# Sections of Review Problems 1 Solutions

Chris Williams

## Exercise 2.3

Method 1 is not random at all, Methods 2 and 3 are systematic but depend on random order of subjects, Methods 4 and 5 use randomization with 4 yielding a random sample size while Method 5 is the typical fixed sample size randomization method.

## Exercise 3.3

First read in the data:

```
data <- read.table("c:/temp/OEx33.txt",header=T)

Ex3.3 <- data.frame(data)
rm(data)

Ex3.3$treat <- as.factor(Ex3.3$treat)
Ex3.3$treat <- relevel(Ex3.3$treat, ref="control")

Ex3.3
```

```
##          treat moisture
## 1         NaCl     80.5
## 2         NaCl     79.3
## 3         NaCl     79.0
## 4   formicAcid     89.1
## 5   formicAcid     75.7
## 6   formicAcid     81.2
## 7     beetPulp     77.8
## 8     beetPulp     79.5
## 9     beetPulp     77.0
## 10     control     76.7
## 11     control     77.2
## 12     control     78.6
```

### Fit the CR model, look at ANOVA table

```
Ex3.3.lm <- lm(moisture ~ treat, data=Ex3.3)

anova(Ex3.3.lm)
```

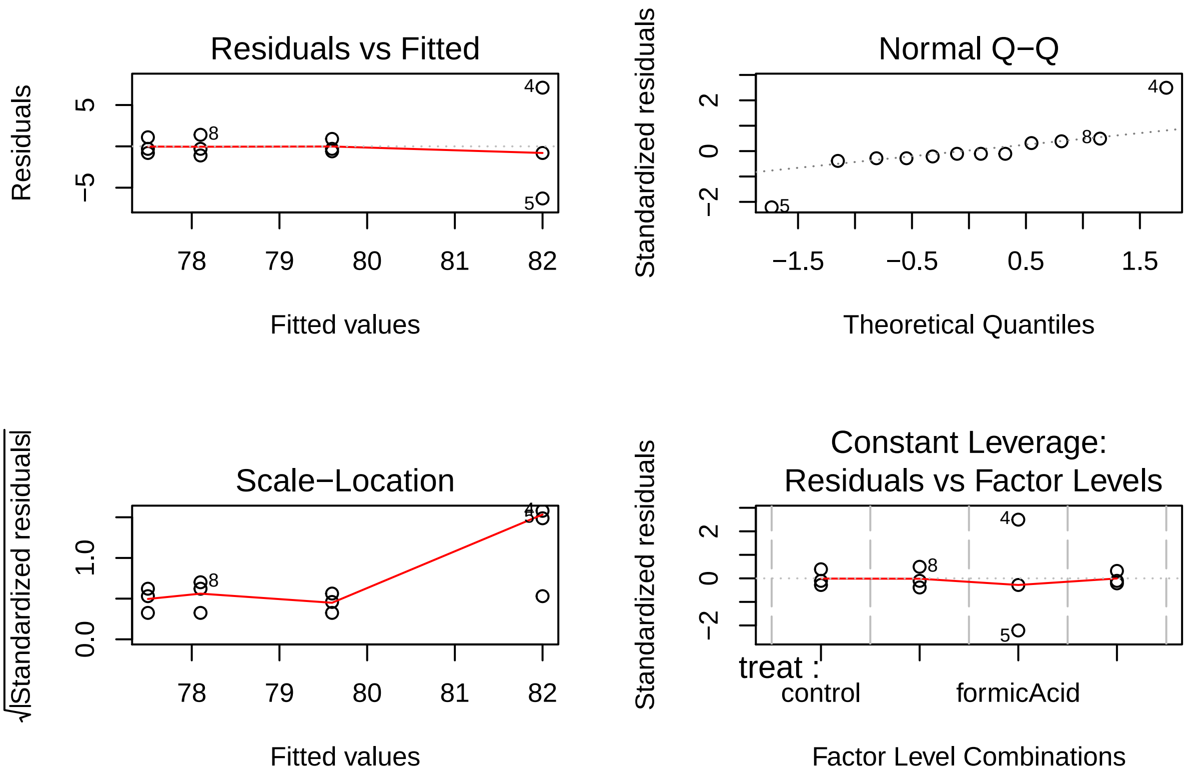
```
## Analysis of Variance Table
##
## Response: moisture
##           Df Sum Sq Mean Sq F value Pr(>F)
```

```
## treat      3  36.18   12.06  0.9926  0.444
## Residuals  8  97.20   12.15
```

```r
summary(Ex3.3.lm)
```

```
##
## Call:
## lm(formula = moisture ~ treat, data = Ex3.3)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -6.30  -0.80  -0.30   0.95   7.10
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        77.500      2.013  38.510 2.27e-10 ***
## treatbeetPulp       0.600      2.846   0.211    0.838
## treatformicAcid     4.500      2.846   1.581    0.153
## treatNaCl           2.100      2.846   0.738    0.482
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.486 on 8 degrees of freedom
## Multiple R-squared:  0.2713, Adjusted R-squared:  -0.002024
## F-statistic: 0.9926 on 3 and 8 DF,  p-value: 0.444
```

```r
 # check model assumptions

par(mfrow=c(2,2))
plot(Ex3.3.lm)
```




```r
par(mfrow=c(1,1))
```

