Statistics 515 - Spring 2001 – Exam 2 Answers

Part I: Answer seven of the following eight questions. If you complete more than seven, I will grade only the first seven. Five points each.

1) Define what is meant by "significance level (α) of a test." **The probability of rejecting the null hypothesis when it is true.** or **The probability of committing a Type I error when H_0 is true**.

2) Define what is meant by "Type II error". **The error of failing to reject the null hypothesis when it is false.**

3) Define what is meant by "the p-value (or observed significance level) of a test". **The probability of observing a test statistic as extreme as the one observed, or more extreme, given that the null hypothesis is true.** or **The smallest alpha level at which the null hypothesis would be rejected.**

4) (Circle the correct answers) A test results in a p-value of 0.034. For α=0.05 we **reject** H_0. For α=0.01 we **accept** H_0.

5) For testing H_0:μ=0 vs. H_A:$\mu \neq 0$, SAS returns a t-statistic of –1.5 and a p-value of 0.086. What would the p-value be for testing H_A:μ<0? **0.086/2 = 0.043 (Draw the picture!)**

6) (Circle the correct answers) If you increase the sample size, you will likely **decrease** the width of a confidence level, and be **more** likely to reject a false H_0 in a test of hypotheses.

7) It is desired to H_0:σ^2=1 vs. H_A:σ^2>1 using a chi-squared test. What would the rejection region be for a sample size of 10 and α=0.10? **Greater than 14.6837 (df=10-1=9)**

8) The output at the bottom of the page was generated using PROC TTEST and PROC INSIGHT on SAS for random samples from two data sets N and S. In terms of μ_N, μ_S, σ^2_N and σ^2_S, what null and alternate hypotheses are being tested by the p-values labeled A and B?

A) H_0: $\mu_N = \mu_S$ H_A: $\mu_N \neq \mu_S$ B) H_0: $\sigma^2_N = \sigma^2_S$ H_A: $\sigma^2_N \neq \sigma^2_S$

```
                         The TTEST Procedure

                               T-Tests

   Variable    Method           Variances      DF     t Value    Pr > |t|

   value       Pooled           Equal          20        1.55     0.1364 ◄ A
   value       Satterthwaite    Unequal      19.8        1.56     0.1336

                         Equality of Variances

       Variable    Method      Num DF    Den DF    F Value    Pr > F

       value       Folded F        11         9       1.18    0.8148  ◄ B
```

Part II: Answer every part of the next two problems. Read each problem carefully, and show your work for full credit. Twenty points each.

1) A manufacturer of alkaline batteries wants to be reasonably certain that fewer than 5% of its batteries are defective. Suppose 300 batteries are randomly selected from a very large shipment; each is tested and 10 defective batteries are found.
A) State the appropriate null and alternate hypotheses to test if fewer than 5% of the batteries are defective.

H_0: p=0.05 (not evidence that fewer than 5% are defective)
H_A: p<0.05 (evidence that fewer than 5% fewer are defective)

B) Demonstrate that the sample size of 300 is large enough to conduct the test in A (do not conduct the test).

np = 300(0.05) =15 ≥ 5 **So the sample size is large enough.**
n(1-p) = 300(0.95) = 285 ≥ 5

C) Instead of performing a test, it may also be desirable simply to form a confidence interval for the actual percent of defective batteries. Construct such a 95% confidence interval.

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.033 \pm 1.96\sqrt{\frac{0.033(1-0.033)}{300}} = 0.033 \pm 0.020 = (0.013, 0.053)$$

2) The EPA is testing whether a manufacturing plant pollutes the nearby river. To test this they gather six samples of water upstream from (before) the plant and six samples of water downstream from (right after) the plant.
A) State the appropriate null and alternate hypothesis for determining whether the EPA has significant evidence that the plant is polluting the river.

H_0: $\mu_{up}=\mu_{down}$ or $\mu_{up}-\mu_{down}=0$ (no evidence of pollution)
H_A: $\mu_{up}<\mu_{down}$ or $\mu_{up}-\mu_{down}<0$ (evidence of pollution)

B) What assumptions must be true in order to test this hypothesis?

The amount of pollution in the water both above and below the plant must follow a normal distribution.
The samples of water above the plant must be taken randomly and the samples of water below the plant must be taken randomly.
The samples of water above the plant must be taken independently of the samples of water below the plant.
The variances of the two samples must be equal (because the sample size is small).

C) Assume that the assumptions in part B are met, that the observations from upstream had a mean of 29.6167 and a standard deviation of 0.961, and the values downstream had a mean of 32.1000 and a standard deviation of 1.302. Test the hypothesis in A at an α=0.05 level.

$$s_p^2 = \frac{(n_{up}-1)s_{up}^2 + (n_{down}-1)s_{down}^2}{n_{up}+n_{down}-2} = \frac{(6-1)(0.961^2)+(6-1)(1.302^2)}{6+6-2} = 1.309$$

$$t_{df=10} = \frac{(\bar{x}_{up}-\bar{x}_{down})-(\mu_{up}-\mu_{down})}{\sqrt{\frac{s_p^2}{n_{up}}+\frac{s_p^2}{n_{down}}}} = \frac{(29.6167-32.100)-0}{\sqrt{\frac{1.309}{6}+\frac{1.309}{6}}} = \frac{-2.4833}{0.6606} = -3.759$$

We compare this to -1.812 and reject the null hypothesis. We do have evidence that the plant is polluting. (If you reversed it and made it down-up, we would have had +3.579 and compared it to +1.812, and still rejected.)