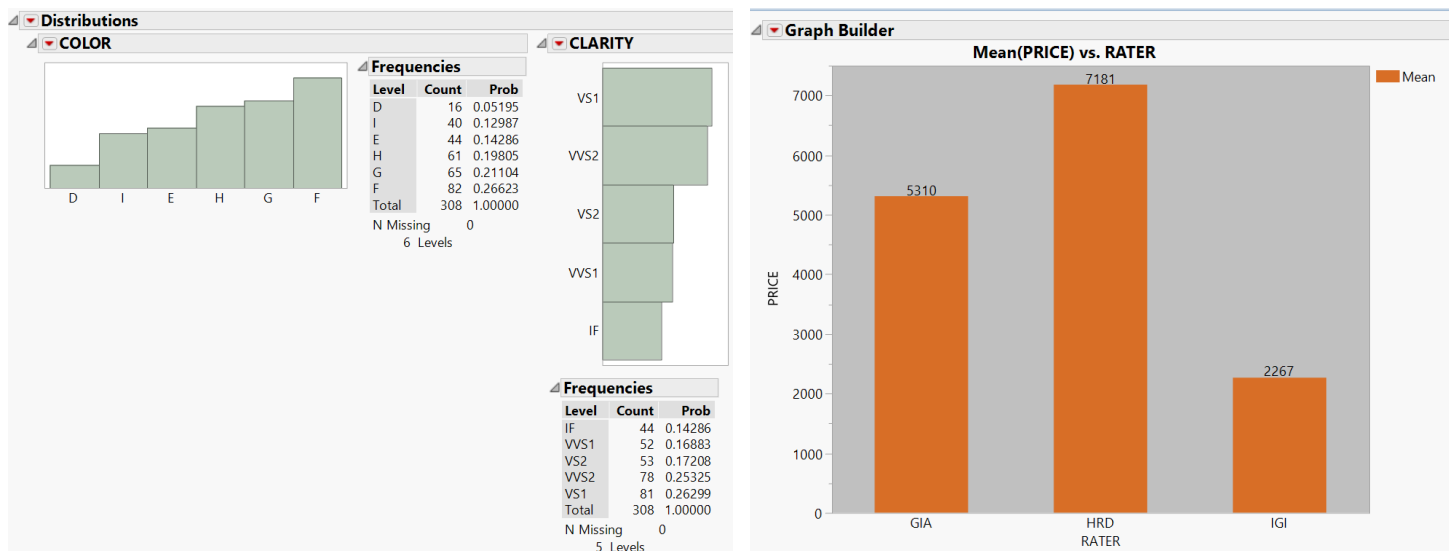


BAN 5733 Exercise 2

Fall 2021 Solution

1. Open JMP Pro and import the Diamonds worksheet from the DIAMONDS.XLS file. **(2 points)**
 - a. Using Analyze > Distribution > Generate bar charts and frequency distribution for the variables **COLOR** (horizontal bar chart) and **CLARITY** (vertical bar chart).
 - i. Change the Layout to order the bars by Ascending Count.
 - b. Using the Graph Builder> Generate a vertical bar chart **RATER** (X) by **PRICE** (Y)
 - i. Ensure the **PRICE** is the mean price.



2. Mosaic Plot and Contingency Table **(2 points)**
 - a. Using Analyze > Fit Y by X generate Contingency Tab for **CLARITY** by **RATER**. The table should have the Count, Row% and Col%.
 - b. Generate a Mosaic plot for the variables **CLARITY** & **RATER**. Display **Percent** in the Mosaic plot.

Contingency Analysis of CLARITY By RATER

Contingency Table

		CLARITY					
		IF	VS1	VS2	VVS1	VVS2	Total
RATER	Count	6	61	36	15	33	151
	Col %	13.64	75.31	67.92	28.85	42.31	
	Row %	3.97	40.40	23.84	9.93	21.85	
	GIA	4	13	15	23	24	79
	HRD	9.09	16.05	28.30	44.23	30.77	
IGI	5.06	16.46	18.99	29.11	30.38		
	34	7	2	14	21	78	
	77.27	8.64	3.77	26.92	26.92		
Total	43.59	8.97	2.56	17.95	26.92		
	44	81	53	52	78	308	

Tests

N	DF	-LogLike	RSquare (U)
308	8	54.295325	0.1116

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	108.591	<.0001*
Pearson	112.747	<.0001*

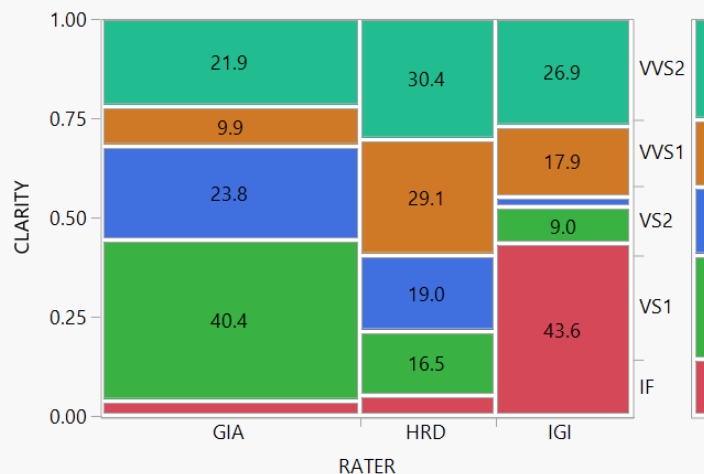
Tests

N	DF	-LogLike	RSquare (U)
308	8	54.295325	0.1116

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	108.591	<.0001*
Pearson	112.747	<.0001*