With an F value of F = .99 on 3 and 8 degrees of freedom and a P value of P = .44, we fail to reject the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$. The diagnostic plots show that two observations in the 'formicAcid' group are far from the group mean and would benefit from further inspection.

# Exercise 4.4

Using the ordering of groups from the problem, to compare chewy with crispy we would use the contrast $w_1 = (1, -1, 1, -1, 1, -1)$. To compare expensive with inexpensive we would use the contrast $w_2 = (1, 1, -1, -1, 1, -1)$. To check orthogonality, since we assume equal sample sizes we multiply the corresponding coefficients and add, giving:

$$\sum_{i=1}^{6} w_{1i}w_{2i} = (1)(1) + (-1)(1) + (1)(-1) + (-1)(-1) + (1)(1) + (-1)(-1) = 2,$$

so the two contrasts are not orthogonal since the sum is not 0.

# Exercise 5.5

Since simultaneous confidence intervals are requested for differences between treatment groups and a control, Dunnetts method is appropriate.

```r
library(multcomp)
```

```
## Loading required package: mvtnorm
```

```
## Loading required package: survival
## Loading required package: TH.data
## Loading required package: MASS
##
## Attaching package: 'TH.data'
## The following object is masked from 'package:MASS':
##
##     geyser
```

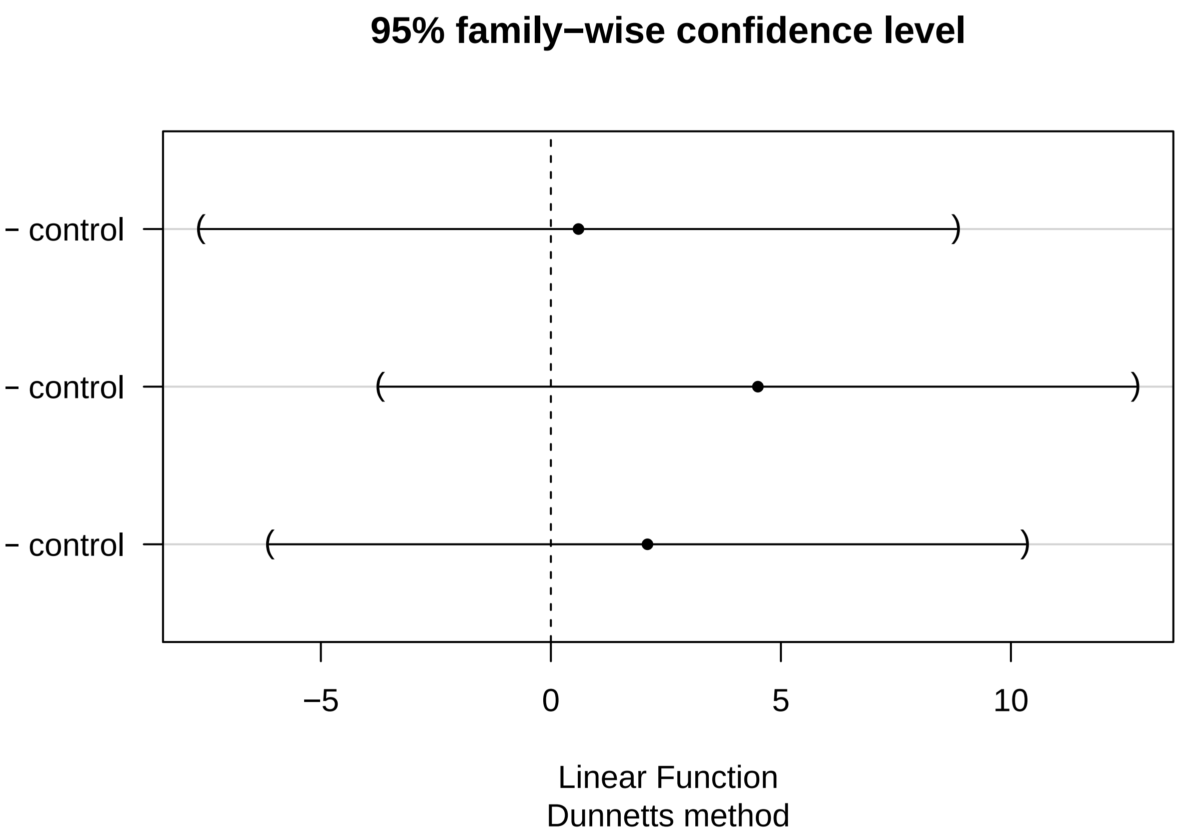
```r
Ex3.3.Dunnett <- glht(Ex3.3.lm,linfct=mcp(treat="Dunnett"))

summary(Ex3.3.Dunnett)
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Dunnett Contrasts
##
##
## Fit: lm(formula = moisture ~ treat, data = Ex3.3)
##
## Linear Hypotheses:
##                             Estimate Std. Error t value Pr(>|t|)
## beetPulp - control == 0        0.600      2.846   0.211    0.993
## formicAcid - control == 0      4.500      2.846   1.581    0.326
## NaCl - control == 0            2.100      2.846   0.738    0.805
## (Adjusted p values reported -- single-step method)
```

```r
confint(Ex3.3.Dunnett)
```

```
##
##   Simultaneous Confidence Intervals
##
## Multiple Comparisons of Means: Dunnett Contrasts
##
##
## Fit: lm(formula = moisture ~ treat, data = Ex3.3)
##
## Quantile = 2.8754
## 95% family-wise confidence level
##
##
## Linear Hypotheses:
##                             Estimate lwr     upr
## beetPulp - control == 0       0.6000  -7.5837  8.7837
## formicAcid - control == 0     4.5000  -3.6837 12.6837
## NaCl - control == 0           2.1000  -6.0837 10.2837
```

```r
plot(Ex3.3.Dunnett,sub="Dunnetts method")
```




