**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?

|  |  |  |
| --- | --- | --- |
| **t-Test: Two-Sample Assuming Unequal Variances** | | |
|  |  |  |
|  | ***Male survival*** | ***Female survival*** |
| **Mean** | **4.794871795** | **5.42481203** |
| **Variance** | **3.578249337** | **4.388243336** |
| **Observations** | **117** | **133** |
| **Hypothesized Mean Difference** | **0** |  |
| **df** | **248** |  |
| **t Stat** | **-2.49831475** |  |
| **P(T<=t) one-tail** | **0.006563411** |  |
| **t Critical one-tail** | **1.651021013** |  |
| **P(T<=t) two-tail** | **0.013126823** |  |
| **t Critical two-tail** | **1.969575654** |  |

**The mean survival length for men is less than that for women, and the small p-value tells us that the difference is significant.**

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

|  |  |  |
| --- | --- | --- |
| **F-Test Two-Sample for Variances** | |  |
|  |  |  |
|  | ***Male Survival*** | ***Female Survival*** |
| **Mean** | **4.794871795** | **5.42481203** |
| **Variance** | **3.578249337** | **4.388243336** |
| **Observations** | **117** | **133** |
| **df** | **116** | **132** |
| **F** | **0.815417255** |  |
| **P(F<=f) one-tail** | **0.130472182** |  |
| **F Critical one-tail** | **0.741679429** |  |
|  |  |  |

**Although the variance for men is less than that for women, the comparatively high p-value (greater than 10%) tells us that the difference is not even marginally significant.**

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 = -0.1595

* 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.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 | 0.066662833 |  |  |  |  |
| Residual | 131 | 564.5103 | 4.309239 |  |  |  |  |  |  |
| Total | 132 | 579.2481 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 90.0%* | *Upper 90.0%* |  |
| Intercept | 6.115391 | 0.414539 | 14.75228 | 1.26E-29 | 5.295334509 | 6.93544751 | 5.428679138 | 6.80210288 |  |
| Age | -0.01457 | 0.007876 | -1.84934 | 0.066663 | -0.03014523 | 0.00101519 | -0.027611819 | -0.001518218 |  |
|  |  |  |  |  |  |  |  |  |  |

**The p-value for the intercept is extremely small, telling us that the intercept of 6.11 is statistically significant. The p-value for the slope is about 6.6% which tells us that the slope is only marginally significant---it is not significant at the 95% level.**

**Our analysis in part a showed that the average survival length was different depending upon gender, so we would want to account for that difference.**

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.

We could use either a t-test or a z-test. I shall show a z-test.

Null Hypothesis: Mean time >= 186.

Sample mean = 181.14

Sample std. dev. = 15.52

Sample size, n = 85

Standard error = 15.52/sqrt(85) = 1.684

Test statistic = [181.14 – 186] / 1.684 = -2.89

Rejection Region: Reject if test statistic < -normsinv(.95) = -1.645

Conclusion: Reject the null hypothesis.

p-value = normsdist(-2.89) = 0.001926 (less than 1%)

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.

