# STAT511 – Exam 2
# Fall 2017

**Honor Pledge:** I have not given, received, or used any unauthorized assistance on this exam.

**Signature:**________________________________________

**Printed Name:**________________________________________

<u>**Instructions:**</u>

- **Open book, open notes, calculator required. No computers or cell phones.**
- **Time limit is 1 hour 50 minutes - strictly enforced!**
- If an answer is in the computer output, use it; don't calculate it by hand.
- Show your work where appropriate. Put your final answer in the box (if provided).
- Make explanations brief and legible.
- All questions are worth 4 points except where noted. Maximum score is 100.
- Computer input/output is provided at the end of the exam.
- The exam contains a total of 11 pages (including computer input/output).
- If you run out of space, you may use the blank area on page 5 or extra paper.

**<u>Questions 1 through 4 (Multiple Testing):</u>** In this group of questions we consider four different multiple testing methods: (1) **Unadjusted (LSD), (2) Tukey (HSD), (3) Bonferroni and (4) Dunnett**. For each of the scenarios below, choose the preferred multiple testing adjustment. **Just state a single preferred method, no need to justify.** A method can be used more than once. (~~3~~ 4 **pts** per question)

1. A study compares 5 treatments using n = 12 per group. Investigators would like to control the experiment wise error rate for the group of all pairwise comparisons (total of 10 tests) but they are also concerned about power. What multiple testing method is preferred?

-2 for Bonf.

Tukey

2. A study compares blood analytes for 2 treatment groups with n = 20 subjects per group. The blood panel includes 11 analytes (Ca, Crea, Cl, etc). <u>For each analyte</u>, a two-sample t-test will be used to compare groups (total of 11 tests). Investigators would like to control the experiment wise error rate. What multiple testing method is preferred?

Bonferoni

3. A study compares 5 active treatments versus a control using n = 16 per group. Investigators are only interested in the comparisons of each of the active treatments versus control (total of 5 tests). They would like to control the experiment-wise error rate but are also concerned about power. What multiple testing method is preferred?

-2 for Bonf

Dunnett.



12

**Questions 1 through 4 (Multiple Testing) continued:**

4. A preliminary study compares 15 treatments but with small sample sizes (n = 4 per group). The goal is to identify a subset of treatments for further study with larger sample sizes. Because this is a preliminary study, investigators are not concerned about the experiment-wise error rate, but they would like to have the highest possible power. What multiple testing method is preferred?

Unadjusted

**Questions 5 through 11 (Observational Study):** In an observational study, **n = 20** subjects are placed into **4 Groups** based on results of a preliminary survey. Then in a lab setting, some response (**Y**) is measured and recorded for each subject. The goal of the study is to compare mean responses between the four groups. The R input and output are labeled **Obs Study**. Use $\alpha = 0.05$. **Note:** The sample sizes were not the same for each group (this is reasonable given that it is an observational study.)

5. After looking the ANOVA table for Model1, the analyst realizes that Model1 is not the correct model for comparing means between the four groups. What did the analyst need to do to get the correct model? (The "fix" corresponds to the missing line of code before Model2.)

Define Group as.factor

InData$Group <- as.factor(InData$Group)

**Note:** All remaining questions are based on **Model2**.

6. Fill in the missing df for Model2.

| | |
|---|---|
| df Group: | 3 |
| df Resid: | 16 |

df Group $= t-1 = 4-1 = 3$

df Resid $= n_T - t = 20-4 = 16$

7. Is the assumption of equal variances satisfied? Briefly Discuss two pieces of evidence provided in the output. As part of the discussion, clearly state the test(s) and/or graph(s) you are considering.

Resids vs Fitted: Looks OK (but not great).

Levene's Test: Large p-value supports equal variances

8. Is the assumption of normality satisfied? Briefly Discuss two pieces of evidence provided in the output. As part of the discussion, clearly state the test(s) and/or graph(s) you are considering.

QQplot: Looks good; points fall close to line.

SW test: Large p-value supports normality.



20

(Dunnett).

9. A colleague considers the results of the Model2 ANOVA Table and the LSMeans1 output. He is puzzled that the F-test is significant (p = 0.0078) but none of the pairwise comparisons are significant (all p-values > 0.05). Briefly explain these results. **Note:** Use only LSMeans1 for this question.