Dunnetts two-sided statistic with 3, 8 df (shown in Table D.9 on page 637) is $d_{.05}(3, 8) = 2.88$. The output shows the standard error of each difference to be 2.84, so the Dunnett intervals are formed via

$$\text{difference} \pm (2.88)(2.84) = \text{difference} \pm 8.18.$$

All intervals include 0 since all P values were above .05.

# Problem 6.1

First read in the data:

```r
data <- read.table("c:/temp/OProb61.txt",header=T)

Ex6.1 <- data.frame(data)
rm(data)

Ex6.1$treat <- as.factor(Ex6.1$treat)
Ex6.1$treat <- relevel(Ex6.1$treat, ref="1-non")
```

## Fit the model, check assumptions

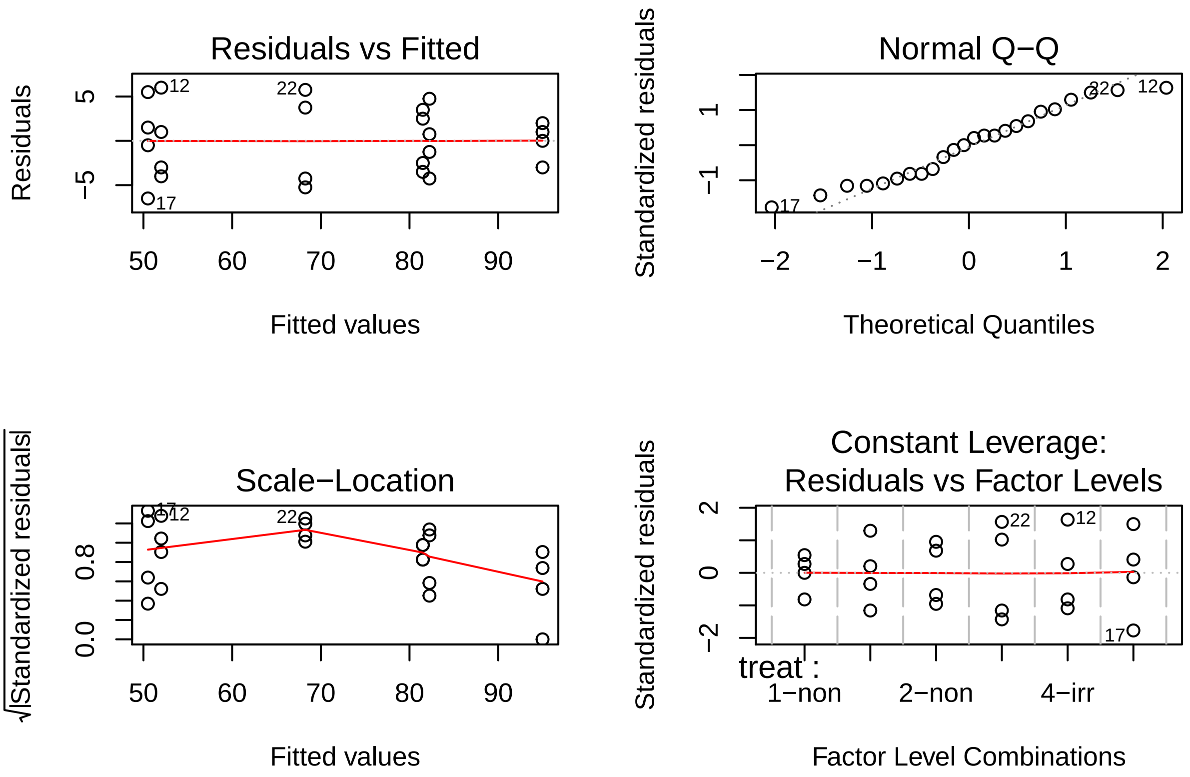
```r
Ex6.1.lm <- lm(pctBluestem ~ treat, data=Ex6.1)

anova(Ex6.1.lm)
```

```
## Analysis of Variance Table
##
## Response: pctBluestem
##           Df Sum Sq Mean Sq F value    Pr(>F)
## treat      5 6398.3 1279.67  71.203 3.197e-11 ***
## Residuals 18  323.5   17.97
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(Ex6.1.lm)
