17.8 Refer to problem 16.7.

(a) Prepare a line plot of the estimated factor level means $\overline{Y}_{i\cdot}$. What does this plot suggest regressing the effect of the level of research and development expenditures on mean productivity improvement?

Answer:

We use the ANOVA model

$$Y_{ij} = \mu_i + \varepsilon_{ij} \tag{1}$$

Where Y_{ij} is the value of the response variable in the jth trial for the ith factor level or treatment, μ_i are the parameters, and ε_{ij} are independent $N(0, \sigma^2)$.

The fitted value for observation Y_{ij}, denoted by $\hat{Y}_{ij}$ for regression models, is simply the corresponding factor level sample mean here:

$$\hat{Y}_{ij} = \overline{Y}_{i\cdot}$$

In our case, the estimated factor level means are:

$$\begin{aligned} \text{low} &= \hat{Y}_{1j} = \overline{Y}_{1\cdot} = 6.878 \\ \text{moderate} &= \hat{Y}_{2j} = \overline{Y}_{2\cdot} = 8.133 \\ \text{high} &= \hat{Y}_{3j} = \overline{Y}_{3\cdot} = 9.2 \end{aligned}$$

where $\overline{Y}_{i\cdot}$ represents the mean of the ith factor level.

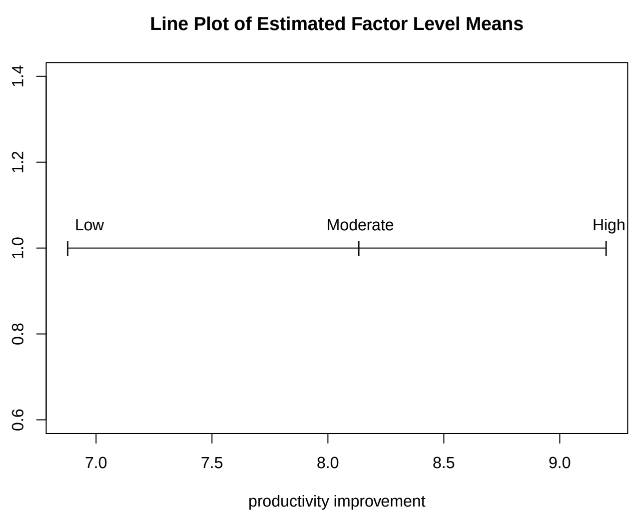

Figure 1: lineplot

(b) Estimate the mean productivity improvement for firms with high research and development expenditures levels: use a 95 percent confidence interval.

Answer: Our $1-\alpha$ confidence limits for μ_{high} are

$$\overline{Y}_3 \pm t(1-\alpha/2, n_t - r)s\left\{\overline{Y}_3\right\}$$

where in our case $n_T = 27$ and $r = 3$ because we have 3 factor levels. So

$$t(.975, 24) = 2.06$$

We note that

$$s^2\left\{\overline{Y}_3\right\} = \frac{\text{MSE}}{n_3} = \frac{0.6401}{6} \Rightarrow s\left\{\overline{Y_3}\right\} = 0.326$$

Table 1: 16.7 ANOVA table

Response	DF	Sum Squares	Mean squares	F Value	Pr(>F)
Development	2	20.125	10.06	15.72	4.331e-05
Residuals	24	15.362	$\boxed{0.6401}$		
Total	26	35.487			

We found the MSE from the ANOVA table, the sum of variance explained due to error. Therefore, our interval for the high research, controlling α risk at 0.05, is

$$9.20 \pm 2.06 \cdot 0.326$$

(c) Obtain a 95 percent confidence interval for $D = \mu_2 - \mu_1$. Interpret your interval estimate.

Answer: The difference between two factor level means is called a pairwise comparison. A point estimator of $D = \mu_i - \mu_{i'}$, denoted by $\hat{D}$ is:

$$\hat{D} = \overline{Y}_i - \overline{Y}_{i'}$$

This is an unbiased point estimator. Since $\overline{Y}_i$ and $\overline{Y}_{i'}$ are independent, the variance of $\hat{D}$ follows:

$$\sigma^2\left\{\hat{D}\right\} = \sigma^2\left\{\overline{Y}_{i\cdot}\right\} + \sigma^2\left\{\overline{Y}_{i'\cdot}\right\} = \sigma^2\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)$$

The estimated variance of $\hat{D}$, denoted by $s^2\left\{\hat{D}\right\}$ is given by

$$s^2\left\{\hat{D}\right\} = \text{MSE}\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)$$

$\hat{D}$ is normally distributed because $\hat{D}$ is a linear combination of independent normal variables. Since we are under the ANOVA model assumption, SSE/σ^2 is distributed as χ^2 with $n_T - r$ degrees of freedom and is independent of $\overline{Y}_{1\cdot}, \ldots, \overline{Y}_{r\cdot}$. From the definition of a t distribution, we note that for ANOVA model

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

that

$$\frac{\hat{D} - D}{s\left\{\hat{D}\right\}} \quad \text{is distributed as } t(n_T - r)$$

Hence, the $1 - \alpha$ confidence limits for D are:

$$\hat{D} \pm t(1 - \alpha/2; n_t - r)s\left\{\hat{D}\right\}$$

Therefore, we find

$$\hat{D} = 1.256 \qquad t(.975, 24) = 2.06 \quad s^2\left\{\hat{D}\right\} = 0.353$$

which as a result means

$$1.255 \pm 0.729$$

(d) Obtain confidence intervals for all pairwise comparisons of the treatment means; use the Tukey procedure and a 90 percent family confidence coefficient. State your findings and prepare a graphic summary by underlining nonsignificant comparisons in your line plot in part (a).