Issue is we are only considering Dunnett Comparisons
- Multiple testing adjustment can eliminate sig. differences
- Only comparing versus Group 1.

10. Now consider the LSMeans2 output. Notice that for the comparison of 1 vs 4 the estimate is -1.98 with p-value = 0.0210. For comparisons 2 vs 3 the estimate is -1.81 with p-value = 0.0055. Your colleague is surprised that the smaller magnitude difference has the smaller (more significant) p-value. Briefly explain these results (to someone with little knowledge of statistics).

-2 for SE

Statistical Sig (p-values) are effected by sample Size! Groups 1 and 4 ($n_1=4, n_4=3$) have smaller sample sizes than Groups 2 and 3 ($n_2=6, n_3=7$).

11. Again considering the LSMeans2 output, your colleague notices that these are unadjusted pairwise comparisons. He suggests that you may get more significant differences if you run the pairwise comparisons with Tukey adjustment. Respond to this suggestion.

No. After Tukey adjustment you will get fewer (or the same #) of sig. differences.

**Questions 12 through 17 (Machines):** A manufacturer produces a part that is supposed to be Y = 10 units in size. They are interested in comparing 2 machines that produce this same part. The R input and output are labeled **Machines**. Use α = 0.05. **Important note:** When I ask about the difference between two tests, you will want to consider the hypotheses being testing and also the assumptions required for the test to be valid. I am NOT asking about statistical significance when I ask about the difference between tests.

12. For Test1, calculate the missing F statistic.

4.120 (OK).

$$F = \left(\frac{1.18}{2.39}\right)^2 = 0.243$$

F Test Statistic: 0.243

13. For Test3, calculate the missing F statistic. **Hint:** This value can be easily calculated using provided output.

$$F = t^2 = (-2.1175)^2 = 4.48$$

$$F = \frac{\text{MS Machine}}{\text{MS Resid}} = \frac{15.986}{3.566} = 4.48$$

F Test Statistic: 4.48

14. Comparing Test1 versus Test3 (both F tests), what is the difference between these two tests?

Test 1 $H_0: \sigma_1^2 = \sigma_2^2$ (test of variances)

Test 3 $H_0: \mu_1 = \mu_2$ (test of means)



24.

**Questions 12 through 17 (Machines) continued:**

15. Comparing Test1 versus Test2, what is the difference between these two tests?

Both Test1 and Test2 are tests of variances.

Test 1 (F-test) requires assumption of normality

Test2 (levene) does NOT.

16. Comparing Test4 versus Test5 (both two-sample t-tests) which is preferred for this data? Justify your response based on result of another formal test from the output.

Test 5 (allowing unequal variances) is preferred.

Small p-values for both F-test and levene's test suggest UNEQUAL variances.

17. Considering the problem description from above, which machine is preferred? Justify your response based on mean and standard deviation.

Machine 1 is preferred because.

Mean (9.94) is closer to stated size (10 units)

Standard deviation (1.18) is smaller than Machine 2.

**Questions 18 through 22 (Free Speech Survey #1):** A recent Denver Post article was titled "College students hostile to free speech" (9/25/17). The article reported on a survey of **n = 1,500** undergraduate students. We will focus on just one question which asked whether "presentation of counterpoints to offensive views is legally required in on-campus events". **930 students agreed with the statement.** Let π be the true population proportion that agree with the statement.

18. The full article reports the proportion of students who agree with several different survey questions. Give the "worst case" 95% Margin of Error.

$$95\%\ ME = 1/\sqrt{n} = 1/\sqrt{1500} = 0.026$$

95%ME: 0.26

19. Estimate the proportion (π) of undergraduate students that agree with the statement.

$$\hat{\pi} = 930/1500 = 0.62$$

Estimate: 0.62

20. Suppose we want to test H0: π≤0.5 versus HA: π>0.5. Give the Z test statistic.

$$Z = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0.62 - 0.50}{\sqrt{\frac{0.5(1-0.5)}{1500}}} = 9.29$$

Z Test Statistic: 9.29

21. Give the p-value corresponding to the test from the previous question.

$$\text{p-value} = P(Z > 9.29)$$
$$= 1 - P(Z \leq 9.29)$$

"off the charts" p-value < 0.001.

p-value: < 0.001



28.

