

# Stat 302, Assignment 2, Due Friday February 19, 2016 at 4:30pm

There are \_\_\_\_\_ multi-part questions and a set of questions based on 5 pages of reading. Please see the Stats Workshop for help, or see me in office hours (Tue 1-2, Thur 3:30-4:30), or e-mail me at [jackd@sfu.ca](mailto:jackd@sfu.ca)

Explanations and relevant computer output should be included, but plots do not need to be. Necessary R code is included for every question, after the reading questions. There are also four practice problems, mostly surrounding review material.

Total / 64  
Q1 / 7  
Q2 / 14  
Q3 / 10  
Q4 / 16  
Reading /17

Name \_\_\_\_\_

Student Number \_\_\_\_\_

1) Consider the dataset A2\_npk.csv , which is a dataset of yield across several different factors for some industrial process.

1a, 2pts) Check this data against unequal variance and imbalance. Are there any problems? How do you know?

*Using the by() function in R, we can find the standard deviation and number of observations of each group.*

Block	Mean	SD	N
A	54.0	7.27	4
B	57.5	2.04	4
C	60.8	6.79	4
D	50.1	8.15	4
E	50.5	1.49	4
F	56.4	2.43	4

*This data is perfectly balanced across block.*

*However, the standard deviation of block D is more than 3 times that of block E, which violates the equal variance assumption.*

*Using a Bartlett test is also an appropriate way to solve this problem.*

1b, 1pt) Using the p-value, use ANOVA test the hypothesis of no difference between the four groups.

*Using a default of alpha = 0.05.*

*ANOVA gives a p-value of 0.086, so there is weak if any evidence of a difference in means. (However,*

*the unequal variance from 1a could be affecting that results.)*

1c, 2pts) Using the obtained F score, test whether there is any difference between the **six** groups. Use the  $\alpha = 0.05$  significance level.

*From the ANOVA, the obtained F is 2.3184*

*From the  $qf()$  function, the critical F-value is 2.772*

*Since the obtained F does not exceed the critical F, we do not have enough evidence against the null to reject it.*

1d, 2pts) How much of the variance in **yield** is explained by the different levels of **block**? Show your work.

*The sum of squares total is  $343 + 533 = 876$ . This is the total variance in the data.*

*Of that,  $343/876 = 0.3915$ , or 39.15% is explained by block.*

---

---

2) Consider the dataset A2\_Toothgrowth.csv, which measures the tooth growth of Guinea pigs when given two different kinds of supplement (Vitamin C and Orange Juice) at three different dosage levels (0.5, 1.0, and 2.0 mg)

2a, 2pts) Conduct a one-way ANOVA using **supplement** as a grouping variable and **growth** as a response variable. Does the response vary between different groups? Use the  $\alpha = 0.05$  significance level.

*This ANOVA gives a p-value of 0.06039. There is not enough evidence that the means are different between supplements.*

2b, 2pts) Conduct a one-way ANOVA using **dosage** as a grouping variable and **growth** as a response variable. Does the response vary between different groups? Use the  $\alpha = 0.05$  significance level.

*This ANOVA gives a p-value of  $9.533 \times 10^{-16}$ . There is very strong evidence that the means are different between dosage levels.*

2c, 3pts) Conduct a two-way ANOVA using the same response both grouping variables from 2a and 2b. Using the p-values, test whether the response changes across different types of **supplement** and of **dosage** changes, holding each other constant. Use the  $\alpha = 0.05$  significance level.

*This ANOVA gives a p-value of 0.00042 for supplement, and  $2.2 \times 10^{-16}$  for dosage. There is strong evidence that the means are different between supplements (holding dosage constant), and between dosage (holding supplement constant).*

2d, 3pts) Using the obtained F scores, conduct the same two tests as you did in 2c.

*The ANOVA gives an F-statistic of 14.02 for supplement and 82.81 for dosage.*

*The critical value for supplement is 4.01 because we use df 1 and 56.*

*The critical value for dosage is 3.16 because we use df 2 and 56.*

*Since both obtained values were larger than the critical values, we reject both null hypotheses. Both supplement and dosage have an effect on tooth growth.*

*(Not needed) Notice that the second df is the same for both tests – the residual df is always used.*

2e, 2pts) Explain the apparent paradox between the results of 2a and 2c. Why does **supplement** appear to be a non-significant (or marginally significant) factor in one answer, and strongly significant in the other?

