36-402/608 Homework #4 Solutions 2/11

1. Pygmalion effect (40 points, 5 each)

   Load the Sleuth's Pygmalion dataset (case1302.csv), which is the second featured case of Chapter 13, and is found on the Sleuth CD. Do the usual EDA to be sure you understand the form of the data. Let's use μ_C and μ_P to represent the population means of score for subjects exposed to control vs. "Pygmalion" treatment, and μ_{C1} through μ_{C10} to represent the population means of score for subjects in the 10 companies.

   (a) Turn in the R `summary(aov())` table that allows you to retain the null hypothesis of no interaction between company and treatment in their effects on score.

```
> summary(aov(score~company*treat,pyg))
             Df Sum Sq Mean Sq F value  Pr(>F)
company       9 670.98   74.55  1.4367 0.29902
treat         1 338.88  338.88  6.5304 0.03092 *
company:treat 9 311.46   34.61  0.6669 0.72212
Residuals     9 467.04   51.89
```

   (b) Turn in the R `summary(aov())` table that checks just the "block" effect of company. Turn in the null hypothesis of interest (formula or word format, being sure to refer to population not sample means), and state whether or not you have sufficient evidence to reject it.

```
> summary(aov(score~company,pyg))
            Df  Sum Sq Mean Sq F value Pr(>F)
company      9  670.98  74.553  1.2677 0.3153
Residuals   19 1117.39  58.810
```

   $H_0 : \mu_{B1} = \cdots = \mu_{B10}$ is retained with p=0.315>0.05.

   (c) Turn in the R `summary(aov())` that tests for the additional effect of treatment after adjusting (correcting) for the effect of company (without interaction). Turn in the null hypothesis of interest and state whether or not you have sufficient evidence to reject it.

```
> summary(aov(score~company+treat,pyg))
            Df Sum Sq Mean Sq F value  Pr(>F)
company      9 670.98   74.55  1.7238 0.15556
treat        1 338.88  338.88  7.8354 0.01186 *
Residuals   18 778.50   43.25
```

$H_0 : \mu_C = \mu_P$ is rejected at p=0.0119≤0.05.

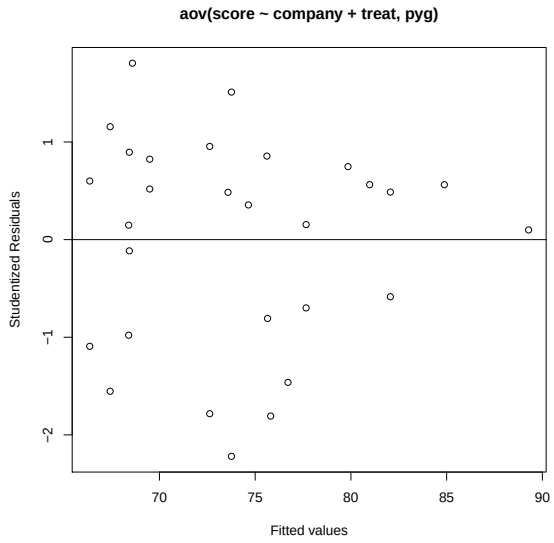
(d) What is the best estimate of σ^2, the within-group error variance, for the model in part c?

MS_R=43.25 is the best estimate of σ^2.

(e) What is the p-value for treatment when it is put in the additive model before company?

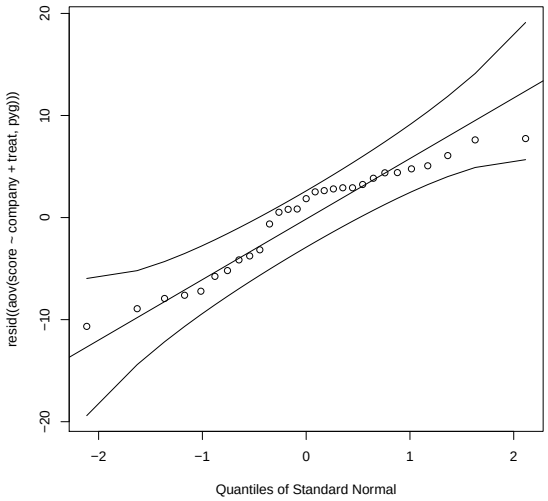
p=0.01314 (which differs from the 0.01186 above).

