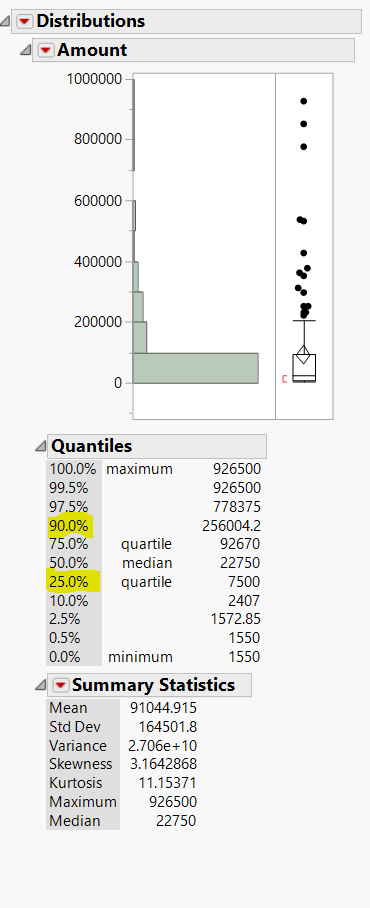
**OPIM 5604 B15 – Predictive Modeling Assignment Meghana Kasula (Net ID=mek15120)**

*“The work contained and presented here is my work and my work alone.”*

**1. Obtaining Descriptive Statistics (10 points)**



|  |  |
| --- | --- |
| sample mean | 91044.915 |
| maximum | 926500 |
| median | 22750 |
| variance | 2.706e + 10 |
| standard deviation | 164501.8 |
| first quartile | 7500 |
| 90th percentile | 256004.2 |
| skewness | 3.1642868 |
| kurtosis | 11.15371 |

Figure (1)

The statistics in Figure (1) was obtained through the following steps :

Analyze> Distribution> summary statistics> customize summary statistics

**2. Used the distribution function in JMP to explorer the data and list at least four facts that relate total amount paid with the other variables. (20 Points)**

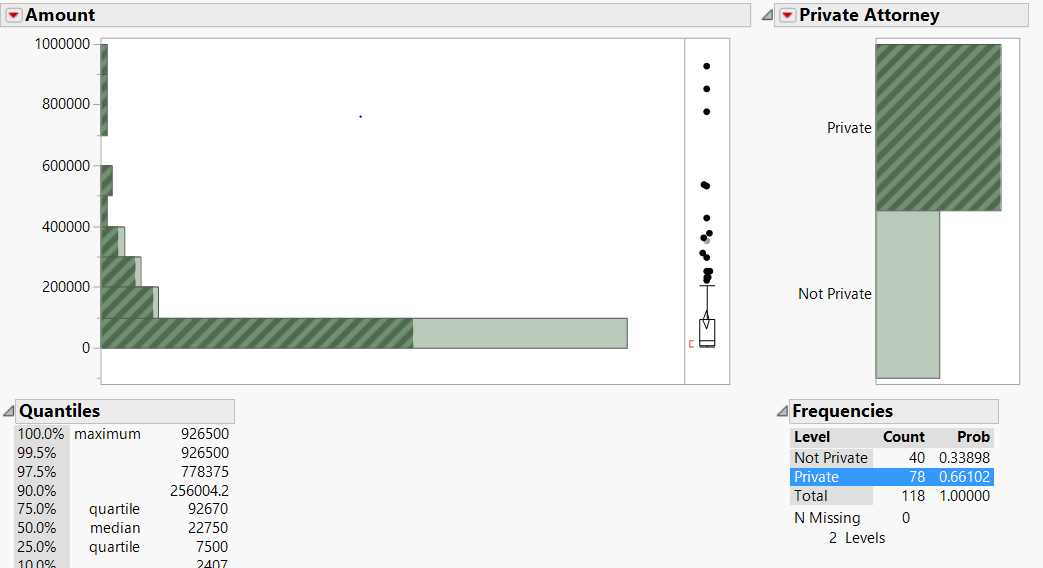
1. Claims above 400000 $ are all from private attorneys (Figure. 2)

