STAT151A HW2 Q6 Solution

Billy Fang

September 16, 2017

```
dat <- read.csv("bodyfat.csv")
```

(a)

We provide two approaches for doing the F-test.

Using Section 9.4.3

We can write the null hypothesis as

$$H_0 : L\beta = 0,$$

where $L = \begin{bmatrix} 0 & 1 & 1 & -1 & -1 \end{bmatrix}$. Using the formula in section 9.4.3, we have

$$F = \frac{(L\hat{\beta})^\top [L(X^\top X)^{-1} L^\top]^{-1} (L\hat{\beta})/1}{S_E^2},$$

where $S_E^2 = \mathrm{RSS}/(n-4-1)$. The first degree of freedom is 1 because we are testing one linear constraint, whiel the second degree of freedom is the usual $n-4-1$ because we have four explanatory variables and one intercept.

```
n <- dim(dat)[1]
k <- 4
X <- cbind(1, dat[,c("Knee", "Thigh", "Hip", "Ankle")])
X <- as.matrix(X)
y <- as.numeric(dat$bodyfat)
fit <- lm(bodyfat ~ Knee + Thigh + Hip + Ankle, data=dat)
beta_hat <- coef(fit)
RegSS <- sum((fitted(fit) - mean(y))^2)
RSS <- sum(resid(fit)^2)
SE2 <- RSS / (n - k - 1)
L <- matrix(c(0, 1,1,-1,-1), 1)
Fstat <- t(L %*% beta_hat) %*% solve(L %*% solve(t(X) %*% X) %*% t(L)) %*% L %*% beta_hat
Fstat <- Fstat / SE2
Fstat
```

```
##             [,1]
## [1,] 0.002813842
```

```
pval <- 1 - pf(Fstat, 1, n - k - 1)
pval
```

```
##           [,1]
## [1,] 0.9577384
```

Thus we do not have enough evidence to reject the null hypothesis.

Using the usual incremental sum of squares interpretation

The unusual formula for the F-statistic in the previous approach is a formula for the general "incremental sum of squares" definition of the F-statistic, as discussed in lecture and lab, which is

$$F = \frac{(\text{RegSS} - \text{RegSS}_0)/1}{\text{RSS}/(n-4-1)} = \frac{(\text{RSS}_0 - \text{RSS})/1}{\text{RSS}/(n-4-1)},$$

where RSS_0 and RegSS_0 denote the quantities for the null model (where the constraint $\beta_1 + \beta_2 = \beta_3 + \beta_4$ is enforced), and the other quantities are for the full model. The first degrees of freedom is 1 because the smaller model is obtained by a single linear constraint, while the second degrees of freedom is the usual $n - 4 - 1$ since we have four explanatory variables and an intercept term.

How do we fit the null model? We can write the null model as

$$\begin{aligned} \text{Bodyfat} &= \beta_0 + \beta_1\text{Knee} + \beta_2\text{Thigh} + \beta_3\text{Hip} + (\beta_1 + \beta_2 - \beta_3)\text{Ankle} && (1)\\ &= \beta_0 + \beta_1(\text{Knee} + \text{Ankle}) + \beta_2(\text{Thigh} + \text{Ankle}) + \beta_3(\text{Hip} - \text{Ankle}) && (2) \end{aligned}$$

Thus, the fit for the null model is simply the fit using only these three new "variables". [Note that below, I create these new variables during my call of `lm()` using the `I()` function. Alternatively/equivalently, one can create a "new" dataset with these new variables as columns, and do a fit using that dataset.]

```
fit0 <- lm(bodyfat ~ I(Knee + Ankle) + I(Thigh + Ankle) + I(Hip - Ankle), data=dat)
RSS0 <- sum(resid(fit0)^2)
RegSS0 <- sum((fitted(fit0) - mean(y))^2)
Fstat2 <- (RegSS - RegSS0) / (RSS / (n - k - 1))
Fstat3 <- (RSS0 - RSS) / (RSS / (n - k - 1))
Fstat2
```

```
## [1] 0.002813842
```

```
Fstat3
```

```
## [1] 0.002813842
```

```
pval <- 1 - pf(Fstat2, 1, n - k - 1)
pval
```

```
## [1] 0.9577384
```

We obtain the same results as before.

(b)

Using the definition of L above, we have

$$\text{Var}(\hat\beta_1 + \hat\beta_2 - \hat\beta_3 - \hat\beta_4) = \text{Var}(L\hat\beta) = L\text{Cov}(\hat\beta)L^\top = \sigma^2 L(X^\top X)^{-1}L^\top.$$

We can estimate σ^2 using $S_E^2 = \text{RSS}/(n-3-1)$. Thus, under the null hypothesis,

$$t = \frac{L\hat\beta}{S_E\sqrt{L(X^\top X)^{-1}L^\top}} = \frac{\hat\beta_1 + \hat\beta_2 - \hat\beta_3 - \hat\beta_4}{\sqrt{\frac{\text{RSS}}{n-4-1}}\sqrt{L(X^\top X)^{-1}L^\top}}$$

follows a t-distribution with $n - 4 - 1$ degrees of freedom.

```r
t <- L %*% beta_hat / (sqrt(SE2) * sqrt(L %*% solve(t(X) %*% X) %*% t(L)))
t
```

```
##            [,1]
## [1,] 0.05304566
```

```r
2 * (1 - pt(t, n - k - 1))
```

```
##           [,1]
## [1,] 0.9577384
```

Thus we do not have enough evidence to reject the null hypothesis.

(c)

$$t^2 = \frac{(L\hat{\beta})^\top (L\hat{\beta})}{S_E^2 L(X^\top X)^{-1} L^\top} = \frac{(L\hat{\beta})^\top [L(X^\top X)^{-1} L^\top]^{-1} (L\hat{\beta})}{S_E^2} = F$$

(Note that we can freely move $L(X^\top X)^{-1}L^\top$ around in the expression because it is a scalar in this case.)

```r
t^2
```

```
##             [,1]
## [1,] 0.002813842
```

```r
Fstat
```

```
##             [,1]
## [1,] 0.002813842
```