

STAT 421 Homework 4

Lecture 9 Problem 1a:

The degrees of freedom of the term $\sum_i^a \sum_j^n (y_{ij} - \mu)^2$ is an . The degrees of freedom is an because there are a total of an different terms in the sum, and the terms inside the sum, $y_{ij} - \mu$, are subject to zero constraints.

Lecture 9 Problem 1b:

The degrees of freedom of the term $\sum_i^a \sum_j^n (y_{ij} - \bar{y}_{..})^2$ is $an - 1$. The degrees of freedom is $an - 1$ because there are a total of an different terms in the sum, and the terms inside the sum, $y_{ij} - \bar{y}_{..}$, are subject to one constraint: $\sum_i^a \sum_j^n (y_{ij} - \bar{y}_{..}) = 0$.

Lecture 9 Problem 1c:

The degrees of freedom of the term $\sum_i^a (\bar{y}_{i.} - \bar{y}_{..})^2$ is $a - 1$. The degrees of freedom is $a - 1$ because there are a total of a different terms in the sum, and the terms inside the sum, $\bar{y}_{i.} - \bar{y}_{..}$, are subject to one constraint: $\sum_i^a (\bar{y}_{i.} - \bar{y}_{..}) = 0$.

Lecture 9 Problem 2:

In this problem, our goal is to show that $E[MS_E] = \sigma_\epsilon^2$, using the identity $E[(\epsilon_{ij} - \bar{\epsilon}_{i.})^2] = \frac{n-1}{n} \sigma_\epsilon^2$. We will start this calculation below.

$$\begin{aligned} E[MS_E] &= E \left[\frac{1}{N - a} SS_E \right] \\ &= \frac{1}{N - a} E[SS_E] \\ &= \frac{1}{N - a} E \left[(n - 1) \sum_i^a S_i^2 \right] \\ &= \frac{1}{N - a} E \left[\frac{(n - 1) \sum_i^a \sum_j^n (y_{ij} - \bar{y}_{i.})^2}{n - 1} \right] \\ &= \frac{1}{N - a} E \left[\sum_i^a \sum_j^n (y_{ij} - \bar{y}_{i.})^2 \right] \end{aligned}$$

In the lecture, we defined our model and its average over the n replications as:

$$\begin{aligned} y_{ij} &= \mu + \tau_i + \epsilon_{ij} \\ \bar{y}_{i.} &= \mu + \tau_i + \bar{\epsilon}_{i.} \end{aligned}$$

We will substitute these into the above equation for $E[MS_E]$ and continue our calculation below.

$$\begin{aligned}
E[MS_E] &= \frac{1}{N-a} E \left[\sum_i^a \sum_j^n (\mu + \tau_i + \epsilon_{ij} - \mu - \tau_i - \bar{\epsilon}_{i.})^2 \right] \\
&= \frac{1}{N-a} E \left[\sum_i^a \sum_j^n (\epsilon_{ij} - \bar{\epsilon}_{i.})^2 \right] \\
&= \frac{1}{N-a} \sum_i^a \sum_j^n E [(\epsilon_{ij} - \bar{\epsilon}_{i.})^2] \\
&= \frac{an}{an-a} E [(\epsilon_{ij} - \bar{\epsilon}_{i.})^2] \\
&= \frac{an}{an-a} \frac{n-1}{n} \sigma_\epsilon^2 \\
&= \frac{an^2 - an}{an^2 - an} \sigma_\epsilon^2 \\
&= \sigma_\epsilon^2
\end{aligned}$$

Hence we have just shown that $E[MS_E] = \sigma_\epsilon^2$, using the identity $E [(\epsilon_{ij} - \bar{\epsilon}_{i.})^2] = \frac{n-1}{n} \sigma_\epsilon^2$.

Lecture 9 Problem 3a:

In this problem we are using the data from exercise 3.22, which are shown below. In this completely randomized experiment, the response time in milliseconds was determined for three different types of circuits that could be used in an automatic valve shutoff mechanism.

Circuit Type		Response Time			
1	9	12	10	8	15
2	20	21	23	17	30
3	6	5	8	16	7

In this part of the problem, we will run a one-way ANOVA F-test to determine if the means vary across the levels of the treatment factor, using the p-value method. The calculation of this p-value is calculated by hand in R below.

```

# Set up number of treatment factors and replications:
a <- 3
n <- 5
N <- a*n

# Set up an empty matrix:
y <- matrix(nrow = 3, ncol = 5)

# Add the data to the matrix:
y[1,] <- c(9, 12, 10, 8, 15)
y[2,] <- c(20, 21, 23, 17, 30)
y[3,] <- c(6, 5, 8, 16, 7)

# Calculate the "effect estimates" for each factor level:
est_effects <- rowMeans(y) - mean(y)

```

```

# Find the sample variance for each treatment factor level:
vars <- numeric(a)
for (i in 1:a) {
  vars[i] <- var(y[i,])
}

# Compute sums of squares and mean squares for the F-test:
SS_treat <- n*sum(est_effects^2)
SS_e <- (n-1)*sum(vars)
MS_treat <- SS_treat / (a-1)
MS_e <- SS_e / (N-a)

# Compute the observed F value:
F_obs <- MS_treat / MS_e

# Compute the p-value of the test:
p_val_F <- pf(F_obs, a-1, N-a, lower.tail = F)

```

As calculated from the above one-way ANOVA F-test, the p-value associated with this test was 0.0004023258. Furthermore, as described in the problem, the significance level is $\alpha = 0.01$. Since the p-value is smaller than the significance level, we reject H_0 in favor of H_1 . Thus we have significant evidence that at least two of the means are different across the levels of the treatment factor.

Lecture 9 Problem 3b

In this problem, we will make qqplots of the response for each of the 3 levels of the treatment factor X and interpret what the plots are telling us. The code for making these qqplots is shown below.

