C2M2_peer_reviewed

February 29, 2024

1 C2M2: Peer Reviewed Assignment

1.0.1 Outline:

The objectives for this assignment:

1. Utilize contrasts to see how different pairwise comparison tests can be conducted.
2. Understand power and why it's important to statistical conclusions.
3. Understand the different kinds of post-hoc tests and when they should be used.

General tips:

1. Read the questions carefully to understand what is being asked.
2. This work will be reviewed by another human, so make sure that you are clear and concise in what your explanations and answers.

2 Problem 1: Contrasts and Coupons

Consider a hardness testing machine that presses a rod with a pointed tip into a metal specimen with a known force. By measuring the depth of the depression caused by the tip, the hardness of the specimen is determined.

Suppose we wish to determine whether or not four different tips produce different readings on a hardness testing machine. The experimenter has decided to obtain four observations on Rockwell C-scale hardness for each tip. There is only one factor - tip type - and a completely randomized single-factor design would consist of randomly assigning each one of the 4×4=16 runs to an experimental unit, that is, a metal coupon, and observing the hardness reading that results. Thus, 16 different metal test coupons would be required in this experiment, one for each run in the design.

[28]:
```
tip    <- factor(rep(1:4, each = 4))
coupon <- factor(rep(1:4, times = 4))
y <- c(9.3, 9.4, 9.6, 10,
       9.4, 9.3, 9.8, 9.9,
       9.2, 9.4, 9.5, 9.7,
       9.7, 9.6, 10, 10.2)
hardness <- data.frame(y, tip, coupon)
hardness
```

A data.frame: 16 × 3

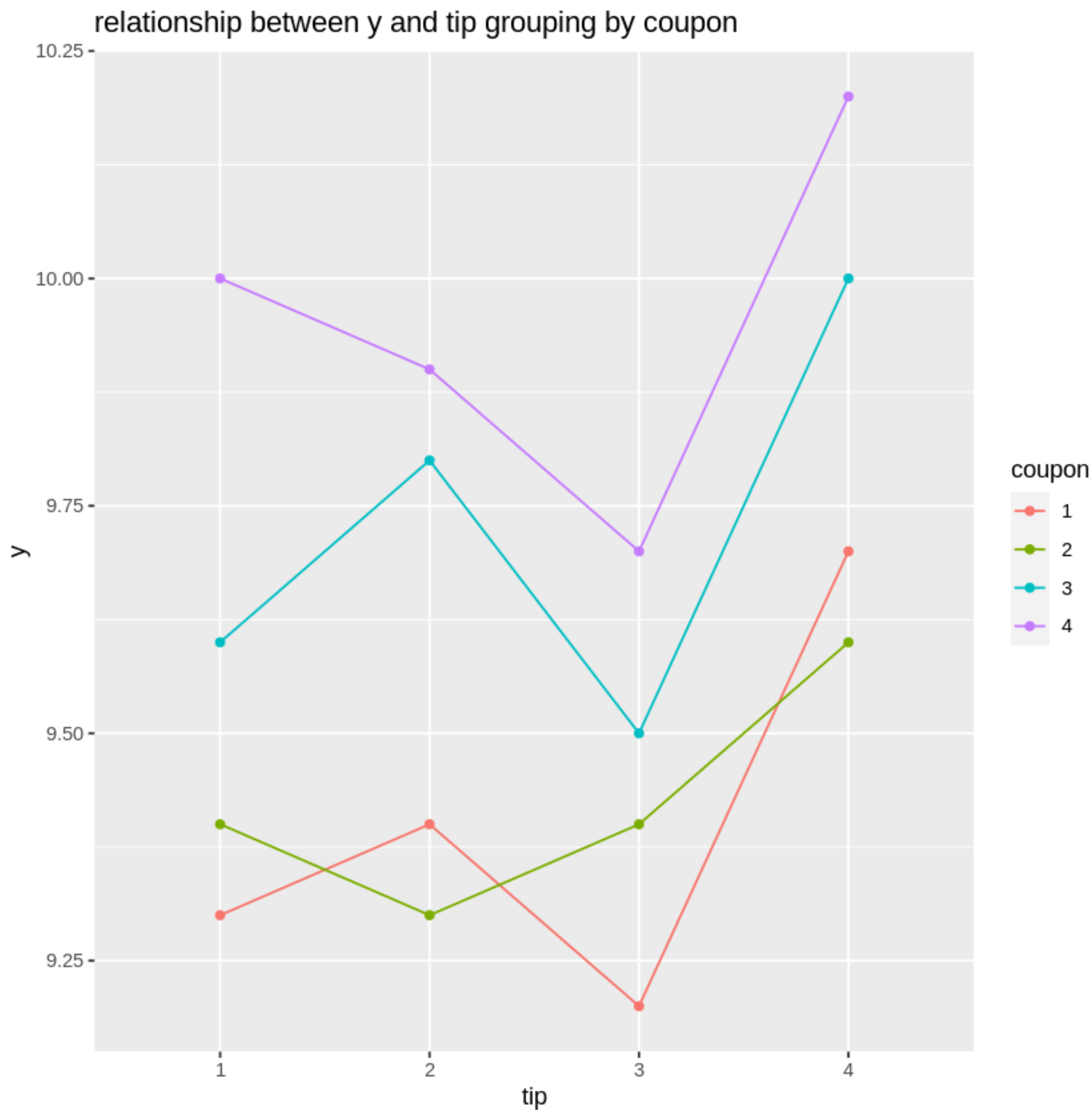
y <dbl>	tip <fct>	coupon <fct>
9.3	1	1
9.4	1	2
9.6	1	3
10.0	1	4
9.4	2	1
9.3	2	2
9.8	2	3
9.9	2	4
9.2	3	1
9.4	3	2
9.5	3	3
9.7	3	4
9.7	4	1
9.6	4	2
10.0	4	3
10.2	4	4