```

```
##
## Call:
## lm(formula = pctBluestem ~ treat, data = Ex6.1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.500 -3.125  0.375  2.750  6.000
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   95.000      2.120  44.818  < 2e-16 ***
## treat1-irr   -12.750      2.998  -4.253 0.000478 ***
## treat2-non   -13.500      2.998  -4.503 0.000275 ***
## treat3-non   -26.750      2.998  -8.924 5.00e-08 ***
## treat4-irr   -43.000      2.998 -14.344 2.72e-11 ***
## treat4-non   -44.500      2.998 -14.845 1.53e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.239 on 18 degrees of freedom
## Multiple R-squared:  0.9519, Adjusted R-squared:  0.9385
## F-statistic:  71.2 on 5 and 18 DF,  p-value: 3.197e-11
```

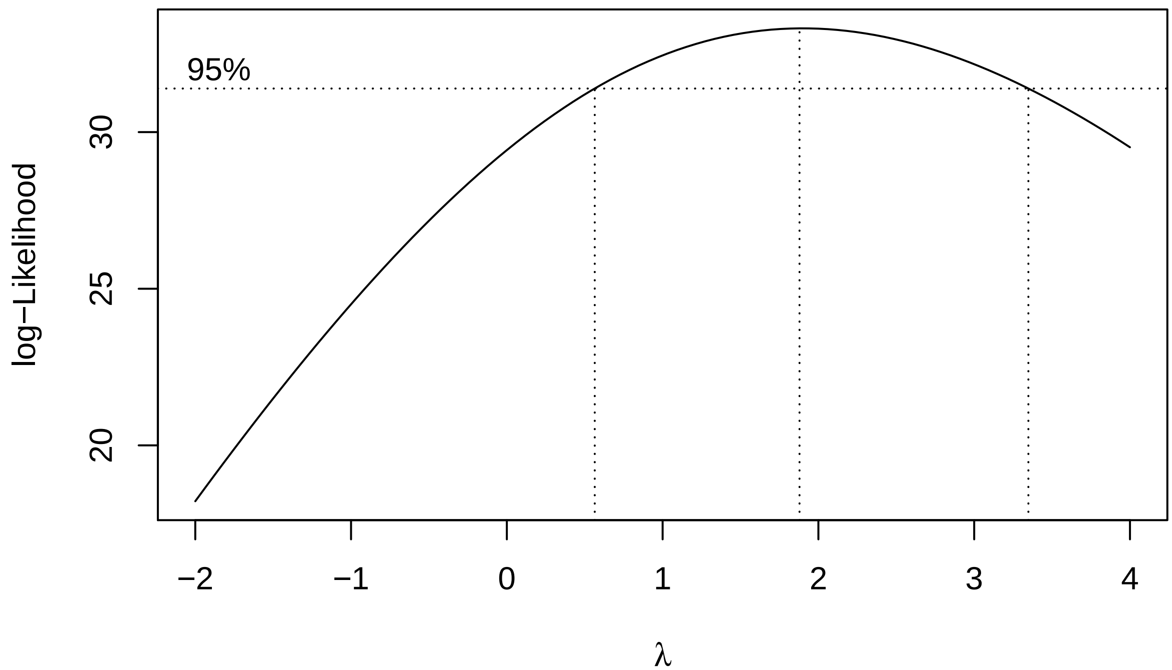
```
 # check model assumptions, part a

par(mfrow=c(2,2))
plot(Ex6.1.lm)
```




```
par(mfrow=c(1,1))
```