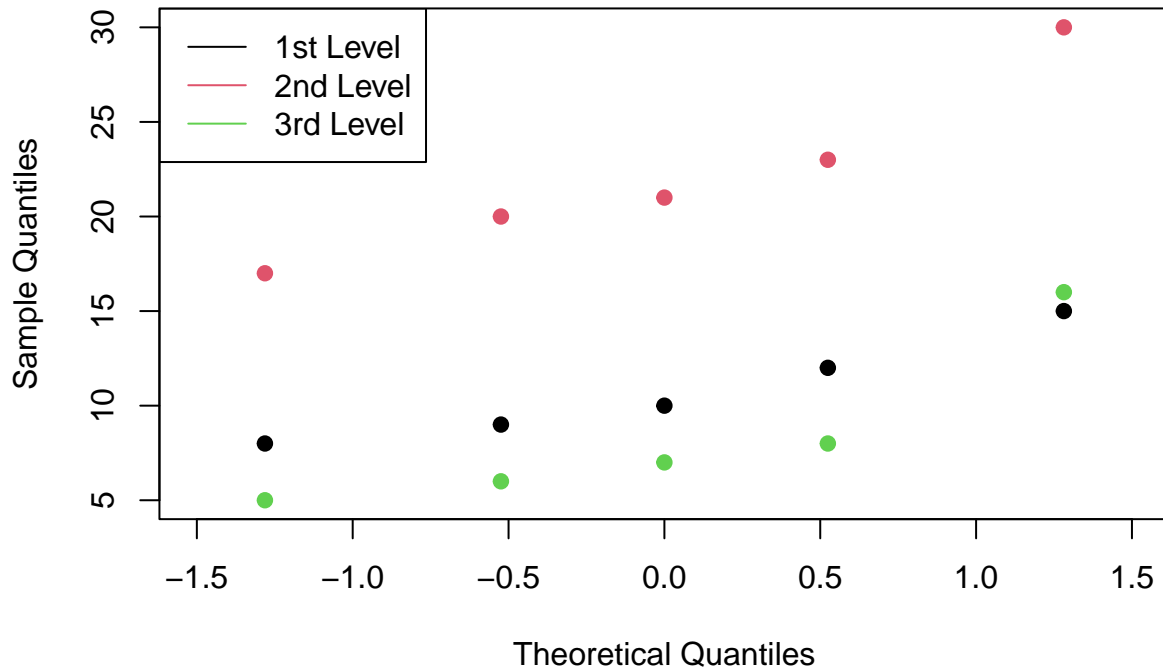
```

# Create qqplots for the response of each treatment level:
plot(c(0,0), cex=0, xlim=c(-1.5,1.5), ylim=range(y), xlab="Theoretical Quantiles",
     ylab="Sample Quantiles", main="QQplots of all 3 Treatment Factor Levels")
for (i in 1:3) {
  probs = seq(.5/n,1-.5/n, length=n)
  Q = qnorm(probs, 0, 1)
  points(Q, sort(y[i,]), col=i, pch=19)
}

# Add legend:
legend("topleft",
      legend=c("1st Level", "2nd Level", "3rd Level"),
      col=c(1, 2, 3),
      lty=1)

```

QQplots of all 3 Treatment Factor Levels



As can be seen from the above qqplot, since all three treatment factor levels are roughly linear, our assumption that all three come from a normal population is valid. Furthermore, since all three treatment factor levels have roughly the same slope, our assumption that all three come from a population with equal variance is also valid. Lastly, the fact that all three treatment factor levels have noticeably different intercept values, shows why the ANOVA F-test rejected H_0 in favor of H_1 .

Lecture 9 Problem 3c:

In this problem, we will find 95% confidence intervals for μ , μ_i , and σ_ϵ^2 , all by hand. The equations used to calculate these intervals are given in order as follows: $\hat{y}_{i.} \pm t_{\frac{\alpha}{2}, an-a} \sqrt{\frac{MSE}{n}}$, $\hat{y}_{i.} - \hat{y}_{j.} \pm t_{\frac{\alpha}{2}, an-a} \sqrt{\frac{2MSE}{n}}$, and $\left[\frac{(N-a)MSE}{\chi^2_{\frac{\alpha}{2}, an-a}}, \frac{(N-a)MSE}{\chi^2_{1-\frac{\alpha}{2}, an-a}} \right]$. These calculations are done in R below.

```
# Set obtain important values for the intervals:
```

```
alpha <- 0.05
```

```
est_grand <- mean(y)
```

```
t_crit <- qt(alpha/2, N-a, lower.tail=F)
```

```
# Create 95% confidence interval for mu:
```

```
lower_mu <- est_grand - t_crit * sqrt(MS_e / N)
```

```
upper_mu <- est_grand + t_crit * sqrt(MS_e / N)
```

```
# Create 95% confidence interval for mu_i:
```

```
lower_mu_i <- numeric(a)
```

```
upper_mu_i <- numeric(a)
```

```
for (i in 1:a) {
```

```

est_mu_i <- mean(y[i,])
lower_mu_i[i] <- est_mu_i - t_crit * sqrt(MS_e / n)
upper_mu_i[i] <- est_mu_i + t_crit * sqrt(MS_e / n)
}

# Create 95% confidence interval for sigma^2_e:
lower_sigma2_e <- (N-a)*MS_e / qchisq(alpha/2, N-a, lower.tail = F)
upper_sigma2_e <- (N-a)*MS_e / qchisq(alpha/2, N-a)

```

The 95% confidence interval for μ is [11.4873088, 16.1126912]. The 95% confidence interval for μ_1 is [6.7943014, 14.8056986]. The 95% confidence interval for μ_2 is [18.1943014, 26.2056986]. The 95% confidence interval for μ_3 is [4.3943014, 12.4056986]. Lastly, the 95% confidence interval for σ_e^2 is [8.6901881, 46.0512578].

Lecture 9 Problem 3d:

In this problem, we will compute the p-value for testing whether μ_1 is different from μ_3 , we will use the two-sample t-test, keeping in mind that the estimate for σ_e^2 is still the MS_E for the full model.

```

# Find sample information:
ybar_1 <- mean(y[1,])
ybar_3 <- mean(y[3,])

# Find the observed test statistic value:
t_obs <- (ybar_1 - ybar_3) / sqrt(2*MS_e / n)

# Find the p-value:
p_val_t <- 2*pt(t_obs, N-a, lower.tail = F)

```

As calculated from the above two-sample t-test, the p-value is 0.374155. At the significance level of $\alpha = 0.01$, since the p-value is greater than the significance level we fail to reject H_0 in favor of H_1 . Thus we have no evidence that the mean of treatment factor 1 is different from the mean of treatment factor 2.

Lecture 10 Problem 1a:

In this problem we are using the data from exercise 3.6, but ignoring the “Sham” column, which are shown below. In this randomized double-blind experiment, the percent loss in bone mineral density is collected for patients wearing pulsing electromagnetic field (PEMF) transducers for three different lengths of time.