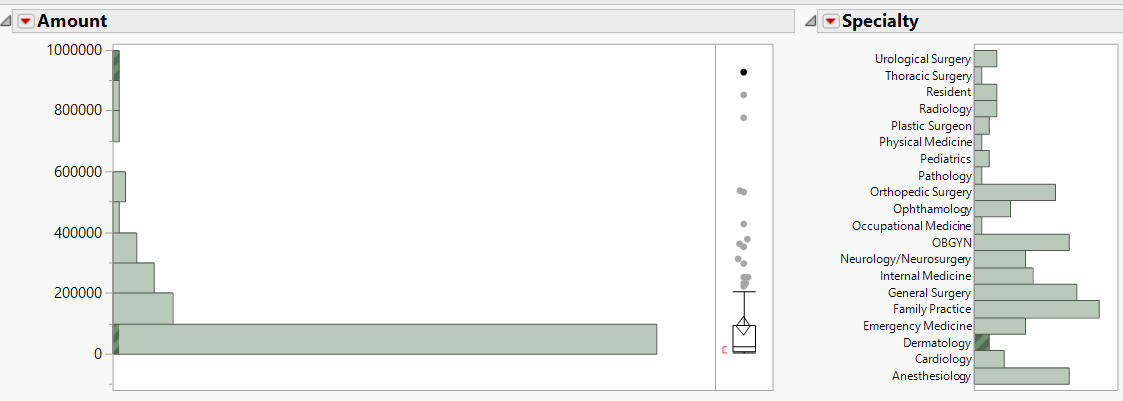
2. Highest claim amount is from dermatology. (Figure 3)

3. Cases with highest severity have paid amount less than 400000 $. (Figure 4)

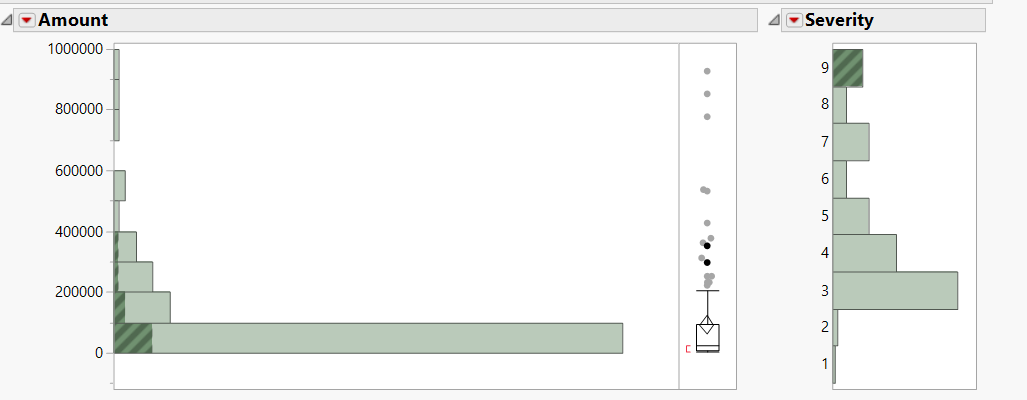
4. Non privat

e firms could make maximum claim amount only up to 400000 $. (Figure 5)

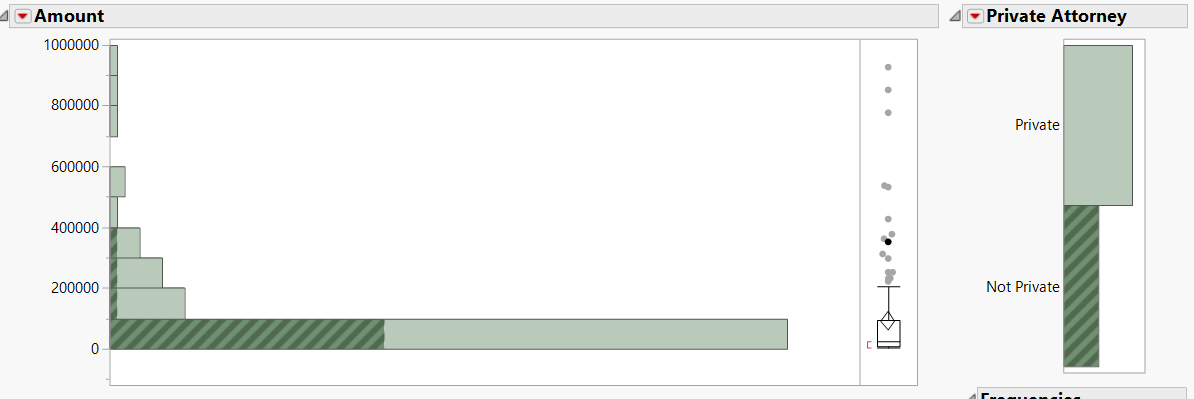
(Figure 2) I observed that the all claims amount above 400000 $ have been handled by only private attorneys.