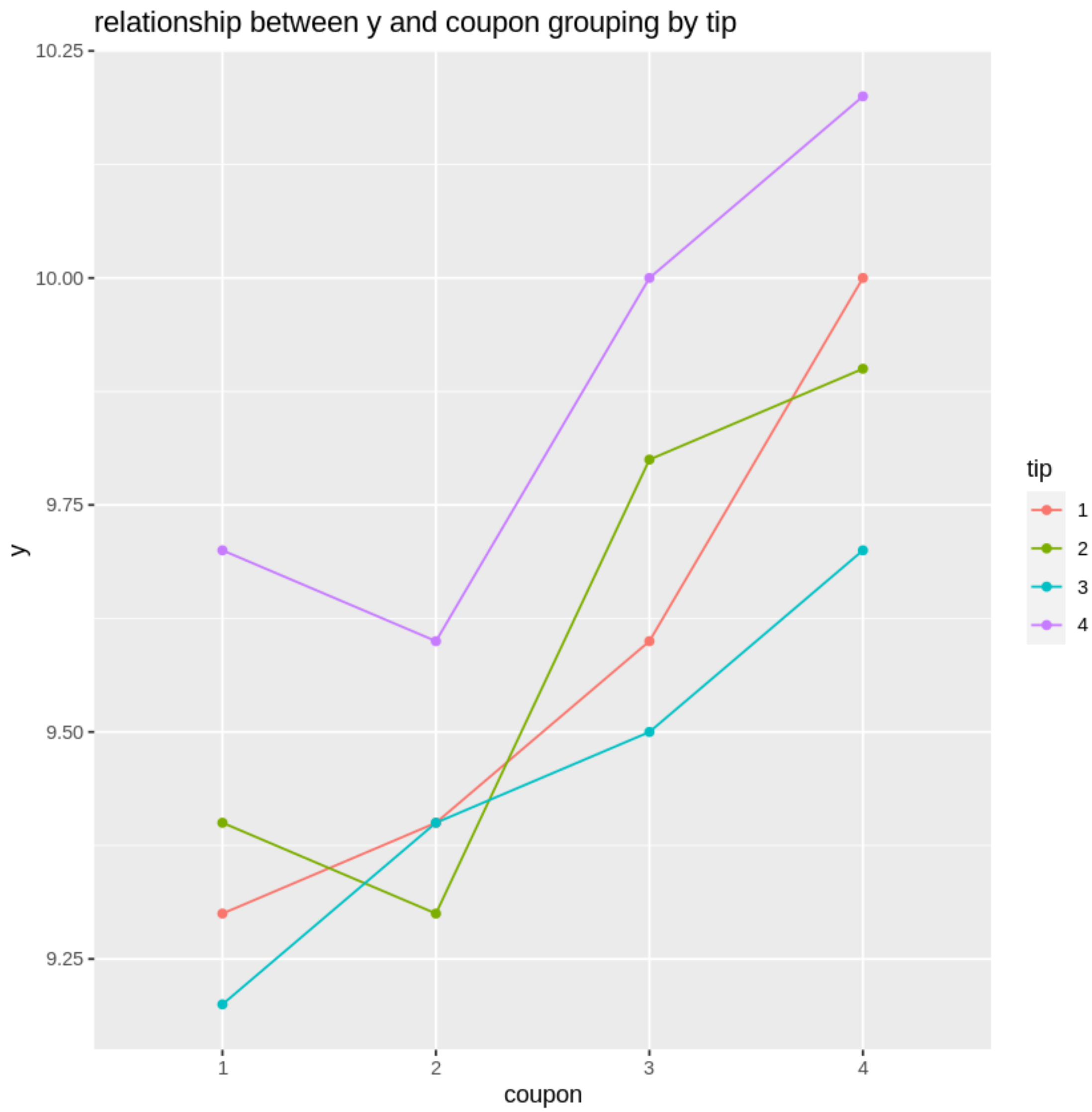
2.0.1 1. (a) Visualize the Groups

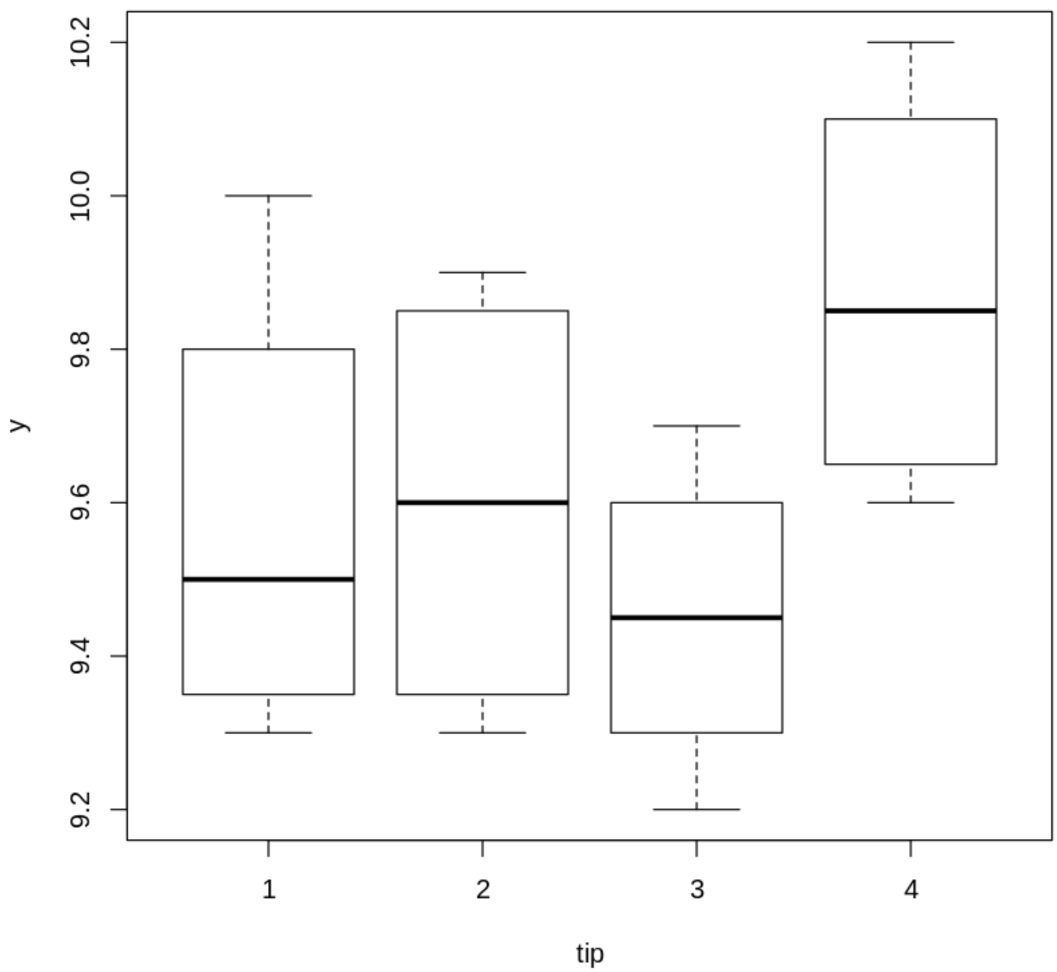
Before we start throwing math at anything, let's visualize our data to get an idea of what to expect from the eventual results.

Construct interaction plots for `tip` and `coupon` using