Sham	PEMF 1 h/day	PEMF 2 h/day	PEMF 4 h/day
4.51	5.32	4.73	7.03
7.95	6.00	5.81	4.65
4.97	5.12	5.69	6.65
3.00	7.08	3.86	5.49
7.97	5.48	4.06	6.98
2.23	6.52	6.56	4.85
3.95	4.09	8.34	7.26
5.64	6.28	3.01	5.92

9.35	7.77	6.71	5.58
6.52	5.68	6.51	7.91
4.96	8.47	1.70	4.90
6.10	4.58	5.89	4.54
7.19	4.11	6.55	8.18
4.03	5.72	5.34	5.42
2.72	5.91	5.88	6.03
9.19	6.89	7.50	7.04
5.17	6.99	3.28	5.17
5.70	4.98	5.38	7.60
5.85	9.94	7.30	7.90
6.45	6.38	5.46	7.91

In this part of the problem, we will perform a one-way ANOVA F-test to see if the number of hours has an effect on bone density. We will report a conclusion based on the p-value, and we will do calculations using R functions as shown below.

```
# Set up number of treatment factors and replications:
a_1 <- 3
n_1 <- 20
N_1 <- a_1*n_1

# Set up an empty matrix:
y_1 <- matrix(nrow = 3, ncol = 20)

# Add the data to the matrix:
y_1[1,] <- c(5.32, 6.00, 5.12, 7.08, 5.48, 6.52, 4.09, 6.28, 7.77, 5.68, 8.47,
            4.58, 4.11, 5.72, 5.91, 6.89, 6.99, 4.98, 9.94, 6.38)
y_1[2,] <- c(4.73, 5.81, 5.69, 3.86, 4.06, 6.56, 8.34, 3.01, 6.71, 6.51, 1.70,
            5.89, 6.55, 5.34, 5.88, 7.50, 3.28, 5.38, 7.30, 5.46)
y_1[3,] <- c(7.03, 4.65, 6.65, 5.49, 6.98, 4.85, 7.26, 5.92, 5.58, 7.91, 4.90,
            4.54, 8.18, 5.42, 6.03, 7.04, 5.17, 7.60, 7.90, 7.91)

# Create the levels of X:
x <- rep(1:a_1, n_1)

# Perform the one-factor ANOVA F-test in R:
summary.aov(lm(as.vector(y_1)~as.factor(x)))

##           Df Sum Sq Mean Sq F value Pr(>F)
## as.factor(x) 2   8.45   4.227    2.01  0.143
## Residuals   57 119.85   2.103
```

As calculated from the one-way ANOVA F-test, the p-value is 0.143. At a self-chosen significance level of $\alpha = 0.05$, since the p-value is greater than the significance level, we fail to reject H_0 in favor of H_1 . Therefore, we have no evidence that the number of hours has an effect on bone density.

Lecture 10 Problem 1b:

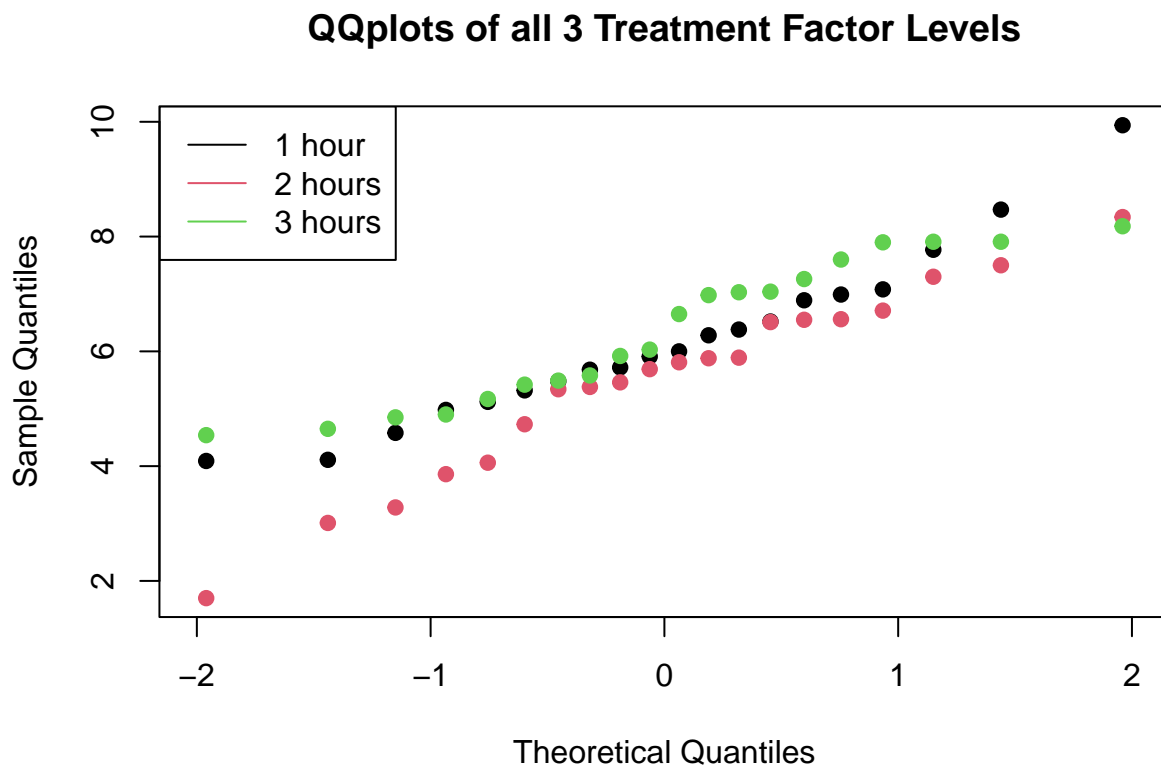
We are skipping this part of the problem.

Lecture 10 Problem 1c:

In this problem, we will make qqplots of the response for each of the 3 levels of the treatment factor X and interpret what the plots are telling us. The code for making these qqplots is shown below.

```
# Create qqplots for the response of each treatment level:
plot(c(0,0), cex=0, xlim=c(-2,2), ylim=range(y_1), xlab="Theoretical Quantiles",
     ylab="Sample Quantiles", main="QQplots of all 3 Treatment Factor Levels")
for (i in 1:a_1) {
  probs = seq(.5/n_1, 1-.5/n_1, length=n_1)
  Q = qnorm(probs, 0, 1)
  points(Q, sort(y_1[i,]), col=i, pch=19)
}

# Add legend
legend("topleft",
      legend=c("1 hour", "2 hours", "3 hours"),
      col=c(1, 2, 3),
      lty=1)
```



As can be seen from the above qqplot, since all three treatment factor levels are roughly linear, our assumption that all three come from a normal population is valid. Furthermore, since all three treatment factor levels have roughly the same slope, our assumption that all three come from a population with equal variance is also valid. Lastly, the fact that all three treatment factor levels have very similar intercept values, shows why the ANOVA F-test failed to reject H_0 in favor of H_1 .

Lecture 10 Problem 1d:

In this problem, we will compute the 95% confidence interval for μ_1 by hand. We will use the equation given

in lecture to compute this interval: $\hat{y}_{i.} \pm t_{\frac{\alpha}{2}, an-a} \sqrt{\frac{MSE}{n}}$. This calculation is done in R below.

```
# Find the sample variance for each treatment factor level:
vars_1 <- numeric(a_1)
for (i in 1:a_1) {
  vars_1[i] <- var(y_1[i,])
}

# Compute sums of squares and mean squares for the interval:
SS_e_1 <- (n_1-1)*sum(vars_1)
MS_e_1 <- SS_e_1 / (N_1-a_1)

# Create important values for the confidence interval:
alpha_1 <- 0.05
est_mu_1 <- mean(y_1[1,])
t_crit_1 <- qt(alpha_1/2, N_1-a_1, lower.tail=F)

# Create 95% confidence interval for mu_1:
lower_mu_1 <- est_mu_1 - t_crit_1 * sqrt(MS_e_1 / n_1)
upper_mu_1 <- est_mu_1 + t_crit_1 * sqrt(MS_e_1 / n_1)
```

As calculated above, the 95% confidence interval for μ_1 is [5.5162317, 6.8147683].

