Zagat survey case study part iii

This part of the case study will explore the application of inferential statistics to the Gate Survey sample data. In Part II, claims were made that attempted to extrapolate the sample data to the population, but they were statistically invalid. So here in Part Ill, I will properly project the sample data Into the population and compare it to the previous methods.

Earlier In the study, the universal estimate used was simply the mean value of the sample taken for the variable Cost. But this does not account for the presence of variation in a sample that could affect the mean.

We can use this value within a standard normal distribution of sample means to try and narrow down the true population mean within a confidence interval of probability. This is preferable because it takes into account the standard error that is intrinsic to all Iterated samples of a population and uses that to capture the population mean (p) within a range. It Is not plausible to calculate the probability that a continuous random variable will assume a specific value in a continuous probability distribution. To establish the boundaries of the confidence interval, we must first select the level of inference (L).

In this case, I will set L= O. 98, its complement being a level of significance (a) of a= O. 02. Since we only have the sample mean and standard deviation, we must use the t-table and find the critical value when a/2= O. 01 (simple differentiation = two-tailed test).

Appendix A establishes the variables and shows the calculation of the Interval boundaries. This allows me to state with 98% confidence that the population mean Cost will fall between $39. 14 and $44. 76. The

management should not be concerned with the number of restaurants included in the sample.

A larger sample would only lower the standard error and the t-value for the level of confidence, thereby narrowing the interval closer to the true population mean.

Also, according to Central Limit Theorem, even though the sample values aren't normally distributed, the mean of the samples will be normally distributed. Using this theorem, we can confidently make predictions about the population without having to capture a large quantity of restaurants in the sample. The margin of error is only about $2. 81, being half of the difference between upper and lower bound values of the confidence interval.

This is not a significant amount, considering the average price of almost $42. 00.

Also, the pre-calculated t-value tables usually stop at 100 degrees of freedom, meaning they only anticipate a sample size of 101 . Further degrees of freedom, from 100 to Infinity, approach an asymptote and only differ by a very insignificant decimal amount. Sampling more than 100 restaurants will narrow the interval, but at a certain point the costs of collecting the data will overshadow the added marginal predictive value. To discover if there is evidence that the mean cost per meal is more than $43, we must set up and perform a null hypothesis test. Sing the 0.

10 level of significance, I can use the sample data to unknown, is strictly greater than $43. This is a right-tailed test, and we will be using the t-table

since the population standard deviation is unknown. The calculations are provided in Appendix B. This says that we can be 90% confident that the true population mean is going to be less than or equal to $43. There is a 10% chance that the null hypothesis was incorrectly accepted (Type 1 Error) and that the true population mean is greater than $43. In summation, there is not strong evidence to purport this claim.

To test the multiple-regression model in Part Sis's significance at 0= 0. 025, we must perform a similar hypothesis test for the regression. This uses the regression based on the sample to test the relationship between the population's independent/dependent variables. Hypothesis testing of a regression uses the F- distribution, which requires the degrees of freedom for both the numerator and denominator. Both of these values, along with the test statistic, are provided in the Excel regression output. An online F-value calculator 1 produced a Critical F-Value of 2.

704.

As outlined in Appendix C, we reject the null hypothesis and conclude that the regression is in fact useful. Using the same null and alternative hypotheses, the same test can be performed for the regression between D©cord and Cost. In this case, we reject the null hypothesis and confirm the alternative, which states that D©cord (being the only independent variable) does in fact affect the Cost. So there is a direct relationship between the D©cord and Cost. The multiple regression model does a good bob of predicting the Cost.

As seen in Part II of the case study, the regression produced a relatively high Adjusted R Square value of . 406, which means that about 76% of the variation in Cost can be explained by the other variables (Appendix D). We can also see from this output that each variable is given a t-Stats value. All of these values are higher than the Critical T-Value (from T-distribution table) for 99 degrees of freedom all the way up to the significance level of 1% (t = 2. 364). And, as shown before, the F-Value hypothesis testing for the regression as a whole proved that the regression was in fact useful.

So all of these claims when used in concert provide strong evidence that this model has a high predictive value.

The best estimation for Cost is going to be the estimate based on the multiple regression. This is because it has been shown in multiple ways to have a strong predictive value, and it takes into account all of the other variables and allows the manipulation of each, responding accordingly. But this is a point estimate, so it would be a stronger prediction if it took into account the standard error and produced a confidence interval. The best estimate for the average cost would be to use the sample mean and standard deviation in a standard normal distribution with a confidence interval.

I think that based on my analyses, I would conclude that there is no doubt the Gate Survey helps consumers make informed decisions about restaurants, as claimed on their website.

Although it may not be a perfect predictor and may contain a small amount of statistical error, it certainly holds a reasonable amount of predictive value,

which allows the consumer to have a rational expectation when walking into

a restaurant; there wouldn't be any big surprises. And considering the

alternatives,