(f) Make a residual plot for the model from part c. Briefly state your conclusions reached from examining this plot.



This plot looks "all clear" to me with no problems with non-linearity (in ANOVA lack of zero residual means really reflects the need for an interaction) and no problem with unequal variance (although one might claim that there is low variance where the fitted values are highest).

(g) Run `qqn()` (from http://www.stat.cmu.edu/∼hseltman/files/qqn.R) on the residuals from same model as part c, and briefly state your conclusions reached from examining this plot.



There is a very slight degree of "light tails", but certainly not to an extent that would affect the p-value in a meaningful way.

(h) Why is the p-value from `summary(aov(score∼company+treat,pyg))` smaller than the p-value from `summary(aov(score∼treat,pyg))`, and how does this relate to improving power? (Substitute your data.frame name for "pyg".)

The SS for treatment is similar for both models but the residual SS is much smaller when some of the total SS is alloted to company differences. This causes a somewhat smaller MS, which raises the F value and therefore lowers the p-value. In general reducing residual error by making more homogenous groups of subjects raises power because most statistical tests are based on comparing the size of treatment effects to the size of the resiudal error.

2. Stepping and heart rate (45 points)

   The study examined in this problem is about the effects of "stepping" exercise on heart rate (HR). Subjects were randomly assigned to two different heights of steps, and three different frequencies. Instructor differences may occur, so instructors were treated as blocks. (We will not use the "order" or "RestHR" variables.)

   Load the data from "stepping.dat" using

```
stp = read.table("stepping.dat", header=TRUE)
dim(stp) # 30 6
sapply(stp, class)
#     Order     Block    Height Frequency    RestHR        HR
# "integer" "integer" "integer" "integer" "integer" "integer"
stp$Block = factor(stp$Block)
stp$Height = factor(stp$Height, labels=c("Low","High"))
stp$Frequency = factor(stp$Frequency, labels=c("Low","Med","High"))
summary(stp)
```

   (a) Run `with(stp, table(Height, Frequency, Block))` and turn in a statement of what you observe.

   Each block shows one subject for each of five of the six possible combinations of height and frequency. The missing combination is in a different position for each block.

   (b) Make a similar table without Block, and turn in a statement of what you observed including either the term "balanced" or "unbalanced".

```
> with(stp, table(Height, Frequency))
      Frequency
Height Low Med High
  Low    5   5    5
  High   5   5    5
```

   Across all blocks the two treatments are balanced with 5 subjects for each combination of treatment levels.

   (c) Run the additive 3-way ANOVA model for the HR outcome with all three explanatory factors. Try different orderings of the variables paying special attention to the SS values and F values (as an easier to read surrogate for the p-values). Explain the pattern when i) comparing Block+Frequency+Height to Frequency+Block+Height and ii) comparing Block+Frequency+Height to Block+Height+Frequency.

```
> summary(aov(HR~Block+Frequency+Height,stp))
            Df Sum Sq Mean Sq F value    Pr(>F)
```

```
Block        5 4510.8   902.2  16.203 1.374e-06 ***
Frequency    2 3035.1  1517.6  27.256 1.459e-06 ***
Height       1 3406.1  3406.1  61.173 1.181e-07 ***
Residuals   21 1169.2    55.7
```

```
> summary(aov(HR~Frequency+Block+Height,stp))
            Df Sum Sq Mean Sq F value    Pr(>F)
Frequency    2 3727.8  1863.9  33.476 2.942e-07 ***
Block        5 3818.1   763.6  13.715 5.126e-06 ***
Height       1 3406.1  3406.1  61.173 1.181e-07 ***
Residuals   21 1169.2    55.7
```

Height SS and F are unchanged because in both cases it is added after block and frequency. Both SS and F change for Block and Frequency depending on which is added first because they are unbalanced, but the sum of the two SS values stays the same.

```
> summary(aov(HR~Block+Height+Frequency,stp))
            Df Sum Sq Mean Sq F value    Pr(>F)
Block        5 4510.8   902.2  16.203 1.374e-06 ***
Height       1 3406.1  3406.1  61.173 1.181e-07 ***
Frequency    2 3035.1  1517.6  27.256 1.459e-06 ***
Residuals   21 1169.2    55.7
```

With the balanced pattern of height and frequency across all blocks, we see the same SS and F values for both orderings.