### need for transformation?

```
library(MASS)

 # plotting the log-likelihood function for the Box-Cox transformation
boxcox(pctBluestem ~ treat, data=Ex6.1,lambda = seq(-2.00, 4.00, length = 50))
```




Although the residual x predicted plot shows some evidence of variance heterogeneity, the 95% ci about the power parameter estimate includes 1, so we do not have strong evidence to need a transformation. A case could be made for transformation, but here we will use the original data.

### contrasts and multiple comparisons, parts c and d

With equal spacing of the nitrogen levels and equal sample sizes, we can use the coefficients for a quadratic effect (shown in Table D.6 on page 630). For the irrigation contrast, we should only use nitrogen levels where both irrigation levels were used.

```
library(lsmeans)
```

```
## Warning: package 'lsmeans' was built under R version 3.6.3

## Loading required package: emmeans

## Welcome to emmeans.
## NOTE -- Important change from versions <= 1.41:
##     Indicator predictors are now treated as 2-level factors by default.
##     To revert to old behavior, use emm_options(cov.keep = character(0))

## The 'lsmeans' package is now basically a front end for 'emmeans'.
## Users are encouraged to switch the rest of the way.
## See help('transition') for more information, including how to
## convert old 'lsmeans' objects and scripts to work with 'emmeans'.
```

```
levels(Ex6.1$treat)
```

```
## [1] "1-non" "1-irr" "2-non" "3-non" "4-irr" "4-non"
```

```r
Ex6.1.lsm <- lsmeans(Ex6.1.lm, "treat")

Contrasts <- list(Quad = c(1, 0, -1, -1, 0, 1),IrrEffect=c(1,-1,0,0,-1,1))

contrast(Ex6.1.lsm,Contrasts,adjust="none")
```

```
##  contrast  estimate   SE df t.ratio p.value
##  Quad         -4.25 4.24 18 -1.003  0.3294
##  IrrEffect    11.25 4.24 18  2.654  0.0162
```

### comparison with the best, part e

Here we are comparing the other treatments to the best one, so a one-sided Dunnett approach is appropriate.

```r
library(multcomp)

Ex6.1.CompBest <- glht(Ex6.1.lm,linfct=mcp(treat="Dunnett"),alternative="less")  # a one-sided Dunnett
                                                                                 # to compare to the bes

summary(Ex6.1.CompBest)
```

