

# PROJECT 2

inference statistical analysis

Group Member

- 1) ZIHAN IZHAR BIN AZAHIR
- 2) MUHAMMAD AFIQ AZMI BIN SUHAIZI
- 3) MUHAMMAD AZIM BIN ARIS
- 4) FIRHAD HAZIMI BIN JAMALUDIN



# Introduction

In this project, we are tasked to conduct an inference statistical analysis that cover the following analysis based on selected dataset. Our group will cover the analysis about the wages by education from 2010 until 2020. Our aim is to investigate the relationship between the level of education and the amount of wages received by them based on average hourly wages. The data pertain to private-sector wage and salary workers aged 18 to 64 who worked at least 20 hours per week and 26 weeks per year. This sample focuses on those with consistent employment. Coverage is defined as employees who received health insurance through their own employment and for which their employer paid at least a portion of the insurans.

# Dataset

This data set is obtained from [www.epi.org](http://www.epi.org). The data set is a secondary data which contains the average wages by education based on the hourly wages of workers disaggregated by the highest level of education attained from the year 2010 to 2020. The population sample is the wages and salary workers aged 16 and above. The respondents must meet these criteria in order to be included in the sample. 1) ages 16 and older, 2) employed in the public or private sector (self-employed were excluded), 3) hours worked were within the valid range in the survey and 4) had either hourly or weekly wages within the valid survey range. The data set contains about 5 variables which are year, highschool, some college, bachelor degree and advance degree.



# Hypothesis One-Sample Test

This one sample testing is to test whether the statement is true that the average wages of men High School in 2020 is 15% greater than average wages of men and women High School in the same year.

Assuming the  $\alpha = 0.1$

Year	Men <u>Highschool</u>	Men and Women Highschool
2020	\$22.55	\$21.03

$$H_0: p = 0.15$$

$$H_1: p > 0.15$$

### Calculations

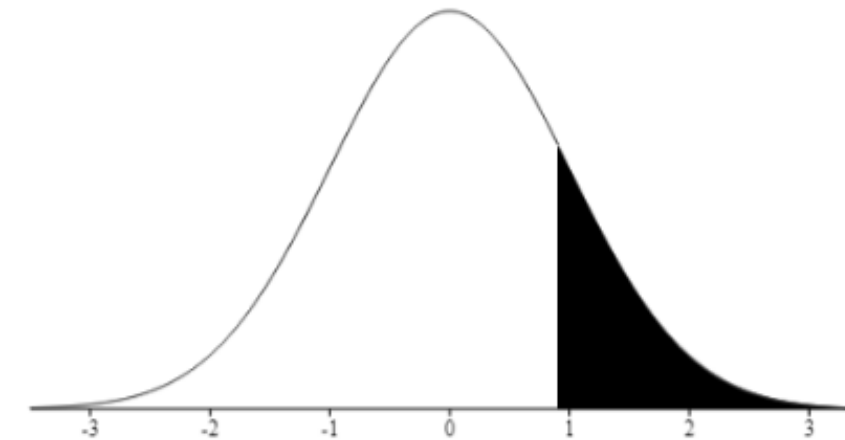
$$\begin{aligned}\hat{p} &= \frac{22.55}{21.03} \\ &= 1.0723\end{aligned}$$

$\alpha$	0.1
hypothesized population mean in the null hypothesis, $\mu$	0.15
sample mean, $\hat{p}$	1.0723
standard deviation, $\sigma$	4.6835

$$\begin{aligned}\sigma &= \sqrt{\frac{(22.55 - 1.0723)^2}{21.03}} \\ &= 4.6835\end{aligned}$$

### Test statistic

$$\begin{aligned}z &= \frac{\hat{p} - \mu}{\sigma / \sqrt{n}} \\ &= \frac{1.0723 - 0.15}{4.6835 / \sqrt{21.03}} \\ &= 0.9031\end{aligned}$$



$$P(z > 0.9031) = 0.31594$$

$$= 0.31594 > 0.15$$

### Test conclusion

Since the test value is 0.31594 which is more than 0.15, therefore we fail to reject the null hypothesis. There is no sufficient evidence to support the statement that average wages of men High School in 2020 is 15% more than average wages of men and women High School in the same year.

