**Quiz #6 (open book, open note) Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Business Analytics Fall, 2016**

1. **(3 points) Disease Survival.** The disease survival data in the Quiz 6 Data file gives, for a random sample of 250 patients who contracted a specific disease, their gender, the age at which they contracted the disease, and how long they survived after contracting the disease. Note: Gender is coded as Female = 1 and Male = 0.
2. (1 point) Is there a significant difference between the average survival length of men and the average survival length of women?

Analyzing with pivot tables, given below is the average survival length for men and women.

|  |  |
| --- | --- |
| **Row Labels** | **Average of Survival length (years)** |
| 0 | 4.794871795 |
| 1 | 5.42481203 |
| **Grand Total** | **5.13** |

1. (1 point) Is the variance in survival length different between men and women?

Analyzing with pivot tables, given below is the variance in survival length for men and women.

|  |  |
| --- | --- |
| **Row Labels** | **Var of Survival length (years)** |
| 0 | 3.578249337 |
| 1 | 4.388243336 |
| **Grand Total** | **4.09246988** |

1. (1 point) What, if any, is the relationship between survival length and age of contraction?
   1. (1/2 point) Using the female data only, construct a scatterplot of survival length vs. age of contraction and compute the correlation coefficient.

Correlation Coefficient: -0.15951

* 1. (1/2 point) Using the female data only, perform a regression analysis using age of contraction as the x or independent variable and survival length as the y or dependent variable. What are the p-values of the slope and the intercept? What can you conclude? Is there any reason (based on these data) we might want to look at males and females separately? Hint: recall your answers to parts a and b.

Intercept: 6.11; p-value: 1.26E-29

Slope: -0.015; p-value: 0.066663

REGRESSION ANALYSIS SUMMARY OUTPUT

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Regression Statistics* | |  |  |  |  |  |  |  |
| Multiple R | 0.159509 |  |  |  |  |  |  |  |
| R Square | 0.025443 |  |  |  |  |  |  |  |
| Adjusted R Square | 0.018004 |  |  |  |  |  |  |  |
| Standard Error | 2.075871 |  |  |  |  |  |  |  |
| Observations | 133 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |  |  |
| Regression | 1 | 14.73783 | 14.73783 | 3.420053 | 0.066663 |  |  |  |
| Residual | 131 | 564.5103 | 4.309239 |  |  |  |  |  |
| Total | 132 | 579.2481 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 6.115391 | 0.414539 | 14.75228 | 1.26E-29 | 5.295335 | 6.935448 | 5.295335 | 6.935448 |
| X Variable 1 | -0.01457 | 0.007876 | -1.84934 | 0.066663 | -0.03015 | 0.001015 | -0.03015 | 0.001015 |

1. **(3 points) Process Improvement**. A particular task in the claims department of a large insurance company is known to have an average time of 186 minutes and a standard deviation of 18 minutes. After a process improvement (Kaizen) event, the task has been re-structured, and it is hoped that the average time and standard deviation will both have decreased. Data from a sample of 85 observations of the new times (after restructuring) are shown in the Quiz 6 data spreadsheet.
2. (1.5 points) Perform an appropriate statistical test to determine whether or not the average time has been reduced. State your null hypothesis, compute your test statistic, determine your rejection region, and state your conclusion. Use a 95% confidence level. Also, compute the p-value.

Null Hypothesis: average time >= 186

Std Error: 18/SQRT(85) = 1.95

Mean of the sample: 181

Test Statistic: (181-186)/1.95 = -2.56

-NORMSINV(0.95): -1.64

Since test statistic is lesser than -1.64, we reject the null hypothesis and conversely we can conclude that the average time has been reduced.

p-value: 0.005234

1. (1.5 points) Perform an appropriate statistical test to determine whether or not the standard deviation (or, equivalently, the variance) has been reduced. State your null hypothesis, compute your test statistic, determine your rejection region, and state your conclusion. Use a 95% confidence level.

Null Hypothesis: variance >= 18^2

Mean: 181

p-value: CHISQ.TEST() = 0.02283

Test Statistic:

**3. (4 Points) Seed Corn**

Use the seed corn data in the quiz 6 spreadsheet to answer this problem.

1. Histogram (2 points) First, for each of the 206 observations, compute the A/F ratio (that is, actual sales divided by forecasted sales). Second, develop a histogram of the A/F ratios using the following “bins”: 0 to 0.1, 0.1 to 0.2, 0.2 to 0.3, and so on out to 3.2 to 3.3. Hint: you need to construct a table showing how many observations have A/F ratios placing them into each of the bins. There is no need to construct a graphical form of the histogram.
2. Normal Distribution Test (2 points). Compute A = sample average of the A/F ratios and S = sample stdev of the A/F ratios. Perform a statistical test to determine whether the A/F ratios are normally distributed (that is, they come from a normal distribution wiih a mean of A and a standard deviation of S. Hint: In part A you have constructed a histogram of Actual Counts of the A/F ratios. You should develop a histogram of Expected Counts of the A/F ratios under the *hypothesis that they come from a normal distribution with a mean of A and a standard deviation of S.* Note that the expected count of A/F ratios in a particular bin will be given by multiplying: [the probability that an observed A/F ratio lies in that particular bin] x [total # of observations----which in this case is 206]. **Report the p-value of your test and state whether you will reject or fail to reject the null hypothesis.**