Lab 03B: Week 9

Contents

The **specific aims** of this lab are:

- practice performing (one-way) ANOVA and interpreting the output
- check ANOVA assumptions using residuals
- conduct post-hoc testing using contrasts
- perform rank based testing of multiple means (Kruskal-Wallis test)

The unit **learning outcomes** addressed are:

- LO1 Formulate domain/context specific questions and identify appropriate statistical analysis.
- LO3 Construct, interpret and compare numerical and graphical summaries of different data types including large and/or complex data sets.
- LO6 Formulate, evaluate and interpret appropriate linear models to describe the relationships between multiple factors.
- LO8 Create a reproducible report to communicate outcomes using a programming language.

1 Recap: multiple comparisons

1.1 Simultaneous confidence intervals

A 95% confidence interval for a *single* population contrast $\sum_{i=1}^{g} c_i \mu_i$ (where $\sum_{i=1}^{g} c_i = 0$) is of the form

$$\underline{g} \qquad \left(\quad \overline{\underline{g} \quad c^2} \right)$$

$$\sum_{i=1} c_i \bar{y}_{i\cdot} \pm m \left(\hat{\sigma} \sqrt{\sum_{i=1} \frac{c_i}{n_i}} \right)$$

where the **multiplier** m is the upper 2.5% quantile from the t_{N-g} distribution (recall N is the total sample size); the quantity in round brackets is the *standard error*. When the model is correct this procedure "works" 95% of the time in repeated experiments.

However if we are constructing several of these at once, while each one individually may work 95% of the time, having *all of them* "work" (simultaneously) is not guaranteed to the same degree. We can fix this by *increasing the multiplier* m. We have discussed 3 different approaches:

1.1.1 The Bonferroni method

If we are constructing k simultaneous $100(1-\alpha)\%$ confidence intervals, instead of using the upper $\alpha/2$ quantile, we use the upper $\alpha/(2k)$ quantile, i.e. as if we were constructing individual $100(1-\alpha/k)\%$ intervals. This procedure is in general *conservative* i.e. the resultant true confidence level is typically greater than desired (i.e. the multiplier is bigger than it needs to be).

1.1.2 Tukey's method

This provides the *exact* multiplier one needs when

- we are looking at all pairwise comparisons and
- the sample sizes are all the same.

When these two conditions hold, it is the best we can do i.e. it gives the smallest multiplier m that does the job. When the sample sizes are *unequal* it is conservative (although possibly less so than the corresponding Bonferroni multiplier.)

1.1.3 Scheffé's method

This provides the *exact* multiplier one needs when considering *all possible contrasts*, and thus permits "unlimited data snooping". The multiplier is taken from the $\sqrt{(g-1)F}$ distribution, where here F denotes the distribution of the corresponding F-statistic i.e. $F_{g-1,N-g}$. This multiplier is thus conservative when considering only a finite number of contrasts, but again may be smaller than the corresponding Bonferroni multiplier.

1.2 Implementation

The Bonferroni correction is straightforward to implement. But in general, the **emmeans** package provides a convenient way to implement a variety of post hoc tests. The generic code below shows how this might be done for a hypothetical dependent variable `y` and factor variable `group`.

```r
# this code won't actually run, we haven't defined y or group
library(emmeans)
one_way = aov(y ~ group)
one_way_em = emmeans(one_way, ~ group)
one_way_pairs = contrast(one_way_em, method = "pairwise", adjust = "bonferroni")
# alternatively, can use pairs()
# one_way_pairs = pairs(one_way_em, adjust = "bonferroni")
plot(one_way_pairs)
# Tukey's method
contrast(one_way_em, method = "pairwise", adjust = "tukey")
# Scheffe's method:
contrast(one_way_em, method = "pairwise", adjust = "scheffe")
```

2 Questions

2.1 Pain thresholds

Recall the pain/hair colour data. Below we change the factor order from alphabetical to "lightest to darkest":