**Questions 22 through 25 (Motorcycles):** A recent Denver Post article was titled "Older, male motorcyclists most likely to die in crashes" (10/26/17).

22. In the article it is stated that "91% of those killed in motorcycle crashes in 2015 were male." Does this mean that a male motorcyclist has a higher chance (probability) of death than a female motorcyclist? Discuss. Hint: What information would you want to put this statistic into perspective?

No. This only tells us that given someone died, they were more likely to be Male.
But this may be explained if majority of motorcyclists were Male (which is probably true).

Need # or prop. female riders

**Questions 23 through 25 (Motorcycles continued):** In this group of questions, we will look at the distribution of motorcycle deaths by age group. Suppose that in 2015 there were **n = 125 motorcycle deaths** in Colorado. Of these, 38 were riders aged 18-34, 30 were riders aged 35-44 and 57 were riders aged 45 or older. We compared these values to the **proportion of registered motorcyclists** in each age group. The data is summarized in the table below.

| | Ages 18 - 34 | Ages 35 - 44 | Ages 45 + | Total |
|---|---|---|---|---|
| # Deaths | 38 | 30 | 57 | 125 |
| Proportion Registered | 0.33 | 0.26 | 0.41 | 1.00 |

```
> GOFtest <- chisq.test(c(38, 30, 57),p = c(0.33, 0.26, 0.41))
> GOFtest
      Chi-squared test for given probabilities
data:  c(38, 30, 57)
X-squared = 1.0935, df = 2, p-value = 0.5788

> GOFtest$observed
[1] 38 30 57
> GOFtest$expected
[1] 41.25 32.50 51.25
```

23. Thinking about the headline "Older…motorcyclists most likely to die in crashes". This statement is true in the sense that the 45+ age group had the largest number of deaths (accounting for 57 out of the total 125 deaths). But use the Chi-square GOF test results to discuss whether this is unusual (or surprising) based on the age distribution of registered motorcycle riders.

Does not seem surprising based on rider registration
$\chi^2$ GOF test* $\boxed{\text{p-value} = 0.578} > \alpha = 0.05$
So we do not have evidence that distb'n of deaths is different from distb'n of registration

24. Which age group has the largest difference between observed and expected values? Were there more or less deaths than expected (if the null hypothesis was true). Note: the observed and expected values are given in the R output.

Obs - Exp
45+ Group = 57 − 51.25 = +5.75
More deaths than expected for this Age Group

One more question on the next page….



12.

25. Suppose we had run the chi-square test using the number of motorcycle <u>injuries</u> instead of motorcycle <u>deaths</u>. There were $n = 2824$ motorcycle injuries (as compared to $n = 125$ motorcycle deaths) in Colorado in 2015. Do you think that the conclusion (statistical significance) of the chi-square GOF test might change? I am looking for a discussion here (not a yes or no answer). Hint: Think about the role that sample size plays in statistical significance.

With increased sample size we increase power, so we may be able to detect a difference.



4.