# REGRESSION ANALYSIS

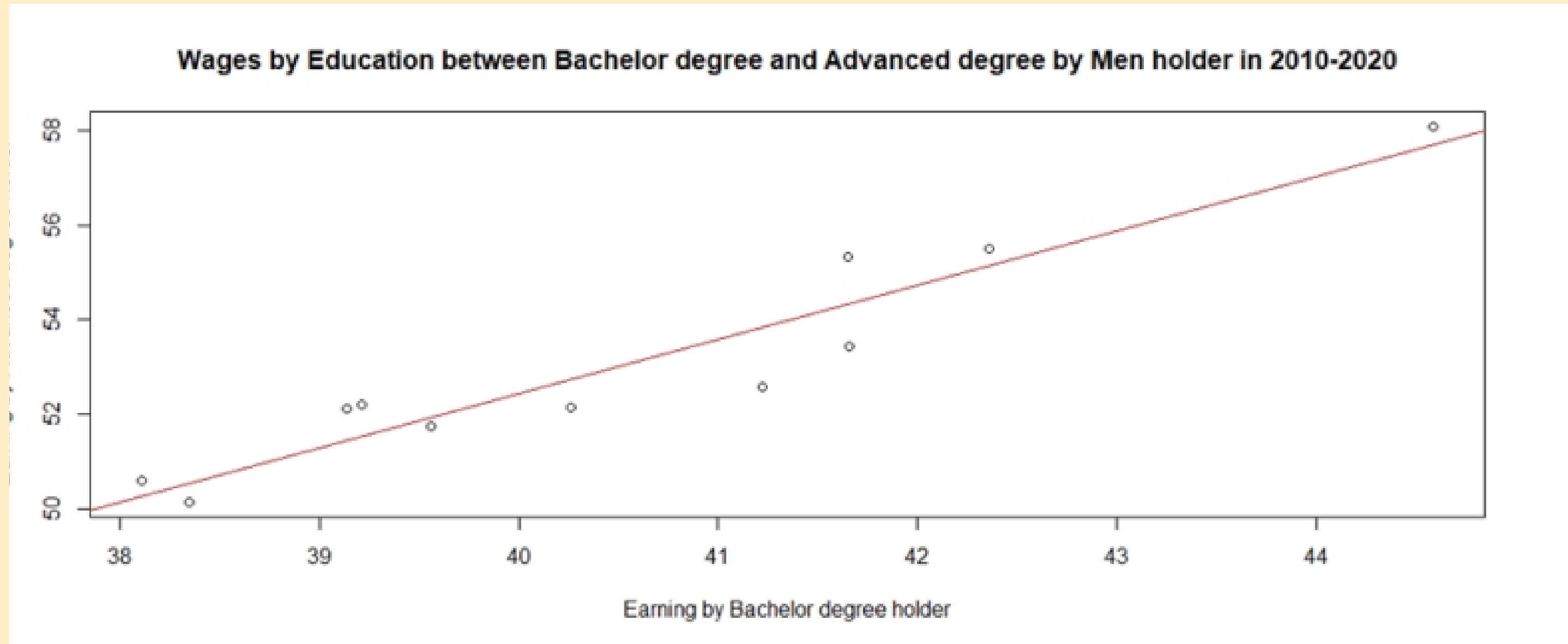
In this analysis, we are using earning by bachelor degree and earning by advanced degree as our variables, where we will test whether the value of earning by advanced degree depend on the value of earning by bachelor degree, using earning by bachelor degree as the independent variable(x) and earning by advanced degree as the dependent variable(y). Our regression model is a linear model, hence simple linear regression is used. The changes in the values of earning by advanced degree are assumed to be caused by the changes in the values of bachelor degree.

# Result

Hypothesis Statement:

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$



```

> edu.wages<-data.frame(x=c(39.56,38.35,39.14,39.21,38.11,40.26,41.66,41.22,41.65,42.36,44.59),y=c(51.74,50.15,52.11,52.19,50.60,52.16,53.45,52.58,55.32,55.50,58.08))
> edu.wages
  x      y
1 39.56 51.74
2 38.35 50.15
3 39.14 52.11
4 39.21 52.19
5 38.11 50.60
6 40.26 52.16
7 41.66 53.45
8 41.22 52.58
9 41.65 55.32
10 42.36 55.50
11 44.59 58.08
> plot(edu.wages$x,edu.wages$y,main='wages by Education between Bachelor degree and Advanced degree by Men holder in 2010-2020',xlab='Earning by Bachelor degree holder',ylab='Earning by Advanced degree holder')
> wages.regression<-lm(y~x, data=edu.wages)
> summary(wages.regression)

Call:
lm(formula = y ~ x, data = edu.wages)

Residuals:
    Min       1Q   Median       3Q      Max
-1.2629 -0.4895  0.3272  0.5115  0.9835

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.5248     4.9613   1.315   0.221
x            1.1479     0.1222   9.394 6.01e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7546 on 9 degrees of freedom
Multiple R-squared:  0.9074,    Adjusted R-squared:  0.8972
F-statistic: 88.24 on 1 and 9 DF,  p-value: 6.008e-06

```

Estimated Regression Model:

The diagram shows the regression equation  $\hat{y}_i = b_0 + b_1x$  with arrows pointing to each component:

- $\hat{y}_i$ : Estimated (or predicted) y value
- $b_0$ : Estimate of the regression intercept
- $b_1$ : Estimate of the regression slope
- $x$ : Independent variable

By using RStudio, we get  $b_1 = 1.1479$ ,  $b_0 = 6.5248$

Substitute the values of  $b_0$  and  $b_1$  into the regression model equation:  $\hat{y}_i = 6.5248 + 1.1479x$

# Analysis result

Find critical value, using  $\alpha = 0.05$ ,  $df = n-2 = 9$

From  $t$ -table, since this is a two-tailed test, there are two critical values:

Lower tail critical value  $-t_{\alpha/2=0.025, df=9} = -2.262$ , Upper tail critical value  $t_{\alpha/2=0.025, df=9} = 2.262$

From RStudio, we also get  $p\text{-value} = 6.008e-06$ . Hence, we reject  $H_0$  if test statistics  $> 2.262$  / test statistics  $< -2.262$

Calculate test statistic by:

$$t = \frac{b_1 - \beta_1}{s_{b_1}} \quad > \quad t \leftarrow (b1-0)/sb1$$

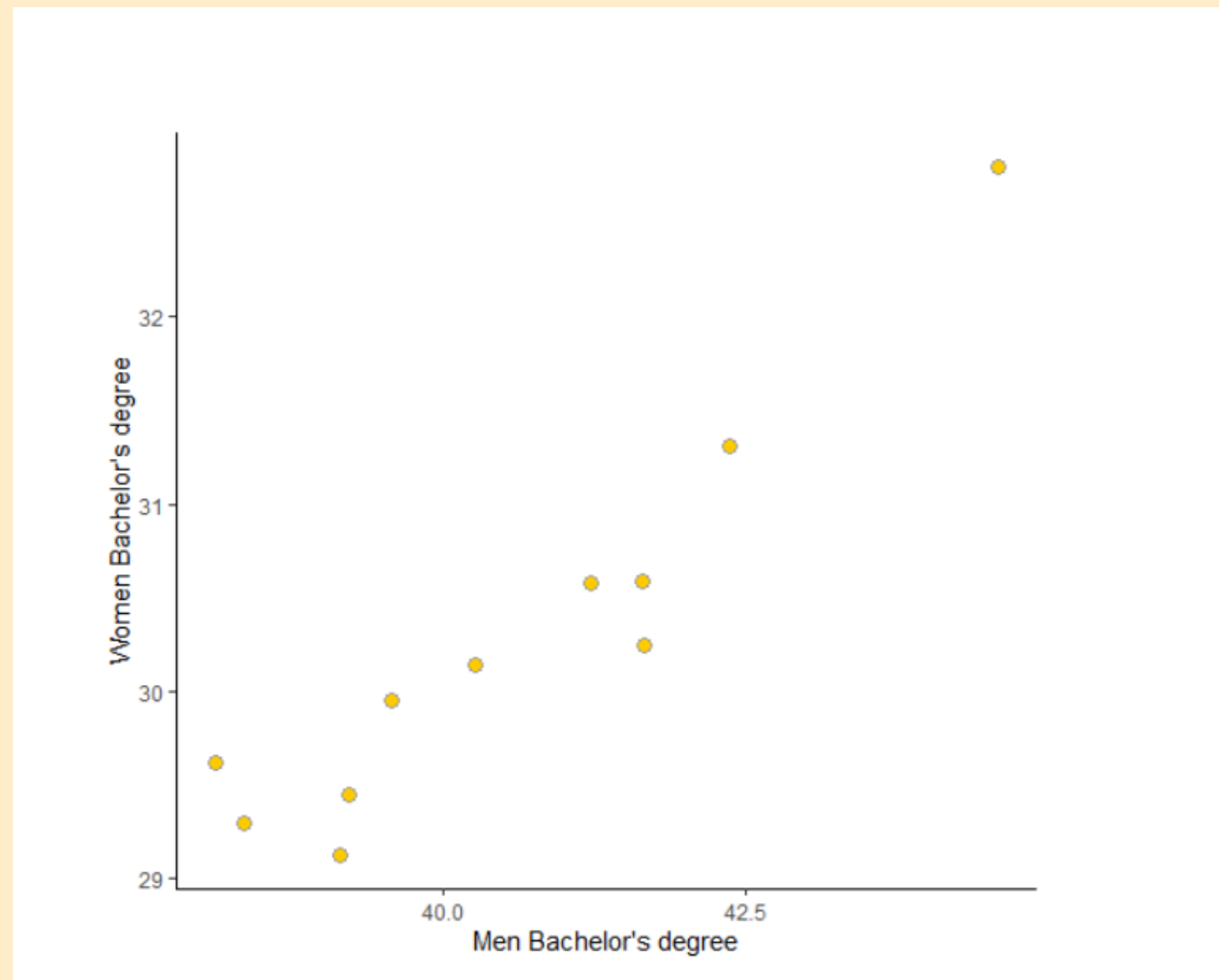
By using RStudio, we get test statistic  $t = 9.394$