Answer:

The Tukey multiple comparison confidence limits for all pairwise comparisons $D = \mu_i - \mu_{i'}$ with family confidence coefficient of at least $1-\alpha$ are as follows:

$$\hat{D} \pm Ts\left\{\hat{D}\right\}$$

where

$$T = \frac{1}{\sqrt{2}}q(1-\alpha; r, n_t - r)$$

where $q(r,v) = \frac{w}{s}$ is the studentized range, where w is the range of set of observations. q distributions given in table B.9. We note$v = 24$ and $r = 3$

$$q(0.10, 3, 24) = 3.05 \Rightarrow T = 2.16$$

We have three cases to consider:

$$\hat{D}_1 = \overline{Y}_{2\cdot} - \overline{Y}_{1\cdot} = 1.256$$
$$\hat{D}_2 = \overline{Y}_{3\cdot} - \overline{Y}_{1\cdot} = 2.32$$
$$\hat{D}_3 = \overline{Y}_{3\cdot} - \overline{Y}_{2\cdot} = 1.07$$

Similarly,

$$s\left\{\hat{D}_1\right\} = \text{MSE} \cdot \left(\frac{1}{n_2} + \frac{1}{n_3}\right) = 0.353$$
$$s\left\{\hat{D}_2\right\} = \text{MSE} \cdot \left(\frac{1}{n_2} + \frac{1}{n_3}\right) = 0.422$$
$$s\left\{\hat{D}_3\right\} = \text{MSE} \cdot \left(\frac{1}{n_2} + \frac{1}{n_3}\right) = 0.400$$

So, We have the following Tukey multiple comparison confidence limits

$$1.256 \pm 0.76 \longrightarrow 0.496 \leq D \leq 2.016$$
$$2.32 \pm 0.909 \longrightarrow 1.411 \leq D \leq 3.23$$
$$1.07 \pm 0.863 \longrightarrow 0.207 \leq D_3 \leq 1.93$$

(e) Is the Tukey procedure in part(d) the most efficient one that could be used here. Explain.
Answer: Since only pairwise comparisons are to be made, the Tukey procedure gives narrower confidence limits than the Scheffe method. The Scheffe method is

$$\hat{L} \pm Ss\left\{\hat{L}\right\}$$

where $S^2 = (r-1)F(1\alpha, r-1, n_T - r)$. In our example,

$$S = \sqrt{(3-1)F(0.9, 2, 24)} = 2.25$$

In comparison with the Bonferroni method,

$$\hat{L} \pm Bs\left\{\hat{L}\right\}$$

where for g linear combinations of L (in this example 3),

$$B = t(1-\alpha/2g; n_t - r) = 2.26$$

The T value from the Tukey is lower than both of these.

17.10 Refer to problem 16.9.

(a) Prepare a line plot of the estimated factor level means $\overline{Y}_{i\cdot}$. What does this plot suggest about the effect of prior physical fitness on the mean time required in therapy?

Answer: Fitted values

$$\begin{aligned} \text{below average} &= \hat{Y}_{1j} = \overline{Y}_{1\cdot} = 38 \\ \text{average} &= \hat{Y}_{2j} = \overline{Y}_{2\cdot} = 32 \\ \text{above average} &= \hat{Y}_{3j} = \overline{Y}_{3\cdot} = 24 \end{aligned}$$

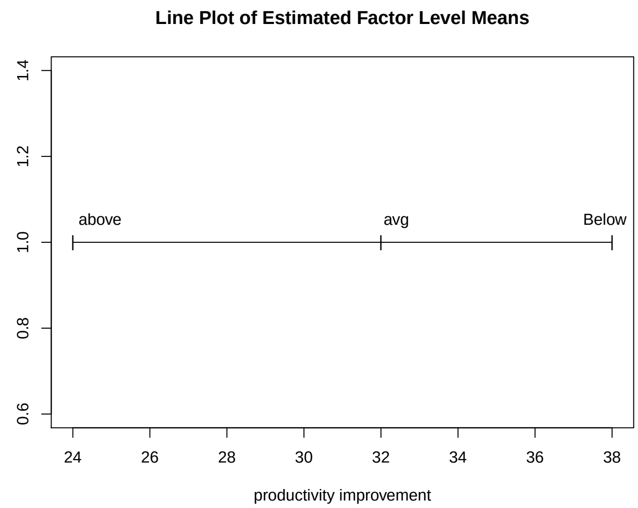

Figure 2: line plot

(b) Estimate with a 99 percent confidence interval the mean number of days required in therapy for persons of average physical fitness.

Answer: Our $1-\alpha$ confidence limits for μ_{high} are

$$\overline{Y}_2 \pm t(1-\alpha/2, n_t - r)s\left\{\overline{Y}_2\right\}$$

where in our case $n_T = 24$ and $r = 3$ because we have 3 factor levels. There are 10 people of average fitness. So

$$t(.995, 21) = 2.83$$

We note that

$$s^2\left\{\overline{Y}_2\right\} = \frac{\text{MSE}}{n_2} = \frac{19.81}{10} \Rightarrow s\left\{\overline{Y}_2\right\} = 1.40$$

We found the MSE from the ANOVA table, the sum of variance explained due to error.

Table 2: ANOVA table

Response	DF	Sum Squares	Mean squares	F Value	Pr(>F)
Fitness	2	672	336.00	16.962	4.129e-05
Residuals	21	416	$\boxed{19.81}$		
Total	23	1088			

Therefore, our interval for the high research, controlling α risk at 0.05, is

$$32 \pm 2.83 \cdot 1.40 \longrightarrow 28.04 \leq Y_2 \leq 35.96$$

(c) Obtain confidence intervals for $D_1 = \mu_2 - \mu_3$ and $D_2 = \mu_1 - \mu_2$: use the Bonferroni procedure with a 95 percent family confidence coefficient. Interpret your results.

