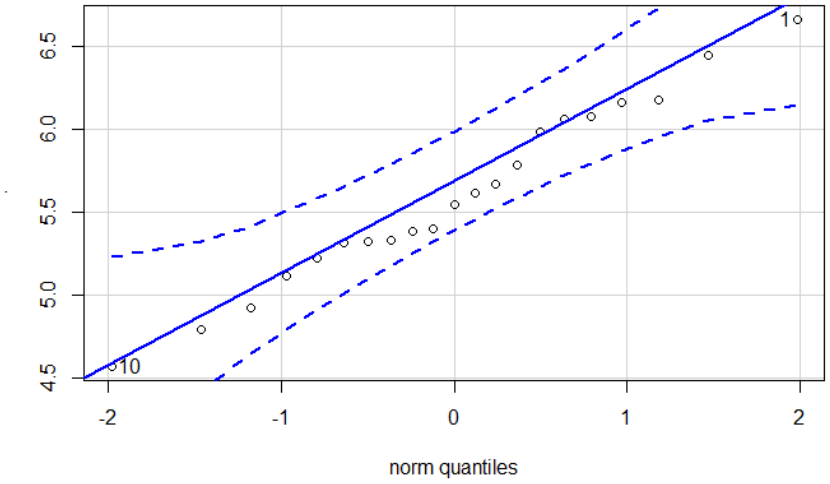
- Scatterplot and mahalanobis distance shows no outliers for the datasets.

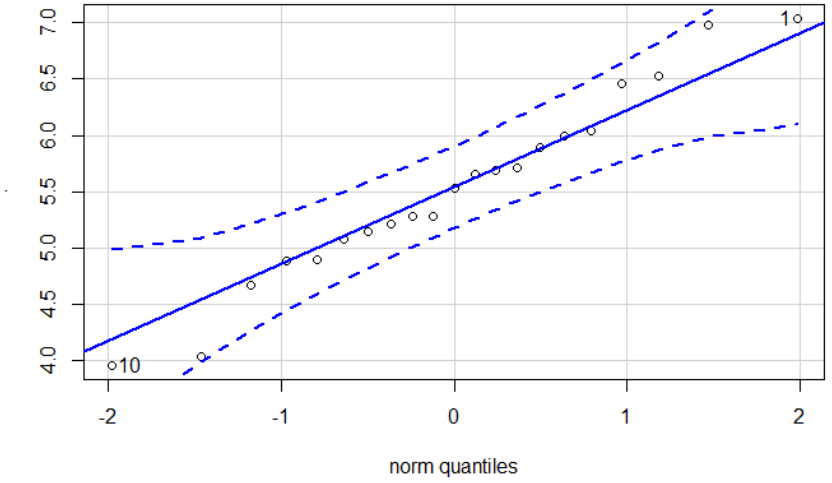




Testing the Assumption of Normality

- The number of observations (n) for each gender is 21.
- Therefore, because n < 30 we need to check the normality of each gender dataset using qqplot.
- It can be seen from the diagram below that no points falling outside the tails of distribution for male and female dataset. Therefore, we assume that the data is normally distributed.





Homogenity of Variance

- Test the assumption of equal variance using the levene's test.
- The levene's test has the following statistical hypothesis:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_A : \sigma_1^2 \neq \sigma_2^2$$

- It is found that the p-value for the Levene's test of equal variance for unemployment rate between males and females was p = 0.12. We find p > 0.05 therefore, we fail to reject H0. Therefore, we are safe to assume equal variance.

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1  2.5247 0.1199
      40
```

Hypothesis testing

- The two-sample t-test for unemployment rate gap between genders has the following statistical hypothesis:

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_A : \mu_1 - \mu_2 \neq 0$$

- μ1 and μ2 above refer to the population mean of males and females unemployment rate respectively. The null hypothesis is simply stating that the difference between the two independent population means is 0. The mean difference between males and females estimated by the sample was 5.59 - 5.52 = 0.07.

Hypothesis testing cont.

- As the test statistics t from the two-sample t-test assuming equal variance was t = 0.34 which was NOT more extreme than the t-critical value -2.02, therefore we fail to reject H0. According to the critical value method, there was no statistically significant difference between male and female unemployment rate.
- The two-tailed p value was reported to be p = 0.74. According to p-value method, as p = 0.74 > α = 0.05, we fail to reject H0. There was no statistically significant difference between male and female unemployment rate.
- Lastly, the 95% CI of the difference between the means (0.07) was calculated and reported by R as 95% CI [-0.37 0.51]. As this interval capture H0, we fail to reject it. There was no statistically significant difference between male and female unemployment rate.

```
        Two Sample t-test

data:  Unemployment rate by Gender
t = 0.33617, df = 40, p-value = 0.7385
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.3651667  0.5108809
sample estimates:
mean in group Female   mean in group Male
            5.594762             5.521905
```

```r
#Two-tailed t-critical value t* with df:
qt(p = 0.025, df = 21 + 21 - 2)
```

```
[1] -2.021075
```

Discussion

- A two-sample t-test was used to test for a significant difference between the mean unemployment rate of males and females. The results of the two-sample t-test assuming equal variance found that there is **no statistically significant** difference between the unemployment rate of males and females in Australia, t (df = 40) = 0.34, p = 0.74, 95% CI for the difference in means [-0.37 0.51].
- This investigation also has strength and weakness. The strength of this investigation is data taken from reliable resources. However, the limitation is that the sample size used is relatively small which less than 30 for each gender.

Discussion cont.

- As we all know that the unemployment rate gap between gender occurs all over the world, especially women are badly affected. Women want to be in an employment, but a persistent set of socio-economic barriers keep them out of the workforce.
- Finally, closing gender gaps in the labor force is not just good for women and their household, but for a country's economy as a whole. Gladly, the result of the hypothesis testing says that Australia is a country that has an equal unemployment rate for both genders. Australia has done a good job in equalizing social justice between genders, which is an amazing achievement as a country that we all should be proud of.

References

- The World Bank, International Labour Organization, accessed 20 October 2021, <https://databank.worldbank.org/reports.aspx?source=2&series=SL.UEM.TOTL.MA.NE.ZS&country=>
- The World Bank, International Labour Organization, accessed 20 October 2021, https://databank.worldbank.org/reports.aspx?source=2&series=SL.UEM.TOTL.FE.ZS&country=