

STAT502 Lab #8

1. The “CH22PR11.txt” gives days to recovery (Y) after knee surgery for 24 young adults. Also recorded are their age (X) and group (Factor A) of physical health (low,med,high). The data can be read into R with

```
data = read.table('CH22PR11.txt')
y = data[,1]
x = data[,4]
A = as.factor(data[,2])
```

- (a) Create scatterplots of the recovery versus age with the commands:

```
plot(x[A==1],y[A==1],xlab='Age',ylab='Recovery',ylim=c(18,45),col='blue')
points(x[A==2],y[A==2],col='red')
points(x[A==3],y[A==3],col='green')
```

Using the notation `x[A==1]` restricts the values to the first group ($A=1$ for the low physical group, etc.). Also, using the `points` command allows us to add additional points without removing those already present.

- (b) Using the plot above, comment on the assumption of equal slopes for the groups. Write down the full and reduced models for testing whether slopes are equal or not. In terms of these model parameters, what are the hypotheses?
- (c) First, center the ages with `x=x-mean(x)`, and then fit the ANCOVA model with unequal slopes (interaction between age and group) using `mod1=anova(lm(y~x+A+x:A))`. Record the sum of squares and degrees of freedom for error. Next, fit the ANCOVA model with equal slopes (no interaction between age and group) using `mod2=anova(lm(y~x+A))`. Record the sum of squares and degrees of freedom for error.
- (d) Carry out the test in part (b) using the full versus reduced model approach. Report the test statistic, degrees of freedom, and conclusion with $\alpha = .05$. What does this mean about the equal-slopes model?
- (e) Refer again to the plot of recovery versus age, and comment on the assumption of an effect due to physical health group. Does it appear that recovery time depends on a person's physical health? Write down the full and reduced models for testing for this effect. In terms of these model parameters, what are the hypotheses?
- (f) For the full model in part (e), save the residuals and fitted values. Check the plot of residuals versus fits and the QQ plot of residuals, and comment on the assumptions of this model.
- (g) Fit the reduced model in part (e), and to carry out the test of an effect due to physical health with $\alpha = .05$. What are your conclusions? Also, comment on the appropriateness of this test, given the results from the test in part (d). *Hint: compare the reduced model in (d) with the full model in this test.*

Note: You may have noticed that the order of terms expressed in the model makes a difference in R's `lm()` function. To see this, note the differences in the output for `anova(lm(y~x+A))` and `anova(lm(y~x+A))`. This is because R uses “Type I” sum of squares, which are defined as the difference in error sum of squares for adding terms sequentially—so order matters. If the last term added is the one we want to test, we can use the given F and p values in the ANOVA table; otherwise we have to fit both full and reduced models separately. To see this, compare the F value from `anova(lm(y~x+A))` with your test statistic from part (g).