Contingency Analysis of CLARITY By RATER

Mosaic Plot



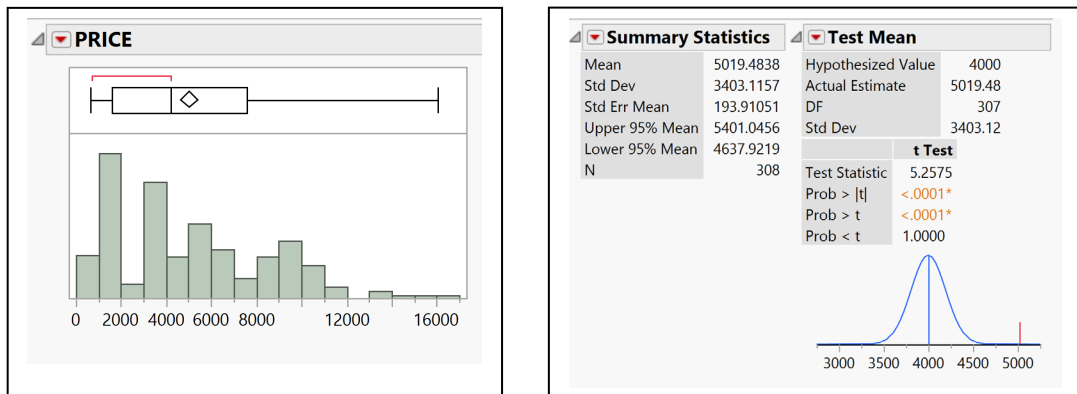
3. The average price in the market for a diamond is \$4,000.00. A jewelry store manager wants to determine if his sales are within the market range. Run the Distribution of **PRICE** and answer the following questions using this information. **(2 points)**
 - a. Write out the manager's problem in testable null and alternative hypothesis terms.
 - b. Test your hypothesis with a two-sided t-test to determine if the price in the dataset is over or under the market average price.
 - c. What suggestion would you give to the jewelry store manager regarding his prices according to this data?

Solution:

The jewelry store manager would like to determine if his sales match the average market price of \$4,000. In order to accomplish this, he must use a one-sample t-test and look at both sides (above and below the average value).

H0: average sale price for diamonds at the store = \$4,000

Ha: average sale price for diamonds at the store \neq \$4,000



From the analysis above, it is clear that the jewelry store is better than the market on average. The average sales for the store are \$5,019 which is significantly higher than the market average of \$4,000. Because the two-sided p-value is less than 0.0001, we can say there is less than a 1 in 10,000 chance of the sample mean falling in the market average's confidence interval. Therefore, we reject the null hypothesis that the store's sales fall in the market range. However, the distribution of sale prices could be affecting this result. Further research needs to be done to determine if outliers are causing the analysis to be inaccurate.

In the meantime, the manager can use this information in various marketing campaigns. He could lower prices to see if sales would increase. He could divide his clientele into buyers of high quality and low quality and market to each or he could more evenly distribute his diamond prices (bringing some up and some down). Before anything long-term or permanent is done with pricing (like liquidating assets), I would advise him to remove any outliers that may exist and re-run the analysis.

4. The jewelry store manager is concerned that one of the three rating companies is undervaluing his diamonds. Run the appropriate test for the following comparisons to see if the manager's concerns are valid regarding the **RATER** and **PRICE** variables: **(2 points)**
 - GIA compared to HRD
 - GIA compared to IGI
 - a. Summarize your findings for the manager and provide a suggestion supported by your results.

Solution:

To address the concerns of the jewelry store manager, the first step is to examine what the average prices are for each rater. We can see the rater HRD has the highest price valuation and IGI has the lowest. However, we must examine these in a slightly different light to see if there truly is a difference. Separate two-sample t-tests were completed for each of the comparisons that the manager wanted to evaluate. This allowed us to say with confidence that there is a statistically significant difference in mean price between the evaluator sets.