Answer: We have three cases to consider:

$$\hat{D}_1 = \overline{Y}_{1.} - \overline{Y}_{2.} = 6$$
$$\hat{D}_2 = \overline{Y}_{2.} - \overline{Y}_{3.} = 8$$

The value of B is 2.41 ($B = t(1 - .05/(2g), 24 - 3)$), where $g = 2$ here because we have two linear combos. We find that $s\left\{\hat{D}_1\right\} = 2.11$ and $s\left\{\hat{D}_2\right\} = 2.298$, from

$$s\left\{\hat{D}_1\right\} = \sqrt{\text{MSE} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$
$$s\left\{\hat{D}_2\right\} = \sqrt{\text{MSE} \cdot \left(\frac{1}{n_2} + \frac{1}{n_3}\right)}$$

We now do

$$\hat{D} \pm Bs\left\{\hat{D}\right\}$$

Therefore,

$$\hat{D}_1 \pm 2.41 \cdot 2.11 \qquad 6 \pm 2.41 \cdot 2.11 = 0.915 \leq D_1 \leq 11.09$$
$$\hat{D}_2 \pm 2.41 \cdot 2.298 \qquad 8 \pm 2.41 \cdot 2.3 = 2.46 \leq D_2 \leq 13.54$$

(d) Would the Tukey procedure have been more efficient?

Answer: No. We need to find the T value, found from $q(0.95, 3, 21)$ where q is the studentized range distribution, which is found from table B.9 ($\nu = 21$ and $r = 3$) and is approximately 3.08. This is greater than the B value from the Bonferroni procedure.

(e) If the researcher also wished to estimate $D_3 = \mu_1 - \mu_3$, still with a 95 percent family confidence coefficient, would the B multiple in part (c) need to be modified? Would this also be the case if the Tukey procedure had been employed?

Answer: B would change because g would now equal 3. T wouldn't change, because α, n_T, and r stay constant.

(f) Test for all pairs of factor level means whether or not they differ: use the Tukey procedure with $\alpha = 0.05$. Set up groups of factor levels whose means do not differ.

Answer: We have three cases to consider: (we change $\hat{L}_2$ to $\hat{L}_3$ to stay consistent with notation from question 1)

$$\hat{D}_1 = \overline{Y}_{1.} - \overline{Y}_{2.} = 6$$
$$\hat{D}_2 = \overline{Y}_{1.} - \overline{Y}_{3.} = 14$$
$$\hat{D}_3 = \overline{Y}_{2.} - \overline{Y}_{3.} = 8$$

We note that $n_1 = 8$, $n_2 = 10$ and $n_3 = 6$, and MSE is 19.81. Therefore, the standard errors

$$s\left\{\hat{D}_1\right\} = \sqrt{\text{MSE} \cdot \left(\frac{1}{n_2} + \frac{1}{n_3}\right)} = 2.11$$
$$s\left\{\hat{D}_2\right\} = \sqrt{\text{MSE} \cdot \left(\frac{1}{n_2} + \frac{1}{n_3}\right)} = 2.40$$
$$s\left\{\hat{D}_3\right\} = \sqrt{\text{MSE} \cdot \left(\frac{1}{n_2} + \frac{1}{n_3}\right)} = 2.298$$

$$2.21 \le D_1 \le 13.79$$
$$7.94 \le D_2 \le 20.06$$
$$0.679 \le D_3 \le 11.32$$

some relevant R-code.

```
data169 = read.table("http://www.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/data/textdatasets/
    KutnerData/Chapter%2016%20Data%20Sets/CH16PR09.txt")

data169$Recover = data169$V1
data169$Fitness = data169$V2
data169<-data169%>%
select(Recover, Fitness)


data169$Fitness[data169$Fitness ==1] = "Below Average"
data169$Fitness[data169$Fitness ==2] = "Average"
data169$Fitness[data169$Fitness ==3] = "Above Average"

anovastuff169=aov(Recover~Fitness,data=data169)

anovatable169=anova(anovastuff169)

TukeyHSD(anovastuff169)
```

17.13 Refer to problem 16.12

(a) Prepare an interval plot of the estimated factor level means $\overline{Y}_{i\cdot}$, where the intervals correspond to the confidence limits in (17.7) with $\alpha = 0.10$. What does this plot suggest about the variation in the mean time lapses for the five agents?

Answer: These are the mean time lapses for the five agents.

$$\text{agent } 1 = \hat{Y}_{1j} = \overline{Y}_1 = 24.55$$
$$\text{agent } 2 = \hat{Y}_{2j} = \overline{Y}_2 = 22.55$$
$$\text{agent } 3 = \hat{Y}_{3j} = \overline{Y}_3 = 11.75$$
$$\text{agent } 4 = \hat{Y}_{4j} = \overline{Y}_3 = 14.8$$
$$\text{agent } 5 = \hat{Y}_{5j} = \overline{Y}_3 = 30.1$$

And the anova table (17.7) tells use the confidence limits for μ_i are

Table 3: ANOVA table

Response	DF	Sum Squares	Mean squares	F Value	Pr(>F)
Agents	4	4430.1	1107.536	147.23	7.98 e-40
Residuals	95	714.65	$\boxed{7.52}$		
Total	99	5144.75			

$$\overline{Y}_i \pm t(1-\alpha/2; n_T - r)s\left\{\overline{Y}_i\right\} \tag{2}$$

The t-value here is $t(.95, 100-5) = 1.66$ Recall,

$$s\left\{\overline{Y}_{i\cdot}\right\} = \sqrt{\frac{\text{MSE}}{n_i}} = \sqrt{7.52/20} = 0.613$$

because each factor level has the same number of observations.
Therefore, we see

$$\overline{Y}_1 \pm 1.66 \cdot 0.613 \longrightarrow 23.53 \leq \mu_1 \leq 25.57$$
$$\overline{Y}_2 \pm 1.66 \cdot 0.613 \longrightarrow 21.53 \leq \mu_2 \leq 23.57$$
$$\overline{Y}_3 \pm 1.66 \cdot 0.613 \longrightarrow 10.73 \leq \mu_3 \leq 12.77$$
$$\overline{Y}_4 \pm 1.66 \cdot 0.613 \longrightarrow 13.78 \leq \mu_4 \leq 15.82$$
$$\overline{Y}_5 \pm 1.66 \cdot 0.613 \longrightarrow 29.08 \leq \mu_4 \leq 31.12$$

Therefore, the variances are all the same, but the means differ for all but means 1 and means 2. Figure 3.