ggplot(). Be sure to explain what you can from the plots.

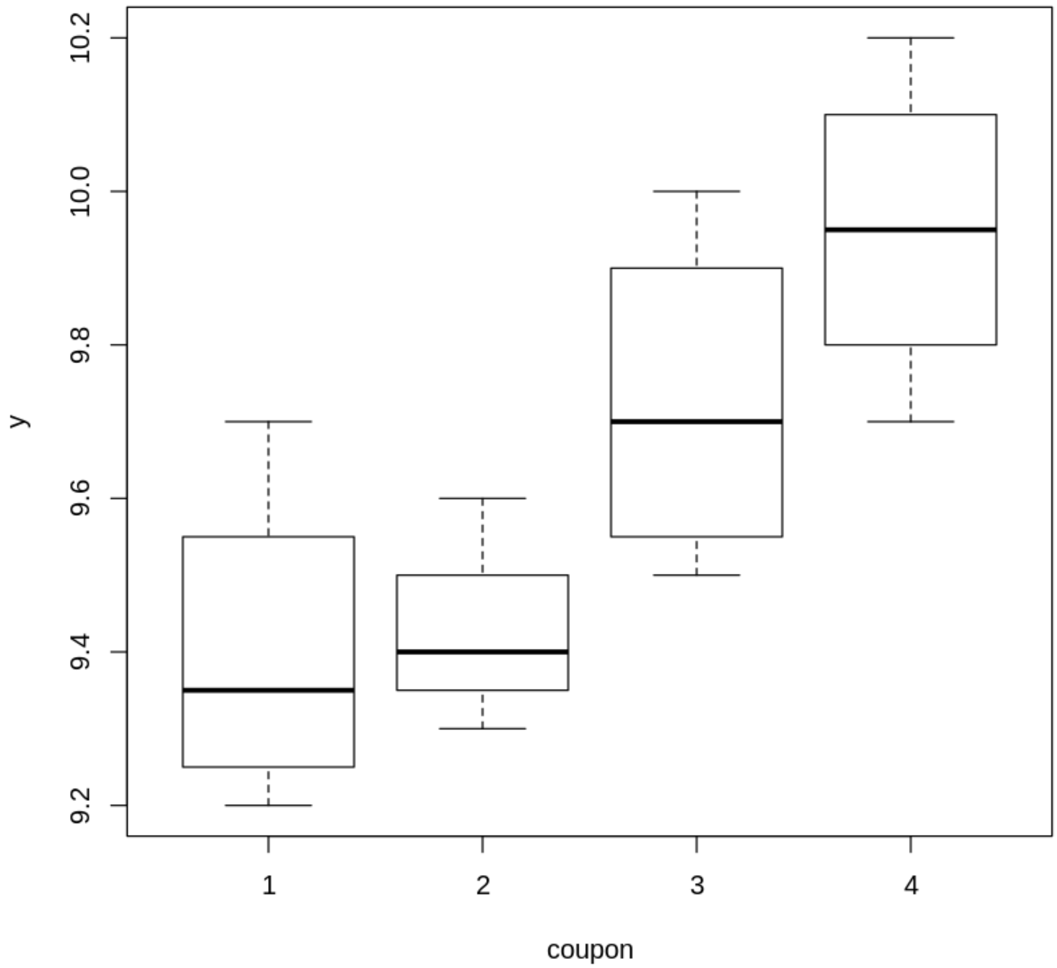
[29]:
```
# Your Code Here
library(ggplot2)
library(tidyverse)

hardness.aggregate = aggregate(y~tip+coupon,hardness,FUN=mean)

ggplot(hardness.aggregate, aes(x=tip, y=y, group=coupon, color=coupon)) +
    geom_line() +
    geom_point() +
    ggtitle('relationship between y and tip grouping by coupon')

ggplot(hardness.aggregate, aes(x=coupon, y=y, group=tip, color=tip)) +
    geom_line() +
    geom_point() +
    ggtitle('relationship between y and coupon grouping by tip')

plot(y~tip+coupon,hardness)
```