(d) Taking any of the tables of part c, we can see the $df_W = 21$, $SS_W = 1169.2$, and $MS_W = 55.7$, where W stands for "within groups". Some programs include lines for Between Groups and Total. What would the values be for $\mathrm{df}_B, \mathrm{SS}_B, \mathrm{MS}_B, \mathrm{df}_T$, and SS_T?

$\mathrm{df}_B = \mathrm{df}_{\mathrm{Block}} + \mathrm{df}_H + \mathrm{df}_F = 5 + 1 + 2 = 8.$
$\mathrm{SS}_B = \mathrm{SS}_{\mathrm{Block}} + \mathrm{SS}_H + \mathrm{SS}_F = 4510.8 + 3406 + 3035.1 = 10952.$
$\mathrm{MS}_B = \mathrm{SS}_B/\mathrm{df}_B = 10952/8 = 1369.$
$\mathrm{df}_T = \mathrm{df}_B + \mathrm{df}_W = 8 + 21 = 29$
$\mathrm{SS}_T = \mathrm{SS}_B + \mathrm{SS}_W = 10952 + 1169.2 = 12121.2.$

(e) Logically the interaction to be most concerned about is that the effect of a change in step height on heart rate depends on (changes with) the specific level of treatment. Fit that model (including blocks as the first factor) and turn in the p-value for the interaction, and your conclusion about interaction.

```
> summary(aov(HR~Block+Frequency*Height,stp))
                 Df Sum Sq Mean Sq F value    Pr(>F)
Block             5 4510.8   902.2 19.7945 6.122e-07 ***
```

```
Frequency          2 3035.1  1517.6 33.2969 6.165e-07 ***
Height             1 3406.1  3406.1 74.7329 5.201e-08 ***
Frequency:Height   2  303.3   151.6  3.3274   0.05768 .
Residuals         19  865.9    45.6
```

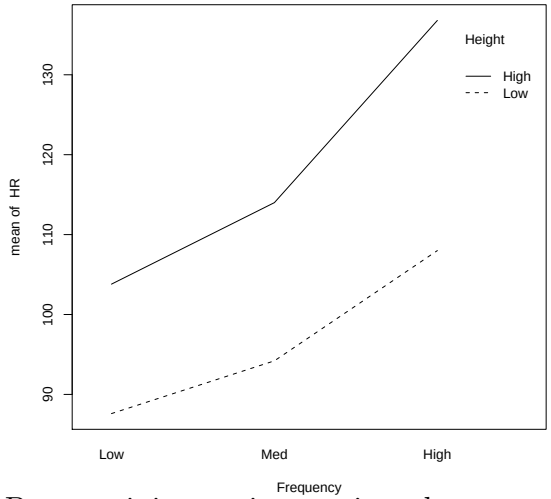
The p-value is 0.058, so using the conventional cutoff value of 0.05, we do not have sufficient evidence to reject the null hypothesis of no iteraction, and we should model the pattern of means across frequencies as being the same for both heights with only a constant difference between them. (But with p only slightly higher than 0.05, we might be concerned about a high chance of a type-2 error regarding the interaction.)

(f) When we have a term in a model that is not statistically significant, it is helpful to get a CI on the effect estimates to allow subject matter experts to determine whether a practically significant effect is likely.

We will examine whether the step height effect is statistically different for low vs. high frequency. First run

```
mi=aov(HR~Block+Frequency+Height+Frequency:Height, stp)
coefficients(mi)
```

to see the estimated coefficients. Next think carefully about how you would use the coefficients to calculate the expected HR for conditions FH&HH, FH&HL, FL&HH, and FL&HL, where, e.g., FH&HH means frequency is high and step height is high.



By examining an interaction plot, you can see that the quantity of interest is (FH&HH-FH&HL) - (FL&HH-FL&HL), which simplifies nicely in this case to $b_{\text{FrequencyHigh:HeightHigh}}$, with an estimate of 9.75. To make a CI, we need to get the standard error of this quantity, using either

```
sqrt(vcov(mi)["FrequencyHigh:HeightHigh", "FrequencyHigh:HeightHigh"])
```

or `summary.lm(mi)`, which both give 6.163. The pertinent df is from the Residual line: 19. The "plus or minus" comes from `qt(0.975, 19)`.