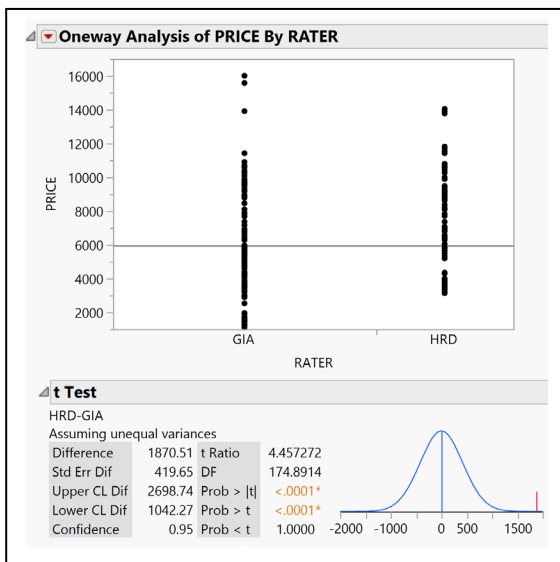
Means and Std Deviations						
Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
GIA	151	5310.4172	3246.9874	264.23609	4788.3117	5832.5227
HRD	79	7180.9241	2897.7118	326.01805	6531.872	7829.9761
IGI	78	2267.1154	2120.6972	240.12167	1788.9721	2745.2587

When comparing GIA to HRD we can see that the mean value is higher for HRD and the two-sample t-test confirms that the mean difference in price is unlikely to be due to chance (mean difference \$1,870.51; p-value<.0001).

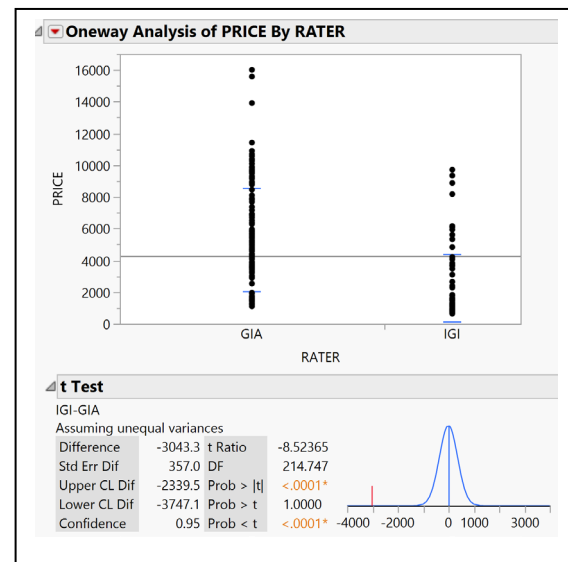
This is not consistent with GIA and IGI raters. The GIA rater is much higher than IGI and, again, this difference is statistically significant (\$3,043.30; p-value<0.0001).

While these results seemingly support moving away from one diamond valuator to HRD, we must take into consideration that we are not looking at other factors involved such as clarity and cut. It may be that the diamond evaluations are perfectly legitimate if we examine the information in this light. Subsequently, a suggestion is made to the manager that additional information be examined before a final decision is made. The subsequent analysis would look at rater while controlling for quality, color, and weight.

GIA compared to HRD



GIA compared to IGI



5. There are assumptions that must be met for most statistical tests to be valid. The one-sample t-test has the following assumptions: **(2 points)**

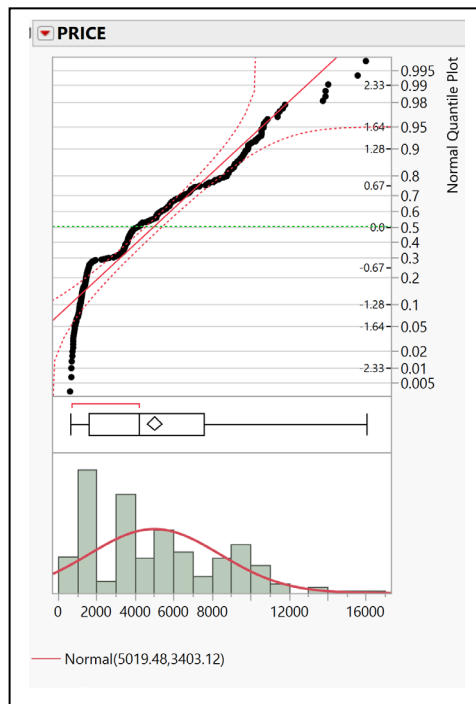
- ✓ The dependent variable must be continuous (interval/ratio).
- ✓ The observations are independent of one another.
- ✓ The dependent variable should be approximately normally distributed.
- ✓ The dependent variable should not contain any outliers.

- a. Does the above analysis meet these assumptions? Talk to each of the assumptions in terms of your analysis.
- b. Would your response to the store manager change based on this information?

Solution:

Question 3 employed the one-sample t-test which looked at two-tails for answer determination. The results for examining the assumptions for that analysis are:

1. The dependent variable must be continuous: Yes, PRICE is a continuous variable. It is ratio-level data.
2. The observations are independent of each other: Yes, the observations are independent of each other. No single observation is based on another in the sample or observations are not connected to one another. Should one observation not be present, it will not affect any other observation in the sample.
3. The dependent variable should be approximately normally distributed: In this instance, the dependent variable is the sale price variable. When examining the variable for normality, a normal fit line was examined on the histogram. A normality test and normal probability plot were also examined. The sale price variable violates the normal distribution assumption. If outliers were removed, this could be a different answer.



4. The dependent variable should not contain any outliers: No outliers exist in the price but some data is skewing the table and needs to be removed.