Lecture 10 Problem 1e:

In this problem, we will compute the 95% confidence interval for $\mu_2 - \mu_3$ by hand. We will use the equation given in lecture to compute this interval: $\hat{y}_{i.} - \hat{y}_{j.} \pm t_{\frac{\alpha}{2}, an-a} \sqrt{\frac{2MSE}{n}}$. This calculation is done in R below.

```
# Create important values for the confidence interval:
est_mu_2 <- mean(y_1[2,])
est_mu_3 <- mean(y_1[3,])

# Create 95% confidence interval for mu_2 - mu_3:
lower_mu_23 <- (est_mu_2 - est_mu_3) - t_crit_1 * sqrt((2*MS_e_1) / n_1)
upper_mu_23 <- (est_mu_2 - est_mu_3) + t_crit_1 * sqrt((2*MS_e_1) / n_1)
```

As calculated above, the 95% confidence interval for $\mu_2 - \mu_3$ is [-1.7907041, 0.0457041].

Lecture 10 Problem 1f:

In this problem, we will use a significance level of $\alpha = 0.05$, the rejection region method, and the contrast method to test $H_0 : \mu_1 = \frac{\mu_2 + \mu_3}{2}$ versus $H_1 : \mu_1 \neq \frac{\mu_2 + \mu_3}{2}$.

The contrast method starts by estimating the contrast $\Gamma = \sum_i^a c_i \mu_i$ with $\hat{\Gamma} = \sum_i^a c_i \bar{y}_{i.}$. Then using the fact that $E[\hat{\Gamma}] = \Gamma$, and the fact that $V[\hat{\Gamma}] = \frac{\sigma_e^2 \sum_i^a c_i^2}{n}$ when the data are balanced and homoscedastic, we can test $H_0 : \Gamma = 0$ versus $H_1 : \Gamma \neq 0$. To do this test we use the test statistic $t = \frac{\hat{\Gamma} - \Gamma}{\sqrt{\frac{MSE \cdot |\vec{c}|^2}{n}}} \sim t_{an-a}$, where $|\vec{c}|^2$ is the squared magnitude of the vector of contrast components. Lastly, as per usual, the rejection region is calculated as $|t| > t_{\frac{\alpha}{2}, an-a}$. We will now calculate the rejection region in R below.

```
# Create contrast constants:
contr_const_1 <- c(2, -1, -1)

# Create vector of estimated mu values:
est_mu_1 <- c(est_mu_1, est_mu_2, est_mu_3)
```



```

# Find the estimated contrast value:
est_contr_1 <- sum(contr_const_1*est_mu_1)

# Find the observed t-statistic:
t_obs_1 <- est_contr_1 / sqrt((MS_e_1 * sum((contr_const_1)^2)) / n_1)

# Find the bounds of the rejection region:
t_rr_lower <- qt(alpha_1/2, N_1-a_1)
t_rr_upper <- qt(alpha_1/2, N_1-a_1, lower.tail=F)

```

As calculated above, the rejection region is $(-\infty, -2.0024655] \cup [2.0024655, \infty)$. Since our observed t-statistic was 0.632705, which is not in the rejection region, we fail to reject the H_0 in favor of H_1 . Hence there is no significant evidence that Γ is nonzero.

Lecture 10 Problem 1g:

In this problem, we will compute the 95% confidence interval for the contrast in part f by hand. We will use the equation given in lecture to compute this interval: $\hat{\Gamma} \pm t_{\frac{\alpha}{2}, an-a} \sqrt{\frac{MSE \cdot |c|^2}{n}}$. This calculation is done in R below.

```

# Create 95% confidence interval for mu_1 = (mu_2 - mu_3)/2:
lower_contr <- est_contr_1 - t_crit_1 * sqrt((MS_e_1 * sum((contr_const_1)^2)) / n_1)
upper_contr <- est_contr_1 + t_crit_1 * sqrt((MS_e_1 * sum((contr_const_1)^2)) / n_1)

```

As calculated above, the 95% confidence interval for Γ is $[-1.0878761, 2.0928761]$. Since 0 is in this interval, we fail to reject the H_0 in favor of H_1 . Hence there is no significant evidence that Γ is nonzero.

Lecture 10 Problem 2a:

In this problem we are using the data from exercise 3.25, which are shown below. In this experiment, four chemists are asked to determine the percentage of methyl alcohol in a certain chemical compound. Each chemist makes three determinations, summarized below

Chemist	Percentage of Methyl Alcohol		
1	84.99	84.04	84.38
2	85.15	85.13	84.88
3	84.72	84.48	85.16
4	84.20	84.10	84.55

In this part of the problem, we will perform a one-way ANOVA F-test to see if the chemists results differ significantly. We will report a conclusion based on the p-value, and we will do calculations using R functions as shown below.

```

# Set up number of treatment factors and replications:
a_2 <- 4
n_2 <- 3
N_2 <- a_2*n_2

# Set up an empty matrix:
y_2 <- matrix(nrow = 4, ncol = 3)

```

```

# Add the data to the matrix:
y_2[1,] <- c(84.99, 84.04, 84.38)
y_2[2,] <- c(85.15, 85.13, 84.88)
y_2[3,] <- c(84.72, 84.48, 85.16)
y_2[4,] <- c(84.20, 84.10, 84.55)

# Calculate the "effect estimates" for each factor level:
est_effects_2 <- rowMeans(y_2) - mean(y_2)

# Find the sample variance for each treatment factor level:
vars_2 <- numeric(a_2)
for (i in 1:a_2) {
  vars_2[i] <- var(y_2[i,])
}

# Compute sums of squares and mean squares for the F-test:
SS_treat_2 <- n_2*sum(est_effects_2^2)
SS_e_2 <- (n_2-1)*sum(vars_2)
MS_treat_2 <- SS_treat_2 / (a_2-1)
MS_e_2 <- SS_e_2 / (N_2-a_2)

# Compute the observed F value:
F_obs_2 <- MS_treat_2 / MS_e_2

# Compute the p-value of the test:
p_val_F_2 <- pf(F_obs_2, a_2-1, N_2-a_2, lower.tail = F)

```

As calculated from the one-way ANOVA F-test, $SS_{Treatment} = 1.0445667$, $SS_E = 0.8582$, $MS_{treatment} = 0.3481889$, $MS_E = 0.107275$, $F_{obs} = 3.2457599$, and lastly the p-value is 0.0812898. At a significance level of $\alpha = 0.1$, since the p-value is less than the significance level, we reject H_0 in favor of H_1 . Therefore, we have significant evidence that the chemists measurements differ.

