C2M2_peer_reviewed

June 29, 2023

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.

[1]:
```
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

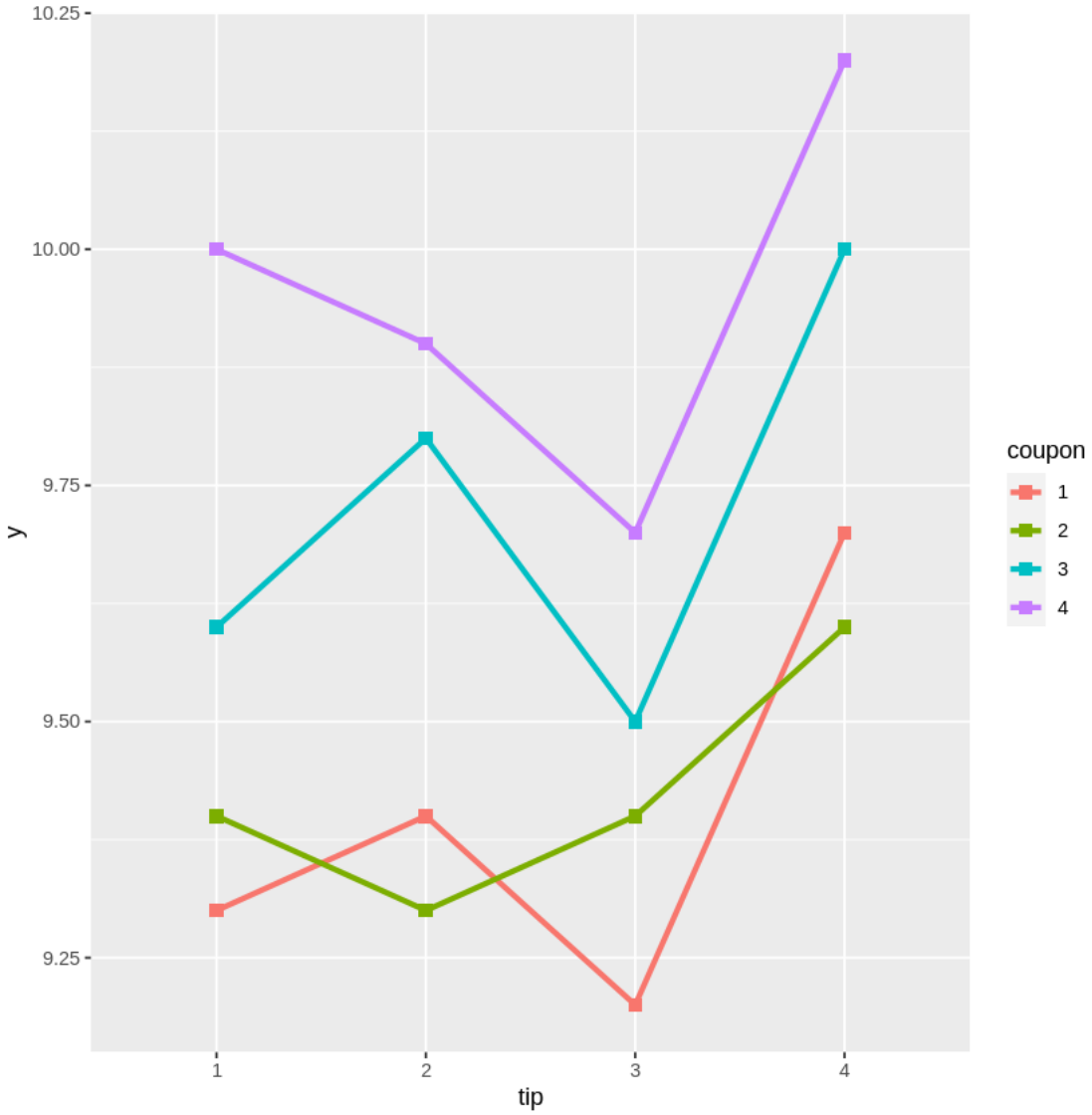
y <dbl>	tip <fct>	coupon <fct>
9.3	1	1
9.4	1	2
9.6	1	3
10.0	1	4
9.4	2	1
9.3	2	2
9.8	2	3
9.9	2	4
9.2	3	1
9.4	3	2
9.5	3	3
9.7	3	4
9.7	4	1
9.6	4	2
10.0	4	3
10.2	4	4

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.

[2]:
```
# Your Code Here
library(ggplot2)
ggplot(data=hardness, aes(x=tip, y=y)) +
    geom_line(size = 1.2, aes(group = coupon, color = coupon)) +
    geom_point(size = 2.6, aes(color = coupon), shape = 15)
```



In general we observe hardness variation across tips and coupons with an exception of tip 4 which seem to yield higher hardness readings for all coupons.

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.

[3]:
```
# Your Code Here
model = lm(y ~ tip + coupon + tip:coupon, data=hardness)
```