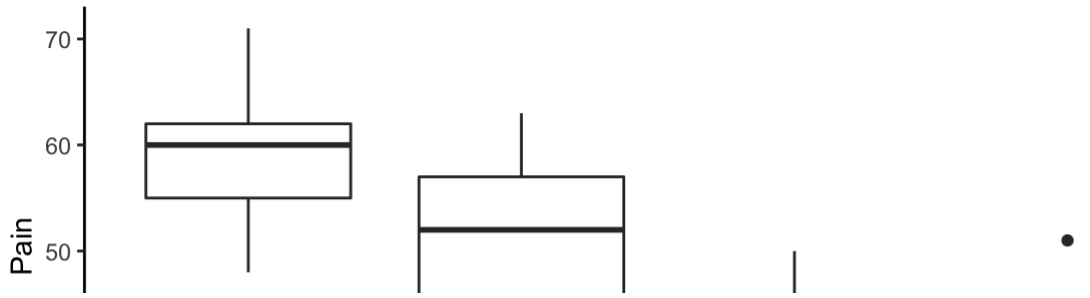
```r
library(tidyverse)
pain = read_tsv("https://raw.githubusercontent.com/DATA2002/data/master/blonds.txt")
glimpse(pain)
```

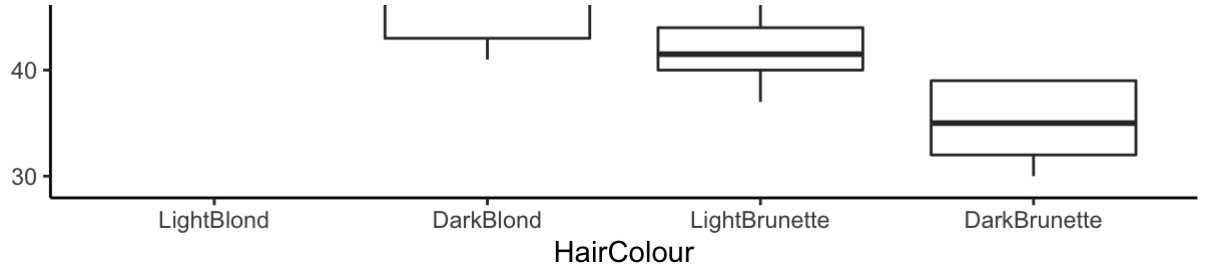
```
Rows: 19
Columns: 2
$ HairColour <chr> "LightBlond", "LightBlond", "LightBlond", "LightB…
$ Pain       <dbl> 62, 60, 71, 55, 48, 63, 57, 52, 41, 43, 42, 50, 4…
```

```r
pain = pain %>%
    mutate(HairColour = factor(HairColour, levels = c("LightBlond", "DarkBlond",
        "LightBrunette", "DarkBrunette")))
levels(pain$HairColour)
```

```
[1] "LightBlond"    "DarkBlond"     "LightBrunette" "DarkBrunette"
```

```r
ggplot(pain, aes(x = HairColour, y = Pain)) + geom_boxplot() + theme_classic()
```






```
pain_sum = pain %>%
    group_by(HairColour) %>%
    summarise(n = n(), ybar = mean(Pain))
pain_sum
```

```
# A tibble: 4 × 3
  HairColour        n  ybar
  <fct>         <int> <dbl>
1 LightBlond        5  59.2
2 DarkBlond         5  51.2
3 LightBrunette     4  42.5
4 DarkBrunette      5  37.4
```

```
ni = pain_sum %>%
    pull(n)
ybar_i = pain_sum %>%
    pull(ybar)
```

```
pain_aov = aov(Pain ~ HairColour, data = pain)
summary(pain_aov)
```

```
            Df Sum Sq Mean Sq F value  Pr(>F)
HairColour   3   1361   453.6   6.791 0.00411 **
Residuals   15   1002    66.8
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

1. Compute the standard error of each pairwise difference (note, there are only two different standard errors over the $\binom{4}{2} = 6$ pairwise differences).

2. Compute t-statistics for all 6 pairwise comparisons. The output below may be useful:

```
diff_mat = outer(ybar_i, ybar_i, "-")
diff_mat
```

```
      [,1]  [,2] [,3] [,4]
[1,]   0.0   8.0 16.7 21.8
[2,]  -8.0   0.0  8.7 13.8
[3,] -16.7  -8.7  0.0  5.1
[4,] -21.8 -13.8 -5.1  0.0
```

3. Using the output below and the Bonferroni method, determine the appropriate multiplier for constructing 6 simultaneous confidence intervals at both the 95% and 99% confidence levels.