Give a careful statement about our confidence in the size of the interaction using the confidence interval (rounded to whole numbers) in a form that could be understood by an exercise physiologist. Start with "We are 95% confident that ....".

qt(0.975, 19) = 2.09.

CI = 9.75 +/- 2.09(6.163) = [-3.15, 22.649]

We are 95% confident that the difference in the rise of heart rate from low to high steps is between 3 beats per minute smaller and 23 bpm larger when comparing high frequency to low frequency.

If the subject matter expert thinks 23 bmp is a small change, then we are happy reporting our no-interaction model. If she thinks it is a large change, then we should run another experiment with higher power, to see if there really is a meaningful interaction or not. Here is seems very likely that the correct conclusion is insufficient power.

(g) Turn in the R `summary(aov())` table for the additive 3-way ANOVA model and your conclusions about whether or not step height and frequency have effects on heart rate.

```
> summary(aov(HR~Block+Frequency+Height,stp))
            Df Sum Sq Mean Sq F value    Pr(>F)
Block        5 4510.8   902.2  16.203 1.374e-06 ***
Frequency    2 3035.1  1517.6  27.256 1.459e-06 ***
Height       1 3406.1  3406.1  61.173 1.181e-07 ***
Residuals   21 1169.2    55.7
```

With very small p-values we conclude that the mean heart rate is not that same for all three frequencies and is not the same for the two heights.

(h) The client asks for the following planned contrast hypotheses: to test whether the "high frequency heart rate differs from the average of the medium and low frequencies" and whether the "medium frequency heart rate differs from the low frequency." Using the additive model, construct the coefficients and carry out the contrast tests using fit.contrasts() in package "gmodels". Turn in the R code, the R output, and a brief summary of your conclusions, being sure to correctly explain the sign of any significant effect(s).

```
> library(gmodels)
> levels(stp$Frequency)  # to check the order of the levels for the contrast st
[1] "Low"  "Med"  "High"
> s0 = aov(HR~Block+Frequency+Height,stp)
> contr = rbind(HvsML = c(-1/2, -1/2, 1), MvsL = c(-1, 1, 0))
```

```
> round( fit.contrast(s0, "Frequency", contr, conf.int=0.95), 3)
                Estimate Std. Error t value Pr(>|t|) lower CI upper CI
FrequencyHvsML   20.250      2.950   6.865        0   14.116   26.384
FrequencyMvsL      9.25      3.406   2.716    0.013    2.167   16.333
```

Heart rate is signficantly higher with high frequency stepping than the average of low and medium frequency (95% CI=[14.1, 26.4] bpm).

Heart rate is signficantly higher with medium frequency stepping than with low frequency (95% CI=[2.2, 16.3] bpm).

(i) Explain why it does not make sense to construct contrasts for Height (hint: the answer is statistical, not from the subject matter.)

The only possible contrast is one proportional to c(1, -1) which is the 1 df test of $\mu_{H1} = \mu_{H2}$, and that is the same null hypothesis as the 1 df F test shown in the Height line of the ANOVA table. ($t^2 = F$)

3. Writing exercise #1 (15 points) **Turn this in on a separate non-stapled piece of paper with your name on it.**

You are the statistician for the Institute of Global Oceanographic Research (IGOR). Your boss, Dr. Frankenstein, is an intelligent, but very busy woman who had only one statistics course 20 years ago. You must write a one page (one side only) executive report for her, summarizing how you analyzed the algal regrowth data and what you conclusions are. You may include any text, tables and graphs that you want, but she will not look beyond a single (one sided) report.

Your goal is to honestly explain what you found, and make certain that she will not get mad at you if someone else tells her some true additional details of your analysis that you left out, and that she will not be embarrassed if she uses your report as part of a talk at a scientific meeting. Assume that she sees dozens of these reports a week, and can't remember the details of each study that she funded.

You should base your report on Breakout # 8, and do not need not do any reanalysis of the data (though you may, if you like). You can use the code in HW4p3.R to regenerate all of the results in the breakout as an aid to generating tables and graphs.

This assignment will be graded on how well you choose the information to present based on what is most important, and how clearly you communicate the ideas (though you will not be penalized for grammar errors or other errors related to non-English first language).