(Figure 3) The highest claim amount is from Dermatology specialty. And also dermatology has the biggest range of claim amount compared to other specialties starting from least to highest claim amount which we can observe in the above distribution.



(Figure 4) As we click on the severity 9 which is the highest one, the claim amount that are highlighted are all below 400000 $.



(Figure 5) When I clicked on Non private attorney, only claim amount below 400000 $ have been highlighted. Which means, the claim amounts for private attorneys are up to 400000 $.

**3. Obtaining Confidence Intervals (20 points)**

**The insurance company claims that the Amount of the claim payment is about $80000 for each claim.**

**The company wants to know if the average amount is significantly different from $ 80000. Complete the**

**following steps in order to get an answer for the manufacturer.**

**a) State the assumption to generate a 95% confidence interval.**

**b) State how you would validate the assumption.**

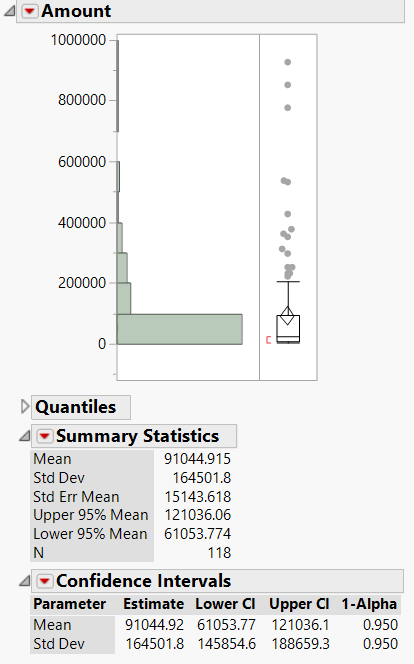
**c) Determine the 95% confidence interval for the mean.**

**d) Interpret this confidence interval in your own words.**

a) The assumption to generate 95% confidence Interval is that

i) the sample is randomly selected;

ii) the sample has symmetric distribution, basically its almost nearly normal.

b) We can validate the assumption by invoking the central Limit Theorem where we need

i) access to the entire population and randomly select samples

ii) Plot Histogram on frequency distribution of means

We observe for Gaussian distribution which will validate our assumption.

c) The 95% Confidence Interval lower and upper limit.

Lower limit = 61053.77

Upper limit =121036.1

d) Since the assumed mean (µ0= $80000) falls into the 95% confidence interval where α= 0.05, it is a plausible value and since we did not reject the null hypothesis ( H0 : µ0= $80000), we can assume that the p value is greater than 0.05.

Figure 6

**4. The insurance company wants to perform a one-sample *t*-test to determine whether the mean**

**Amount is significantly different from $80000. Be sure to validate any assumptions that are**

**associated with the test. Continue to use MedicalMalpractice.jmp.**

**a) Perform the analysis to validate the assumptions of a one-sample *t*-test graphically. Does the**

**data appear normally distributed using the normal curve and the normal quantile plot?**

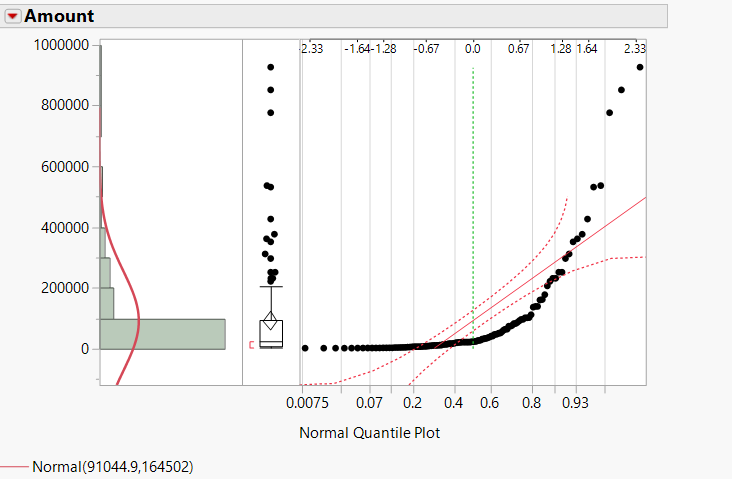
**b) Before conducting the test, complete the hypotheses. Is this a one-sided or two-sided test?**

