

Question 1: One-Way Within-Subjects ANOVA (55 points)

Question 2: Two-Way Mixed ANOVA (45 points)

# 36-309 / 36-749 Homework 8: Within-Subject Designs

Code ▼

Due Wednesday, November 16, 11:59pm

Sanaz Saadatifar

## Question 1: One-Way Within-Subjects ANOVA (55 points)

This problem is loosely based on “Choking under pressure and working memory capacity (WMC): When performance pressure reduces fluid intelligence” (<https://link.springer.com/article/10.3758/BF03213916>) by David Gimmig, et al. (2006).

The participants were 67 undergraduates, and the purpose of the study was to understand students’ performance on academic tasks of varying “pressure.” In the experiment, each subject was given two “easy” tasks and two “hard” tasks in random order. The accuracy of each task was measured on a 0-to-100 scale, which acts as the outcome for this study.

Here’s the dataset for this problem:

Hide

```
choke = read.csv("https://raw.githubusercontent.com/zjbranson/stat309fall2022/main/choke.csv")
```

In this dataset, there are 67 subjects with one row for each subject. Thus, the dataset is in a *wide* format. The four columns in this dataset are `easy1` , `easy2` , `hard1` , and `hard2` , corresponding to the two easy tasks and two hard tasks.

- a. (10pts) For this part, answer the following three questions.
  - (3pts) In the setup of this problem, it was stated that “each subject was given two easy and two hard tasks in random order.” What is the name of this experimental design technique (where we randomize the order of the treatments among subjects)? Furthermore, what is the purpose of doing this experimental design technique? Explain in 1-3 sentences.

**[It is called “counter-balanced within-subjects design.” the purpose of doing this experimental design technique is Randomizing the order that subjects receive treatments, so that the pitfalls of standard repeated experiment designs is avoided and confounding variables are removed from the experiment. ]**

- (3pts) First, we will use a version of response simplification to analyze these data. **In particular, we'll use a paired t-test to compare (1) the average of the two easy task scores to (2) the average of the two hard task scores, to see if they are different from each other.** Using the symbols  $\mu_{E1}$ ,  $\mu_{E2}$ ,  $\mu_{H1}$ ,  $\mu_{H2}$  for the population means of the four task scores, write the null hypothesis for this comparison. When writing your null hypothesis, please write it such that 0 is on the right-hand side of the equal sign.

**[H0:  $(1/2)\mu_{E1} + (1/2)\mu_{E2} - (1/2)\mu_{H1} - (1/2)\mu_{H2} = 0$ ]**

- (4pts) To run the paired t-test, we need to create two new columns in the `choke` dataset:
  1. `easyMean` : The average of the two easy task scores (i.e., the average of `easy1` and `easy2` )
  2. `hardMean` : The average of the two hard task scores (i.e., the average of `hard1` and `hard2` )

In other words, *for each subject in the dataset*, we want a column that tells us the average of their two easy task scores and a column that tells us the average of their two hard task scores.

Here you'll write a line of code to create each of these columns:

Hide

```
choke$easyMean = (choke$easy1+choke$easy2)/2
choke$hardMean = (choke$hard1+choke$hard2)/2
```

For this part, you just need to fill in the ?s (and uncomment the two lines of code after you've written your code). **Hint:** You *should not* use the `mean()` function. If you're having trouble with this part, look back at your answer for the null hypothesis question in the previous bullet point - it should suggest how to write your code here.

After you've done this, you should see below if `easyMean` and `hardMean` are properly defined in the dataset (you should see that there are two new columns, `easyMean` and `hardMean` , if you did the above correctly):

Hide

```
head(choke)
```

```
##   easy1 easy2 hard1 hard2 easyMean hardMean
## 1    58    60    62    64     59.0     63.0
## 2    75    74    60    57     74.5     58.5
## 3    57    73    67    63     65.0     65.0
## 4    42    59    50    50     50.5     50.0
## 5    68    70    73    76     69.0     74.5
## 6    64    55    67    54     59.5     60.5
```

- (10pts) Now we are going to run a paired t-test, which is the appropriate analysis for an experiment where each subject has *two measurements*. Specifically, we'll test if `easyMean` is significantly different from `hardMean` . Answer the following three questions.
  - (3pts) First, appropriately run the paired t-test between `easyMean` and `hardMean` .

