**Analytics Types to be performed:**

* **Provide your analytical insight on sampling, bias & relevant matters**

Sample contains 4856 records but all of the individuals are with in age 30-50 range. There for we can clearly see a bias when considering age in to account.

Also more than 60% of the individuals in the sample are married and there also this can be due to the bias since never married individuals are only 14% of the total sample but cannot exactly confirm without looking at the population statistics about the marital status.

Considering the Educational level we noticed that ~40% of the individuals are from education level 12. However there were only ~10% of the individuals that have lower education levels (0-10). Here also we are noticing that this sample can be bias towards higher education levels. But cannot exactly confirm without knowing the population statistics about the educational level.

We also noticed there are ~25% of the individuals without earnings. But since our sample contains individuals with 30-50. These people most probably are non-respondent’s for the earnings questions in this survey. Also US unemployment rate in ~4% we can’t assume these are unemployed people other than if this data set has a large sampling bias when selecting the individuals. Same issue is there in the working hours as well. Since we assumed these are non-respondents we excluded them from our analysis on earnings and hours.

* **Provide your analytical insights on Hypothesis, Confident Intervals & Sample statistics**

**Analysis 1 Hypothesis: People with age over 40 years earnings are different than the people with age less than 40 years**

H0: Average earnings of a person with age less than 40 is equal to Average earning of a person with age greater than or equal to 40 years

H1:Average earnings of a person with age less than 40 is not equal to Average earning of a person with age greater than or equal to 40 years

Observed P value - 2.506 e-06

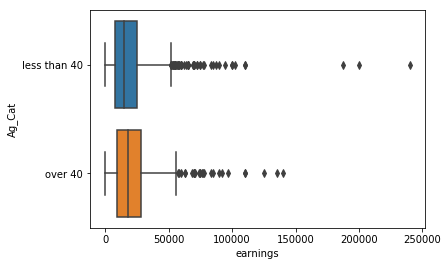
Since the P value of the T test is less than 0.05 we reject H0 with a 95% level of confidence

Therefore we can conclude there is a clear difference between average earnings of a person with age less than 40 and average earnings of a person with age greater than or equal to 40.

According to the sample summary statistics Average earnings of a person with age less than 40 is equal to $17901 and Average earnings of a person with age greater than or equal to 40 is equal to $20401. Which shows that people with age over 40 years earns more than people with age less than 40 people.

95% confidence interval for Average earnings (per year) of a person with age greater than or equal to 40 years is ($17232, $18571). We can say if a person’s age greater than or equal to 40 his salary should be in the interval ($17232, $18571) at a 95% level of confidence.

95% confidence interval for Average earnings (per year) of a person with age less than 40 years is ($19605, $21198). We can say if a person’s age less than 40 his salary should be in the interval ($19605, $21198) at a 95% level of confidence.

****

**Analysis 2 Hypothesis: People with age over 40 years working hours are different than the people with age less than 40 years**

H0: Average working hours(per year) of a person with age less than 40 is equal to Average working hours of a person with age greater than or equal to 40 years

H1:Average working hours(per year) of a person with age less than 40 is not equal to Average working hours(per year) of a person with age greater than or equal to 40 years

Observed P value - 2.896 e-05

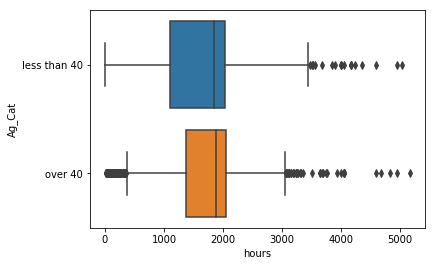
Since the p value of the T test is less than 0.05 we can reject H0 with a 95% level of confidence.

Therefore we can conclude there is a clear difference between average working hours of a person with age less than 40 and average working of a person with age greater than or equal to 40.

According to the sample summary statistics Average working hours(per year) of a person with age less than 40 is equal to 1594 According to the sample Average working hours(per year) of a person with age greater than or equal to 40 is equal to 1696. Which shows that people with age over 40 years works more hours per year than people with age less than 40

95% confidence interval for Average working hours (per year) of a person with age greater than or equal to 40 years is (1660, 1732)

95% confidence interval for Average working hours (per year) of a person with age less than 40 years is (1562, 1625).



**Analysis 3 Hypothesis: Earnings of people with high educational levels are different than earnings of the people with low educational levels**

H0: Average earning of a person with High Educational Level is equal to Average earning of a person with Low Educational Level

H1: Average earnings of a person with High Educational Level is not equal to Average earning of a person with Low Educational Level

Observed P value - 7.713 e-95

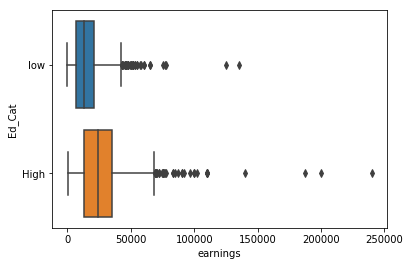
Since the p value of the T test is less than 0.05 we can reject H0 with a 95% level of confidence

Therefore we can conclude there is a clear difference between average earning of a person with high educational level and average earnings of a person with low educational level.