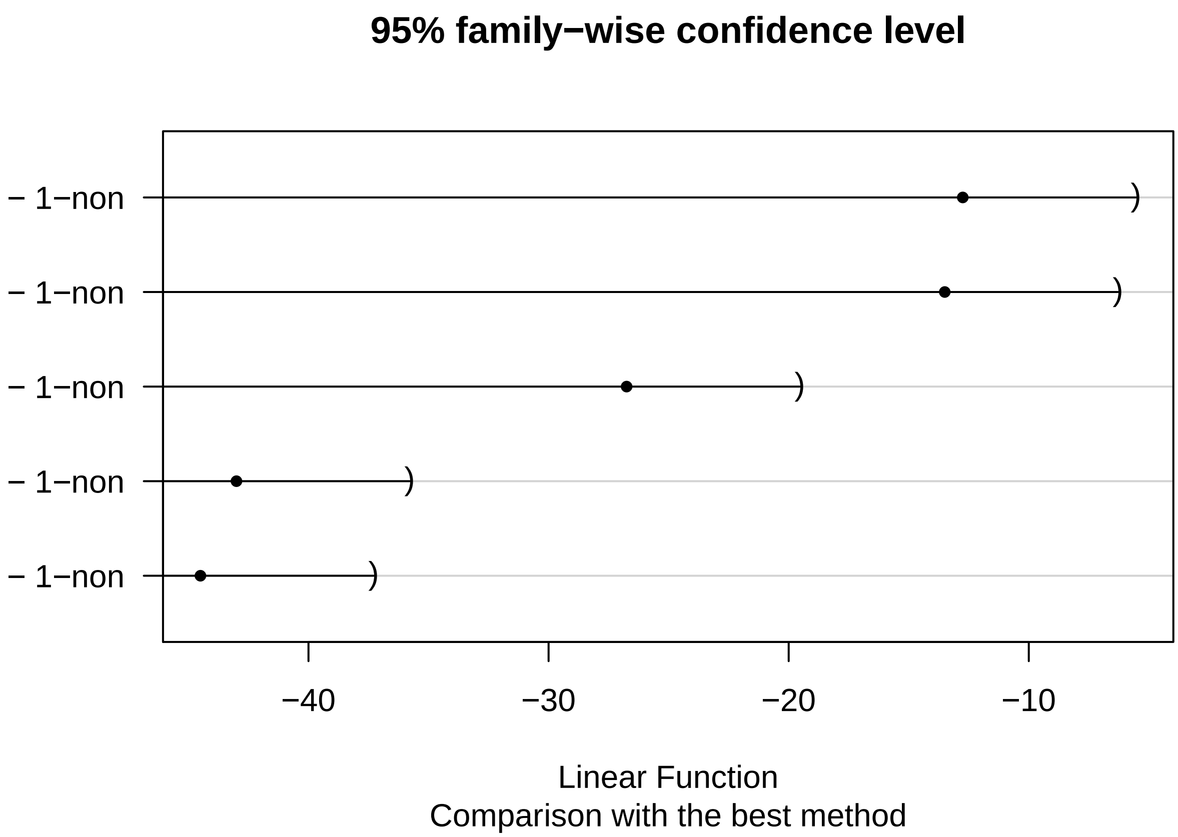
```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Dunnett Contrasts
##
##
## Fit: lm(formula = pctBluestem ~ treat, data = Ex6.1)
##
## Linear Hypotheses:
##                      Estimate Std. Error t value Pr(<t)
## 1-irr - 1-non >= 0   -12.750      2.998  -4.253 <0.001 ***
## 2-non - 1-non >= 0   -13.500      2.998  -4.503 <0.001 ***
## 3-non - 1-non >= 0   -26.750      2.998  -8.924 <0.001 ***
## 4-irr - 1-non >= 0   -43.000      2.998 -14.344 <0.001 ***
## 4-non - 1-non >= 0   -44.500      2.998 -14.845 <0.001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

```r
confint(Ex6.1.CompBest)
```

```
##
##   Simultaneous Confidence Intervals
##
## Multiple Comparisons of Means: Dunnett Contrasts
##
##
## Fit: lm(formula = pctBluestem ~ treat, data = Ex6.1)
##
## Quantile = 2.4082
## 95% family-wise confidence level
##
##
## Linear Hypotheses:
```

```
##                        Estimate lwr      upr
## 1-irr - 1-non >= 0 -12.7500     -Inf  -5.5309
## 2-non - 1-non >= 0 -13.5000     -Inf  -6.2809
## 3-non - 1-non >= 0 -26.7500     -Inf -19.5309
## 4-irr - 1-non >= 0 -43.0000     -Inf -35.7809
## 4-non - 1-non >= 0 -44.5000     -Inf -37.2809
```