*In 2a, we are just checking if growth changes by supplement without considering the possible effects from other factors. However, the dosage also matters, and in 2c we check if growth changes while holding dosage constant.*

*(Not needed) In other words, dosage explains enough of the remaining variation in growth that we can see that the differences in supplement aren't simply the result of unexplained variation.*

2f, 2pts) How much of the variance in **growth** is explained by **supplement** and by **dosage**? Show your

work.

The sum of squares total is  $205.35 + 2426.43 + 820.43 = 3452.21$

Of this,  $205.35 / 3452.21$  , or 5.95% of the total variation is explained by supplement type,

And  $2426.43 / 3452.21$  , or 70.29% of the total variation is explained by dosage.

(Not needed) In total, 76.24% of the variation is explained

---

---

3) Consider the dataset A2\_chickwgt.csv, which measures the weight of chickens that were given six different kinds of feed supplement.

3a, 1 pt) Conduct a one-way ANOVA using **feed** as a grouping variable and **weight** as a response variable. Does the response vary between different groups? Use the  $\alpha = 0.05$  significance level.

From the ANOVA, the  $p$ -value is  $5.94 \times 10^{-10}$ , so there is more than enough evidence to say weight changes with the feed type

3b, 2 pts) If we were to compare each pair of means individually using a t-test and a Bonferroni correction with a family/experiment-wise  $\alpha$  of 0.05, what is the alpha used for each pair to determine significance?

There are 6 groups.

That makes a total of  $5 + 4 + 3 + 2 + 1 = 15$  pairs to compare.

The alpha for each pair would be  $0.05 / 15 = 0.00333$

3c, 4 pts) Using Tukey's Honestly Significant Difference analysis, which pairs of means show a significant difference? Use a family/experiment-wise  $\alpha$  of 0.05.

Pair	P-value, adjusted
soybean-horsebean	0.0042167
meatmeal-horsebean	0.0001062
casein-horsebean	0.0000000
sunflower-horsebean	0.0000000
casein-linseed	0.0002100
sunflower-linseed	0.0000884
casein-soybean	0.0083653
sunflower-soybean	0.0038845

3d, 3 pts) Using the results from 3c, arrange the groups into clusters. Label the group means that aren't honestly different from the highest

We start with the highest mean group, and compare everything with a lower mean to that first. Sunflower is significantly different from Soybean, Linseed, and Horsebean. Therefore, it must share a cluster with Casein and Meatmeal.

Group	Mean	Cluster
Sunflower	328.9	A
Casein	323.6	A
Meatmeal	276.9	A
Soybean	246.4	
Linseed	218.8	
Horsebean	160.2	

Moving onto the second mean, Casein is significantly different from Soybean, Linseed, and Horsebean as well. Therefore, it shares a cluster with Meatmeal. This is already covered by cluster A.

The next one, Meatmeal, is only different from Horsebean. It shares a cluster with Soybean and Linseed.

Group	Mean	Cluster
Sunflower	328.9	A
Casein	323.6	A
Meatmeal	276.9	AB
Soybean	246.4	B
Linseed	218.8	B
Horsebean	160.2	

Soybean and Horsebean are significantly different. That leaves Soybean and Linseed in the same cluster. However, this is already covered by cluster B.

Finally, Linseed and Horsebean are not significantly different, so they share a cluster.

Group	Mean	Cluster
Sunflower	328.9	A
Casein	323.6	A
Meatmeal	276.9	AB
Soybean	246.4	B
Linseed	218.8	BC
Horsebean	160.2	C

Here is an alternate solution if you apply the top-down compare method without removing redundant clusters.

<i>Group</i>	<i>Mean</i>	<i>Cluster</i>
<i>Sunflower</i>	<i>328.9</i>	<i>A</i>
<i>Casein</i>	<i>323.6</i>	<i>AB</i>
<i>Meatmeal</i>	<i>276.9</i>	<i>ABC</i>
<i>Soybean</i>	<i>246.4</i>	<i>CD</i>
<i>Linseed</i>	<i>218.8</i>	<i>CDE</i>
<i>Horsebean</i>	<i>160.2</i>	<i>E</i>

---

---

4) Consider the dataset A4\_viral.csv, which is loosely based on real data of HIV viral load in blood in response to three different treatments and a control.

4a 2 pts) Use **load** as a response and **trt** as a grouping variable. Check the ANOVA against unequal variance and imbalance. Are there any problems? How do you know?

*Using the by() function in R, we can find the standard deviation and number of observations of each group.*

Block	Mean	SD	N
ART1	6318	5995	10
ART2	30490	52210	10
ART3	16750	22150	10
Control	121900	232600	10

*The groups are balanced, but there is massive inequality in the variances. Any ANOVA results can't be trusted.*

4b 1 pt) From a boxplot, would you say there is a difference in viral load between the groups? Briefly explain.

*The boxplot of the control group has a much higher viral load than the other groups. I would expect any reasonable test to find a difference in the means.*

4c 2 pts) Ignore any problems and conduct a one-way ANOVA on **load** as a response to **trt**. Using the p-value, test the hypothesis of no difference between the groups. Does this match your expectations from the boxplot?

*The p-value of the ANOVA is 0.138, meaning there is no evidence of a difference between the four means. This runs contrary to what the boxplot showed.*

4d 3 pts) Using Tukey's Honestly Significant Difference analysis, which pairs of means, if any, show a significant difference? Use a family/experiment-wise  $\alpha$  of 0.05, and list p-values of significant differences. Given your answer in 4c, is this surprising?

*None of the pairs showed a significant difference.*

*Since the ANOVA failed to detect any differences between the groups, looking at individual pairs should also fail to find any significant difference.*

*(not needed) If the ANOVA fails to find a difference, there isn't really a point to doing a Tukey HSD Analysis.*

4e 2 pts) Now use **logload** (the log-transform of **load**) as a response and **trt** as a grouping variable. Check this new ANOVA against unequal variance and imbalance. Comment briefly.

*ed.*

Using the `by()` function in R again, but on `logload`, we find:

Block	Mean	SD	N
ART1	8.36	1.003	10
ART2	9.63	1.089	10
ART3	8.97	1.341	10
Control	10.14	1.909	10

The groups are still balanced, and no standard deviation is three times as large as any other. The unequal variance problem has been resolved.

4f 1 pt) From a boxplot, would you say there is a difference in **log(viral load)** between the groups? Briefly explain.

*There are some differences between the boxes, but it is unclear now if these are large enough to say the means are different.*

4g 2 pts) Repeat the one-way ANOVA for this new, log-transformed response. Test the hypothesis of equal means again. Does this match your expectations from the new boxplot?

*The ANOVA table gives a p-value of 0.036, which shows some evidence of a difference between the means, but no overwhelming evidence. Our uncertainty in 4f was founded.*

4h 3 pts) Repeat the Tukey's HSD with the log-transformed response. Use a family/experiment-wise  $\alpha$  of 0.05.

*The Control group and the ART1 group show an honestly significant difference ( $p_{adj} = 0.031$ ). Nothing else does.*

---

Reading questions)

These questions pertain to "The Insignificance of Significance Testing", by Neville Nicholls.  
The Annals of Applied Statistics  
May 2000, Vol. 81, No. 5, 981-985

The answer to R1 appears first in the text, and so on. Every question can be answered in 25 words or fewer.

R1, 2 pts) Give an example of a data set that could have physical significance, but not statistical significance.

*Two variables with a sample correlation of 0.60, but only 10 data points.*

R2, 2 pts) Cohen mentions some probability when talking about an  $n=50$  sample from a population with population correlation of  $\rho = 0.30$ . What is the name for this probability? (The name isn't in the paper, it IS in our notes on hypothesis testing)

*Cohen is talking about the chance of correctly rejecting the null hypothesis when it is false. In other words, the probability of detecting an effect when there really is one. This is known as the POWER.*

*(A single word answer of 'power' is also acceptable).*

R3, 2 pts) The 1979-95 data used to calculate climate trends has no uncertainty from sampling, How is this possible?

