Stat 242 Quiz – Topics Drawn from Sections 2.3, 5.2, and 6.2

What's Your Name? ______________________________

In the 1970's, it was found that giving terminal cancer patients a dietary supplement of ascorbate could prolong their lives. As a follow-up question, researchers wondered if the effect of the ascorbate was different when different organs were affected by the cancer. We have a data set with records for 64 different cancer patients. Each patient had cancer in one of 5 different organs: breast, bronchus, colon, ovary, or stomach. For each patient, we have recorded their cancer type and their survival time. Here we will treat the log of survival time as the response variable. The R code below calculates the mean log survival time for each group of patients.

```r
cancer %>%
  group_by(Organ) %>%
  summarize(
    mean_log_survival_time = mean(Survival)
  )
```

```
## # A tibble: 5 x 2
##   Organ    mean_log_survival_time
##   <chr>                     <dbl>
## 1 Breast                    1396.
## 2 Bronchus                   212.
## 3 Colon                      457.
## 4 Ovary                      884.
## 5 Stomach                    286
```

Define the parameters as follows:

- μ_1 is the mean log survival time for breast cancer patients similar to those enrolled in this study;
- μ_2 is the mean log survival time for bronchus cancer patients similar to those enrolled in this study;
- μ_3 is the mean log survival time for colon cancer patients similar to those enrolled in this study;
- μ_4 is the mean log survival time for ovarian cancer patients similar to those enrolled in this study; and
- μ_5 is the mean log survival time for stomach cancer patients similar to those enrolled in this study

1. We would like to see if the average log survival time for breast cancer is different from the average log survival time for stomach cancer. State the hypotheses for this test as a linear combination of the means for the five groups.

$H_0 : \mu_1 - \mu_5 = 0$

$H_A : \mu_1 - \mu_5 \neq 0$

2. Show how you could find an estimate of the linear combination specified in the hypotheses for part 1. (You can just do the set up, no need to simplify to find an actual number.)

1396 - 286

(See question 3 on other side!)

3. We would like to see if the average log survival time for breast and ovary cancers (treated as a group) are different from the average log survival time for the other three types. State the hypotheses for this test as a linear combination of the means for the five groups.

For your reference, here are our parameter definitions again:

- μ_1 is the mean log survival time for breast cancer patients similar to those enrolled in this study;
- μ_2 is the mean log survival time for bronchus cancer patients similar to those enrolled in this study;
- μ_3 is the mean log survival time for colon cancer patients similar to those enrolled in this study;
- μ_4 is the mean log survival time for ovarian cancer patients similar to those enrolled in this study; and
- μ_5 is the mean log survival time for stomach cancer patients similar to those enrolled in this study

$H_0 : \frac{1}{2}(\mu_1 + \mu_4) - \frac{1}{3}(\mu_2 + \mu_3 + \mu_5) = 0$

$H_A : \frac{1}{2}(\mu_1 + \mu_4) - \frac{1}{3}(\mu_2 + \mu_3 + \mu_5) \neq 0$

. . . or. . .

$H_0 : \frac{1}{2}\mu_1 - \frac{1}{3}\mu_2 - \frac{1}{3}\mu_3 + \frac{1}{2}\mu_4 - \frac{1}{3}\mu_5 = 0$

$H_A : \frac{1}{2}\mu_1 - \frac{1}{3}\mu_2 - \frac{1}{3}\mu_3 + \frac{1}{2}\mu_4 - \frac{1}{3}\mu_5 \neq 0$

Stat 242 Quiz – Topics Drawn from Sections 2.3, 5.2, and 6.2

What's Your Name? ________________________________

In the 1970's, it was found that giving terminal cancer patients a dietary supplement of ascorbate could prolong their lives. As a follow-up question, researchers wondered if the effect of the ascorbate was different when different organs were affected by the cancer. We have a data set with records for 64 different cancer patients. Each patient had cancer in one of 5 different organs: breast, bronchus, colon, ovary, or stomach. For each patient, we have recorded their cancer type and their survival time. Here we will treat the log of survival time as the response variable.

The R code below produces output for a confidence interval and a hypothesis test for the hypotheses

$H_0 : \frac{1}{2}(\mu_1 + \mu_4) - \frac{1}{3}(\mu_2 + \mu_3 + \mu_5) = 0$

$H_A : \frac{1}{2}(\mu_1 + \mu_4) - \frac{1}{3}(\mu_2 + \mu_3 + \mu_5) \neq 0$

Here, the parameters are as follows:

- μ_1 is the mean log survival time for breast cancer patients similar to those enrolled in this study;
- μ_2 is the mean log survival time for bronchus cancer patients similar to those enrolled in this study;
- μ_3 is the mean log survival time for colon cancer patients similar to those enrolled in this study;
- μ_4 is the mean log survival time for ovarian cancer patients similar to those enrolled in this study; and
- μ_5 is the mean log survival time for stomach cancer patients similar to those enrolled in this study

```r
anova_fit <- lm(log_survival_time ~ Organ, data = cancer)
fit.contrast(anova_fit, "Organ", c(1/2, -1/3, -1/3, 1/2, -1/3), conf.int = 0.95)
```

```
##      Estimate Std. Error t value Pr(>|t|) lower CI upper CI
## [1,]     1.13      0.351    3.23  0.00204     0.43     1.83
```

(a) Draw a conclusion about the strength of evidence against the null hypothesis. Your statement should be in the context of this question.