```
upper.tail.area = c(0.05, 0.025, 0.05/6, 0.025/6, 0.01, 0.005, 0.01/6,
    0.005/6)
t.quantile = qt(1 - upper.tail.area, df = 15)
cbind(upper.tail.area, t.quantile)
```

```
     upper.tail.area t.quantile
[1,]    0.0500000000   1.753050
[2,]    0.0250000000   2.131450
[3,]    0.0083333333   2.693739
[4,]    0.0041666667   3.036283
[5,]    0.0100000000   2.602480
[6,]    0.0050000000   2.946713
[7,]    0.0016666667   3.483677
[8,]    0.0008333333   3.821973
```

4. Which differences are significant at the

- 5% level
- 1% level

5. Check your answers using the **emmeans** package. Do your conclusions change if you use Tukey's or Scheffe's method?

2.2 Tablet

This data contains the level of chlorpheniramine maleate in tablets from seven labs (Rice 1995, 443–44). The purpose of the experiment was to study the consistency between labs. For each of four manufacturers, composites were prepared by grinding and mixing together tablets in order to measure the amount of chlorpheniramine maleate. Seven labs were asked to make 10 determinations on each composite (Kirchhoefer 1979). The data for the 7 labs are provided in the file `tablet1.txt`.

We start by reading in the data and use the `pivot_longer()` function from the **tidyr** package to reshape the data from *wide* to *long* format.

```
library(tidyverse)
tablet = read_tsv("https://raw.githubusercontent.com/DATA2002/data/master/tablet1.txt")
glimpse(tablet)
```

```
Rows: 10
Columns: 7
$ Lab1 <dbl> 4.13, 4.07, 4.04, 4.07, 4.05, 4.04, 4.02, 4.06, 4.10, 4…
$ Lab2 <dbl> 3.86, 3.85, 4.08, 4.11, 4.08, 4.01, 4.02, 4.04, 3.97, 3…
$ Lab3 <dbl> 4.00, 4.02, 4.01, 4.01, 4.04, 3.99, 4.03, 3.97, 3.98, 3…
$ Lab4 <dbl> 3.88, 3.88, 3.91, 3.95, 3.92, 3.97, 3.92, 3.90, 3.97, 3…
```

```
$ Lab5 <dbl> 4.02, 3.95, 4.02, 3.89, 3.91, 4.01, 3.89, 3.89, 3.99, 4…
$ Lab6 <dbl> 4.02, 3.86, 3.96, 3.97, 4.00, 3.82, 3.98, 3.99, 4.02, 3…
$ Lab7 <dbl> 4.00, 4.02, 4.03, 4.04, 4.10, 3.81, 3.91, 3.96, 4.05, 4…
```

```
tabdat = tablet %>%
    pivot_longer(cols = everything(), names_to = "lab", values_to = "measurement")
glimpse(tabdat)
```

```
Rows: 70
Columns: 2
$ lab         <chr> "Lab1", "Lab2", "Lab3", "Lab4", "Lab5", "Lab6", …
$ measurement <dbl> 4.13, 3.86, 4.00, 3.88, 4.02, 4.02, 4.00, 4.07, …
```

Produce some basic summary statistics and generate side by side box plots.