Hide

```
t.test(x = choke$easyMean, y = choke$hardMean, paired = TRUE)
```

```
##
## Paired t-test
##
## data: choke$easyMean and choke$hardMean
## t = 5.6528, df = 66, p-value = 3.653e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.697379 7.735456
## sample estimates:
## mean of the differences
##                5.716418
```

- (4pts) What is the point estimate for the mean difference between `easyMean` and `hardMean` (i.e., `easyMean` minus `hardMean`)? Furthermore, what is the 95% confidence interval for this mean difference?

**[The point estimate or the mean of the differences is 5.716418 which shows the [mean `easyMean` – mean `hardMean`]. Moreover, the 95% confidence interval for this mean difference (3.697379 , 7.735456)]**

- (3pts) What is your scientific conclusion according to this test? Explain your reasoning in 1-3 sentences.

**[Since the p-value is less than 0.05 (3.653e-07), then we reject the null hypothesis of means of `hardmean` and `easy mean` being equal. ]**

- (10pts) Now we'll consider what would happen if we ran the *incorrect* independent-samples t-test for these data, where we pretend that the 67 `easyMean` measurements and 67 `hardMean` measurements are from different sets of people. Answer the following three questions.
- (3pts) First, implement the *incorrect* independent-samples t-test mentioned above. When implementing the t-test, set `var.equal = TRUE`. For this part, you just need to display the output from the test.

Hide

```
t.test(x = choke$easyMean, y = choke$hardMean, var.equal = TRUE)
```

```
##
## Two Sample t-test
##
## data: choke$easyMean and choke$hardMean
## t = 4.7269, df = 132, p-value = 5.765e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.324211 8.108625
## sample estimates:
## mean of x mean of y
##  71.55224  65.83582
```

- (3pts) Now look at the 95% confidence interval for this independent samples t-test, and compare it to the 95% confidence interval you got from the paired t-test in Part B. In particular, which confidence

interval is *wider* - the independent-samples t-test one (the one you got here) or the paired t-test one (the one from Part B)?

**[Independent sample t-test's CI (3.324211 , 8.108625) is wider than paired t-test's CI (3.697379 , 7.735456). ]**

- (4pts) Now explain in 1-2 sentences why one confidence interval is wider than another (i.e., why the confidence interval in this part is of a different width than the confidence interval in Part B).

**[CI is point estimate +/- a number that is relevant to sigma squares. Since paired t-test gets rid of subject-to-subject variability, it actually reduces the variance or sigma squared. Since sigma squared is decreased, then the CI's range would be more closer to the point estimate compared to the time when sigma squared is fixed]**

- d. (10pts) Now we will use repeated measures ANOVA to analyze these data, instead of aggregating the data into an *easyMean* score and a *hardMean* score. To aid you in this task, here is the *tall* version of this dataset:

Hide

```
chokeTall = read.csv("https://raw.githubusercontent.com/zjbranson/stat309fall2022/main/chokeTall.csv")
#ensure that categorical variables are factors:
chokeTall$subjectID = factor(chokeTall$subjectID)
chokeTall$task = factor(chokeTall$task)
```

In this dataset, there are three columns:

1. *subjectID* : The ID number (identification number) for each subject (ranging from 1 to 67).
2. *task* : Either "easy1", "easy2", "hard1", or "hard2" (corresponding to the type of task for each row).
3. *score* : The score of the task.

Note that this dataset has the exact same information as the previous dataset, except that it is rearranged as a *tall* format. (At first, I thought about asking you all to make this dataset yourselves, because I thought it would be a good character-building exercise (<https://riteshjsr.files.wordpress.com/2011/10/ch890102.gif>); however, at this point I think you all have good enough character for me to just give you this dataset. But in return, I ask you to take a moment to understand how this dataset is arranged, e.g., by typing `chokeTall` on the Console.)

For this part, answer the following two questions.

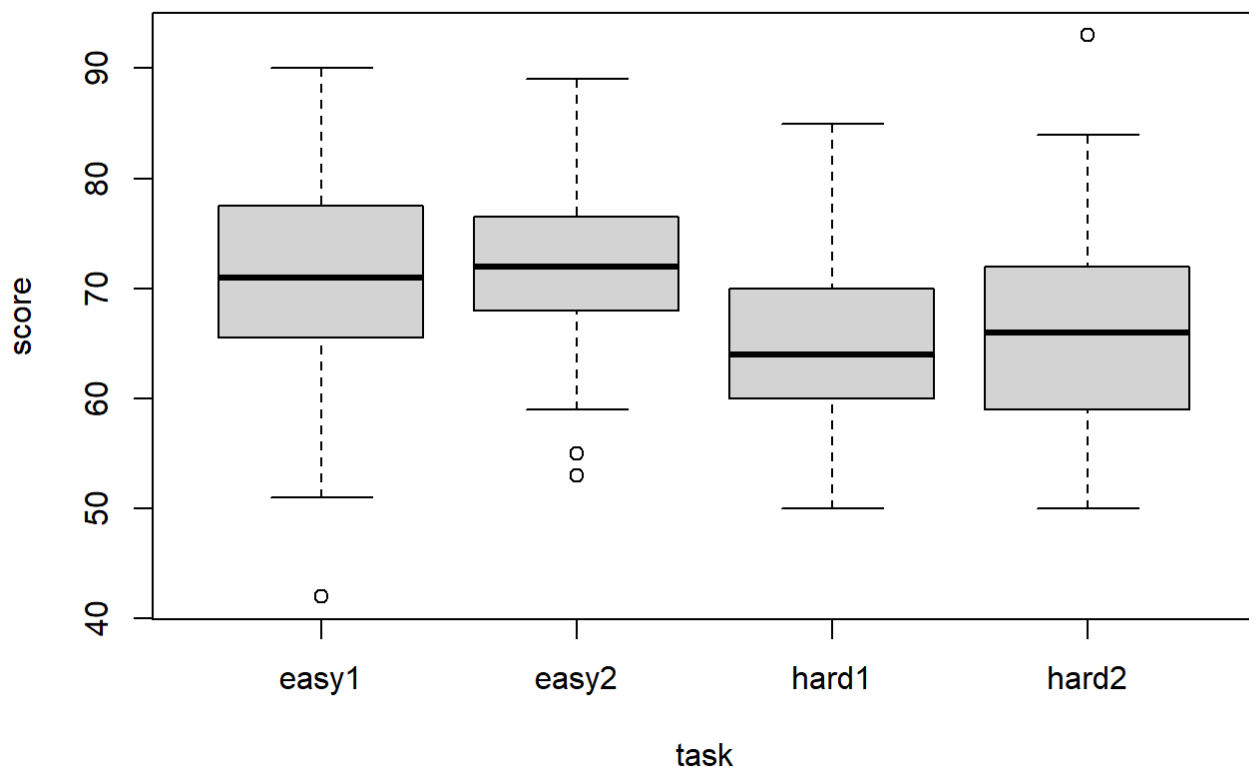
- (4pts) Using the symbols  $\mu_{E1}$ ,  $\mu_{E2}$ ,  $\mu_{H1}$ ,  $\mu_{H2}$  for the population means of the score for the four tasks, write the overall null hypothesis for repeated measures ANOVA.

**[H0:  $\mu_{E1}=\mu_{E2}=\mu_{H1}=\mu_{H2}$  NULL says mean is equal across all 4 categories ]**

- (6pts) As some EDA for this dataset, create a side-by-side boxplot, where the score is on the y-axis and there is a box for *easy1* , *easy2* , *hard1* , *hard2* . From the plot, which of the two categories (easy or hard) has a higher median score? Within each of the two categories (easy and hard), which task (the first or second task) has a higher median score?

Hide

```
boxplot(score ~ task, data = chokeTall)
```



[Easy category seems to have a higher median score than the hard category. Within each of the two categories (easy and hard), for easy ones the median score is approximately similar comparing easy1 and easy2, however in terms of hard ones, hard 2 seems to have a slightly higher median score compared to hard 1.]

e. (15pts) For this part, please complete the following **three tasks** using the `chokeTa11` dataset:

- (5pts) Using the `anova_test()` function within the `rstatix` library, run repeated measures ANOVA on the `chokeTa11` dataset. Remember to write `library(rstatix)` (which you need to have installed) before you write anything with `anova_test()`. For this task, you just need to write the code that successfully displays the output from `anova_test()`.