**Ho: \_\_\_\_\_\_\_\_\_\_\_\_**

**Ha: \_\_\_\_\_\_\_\_\_\_\_\_**

**c) Assuming 95% confidence interval, determine whether the mean Amount is significantly**

**different from $80000.**

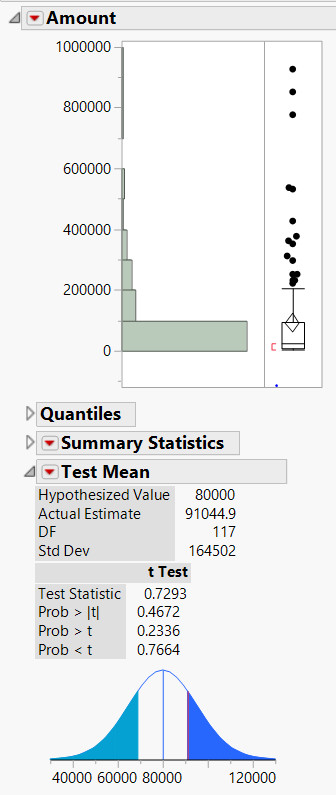
1. When I selected Normal quantile Plot and Normal Plot, I realized that the plot is not normal.
2. It is a two sided test, since we just have to determine if the mean amount is different from the given mean of $80000(µ0).

H0 : µ0 = 80000

Ha : µ0 ≠ 80000

Figure 7

Figure 8

1. My hypothesis would go by

H0 : µ0 = 80000

Ha : µ0 ≠ 80000

So, I selected the test mean from the distribution table and tested mean hypothesis for $80000.

I got the two tailed p value (Prob > |t|) as 0.4672 .

Our Confidence Interval is 95% , hence α = 0.05

P value > > α

Therefore, I fail to reject the null Hypothesis.

Conclusion: We do not have significance evidence that the mean amount is different from $80000 at p value – 0.4672

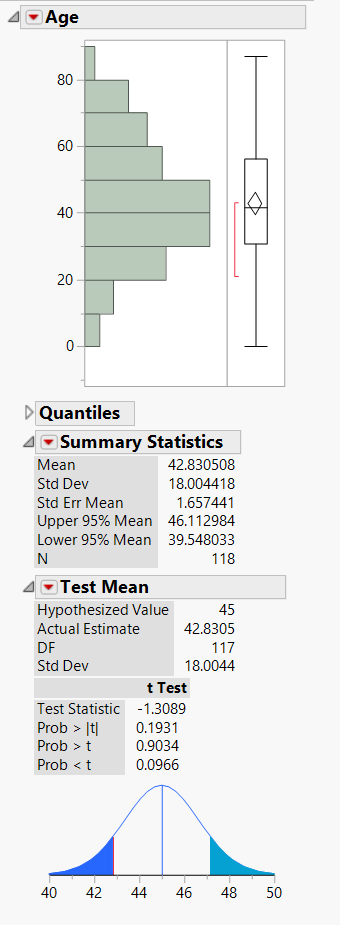
**5. Test the following claims (or hypotheses). (10 points)**

Figure 9

**Claim1: Average Profit Age is greater than 45. Before conducting the test, complete the**

**hypotheses. Is this a one-sided or two-sided test?**

**Ho: \_\_\_\_\_\_\_\_\_\_\_\_**

**Ha: \_\_\_\_\_\_\_\_\_\_\_\_**

**Result (reject or not reject):**

**p-value:**

It is a one-sided test because of the expressions “Average Profit Age is greater than 45”. It is a right tailed test.

H0 : µ0 ≤ 45

Ha : µ0 > 45

I got the two tailed p value (Prob < t) as 0.9034 . (Figure 9)

We will consider confidence Interval is 95% , hence α = 0.05,

P Value = 0.9034

P value >> α

Hence, we fail to reject the null hypothesis, since there is no enough evidence to prove that the mean amount is less than or equal to 45.