Since test statistics  $t = 9.394 >$  upper-tail critical value  $t_{\alpha/2=0.025, df=9} = 2.262$ , we reject the null hypothesis. There is sufficient evidence that wages for advanced degree holder affect by wages for bachelor degree.



# Correlation Analysis

In the correlation test, we conducted a test to investigate the strength of the relationship between the average wage of Male Bachelor's degree and the average wage of Female Bachelor's degree from 2010-2020. We conducted the test with a significance level of 0.05.



$$r = \frac{13529.25 - \left(\frac{148608.163}{11}\right)}{\sqrt{\left[\left(18130.32\right) - \frac{(446.11^2)}{11}\right] \left[\left(10099.22\right) - \frac{(333.12^2)}{11}\right]}} = 0.9424$$

This indicates that the relationship between the average wage of Male Bachelor's degree and average wage of Female Bachelor's degree is a relatively strong positive linear relationship.

### Hypothesis Statement

$H_0 : P=0$  (no linear correlation)

$H_1 : P \neq 0$  (linear correlation exists)

### Critical Value

By using Rstudio, we obtain the results that  $t=8.4657$ ,  $df=9$  and  $t_{0.025, 9} = 2.262$ .

### Test Statistic

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

$$= \frac{0.9424}{\sqrt{\frac{1-0.9424^2}{11-2}}} = 3$$

### Conclusion

In conclusion, since  $t = 3 > t_{0.025, 9} = 2.262$ , we reject  $H_0$ . There is sufficient evidence to conclude that there is a linear relationship between the average wage of Male Bachelor's degree students and the average wage of Female Bachelor's degree students at the 5% significance level.





New Tab

Using this formula  $\frac{(O - E)^2}{E}$

Year	2017	2018	2019	2020	2021	Total
Observed Wages, O	\$35.80	\$36.00	\$36.69	\$38.57	\$38.26	\$185.32
Expected Wages, E	\$37.06	\$37.06	\$37.06	\$37.06	\$37.06	\$185.32
$((O-E)^2)/E$	0.04284	0.03032	0.00369	0.06152	0.03886	0.17723

### Critical Value

Degree of Freedom =  $5 - 1 = 4$ ,  
with significance level of 0.05%

$\chi^2_{4,0.05} = 9.48$

$$\chi^2 = \sum \frac{(O - E)^2}{E} = 0.04284 + 0.03032 + 0.00369 + 0.06152 + 0.03886$$

$$= 0.17723$$

### Conclusion

Test Statistic falls within the critical region, therefore we fail to reject hypothesis null, H0. That is, we fail to reject the claim that the monthly wages for bachelor's degree education level is in equal proportions within the 5 years (2017-2021)