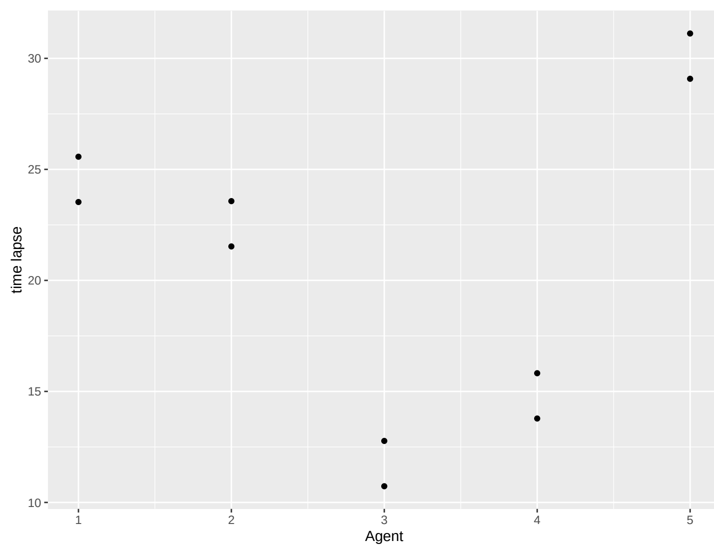

Figure 3: Plot 1

(b) Test for all pairs of factor level means whether or not they differ; use the Tukey procedure with $\alpha = 0.10$. Set up groups of factor levels whose means do not differ. Use a paired comparison plot to summarize the results.

Answer: We have $\binom{5}{2}$ pairs of factor level means. The hypothesis tests are of the form (skipping most of them)

$$H_0 : \mu_1 = \mu_2$$
$$H_a : \mu_1 \neq \mu_2$$

We repeat this test for the ten combos. The values are

$$\hat{D}_1 = \mu_1 - \mu_2 = 24.55 - 22.55 = 2 \qquad \hat{D}_6 = \mu_2 - \mu_4 = 22.55 - 14.8 = 7.75$$
$$\hat{D}_2 = \mu_1 - \mu_3 = 24.55 - 11.75 = 12.8 \qquad \hat{D}_7 = \mu_2 - \mu_5 = 22.55 - 30.1 = -7.55$$
$$\hat{D}_3 = \mu_1 - \mu_4 = 24.55 - 14.8 = 9.75 \qquad \hat{D}_8 = \mu_3 - \mu_4 = 11.75 - 14.8 = -3.05$$
$$\hat{D}_4 = \mu_1 - \mu_4 = 24.55 - 30.1 = -5.55 \qquad \hat{D}_9 = \mu_3 - \mu_5 = -18.35$$
$$\hat{D}_5 = \mu_2 - \mu_3 = 22.55 - 11.75 = 10.8 \qquad \hat{D}_{10} = \mu_4 - \mu_5 = -15.3$$

Now, the test statistic we use to conduct the pairwise test is

$$q^* = \frac{\sqrt{2}\hat{D}}{s\left\{\hat{D}\right\}}$$

If $|q^*| \leq q(1-\alpha; r; n_T - r) = q(.9, 5, 95) \approx 3.54$, found from table B.8 with $\nu = 95$ and $r = 5$. Recall, (using the MSE from the anova table in part(a) and also that there are 20 samples for each agent)

$$s\left\{\hat{D}\right\} = \sqrt{s^2\left\{\overline{Y}_{i\cdot}\right\} + s^2\left\{\overline{Y}_{i'\cdot}\right\}} = \sqrt{\text{MSE}\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)} = \sqrt{7.52 \cdot \left(\frac{1}{20} + \frac{1}{20}\right)} = 0.867$$

Now, we perform our test, whether or not $|q^*| \leq 3.54$.

Table 4: Hypothesis tests

Test	q^*
1	$\sqrt{2}\frac{2}{.867}$=3.26
2	$\sqrt{2}\frac{12.8}{.867}$=20.86
3	$\sqrt{2}\frac{9.75}{.867} = 15.89$
4	$\sqrt{2}\frac{-5.55}{.867} = -9.04$
5	$\sqrt{2}\frac{10.8}{.867} = 17.60$
6	$\sqrt{2}\frac{7.75}{.867} = 12.63$
7	$\sqrt{2}\frac{-7.55}{.867} = -12.31$
8	$\sqrt{2}\frac{3.05}{.867} = 4.97$
9	$\sqrt{2}\frac{-18.35}{.867} = -29.91$
10	$\sqrt{2}\frac{-15.3}{.867} = -24.94$

Therefore, we conclude the null hypothesis for the **1st pairwise test**, and conclude h_a for the rest. We now summarize the family of pairwise tests by setting up groups of factor levels whose means do no differ according to the single degree of freedom test. Therefore, we have 4 groups (because mean 1 and 2 do not differ at $\alpha = .1$)

Table 5: Groups

Group 1		Group 2	Group 3	Group 4
Agent 1	$\hat{Y}_{1\cdot} = 24.55$	Agent 3 $\hat{Y}_{3\cdot}$=14.8	Agent 4 $\hat{Y}_{4\cdot}$=30.1	Agent 5 $\hat{Y}_{5\cdot}$=11.75
Agent 2	$\hat{Y}_{2\cdot}$=22.55			