## Obs Study (Questions 5 through 11)

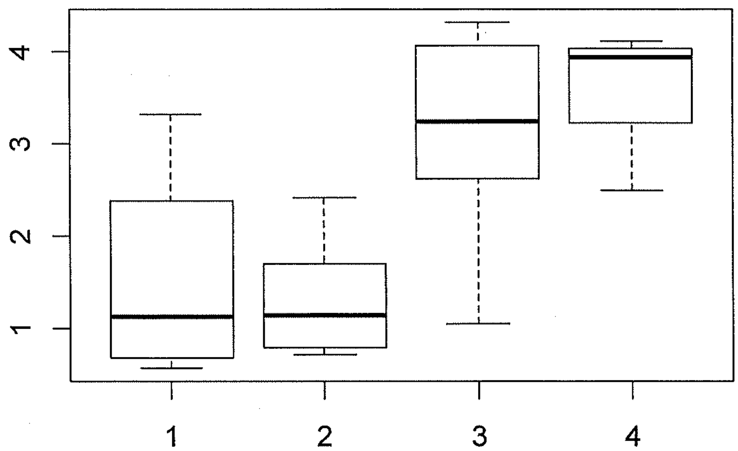
```
library(plyr)
library(car)
library(lsmeans)
str(InData)
SumStats <- ddply(InData, c("Group"), summarize,
                  n = length(Y),
                  mean = mean(Y),
                  sd = sd(Y),
                  SE = sd/sqrt(n))
SumStats
boxplot(Y ~ Group, data = InData)
Model1 <- lm(Y ~ Group, data = InData)
anova(Model1)
[redacted]
Model2 <- lm(Y ~ Group, data = InData)
anova(Model2)
par(mfrow=c(1, 2))
plot(Model2)
leveneTest(Y ~ Group, data = InData)
shapiro.test(residuals(Model2))
#LSMeans1
lsmeans(Model2, dunnett ~ Group)
#LSMeans2
lsmeans(Model2, pairwise ~ Group, adjust = "none")
```

---

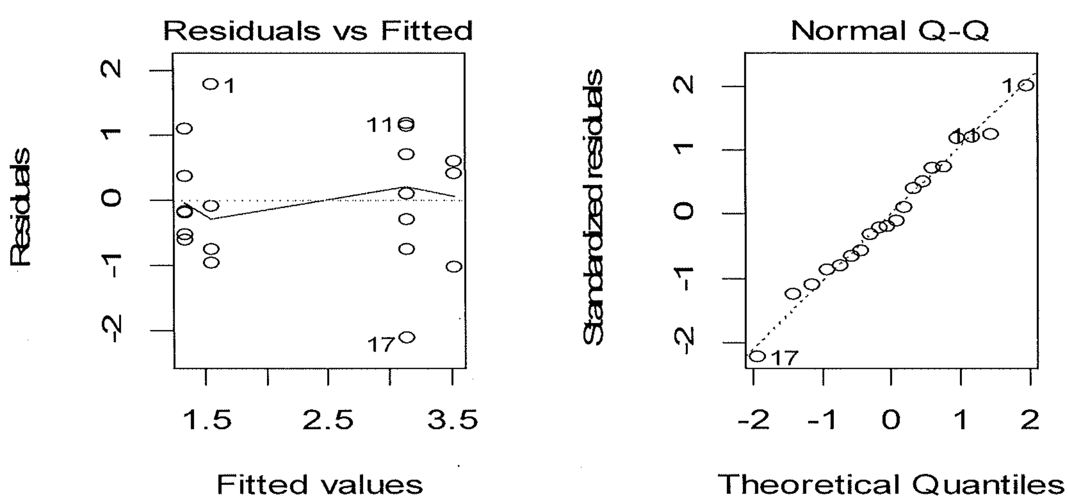
```
> library(plyr)
> library(car)
> library(lsmeans)
> library(multcompView)
> str(InData)
'data.frame':     20 obs. of  2 variables:
 $ Group: num  1 1 1 1 2 2 2 2 2 2 ...
 $ Y    : num  3.323 0.794 1.461 0.583 0.8 ...
> SumStats <- ddply(InData, c("Group"), summarize,
+                   n = length(Y),
+                   mean = mean(Y),
+                   sd = sd(Y),
+                   SE = sd/sqrt(n))
> SumStats
  Group n     mean        sd        SE
1     1 4 1.540062 1.2459275 0.6229638
2     2 6 1.324823 0.6406817 0.2615572
3     3 7 3.138321 1.1696318 0.4420793
4     4 3 3.524074 0.8875883 0.5124494
```

# Obs Study continued (Questions 5 through 11)

```
> boxplot(Y ~ Group, data = InData)
```




```
> Model1 <- lm(Y ~ Group, data = InData)
> anova(Model1)
Analysis of Variance Table
Response: Y
          Df Sum Sq Mean Sq F value   Pr(>F)
Group      1 13.442 13.4419  11.825 0.002929 **
Residuals 18 20.461  1.1367

> [redacted]
> Model2 <- lm(Y ~ Group, data = InData)
> anova(Model2)
Analysis of Variance Table
Response: Y
          Df Sum Sq Mean Sq F value   Pr(>F)
Group     [redacted] 17.409  5.8031  5.6296 0.007884 **
Residuals [redacted] 16.493  1.0308

> plot(Model2)
```



