C2M2_peer_reviewed

July 6, 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
```

y	tip	coupon
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]:
```
library(ggplot2)

# Construct interaction plot
interaction_plot <- ggplot(hardness, aes(x = tip, y = y, color = coupon)) +
  geom_point() +
  geom_line(aes(group = coupon)) +
  xlab("Tip Type") +
  ylab("Hardness") +
  ggtitle("Interaction Plot: Hardness by Tip Type and Coupon") +
  theme_bw()

# Display the interaction plot
print(interaction_plot)
```

In this code, we first load the ggplot2 package. Then, we use the ggplot() function to create a plot object and map the x-axis to "tip", the y-axis to "y", and color the plot by "coupon". We add the geom_point() function to display individual data points and geom_line() to connect them with lines for each coupon. The x-axis represents the tip type, the y-axis represents the hardness reading, and different colors represent different coupons.

The resulting plot shows the interaction between tip type and coupon on the hardness readings. Each line represents a different coupon, and the points on the lines indicate the hardness readings

for each tip type. By observing the plot, we can gain insights into the nature of the interaction between the variables.

Specifically, we can observe:

If the lines for different coupons are parallel, it suggests that there is no interaction between tip type and coupon, and the effect of tip type on hardness is consistent across all coupons. If the lines are not parallel and cross each other, it indicates an interaction between tip type and coupon. The effect of tip type on hardness may differ depending on the coupon. By examining the plot, we can visually assess any interaction effects between tip type and coupon on hardness and gain insights into the relationship between the variables.

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]:
```
# Fit the linear model with interaction
model <- lm(y ~ tip * coupon, data = hardness)

# Display the summary
summary(model)
```

By fitting the linear model with the interaction term, we can assess the significance and interpret the coefficients associated with the interaction. The summary output will provide information about the estimated coefficients, their standard errors, t-values, and p-values.

Analyzing the summary output, we can determine whether the interaction term makes sense for this data. Specifically, we can look at the p-value associated with the interaction term. If the p-value is below a predetermined significance level (e.g., 0.05), it suggests that there is evidence of an interaction between tip and coupon.

In addition to the statistical significance, it is important to consider the context and theoretical understanding of the experiment. If there are scientific reasons or prior knowledge that suggest the possibility of an interaction between the different metals (coupons) and tip types, it further supports the inclusion of the interaction term in the model.

However, if the p-value for the interaction term is not statistically significant, or there are no theoretical reasons to expect an interaction, it might suggest that the effect of tip type on hardness does not depend on the specific metal (coupon) used. In such cases, a simpler model without the interaction term could be appropriate.

Interpreting the output and considering the significance of the interaction term in the context of the experiment will help determine whether including the interaction makes sense for this data.

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]))
```

0.0250000000000021

Interpreting the results, if the p-value for this contrast is below a predetermined significance level (e.g., 0.05), it suggests that there is evidence to reject the null hypothesis and conclude that there is a significant difference between the means associated with tip4 and tip3.

On the other hand, if the p-value is not statistically significant, it suggests that there is not enough evidence to reject the null hypothesis, and we cannot conclude a significant difference between the means of tip4 and tip3.

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?

[11]:
```
# Perform Tukey pairwise tests
tukey_result <- TukeyHSD(lmod, "tip")

# Display the results
print(tukey_result)
```

The TukeyHSD() function calculates the pairwise comparisons between treatments and adjusts the p-values for multiple comparisons using the Tukey method. The resulting tukey_result object contains the pairwise comparisons along with the adjusted p-values.

By examining the output, you can interpret the results and draw conclusions about the differences in means between the different tips. The output will provide information about the estimated differences, standard errors, confidence intervals, and adjusted p-values for each pairwise comparison.

Specifically, you can focus on the adjusted p-values to determine the significance of each pairwise comparison. If the adjusted p-value is below a predetermined significance level (e.g., 0.05), it suggests a significant difference between the means of the corresponding tips. On the other hand, if the adjusted p-value is not statistically significant, it indicates that there is not enough evidence to conclude a significant difference between the means.

By examining the Tukey pairwise test results, you can identify which specific pairwise comparisons have significant differences in mean hardness readings and draw conclusions about the relative performances of different tip types in the hardness testing machine.

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?

In the statistical context, power refers to the probability of correctly detecting a true effect or rejecting a false null hypothesis in a statistical test. It measures the ability of a statistical test to detect a meaningful difference or relationship between variables when it exists.

Power is important because it allows us to assess the sensitivity of a statistical test. It helps us determine the likelihood of finding a statistically significant result when there is a true effect present. A high power indicates a greater chance of detecting a true effect, while a low power suggests a higher probability of failing to detect a true effect, leading to a Type II error (false negative).

Ignoring or not including power calculations in statistical analyses can have several potential consequences. Firstly, it can lead to underpowered studies, where the sample size is insufficient to detect the effect of interest. This can result in low statistical power, making it difficult to draw reliable conclusions from the study and potentially leading to false negative results.

Another consequence is that underpowered studies may waste resources, both in terms of time and money, as well as the efforts of participants and researchers. These studies may produce inconclusive or misleading results, hindering the progress of scientific knowledge and potentially leading to misguided decisions in various fields.

Additionally, if power calculations are not considered, researchers may underestimate the sample size needed to achieve adequate power. This can result in studies that are underpowered, unable to detect meaningful effects, and thus lacking scientific rigor and validity.

Overall, including power calculations in statistical analyses is crucial for designing studies with appropriate sample sizes, ensuring that the study has a high chance of detecting meaningful effects if they exist. It helps researchers make informed decisions about study design, improve the reliability of their findings, and avoid wasted resources and potential misinterpretation of results.

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.

Multiple post-hoc tests exist because different tests have different underlying assumptions and statistical properties. Researchers have developed various post-hoc tests to address specific research questions, account for different experimental designs, and control for Type I error rates appropriately.

When choosing between Tukey's Method and the Bonferroni correction, the decision depends on the specific research context and the desired balance between Type I and Type II errors:

1. Tukey's Method (Tukey's HSD): Tukey's Method is a post-hoc test commonly used for pairwise comparisons in ANOVA. It controls the family-wise error rate (FWER), which is the probability of making at least one Type I error across all pairwise comparisons. Tukey's Method is suitable when the primary goal is to identify significant pairwise differences while controlling the overall Type I error rate. It is particularly effective when there are many pairwise comparisons to be made.

2. Bonferroni Correction: The Bonferroni correction is a conservative method that adjusts the significance threshold for individual comparisons to control the family-wise error rate. It divides the desired overall significance level (e.g., 0.05) by the number of comparisons being made. The Bonferroni correction is suitable when the focus is on controlling the Type I error rate strictly and making individual comparisons more conservative. However, it can result in reduced power and an increased likelihood of Type II errors, especially when there are many comparisons.

Other common post-hoc tests include:

- Dunnett's Test: Dunnett's test is used when comparing multiple treatments or groups to a control group. It adjusts for multiple comparisons by using a single control group as a reference.

- Scheffé's Method: Scheffé's method is a conservative post-hoc test that controls the family-wise error rate for all possible comparisons. It is more powerful than the Bonferroni correction but tends to be more conservative than Tukey's Method.

- Fisher's Least Significant Difference (LSD): The LSD test is a simple post-hoc test that compares means in pairs, similar to Tukey's Method. It does not control the family-wise error

rate but can be used as an exploratory analysis when the number of pairwise comparisons is relatively small.

The choice of post-hoc test depends on the specific research question, study design, desired control of Type I error rate, and the number of pairwise comparisons to be made. Researchers should consider the assumptions and properties of each test and select the one that best suits their needs.