Lecture 10 Problem 2b:

We are skipping this part of the problem.

Lecture 10 Problem 2c:

In this problem, if we assume that chemist 2 is a new employee, we must construct a meaningful set of zero-sum orthogonal contrasts. In particular, we will begin with one sensible zero-sum contrast vector, and then find two more zero-sum orthogonal contrasts. After that we will find the p-value for a two-sided test of each contrast, and state our conclusions at $\alpha = 0.1$.

Hypothetically, if we assume that chemist 2 is a new employee, then it follows that their measurements are most likely going to be the most flawed. Alternatively, if we assume that chemist 2 had an equal amount of training from all of the other three chemists, then it follows that a sensible zero-sum contrast for this situation is $\Gamma_1 = -\mu_1 + 3\mu_2 - \mu_3 - \mu_4$. In other words, the average reading for chemist 2 is simply the average of the other 3 chemist's average readings.

Hence it follows that two more zero-sum orthogonal contrasts are: $\Gamma_2 = 2\mu_1 - \mu_3 - \mu_4$ and $\Gamma_3 = \mu_3 - \mu_4$. We will now Additionally, find the p-value for a two-sided test of each contrast, and state our conclusions at $\alpha = 0.1$. We will first run the two-sided test for contrast 1.

```

# Create contrast constants:
contr_const_2 <- c(-1, 3, -1, -1)

# Create vector of estimated mu values:

```

```

est_mu_2 <- rowMeans(y_2)

# Find the estimated contrast value:
est_contr_2 <- sum(contr_const_2*est_mu_2)

# Find the observed t-statistic:
t_obs_2 <- est_contr_2 / sqrt((MS_e_2 * sum((contr_const_2)^2)) / n_2)

# Find the p-value associated with t_obs:
p_val_t_2 <- 2*pt(t_obs_2, N_2-a_2, lower.tail=F)

```

As calculated in the above contrast test, the p-value was 0.0385258. Since the p-value is less than the significance level $\alpha = 0.1$ it follows that we can reject H_0 in favor of H_1 . Hence we have significant evidence that $-\mu_1 + 3\mu_2 - \mu_3 - \mu_4 \neq 0$. In other words, we have evidence to say that the mean of chemist 2 is not equal to the mean of the means from the other three chemists. We will now move onto testing our second contrast.

```

# Create contrast constants:
contr_const_3 <- c(2, 0, -1, -1)

# Find the estimated contrast value:
est_contr_3 <- sum(contr_const_3*est_mu_2)

# Find the observed t-statistic:
t_obs_3 <- est_contr_3 / sqrt((MS_e_2 * sum((contr_const_3)^2)) / n_2)

# Find the p-value associated with t_obs:
p_val_t_3 <- 2*pt(t_obs_3, N_2-a_2, lower.tail=F)

```

As calculated in the above contrast test, the p-value was 0.7860934. Since the p-value is greater than the significance level $\alpha = 0.1$ it follows that we fail to reject H_0 in favor of H_1 . Hence we do not have significant evidence to say that $2\mu_1 - \mu_3 - \mu_4 \neq 0$. In other words, we do not have evidence to say that the mean of chemist 1 is not the mean of the means from the third or fourth chemist. We will now move onto testing our final contrast.

```

# Create contrast constants:
contr_const_4 <- c(0, 0, 1, -1)

# Find the estimated contrast value:
est_contr_4 <- sum(contr_const_4*est_mu_2)

# Find the observed t-statistic:
t_obs_4 <- est_contr_4 / sqrt((MS_e_2 * sum((contr_const_4)^2)) / n_2)

# Find the p-value associated with t_obs:
p_val_t_4 <- 2*pt(t_obs_4, N_2-a_2, lower.tail=F)

```

As calculated in the above contrast test, the p-value was 0.096587. Since the p-value is less than the significance level $\alpha = 0.1$ it follows that we reject H_0 in favor of H_1 . Hence we have significant evidence to say that $\mu_3 - \mu_4 \neq 0$. In other words, we have evidence that the mean of chemist 3 is not equal to the mean of fourth chemist.

Lecture 11 Problem 1e:

In this problem, which is a continuation of Lecture 9 problem 3, we are interested in whether $\mu_2 = \frac{\mu_1 + \mu_3}{2}$.

Furthermore, we will construct an orthogonal zero-sum contrast and confirm that the contrast sum-of-squares of the two orthogonal contrasts add up to the $SS_{Treatment}$ found in Lecture 9 Problem 3b. We will do this by hand in R, which is shown below.

```
# Create contrast constants for main contrast:
main_const <- c(-1, 2, -1)

# Find estimated contrast for main contrast:
main_contr <- sum(main_const * rowMeans(y))

# Find contrast sum-of-squares for main contrast:
main_ssc <- main_contr^2 / (sum(main_const^2) / n)

# Create contrast constants for orthogonal contrast:
orth_const <- c(1, 0, -1)

# Find estimated contrast for main contrast:
orth_contr <- sum(orth_const * rowMeans(y))

# Find contrast sum-of-squares for main contrast:
orth_ssc <- orth_contr^2 / (sum(orth_const^2) / n)

# Find the sum of the two contrasts:
orth_sum <- main_ssc + orth_ssc
```

As calculated above, the sum of the two orthogonal contrast sum-of-squares is 543.6. As calculated in Lecture 9 Problem 3b, $SS_{Treatment} = 543.6$. As expected, these values are equal.

Lecture 11 Problem 1f:

In this problem we are concerned about what conclusions we can make from the results of part e. In particular, we want to know what we can conclude about “why” the ANOVA F-test was significant. From part e, since the contrast sum-of-squares of the original contrast (529.2) is much larger than the orthogonal contrast (14.4), then we know that when ANOVA rejected its H_0 , it was mostly because $\mu_2 \neq \frac{\mu_1 + \mu_3}{2}$.

Lecture 11 Problem 1g:

In this problem, which is a continuation of Lecture 9 problem 3, we are interested in whether $\mu_2 = \frac{\mu_1 + \mu_3}{2}$. Furthermore, we will construct a non-orthogonal contrast and confirm that the contrast sum-of-squares of the two contrasts do not add up to the $SS_{Treatment}$ found in Lecture 9 Problem 3b. We will do this by hand in R, which is shown below.

```
# Create contrast constants for non-orthogonal contrast:
non_orth_const <- c(1, 1, -1)

# Find estimated contrast for non-orthogonal contrast:
non_orth_contr <- sum(non_orth_const * rowMeans(y))

# Find contrast sum-of-squares for non-orthogonal contrast:
non_orth_ssc <- non_orth_contr^2 / (sum(non_orth_const^2) / n)

# Find the sum of the two contrasts:
non_orth_sum <- main_ssc + non_orth_ssc
```

As calculated above, the sum of the two orthogonal contrast sum-of-squares is 1537.8. As calculated in Lecture 9 Problem 3b, $SS_{Treatment} = 543.6$. As expected, these values are not equal.