```
plot(Ex6.1.CompBest,sub="Comparison with the best method")
```




All of the P values are below .05 so all other groups have different means than the best group.

## Exercise 7.2

The problem indicates that there are three groups, that the error variance is assumed to be $\sigma^2 = 4$, and that $\alpha = .05$. The group means are assumed to be 10, 11, and 11, giving a grand mean of 10.67. Then the key ratio needed for the sample size calculation is:

$$\sum_i \frac{\alpha_i^2}{\sigma^2} = \frac{(10 - 10.67)^2 + (11 - 10.67)^2 + (11 - 10.67)^2}{4} = .167.$$

We can now use our power function to calculate the result:

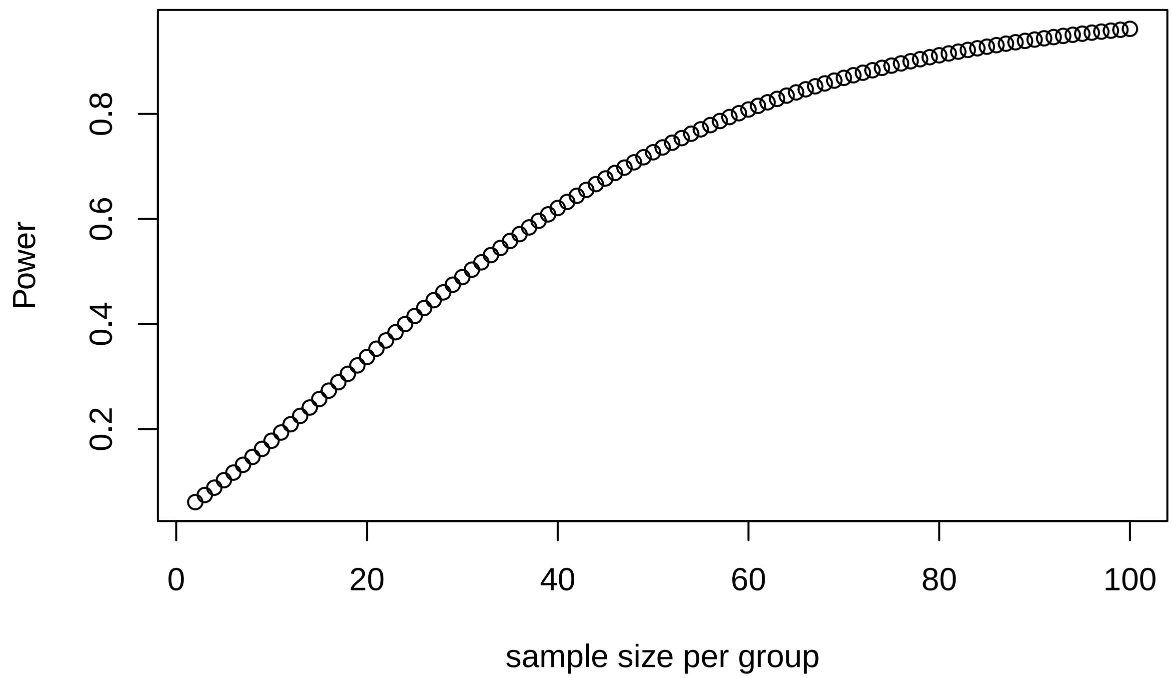
```
powercr <- function(ngrp,rat,alpha,n1,nlast,ninc)
     {
        n       <- seq(n1,nlast,by=ninc)
        power   <- numeric(length(n))
        ndf     <-  ngrp -1
        ddf         <-  ngrp*(n-1)
```

```r
        fcr          <-  qf(1-alpha,ndf,ddf)
        lambda       <-  n*rat
       # print(c(n,ddf,fcr,lambda))
        power    <-  1 - pf(fcr,ndf,ddf,lambda)
        print(cbind(n,ddf,lambda,fcr,power))
        plot(n,power,type="b",xlab="sample size per group",ylab="Power")
      }

 # for the homework problem using the powercr function

powercr(3,.1667,.05,2,100,1)
```

```
##           n ddf  lambda      fcr      power
##    [1,]   2   3  0.3334 9.552094 0.06087451
##    [2,]   3   6  0.5001 5.143253 0.07442152
##    [3,]   4   9  0.6668 4.256495 0.08833611
##    [4,]   5  12  0.8335 3.885294 0.10255753
##    [5,]   6  15  1.0002 3.682320 0.11707913
##    [6,]   7  18  1.1669 3.554557 0.13188256
##    [7,]   8  21  1.3336 3.466800 0.14694210
##    [8,]   9  24  1.5003 3.402826 0.16222869
##    [9,]  10  27  1.6670 3.354131 0.17771211
##   [10,]  11  30  1.8337 3.315830 0.19336199
##   [11,]  12  33  2.0004 3.284918 0.20914852
##   [12,]  13  36  2.1671 3.259446 0.22504266
##   [13,]  14  39  2.3338 3.238096 0.24101646
##   [14,]  15  42  2.5005 3.219942 0.25704306
##   [15,]  16  45  2.6672 3.204317 0.27309685
##   [16,]  17  48  2.8339 3.190727 0.28915345
##   [17,]  18  51  3.0006 3.178799 0.30518974
##   [18,]  19  54  3.1673 3.168246 0.32118386
##   [19,]  20  57  3.3340 3.158843 0.33711522
##   [20,]  21  60  3.5007 3.150411 0.35296445
##   [21,]  22  63  3.6674 3.142809 0.36871340
##   [22,]  23  66  3.8341 3.135918 0.38434514
##   [23,]  24  69  4.0008 3.129644 0.39984386
##   [24,]  25  72  4.1675 3.123907 0.41519492
##   [25,]  26  75  4.3342 3.118642 0.43038475
##   [26,]  27  78  4.5009 3.113792 0.44540087
##   [27,]  28  81  4.6676 3.109311 0.46023181
##   [28,]  29  84  4.8343 3.105157 0.47486709
##   [29,]  30  87  5.0010 3.101296 0.48929719
##   [30,]  31  90  5.1677 3.097698 0.50351350
##   [31,]  32  93  5.3344 3.094337 0.51750831
##   [32,]  33  96  5.5011 3.091191 0.53127471
##   [33,]  34  99  5.6678 3.088240 0.54480662
##   [34,]  35 102  5.8345 3.085465 0.55809873
##   [35,]  36 105  6.0012 3.082852 0.57114643
##   [36,]  37 108  6.1679 3.080387 0.58394582
##   [37,]  38 111  6.3346 3.078057 0.59649366
##   [38,]  39 114  6.5013 3.075853 0.60878732
##   [39,]  40 117  6.6680 3.073763 0.62082476
##   [40,]  41 120  6.8347 3.071779 0.63260448
##   [41,]  42 123  7.0014 3.069894 0.64412553
```

