

Instructions

For the following problem, an exercise from the textbook (OpenIntro Statistics) has been worked, with any requisite R (or RStudio) commands provided. Additional questions and/or steps have been added for a more complete introduction to one-way ANOVA. Use it as a reference when repeating the analysis for the study identified under *On Your Own*. The inputs that will require new values are highlighted in yellow. All input will be in the RStudio console, while output will be generated in both the console and plots windows of RStudio. It is recommended that section 5.5 on pp. 246 - 256 be read, or at least reviewed, prior to completing this lab assignment.

1. *Exercise 5.43 on page 269*

In Exercises 5.31 and 5.33 we compared the effects of two types of feed at a time. In 5.31 it was linseed vs. horsebean, and in 5.33 it was casein vs. soybean.

*** Description of experiment from Exercise 5.31 on page 266 ***

Chicken farming is a multi-billion dollar industry, and any methods that increase the growth rate of young chicks can reduce consumer costs while increasing company profits, possibly by millions of dollars. An experiment was conducted to measure and compare the effectiveness of various feed supplements on the growth rate of chickens. Newly hatched chicks were randomly allocated into six groups, and each group was given a different feed supplement.

A two-sample t-test was performed in in both exercises to determine if there was a significant difference in the mean growth rate of the chickens between the two feed types. As discussed on page 255, if we consider all possible pairwise comparisons of the ($k = 6$) five feed types, we would have to perform a total of $K = 15$ comparisons, which would have increased the Type 1 Error rate from 5%, i.e., $\alpha = 0.05$, for a single two-sample t-test, to a rather unacceptable 53.67%. The respective calculations follow below.

$$K = \frac{k(k-1)}{2} = \frac{6(6-1)}{2} = 15$$

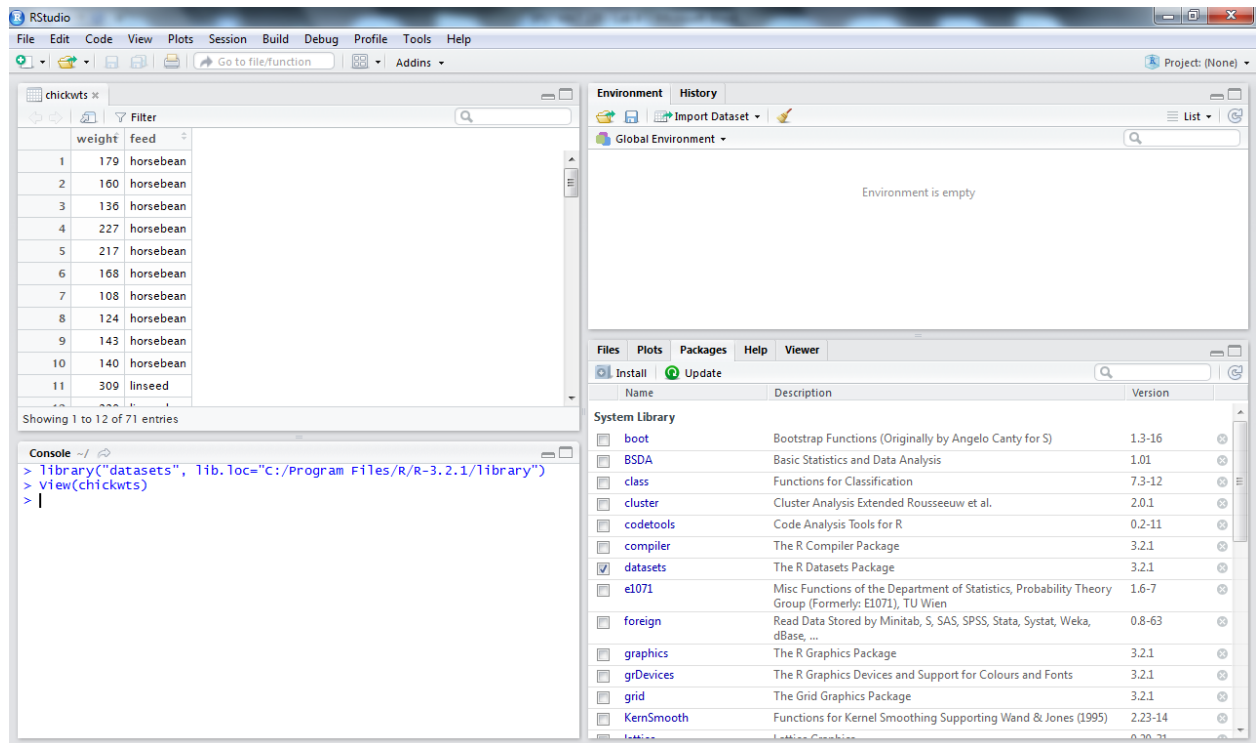
$$P(\text{at least 1 Type 1 Error}) = 1 - (1 - \alpha)^K = 1 - (1 - 0.05)^{15} = 0.5367088$$

A better analysis would first consider all feed types at once: casein, horsebean, linseed, meat meal, soybean, and sunflower.

In Exercise 5.43, we are asked to perform a one-way Analysis of Variance (ANOVA) to determine if these data provide convincing evidence that the average weight of chicks varies across some (or all) groups. Generally, this includes verifying relevant conditions are met, conducting an overall F-test to determine if the means differ across two or more groups, and (if appropriate) applying an appropriate multiple comparisons procedure to identify which means differ and by how much, while controlling the Type 1 Error rate.

Note: A boxplot and the corresponding summary statistics for growth rate across the five feed types are provided in the exercise on page 269, however, since this dataset is actually stored in the standard datasets package for R, we will use it instead in order to perform a more complete and efficient analysis using the commands available in RStudio.

Before we step through the process of performing a one-way ANOVA, let's first load the dataset into RStudio. This can be done by clicking on the **Packages** tab in the lower right window of RStudio and then selecting the checkbox labeled **datasets**. Upon doing so, you will see the `library` command automatically entered into the console on the lower left. To view the contents of the data set, called `chickwts`, we will use the `View` command and enter `View(chickwts)` into the console (and press enter). A spreadsheet view of the dataset should appear in the upper left window. I have included a screenshot of RStudio for reference.



We could have also just entered the aforementioned commands (also listed below) directly into the console to load the package and view the dataset.

```
# load the package
library(datasets)
```

```
# open the spreadsheet-style data viewer
View(chickwts)
```

(a) *State hypotheses.*

$$H_0: \mu_{\text{casein}} = \mu_{\text{horsebean}} = \mu_{\text{linseed}} = \mu_{\text{meatmeal}} = \mu_{\text{soybean}} = \mu_{\text{sunflower}}$$

$$H_A: \text{the mean weight of chickens varies across some (or all) feed types}$$

where μ_i is the mean weight of chickens whose feed contained the supplement i , which can take one of six values: casein, horsebean, linseed, meatmeal, soybean or sunflower.

(b) *Check relevant conditions.*

As stated on page 252, under section 5.5.4, “there are three conditions we must check for an ANOVA analysis: all observations must be independent, the data in each group must be nearly normal, and the variance within each group must be approximately equal.”

Independence. If the data are a simple random sample from less than 10% of the population, this condition is satisfied. For processes and experiments, carefully consider whether the data may be independent (e.g. no pairing).

Explanation: In the problem statement, it states “an experiment was conducted ...newly hatched chicks were randomly allocated into six groups, and each group was given a different feed supplement.” Since an experiment was performed whereby the experimental units (newly hatched chicks) were randomly assigned to a particular treatment (feed type), we can conclude the observed weights are independent; this condition is met. *Aside:* this experimental design is commonly referred to as a *completely randomized design*.

Approximately normal. As with one- and two-sample testing for means, the normality assumption is especially important when the sample size is quite small. The normal probability plots for each group...are shown. Sometimes in ANOVA there are so many groups or so few observations per group that checking normality for each group isn't reasonable. See the footnote³³ (on page 252) for guidance on how to handle such instances.³³ First calculate the residuals by taking the observed values and subtracting the corresponding group means. Then to check the normality condition, create a normal probability plot using all the residuals simultaneously.

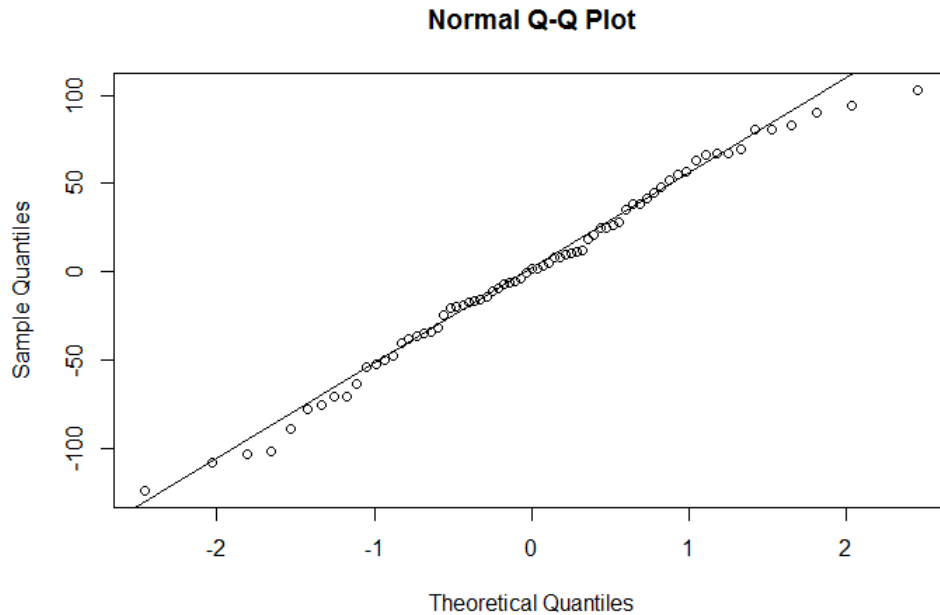
Explanation: If we use the `table` command on the `feed` column in the `chickwts` dataset, we see that the number of observations in each of the five groups fluctuate between 10 and 14 observations.

```
table(chickwts$feed)
casein horsebean linseed meatmeal soybean sunflower
      12         10         12         11         14         12
```

Given these small to moderate (sample) sizes and the fact we have 6 groups (as opposed to say 3 or 4), we will be efficient and check this condition using the a single normal probability plot of the residuals per the footnote. To generate such a plot in RStudio we will actually need to run the analysis first, using the `aov` command, and then construct the plot using the `qqnorm` and `qqline` commands. Enter the following commands into the console in RStudio to generate the subsequent normal probability plot shown.

```
fit <- aov(weight ~ feed, data=chickwts)
qqnorm(fit$residuals)
qqline(fit$residuals)
```

In regards to the `aov` command, notice the column name (`weight`) for the response variable is listed first, followed by the column name (`feed`) for the explanatory (or group) variable, and the `data=` input specifies the name of the dataset (`chickwts`).

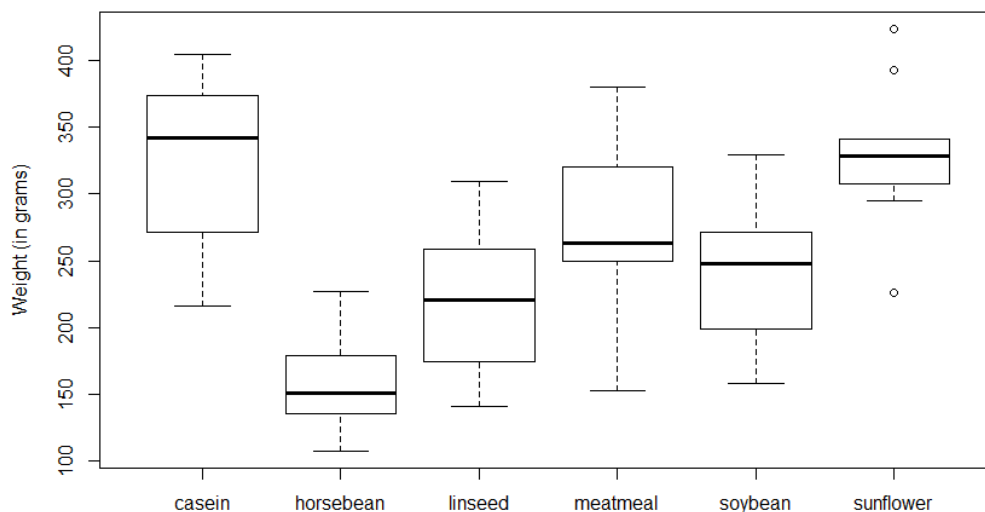


Recall from page 137, where the normal probability (or quantile-quantile) plot is first introduced, “the closer the points are to a perfect straight line, the more confident we can be that the data follow the normal model.” Inspecting the above plot, we find that the assumption of normality appears to be satisfied as there is only slight deviation from the straight line; really only a few points on the upper right and possibly one on the lower left, but they are not particularly extreme.

Constant variance. The last assumption is that the variance in the groups is about equal from one group to the next. This assumption can be checked by examining a side-by-side box plot of the outcomes across the groups.

Explanation: To [re]produce the boxplot of observed weights across each type of feed supplement, copy/paste the following `boxplot` command in the RStudio console.

```
boxplot(chickwts$weight ~ chickwts$feed, ylab="Weight (in grams)")
```



A quick inspection of the boxplots for the observed weights of newly hatched chicks for each feed supplement - in particular the range (whisker span) and IQR (box span) - suggests some inconsistency in the variation across groups.

For some, such a subjective approach to assessing homoscedasticity (or homogeneity of variance) - technical term(s) for the constant variance assumption - is not all that desirable. Fortunately, there are other methods. A common rule-of-thumb used to determine if the variance across groups is relatively constant [enough] to see if the ratio of extreme values for the observed sample standard deviations does not exceed two. In other words, $\max(SD)/\min(SD) \leq 2$; the largest standard deviation is no more than double the smallest.

While this particular exercise provides a table of summary statistics (mean, standard deviation and sample/group size) for this experiment, entering (copy/paste) the following commands into the RStudio console will generate the same table. For convenience, we defined a custom, reusable function called `summary.by` to perform this task.

```
# generate summary statistics by group
summary.by <- function(x, group)
{
  data.frame(
    Mean = tapply(x, group, FUN=mean),
    SD    = tapply(x, group, FUN=sd),
    n     = tapply(x, group, FUN=length)
  )
}

summary.by(chickwts$weight, chickwts$feed)
```

the resulting output follows below

	Mean	SD	n
casein	323.5833	64.43384	12
horsebean	160.2000	38.62584	10
linseed	218.7500	52.23570	12
meatmeal	276.9091	64.90062	11
soybean	246.4286	54.12907	14
sunflower	328.9167	48.83638	12

We see that the largest standard deviation (64.90062) was observed in the **meatmeal** group, while the smallest (38.62584) occurred in the **horsebean** group. Since their ratio ($64.90062/38.62584 = 1.680238$) is less than two, we can conclude the constant variance assumption is reasonably satisfied.

(c) *Perform the overall F-test.*

Similar to the one- and two-sample *t*-tests studied in Lab 2 and 3, an appropriate test statistic and corresponding P-value are calculated to determine if there is sufficient evidence to support the alternative hypothesis; i.e., is the observed difference in mean weight of the chickens by feed type likely not due to random chance?

As discussed in more detail in section 5.5.2 on pages 250 - 251, this test statistic - called the F statistic - is the ratio of the estimated between group variability to the variability within each of the groups. The value of the F statistic and its P-value are typically given in the upper right of the [statistical software-generated] ANOVA table. Again, while this particular exercise provides this table in the problem description found on page 269, we can [re]produce this table using the `anova` command on the `fit` object defined earlier, which contains the results of the analysis of variance and was initially created to compute the residuals for verifying the normality condition. Type or copy/paste the single command below to obtain the output that follows.

COMMAND: `anova(fit)`

OUTPUT:

Analysis of Variance Table

Response: weight

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
feed	5	231129	46226	15.365	5.936e-10 ***
Residuals	65	195556	3009		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

No significance level was provided, therefore we shall assume $\alpha = 0.05$.

Decision: The p-value (5.936×10^{-10}) $< \alpha$ (0.05), reject the null hypothesis, H_0 .

Conclusion: At the 5% significance level, there is statistically significant evidence the mean weight of chickens varies across some (or all) feed types.

(d) *Conduct pairwise tests to determine which groups are different from each other.*

Having rejected the null hypothesis for the overall F test in the ANOVA analysis, we can now proceed to determine which feed supplements resulted in chickens with statistically different mean weights. Section 5.5.5 on page 253 proposes using the familiar two-sample t -test for each of the $K = 15$ possible pairwise comparisons (of group means) using the pooled estimate of the standard deviations across groups, which is $\sqrt{MSE}$, and a *Bonferroni* corrected significance level, α^* , in order to control the Type 1 Error rate. Per the discussion on page 255, the Bonferroni correction suggests using

$$\alpha^* = \frac{\alpha}{K} = \frac{0.05}{15} = 0.00\bar{3}$$

While the Bonferroni procedure is an effective (and powerful) all around tool for controlling the Type 1 error rate across multiple tests, another procedure called [Tukey's Honest Significant Difference](#), or Tukey HSD, performs slightly better when all possible pairwise comparisons of [group] means are being considered. We will use the available commands (functions) in R to apply both procedures - Bonferroni first, followed by Tukey HSD, but refer to the results of the Tukey method to continue with our [post-hoc analysis](#).

Bonferroni

We can create a table (matrix) of [Bonferroni corrected] p-values for each of the 15 two-sample *t*-tests to compare the mean weights across all 6 feed types using the `pairwise.t.test` command in RStudio. The first input, `x`, identifies the column containing the observed values for the response variable (**weight**), while the second input, `g`, identifies the column containing the values for the explanatory (group) variable (**feed**). The third input, `p.adjust.method`, specifies the correction method. Type or copy/paste the single command below to obtain the output that follows. Due to the large number of characters in the command, it is written across several lines for easier reading.

COMMAND:

```
pairwise.t.test(
  x = chickwts$weight,
  g = chickwts$feed,
  p.adjust.method = "bonferroni"
)
```

OUTPUT:

Pairwise comparisons using t tests with pooled SD

data: chickwts\$weight and chickwts\$feed

	casein	horsebean	linseed	meatmeal	soybean
horsebean	3.1e-08	-	-	-	-
linseed	0.00022	0.22833	-	-	-
meatmeal	0.68350	0.00011	0.20218	-	-
soybean	0.00998	0.00487	1.00000	1.00000	-
sunflower	1.00000	1.2e-08	9.3e-05	0.39653	0.00447

P value adjustment method: bonferroni

If we were to use this output to determine which group means differed (significantly), we would identify all p-values in the above table that were less than the familywise Type 1 Error rate (significance level) of 0.05 or whatever was chosen by the researcher prior to the experiment/study. For convenience, I have highlighted them in gray. The respective row and column name identifies which feed types are being compared. For example, the p-value of 0.00447 in the lower right of the table, which is less than 0.05, indicates there is sufficient (strong enough) evidence to conclude the mean weight of chickens fed with soybean supplement differs from that of chickens fed with sunflower supplement; i.e., $\mu_{\text{soybean}} \neq \mu_{\text{sunflower}}$. It is important to note that this approach is slightly different than the method discussed on page 255, which compares the *unadjusted* P-value from each two-sample *t*-test to the *modified* significance level, $\alpha^* = 0.00\bar{3}$. If we were to apply the method put in practice in Example 5.49, we find the *T* score for testing μ_{soybean} vs $\mu_{\text{sunflower}}$ to be

$$T = \frac{\bar{x}_{\text{soybean}} - \bar{x}_{\text{sunflower}}}{\sqrt{\frac{MSE}{n_{\text{soybean}}} + \frac{MSE}{n_{\text{sunflower}}}}} = \frac{246.4286 - 328.9167}{\sqrt{\frac{3009}{14} + \frac{3009}{12}}} \approx -3.822506$$

The associated p-value on $df = 65$, which is found in the ANOVA table along the row labeled `Residuals`, can then be calculated using the `pt` command in RStudio, which yields the following

```
> 2*pt(-3.822506, df=65)
[1] 0.0002983229
```

Since this p-value (0.0002983229) is less than the modified significance level, $\alpha^* = 0.003$, we reject the null hypothesis. In fact, since the Bonferroni correction divided the family-wise significance level, $\alpha = 0.05$, by the number of comparisons, $K = 15$, if we instead multiply this p-value by K , we obtain the corresponding *adjusted* P-value in the table generated by the `pairwise.t.test` command; that is $0.0002983229 \times 15 = 0.004474843 \approx 0.00447$.

Tukey

To generate a list of *adjusted* p-values for each of the 15 pairwise comparisons of group means using the [studentized range](#) distribution proposed by Tukey, we use the `TukeyHSD` command in RStudio on the `fit` object, which contains the results of the ANOVA analysis (or fitted model) performed earlier. The command requires a minimum of two inputs; the first is the aforementioned `fit` object containing the output of the `aov` command. The second input is the [column] name of the explanatory (group) variable; in this case it's "feed". Another useful input is `conf.level`, which allows the user to specify the family-wise confidence level for the procedure. By default, it is 0.95, which corresponds to a Type 1 Error rate, α , of 0.05. Enter (copy/paste) the following commands into the console in RStudio to generate the subsequent output.

COMMANDS:

```
tukey <- TukeyHSD(fit, "feed", conf.level=0.95)
print(tukey)
```

OUTPUT:

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

Fit: aov(formula = weight ~ feed, data = chickwts)

$feed
              diff      lwr      upr    p adj
horsebean-casein -163.383333 -232.346876 -94.41979 0.0000000
linseed-casein   -104.833333 -170.587491 -39.07918 0.0002100
meatmeal-casein  -46.674242  -113.906207  20.55772 0.3324584
soybean-casein   -77.154762 -140.517054 -13.79247 0.0083653
sunflower-casein   5.333333  -60.420825  71.08749 0.9998902
linseed-horsebean  58.550000  -10.413543 127.51354 0.1413329
meatmeal-horsebean 116.709091   46.335105 187.08308 0.0001062
soybean-horsebean  86.228571   19.541684 152.91546 0.0042167
sunflower-horsebean 168.716667   99.753124 237.68021 0.0000000
meatmeal-linseed   58.159091   -9.072873 125.39106 0.1276965
soybean-linseed    27.678571  -35.683721  91.04086 0.7932853
sunflower-linseed  110.166667   44.412509 175.92082 0.0000884
soybean-meatmeal   -30.480519  -95.375109  34.41407 0.7391356
sunflower-meatmeal  52.007576  -15.224388 119.23954 0.2206962
sunflower-soybean  82.488095   19.125803 145.85039 0.0038845
```

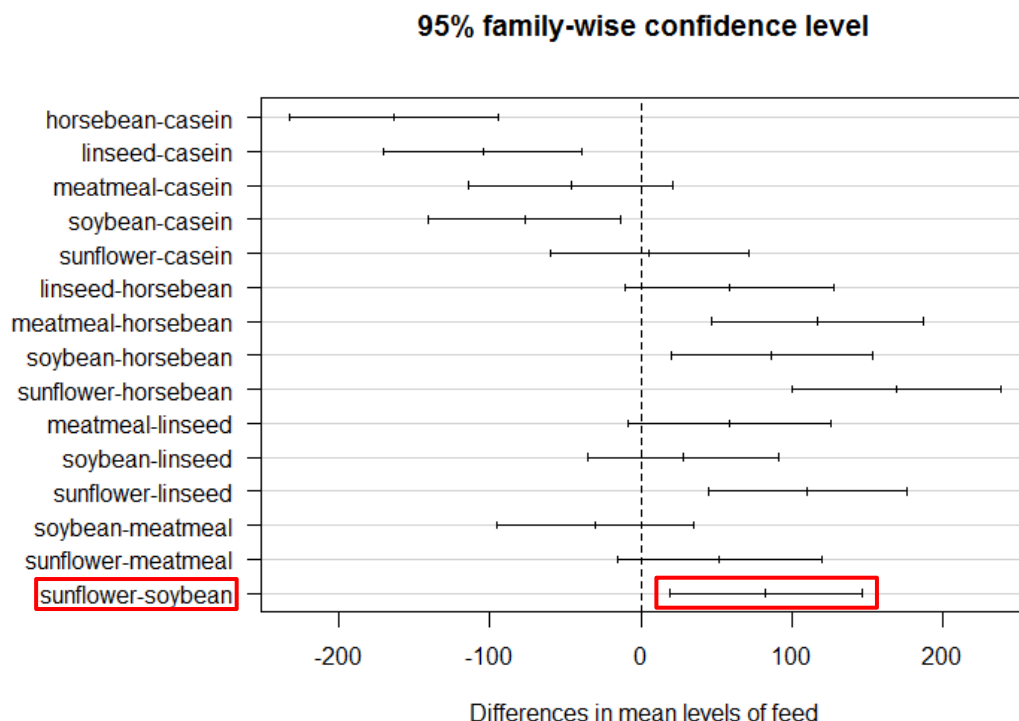
In a manner similar to the one described for the Bonferroni procedure earlier, to determine which group means differed (significantly), we identify all p-values in the far right column (p_{adj}) of the above table that are less than the familywise Type 1 Error rate (significance level) of 0.05. For convenience, I have highlighted them in gray. The corresponding entry in the first column indicates which two group means are being compared, as well as the definition of the difference; e.g., $\mu_A - \mu_B$ or $\mu_B - \mu_A$. If we compare the two procedures, we find that both approaches identified the same eight (8) pairwise mean comparisons that have a statistically significant difference between the respective observed sample mean weights.

An additional benefit of the TukeyHSD command is the computation of the point (difference in sample means) and simultaneous confidence interval estimates for each of the 15 pairwise mean differences. This information can also be displayed graphically using the `par` and `plot` commands. The `mar=` input in the `par` command sets the *bottom*, *left*, *top* and *right* margins, respectively, of the plot region in terms of number of lines of text, while the `las=` input changes the orientation of the y-axis labels to horizontal. This step is needed to make sure the y-axis label for each pairwise difference is readable on the graph. Notice the `plot` command accepts the results of the TukeyHSD command - the `tukey` object - as its input. Enter (copy/paste) the following commands into the console in RStudio to generate the subsequent plot.

COMMANDS:

```
par(mar=c(4,9,4,4), las=1)
plot(tukey)
```

OUTPUT:



Recall a [two-sided] confidence interval indicates there is statistically significant evidence (at $\alpha = 1 - C$) of a difference between two means if the hypothesized difference of zero (0) falls outside the limits of the interval. Therefore any interval (horizontal line with three tick marks) in the plot that does not include the vertical, dashed line at 0 would indicate a statistically significant difference. If the interval lies completely on the right-side of the zero line, then the group mean listed first in the y-axis label is larger, whereas if the interval lies completely on the left-side of the zero line, then the group mean listed second in the y-axis label is larger.

Returning to the μ_{soybean} vs. $\mu_{\text{sunflower}}$ comparison cited earlier in the discussion of the Bonferroni procedure, we see that its corresponding interval in the plot (bottom row) is to the right of the zero line; again, suggesting there is sufficient evidence of a difference in the mean weight of chickens for the two feed types. Further, we can use the confidence interval limits for this pairwise comparison provided in the table output of the TukeyHSD command to quantify said difference; specifically, we are 95% confident newly hatched chickens whose feed is supplemented with sunflower are between 19 grams and 146 grams heavier (on average), over the same time period, than those whose feed is supplemented with soybean. Notice the rather large margin of error for this interval estimate (about 63.5 grams), which is an artifact of the fairly small group (sample) sizes - 12 and 14, the confidence level - 95%, moderate number of comparisons made - 15, and inherent variation in the observed weights of the newly hatched chickens in each group. The first three we have control over, and as will be evident in the discussion that follows, an increase in the group (sample) sizes is recommended to increase the precision of the estimates (and/or reduce the effect size for the tests), should the experiment be repeated. This method of interpretation can be applied to the remaining seven (7) statistically significant pairwise mean comparisons and is left to the student to complete on their own.

Once the statistically significant pairwise differences of mean weights are identified, it is often desirable to succinctly summarize the results of the analysis. This can be communicated both symbolically and literally.

Symbolically

Using the table of summary statistics (in particular the sample means) as a guide, list the group means in ascending (or descending order) using their respective parametric symbols. In the table below, the first row is the respective sample mean weight for each feed type. Although not required, it [list of sample means], along with the boxplot, can be useful when attempting to summarize the results of the analysis and/or draw appropriate conclusions from the experiment/study.

160.20	218.75	246.43	276.91	323.58	328.92
$\mu_{\text{horsebean}}$	μ_{linseed}	μ_{soybean}	μ_{meatmeal}	μ_{casein}	$\mu_{\text{sunflower}}$

Referencing either the printed output of confidence interval limits and/or adjusted p-values for each of the pairwise comparisons of group means from the TukeyHSD command, or the plot of simultaneous confidence intervals, we underline the parameters in the above table that do **NOT** have sufficient evidence of a statistically significant difference. Depending on the number of comparisons, or number of groups in the experiment/study, this may take some time.

Ideally, no parameter would be underlined more than once. However, when such an event happens - twice in the above table - we know a Type 2 Error has occurred. For instance, notice that μ_{meatmeal} is included with μ_{linseed} and μ_{soybean} on one line, but is also included with μ_{casein} and $\mu_{\text{sunflower}}$ on another. On one hand, this suggests that all five mean weights are not significantly different, i.e., $\mu_{\text{soybean}} = \mu_{\text{linseed}} = \mu_{\text{meatmeal}} = \mu_{\text{casein}} = \mu_{\text{sunflower}}$. However, both μ_{casein} and $\mu_{\text{sunflower}}$ are not included on a line with either μ_{linseed} or μ_{soybean} , indicating a statistically significant difference in the respective means between the two sets. So (at minimum) either $\mu_{\text{linseed}} = \mu_{\text{soybean}} = \mu_{\text{meatmeal}}$ or $\mu_{\text{meatmeal}} = \mu_{\text{casein}} = \mu_{\text{sunflower}}$, since there is sufficient evidence that $\mu_{\text{casein}} = \mu_{\text{sunflower}} \neq \mu_{\text{soybean}}$ and $\mu_{\text{casein}} = \mu_{\text{sunflower}} \neq \mu_{\text{linseed}}$.

Literally

While we could methodically go through each pairwise comparison of group means and provide the standard statistical language regarding the interpretation of a confidence interval ("We are 95% confident...") or conclusions to a hypothesis test ("At the 5% family-wise significance level, there is [in] sufficient evidence to conclude..."), it is just as important to concisely communicate the [technical] results of the analysis to a more general audience. So let's return to the objective of the experiment/study and try to summarize the key takeaways and/or possible recommendations as a result of the information gained from the formal analysis.

Recall, the first sentence in the problem description for Exercise 5.31 says "*Chicken farming is a multi-billion dollar industry, and any methods that increase the growth rate of young chicks can reduce consumer costs while increasing company profits, possibly by millions of dollars.*" From the ANOVA and post-hoc analyses, it appears that feeds supplemented with either **casein** or **sunflower** will produce chickens with the highest mean weight among the types evaluated in the experiment; at least when compared to horsebean, linseed and soybean supplements. The differences in their respective mean weights are estimated to range from as little as approximately (14 g, 141 g) when compared to soybean, up to (100 g, 238 g) when compared to horsebean. We could not conclude a statistically significant difference in the mean weights of chickens among those fed with casein, sunflower and meatmeal supplements, nor those fed with meatmeal, soybean and linseed. Given these contradictory results, it is recommended that [should the experiment be repeated] larger group sizes be used and possibly drop one or more of less effective feed supplements, such as horsebean, to increase the power of the tests in the post-hoc analysis.

On Your Own

Repeat this one-way ANOVA analysis for the *Pain Thresholds of Blondes and Brunettes* study described at the following URL:

<http://www.statsci.org/data/oz/blonds.html>

To load the dataset into RStudio, you will need use the following `read.delim` command.

```
blonde <- read.delim("http://www.statsci.org/data/oz/blonds.txt")
```