relationship between y and tip grouping by coupon
10.25
10.00
9.75
9.50
9.25
y
1
2
3
4
tip
coupon
1
2
3
4


relationship between y and coupon grouping by tip
10.25
10.00
9.75
9.50
9.25
y
1
2
3
4
coupon
tip
1
2
3
4



10.2
10.0
9.8
9.6
9.4
9.2
y
1
2
3
4
tip



1. (a) Answer: The graphs of the interaction plots show that there is likely an interaction between the groups, but it depends on which group is observed. For instance, when grouping by coupon, coupons 1 and 3 have the same slopes, which would imply there is no interaction there. Likewise, only looking at tips 1 and 4 in the second graph that groups by tip would result in the same conclusion. This does not apply to other pairs in the respective graphs. So, overall, the graphs suggest that there is an interaction term.

2.0.2 1. (b) Interactions

Should we test for interactions between `tip` and `coupon`? Maybe there is an interaction between the different metals that goes beyond our current scientific understanding!

Fit a linear model to the data with predictors `tip` and `coupon`, and an interaction between the two.

Display the summary and explain why (or why not) an interaction term makes sense for this data.

[30]:
```r
# Your Code Here
lmod = lm(y ~ tip + coupon, data=hardness)
print("lmod summary")
summary(lmod)

lmod.interaction = lm(y ~ tip + coupon + tip:coupon, data=hardness)
print("lmod.interaction summary")
summary(lmod.interaction)
```

```
[1] "lmod summary"

Call:
lm(formula = y ~ tip + coupon, data = hardness)

Residuals:
     Min       1Q   Median       3Q      Max
-0.10000 -0.05625 -0.01250  0.03125  0.15000

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.35000    0.06236 149.934  < 2e-16 ***
tip2         0.02500    0.06667   0.375 0.716345
tip3        -0.12500    0.06667  -1.875 0.093550 .
tip4         0.30000    0.06667   4.500 0.001489 **
coupon2      0.02500    0.06667   0.375 0.716345
coupon3      0.32500    0.06667   4.875 0.000877 ***
coupon4      0.55000    0.06667   8.250 1.73e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09428 on 9 degrees of freedom
Multiple R-squared:  0.938,Adjusted R-squared:  0.8966
F-statistic: 22.69 on 6 and 9 DF,  p-value: 5.933e-05
```

```
[1] "lmod.interaction summary"

Call:
lm(formula = y ~ tip + coupon + tip:coupon, data = hardness)

Residuals:
ALL 16 residuals are 0: no residual degrees of freedom!

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.300e+00         NA      NA       NA
```

```
tip2           1.000e-01         NA      NA       NA
tip3          -1.000e-01         NA      NA       NA
tip4           4.000e-01         NA      NA       NA
coupon2        1.000e-01         NA      NA       NA
coupon3        3.000e-01         NA      NA       NA
coupon4        7.000e-01         NA      NA       NA