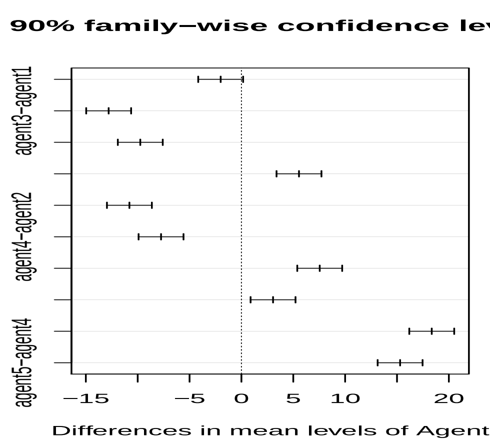

Figure 4: Tukey plot.

(c) Construct a 90 percent confidence interval for the mean time lapse for agent 1.

Answer: Our $1-\alpha$ confidence limits for μ_{high} are (noting $t(.95, 95) = 1.66$)

$$\overline{Y}_1 \pm t(1-\alpha/2, n_t - r)s\left\{\overline{Y}_1\right\} = 24.55 \pm 1.66 \cdot 0.613 \rightarrow 23.53 \leq \mu_1 \leq 25.57$$

(d) Obtain a 90 percent confidence interval for $D = \mu_2 - \mu_1$. Interpret your interval estimate.

Answer: Recall, $s\left\{\hat{D}_i\right\} = 0.867$. Therefore, since $\hat{D} = -2$, noting that $t(1-\alpha; n_t - r) \approx 1.66$ and we obtain a 90 percent confidence interval

$$-2 \pm 1.66 \cdot 0.867 \rightarrow -3.44 \leq D \leq 0.561$$

(e) The marketing director wishes to compare the mean time lapses for agents 1,3, and 5. Obtain confidence intervals for all pairwise comparisons among these three treatment means; use the Bonferroni procedure with a 90 percent family confidence coefficient. Interpret your results and present a graphic summary by preparing a line plot of the estimated factor level means with non significant differences underlined. Do your conclusions agree with those in part (a)?

Answer: We have three cases to consider:

$$\hat{D}_1 = \overline{Y}_{1\cdot} - \overline{Y}_{3\cdot} = 12.8$$
$$\hat{D}_2 = \overline{Y}_{1\cdot} - \overline{Y}_{5\cdot} = -5.55$$
$$\hat{D}_3 = \overline{Y}_{3\cdot} - \overline{Y}_{5\cdot} = -18.35$$

The value of B is 2.16 since $B = t(1 - .05/(2g), 100-5)$, where $g = 3$ here because we have two linear combos.

We now do

$$\hat{D} \pm Bs\left\{\hat{D}\right\}$$

Therefore,

$$\begin{aligned} &\hat{D}_1 \pm 2.16 \cdot 0.867 \qquad && 10.93 \leq D_1 \leq 14.67 \\ &\hat{D}_2 \pm 2.16 \cdot 0.867 && -7.422 \leq D_2 \leq -3.67 \\ &\hat{D}_3 \pm 2.16 \cdot 0.867 && -20.22 \leq D_3 \leq -16.48 \end{aligned}$$

(f) Would the Tukey procedure have been more efficient to use in part (e) than the Bonferroni procedure?

Answer: No, the Tukey value was 2.51, from part (d), which is bigger than the B value.

17.14 Refer to problem 16.7.

(a) Estimate the difference in mean productivity improvement between firms with low or moderate research and development expenditures and firms with high expenditures; use a 95 percent confidence interval. Employ an unweighted mean for the low and moderate expenditures groups. Interpret your interval estimate.

Answer: To do this, we take the mean of the low and moderate groups and then subtract the high group. Then

$$\hat{L} = \frac{\overline{Y}_{1\cdot} + \overline{Y}_{2\cdot}}{2} - \overline{Y}_{3\cdot} = -1.69$$

Note,

$$s^2\left\{\hat{L}\right\} = \text{MSE}\sum_{i=1}^{r} c_i^2/n_i$$

In our case, $r = 3$, $n_1 = 9$, $n_2 = 12$, and $n_3 = 6$, $c_1 = c_2 = 1/2$, and $c_3 = 1$. Of course, we calculated the MSE in the first question, which was 0.64. So $s\left\{\hat{L}\right\} = 0.371$. For our confidence test, at $\alpha = 0.05$, we have that $t(1-\alpha/2, n_T - r) = t(0.975, 27-3) = 2.06$. Therefore, our interval is

$$-1.69 \pm 2.06 \cdot 0.371 \longrightarrow -2.46 \leq L \leq -0.928$$

(b) The sample sizes for the three factor levels are proportional to the population sizes. The economist wishes to estimate the mean productivity gain last year for all firms in the population. Estimate this overall mean productivity improvement with a 95 percent confidence interval.

