

Homework 6 - STATS 513

Saad Ahmad

Ankylosing spondylitis is a chronic form of arthritis. A study was conducted to determine whether daily stretching of the hip tissues would improve mobility. The data are found in `hips`. The flexion angle of the hip before the study (`fbef`) is a predictor and the flexion angle after the study (`faft`) is the response. (The variables `rbef` and `raft` won't be used.)

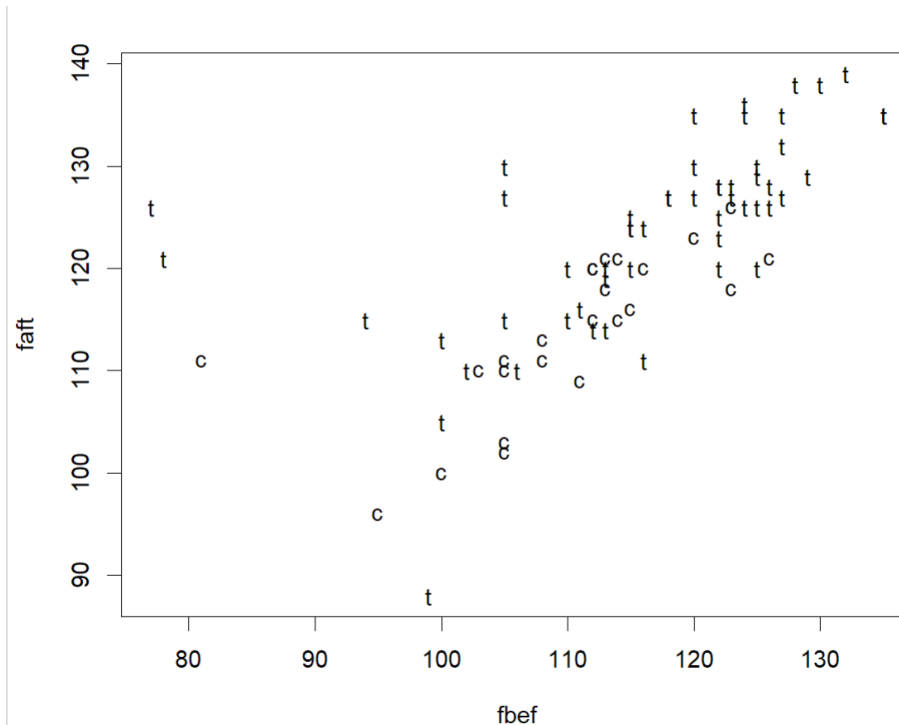
```
library(faraway)
?hips
data(hips)
summary(hips)
```

```
> summary(hips)
```

fbef	faft	rbef	raft	grp	side	person
Min. : 77.0	Min. : 88.0	Min. : 2.00	Min. : 2.00	control:24	right:39	1 : 2
1st Qu.:108.0	1st Qu.:115.0	1st Qu.:20.25	1st Qu.:25.00	treat :54	left :39	2 : 2
Median :115.0	Median :121.0	Median :25.00	Median :30.00			3 : 2
Mean :114.5	Mean :120.8	Mean :24.85	Mean :29.71			4 : 2
3rd Qu.:123.0	3rd Qu.:127.0	3rd Qu.:31.50	3rd Qu.:36.00			5 : 2
Max. :135.0	Max. :139.0	Max. :48.00	Max. :50.00			6 : 2
						(other):66

(1). Plot the data using different plotting symbols for the treatment and the control status.

```
plot(fbef ~ grp, hips)
plot(faft~fbef, pch=as.character(grp), hips)
```



(2). Fit a model to determine whether there is a treatment effect. (Hints: Please use model selection to select the model. For example, you may need to determine whether the interaction term should be included)

Without interaction term

```
linearModel1 = lm(faft ~ fbef + grp, data = hips)
summary(linearModel1)
call:
lm(formula = faft ~ fbef + grp, data = hips)

Residuals:
    Min       1Q   Median       3Q      Max
-27.4050  -3.1586   0.0214   3.1850  21.3768

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  59.86141    7.77213   7.702 4.39e-11 ***
fbef          0.49008    0.06949   7.053 7.39e-10 ***
grptreat      7.02553    1.77085   3.967 0.000165 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.976 on 75 degrees of freedom
Multiple R-squared:  0.5327,    Adjusted R-squared:  0.5202
F-statistic: 42.74 on 2 and 75 DF,  p-value: 4.087e-13
```

With interaction term

```
linearModel2 = lm(faft ~ fbef + grp + fbef*grp, data = hips)
summary(linearModel2)
call:
lm(formula = faft ~ fbef + grp + fbef * grp, data = hips)

Residuals:
    Min       1Q   Median       3Q      Max
-27.7919  -3.2746   0.1585   3.0270  20.5036

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  50.8476    16.7102   3.043 0.003242 **
fbef          0.5721     0.1514   3.778 0.000317 ***
grptreat     18.6146    19.0800   0.976 0.332436
fbef:grptreat -0.1041     0.1706  -0.610 0.543696
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.006 on 74 degrees of freedom
Multiple R-squared:  0.535,    Adjusted R-squared:  0.5161
F-statistic: 28.38 on 3 and 74 DF,  p-value: 2.547e-12
```

(3). Compute the difference between the flexion before and after and test whether this difference varies between treatment and control. Contrast this approach to your previous model. (use `var.equal = T` in the test)

```
difference = (hips$faft - hips$fbef)
t.test(difference[grp=="treat"], difference[grp=="control"])
```

welch Two Sample t-test

```
data: difference[grp == "treat"] and difference[grp == "control"]
t = 1.901, df = 62.271, p-value = 0.06193
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.1898407  7.5694703
sample estimates:
mean of x mean of y
 7.481481  3.791667
```

Here the p-value = 0.06193 being not-significant hence there is no treatment effect because of the angle difference.

(4). What is the estimated size of the treatment effect? Give a 95% confidence interval.

```
linearModel3 = lm(faft ~ fbef + grp, data = hips)
confint(linearModel3)
```

```
> confint(linearModel3)
```

	2.5 %	97.5 %
(Intercept)	44.3785299	75.3442960
fbef	0.3516589	0.6285049
grptreat	3.4978116	10.5532467

The 95% C.I. is (3.497, 10.553).

(5). Notice that both legs of each subject have been included in the study as separate observations. Explain what difficulties this causes with the model assumptions.

It could cause some correlated errors as we are using both legs of the same person as a separate observation.

(6). Compute the average angles for each subject and repeat the modeling with this reduced data set (`hips_new`). Point out differences in the conclusions if any.

```
hips_new<-aggregate(list(fbef=hips$fbef, faft=hips$faft),by=list(person=hips$person),
mean)
hips_new$grp=hips$grp[match(hips_new$person,hips$person)]
plot(fbef ~ grp, hips_new)
plot(faft~fbef, pch=as.character(grp), hips_new)
```