According to the sample summary statistics Average earnings of a person with High Educational Level (14 to 17) is equal to $26087. According to the sample summary statistics Average earnings of a person with Low Educational Level(less than 14) is equal to $14901. Which shows that people with High Educational levels earns more than the people with low educational levels

95% confidence interval for Average earnings (per year) of a person with High Educational Level ($24995, $27180)

95% confidence interval for Average earnings (per year) of a person with Low Educational Level ($14434, $15370)



**Analysis 4 Hypothesis: Working hours of the people with high educational levels are different than the working hours of the people with low educational levels**

H0: Average working hours (per year) of a person with High Educational Level is equal to Average working hours of a person with Low Educational Level

H1: Average working hours (per year) of a person with High Educational Level is not equal to Average working hours (per year) of a person with Low Educational Level

Observed P value - 0.0941

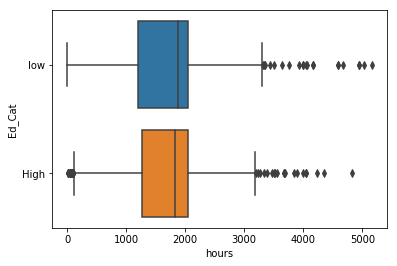
Since the p value of the T test is greater than 0.05 we do not reject H0 with a 95% level of confidence

Therefore we cannot conclude that there is a clear difference between average working hours of a person with high educational level and average working hours of a person with low educational level.

According to the sample summary statistics average working hours (per year) of a person with High Educational is equal to 1663 and average working hours (per year) of a person with Low Educational is equal to 1620. This shows that there is no significant difference between working hours of a person with high educational level and a person with a low educational level.

95% confidence interval for Average working hours (per year) of a person with High Educational Level is (1624, 1703)

95% confidence interval for Average working hours (per year) of a person with Low Educational Level is (1590, 1651)



**Analysis 5 Hypothesis: There is a difference between earnings of married and other people**

H0: Married and Other Peoples Average Earnings are equal

H1: Married and Other Peoples Average Earnings are not equal

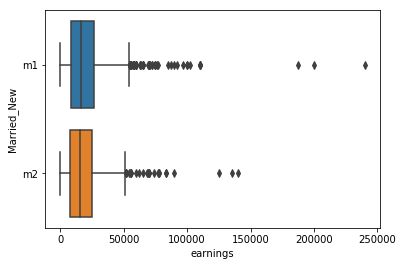
Observed P value - 0.0275155

Since the p value of the T test is less than 0.05 we can reject H0 with a 95% level of confidence

It concludes that average earnings of married and other people is not equal and by looking at the summary stats we noticed that average earnings of married people ($19389) is higher than average earnings of other people ($18128)

95% confidence interval for Average earnings (per year) for married people ($18720, $20058)

95% confidence interval for Average earnings (per year) other people ($17269, $18989)



**Analysis 6 Hypothesis: There is a difference between earnings of people with 2 or less than 2 kids and people with more than 2 kids**

H0: Average earnings for people with 2 or less than 2 kids

H1: Average earnings for people with more than 2 kids

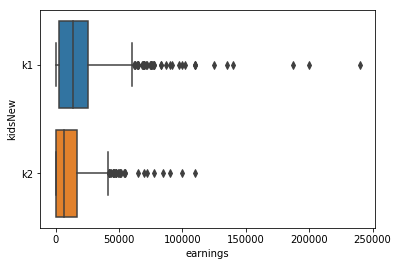
Decision: According to the 2 -sample t test it can be seen that p value is less than 0.05. Therefore we reject h0 at 0.05 significance level

Conclusion: Average earnings for people with 2 or less than 2 kids and average earnings for people with more than 2 kids have significant difference

`

95% confidence interval for Average earnings (per year) for people with 2 or less than 2 kids ($20208, $21591)

95% confidence interval for Average earnings (per year) for people with more than 2 kids ($14238, $15681)



**Analysis 7 Hypothesis: There is no correlation between Earning and Education level**

H0: There is no correlation between earnings and education level

H1: There is a correlation between earnings and education level

Observed correlation values between earnings and educations level. Correlation value between earnings and education level was 0.379. P-value for this was 3.686745979111728x10-155.This is lower than 0.05. Therefore we can reject the null hypothesis. That is there is a moderate linear relationship between two variables.

**Analysis 8 Hypothesis: There is no correlation between Earning and Work**

H0: There is no correlation between earnings and work

H1: There is a correlation between earnings and work

Observed correlation values between earnings and work. Correlation value between earnings and work was 0.6272. P-value for this was 3.6867459791x10-155.This is lower than 0.05. Therefore we can reject the null hypothesis. There is a high positive linear relationship between the two variables

**Analysis 9 Hypothesis: There is no correlation between Earning and Age**

H0: There is no correlation between earnings and age

H1: There is a correlation between earnings and age

Observed correlation values between earnings and work. Correlation value between earnings and age was 0.0924. P-value for this was 3.6867459791x10-155.This is lower than 0.05. Therefore we can reject the null hypothesis. There is a no linear relationship between two variables.