Lecture 11 Problem 2:

In this problem, we are considering the means model, which is defined as $y_{ij} = \mu_i + \epsilon_{ij}$, where $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$. With this model in mind, the main goal of this problem is to find the maximum likelihood estimates of μ_i and σ_ϵ .

To start off, we need to know the distribution of y_{ij} . Since $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$, it follows that $y_{ij} \sim N(\mu_i, \sigma_\epsilon^2)$. Now that we have the distribution of the y_{ij} 's, if we assume that the samples are independent and identically distributed (within the treatment levels), then we can write the likelihood function as the product of the PDF of y_{ij} evaluated at the given data. Below we will simplify this likelihood function into something we can maximize.

$$\begin{aligned}
 L(\mu_i, \sigma_\epsilon) &= \prod_{i=1}^a \prod_{j=1}^n \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} e^{-\frac{1}{2} \left(\frac{y_{ij} - \mu_i}{\sigma_\epsilon} \right)^2} \\
 &= \prod_{i=1}^a \prod_{j=1}^n e^{\ln \left(\frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \right)} e^{-\frac{1}{2} \left(\frac{y_{ij} - \mu_i}{\sigma_\epsilon} \right)^2} \\
 &= \prod_{i=1}^a \prod_{j=1}^n e^{-\frac{1}{2} \left(\frac{y_{ij} - \mu_i}{\sigma_\epsilon} \right)^2 - \frac{1}{2} \ln(2\pi\sigma_\epsilon^2)} \\
 &= e^{\sum_{i=1}^a \sum_{j=1}^n \left(-\frac{1}{2} \left(\frac{y_{ij} - \mu_i}{\sigma_\epsilon} \right)^2 - \frac{1}{2} \ln(2\pi\sigma_\epsilon^2) \right)} \\
 &= e^{\sum_{i=1}^a \sum_{j=1}^n -\frac{1}{2} \left(\frac{y_{ij} - \mu_i}{\sigma_\epsilon} \right)^2 - \frac{1}{2} an \ln(2\pi\sigma_\epsilon^2)}
 \end{aligned}$$

As said in lecture, the likelihood function is maximized when the exponent of the above exponential is minimized. Thus we must take the derivative of the exponent with respect to μ_i and σ_ϵ and find the values of the two parameters that simultaneously minimizes the exponent. We will start with the derivative with respect to μ_i , this calculation is shown below.

$$\begin{aligned}
 \frac{\partial}{\partial \mu_i} [\text{exponent}] &= \frac{\partial}{\partial \mu_i} \left[\sum_{i=1}^a \sum_{j=1}^n -\frac{1}{2} \left(\frac{y_{ij} - \mu_i}{\sigma_\epsilon} \right)^2 - \frac{1}{2} an \ln(2\pi\sigma_\epsilon^2) \right] \\
 &= \sum_{i=1}^a \sum_{j=1}^n \frac{-1}{2} \frac{\partial}{\partial \mu_i} \left(\frac{y_{ij} - \mu_i}{\sigma_\epsilon} \right)^2 - \frac{1}{2} an \frac{\partial}{\partial \mu_i} \ln(2\pi\sigma_\epsilon^2) \\
 &= \sum_{j=1}^n \frac{-y_{ij} + \mu_i}{\sigma_\epsilon} \cdot \frac{-1}{\sigma_\epsilon^2} \\
 &= \sum_{j=1}^n \frac{y_{ij} - \mu_i}{\sigma_\epsilon^3}
 \end{aligned}$$

Equating this above equation to zero and solving for μ_i we obtain the following.

$$\begin{aligned}
 0 &= \sum_{j=1}^n \frac{y_{ij} - \mu_i}{\sigma_\epsilon^3} \\
 &= \sum_{j=1}^n (y_{ij} - \mu_i) \\
 &= y_{i.} - n\mu_i \\
 n\mu_i &= y_{i.} \\
 \hat{\mu}_i &= \bar{y}_{i.}
 \end{aligned}$$

Hence as we can see above, our MLE for μ_i is $\hat{\mu}_i = \bar{y}_{i.}$. We will now find the MLE for σ_ϵ , as shown below.

$$\begin{aligned}\frac{\partial}{\partial \sigma_\epsilon}[\text{exponent}] &= \frac{\partial}{\partial \sigma_\epsilon} \left[\sum_{i=1}^a \sum_{j=1}^n -\frac{1}{2} \left(\frac{y_{ij} - \mu_i}{\sigma_\epsilon} \right)^2 - \frac{1}{2} an \ln(2\pi\sigma_\epsilon^2) \right] \\ &= \sum_{i=1}^a \sum_{j=1}^n \frac{-1}{2} \frac{\partial}{\partial \sigma_\epsilon} \left(\frac{y_{ij} - \mu_i}{\sigma_\epsilon} \right)^2 - \frac{1}{2} an \frac{\partial}{\partial \sigma_\epsilon} \ln(2\pi\sigma_\epsilon^2) \\ &= \sum_{i=1}^a \sum_{j=1}^n \frac{(y_{ij} - \mu_i)^2}{\sigma_\epsilon^3} - \frac{an}{\sigma_\epsilon}\end{aligned}$$

Equating this above equation to zero and solving for σ_ϵ we obtain the following.

$$\begin{aligned}0 &= \sum_{i=1}^a \sum_{j=1}^n \frac{(y_{ij} - \mu_i)^2}{\sigma_\epsilon^3} - \frac{an}{\sigma_\epsilon} \\ \frac{an}{\sigma_\epsilon} &= \sum_{i=1}^a \sum_{j=1}^n \frac{(y_{ij} - \mu_i)^2}{\sigma_\epsilon^3} \\ an\sigma_\epsilon^2 &= \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \mu_i)^2 \\ \sigma_\epsilon^2 &= \frac{\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \mu_i)^2}{an} \\ &= \frac{\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \mu_i)^2}{N} \\ &= \frac{\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2}{N} \quad (\text{Since } \hat{\mu}_i = \bar{y}_{i.}) \\ &= \frac{SS_E}{N} \\ &\simeq MS_E\end{aligned}$$

Hence as we can see above, our MLE for σ_ϵ is $\hat{\sigma}_\epsilon = \sqrt{MS_E}$, which is the same for the effects model.

