C2M2_peer_reviewed

February 8, 2022

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 $4 \times 4 = 16$ 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.

[5]:
```
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

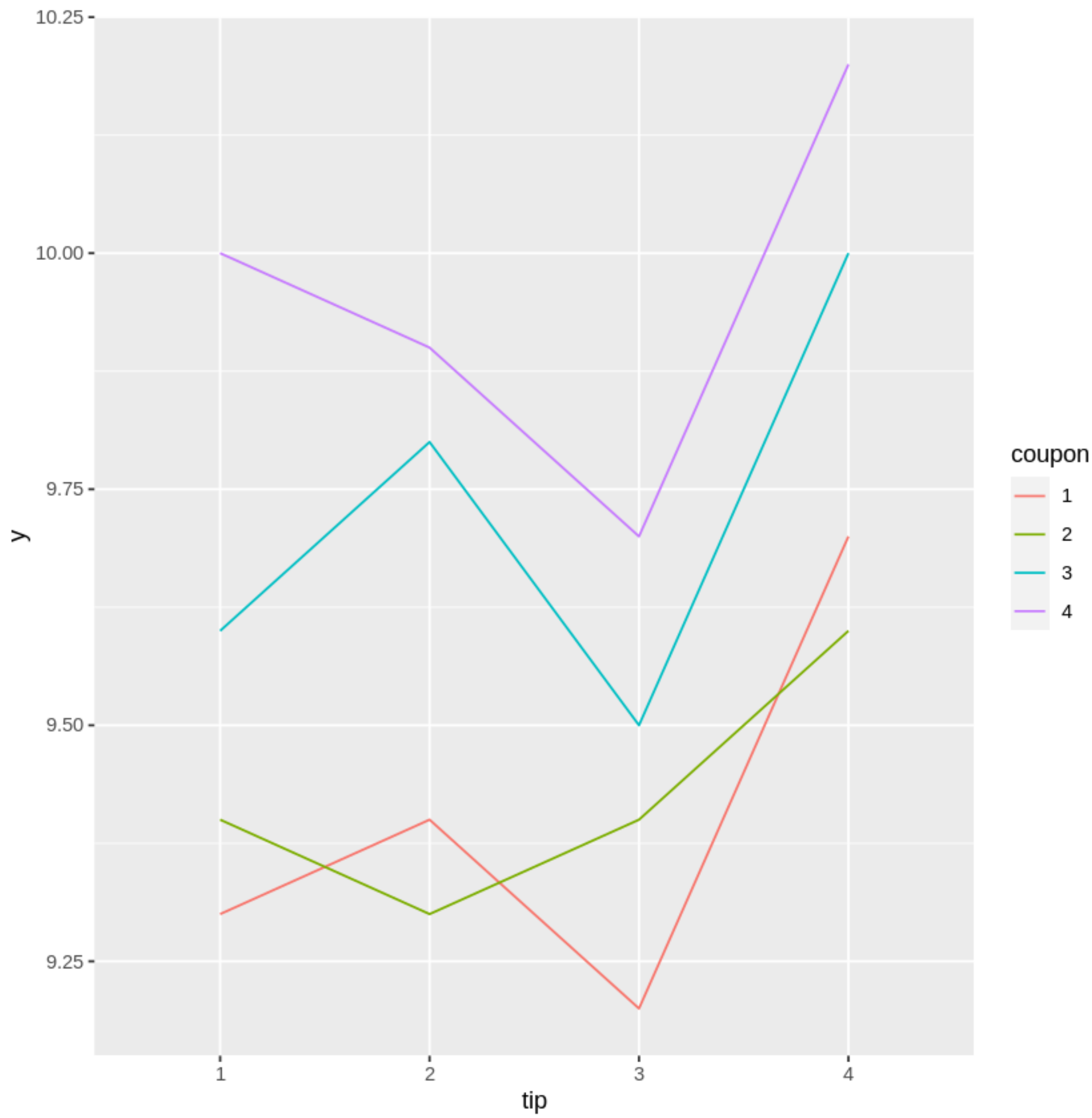
y	tip	coupon
<dbl>	<fct>	<fct>
9.3	1	1
9.4	1	2
9.6	1	3
10.0	1	4
9.4	2	1
9.3	2	2
9.8	2	3
9.9	2	4
9.2	3	1
9.4	3	2
9.5	3	3
9.7	3	4
9.7	4	1
9.6	4	2
10.0	4	3
10.2	4	4

2.0.1 1. (a) Visualize the Groups

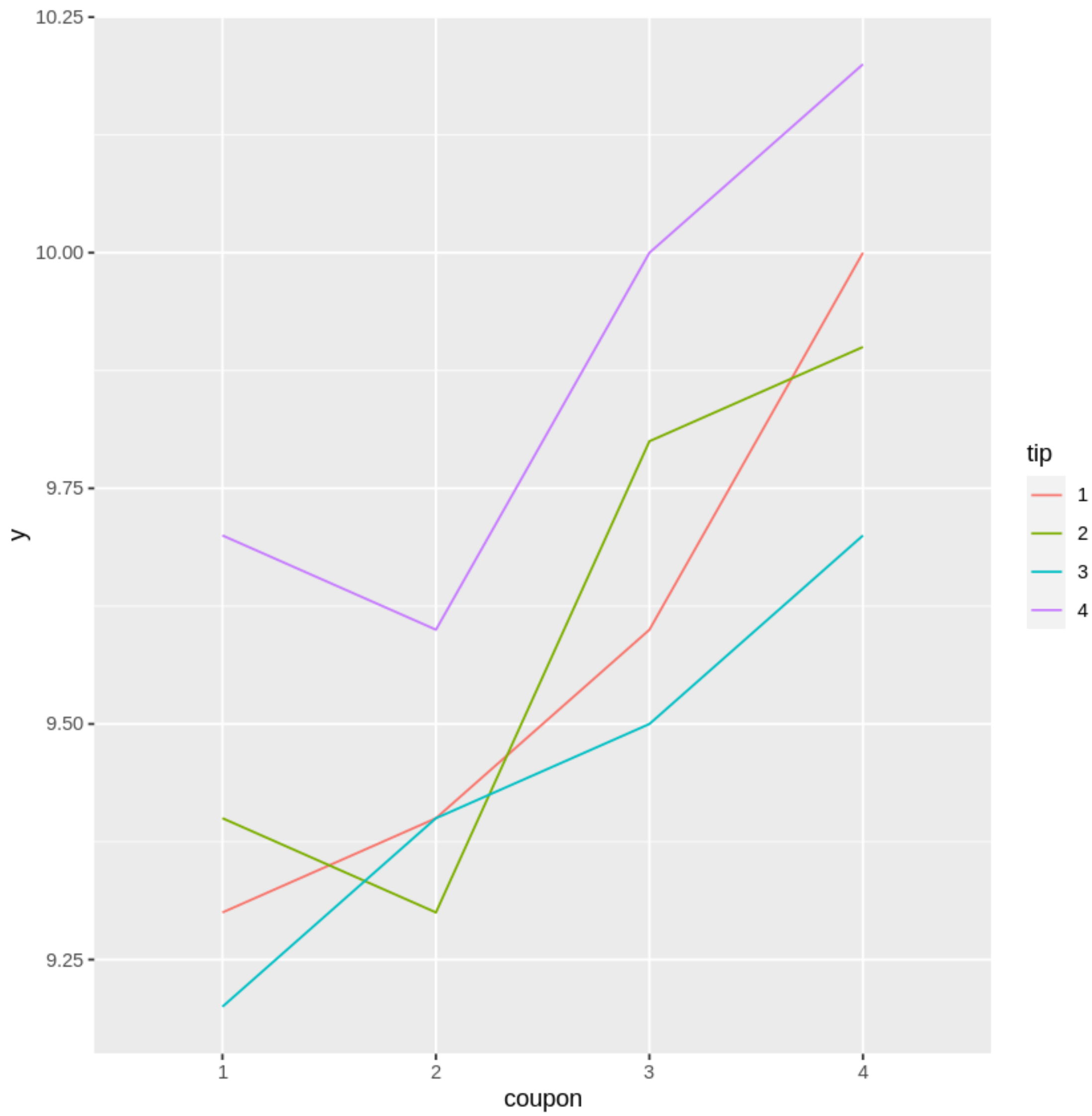
Before we start throwing math at anything, let's visualize our data to get an idea of what to expect from the eventual results.

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

[17]:
```
# Your Code Here
library(ggplot2)
plot1 = ggplot(data=hardness, aes(x=tip, y=y, group = coupon))
plot1 = plot1 + geom_line(aes(color = coupon))
plot1
```



[18]:
```
plot2 = ggplot(data=hardness, aes(x=coupon, y = y, group = tip))
plot2 = plot2 + geom_line(aes(color = tip))
plot2
```



These plots don't prove an interaction is needed. The information is inconsistent as some levels suggest an interaction while others don't. For example, in the first plot (Y vs Tip), the lines for coupons 1, 3 and 4 never cross; in fact, they seem to have a consistently similar slope throughout the plot. This suggets no interaction. However, coupons 2 and 4 seems to have some interaction as they both seems to have a larger hardness than the other for 2 of the 4 tips.

Also, another thing to note is that there is only one data point for each combination of coupons and tips. Collecting more data could generate a clearer picture.

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.

[9]:
```
# Your Code Here
lm1 = lm(y ~ tip + coupon + tip:coupon, data=hardness)
summary(lm1)
```

