|  |  |  |
| --- | --- | --- |
| STATS 424 - EXAM #1 | **SHOW YOUR WORK!** | NAME\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |

1. 1 pt each) Identify each of the following statements as True or False.
   1. A t-test for independent samples can be used to analyze data collected in a paired experiment.

**False**

**It is confusing at first, because I think they are t- tests, just different calculation methods.**

**The same way of t-test can be conducted for …**

* 1. Experiment-wise type I error rates can never exceed those for individual tests.

**True**

**Never lower than the individual test**

* 1. There is probably no trend in data if the sample autocorrelation coefficient is positive and large.

**False**

* 1. Blocking controls for important confounding variables that are known prior to experimentation.

**True**

* 1. Randomization tests assume normal populations.

**False**

1. A trend occurs when a response increases or decreases over time for reasons unrelated to experimental factors.
2. 2 pts) Describe how a trend may develop in experimentation.

**A positive trend might develop if technicians get better at performing experiments. Or a downward trend may develop from equipment wear.**

1. 3 pts) Describe how a trend may be confounded with an experimental factor if run order is not randomized.

**If all the low values of a factor are implemented first, followed by all the high values, then the effect of that factor will be confounded with a trend.**

1. The numbers of vehicles entering a city expressway on three consecutive ramps is given by the random variables X1, X2 and X3. These have the vector of means and variance-covariance matrix given below. A traffic engineer wants to model the total number of vehicles entering the expressway in an hour, given by the linear combination L = X1 + X2 + X3.

1. 2 pts) Find the mean of L.

**The mean of L is given by.**

1. 2 pts) Find the standard deviation of L.

**The variance of L is given by. The standard deviation is therefore**

1. 2 pts) Assume that L is normally distributed. What interval contains the central 75% of outcomes of L?

**The value of z.125 = 1.15, so the confidence interval is given by 45 ± 1.15(9.06) = 45 ± 10.4 = (34.6, 55.9).**

1. 2 pts) What are the expected value and standard deviation for two hours of traffic?

**Let S = L1 + L2. The mean of S is 2(45) = 90, and the standard deviation is**

1. 2 pts) Note that COV(X2, X3) = 0. Does this mean there is no relationship between X1 and X2? Explain.

**No, it only means that they are linearly independent. They could still have a non-linear relationship.**

**Only under normal assumption, the correlation =0 will imply that two variables are independent.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Pain Reliever | 1 | 2 | 3 | 4 |
| Sample size | 6 | 6 | 6 | 6 |
| Mean | 7.8 | 9.1 | 6.7 | 11.8 |
| Std. Dev. | 2.5 | 2.2 | 1.8 | 2.8 |

1. A randomized experiment compared four pain relievers. The response was the time (minutes) until noticeable relief from headache pain. The results of the experiment are summarized in the table at right.
2. 2 pts) Calculate the grand mean.

**The grand mean is**

1. 2 pts) Calculate the treatment sum of squares.

**SST**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source | df | SS | MS | F |
| Treatment | **3** | **86.94** | **28.98** | **5.23** |
| Error | **20** | **110.85** | **5.54** |  |
| Total | **23** |  |  |  |

1. 2 pts) Calculate the residual sum of squares.

**SSE**

1. 3 pts) Use the results to complete the ANOVA table.
2. 2 pts) Perform an F test at the 5% level of significance to determine if there are any significant differences among the four pain relievers.

**The F statistic is 4.09, which exceeds the table value of F3,20;.05 = 3.10. Therefore, we reject the null hypothesis, and conclude that there is a difference among the four means.**

1. 4 pts) Regardless of the results of the F test, calculate 95% confidence intervals, and carefully compare them to determine significant differences among the mean relief times.

**The standard error is , and the table critical value is t20;.025 = 2.086, giving a 95% margin of error of 2.086(.96) = 2.0. The confidence intervals for pain relievers 1, 2, 3 and 4 are ± 2.0 = (5.8, 9.8), (7.1, 11.1), (4.7, 8.7), and (9.8, 13.8). The first three pain relievers do not differ significantly, while pain reliever 4 differs significantly from reliever 3, and from reliever 1 with borderline significance.**

1. 3 pts) Construct a 95% confidence interval for the difference in mean relief time between pain relievers 3 and 4.

**The standard error for comparing two means is The confidence interval is given by. We are 95% confident that the mean for pain reliever 4 exceeds the mean for pain reliever 3 by 2.27 to 7.93. In other words, so the means of the two pain relievers differ significantly at the  = .05 level of significance.**

1. 2 pts) Test for equal variances for pain relievers 3 and 4 at the 10% level of significance. Is it reasonable to conclude that these two variances are equal? (Don’t forget to state the hypotheses.)

**H0: HA:**

**The test statistic is The 5% critical value is F5,5;.05 = 5.05, so we retain the null hypothesis and conclude that the variances do not differ at the 10% level of significance.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Cell Phone Provider** | | |  |
|  | **Store** | **Online** | **Other** | **Total** |
| **Male** | **55 (50)** | **30 (40)** | **15 (10)** | **100** |
| **Female** | **45 (50)** | **50 (40)** | **5 (10)** | **100** |
| **Total** | **100** | **80** | **20** | **200** |

1. Random samples of 100 male and 100 female university students were selected. They were asked where they shopped most for textbooks, (book store, online, or other). The results are given in the table at right. The study wanted to determine if the proportions differed by gender.
   1. 3 pts.) Write the expected cell counts in the table.
   2. 3 pts.) Answer the research question by conducting a test at the  = .05 level of significance.

**The critical value for the test is . We retain the null hypothesis. The data do not indicate different proportions by gender for textbook source. These two populations have equal proportions.**

* 1. 2 pts.) Is this a test of homogeneity or independence? Explain your answer.

**This is a test of homogeneity because independent samples were taken from each population.**

* 1. 2 pts.) Calculate the probability that books are purchased online for a male. Do the same for females. How do they compare?

**P{Online | male} = 30/100 = .3. For females, P{Online | female} = 50/100 = .5. The females are more likely than males to buy a textbook online.**