Lecture 11 Problem 3a:

In this problem we will solve for $\hat{\mu}$ and $\hat{\tau}_i$ in the normal equations derived from the effects model, $y_{ij} = \mu + \tau_i + \epsilon_{ij}$. Furthermore, in solving these equations simultaneously, we will use the constraint that $\hat{\tau}_\cdot = 0$, and the assumption that $a = 3$. These normal equations are shown below.

$$y_{..} - 3n\hat{\mu} - n\hat{\tau}_\cdot = 0$$

$$y_{i.} - n\hat{\mu} - n\hat{\tau}_i = 0$$

We will start with the by solving the first equation below.

$$\begin{aligned}0 &= y_{..} - 3n\hat{\mu} - n\hat{\tau}_\cdot \\ &= y_{..} - 3n\hat{\mu} \quad (\text{Since } \hat{\tau}_\cdot = 0) \\ 3n\hat{\mu} &= y_{..} \\ \hat{\mu} &= \frac{y_{..}}{3n} \\ \hat{\mu} &= \bar{y}_{..}\end{aligned}$$

Hence, as shown above, $\hat{\mu} = \bar{y}_{..}$. We will now solve the second equation below.

$$\begin{aligned}
0 &= y_{i.} - n\hat{\mu} - n\hat{\tau}_i \\
&= y_{i.} - n\bar{y}_{..} - n\hat{\tau}_i \quad (\text{Since } \hat{\mu} = \bar{y}_{..}) \\
n\hat{\tau}_i &= y_{i.} - n\bar{y}_{..} \\
\hat{\tau}_i &= \frac{y_{i.}}{n} - \bar{y}_{..} \\
\hat{\tau}_i &= \bar{y}_{i.} - \bar{y}_{..}
\end{aligned}$$

Hence, as shown above, $\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$.

Lecture 11 Problem 3b:

In this problem, using the results from part a, we will estimate the contrasts $\tau_1 - \tau_2$, $\tau_1 - \tau_3$, and $\tau_2 - \tau_3$. These three estimates are calculated and shown below.

$$\begin{aligned}
\hat{\tau}_1 - \hat{\tau}_2 &= \bar{y}_{1.} - \bar{y}_{..} - (\bar{y}_{2.} - \bar{y}_{..}) \\
&= \bar{y}_{1.} - \bar{y}_{..} - \bar{y}_{2.} + \bar{y}_{..} \\
&= \bar{y}_{1.} - \bar{y}_{2.}
\end{aligned}$$

Hence we can see that the estimate for $\tau_1 - \tau_2$ is $\bar{y}_{1.} - \bar{y}_{2.}$.

$$\begin{aligned}
\hat{\tau}_1 - \hat{\tau}_3 &= \bar{y}_{1.} - \bar{y}_{..} - (\bar{y}_{3.} - \bar{y}_{..}) \\
&= \bar{y}_{1.} - \bar{y}_{..} - \bar{y}_{3.} + \bar{y}_{..} \\
&= \bar{y}_{1.} - \bar{y}_{3.}
\end{aligned}$$

Hence we can see that the estimate for $\tau_1 - \tau_3$ is $\bar{y}_{1.} - \bar{y}_{3.}$.

$$\begin{aligned}
\hat{\tau}_2 - \hat{\tau}_3 &= \bar{y}_{2.} - \bar{y}_{..} - (\bar{y}_{3.} - \bar{y}_{..}) \\
&= \bar{y}_{2.} - \bar{y}_{..} - \bar{y}_{3.} + \bar{y}_{..} \\
&= \bar{y}_{2.} - \bar{y}_{3.}
\end{aligned}$$

Hence we can see that the estimate for $\tau_2 - \tau_3$ is $\bar{y}_{2.} - \bar{y}_{3.}$.

Lecture 11 Problem 3c:

In this problem we will solve for $\hat{\mu}$ and $\hat{\tau}_i$ in the normal equations derived from the effects model, $y_{ij} = \mu + \tau_i + \epsilon_{ij}$. Furthermore, in solving these equations simultaneously, we will use the constraint that $\hat{\tau}_3 = 0$. These normal equations are shown below.

$$\begin{aligned}
y_{..} - 3n\hat{\mu} - n\hat{\tau}_3 &= 0 \\
y_{1.} - n\hat{\mu} - n\hat{\tau}_1 &= 0 \\
y_{2.} - n\hat{\mu} - n\hat{\tau}_2 &= 0 \\
y_{3.} - n\hat{\mu} - n\hat{\tau}_3 &= 0
\end{aligned}$$

We will start by solving the fourth of these equations below.

$$\begin{aligned}
0 &= y_{3.} - n\hat{\mu} - n\hat{\tau}_3 \\
&= y_{3.} - n\hat{\mu} \quad (\text{Since } \hat{\tau}_3 = 0) \\
n\hat{\mu} &= y_{3.} \\
\hat{\mu} &= \frac{y_{3.}}{n} \\
\hat{\mu} &= \bar{y}_{3.}
\end{aligned}$$

Hence as shown above $\hat{\mu} = \bar{y}_{3.}$. We will now solve the second and third equations below.

$$\begin{aligned}
0 &= y_{1.} - n\hat{\mu} - n\hat{\tau}_1 \\
&= y_{1.} - n\bar{y}_{3.} - n\hat{\tau}_1 \quad (\text{Since } \hat{\mu} = \bar{y}_{3.}) \\
n\hat{\tau}_1 &= y_{1.} - n\bar{y}_{3.} \\
\hat{\tau}_1 &= \frac{y_{1.}}{n} - \bar{y}_{3.} \\
&= \bar{y}_{1.} - \bar{y}_{3.}
\end{aligned}$$

Hence as shown above, $\hat{\tau}_1 = \bar{y}_{1.} - \bar{y}_{3.}$.

$$\begin{aligned}
0 &= y_{2.} - n\hat{\mu} - n\hat{\tau}_2 \\
&= y_{2.} - n\bar{y}_{3.} - n\hat{\tau}_2 \quad (\text{Since } \hat{\mu} = \bar{y}_{3.}) \\
n\hat{\tau}_2 &= y_{2.} - n\bar{y}_{3.} \\
\hat{\tau}_2 &= \frac{y_{2.}}{n} - \bar{y}_{3.} \\
&= \bar{y}_{2.} - \bar{y}_{3.}
\end{aligned}$$

Hence as shown above, $\hat{\tau}_2 = \bar{y}_{2.} - \bar{y}_{3.}$. We will now solve the last equation below, if all of our estimates are correct, we should get $0 = 0$.

$$\begin{aligned}
0 &= y_{..} - 3n\hat{\mu} - n\hat{\tau} \\
&= y_{..} - 3n\bar{y}_{3.} - n(\hat{\tau}_1 + \hat{\tau}_2 + \hat{\tau}_3) \quad (\text{Since } \hat{\mu} = \bar{y}_{3.} \text{ and } \hat{\tau} = \sum_{i=1}^3 \hat{\tau}_i) \\
&= y_{..} - 3n\bar{y}_{3.} - n(\bar{y}_{1.} - \bar{y}_{3.}) - n(\bar{y}_{2.} - \bar{y}_{3.}) \quad (\text{Since } \hat{\tau}_3 = 0, \hat{\tau}_1 = \bar{y}_{1.} - \bar{y}_{3.}, \text{ and } \hat{\tau}_2 = \bar{y}_{2.} - \bar{y}_{3.}) \\
&= y_{..} - 3n\bar{y}_{3.} - n\bar{y}_{1.} - n\bar{y}_{2.} + n\bar{y}_{3.} \\
&= y_{..} - n\bar{y}_{3.} - n\bar{y}_{1.} - n\bar{y}_{2.} \\
&= \frac{y_{..}}{n} - \bar{y}_{3.} - \bar{y}_{1.} - \bar{y}_{2.} \\
&= \bar{y}_{..} - \bar{y}_{3.} - \bar{y}_{1.} - \bar{y}_{2.} \\
&= \bar{y}_{..} - \bar{y}_{..} \\
&= 0
\end{aligned}$$

Hence we have shown that $\hat{\mu} = \bar{y}_{3.}$, $\hat{\tau}_1 = \bar{y}_{1.} - \bar{y}_{3.}$, and $\hat{\tau}_2 = \bar{y}_{2.} - \bar{y}_{3.}$.