*Sampling uncertainty comes from sampling a small part of a population, and applying your findings to the entire population. The 1979-95 data is from all over the world; essentially the entire population has been measured already.*

*Thus, there is no uncertainty introduced from making the leap from sample to population.*

R4, 3 pts) What deviations from a well-behaved distribution could affect the correlation between SOI and snowfall?

*From the paper "Serial correlation,... , skewness, and bimodality"*

*Other possibilities are non-normality, and non-independence.*

R5, 3 pts) What are three alternatives to null hypothesis testing?

*Confidence intervals, permutation tests, and cross validation (and other 'a posteriori' testing)*

R6, 2 pts) What is another term for the "repeated investigations" issue? That is, when a so-called insignificant result is left unpublished until someone repeats it and finds significance simply by chance? (Again, the term is in our notes, think Tukey)

*Multiple testing.*

*(Not needed) "Repeated investigations" essentially means doing a test over and over with new data until it shows up as significant. In other terms, testing multiple times, but acting like you only tested once.*

R7, 3 pts) A permutation test produces a value that works very similarly to a p-value. However, a permutation test has one major advantage over a classic hypothesis test (and confidence intervals). What is it? (Hint: Spearman correlation and the median have this same advantage).

*Robustness! Permutation tests better handle "the issue of the nonrandomness of samples in the atmospheric sciences".*

*(Not needed) The Spearman correlation and median are used when the usual assumptions of correlation and normality, respectively, don't apply. A permutation test is what you use when the assumptions of a traditional hypothesis test don't apply.*

---

```
##### EXAMPLE CODE QUESTION 1
```

```
## Get the dataset npk, which is embedded into R  
Q1 = npk
```

```
## One-way ANOVA on block level  
mod = lm(yield ~ block, data=Q1)  
anova(mod)
```

```
### Get the standard deviation and N from each treatment group  
by(Q1$yield, Q1$block, sd)  
by(Q1$yield, Q1$block, length)  
boxplot(yield ~ block, data=Q1)
```

```
## Get the critical value of f.
```

```
##### EXAMPLE CODE QUESTION 2
```

```
## Get the dataset ToothGrowth, which is embedded into R  
Q2 = ToothGrowth  
Q2$dose = as.factor(Q2$dose) # Treat the dosage like a categorical  
variable
```

```
## One-way ANOVA on supplement type  
mod1 = lm(len ~ supp, data=Q2)  
anova(mod1)
```

```
## One-way ANOVA on dosage level  
mod2 = lm(len ~ dose, data=Q2)  
anova(mod2)
```

```
## Two-way ANOVA on both dosage AND supplement  
mod12 = lm(len ~ supp + dose, data=Q2)  
anova(mod12)
```

```
##### EXAMPLE CODE QUESTION 3
```

```
## Get the dataset chickwts, which is embedded into R  
Q3 = chickwts
```

```
## One-way ANOVA on feed type  
mod = lm(weight ~ feed, data=Q3)  
anova(mod)
```

```
## Tukey's HSD Analysis  
mod = aov(weight ~ feed, data=Q3)  
TukeyHSD(mod, ordered=TRUE)
```

```
##### EXAMPLE CODE QUESTION 4
```

```
### Get the dataset viral, which is made up for this dataset  
Q4 = read.csv("A2_viral.csv")
```

```
### Get the mean, standard deviation, and N from each treatment group  
by(Q4$load, Q4$trt, mean)  
by(Q4$load, Q4$trt, sd)  
by(Q4$load, Q4$trt, length)
```

```
### Produce a boxplot of load by treatment group  
boxplot(load ~ trt, data=Q4)
```

```
### One-way ANOVA of load by treatment group  
mod = lm(load ~ trt, data=Q4)  
anova(mod)
```

```
## Tukey's HSD Analysis  
mod = aov(load ~ trt, data=Q4)  
TukeyHSD(mod, ordered = TRUE)
```

```
### Now repeat the same things with logload instead of load  
by(Q4$logload, Q4$trt, sd)  
by(Q4$logload, Q4$trt, mean)
```

```
boxplot(logload ~ trt, data=Q4)
```

```
mod = lm(logload ~ trt, data=Q4)  
anova(mod)
```

```
mod = aov(logload ~ trt, data=Q4)  
TukeyHSD(mod, ordered = TRUE)
```