Answer: Here the weighted mean contrast will be

$$\hat{L} = \frac{n_1}{n_T}\overline{Y}_{1\cdot} + \frac{n_2}{n_T}\overline{Y}_{2\cdot} + \frac{n_3}{n_T}\overline{Y}_{3\cdot} = 7.95$$

Similarly, we can calculate $s\left\{\hat{L}\right\}$ like last question.

$$s\left\{\hat{L}\right\} = 0.154$$

Again we use the t- distribution to calculate our interval. Like in part (a), $t(0.975, 24) = 2.06$, so

$$7.95 \pm 0.154 \cdot 2.06 \longrightarrow 7.633 \le L \le 8.267$$

(c) Using the Scheffe procedure, obtain confidence intervals for the following comparisons with 90 percent family confidence coefficient:

$$\begin{aligned} D_1 &= \mu_3 - \mu_2 & D_3 &= \mu_2 - \mu_1 \\ D_2 &= \mu_3 - \mu_1 & L_1 &= \frac{\mu_1 + \mu_2}{2}\mu_3 \end{aligned}$$

Answer: The Scheffe confidence intervals for the family of contrasts L are of the form

$$\hat{L} \pm Ss\left\{\hat{L}\right\}$$

where $S^2 = (r-1)F(1-\alpha; r-1, ; n_T - r)$ and $S = 2 \cdot \sqrt{(}2.538) = 2.253$. In part (a) we calculated $\hat{L} = -1.69$ and $s\left\{\hat{L}_1\right\} = 0.371$, and in question 1 we saw

$$\begin{aligned} \hat{D}_1 &= \overline{Y}_{2\cdot} - \overline{Y}_{1\cdot} = 1.256 \\ \hat{D}_2 &= \overline{Y}_{3\cdot} - \overline{Y}_{1\cdot} = 2.32 \\ \hat{D}_3 &= \overline{Y}_{3\cdot} - \overline{Y}_{2\cdot} = 1.07 \end{aligned}$$

and

$$\begin{aligned} s\left\{\hat{D}_1\right\} &= \text{MSE} \cdot \left(\frac{1}{n_2} + \frac{1}{n_3}\right) = 0.353 \\ s\left\{\hat{D}_2\right\} &= \text{MSE} \cdot \left(\frac{1}{n_2} + \frac{1}{n_3}\right) = 0.422 \\ s\left\{\hat{D}_3\right\} &= \text{MSE} \cdot \left(\frac{1}{n_2} + \frac{1}{n_3}\right) = 0.400 \end{aligned}$$

So, our Scheffe intervals are

$$\begin{aligned} 1.256 \pm 2.253 \cdot 0353 & \qquad 0.461 \le D_1 \le 2.05 \\ 2.32 \pm 2.253 \cdot 0.422 & \qquad 2.37 \le D_2 \le 3.22 \\ 1.07 \pm 2.253 \cdot 0.40 & \qquad 0.17 \le D_3 \le 1.97 \\ -1.69 \pm 2.253 \cdot 0.371 & \qquad -2.525 \le L \le -0.855 \end{aligned}$$

17.15 Refer to problem 16.9.

(a) Estimate the contrast $L = (\mu_1 - \mu_2) - (\mu_2 - \mu_3)$ with a 99 percent confidence interval. Interpret your interval estimate.

Answer: The fitted values are Fitted values

$$\begin{aligned} \text{below average} &= \hat{Y}_{1j} = \overline{Y}_{1\cdot} = 38 \\ \text{average} &= \hat{Y}_{2j} = \overline{Y}_{2\cdot} = 32 \\ \text{above average} &= \hat{Y}_{3j} = \overline{Y}_{3\cdot} = 24 \end{aligned}$$

This is equivalently

$$\hat{L} = \overline{Y}_1 + \overline{Y}_3 - 2\overline{Y}_2 = -2$$

We calculate (from $s^2\left\{\hat{L}\right\} = \text{MSE}\sum_{i=1}^{r} c_i^2/n_i$ and the MSE from table 2

$$s\left\{\hat{L}\right\} = \sqrt{(19.81)\left(\frac{1}{8} + \frac{4}{10} + \frac{1}{6}\right)} = 3.70$$

and we use the usual t distribution to help us calculate our interval $t(.995, 24-3) = 2.83$:

$$-2 \pm 3.70 \cdot 2.83 \longrightarrow -12.48 \le L \le 8.48$$

Thus, we say with confidence coefficient of .99 that the true value of our contrast is between -12.48 and 8.48.

(b) Estimate the following comparisons using the Bonferroni procedure with 95 percent family confidence coefficient:

Answer:

$$\begin{aligned} D_1 &= \mu_1 - \mu_2 & D_3 &= \mu_2 - \mu_3 \\ D_2 &= \mu_1 - \mu_3 & L_1 &= D_1 - D_3 \end{aligned}$$

Our comparison values are

$$\begin{aligned} \hat{D}_1 &= 38 - 32 = 6 & \hat{D}_3 &= 32 - 24 = 8 \\ \hat{D}_2 &= 38 - 24 = 14 & \hat{L}_1 &= 6 - 8 = -2 \end{aligned}$$

And the standard errors are (from question 17.10 and part (a))

$$\begin{aligned} s\left\{\hat{D}_1\right\} &= 2.11 & s\left\{\hat{D}_3\right\} &= 2.298 \\ s\left\{\hat{D}_2\right\} &= 2.40 & s\left\{\hat{L}_1\right\} &= 3.7 \end{aligned}$$

Here, $g = 4$, because we have 4 linear combinations. So our Bonferroni procedure value B is $B(1 - .05/8, 24-3) = 2.73$.

Then, we find that our intervals our defined by our comparison/contrast plus/minus $Bs\left\{\hat{D}\right\}$. This yields

$$\begin{aligned} 6 \pm 2.73 \cdot 2.11 &\longrightarrow 0.240 \le D_1 \le 11.76 \\ 14 \pm 2.73 \cdot 2.4 &\longrightarrow 7.45 \le D_2 \le 20.55 \\ 8 \pm 2.73 \cdot 2.3 &\longrightarrow 1.72 \le D_3 \le 14.28 \\ -2 \pm 2.73 \cdot 3.7 &\longrightarrow -12.1 \le L_1 \le 8.10 \end{aligned}$$

(c) Would the Scheffe procedure have been more efficient?

Answer: The Scheffe value is

$$S = \sqrt{(r-1)F(1-\alpha; r-1; n_T - r)} = \sqrt{2 \cdot F(0.95; 2; 24-3)} = \sqrt{2 \cdot 3.46} = 2.63$$

Since this is less than the value of B from part (b), the Scheffe procedure would indeed have been more efficient.