Lecture 11 Problem 3d:

In this problem, using the results from part c we will estimate the contrasts $\tau_1 - \tau_2$, $\tau_1 - \tau_3$, and $\tau_2 - \tau_3$. These three estimates are calculated and shown below.

$$\begin{aligned}
\hat{\tau}_1 - \hat{\tau}_2 &= \bar{y}_{1.} - \bar{y}_{3.} - (\bar{y}_{2.} - \bar{y}_{3.}) \\
&= \bar{y}_{1.} - \bar{y}_{3.} - \bar{y}_{2.} + \bar{y}_{3.} \\
&= \bar{y}_{1.} - \bar{y}_{2.}
\end{aligned}$$

Hence we can see that the estimate for $\tau_1 - \tau_2$ is $\bar{y}_{1.} - \bar{y}_{2.}$.

$$\begin{aligned}
\hat{\tau}_1 - \hat{\tau}_3 &= \bar{y}_{1.} - \bar{y}_{3.} - 0 \\
&= \bar{y}_{1.} - \bar{y}_{3.}
\end{aligned}$$

Hence we can see that the estimate for $\tau_1 - \tau_3$ is $\bar{y}_{1.} - \bar{y}_{3.}$.

$$\begin{aligned}
\hat{\tau}_2 - \hat{\tau}_3 &= \bar{y}_{2.} - \bar{y}_{3.} - 0 \\
&= \bar{y}_{2.} - \bar{y}_{3.}
\end{aligned}$$

Hence we can see that the estimate for $\tau_2 - \tau_3$ is $\bar{y}_{2.} - \bar{y}_{3.}$.

Lecture 11 Problem 3e:

In this problem we will estimate $\mu + \tau_1$, $2\tau_1 - \tau_2 - \tau_3$, and $\mu + \tau_1 + \tau_2$, using the solutions from both parts a and c. Furthermore, given that contrasts that do not depend on the choice of the constraint are called estimable, we will also decide which of the three contrasts are estimable.

We will start with the estimate of $\mu + \tau_1$ using the solutions from part a, then part c. These calculations are shown below.

$$\begin{aligned}\hat{\mu} + \hat{\tau}_1 &= \bar{y}_{..} + \bar{y}_{1.} - \bar{y}_{..} \\ &= \bar{y}_{1.}.\end{aligned}$$

Hence the estimate for $\mu + \tau_1$ using the solutions from part a is $\bar{y}_{1.}$.

$$\begin{aligned}\hat{\mu} + \hat{\tau}_1 &= \bar{y}_{3.} + \bar{y}_{1.} - \bar{y}_{3.} \\ &= \bar{y}_{1.}.\end{aligned}$$

Hence the estimate for $\mu + \tau_1$ using the solutions from part c is $\bar{y}_{1.}$.

We will now compute with the estimate of $2\tau_1 - \tau_2 - \tau_3$ using the solutions from part a, then part c. These calculations are shown below.

$$\begin{aligned}2\hat{\tau}_1 - \hat{\tau}_2 - \hat{\tau}_3 &= 2(\bar{y}_{1.} - \bar{y}_{..}) - (\bar{y}_{2.} - \bar{y}_{..}) - (\bar{y}_{3.} - \bar{y}_{..}) \\ &= 2\bar{y}_{1.} - 2\bar{y}_{..} - \bar{y}_{2.} + \bar{y}_{..} - \bar{y}_{3.} + \bar{y}_{..} \\ &= 2\bar{y}_{1.} - \bar{y}_{2.} - \bar{y}_{3.}.\end{aligned}$$

Hence the estimate for $2\tau_1 - \tau_2 - \tau_3$ using the solutions from part a is $2\bar{y}_{1.} - \bar{y}_{2.} - \bar{y}_{3.}$.

$$\begin{aligned}2\hat{\tau}_1 - \hat{\tau}_2 - \hat{\tau}_3 &= 2(\bar{y}_{1.} - \bar{y}_{3.}) - (\bar{y}_{2.} - \bar{y}_{3.}) - 0 \\ &= 2\bar{y}_{1.} - 2\bar{y}_{3.} - \bar{y}_{2.} + \bar{y}_{3.} \\ &= 2\bar{y}_{1.} - \bar{y}_{2.} - \bar{y}_{3.}.\end{aligned}$$

Hence the estimate for $2\tau_1 - \tau_2 - \tau_3$ using the solutions from part c is $2\bar{y}_{1.} - \bar{y}_{2.} - \bar{y}_{3.}$.

Lastly, we will compute with the estimate of $\mu + \tau_1 + \tau_2$ using the solutions from part a, then part c. These calculations are shown below.

$$\begin{aligned}\hat{\mu} + \hat{\tau}_1 + \hat{\tau}_2 &= \bar{y}_{..} + \bar{y}_{1.} - \bar{y}_{..} + \bar{y}_{2.} - \bar{y}_{..} \\ &= \bar{y}_{1.} + \bar{y}_{2.} - \bar{y}_{..}.\end{aligned}$$

Hence the estimate for $\mu + \tau_1 + \tau_2$ using the solution from part a is $\bar{y}_{1.} + \bar{y}_{2.} - \bar{y}_{..}$.

$$\begin{aligned}\hat{\mu} + \hat{\tau}_1 + \hat{\tau}_2 &= \bar{y}_{3.} + \bar{y}_{1.} - \bar{y}_{3.} + \bar{y}_{2.} - \bar{y}_{3.} \\ &= \bar{y}_{1.} + \bar{y}_{2.} - \bar{y}_{3.}.\end{aligned}$$

Hence the estimate for $\mu + \tau_1 + \tau_2$ using the solution from part c is $\bar{y}_{1.} + \bar{y}_{2.} - \bar{y}_{3.}$.

Hence, as shown above, $\mu + \tau_1$ and $2\tau_1 - \tau_2 - \tau_3$ are estimable, while $\mu + \tau_1 + \tau_2$ is not estimable.