

Statistic- STATA Sample Assignment

Answer 1:

- a) Population of interest: We are interested in adults between 30 and 39 years who completed atleast a bachelor's degree.
- b) The following table shows that final sample includes adults between 30 and 39 years who completed atleast a bachelor's degree

	Freq.	Percent	Cum.
Age group			
30 to 34 years	2,500	50.98	50.98
35 to 39 years	2,404	49.02	100
Education			
Bachelor's degree	3,415	69.64	69.64
Above Bachelor's degree	1,489	30.36	100

- c) The count of the final sample which includes adults between 30 and 39 years who completed atleast a bachelor's degree is 4904. Of the sub-sample, about 61% are females and the remaining 39% are males.

	Sex		
	Freq.	Percent	Cum.
Male	1,927	39.29	39.29
Female	2,977	60.71	100
Total	4,904	100	

Answer 2:

- a) Variable generated in Stata.
- b) Both the measures of central tendency, the mean and median for annual wages for males are much higher than females.

	Mean Annual Wages	Median Annual Wages
Male	70723.41	66994.2
Female	59365.66	58500

- c) Since both the measure of central tendency, the mean and median for annual wages for males are higher than females, there seems to be association between gender and

annual wages. Moreover, the average annual wages for males are higher by 11,357.75 than average annual wages for females which means that males earn, on an average 19.13% more than females.

d)

Null hypothesis states that there does not exist any difference in the annual wages based on gender.

Alternate hypothesis states that there exists a statistically significant difference in the annual wages based on gender.

Alpha level: 5% alpha level is chosen.

Test statistic and p-value: The value of test statistic is 11.5099 whereas the value of p is close to zero. The critical value of t at 3868 degrees of freedom is 1.96. Since the value of t-test is larger than the critical value of t, we reject the null hypothesis and conclude that there exists a statistically significant difference in the annual wages based on gender.

Moreover, the value of t-stat is positive, this suggests that the annual mean wages of males are higher than females.

Similarly, the p-value is less than alpha level of 0.05 and thus, it can be concluded that there exists a statistically significant difference in the annual wages based on gender. Moreover, the p-value suggests that had our sample been drawn from any other population where there exists no association between annual wages and gender, the chances of getting a high t-stat value as in our sample would have been almost zero. Thus, the results can be extended to broader population of interest.

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
Male	1543	70723.41	848.5011	33330.01	69059.07	72387.75
Female	2327	59365.66	573.7045	27674.93	58240.64	60490.69
combined	3870	63894.09	491.3025	30563.6	62930.85	64857.33
diff		11357.75	986.7844		9423.079	13292.41

```
diff = mean(Male) - mean(Female)                                t = 11.5099
Ho: diff = 0                                                    degrees of freedom = 3868

Ha: diff < 0                                                    Ha: diff != 0                                                    Ha: diff > 0
Pr(T < t) = 1.0000        Pr(|T| > |t|) = 0.0000        Pr(T > t) = 0.0000
```

Answer 3:

a) Variable generated in Stata.

b) Both the measures of central tendency, the mean and median for annual wages for males are much higher than females for respondents with no children.

	Mean Annual Wages	Median Annual Wages
Male	65406.78	63008.24
Female	58315.93	55993.6

c) Since both the measure of central tendency, the mean and median for annual wages for males are much higher than females for respondents with no children, there seems to be association between gender and annual wages. Moreover, the annual wages for males are higher on average by about 7,090.85 than annual wages for females which means that males earn, on an average 12.16% more than females for respondents with no children.

Comparing the two samples, with entire data and childless respondents, we can see that while males earn on the average 19.13% more than female respondents in total sample, they earn on the average 12.16% more than the female respondents with no children. Going deeper in the data, we see that while average income of females remains more or less the same (only a difference of) between total sample and childless respondents. We see that there is a relatively larger difference of between the males respondents of two samples. This indicates that larger difference in the average salaries between males and females in the first case is mainly due to the larger increase in the average wages of the males.

d) Null hypothesis states that there does not exist any difference in the annual wages based on gender for respondents with no children.

Alternate hypothesis states that there exists a statistically significant difference in the annual wages based on gender for respondents with no children.

Alpha level: 5% alpha level is chosen.

Test statistic and p-value: The value of test statistic is 5.16 whereas the value of p is close to zero. The critical value of t at 1668 degrees of freedom is 1.96. Since the value of t-test is larger than the critical value of t, we reject the null hypothesis and conclude that there exists a statistically significant difference in the annual wages based on gender for respondents with no children. Moreover, the value of t-stat is positive, this suggests that the annual mean wages of males are higher than females for respondents with no children. Similarly, the p-value is less than alpha level of 0.05 and thus, it can be concluded that there exists a statistically significant difference in the annual wages based on gender for respondents with no children.