tip2:coupon2 -2.000e-01         NA      NA       NA
tip3:coupon2  1.000e-01         NA      NA       NA
tip4:coupon2 -2.000e-01         NA      NA       NA
tip2:coupon3  1.000e-01         NA      NA       NA
tip3:coupon3 -3.758e-15         NA      NA       NA
tip4:coupon3 -3.869e-15         NA      NA       NA
tip2:coupon4 -2.000e-01         NA      NA       NA
tip3:coupon4 -2.000e-01         NA      NA       NA
tip4:coupon4 -2.000e-01         NA      NA       NA

Residual standard error: NaN on 0 degrees of freedom
Multiple R-squared:      1,Adjusted R-squared:    NaN
F-statistic:   NaN on 15 and 0 DF,  p-value: NA
```

1. (b) Answer: Looking at the summary results, it appears that a reaction term for this data does not make sense. The summary of the linear model with a reaction term has 0 degress of freedom which could mean that there is not enough data for a regression model of its size, or possibly that the data is linearly dependent and cannot be regressed on. The first option seems more likely, and there could be more reasons that this is happening, but I have only mentioned a few. The linear model without the interaction term shows that tip 4, coupon 3, and coupon 4 are significantly different and thus there is an interaction.

2.0.3 1. (c) Contrasts

Let's take a look at the use of contrasts. Recall that a contrast takes the form

$$\sum_{i=1}^{t} c_i \mu_i = 0,$$

where $\mathbf{c} = (c_1, ..., c_t)$ is a constant vector and $\mu = (\mu_1, ..., \mu_t)$ is a parameter vector (e.g., μ_1 is the mean of the i^{th} group).

We can note that $\mathbf{c} = (1, -1, 0, 0)$ corresponds to the null hypothesis $H_0 : \mu_2 - \mu_1 = 0$, where μ_1 is the mean associated with tip1 and μ_2 is the mean associated with tip2. The code below tests this hypothesis.

Repeat this test for the hypothesis $H_0 : \mu_4 - \mu_3 = 0$. Interpret the results. What are your conclusions?

[31]:
```
library(multcomp)
lmod = lm(y~tip+coupon, data=hardness)
fit.gh2 = glht(lmod, linfct = mcp(tip = c(1,-1,0,0)))

#estimate of mu_2 - mu_1
with(hardness, sum(y[tip == 2])/length(y[tip == 2]) -
     sum(y[tip == 1])/length(y[tip == 1]))
```

```
Loading required package: mvtnorm

Loading required package: survival

Loading required package: TH.data

Loading required package: MASS


Attaching package: 'MASS'


The following object is masked from 'package:dplyr':

    select



Attaching package: 'TH.data'


The following object is masked from 'package:MASS':

    geyser
```

0.0250000000000021

[33]:
```
#estimate of mu_4 - mu_3
with(hardness, sum(y[tip == 4])/length(y[tip == 4]) -
     sum(y[tip == 3])/length(y[tip == 3]))
```

0.425000000000001

1. (c) Answer: The test of the hypothesis $H_0 : \mu_4 - \mu_3 = 0$ suggests that the null hypothesis should not be rejected with a result of 0.425000.

