## STAT 350 Exam 2 Review Problems

**1.** A random sample of 26 offshore oil workers took part in a simulated escape exercise, and their times (in seconds) to complete the escape were recorded. The sample mean is 370.69 sec. and the sample standard deviation is 24.36 sec. Construct a 95% lower confidence bound for the true average escape time. Interpret your interval and write down the critical value. Do the problem assuming that you don't know the data (be sure to include the code and the value and code of the critical value) and using the R output below (write down the code that was used to generate the output):

```
            One Sample t-test

data:  escape$time
95 percent confidence interval:
 362.53        Inf
```

**2.** The life in hours of a battery is known to be approximately normally distributed. The manufacture claims that the average battery life exceeds 40 hours. A random sample of ten batteries has a mean life of 40.5 hours and sample standard deviation of 1.25 hours. Carry out the appropriate hypothesis test with a significance level of 0.05. Do the problem assuming that you don't know the data (be sure to include the code) and using the R output below (write down the code that was used to generate the output):

```
            One Sample t-test

data:  battery$lifetime
t = 1.2649, df = ?, p-value = 0.1188
```

**3.** The overall distance traveled by a golf ball is tested by hitting the ball with Iron Byron, a mechanical golfer with a swing that is said to emulate the legendary champion, Byron Nelson. Ten randomly selected balls of two different brands are tested and the overall distance measured to determine if the two brands are different. Please provide the code or work for the appropriate parts. The data (in yards) follows:

Brand 1: 275, 286, 287, 271, 283, 271, 279, 275, 263, 267
Brand 2: 258, 244, 260, 265, 273, 281, 271, 270, 263, 268

The data is summarized in the following table:

| | n | $\bar{x}$ | s |
|---|---|---|---|
| Brand 1 | 10 | 275.7 | 8.03 |
| Brand 2 | 10 | 265.3 | 10.045 |
| 1 - 2 | 10 | 10.4 | 15.005 |

If the situation is two-sample independent, use the following data:

```
distance <- c(275, 286, 287, 271, 283, 271, 279, 275, 263, 267, 258, 244,
   260, 265, 273, 281, 271, 270, 263, 268)
brand <- c(rep("Brand1",10),rep("Brand2",10))
```

If the situation is two-sample paired, use the following data:

```
Brand1 <- c(275, 286, 287, 271, 283, 271, 279, 275, 263, 267)
Brand2 <- c(258, 244, 260, 265, 273, 281, 271, 270, 263, 268)
```

a) Which procedure is the most appropriate, two-sample independent or two-sample paired? Please explain your answer. If you choose a matched pairs procedure, please state the common characteristic that makes these data paired.

b) Should you use a one-sided or two-sided alternative hypothesis? Please explain your answer.

c) Use the four-step procedure to carry out a hypothesis test to determine whether the mean overall distance for brand 1 and brand 2 are different. Assume a significance level of 0.05.
d) Find and interpret the appropriate 95% confidence interval or bound that corresponds with part c). Be sure to provide the value and code for the critical value.
e) Why are parts c) and d) saying the same thing?
f) In practical terms, are the two brands different? Additional information is required.

**4.** The Indiana State Police wish to estimate the average mph being traveled on the Interstate Highways which cross the state. If the estimate needs to be within ±5 mph of the true mean with 95% confidence and the estimated population standard deviation is 25 mph, how large a sample size must be taken? Please provide the code or work.

**5.** A laboratory is testing the concentration level in mg/mL for the active ingredient found in a pharmaceutical product. In a random sample of ten vials of the product, the mean and the sample standard deviation of the concentrations are 2.58 mg/mL and 0.09 mg/mL, respectively. Find a 95% confidence interval for the mean concentration level in mg/mL for the active ingredient found in this product. Please interpret your result. Do the problem assuming that you don't know the data (be sure to include the code and the value and code of the critical value) and using the R output below (write down the code that was used to generate the output):

```
                One Sample t-test

data:  escape$time
95 percent confidence interval:
2.515617  2.64438
```

**6.** An investigator wishes to estimate the difference between two population mean lifetimes of two different brands of batteries under specified conditions. If the population standard deviations are both roughly 2 hr. and the sample size from the first brand is twice the sample size from the second brand, what values of the sample sizes will be necessary to estimate the half-width to within 0.5 hours with 99% confidence?
This is a difficult question.