Moreover, the p-value suggests that had our sample been drawn from any other population where there exists no association between annual wages and gender for respondents with no children, the chances of getting a high t-stat value as in our sample would have been almost zero. Thus, the results can be extended to broader population of interest.

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
Male	742	65406.78	1073.213	29233.98	63299.88	67513.68
Female	928	58315.93	879.7301	26799.3	56589.44	60042.42
combined	1670	61466.48	688.121	28120.5	60116.81	62816.15
diff		7090.852	1374.35		4395.219	9786.485

diff = mean(Male) - mean(Female) t = 5.1594
 Ho: diff = 0 degrees of freedom = 1668

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
 Pr(T < t) = 1.0000 Pr(|T| > |t|) = 0.0000 Pr(T > t) = 0.0000

- e) From the results we see that there is a statistically significant difference in the annual wages based on gender for respondents with no children. Statistical test indicate that there is a statistical significance in wages disparity between males and females respondents of both the samples.

Even though there is disparity in the wages between males and females, males on an average earn 12.16% more than females for respondents with no children. Comparing the two samples, with entire data and childless respondents, we can see that while males earn on the average 19.13% more than female respondents in total sample, they earn on the average 12.16% more than the female respondents with no children. Thus we can infer that there is a possibility that women tend to earn lower wages than men in part because they shoulder more of the child-reading responsibilities.

Answer 4:

- a) The table showing the relationship between gender and part-versus full-time work is presented below:

	Full time	Part time
Male	1684(93%)	118(7%)
Female	2076(80%)	504(20%)

- b) The results show that on one hand, only 7% of male respondents work part time. This figure is higher for female respondents where about 20% of the female respondents work part time. This suggests that there exists a substantive difference between gender and part-time work.
- c) Null hypothesis states that there does not exist any relationship between gender and part-time work.

Alternate hypothesis states that there exists a statistically significant relationship between gender and part-time work.

Alpha level: 5% alpha level is chosen.

Test statistic and p-value: The value of test statistic is 146.91 whereas the value of p is close to zero. The critical value of chi-square at 1 degrees of freedom is 3.84. Since the value of chi-test is larger than the critical value, we reject the null hypothesis and conclude that there exists a statistically significant relationship between gender and part-time work. Moreover, the results show that only 7% of male respondents work part time. This figure is higher for female respondents where about 20% of the female respondents work part time. Similarly, the p-value is less than alpha level of 0.05 and thus, it can be concluded that there exists a statistically significant relationship between gender and part-time work.

Sex of respondent	Part- or full-time job		Total
	Full-time	Part-time	
Male	1,684 93.45	118 6.55	1,802 100.00
Female	2,076 80.47	504 19.53	2,580 100.00
Total	3,760 85.81	622 14.19	4,382 100.00

Pearson chi2(1) = 146.9129 Pr = 0.000

Moreover, the p-value suggests that had our sample been drawn from any other population where there exists no association between part-time work and gender, the chances of getting a high chi value as in our sample would have been almost zero. Thus, the results can be extended to broader population of interest.

- d) The table showing the relationship between gender and part-versus full-time work in childless couples is presented below:

	Full time	Part time
Male	783(93%)	62(7%)
Female	864(85%)	147(15%)

The results show the percentage of male respondents working part time is same at 7%. This figure is lower for female respondents where about 15% of the female respondents

who are childless work part time. This suggests that in childless respondents when there are no child-rearing responsibilities women are more likely to take up full time job.

- e) The results are statistically significant for the relationship between gender and part-time work for childless respondents. Moreover, the p-value suggests that had our sample been drawn from any other population where there exists no association between part-time work and gender for respondents with no children, the chances of getting a high chi value as in our sample would have been almost zero. Thus, the results can be extended to broader population of interest.

Sex of respondent	Part- or full-time job		Total
	Full-time	Part-time	
Male	783 92.66	62 7.34	845 100.00
Female	864 85.46	147 14.54	1,011 100.00
Total	1,647 88.74	209 11.26	1,856 100.00

Pearson chi2(1) = 23.8972 Pr = 0.000

- f) From the results we see that statistical test indicate that there is a statistical significance difference in the gender and part-time work of both the samples. However, we see that comparing the two samples, with entire data and childless respondents, while the percentage of males who work part time remains the same at 7%, the percentage of female in the entire sample who work part time are 20% and this decreases to 15% for female respondents with no children. Thus we can infer that there is a likelihood that gender disparities in full-versus part-time work may stem in part from the fact than women take on more of the child-reading responsibilities.