## Obs Study continued (Questions 5 through 11)

```
> leveneTest(Y ~ Group, data = InData)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  0.5624 0.6475
      16
> shapiro.test(residuals(Model2))
      Shapiro-Wilk normality test
data:  residuals(Model2)
W = 0.98168, p-value = 0.9539
```

```
> #LSMeans1
> lsmeans(Model2, dunnett ~ Group)
$lsmeans
 Group   lsmean        SE df  lower.CL upper.CL
 1     1.540062 0.5076482 16 0.4638963 2.616228
 2     1.324823 0.4144930 16 0.4461374 2.203509
 3     3.138321 0.3837460 16 2.3248163 3.951826
 4     3.524074 0.5861816 16 2.2814242 4.766723

Confidence level used: 0.95

$contrasts
 contrast  estimate        SE df t.ratio p.value
 2 - 1    -0.215239 0.6553710 16  -0.328  0.9563
 3 - 1     1.598259 0.6363707 16   2.512  0.0606
 4 - 1     1.984011 0.7754454 16   2.559  0.0554

P value adjustment: dunnettx method for 3 tests
```

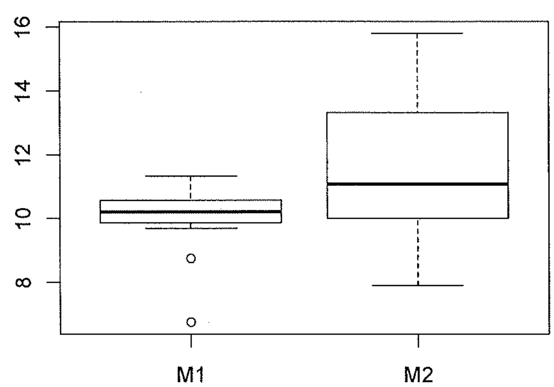
```
> #LSMeans2
> lsmeans(Model2, pairwise ~ Group, adjust = "none")
$lsmeans
 Group   lsmean        SE df  lower.CL upper.CL
 1     1.540062 0.5076482 16 0.4638963 2.616228
 2     1.324823 0.4144930 16 0.4461374 2.203509
 3     3.138321 0.3837460 16 2.3248163 3.951826
 4     3.524074 0.5861816 16 2.2814242 4.766723

Confidence level used: 0.95

$contrasts
 contrast   estimate        SE df t.ratio p.value
 1 - 2     0.2152390 0.6553710 16   0.328  0.7469
 1 - 3    -1.5982590 0.6363707 16  -2.512  0.0231
 1 - 4    -1.9840114 0.7754454 16  -2.559  0.0210
 2 - 3    -1.8134980 0.5648588 16  -3.211  0.0055
 2 - 4    -2.1992504 0.7179230 16  -3.063  0.0074
 3 - 4    -0.3857524 0.7006211 16  -0.551  0.5895
```

# Machines (Questions 12 through 17)

```
library(plyr)
library(car)
str(MachineData)
boxplot(Y ~ Machine, data = MachineData)
SumStats <- ddply(MachineData, c("Machine"), summarize,
                  n = length(Y),
                  mean = mean(Y),
                  sd = sd(Y),
                  SE = sd/sqrt(n))
SumStats
#Test1
var.test(Y~Machine, data = MachineData)
#Test2
leveneTest(Y ~ Machine, data = MachineData)
#Test3
Model <- lm(Y ~ Machine, data = MachineData)
anova(Model)
#Test4
t.test(Y ~ Machine, var.equal = TRUE, data = MachineData)
#Test5
t.test(Y ~ Machine, data = MachineData)
```