**7.** The following are summary data on the proportional stress limits for two different types of woods, red oak and Douglas fir. We are interested if the proportional stress limits are different. Note that only the summary data is available for this question.

| Type of Wood | Sample Size | Sample Mean | Sample Standard Deviation |
|---|---|---|---|
| Red oak | 50 | 8.51 | 1.52 |
| Douglas fir | 62 | 7.69 | 3.25 |

a) Which procedure is the most appropriate, two-sample independent or two-sample paired? Please explain your answer. Stating that the sample sizes are different or there is no information for the paired situation is an incorrect answer. If you choose a matched pairs procedure, please state the common characteristic that makes these data paired.
b) Should you use a one-sided or two-sided alternative hypothesis? Explain.
c) Perform a hypothesis test $\alpha = 0.05$ to determine if the stress limits are different for the two types of woods.
d) Find the appropriate 95% confidence interval or bound that corresponds to part c). Please interpret your result and write down the critical value.

e) Why are parts c) and d) saying the same thing?
f) What practical answer would you tell your supervisor concerning the difference between the average proportional stress limits for the two types of trees?

**8.** The accompanying summary data on the ratio of strength to cross-sectional area for knee extensors is from the article "Knee Extensor and Knee Flexor Strength: Cross Sectional Area Ratios in Young and Elderly Men": I assuming that these are self-identified men. Note that only the summary data is available for this question.

| Group | Sample Size | Sample Mean | Sample Standard Deviation |
|---|---|---|---|
| Young Men | 50 | 7.47 | 0.44 |
| Elderly Men | 45 | 6.71 | 0.56 |

a) Which procedure is the most appropriate, two-sample independent or two-sample paired? Explain. Stating that the sample sizes are different or there is no information for the paired situation is an incorrect answer. If you choose a matched pairs procedure, please state the common characteristic that makes these data paired.
b) Does the data suggest that the true average ratio for young men exceeds that for elderly men? Carry out a test of significance using $\alpha = 0.01$.
c) Find and interpret the appropriate confidence interval or bound at a 99% confidence level. Please calculate and write down the critical value.
d) Why are parts a) and b) saying the same thing?
e) What practical answer would you tell the researcher concerning the difference in the ratio for young men versus elderly men?

**9.** Coronary heart disease (CHD) begins in young adulthood and is the fifth leading cause of death among adults aged 20 to 24 years. Studies of serum cholesterol levels among college students, however, are very limited. A 1999 study looked at a large sample of students from a large southeastern university and reported that the mean serum cholesterol level among women is **168 mg/dL** with a **population standard deviation** of **27 mg/dL**.
A more recent study at a southern university investigated the lipid levels in a cohort of sedentary university students. The mean total cholesterol level among ***n* = 71** women was **$\bar{x}$ = 173.7**. You may assume that the population standard deviation is known with a value of **27 mg/dL** and remains the same between the two studies. Is there evidence that the mean cholesterol level among sedentary university women has increased from the average in 1999?

a) Use the four-step procedure to carry out a test of significance at $\boldsymbol{\alpha = 0.05}$. Note that the summary data is all that is available for this question.
b) Researchers are concerned that the power may be too low to detect an increase of 5 mg/dl.
   i. Determine the cutoff value for $\alpha = 0.05$.
   ii. Determine the power associated with an alternative of 173 mg/dL.
   i. How can the researchers obtain larger power in a follow up study if they require the same significance level?

**10.** In continuation of the investigation into cholesterol levels among university students, an additional follow-up study is planned to be conducted with greater statistical power.

Given that the study wants to detect an **increase of 5 mg/dL (alternative of 173 mg/dL)** from the mean cholesterol level among sedentary university women from the 1999 average of **168 mg/dL**, with a **power of 90%**. **Determine the sample size** required to achieve this level of statistical power? You may assume the **population standard deviation of 27 mg/dL** remains unchanged and a significance level of $\boldsymbol{\alpha = 0.01}$.

**11.** Fifteen self-identified adult males between the ages of 35 and 45 participated in a study to evaluate the effect of diet and exercise on blood cholesterol levels. The total cholesterol was measured in each subject initially. Each subject then spent three months participating in an aerobic exercise program and switched to a low-fat diet. After the three months, the total cholesterol level was again measured. It is hoped that the cholesterol level decreased after the three months is over. Please provide the code or work for the appropriate parts. The data are shown in the accompanying tables.