```
tabdat %>%
  ggplot() +
  aes(x = lab, y = measurement, fill = lab) +
  geom_boxplot() +
  theme_classic() +
  labs(y = "Chlorpheniramine maleate (mg)",
       x = "Lab", fill = "")
```

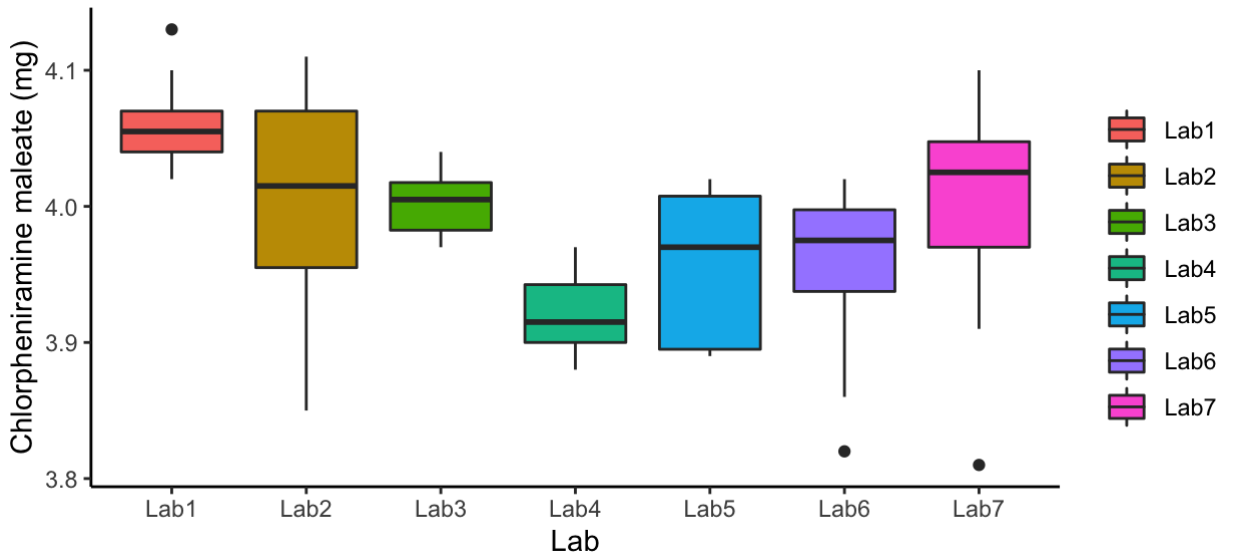


Figure 1: Figure 1: Boxplots of determinations of amounts of chlorpheniramine maleate in tablets by seven laboratories.

The boxplots in Figure 1 show some differences in the medians. Are these differences significant? We can address this by asking a variety of questions.

1. Is the mean level of chlorpheniramine maleate in tablets from Lab 1 different from 4.0? State the null hypothesis.

2. Is the mean level of chlorpheniramine maleate in tablets from Lab 1 different from that from Lab 3?

3. Perform a one-way ANOVA to test if the mean levels of chlorpheniramine maleate differ across the seven labs.
4. Obtain a QQ plot of the residuals and comment on the validity of the ANOVA assumptions.
5. Perform pairwise post hoc tests to determine which labs are significantly different.
6. Use a **rank** based approach to testing for a difference between the means of the 7 labs.
7. Use a **permutation** based approach to testing for a difference between the means of the 7 labs.

3 For practice after the computer lab

- Use a rank based approach to testing for a difference between the means for the pain threshold data.
- In Larsen and Marx (2012) work through Case Study 12.3.1 and Case Study 12.4.1 and then consider questions 12.3.3 and 12.4.4.
- You can also attempt the DataCamp chapter on comparing many means.

References

Kirchhoefer, R D. 1979. "Semautomated Method for the Analysis of Chlorpheniramine Maleate Tablets: Collaborative Study." *Journal - Association of Official Analytical Chemists* 62 (6): 1197–1201.

Larsen, Richard J., and Morris L. Marx. 2012. *An Introduction to Mathematical Statistics and Its Applications*. 5th ed. Boston, MA: Prentice Hall.

Rice, John A. 1995. *Mathematical Statistics and Data Analysis*. Belmont, CA: Duxbury Press.