```
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
```

We do not have any information regarding f-tests, t-tests and p-value. This is because creating dummy variable caused us to have as many data points as paraemter: this results in 0 degrees of freedom. We can't test any data; therefore, we shouldn't use an interaction term. We won't be able to test if it is significant or not.

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?

[10]:
```
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: 'TH.data'


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

    geyser
```

0.0250000000000021

[19]:
```
fit.gh3 = glht(lmod, linfct = mcp(tip = c(0,0,-1,1)))
summary(fit.gh3)

with(hardness, sum(y[tip == 4])/length(y[tip == 4]) -
     sum(y[tip == 3])/length(y[tip == 3]))
```

```
 Simultaneous Tests for General Linear Hypotheses
```

```
Multiple Comparisons of Means: User-defined Contrasts


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

Linear Hypotheses:
       Estimate Std. Error t value Pr(>|t|)
1 == 0  0.42500    0.06667   6.375 0.000129 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

0.425000000000001

We have a p-value of 0.00129 which is less than our significance level of 0.05. We reject the null hypothesis that there is no difference in mean. We have statistical evidence that the mean of u4 and u3 does not equal 0.

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?

[20]:
```
# Your Code Here
fit.gh3 = glht(lmod, linfct = mcp(tip = "Tukey"))
summary(fit.gh3)
```

```
 Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


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

Linear Hypotheses:
           Estimate Std. Error t value Pr(>|t|)
2 - 1 == 0  0.02500    0.06667   0.375   0.9809
3 - 1 == 0 -0.12500    0.06667  -1.875   0.3027
4 - 1 == 0  0.30000    0.06667   4.500   0.0066 **
3 - 2 == 0 -0.15000    0.06667  -2.250   0.1812
4 - 2 == 0  0.27500    0.06667   4.125   0.0114 *
4 - 3 == 0  0.42500    0.06667   6.375   <0.001 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Adjusted p values reported -- single-step method)
```

At an 0.05 significance level, the following tips result in different means of hardness: - 4 and 1 - 4 and 2 - 4 and 3

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?

- Power is when we reject the null hypohtesis when the alternative is true. In other words, the probability we will find a true value when there is one to find.
- The higher power we have, the less likely we are to commit a Type II error. Therefore we are more likely to detect an effect (reject the null hypothesis) when we are supposed to.
- Ignoring the power of statistical analyses can have many consequences. If you obtain a mind breaking statistical result, but have low power, then your result may not be practical. On the other hand, with low power you have a higher probability to avoid find a relationship between levels when there is actually one. This can be irresponsible in social aspects. For example, say we ran a test on racial discrimination. We had a null hypothesis stating there wasn't any and we rejected. If we had a low statistical power, we more than likely had an invalid statistically significant result. This can cause policy reforms, etc.

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.

- Because there is no exact way to adjust for multiple statistical tests. Different methods adjust the level of significance in different ways.
- We want to use the Bonferroni method when we want a a lower type I error compared to a type II error. The Tukey method is best when we are concered about type II error.
- There is the Scheffe's Test. It is Similar to the Tukey's method as but a little more conservative. This results ain a consistently low Type I error; however, this comes at a cost of the statistical power. As each test of means is carried out, it corrects for alpha

[ ]: