Chapter Two

Ryan B. Honea

Exercise One

Question

Genetically similar seeds are randomly assigned to be raised in either a nutritionally entriched environment (treatment group) or standard conditions (control group) usig a completely randomized experimental design. After a predetermined time, all plants are harvested, dried and weighed. The results, expressed in grams, for 20 plants in each group are shown in the table below.

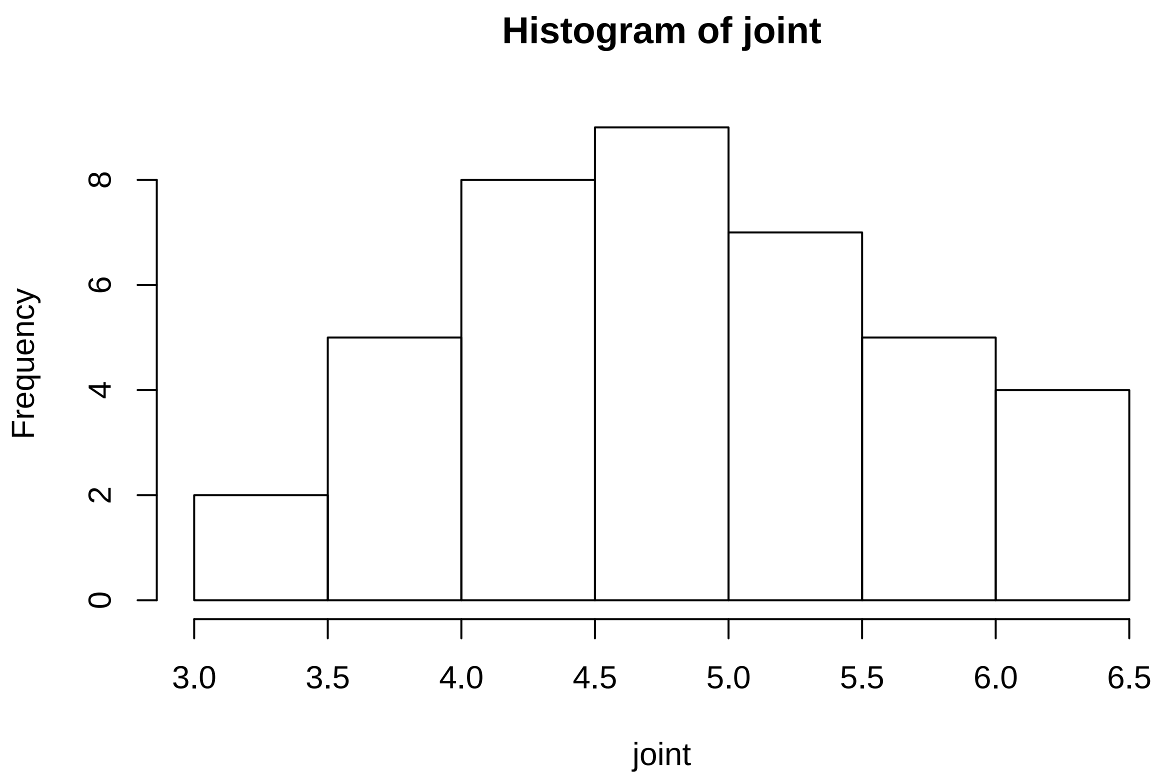
Treatment Group		Control Group	
4.81	5.36	4.17	4.66
4.17	3.48	3.05	5.58
4.41	4.69	5.18	3.66
3.59	4.44	4.01	4.50
5.87	4.89	6.11	3.90
3.83	4.71	4.10	4.61
6.03	5.48	5.17	5.62
4.98	4.32	3.57	4.53
4.90	5.15	5.33	6.05
5.75	6.34	5.59	5.14

We want to test whether there is any difference in yield between the two groups. Let Y_{jk} denote the kth observation in the jth group where $j = 1$ for the treatment group, $j = 2$ for the control group and $k = 1, ..., 20$ for both groups. Assume that the Y_{jk}'s are independent random variables with $Y_{jk} \sim \mathrm{N}(\mu_j, \sigma^2)$. The null hypothesis $H_0 : \mu_1 = \mu_2 = \mu$, that there is no differenc, is to be compared with the alternative hypothesis $H_1 : \mu_1 \neq \mu_2$.

Solution

(a): Conduct a exploratory analysis of the data looking at the distributions for each group (e.g. using dot plots, stem and leaf plots or Normal probability plots) and calculate summary statistics (e.g., means, medians, standard derivations, maxima and minima). What can you infer from these investigations?

Solution: We begin by making a histogram of the joint data from treatment and control, followed by an Anderson Darling test to assess the normality of the joint set of data.




```
##
##  Shapiro-Wilk normality test
##
## data:  joint
## W = 0.98419, p-value = 0.8387
```

With a p-value of .84, we can assume this data behaves normally (which the histogram also seems to indicate). We now find the mean and standard deviations of the treatment and control vectors.

```
## Mean of treatment:  4.8555
## Standard Deviation of treatment:  0.7904594
```

```
##
## Mean of control:  4.7265
## Standard Deviation of control:  0.8635257
```

The initial assumption is that these means are not significiantly different, but that will potentially be confirmed in later tests.

(b): Perform an unpaired t-test on these data and calculate a 95% confidence interval for the difference between the group means. Interpret these results.

Solution:

```
##
##  Welch Two Sample t-test
##
## data:  treatment and control
## t = 0.49279, df = 37.707, p-value = 0.625
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.4010672  0.6590672
## sample estimates:
```

```
## mean of x mean of y
##    4.8555    4.7265
```

The results of this t-test suggest that there is so significant statistical difference between the means. With a confidence interval that includes 0, this is especially clear.

(c): The following models can be used to test the null hypothesis H_0 against the alternative hypothesis H_1, where

$$
\begin{aligned}
H_0 &: E(Y_{jk}) = \mu; \quad Y_{jk} \sim \mathrm{N}(\mu, \sigma^2), \\
H_1 &: E(Y_{jk}) = \mu_j; \quad Y_{jk} \sim \mathrm{N}(\mu_j, \sigma^2),
\end{aligned}
$$

for $j = 1, 2$ and $k = 1, ..., 20$. Find the maximum likelihood and least squares estimate of the parameters μ, μ_1, μ_2 assuming σ^2 is a known constant.

Solution: We begin by utilizing the most-likely estimate. Note the following probability density functions.

$$
\begin{aligned}
f(\mu; Y_{jk}) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_{jk} - \mu)^2}{2\sigma^2}\right) \\
f(\mu_1; Y_{1k}) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_{jk} - \mu_1)^2}{2\sigma^2}\right) \\
f(\mu_2; Y_{2k}) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_{jk} - \mu_2)^2}{2\sigma^2}\right)
\end{aligned}
$$

Using the log-likelihood, we have the following formulas

$$
\begin{aligned}
l_0 &= \frac{1}{2} \sum_{j=1}^{2} \sum_{k=1}^{20} \left[\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(Y_{jk} - \mu)^2\right] &&= 20\log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^{2} \sum_{k=1}^{20} (Y_{jk} - \mu)^2 \\
l_1 &= \frac{1}{2} \sum_{j=1}^{2} \sum_{k=1}^{20} \left[\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(Y_{jk} - \mu_j)^2\right] &&= 20\log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \left(\sum_{k=1}^{20} (Y_{1k} - \mu_1)^2 + \sum_{k=1}^{20} (Y_{2k} - \mu_2)^2\right)
\end{aligned}
$$

When taking the derivatives in respect to μ, μ_1, and μ_2 and setting them to zero, we gain

$$
\begin{aligned}
\frac{\partial l_0}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{j=1}^{2} \sum_{k=1}^{20} (Y_{jk} - \mu) = 0 \\
&\implies \sum_{j=1}^{2} \sum_{k=1}^{20} Y_{jk} - 40\mu = 0 \\
&\implies \mu_{MLE} = \frac{\sum_{j=1}^{2} \sum_{k=1}^{20} Y_{jk}}{40}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial l_1}{\partial \mu_1} &= \frac{1}{\sigma^2} \sum_{k=1}^{20} (Y_{1k} - \mu_1) = 0 \\
&\implies \mu_{1MLE} = \frac{\sum_{k=1}^{20} Y_{1k}}{20}
\end{aligned}
\qquad
\begin{aligned}
\frac{\partial l_1}{\partial \mu_2} &= \frac{1}{\sigma^2} \sum_{k=1}^{20} (Y_{2k} - \mu_2) = 0 \\
&\implies \mu_{2MLE} = \frac{\sum_{k=1}^{20} Y_{2k}}{20}
\end{aligned}
$$

These all result in the sample means. The least squares estimates result in the same below.

$$\begin{aligned}
S_0 &= \sum_{j=1}^{2}\sum_{k=1}^{20}(Y_{jk}-\mu) \\
S_1 &= \sum_{k=1}^{20}(Y_{1k}-\mu_1)^2 + \sum_{k=1}^{20}(Y_{2k}-\mu_2)^2
\end{aligned}$$

Similar to the MLE, we take the derivative of these in respect to μ, μ_1, and μ_2 and set them to zero to find the best estimates.

$$\begin{aligned}
\frac{\partial S_0}{\partial \mu} &= -2\sum_{j=1}^{2}\sum_{k=1}^{20}(Y_{jk}-\mu) = 0 \\
\implies & \sum_{j=1}^{2}\sum_{k=1}^{20} Y_{jk} = 40\mu \\
\implies & \mu = \frac{\sum_{j=1}^{2}\sum_{k=1}^{20} Y_{jk}}{40}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial S_1}{\partial \mu_1} &= -2\sum_{k=1}^{20}(Y_{1k}-\mu_1) = 0 \\
\implies \mu_{1LSE} &= \frac{\sum_{k=1}^{20} Y_{1k}}{20}
\end{aligned}
\qquad
\begin{aligned}
\frac{\partial S_1}{\partial \mu_2} &= -2\sum_{k=1}^{20}(Y_{2k}-\mu_2) = 0 \\
\implies \mu_{2LSE} &= \frac{\sum_{k=1}^{20} Y_{2k}}{20}
\end{aligned}$$

(d): Show that the minimum values of the least squares criteria are

$$\begin{aligned}
&\text{for } H_0, \quad \hat{S}_0 = \sum\sum(Y_{jk}-\overline{Y})^2, \text{ where } \overline{Y} = \sum_{j=1}^{2}\sum_{k=1}^{K} Y_{jk}/40; \\
&\text{for } H_1, \quad \hat{S}_1 = \sum\sum(Y_{jk}-\overline{Y_j})^2, \text{ where } \overline{Y_j} = \sum_{k=1}^{K} Y_{jk}/20;
\end{aligned}$$

for $j = 1, 2$.

Solution: From **(c)**, we know that the

$$\begin{aligned}
\frac{\partial S_0}{\partial \mu} &= -2\sum_{j=1}^{2}\sum_{k=1}^{20}(Y_{jk}-\mu) \\
\frac{\partial S_1}{\partial \mu_j} &= -2\sum_{k=1}^{20}(Y_{jk}-\mu_j)
\end{aligned}$$

The estimates that resulted from these are equivelant to $\overline{Y}$ and $\overline{Y_j}$. If it can be shown that the second derivative is positive, then we know that these are the minimum values for $\hat{S}_0$ and $\hat{S}_1$.

$$\frac{\partial^2 S_0}{\partial \mu^2} = 2 \qquad\qquad \frac{\partial^2 S_1}{\partial \mu_j^2} = 2$$

As they are positive, then we know that $\hat{S_0}$ and $\hat{S_1}$ are minimized.

(e): Using the results of Exercise 1.4 show that

$$\frac{1}{\sigma^2}\hat{S_1} = \frac{1}{\sigma^2}\sum_{j=1}^{2}\sum_{k=1}^{20}(Y_{jk}-\mu_j)^2 - \frac{20}{\sigma^2}\sum_{k=1}^{20}(\overline{Y_j}-\mu_j)^2,$$

and deduce that if H_1 is true

$$\frac{1}{\sigma^2}\hat{S_1} \sim \chi^2(38).$$

Similarly show that

$$\frac{1}{\sigma^2}\hat{S_0} = \frac{1}{\sigma^2}\sum_{j=1}^{2}\sum_{k=1}^{20}(Y_{jk}-\mu)^2 - \frac{40}{\sigma^2}\sum_{j=1}^{2}(\overline{Y}-\mu)^2$$

and if H_0 is true then

$$\frac{1}{\sigma^2}\hat{S_0} \sim \chi^2(39).$$

Solution: (Note that handwaving is definitely occurring here) From Exercise 1.4, we know that

$$\sum_{i=1}^{n}(Y_i-\overline{Y})^2 = \sum_{i=1}^{n}(Y_i-\mu)^2 - n(\overline{Y}-\mu)^2$$

Similarly, in this case where $n = K$, we have the following:

$$\hat{S_0} = \sum_{j=1}^{2}\sum_{k=1}^{20}(Y_{jk}-\overline{Y})^2 = \sum_{j=1}^{2}\left[\sum_{k=1}^{20}(Y_{jk}-\mu)^2 - 20(\overline{Y}-\mu)^2\right]$$

and

$$\hat{S_1} = \sum_{j=1}^{2}\sum_{k=1}^{20}(Y_{jk}-\overline{Y_j})^2 = \sum_{j=1}^{2}\left[\sum_{k=1}^{20}(Y_{jk}-\mu_j)^2 - 20(\overline{Y_j}-\mu_j)^2\right]$$

which with the $\frac{1}{\sigma^2}$ term will become below

$$\frac{1}{\sigma^2}\hat{S_0} = \frac{1}{\sigma^2}\sum_{j=1}^{2}\sum_{k=1}^{20}(Y_{jk}-\mu)^2 - \frac{40}{\sigma^2}\sum_{j=1}^{2}(\overline{Y}-\mu)^2$$
$$\frac{1}{\sigma^2}\hat{S_1} = \frac{1}{\sigma^2} sum_{j=1}^{2}\sum_{k=1}^{20}(Y_{jk}-\mu_j)^2 - \frac{20}{\sigma^2}\sum_{k=1}^{20}(\overline{Y_j}-\mu_j)^2$$

In the case of S_0, we only estimate one parameter, so $S_0/\sigma^2 \sim \chi^2(JK-1)$, while for S_0 we estimate two parameters so $S_1/\sigma^2 \sim \chi^2(JK-2)$. Therefore

$$S_0/\sigma^2 \sim \chi^2(39)$$
$$S_1/\sigma^2 \sim \chi^2(38)$$

(f): Use an argument similar to the one in Example 2.2.2 and the results from (e) to deduce that the statistic

$$F = \frac{\hat{S}_0 - \hat{S}_1}{\hat{S}_1/38}$$

has the central F-distribution $F(1, 38)$ if H_0 is true and a non-central distribution if H_0 is not true.

Solution: If H_0 is correct, that is the means of the control and treatment are equal (and so difference is approximately 0 and thus centralized), then $\frac{1}{\sigma^2}(\hat{S}_0 - \hat{S}_1) \sim \chi^2(1)$. If H_0 is not correct, however, then $\frac{1}{\sigma^2}(\hat{S}_0 - \hat{S}_1)$ is not centralized and cannot be compared and so we use S_1 with it's central chi-squared distribution. This results in

$$\frac{(\hat{S}_0 - \hat{S}_1)/1}{\hat{S}_1/38} \sim F(1, 38)$$

(g): Calculate the F-statistic from (f) and use it to test H_0 against H_1. What do you conclude?

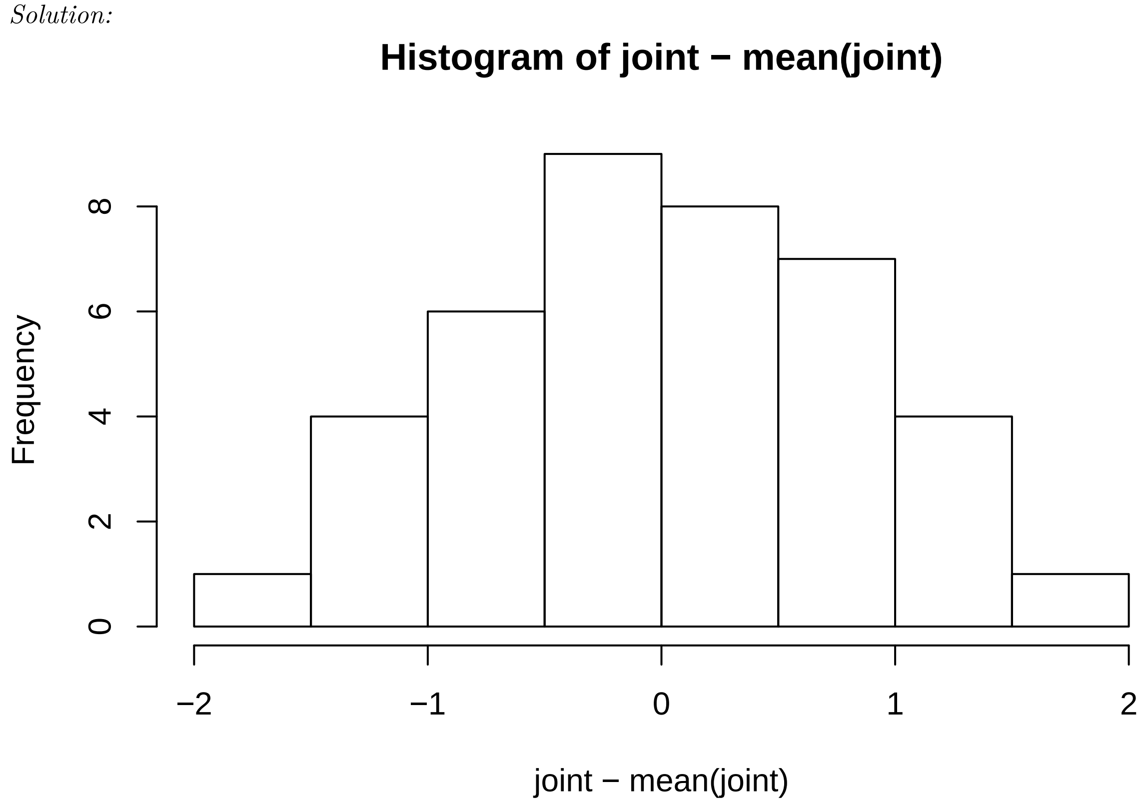
```
## The F-Statistic is  0.2428452

##
## The p-value is  0
```

(h): Compare the value of F-statistic from (g) with the t-statistic from (b), recalling the relationship between the t-distribution and the F-distribution (see Section 1.4.4). Also compare the conclusion from (b) and (g).

Solution: The p-value from the F-test was $\approx .75$ while the p-value from the t-test was $\approx .65$. The t-statistic was $\approx .50$ while the F-statistic was $\approx .25$. This makes sense because $T^2 \sim F(1, 40)$ which is close to the final distribution, and $T^2 \approx .25$.

(i): Calculate residuals from the model for H_0 and use them to explore the distributional assumptions.

Solution:




```
##
##  Shapiro-Wilk normality test
```

```
##
## data:  joint - mean(joint)
## W = 0.98419, p-value = 0.8387
```

Part of the assumption of the normal distribution is the normality of the residuals, and they indeed distributed approximately normally which is confirmed by the shapiro test and histogram.

Exercise Two

Question

The weights, in kilograms, of twenty men before and after participation in a "waist loss" program are shown in Table 2.8 (Egger et al. 1999). We want to know if, on average, they retain a weight loss twelve months after the program.

Man	Before	After	Man	Before	After
1	100.8	97	11	105	105
2	102	107.5	12	85	82.4
3	105.9	97	13	107.2	98.2
4	108	108	14	80	83.6
5	92	84	15	115.1	115
6	116.7	111.5	16	103.5	103
7	110.2	102.5	17	82	80
8	135	127.5	18	101.5	101.5
9	123.5	118.5	19	103.5	102.6
10	95	94.2	20	93	93

Let Y_{jk} denote the weight of the kth man at the jth time, where $j = 1$ before the program and $j = 2$ twelve months later. Assume the Y_{jk}'s are independent random variables $Y_{jk} \sim N(\mu_j, \sigma^2)$ for $j = 1, 2$ and $k = 1, ..., 20$.

(a): Use an unpaired t-test to test the hypothesis

$$H_0 : \mu_1 = \mu_2 \quad \text{versus} \quad H_1 : \mu_1 \neq \mu_2.$$

Solution: I'll begin by entering the data into R.

```
##
##  Welch Two Sample t-test
##
## data:  before and after
## t = 0.65471, df = 37.749, p-value = 0.5166
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.639874 11.029874
## sample estimates:
## mean of x mean of y
##   103.295   100.600
```

(b): Let $D_k = Y_{1k} - Y_{2k}$, for $k = 1, ..., 20$. Formulate models for testing H_0 against H_1 using the D_k's. Using analogous methods to Exercise 2.1 above, assuming σ^2 is a known constant, test H_0 against H_1.

Unsure of where to go from here.

(c): The analysis in (b) is a paired t-test which uses the natural relationship between weights of the *same* person before and after the program. Are the conclusions the same from (a) and (b)?

Unsure where to go from here.

(d): List the assumptions made for (a) and (b). Which analysis is more appropriate for these data?

Unsure where to go from here.

Exercise Three

Question

For model (2.7) for the data on birthweight and gestational age, using methods similar to those for Exercise 1.4, show

$$
\begin{aligned}
\hat{S}_1 &= \sum_{j=1}^{J}\sum_{k=1}^{K}(Y_{jk} - a_j - b_j x_{jk})^2 \\
&= \sum_{j=1}^{J}\sum_{k=1}^{K}[(Y_{jk} - (a_j + \beta_j x_{jk})]^2 - K\sum_{j=1}^{J}(\overline{Y_j} - \alpha_j - \beta_j \overline{x_j})^2 \\
&- \sum_{j=1}^{J}(b_j - \beta_j)^2(\sum_{k=1}^{K} x_{jk}^2 - K\overline{x_j^2})
\end{aligned}
$$

and that the random variables Y_{jk}, $\overline{Y_j}$ and b_j are all independent and have the following distributions

$$
\begin{aligned}
Y_{jk} &\sim \mathrm{N}(\alpha_j + \beta_j x_{jk}, \sigma^2), \\
\overline{Y_j} &\sim \mathrm{N}(\alpha_j + \beta_j \overline{x_j}, \sigma^2/K), \\
b_j &\sim \mathrm{N}(\beta_j, \sigma^2/(\sum_{k=1}^{K} x_{jk}^2 - K\overline{x_j^2})).
\end{aligned}
$$

Solution

We begin by showing the equivelance, but will first note some terms and identities of importance.

$$
\mu_{jk} = \alpha_j + \beta_j x_{jk} \qquad \sum_{k=1}^{K} Y_{jk} = K\overline{Y_j} \qquad \overline{Y_j} = a_j + b_j x_{jk}
$$

Now, our final result uses α_j and β_j so we need a term that uses these, and so we add $-\mu_{jk} + \mu_{jk}$ to the left hand side.

$$
\begin{aligned}
&\sum_{j=1}^{J}\sum_{k=1}^{K}(Y_{jk} - \mu_{jk} + \mu_{jk} - a_j - b_j x_{jk})^2, \quad A = Y_{jk} - \mu_{jk}, \quad B = \mu_{jk} - a_j - b_j x_{jk} \\
=&\sum_{j=1}^{J}\sum_{k=1}^{K}(A + B)^2 \\
=&\sum_{j=1}^{J}\sum_{k=1}^{K} A^2 + 2AB + B^2
\end{aligned}
$$

Note that A^2 is equal to the first term of the right hand side of the equation (because $A^2 = (Y_{jk} - \mu_{jk})^2 = (Y_{jk} - (\alpha_j + \beta_j x_{jk}))^2$. So, we need to show that $2AB + B^2$ is equal to the remaining two terms.

$$
\begin{aligned}
2AB &= 2(Y_{jk}-\mu_{jk})(\mu_{jk}-a_j-b_jx_{jk}) \\
&= 2Y_{jk}\mu_{jk}-2Y_{jk}a_j-2Y_{jk}b_jx_{jk}-2\mu_{jk}^2+2\mu_{jk}a_j+2b_jx_{jk}\mu_{jk}) \\
B^2 &= (\mu_{jk}-a_j-b_jx_{jk})(\mu_{jk}-a_j-b_jx_{jk}) \\
&= \mu_{jk}^2-a_j\mu_{jk}-b_jx_{jk}\mu_{jk}-a_j\mu_{jk}+a_j^2+a_jb_jx_{jk}-b_jx_{jk}\mu_{jk}+a_jb_jx_{jk}+b_j^2x_{jk}^2 \\
&= \mu_{jk}^2-2a_j\mu_{jk}-2b_jx_{jk}\mu_{jk}+2a_jb_jx_{jk}+a_j^2+b_j^2x_{jk}^2 \\
2AB+B^2 &= 2Y_{jk}\mu_{jk}-2Y_{jk}a_j-2Y_{jk}b_jx_{jk}-\mu_{jk}^2+2a_jb_jx_{jk}+a_j^2+b_j^2x_{jk}^2 \\
&= 2Y_{jk}\alpha_j+2Y_{jk}\beta_jx_{jk}-2Y_{jk}a_j-2Y_{jk}b_jx_{jk}-\alpha_j^2-2\alpha_j\beta_jx_{jk}-\beta_j^2x_{jk}^2+2a_jb_jx_{jk}+a_j^2+b_j^2x_{jk}^2
\end{aligned}
$$

From this point, a majority of the terms in $2AB+B^2$ can be simplified using the identities noted in the beginning. However, the lack of a method to handle terms like $\sum_{j=1}^J\sum_{k=1}^K x_{jk}Y_{jk}$ remains a problem for simplifying our terms. We utilize normal equations on Page 26 (3rd ed) to find a substitution for $\sum_{j=1}^J\sum_{k=1}^K x_{jk}Y_{jk}$.

$$
\begin{aligned}
&& b_j &= \frac{K\sum_K x_{jk}Y_{jk}-(\sum_K x_{jk})(\sum_K Y_{jk})}{K\sum_K x_{jk}^2-(\sum_K x_{jk})^2} \\
\implies && b_j(K\sum_K x_{jk}^2-(\sum_K x_{jk})^2) &= K\sum_K x_{jk}Y_{jk}-(\sum_K x_{jk})(\sum_K Y_{jk}) \\
\implies && K\sum_K x_{jk}Y_{jk} &= b_j(K\sum_K x_{jk}^2-(\sum_K x_{jk})^2)+(\sum_K x_{jk})(\sum_K Y_{jk}) \\
\implies && K\sum_K x_{jk}Y_{jk} &= b_j(K\sum_K x_{jk}^2-K^2\overline{x_j})+K^2\overline{x_j}\overline{Y_j} \\
\implies && \sum_K x_{jk}Y_{jk} &= b_j(\sum_K x_{jk}^2-K\overline{x_j})+K\overline{x_j}\overline{Y_j}
\end{aligned}
$$

With this information, we can attempt to obtain our desired result.

$$
\begin{aligned}
&= \sum_{j=1}^J\sum_{k=1}^K(2AB+B^2) \\
&= \sum_{j=1}^J\sum_{k=1}^K 2Y_{jk}\alpha_j+2Y_{jk}\beta_jx_{jk}-2Y_{jk}a_j-2Y_{jk}b_jx_{jk}-\alpha_j^2-2\alpha_j\beta_jx_{jk}-\beta_j^2x_{jk}^2+2a_jb_jx_{jk}+a_j^2+b_j^2x_{jk}^2 \\
&= \sum_{j=1}^J[2K\overline{Y_j}\alpha_j-2K\overline{Y_j}a_j-2K\alpha_j\beta_j\overline{x_j}+2Ka_jb_j\overline{x_j}-K\alpha_j^2+Ka_j^2+2K\overline{Y_j}\beta_j\overline{x_j}-2K\overline{Y_j}\beta_j\overline{x_j} \\
&\quad +2\beta_jb_j(\sum_{k=1}^K x_{jk}^2-K\overline{x_j}^2)-2b_j^2(\sum_{k=1}^K x_{jk}^2-K\overline{x_j}^2)-\beta_j^2x_{jk}^2+b^2x_{jk}^2] \\
&= \sum_{j=1}^J[2K\overline{Y_j}\alpha_j-2K\overline{Y_j}a_j-2K\alpha_j\beta_j\overline{x_j}+2Ka_jb_j\overline{x_j}-K\alpha_j^2+Ka_j^2+2K\overline{Y_j}\beta_j\overline{x_j}-2K\overline{Y_j}\beta_j\overline{x_j} \\
&\quad +2\beta_jb_j(\sum_{k=1}^K x_{jk}^2-K\overline{x_j}^2)-2b_j^2(\sum_{k=1}^K x_{jk}^2-K\overline{x_j}^2)-\beta_j^2x_{jk}^2+b^2x_{jk}^2]
\end{aligned}
$$

If we add the following zero term to the above equation

$$
Kb^2\overline{x_j}^2-Kb^2\overline{x_j}^2+K\beta^2\overline{x_j}^2-K\beta^2\overline{x_j}^2
$$

then much of the problem can be reduced.

$$
\begin{aligned}
=&K\sum_{j=1}^{J}\left[2\overline{Y_j}\alpha_j - 2\overline{Y_j}a_j - 2\alpha_j\beta_j\overline{x_j} + 2a_jb_j\overline{x_j} - \alpha_j^2 + a_j^2 + 2\overline{Y_j}\beta_j\overline{x_j} - 2\overline{Y_j}\beta_j\overline{x_j} - K\beta^2\overline{x_j}^2 + Kb^2\overline{x_j}^2\right]\\
&+\sum_{j=1}^{J}\left[2\beta_jb_j(\sum_{k=1}^{K}x_{jk}^2 - K\overline{x_j}^2) - 2b_j^2(\sum_{k=1}^{K}x_{jk}^2 - K\overline{x_j}^2) - \beta_j^2x_{jk}^2 + b_j^2x_{jk}^2 - Kb_j^2\overline{x_j}^2 + K\beta_j^2\overline{x_j}^2\right]\\
=&K\sum_{j=1}^{J}\left[2K\overline{Y_j}\alpha_j + 2\overline{Y_j}\beta_j\overline{x_j} - 2\overline{Y_j}(\alpha_j + \beta_j\overline{x_j}) - \alpha_j^2 - 2\alpha_j\beta_j\overline{x_j} - \beta_j^2\overline{x_j}^2 + (\alpha_j + \beta_j\overline{x_j})^2\right]\\
&+\sum_{j=1}^{J}\left[-b_j^2(\sum_{k=1}^{K}x_{jk}^2 - K\overline{x_j}^2) - 2b_j\beta_j(\sum_{k=1}^{K}x_{jk}^2 - K\overline{x_j}^2) - \beta^2(\sum_{k=1}^{K}x_{jk}^2 - K\overline{x_j}^2)\right]\\
=&-K\sum_{j=1}^{J}\left[\overline{Y_j}^2 - 2\alpha_j\overline{Y_j} - 2\beta_j\overline{x_j}\overline{Y_j} + 2\alpha_j\beta_j\overline{x_j} + \alpha_j^2 + \beta_j^2\overline{x_j}^2\right]\\
&-\sum_{j=1}^{J}\left[(b_j - \beta_j)^2(\sum_{k=1}^{K}x_{jk}^2 - K\overline{x_j}^2)\right]\\
=&-K\sum_{j=1}^{J}(\overline{Y_j} - \alpha_j - \beta_j\overline{x_j})^2\\
&-\sum_{j=1}^{J}\left[(b_j - \beta_j)^2(\sum_{k=1}^{K}x_{jk}^2 - K\overline{x_j}^2)\right]
\end{aligned}
$$

So we have shown that they are equal.

In regards to showing how they are distributed, we can recognize that $\hat{S}_1/\sigma^2$ is now the sum of three χ^2 random variables. So, as the three are also χ^2 random variables with behavior similar to residuals, then the variables within the square behave normally.

Exercise Four

Suppose you have the following data:

x:	1.0	1.2	1.4	1.6	1.8	2.0
y:	3.15	4.85	6.50	7.20	8.25	16.50

and you want to fit a model with

$$\mathrm{E}(Y) = \ln(\beta_0 + \beta_1 x + \beta_2 x^2).$$

Write this model in the form of (2.13) specifying the vector $\mathbf{y}$ and β and the matrix $\mathbf{X}$.

Solution

$$\begin{bmatrix} e^{3.15} \\ e^{4.85} \\ e^{6.50} \\ e^{7.20} \\ e^{8.25} \\ e^{16.50} \end{bmatrix} \approx \begin{bmatrix} 1 & 1.0 & 1.00 \\ 1 & 1.2 & 1.44 \\ 1 & 1.4 & 1.96 \\ 1 & 1.6 & 2.56 \\ 1 & 1.8 & 3.24 \\ 1 & 2.0 & 4.00 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

Exercise Five

The model for two-factor analysis of variance with two levels of one factor, three levels of the other ad no replication is

$$\mathrm{E}(Y_{jk}) = \mu_{jk} = \mu + \alpha_j + \beta_k; \qquad Y_{jk} \sim \mathrm{N}(\mu_{jk}, \sigma^2),$$

where $j = 1, 2; k = 1, 2, 3$ and, using the sum-to-zero constraints, $\alpha_1 + \alpha_2 = 0, \beta_1 + \beta_2 = 0$. Also the $Y_{jk}'s are assumed to be independent.

Write the equation for $\mathrm{E}(Y_{jk})$ in matrix notation. (Hint: Let $\alpha_2 = -\alpha_1$ and $\beta_3 = -\beta_1 - \beta_2$)

Solution

$$\begin{bmatrix} E[Y_1] \\ E[Y_2] \\ E[Y_3] \\ E[Y_4] \\ E[Y_5] \\ E[Y_6] \end{bmatrix} \approx \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$