Hide

```
library(rstatix)
```

```
##
## Attaching package: 'rstatix'
```

```
## The following object is masked from 'package:stats':
##
## filter
```

Hide

```
anova_test(dv = score, wid = subjectID, within = task, data = chokeTall)
```

```
## ANOVA Table (type III tests)
##
## $ANOVA
##   Effect DFn DFd      F      p p<.05    ges
## 1   task   3 198 13.13 7.44e-08    * 0.102
##
## $`Mauchly's Test for Sphericity`
##   Effect      W      p p<.05
## 1   task 0.911 0.302
##
## $`Sphericity Corrections`
##   Effect   GGe      DF[GG]   p[GG] p[GG]<.05   HFe      DF[HF]   p[HF]
## 1   task 0.941 2.82, 186.36 1.62e-07          * 0.988 2.96, 195.55 8.76e-08
##   p[HF]<.05
## 1          *
```

- (5pts) Your `anova_test()` output should include something called “Mauchly’s Test for Sphericity”. What do you conclude from the “Mauchly’s Test for Sphericity” output? Explain your answer in 1-2 sentences, being sure to state a null hypothesis and p-value in your answer.

**[Since the p-value is larger than 0.05 (0.302), then we fail to reject the null that sphericity holds (meaning that each treatment has the same variance and each pair of treatment has the same correlation). then Sphericity holds and we need to use p-value assuming sphericity]**

- (5pts) Based on the `anova_test()` output, what do you conclude in terms of the null hypothesis you wrote in Part D? In your answer, please state the specific p-value you used to make this conclusion and whether you reject or fail to reject the null hypothesis.

**[Since Sphericity holds we need to use p-value assuming sphericity (use the p-value of the \$ANOVA, the first part). The conclusion is that We have enough evidence to reject the null hypothesis mentioned in part D as the p-value is less than 0.05 (7.44e-08).]**

## Question 2: Two-Way Mixed ANOVA (45 points)

Professor J. Birgisson (<https://en.wikipedia.org/wiki/J%C3%B3nsi>) was recently hired by Cabbage Corp ([https://avatar.fandom.com/wiki/Cabbage\\_Corp](https://avatar.fandom.com/wiki/Cabbage_Corp)), which owns a large number of restaurants. Cabbage Corp is considering a new “20% off appetizers” deal, but they’re not sure if they should make the deal on a Monday, Tuesday, Wednesday, or Thursday. They decide to recruit 20 of their restaurants to run an experiment. (These restaurants are all totally different - i.e., not a franchise. Cabbage Corp is a huge company.)

The experiment is run for four weeks. Each week, each restaurant is randomized to get the “20% off appetizers” deal on one day of the week. The morning of that day, they advertise the deal on social media, and see if revenue increases that day as a result of the deal. At the end of the four weeks, each restaurant experienced the deal on Monday, Tuesday, Wednesday, and Thursday, but in a random order.

Here’s the dataset resulting from the experiment:

Hide

```
restaurants = read.csv("https://raw.githubusercontent.com/zjbranson/stat309fall2022/main/restaurantData.csv")
#ensure categorical variables are factors
restaurants$id = factor(restaurants$id)
restaurants$day = factor(restaurants$day)
restaurants$type = factor(restaurants$type)
```

Here are the variables in the dataset:

- `id` : The ID of each restaurant (from 1 to 20).
- `day` : The day the restaurant experienced the deal ("mon", "tues", "wed", "thurs")
- `type` : The type of restaurant ("chinese", "italian", "seafood", or "vegetarian")
- `revenueAdded` : The revenue (in dollars) on that day, compared to an "average day" for that restaurant. Thus, `revenueAdded = 1000` means that revenue was \$1000 higher than the average day.

a. (9pts) For this part, answer the following two questions.