**6. Is it appropriate to conduct a two-sample *t*-test to test if a difference exists between the averages**

**Amount paid when Private Attorney is used vs Not Private?**

**a) If appropriate, generate a two-sample *t*-test for both a) and state the null and alternative**

**hypotheses.**

**Ho: \_\_\_\_\_\_\_\_\_\_\_\_**

**Ha: \_\_\_\_\_\_\_\_\_\_\_\_**

**Result (reject or not reject):**

**p-value:**

**b) What conclusion do you reach?**

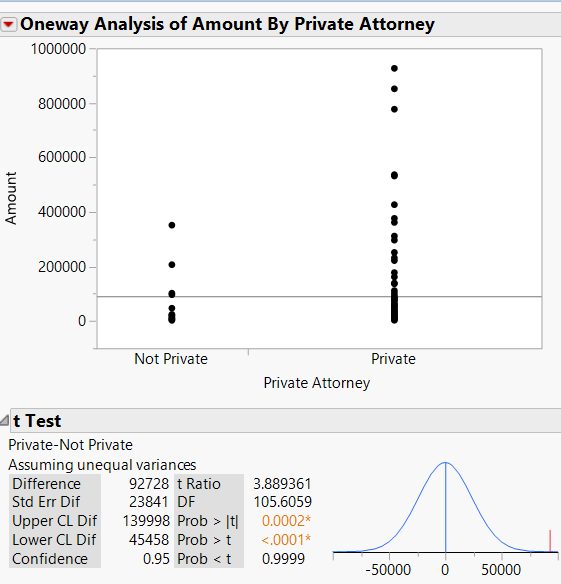
a) Yes, two sample test can be done to test the mean in both samples of similar parameter.

Figure 10

The plot beside gives us a two- sample test of amount paid by private and non-private attorneys.

Here let µ1 be estimate mean of private attorneys

µ2 be estimate mean of non private attorneys

H0 : µ1 = µ2

Ha : µ1 ≠ µ2

P value = 0.0002, α =0.05 (Figure 10)

Since P value is << α, we reject the null hypothesis.

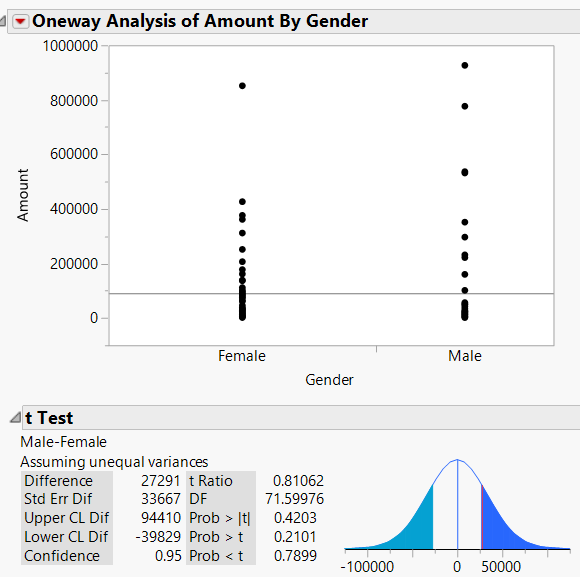
b) Since p value is less than α , there is sufficient evidence to prove that the means of amount paid to private and non -private attorney is different.

**7. Is it appropriate to conduct a two-sample *t*-test to test if a difference exists between the average**

**Amount paid when Female claimant vs male claimant?**

**a) If appropriate, generate a two-sample *t*-test for both a) and State the null and alternative**

**hypotheses. Assume that you are willing to accept a 5% probability of incorrectly rejecting the**

**null hypothesis.**

**Ho: \_\_\_\_\_\_\_\_\_\_\_\_**

**Ha: \_\_\_\_\_\_\_\_\_\_\_\_**

**Result (reject or not reject):**

**p-value:**

**b) What conclusion do you reach?**

a) Yes, two sample test can be done to test the mean in both samples of similar parameter.

The plot beside gives us a two- sample test of amount paid by private and non-private attorneys.

Here let µ1 be estimate mean of amount paid by men

µ2 be estimate mean of amount paid by women

Figure 11

H0 : µ1 = µ2

Ha : µ1 ≠ µ2

P value = 0.4203,α =0.05 (5 % error )(Figure 11)

Since P value is >> α, we fail to reject the null hypothesis.

b) Hence there is insufficient evidence to prove that the amount paid by men is different from amount paid by women.

**8. Analyzing Data from Two Independent Samples (10 Points)**

**You own two plants and have taken 100 random samples of the output of each plant. If you**

**want to test the hypothesis that the plants are performing equally well, what test would you**

**utilize? Where is the test found in JMP?**

Here our null hypothesis would be that there is no difference in the mean of output and the alternate hypothesis is that there is an existing difference.

H0 : difference in mean is 0.

H1 : difference in mean in not equal to 0.

To know that the plants are doing equally well we would go to through the following steps.

JMP file > Analyze > Matched pairs.