```
summary(model)
```

```
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
```

It does not make sense to include interaction as the number of parameters equals the number of data points so we're clearly overfitting as reflected by R-squared = 1.

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?

[4]:
```
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]))

#estimate of mu_4 - mu_3
fit.gh3 = glht(lmod, linfct = mcp(tip = c(0, 0, -1, 1)))
with(hardness, sum(y[tip == 4])/length(y[tip == 4]) -
     sum(y[tip == 3])/length(y[tip == 3]))

summary(fit.gh2)
summary(fit.gh3)
```

```
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

0.425000000000001

```
 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 -0.02500    0.06667  -0.375    0.716
(Adjusted p values reported -- single-step method)


 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)
```

We can't reject the null for $H_0 : \mu_2 - \mu_1 = 0$ as the corresponding p-value is 0.716 and is greater than the usual 0.05 significance level.

At the same time we have to reject the null for $H_0 : \mu_4 - \mu_3 = 0$ as the corresponding p-value is 0.000129 which is lower than the usual 0.05 significance level.

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?

[5]:
```
# Your Code Here
TukeyHSD(aov(lm(y ~ tip, data=hardness)))
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

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

$tip
      diff       lwr     upr     p adj
```

```
2-1  0.025 -0.55152 0.60152 0.9991931
3-1 -0.125 -0.70152 0.45152 0.9157208
4-1  0.300 -0.27652 0.87652 0.4431278
3-2 -0.150 -0.72652 0.42652 0.8653399
4-2  0.275 -0.30152 0.85152 0.5135079
4-3  0.425 -0.15152 1.00152 0.1815685
```

Once we adjust for multiple comparisons, all the null hypothesis comparing tip pairs can't be rejected as none of the adjusted p-values are below 0.05.

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 of a test is the ability to reject a null hypothesis when the null hypothesis is false i.e. P(reject H0 | false H0).

- Usually the null hypothesis states there's no effect and alternative is there is an effect. Then rejection of the null with false null represents the ability of the test to detect an effect when the effect is truly present.

- Tests with low power are unlikely to be reproducible.

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.

- There's no consensus on how to deal with post-hoc tests hence many approaches (tests).

- Tukey test is a generalization of the t-test for pairwise group comparisons to account for multiple testing. Bonferroni correction is applicable to a more general setting where more than one hypothesis needs to be tested simultaneously (not necessarily pairwise comparisons but obviously could be pairwise group comparisons). Bonferroni correction tends to be more conservative (especially for a large number of hypotheses to be tested) meaning it will increase type II error rates (not rejecting null when null is false or in other words not detecting an

effect when there is an effect). It makes sense to use Tukey when testing pairwise comparisons and Bonferroni for other types of hypotheses.

- There's also Fisher's Least Significant Difference Test that could be used for pairwise group comparisons. It works in two stages: 1. F-test is run to check null hypothesis stating that all group means are equal and is done at given significance level; 2. If the F-test rejects the null, one can run individual t-test for pairwise comparisons at the same significance level as the F-test in step 1. Then there's Scheffé's method where simultaneous hypothesis testing is done for arbitrary contracts of means of groups, not just pairwise comparisons like in Tukey's method.