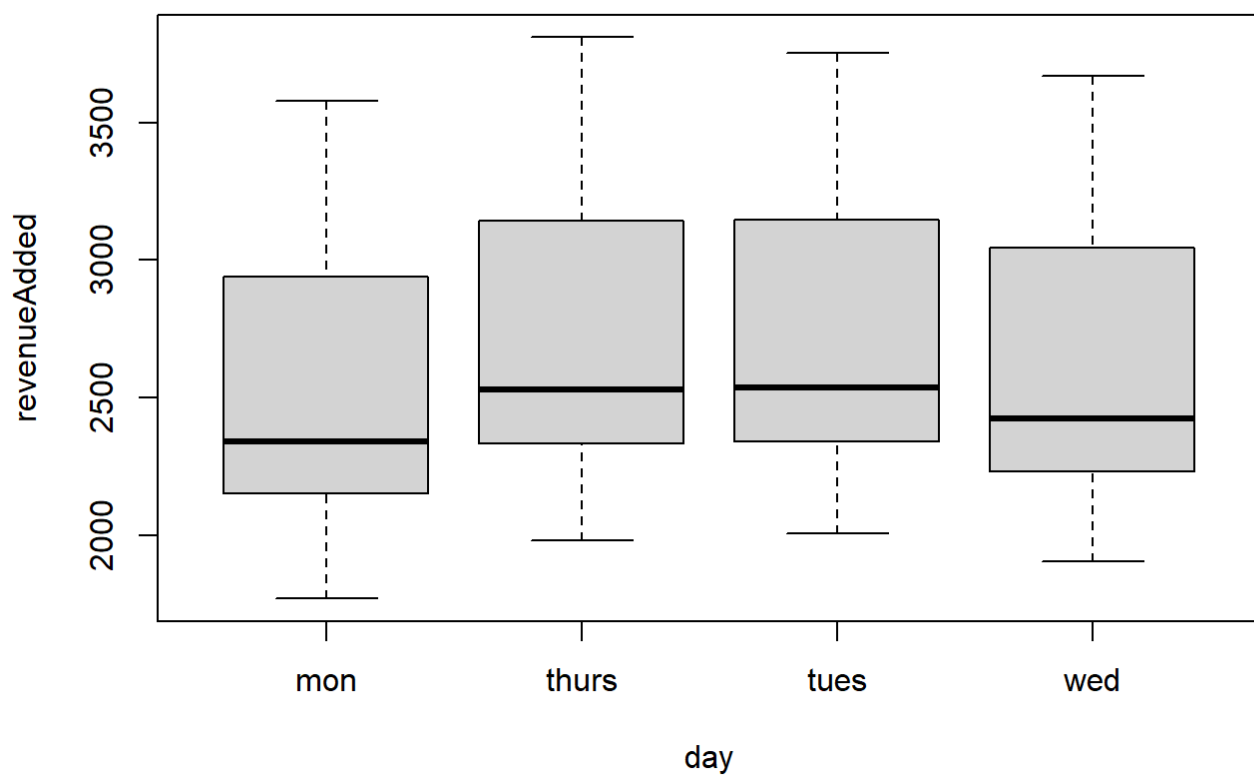
- (4pts) Professor Birgisson recognizes that you'll have to do a mixed within- and between-subjects ANOVA, but that's about all he knows for these data. In this dataset, which variable is the outcome? Furthermore, which variable is the "between-subjects factor" and which is the "within-subjects factor"? For the second question, please give an explanation for your answer in 1-2 sentences.

**[revenueAdded variable is the outcome. type variable is the "between-subjects factor" because it is same for a subject id across its all days (within factor). day is the "within-subjects factor" because a single subject id has multiple days.]**

- (5pts) As some initial EDA, create two boxplots: (1) A boxplot where `revenueAdded` is on the y-axis and `day` is on the x-axis, and (2) a boxplot where `revenueAdded` is on the y-axis and `type` is on the x-axis. After creating these boxplots, interpret each one in 1-3 sentences. In particular, discuss to what extent `day` and `type` are associated with `revenueAdded`, based on the boxplots.

Hide

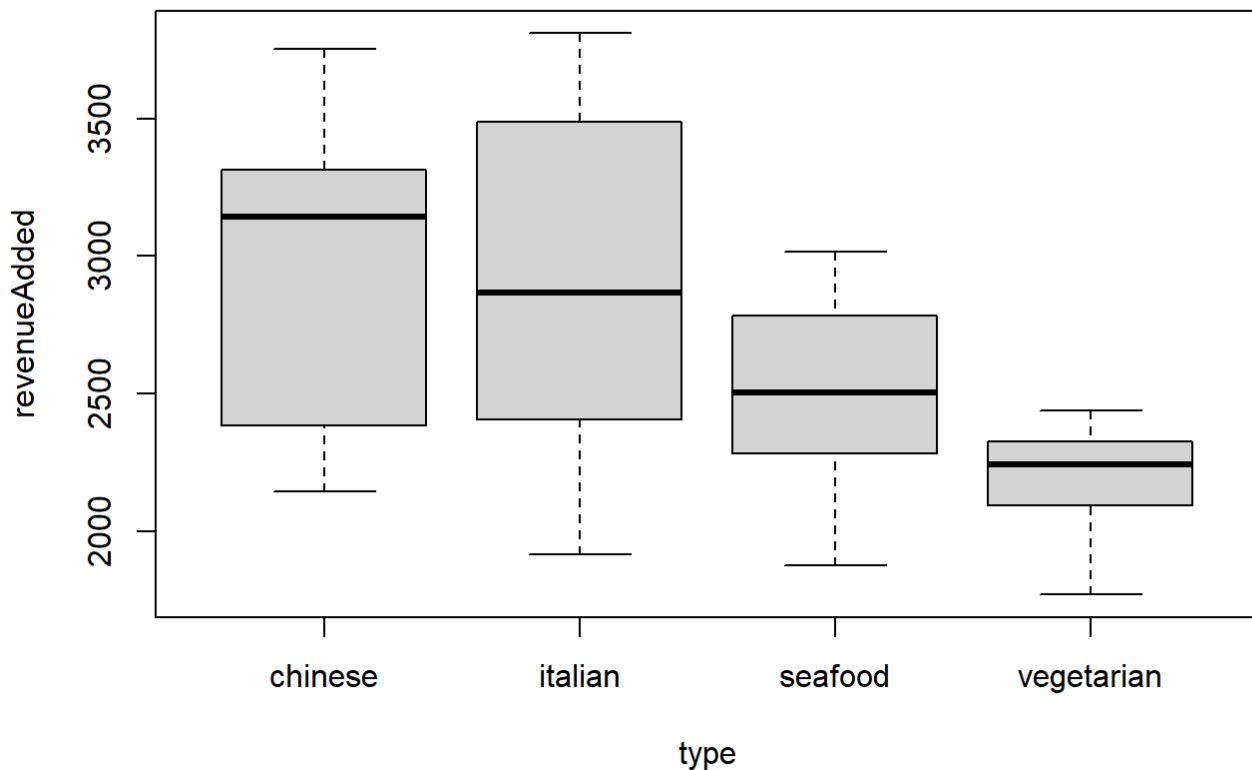
```
boxplot(revenueAdded ~ day, data = restaurants)
```



Hide

```
boxplot(revenueAdded ~ type, data = restaurants)
```





[The first EDA shows that revenue added does not differ extensively across different days, maybe just we can say that Mondays' revenues added are less compared to other days but not that much. however the second EDA shows that revenue added considerably changes across different types of the restaurants. Chinese's type has the highest median revenue added, followed by Italian and then followed by seafood. Vegetarian type of restaurant has the least revenue added compared to other types.]

b. (13pts) For this part, answer the following three questions.

- (4pts) Run the appropriate mixed within- and between-subjects ANOVA for this dataset. For this task, you just need to write the code that runs this analysis (and displays the resulting output).

Hide

```
anova_test(dv = revenueAdded, wid = id, within = day, between = type, data = restaurants)
```

```
## ANOVA Table (type II tests)
##
## $ANOVA
##      Effect DFn DFd      F      p p<.05      ges
## 1      type   3  16   2.591 8.90e-02    0.327000
## 2       day   3  48 716.431 7.75e-40    * 0.034000
## 3 type:day   9  48   0.858 5.68e-01    0.000125
##
## $`Mauchly's Test for Sphericity`
##      Effect      W      p p<.05
## 1       day 0.386 0.016      *
## 2 type:day 0.386 0.016      *
##
## $`Sphericity Corrections`
##      Effect GGe      DF[GG]      p[GG] p[GG]<.05      HFe      DF[HF]      p[HF]
## 1       day 0.736 2.21, 35.34 6.45e-30      * 0.859 2.58, 41.24 1.52e-34
## 2 type:day 0.736 6.63, 35.34 5.44e-01      0.859 7.73, 41.24 5.56e-01
##      p[HF]<.05
## 1      *
## 2
```

- (3pts) Your output should include output labeled “ANOVA” and “Sphericity Corrections”. Professor Birgisson notices that there is a `type` row in the “ANOVA” output but not the “Sphericity Corrections” output. Explain in 1-3 sentences to Professor Birgisson why that is.

**[This is because sphericity only makes sense for within-subjects factors, and type is not a within subject factor.]**

- (6pts) Both the “ANOVA” and “Sphericity Corrections” output include a row for `day` and a row for `type:day` (or `day:type`). For each of these rows, what is the null hypothesis? For the `day` row, please write out the null hypothesis mathematically using  $\mu$  symbols. In your answer, clearly define what your  $\mu$  symbols represent. For the `type:day` row, you can write out the null hypothesis in words.

**[For the day row: null hypothesis is that mean revenue added is same across all four days  $\mu$  represents mean revenue added**

**( $H_0: \mu(\text{monday}) = \mu(\text{thursday}) = \mu(\text{tuesday}) = \mu(\text{wednesday})$ ). For the type:day row, null hypothesis is that there is no interaction between day and type variables that would affect each other's effect on outcome variable.]**

- (8pts) Using the “ANOVA” and/or “Sphericity Corrections” output from Part B, state your scientific conclusions within the context of this dataset. In your answer, please specify the p-values you used to arrive at your conclusions, and explain why you used those p-values specifically.

**[Based on Mauchly's test row, we can conclude that Sphericity does not hold as P-value is less than 0.05 (0.016). so for the day and day:type scientific conclusions we will be using Sphericity corrections table. And since sphericity does not make sense for type alone, we will be using Anova table for scientific conclusion about the type. For day we can reject the null as p-value is less than 0.05 (6.45e-30 for GG, and 1.52e-34 for HF). Also there is no interaction between the type and day as we reject the null for that because of p-value being above 0.05 (5.44e-01 for GG, and 5.56e-01 for HF). In terms of Type, we fail to reject the null saying that revenue added is same across all restaurant types because the p-value is above 0.05 (0.327). ]**

- d. (15pts) At this point, Professor Birgisson decides to run a simpler analysis. He shows you the following code and output:

Hide

```
summary(aov(revenueAdded ~ day, data = restaurants))
```

```
##           Df    Sum Sq Mean Sq F value Pr(>F)
## day         3    540513   180171    0.592  0.622
## Residuals   76 23127749   304312
```

Using this output, answer the following three questions.

- (5pts) First, as a review of old material: What type of analysis did Professor Birgisson run here? After stating what kind of analysis he ran, explain to Professor Birgisson why this is not an appropriate analysis for these data.

**[He ran one-way anova, but this is not an appropriate analysis, because one of the main conditions of one-way anova is the independent subjects, however we know that this data set does not hold for that as each subject experiences different treatments (days). So repeated measure anova should have run, and even better than that mixed anova. As there was one other between-subject parameter called type. ]**

- (4pts) In the above output, there is a `day` row, and there is a null hypothesis that corresponds to this row (and thus a p-value, which we can see is 0.622 above). Is this null hypothesis the same or different from the null hypothesis for the `day` row that you discussed in Part B? State whether it's the Same or Different, and then explain your answer in 1-2 sentences.

**[null hypothesis  $\mu$  represents mean revenue added**

**( $H_0: \mu(\text{monday}) = \mu(\text{thursday}) = \mu(\text{tuesday}) = \mu(\text{wednesday})$ ) is the same both for part B hypothesis testing and here one-way anova. The reason is that in both cases we measure whether mean revenue added is the same across all four days between subjects, however due to the fact that subjects here are not independent, then this fact should be mentioned in the hypothesis testing itself, however both hypothesis tests are testing the same null hypothesis. ]**

- (6pts) Now look back at the p-value you got for the `day` row in the mixed ANOVA analysis from Part B. Which `day` p-value is *smaller*: the p-value from Professor Birgisson's analysis here (0.622), or the p-value from the mixed ANOVA analysis from Part B? After answering that, explain in 1-3 sentences *why* the p-value is smaller for one of the analyses. (For example: If the p-value is smaller for Professor Birgisson's analysis, why is it that the p-value is smaller for his analysis, compared to the mixed ANOVA analysis? Or alternatively, if the mixed ANOVA analysis p-value is smaller, why is that p-value smaller, compared to Professor Birgisson's here?)

**[the p-value from the mixed ANOVA analysis from Part B is smaller, because the mixed anova accounts for non-independent within-subject samples by removing subject-to-subject variability (decreasing sigma squared) which will eventually increase the power, and we will see lower p-values.]**