**Note that variance = [standard deviation] ^ 2**

**Null Hypothesis: variance >= 18^2 = 324**

**Sample variance = 15.52 ^2 = 241 (approximately)**

**Test statistic = (n-1) x sample variance/hypothesized variance =**

**= 84 x 241/324 = 62.48 (approx.)**

**Critical value = chisq.inv(.05, 84) = chiinv(0.95, 84) = 63.87**

**Rejection Region: reject for all values of the test statistic < 63.87.**

**Since the test statistic lies in the rejection region, we should reject the null hypothesis.**

**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.

|  |  |  |
| --- | --- | --- |
| **Bottom of bin** | **Top of bin** | **Actual count** |
| 0 | 0.1 | 4 |
| 0.1 | 0.2 | 7 |
| 0.2 | 0.3 | 17 |
| 0.3 | 0.4 | 13 |
| 0.4 | 0.5 | 16 |
| 0.5 | 0.6 | 13 |
| 0.6 | 0.7 | 17 |
| 0.7 | 0.8 | 23 |
| 0.8 | 0.9 | 21 |
| 0.9 | 1 | 25 |
| 1 | 1.1 | 21 |
| 1.1 | 1.2 | 9 |
| 1.2 | 1.3 | 9 |
| 1.3 | 1.4 | 2 |
| 1.4 | 1.5 | 3 |
| 1.5 | 1.6 | 0 |
| 1.6 | 1.7 | 0 |
| 1.7 | 1.8 | 1 |
| 1.8 | 1.9 | 0 |
| 1.9 | 2 | 1 |
| 2 | 2.1 | 0 |
| 2.1 | 2.2 | 0 |
| 2.2 | 2.3 | 0 |
| 2.3 | 2.4 | 0 |
| 2.4 | 2.5 | 1 |
| 2.5 | 2.6 | 1 |
| 2.6 | 2.7 | 1 |
| 2.7 | 2.8 | 0 |
| 2.8 | 2.9 | 0 |
| 2.9 | 3 | 0 |
| 3 | 3.1 | 0 |
| 3.1 | 3.2 | 0 |
| 3.2 | 3.3 | 1 |

1. 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 with 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.**

**In the table on the following page, I used the following function to compute probabilities for the various bins:**

**Bin Probability =**

**normdist(top of bin, mean, stdev,1) – normdist(bottom of bin, mean, stdev,1)**

**where**

**mean = 0.77705747, and**

**stdev = 0.4411297**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Bottom of bin** | **Top of bin** | **Actual count** | **Probability** | **Expected Count** |
| 0 | 0.1 | 4 | 0.0233378 | 4.807583 |
| 0.1 | 0.2 | 7 | 0.0330004 | 6.798084 |
| 0.2 | 0.3 | 17 | 0.044336 | 9.133209 |
| 0.3 | 0.4 | 13 | 0.056594 | 11.65836 |
| 0.4 | 0.5 | 16 | 0.0686376 | 14.13934 |
| 0.5 | 0.6 | 13 | 0.0790917 | 16.29288 |
| 0.6 | 0.7 | 17 | 0.0865919 | 17.83793 |
| 0.7 | 0.8 | 23 | 0.0900744 | 18.55532 |
| 0.8 | 0.9 | 21 | 0.0890231 | 18.33876 |
| 0.9 | 1 | 25 | 0.0835953 | 17.22064 |
| 1 | 1.1 | 21 | 0.0745828 | 15.36406 |
| 1.1 | 1.2 | 9 | 0.0632227 | 13.02388 |
| 1.2 | 1.3 | 9 | 0.0509196 | 10.48943 |
| 1.3 | 1.4 | 2 | 0.0389649 | 8.026771 |
| 1.4 | 1.5 | 3 | 0.0283296 | 5.835889 |
| 1.5 | 1.6 | 0 | 0.0195696 | 4.031347 |
| 1.6 | 1.7 | 0 | 0.0128441 | 2.645881 |
| 1.7 | 1.8 | 1 | 0.0080094 | 1.649935 |
| 1.8 | 1.9 | 0 | 0.0047454 | 0.977552 |
| 1.9 | 2 | 1 | 0.0026713 | 0.550287 |
| 2 | 2.1 | 0 | 0.0014287 | 0.294316 |
| 2.1 | 2.2 | 0 | 0.000726 | 0.14956 |
| 2.2 | 2.3 | 0 | 0.0003505 | 0.072209 |
| 2.3 | 2.4 | 0 | 0.0001608 | 0.033124 |
| 2.4 | 2.5 | 1 | 7.008E-05 | 0.014437 |
| 2.5 | 2.6 | 1 | 2.902E-05 | 0.005978 |
| 2.6 | 2.7 | 1 | 1.142E-05 | 0.002352 |
| 2.7 | 2.8 | 0 | 4.268E-06 | 0.000879 |
| 2.8 | 2.9 | 0 | 1.516E-06 | 0.000312 |
| 2.9 | 3 | 0 | 5.115E-07 | 0.000105 |
| 3 | 3.1 | 0 | 1.64E-07 | 3.38E-05 |
| 3.1 | 3.2 | 0 | 4.995E-08 | 1.03E-05 |
| 3.2 | 3.3 | 1 | 1.446E-08 | 2.98E-06 |

**If we use the chitest function to compute the p-value for the test we get:**

**p-value = chitest(actual count range, expected count range = 0.**

**This tells us we should definitely reject the null hypothesis.**