**Table I: Blood Cholesterol Levels for 15 Adult self-identified Males**

| Subject | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Before | 265 | 240 | 258 | 296 | 251 | 245 | 287 | 314 | 260 | 279 | 283 | 240 | 238 | 225 | 247 |
| After | 229 | 231 | 227 | 240 | 238 | 241 | 234 | 256 | 247 | 239 | 246 | 218 | 219 | 226 | 233 |

| | **N** | **Mean** | **StDev** |
|---|---|---|---|
| **Before** | **15** | **261.80** | **24.96** |
| **After** | **15** | **234.93** | **10.48** |
| **Diff (Before - After)** | **15** | **26.87** | **19.04** |

If the situation is two-sample independent, use the following code:

```
cholesterol <- c(275, 286, 287, 271, 283, 271, 279, 275, 263, 267, 258,
   244, 260, 265, 273, 281, 271, 270, 263, 268)
time <- c(rep("before",15),rep("after",15))
```

If the situation is two-sample paired, use the following code:

```
before <- c(265, 240, 258, 296, 251, 245, 287, 314, 260, 279, 283, 240,
   238, 225, 247)
after <- c(229, 231, 227, 240, 238, 241, 234, 256, 247, 239, 246, 218, 219,
   226, 233)
```

a) Which procedure is the most appropriate, two-sample independent or two-sample paired? Please explain your answer. Stating that the sample sizes are different or there is no information for the paired situation is an incorrect answer. If you choose a matched pairs procedure, please state the common characteristic that makes these data paired.

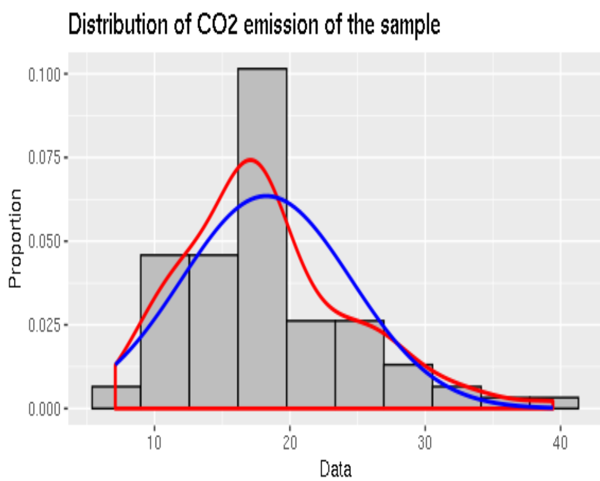
b) Should you use a one-sided or two-sided alternative hypothesis? Please explain your answer.

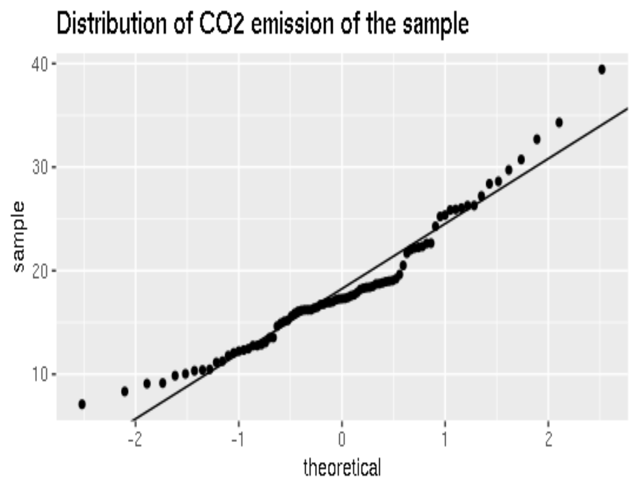
c) Carry out a hypothesis test to determine if the data support the claim that the low-fat diet and aerobic exercise are of value in reducing the mean blood cholesterol levels? Use α=0.05.

d) Find and interpret the appropriate confidence interval or bound at a 95% confidence level. Please calculate and write down the critical value.

e) What practical answer would you tell the researcher concerning the effect of aerobic exercise and a low fat diet on the cholesterol level?

**12.** A simple random sample of 85 automobiles was obtained and the $CO_2$ emissions from each was measured (in hectogram/mi). The following graphs were generated from the above sample.




Using the above figures, is the appropriate assumption valid so that you can perform statistical inference for the population mean $CO_2$ emission? Please explain your answer. Be sure to state the assumption that is being shown from the graphs.