17.18 Refer to problem 16.12, Agents 1 and agents 2 distribute merchandise only, agents 3 and 4 distribute cash-value coupons only, and agent 5 distributes both merchandise and coupons.

(a) Estimate the contrast:

$$L = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$

with a 90 percent confidence interval Interpret your interval estimate.

Answer: These are the mean time lapses for the five agents.

$$\text{agent } 1 = \hat{Y}_{1j} = \overline{Y}_1 = 24.55$$
$$\text{agent } 2 = \hat{Y}_{2j} = \overline{Y}_2 = 22.55$$
$$\text{agent } 3 = \hat{Y}_{3j} = \overline{Y}_3 = 11.75$$
$$\text{agent } 4 = \hat{Y}_{4j} = \overline{Y}_3 = 14.8$$
$$\text{agent } 5 = \hat{Y}_{5j} = \overline{Y}_3 = 30.1$$

Therefore,

$$\hat{L} = \frac{24.55 + 22.55}{2} - \frac{11.75 + 14.8}{2} = 10.725$$

Recall, $s\left\{\hat{L}\right\} = \sqrt{\text{MSE} \sum_{i=1}^{r} \frac{c_i^2}{n_i}}$. In our case, $c_1 = c_2 = 1/2$ and $c_3 = c_4 = -1/2$, with $c_5 = 0$. $n_i = 20$ in this example. So, we have $4 \cdot \frac{1/4}{20}$ for our sum

$$s\left\{\hat{L}\right\} = \sqrt{7.52 \cdot (\frac{1}{20})} = 0.614$$

We now calculate the appropriate value from the t distribution: $t(1 - .1/2; 100 - 5) = 1.99$, i.e. we are about 2 standard deviations (in units of $s\left\{\hat{L}\right\}$ from our true value) Therefore,

$$10.725 \pm 1.99 \cdot 0.614 \longrightarrow 9.26 \le L \le 11.29$$

Therefore, with confidence coefficient 0.90, we conclude the mean time lapse of merchandise fall below the means for cash-value coupons by some value in this range. Here is some R-code:

```
meanagent1=mean(data1612$Premium[data1612$Agent=="agent1"])
meanagent2=mean(data1612$Premium[data1612$Agent=="agent2"])
meanagent3=mean(data1612$Premium[data1612$Agent=="agent3"])
meanagent4=mean(data1612$Premium[data1612$Agent=="agent4"])
meanagent5=mean(data1612$Premium[data1612$Agent=="agent5"])
L1=(meanagent1+meanagent2)/2-(meanagent3+meanagent4)/2
sL1=sqrt(mse1612*(1/20))
tvalue=qt(.95,95)

ran=c(L1-tvalue*sL1,L1+tvalue*sL1)
```

(b) Estimate the following comparisons with 90 percent family confidence coefficient; use the Scheffe procedure:

Answer:

$$D_1 = \mu_1 - \mu_2 \qquad L_1 = \frac{\mu_1 + \mu_2}{2} - \mu_5$$
$$D_2 = \mu_3 - \mu_4 \qquad L_2 = \frac{\mu_3 + \mu_4}{2} - \mu_5$$
$$L_3 = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$

Our Scheffe procedure is:

$$\hat{L} \pm Ss\left\{\hat{L}\right\}$$

where $S^2 = (r-1)F(1-\alpha; r-1,; n_T - r)$ and $S = \sqrt{4\cdot(F(1-.1;5-1,100-5)} = \sqrt{8} = 2.83$.
We have three cases to consider:

$$\hat{D}_1 = \overline{Y}_{1.} - \overline{Y}_{2.} = 2$$
$$\hat{D}_2 = \overline{Y}_{3.} - \overline{Y}_{4.} = -3.05$$

Conveniently, the value of $s\left\{\hat{D}_i\right\}$ is constant because n_i are all constant. We calculated $s\left\{\hat{D}_i\right\} = 0.867$ in problem 17.13.
We now do

$$\hat{D} \pm Bs\left\{\hat{D}\right\}$$

Therefore,

$$\hat{D}_1 \pm 2.83\cdot 0.867 \qquad -0.45 \le D_1 \le 4.45$$
$$\hat{D}_2 \pm 2.83\cdot 0.867 \qquad -5.5 \le D_2 \le -.6$$

As for the contrasts, (we calculated $\hat{L}_3$ in part (a)):

$$\hat{L}_1 = -6.55 \longrightarrow s\left\{\hat{L}_1\right\} = 0.755$$
$$\hat{L_2} = -16.83 \longrightarrow s\left\{\hat{L}_2\right\} = 0.755$$
$$\hat{L}_3 = 10.275 \longrightarrow s\left\{\hat{L}_3\right\} = 0.613$$

We calculated in same fashion as before. Therefore, using the same Scheffe value to calculate how many deviations we are from true value, we have

$$\hat{L}_1 = -6.55 \pm 2.83\cdot 0.755 - 8.69 \le L_1 \le -4.41$$
$$\hat{L}_2 = -16.83 \pm 2.83\cdot 0.755 \longrightarrow -18.97 \le L_2 \le -14.69$$
$$\hat{L}_3 = 10.275 \pm 2.83\cdot 0.613 \longrightarrow 8.54 \le L_3 \le 12.01$$

(c) Of all premium distributions, 25 percent are handled by agent 1, 20 percent by agent 2, 20 percent by agent 3, 20 percent by agent 4, and 15 percent by agent 5. Estimate the overall mean time lapse for premium distributions with a 90 percent confidence interval.