Since the p-value for the test is 0.002, there is strong evidence against the null hypothesis that the mean log survival time for patients with breast cancer or ovarian cancer is the same as the mean log survival time for patients with bronchus, colon, or stomach cancer, in a population of cancer patients similar to those patients enrolled in this study.

(b) Interpret the confidence interval in context. Include a description of the meaning of the phrase "95% confident" as part of your answer.

We are 95% confident that in a population of cancer patients similar to those enrolled in this study, the difference in mean log survival time for patients with breast cancer or ovarian cancer and the mean log survival time for patients with either bronchus, colon, or stomach cancer is in the interval [0.43, 1.83]. For 95% of samples, an interval computed in this way would contain the difference in average log survival times between patients in those two groups.

(this page left blank intentionally)

Stat 242 Quiz – Topics Drawn from Sections 2.3, 5.2, and 6.2

What's Your Name? ______________________________

In the 1970's, it was found that giving terminal cancer patients a dietary supplement of ascorbate could prolong their lives. As a follow-up question, researchers wondered if the effect of the ascorbate was different when different organs were affected by the cancer. We have a data set with records for 64 different cancer patients. Each patient had cancer in one of 5 different organs: breast, bronchus, colon, ovary, or stomach. For each patient, we have recorded their cancer type and their survival time. Here we will treat the log of survival time as the response variable.

The researchers were interested in seeing whether or not the difference in mean log survival times between breast cancer patients and ovarian cancer patients is 0, among patients similar to those enrolled in this study. They expressed this question in the following hypothesis test:

$H_0 : \mu_1 - \mu_4 = 0$

$H_A : \mu_1 - \mu_4 \neq 0$

Here, μ_1 represented the mean for breast cancer patients and μ_4 represented the mean for ovarian cancer patients.

Based on their sample data, an estimate of $\mu_1 - \mu_4$ was 0.408, with a standard error of 0.607.

(a) Show the set up for calculating a t statistic to use for this test, based on the estimate and standard error above. No need to simplify to get an actual number.

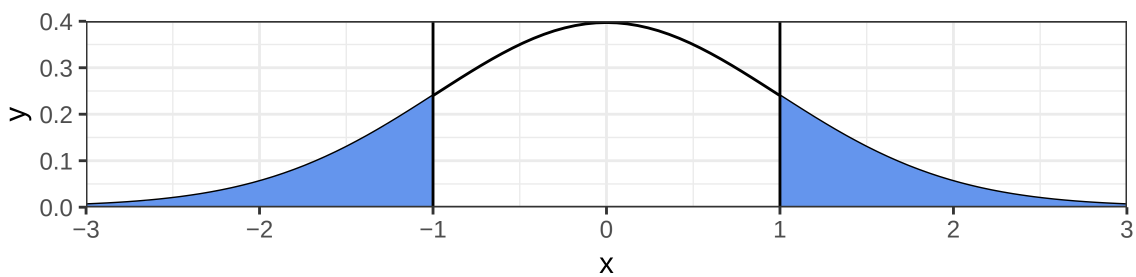
$$\frac{0.408}{0.607}$$

(b) What are the degrees of freedom for the distribution of the t statistic?

$$64 - 5 = 59$$

(c) Suppose the t statistic above worked out to be 1 (it wouldn't, this is just for illustration). Draw a picture illustrating how the p-value for the test would be calculated.

The below is a picture of a t distribution with 59 degrees of freedom. The p-value is the shaded area (the probability of obtaining a t statistic at least as extreme as 1).



(this page left blank intentionally)

Stat 242 Quiz – Topics Drawn from Sections 2.3, 5.2, and 6.2

What's Your Name? ________________________________

In the 1970's, it was found that giving terminal cancer patients a dietary supplement of ascorbate could prolong their lives. As a follow-up question, researchers wondered if the effect of the ascorbate was different when different organs were affected by the cancer. We have a data set with records for 64 different cancer patients. Each patient had cancer in one of 5 different organs: breast, bronchus, colon, ovary, or stomach. For each patient, we have recorded their cancer type and their survival time. Here we will treat the log of survival time as the response variable.

The researchers were interested in the difference in mean log survival times between breast cancer patients and ovarian cancer patients, among patients similar to those enrolled in this study. They expressed this difference as $\mu_1 - \mu_4$, where μ_1 represented the mean for breast cancer patients and μ_4 represented the mean for ovarian cancer patients.

Based on their sample data, an estimate of $\mu_1 - \mu_4$ was 0.408, with a standard error of 0.607.

Use the R output below to answer the following question.

```r
qt(0.975, df = 64)
```

```
## [1] 1.99773
```

```r
qt(0.975, df = 63)
```

```
## [1] 1.998341
```

```r
qt(0.975, df = 62)
```

```
## [1] 1.998972
```

```r
qt(0.975, df = 59)
```

```
## [1] 2.000995
```

(a) What is the degrees of freedom for the t statistic?

$$64 - 5 = 59$$

(b) Show the set up for calculating a 95% confidence interval for $\mu_1 - \mu_4$, based on the estimate and standard error above. No need to simplify to get numbers, just show the set up. You will need the output from just one of the calls to qt above.

$$[0.408 - 2 \times 0.607, 0.408 + 2 \times 0.607]$$

(this page left blank intentionally)