```
## [42,]  43 126  7.1681 3.068100 0.65538743
## [43,]  44 129  7.3348 3.066391 0.66639015
## [44,]  45 132  7.5015 3.064761 0.67713411
## [45,]  46 135  7.6682 3.063204 0.68762011
## [46,]  47 138  7.8349 3.061716 0.69784933
## [47,]  48 141  8.0016 3.060292 0.70782331
## [48,]  49 144  8.1683 3.058928 0.71754389
## [49,]  50 147  8.3350 3.057621 0.72701322
## [50,]  51 150  8.5017 3.056366 0.73623370
## [51,]  52 153  8.6684 3.055162 0.74520800
## [52,]  53 156  8.8351 3.054004 0.75393902
## [53,]  54 159  9.0018 3.052891 0.76242986
## [54,]  55 162  9.1685 3.051819 0.77068380
## [55,]  56 165  9.3352 3.050787 0.77870430
## [56,]  57 168  9.5019 3.049792 0.78649496
## [57,]  58 171  9.6686 3.048833 0.79405954
## [58,]  59 174  9.8353 3.047906 0.80140189
## [59,]  60 177 10.0020 3.047012 0.80852598
## [60,]  61 180 10.1687 3.046148 0.81543585
## [61,]  62 183 10.3354 3.045312 0.82213564
## [62,]  63 186 10.5021 3.044504 0.82862954
## [63,]  64 189 10.6688 3.043722 0.83492179
## [64,]  65 192 10.8355 3.042964 0.84101668
## [65,]  66 195 11.0022 3.042230 0.84691851
## [66,]  67 198 11.1689 3.041518 0.85263162
## [67,]  68 201 11.3356 3.040828 0.85816035
## [68,]  69 204 11.5023 3.040158 0.86350904
## [69,]  70 207 11.6690 3.039508 0.86868202
## [70,]  71 210 11.8357 3.038877 0.87368363
## [71,]  72 213 12.0024 3.038264 0.87851816
## [72,]  73 216 12.1691 3.037667 0.88318989
## [73,]  74 219 12.3358 3.037088 0.88770307
## [74,]  75 222 12.5025 3.036524 0.89206190
## [75,]  76 225 12.6692 3.035975 0.89627055
## [76,]  77 228 12.8359 3.035441 0.90033312
## [77,]  78 231 13.0026 3.034921 0.90425369
## [78,]  79 234 13.1693 3.034414 0.90803627
## [79,]  80 237 13.3360 3.033920 0.91168482
## [80,]  81 240 13.5027 3.033439 0.91520321
## [81,]  82 243 13.6694 3.032969 0.91859529
## [82,]  83 246 13.8361 3.032512 0.92186482
## [83,]  84 249 14.0028 3.032065 0.92501549
## [84,]  85 252 14.1695 3.031629 0.92805093
## [85,]  86 255 14.3362 3.031203 0.93097469
## [86,]  87 258 14.5029 3.030788 0.93379027
## [87,]  88 261 14.6696 3.030382 0.93650107
## [88,]  89 264 14.8363 3.029985 0.93911043
## [89,]  90 267 15.0030 3.029597 0.94162161
## [90,]  91 270 15.1697 3.029218 0.94403780
## [91,]  92 273 15.3364 3.028847 0.94636213
## [92,]  93 276 15.5031 3.028485 0.94859762
## [93,]  94 279 15.6698 3.028130 0.95074724
## [94,]  95 282 15.8365 3.027783 0.95281389
## [95,]  96 285 16.0032 3.027443 0.95480037
```

```
##   [96,]  97 288 16.1699 3.027111 0.95670944
##   [97,]  98 291 16.3366 3.026785 0.95854376
##   [98,]  99 294 16.5033 3.026466 0.96030594
##   [99,] 100 297 16.6700 3.026153 0.96199849
```




We only get to 90% power when n = 77 (per group), so a large sample is needed to detect a difference under these conditions.