Answer: This is similar to 17.14(b). The contrasted mean is a weighted sum:

$$\hat{L} = .25\cdot\mu_1 + .2\cdot\mu_2 + .2\cdot\mu_3 + .2\cdot\mu_4 + .15\cdot\mu_5 = 20.47$$

With a 90 percent confidence interval, we found $t(1-.1/2;100-5) = 1.99$. We calculate as before:

$$s\left\{\hat{L}\right\} = 0.278$$

So, our interval is

$$\hat{L} \pm 1.99\cdot 0.278 \longrightarrow 20.01 \le L \le 20.99$$

17.26 Refer to problem 16.12. Suppose primary interest is in estimating the following comparisons:

$$L_1 = \mu_1 - \mu_2 \qquad L_3 = \frac{\mu_1+\mu_2}{2} - \mu_5$$
$$L_2 = \mu_3 - \mu_4 \qquad L_4 = \frac{\mu_1+\mu_2}{2} - \frac{\mu_3+\mu_4}{2}$$

What would be the required sample sizes if the precision of each of the estimated comparisons is not to exceed ± 1.0 day, using the most efficient multiple comparison procedure with a 90 percent family confidence coefficient?

Answer: We use the estimated variance of an estimated contrast. Because all the sample sizes are equal, $s^2\left\{\hat{L}_i\right\} = \frac{\text{MSE}}{n}\sum_i c_i^2$. Therefore, using the MSE as our estimate of σ^2,

$$s^2\left\{\hat{L}_1\right\} = \frac{2\cdot 7.52}{n} = \frac{15.04}{n}$$
$$s^2\left\{\hat{L}_2\right\} = \frac{2\cdot 7.52}{n} = \frac{15.04}{n}$$
$$s^2\left\{\hat{L}_3\right\} = \frac{1.5\cdot 7.52}{n} = \frac{11.28}{n}$$
$$s^2\left\{\hat{L}_4\right\} = \frac{1\cdot 7.52}{n} = \frac{7.52}{n}$$

$\hat{L}_1$ (and equivalently $\hat{L}_2$) has the widest width, so that is all we need to check. We compare the multiples. We do this at $n_T = 100$, the total sample size, to give us an idea as to which comparison procedure is most efficient.

$$B = t(1-.1/(2*4), 95) = 2.27$$
$$S = \sqrt{(r-1)\cdot F(.9,4,95)} = 2.83$$
$$T = \frac{1}{\sqrt{2}}q(0.9;4,95) = 2.32$$

Relevant R-code:

```
qt(1-0.0125,95)
sqrt(4*qf(1-.1, 4, 95))
qtukey(.9,4,95)/(sqrt(2))
```

Bonferroni is the way to go. Therefore, we find n such that

$$Bs\{\hat{L}_1\} \approx\leq 1 \longrightarrow 2\cdot\sqrt{15.04} < \sqrt{n}$$

This means $n < 77.4$. We now assume $\boxed{77}$ is the correct answer. This means that $n_T = 5\cdot 77 = 385$. We thus find that the Bonferroni multiple is $t(.9875, 385) = 2.25$, which yields a precision of 0.988. Using n=76 yields 1.002, slightly outside our window.

17.27 Refer to problem 16.9. Suppose that primary interest is in comparing the below-average and above-average physical fitness groups, respectively, with the average physical fitness group. The two comparisons are of interest:

$$L_1 = \mu_1 - \mu_2 \qquad L_2 = \mu_3 - \mu_2$$

Assume that a reasonable planning value for the error standard deviation is $\sigma = 4.5$ days.

(a) It has been decided to use equal samples sizes (n) for the below average and above-average groups. If the twice the sample size (2n) were to be used for the average physical fitness group, what would be the required sample sizes if the precision of each pairwise comparison is to be ± 2.5 days, using the Bonferroni procedure and a 90 percent family confidence coefficient?

Answer: We note that for both cases, because twice the sample size is used to the average physical fitness group, μ_2, noting that $n_3 = 2n$ and $n_2 = n$, using the equation for variance of an estimated difference $\hat{L}_i$:

$$\sigma^2\left\{\hat{L}_i\right\} = \sigma^2\left(\frac{1}{n_2} + \frac{1}{n_3}\right) = \frac{3\sigma^2}{2n}$$

With $\sigma^2 = 4.5^2$ and $n = 10$, we find $n_T = 10 + 10 + 2(10) = 40$, and $B = (1 - .9/(2*2), 40 - 3) = 2.02$. This yields precision (remembering to take square root of $\sigma^2 \left\{\hat{L}_1\right\}$) $B\sigma\left\{\hat{L}_1\right\} = 3.5$, which is above our threshold. If we try $n = 20$, $n_T = 80$, and we get $B\sigma\left\{\hat{L}_1\right\} = 2.49$. So $n = 20$, and $n_1 = n_3 = 20$, $n_2 = 40$, and $n_T = 40$.

(b) Repeat the calculations in part (a) if the sample size for the average physical fitness group is to be (1) n and (2) 3n, all other specifications remaining the same.

Answer: We follow the same method of guess and check, except now, our Bonferroni multiple changes because n_T is $3n$ in case (1) and $5n$ in case (2). For case (1), if $n = 20$, our B is now equal to 2.00, which leads us outside our precision. Our $\sigma^2\left\{\hat{L}_1\right\} = \frac{2\sigma^2}{n}$, which leads to $\sigma\left\{\hat{L}_1\right\} = 2.02$, and after multiplying by B puts us outside our required precision. We continue to try different n and find $\boxed{n = 26}$.

In case(2),

$$\sigma^2\left\{\hat{L}_1\right\} = \frac{4\sigma^2}{3n}$$

Following the same procedure, we find $\boxed{n = 18}$.

(c) Compare your results in part (a) and part (b). Which design leads to the smallest total sample size here?

Answer: In part (a), the total sample size was $n_T = 80$, in part (b) case (1), $n_T = 78$, and in case (2) $n_T = 90$. So equal sample size led to the smallest total sample size.