2.0.4 1. (d) All Pairwise Comparisons

What if we want to test all possible pairwise comparisons between treatments. This can be done by setting the treatment factor (`tip`) to "Tukey". Notice that the p-values are adjusted (because we are conducting multiple hypotheses!).

Perform all possible Tukey Pairwise tests. What are your conclusions?

[51]:
```
# Your Code Here
lmod = lm(y~tip+coupon, data=hardness)
fit.gh2 = glht(lmod, linfct = mcp(tip = "Tukey"))

#estimate of mu_4 - mu_1
with(hardness, sum(y[tip == 4])/length(y[tip == 4]) -
     sum(y[tip == 1])/length(y[tip == 1]))

#estimate of mu_3 - mu_1
with(hardness, sum(y[tip == 3])/length(y[tip == 3]) -
     sum(y[tip == 1])/length(y[tip == 1]))

#estimate of mu_2 - mu_1
with(hardness, sum(y[tip == 2])/length(y[tip == 2]) -
     sum(y[tip == 1])/length(y[tip == 1]))

#estimate of mu_4 - mu_2
with(hardness, sum(y[tip == 4])/length(y[tip == 4]) -
     sum(y[tip == 2])/length(y[tip == 2]))

#estimate of mu_4 - mu_3
with(hardness, sum(y[tip == 4])/length(y[tip == 4]) -
     sum(y[tip == 3])/length(y[tip == 3]))


#estimate of mu_3 - mu_2
with(hardness, sum(y[tip == 3])/length(y[tip == 3]) -
     sum(y[tip == 2])/length(y[tip == 2]))
```

0.300000000000001

-0.125

0.0250000000000021

0.274999999999999

0.425000000000001

-0.150000000000002

1. (d) Answer: The only significant difference to come out of all pairwise test is between tip 2 and 1. Aside: surely there was a better way to do this. I did not see the Tukey adustment in the p-values.

3 Problem 2: Ethics in my Math Class!

In your own words, answer the following questions:

- What is power, in the statistical context?
- Why is power important?
- What are potential consequences of ignoring/not including power calculations in statistical analyses?

2. Answer: Power is the probabilty that the null hypothesis is correctly rejected, which is a decent measure of a test. Power is important becuase, ideally, the null hypothesis is rejected when it is false. Additionally, an effective test has enough power to make conclusions on a population accurately. Neglecting the power of a statistical test can result in experiments that are too powerful and do not reflect practical conclusions for the real-world. Likewise, the power could be too little resulting in an equally inefficient test as more resources would be needed to arrive at the same conclusion as a properly tuned statistical test. On a large scale, improper power analysis can critically affect the body of scientific knowledge.

4 Problem 3: Post-Hoc Tests

There's so many different post-hoc tests! Let's try to understand them better. Answer the following questsions in the markdown cell:

- Why are there multiple post-hoc tests?
- When would we choose to use Tukey's Method over the Bonferroni correction, and vice versa?
- Do some outside research on other post-hoc tests. Explain what the method is and when it would be used.

3. Answer: There are multiple post-hoc test because they are designed to tune statistical experiments differently. Tukey's Method allows us to keep the level of false positives equal to the chosen alpha level. Bonferroni's correction instead adjusts the p-values to account for the elevated risk of Type I errors, and but increases Type II errors. We should choose Tukey's Method when we want to make a lot of pairwise comparisons, and we should use the Bonferroni correction when there is a small, planned set of comparisons. A third option is the Sheffe Method, which should be used when we are interested in *all* group comparisons included more complex ones. It's advantage is the its flexibility but the tradeoff is power.

```
[ ]:
```