Answer 5:

- a) The table showing the relationship between gender and occupation of a senior management position is presented below:

	Does not occupy senior position	Occupies a senior position
Male	1851(99.52%)	9(0.48%)
Female	2695(99.85%)	4(0.15%)

- b) The percentage of either of gender who occupies a senior management position is very low and the results show that only 0.48% of male respondents occupies a senior

management position. This figure is much lower for female respondents where about 0.15% of the female respondents occupies a senior management position.

c) Null hypothesis states that there does not exist any relationship between gender and occupies a senior management position.

Alternate hypothesis states that there exists a statistically significant relationship between gender and occupies a senior management position.

Alpha level: 5% alpha level is chosen.

Sex of respondent	Occupation of a senior management position		Total
	Occupies	Occupies	
Male	1,851 99.52	9 0.48	1,860 100.00
Female	2,695 99.85	4 0.15	2,699 100.00
Total	4,546 99.71	13 0.29	4,559 100.00

Pearson chi2(1) = 4.3635 Pr = 0.037

Test statistic and p-value: The value of test statistic is 4.36 whereas the value of p is close to 0.037. The critical value of chi-square at one degree of freedom is 3.84. Since the value of chi-test is larger than the critical value, we reject the null hypothesis and conclude that there exists a statistically significant relationship between gender and occupies a senior management position. Moreover, the results show that only 0.48% of male respondents occupies a senior management position. This figure is much lower for female respondents where about 0.15% of the female respondents occupies a senior management position.

Similarly, the p-value is less than alpha level of 0.05 and thus, it can be concluded that there exists a statistically significant relationship between gender and occupies a senior management position.

d) The table showing the relationship between gender and occupation of a senior management position for childless respondents is presented below:

	Does not occupy senior position	Occupies a senior position
Male	880(99.77%)	2(0.23%)
Female	1053(99.81%)	2(0.19%)

The percentage of either of gender who occupies a senior management position is very low and the results show that only 0.23% of male respondents occupies a senior management position for childless respondents. This figure is much lower for female respondents for childless respondents where about 0.19% of the female respondents occupies a senior management position. The percentage of males, who occupies a senior position and are childless, has declined. However, percentage of females who occupies a senior position are childless has increased.

e)

The value of chi-test is smaller than the critical value, we cannot reject the null hypothesis and conclude that there does not exist a statistically significant relationship between gender and occupies a senior management position for childless respondents.

Sex of respondent	Occupation of a senior management position		Total
	Occupies	Occupies	
Male	880 99.77	2 0.23	882 100.00
Female	1,053 99.81	2 0.19	1,055 100.00
Total	1,933 99.79	4 0.21	1,937 100.00

Pearson chi2(1) = 0.0322 Pr = 0.858

The p-value suggests that had our sample been drawn from any other population where there exists no association between occupies a senior management and gender for respondents with no children, the chances of getting a low chi value as in our sample would have been almost zero. Thus, the results can be extended to broader population of interest.

- f) From the results we see that statistical test indicate that there is a statistical significance difference in the gender and occupies a senior management position in full sample. However, there is no statistical significance difference in the gender and occupies a senior management position in childless respondents.

However, we see that comparing the two samples, with entire data and childless respondents, the percentage of males who occupies a senior management position are higher in the entire sample (0.48%) in comparison to the childless males (0.23%). On the other hand, fewer percentage of females is tend to take up senior position jobs (0.15%) in the entire sample in comparison to the females with no children (0.19%). Thus we can

infer that there is a likelihood that gender disparities in occupation may stem in part from the fact than women take on more of the child-reading responsibilities.

Answer 6:

We can see from our analyses in previous questions that income disparities between males and females get more pronounced for respondents when data of respondents with children are included with males earning significantly more than what females earn when data of respondents with children are included as compared to the difference in incomes of males and females for childless respondents. We also see that females tend to take less full time work and also tend to take less senior management positions as compared to their male counterparts after having children. This suggests that gender inequality stems partly from the fact that women tend to play larger role in child rearing and take up more child-rearing responsibilities.

This indicates that with children women tend to focus less on their careers by taking more part-time work and avoid taking senior management positions that require bigger responsibilities and hours in turn earning less wages as they take up substantial child rearing responsibilities that requires additional time and efforts on their parts.