---

```
> str(MachineData)
'data.frame':      24 obs. of  2 variables:
 $ Machine: Factor w/ 2 levels "M1","M2": 1 1 1 1 1 1 1 1 1 1 ...
 $ Y      : num  10.04 8.76 10.71 9.71 10.05 ...
> boxplot(Y ~ Machine, data = MachineData)
```




```
> SumStats <- ddply(MachineData, c("Machine"), summarize,
+                   n = length(Y),
+                   mean = mean(Y),
+                   sd = sd(Y),
+                   SE = sd/sqrt(n))
> SumStats
  Machine  n      mean       sd        SE
1      M1 12  9.940791 1.180941 0.3409083
2      M2 12 11.573087 2.395075 0.6913986
```

# Machines continued (Questions 12 through 17)

```
> #Test1
> var.test(Y~Machine, data = MachineData)
      F test to compare two variances
data:  Y by Machine
F = [redacted], num df = 11, denom df = 11, p-value = 0.02721
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.06998839 0.84452087
sample estimates:
ratio of variances
         0.2431186
> #Test2
> leveneTest(Y ~ Machine, data = MachineData)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  1  4.4643 0.04619 *
      22
> #Test3
> Model <- lm(Y ~ Machine, data = MachineData)
> anova(Model)
Analysis of Variance Table
Response: Y
          Df Sum Sq Mean Sq F value   Pr(>F)
Machine    1 15.986 15.9863 [redacted] 0.04576 *
Residuals 22 78.441  3.5655
> #Test4
> t.test(Y ~ Machine, var.equal = TRUE, data = MachineData)
      Two Sample t-test
data:  Y by Machine
t = -2.1175, df = 22, p-value = 0.04576
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.23099596 -0.03359635
sample estimates:
mean in group M1 mean in group M2
        9.940791        11.573087
> #Test5
> t.test(Y ~ Machine, data = MachineData)
      Welch Two Sample t-test
data:  Y by Machine
t = -2.1175, df = 16.05, p-value = 0.05018
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.266066528  0.001474215
sample estimates:
mean in group M1 mean in group M2
        9.940791        11.573087
```

**Exam2 Extra Results**

**Note:** This information was NOT provided during the original exam!

**#6 ANOVA Table**

```
> anova(Model2)
Analysis of Variance Table
Response: Y
          Df Sum Sq Mean Sq F value   Pr(>F)
Group      3 17.409  5.8031  5.6296 0.007884 **
Residuals 16 16.493  1.0308
```

**#11 Tukey Adjustment (now only 2 significant differences)**

```
> lsmeans(Model2, pairwise ~ Group)
$contrasts
 contrast   estimate        SE df t.ratio p.value
 1 - 2     0.2152390 0.6553710 16   0.328  0.9873
 1 - 3    -1.5982590 0.6363707 16  -2.512  0.0960
 1 - 4    -1.9840114 0.7754454 16  -2.559  0.0882
 2 - 3    -1.8134980 0.5648588 16  -3.211  0.0253
 2 - 4    -2.1992504 0.7179230 16  -3.063  0.0338
 3 - 4    -0.3857524 0.7006211 16  -0.551  0.9450

P value adjustment: tukey method for comparing a family of 4 est
imates
```

**#12 Test1**

```
> var.test(Y~Machine, data = MachineData)
      F test to compare two variances
data:  Y by Machine
F = 0.24312, num df = 11, denom df = 11, p-value = 0.02721
```

**#13 Test3**

```
> Model <- lm(Y ~ Machine, data = MachineData)
> anova(Model)
Analysis of Variance Table
Response: Y
          Df Sum Sq Mean Sq F value  Pr(>F)
Machine    1 15.986 15.9863  4.4836 0.04576 *
Residuals 22 78.441  3.5655
```

**#18-22**

```
> prop.test(930, 1500, alternative = "greater", correct = FALSE)
      1-sample proportions test without continuity correction
data:  930 out of 1500, null probability 0.5
X-squared = 86.4, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is greater than 0.5
95 percent confidence interval:
 0.599187 1.000000
sample estimates:
   p
0.62
```

**#25** (multiply original counts by 20 for n = 2500)

```
> GOFtest <- chisq.test(c(760, 600, 1140),p = c(0.33, 0.26, 0.41
))
> GOFtest

      Chi-squared test for given probabilities

data:  c(760, 600, 1140)
X-squared = 21.87, df = 2, p-value = 1.783e-05
```