

# HW3

Timothy Schwieg

## Question 1

a

The Null hypothesis that sons are the same height as their fathers is the joint hypothesis that  $\beta_1 = 0$  and  $\beta_2 = 1$ . This would imply that  $\mathbb{E}[Y_n] = x_n$  which is the result we are seeking.

b

```
galtonData <- read.table( "Galton.dat", header=FALSE )
galtonData$V2 <- galtonData$V2 + 60
galtonData$V3 <- galtonData$V3 + 60
galtonData$V5 <- galtonData$V5 + 60
names( galtonData )

## [1] "V1" "V2" "V3" "V4" "V5"

sons <- galtonData[galtonData$V4 == 1,]

testFStat <- function( X, Y, R, r ){
  #Calculate The matrix in the most numerically sound method possible
  Q <- nrow( R )
  #No Numerical problems here officer
  betaHat <- solve( t(X)%*%X)%*%(t(X)%*%Y)
  E <- Y - X%*%betaHat
  #Seriously why is matrix multiplication so ugly in R
  sSquared <- (t(E)%*%E)/(nrow( sons ) - length(betaHat) )
  #print( as.numeric(sSquared)*(solve(t(X)%*%X)))
  #Bang out this bad boy
  Fstat <- (t(R%*%betaHat - r )%*%solve(R%*%solve(t(X)%*%X)%*%t(R))
    %*%(R%*%betaHat -r))/(Q*sSquared)

  #Finally get what we came for. After all only p-values matter
  pValue <- pf( Fstat, lower.tail = FALSE, df1 = Q, df2 = (nrow(sons)-length( betaHat)) )
  return( c( pValue, Fstat, betaHat ) )
}
sons$V6 <- rep( 1, nrow( sons ) )
```

```

X <- data.matrix( sons[,c(6,2)] )
Y <- data.matrix( sons$V5 )

R <- diag(2)
r <- c(0,1)

data <- testFStat( X, Y, R, r )

#P-Value
print( data[1] )

## [1] 6.878526e-27

#F-Stat
print( data[2] )

## [1] 68.49491

#Beta Hat
print( data[3:length(data)] )

## [1] 38.4753827 0.4448583

testReg <- lm( sons$V5 ~ sons$V2 )
nullReg <- lm( sons$V5 ~ sons$V2 - 1)

testStat <- 2*( logLik( testReg )[1] - logLik( nullReg )[1] )
pValue <- pchisq( testStat, df = 1, lower.tail = FALSE )
pValue

## [1] 4.158158e-28

```

As we can see, we can reject the null hypothesis at pretty much any significance level. This leads us to believe that there is some regression to the mean occurring, and that taller fathers have shorter sons and shorter fathers have taller sons.

## Question 2

**a**

We may first note that since the arithmetic mean is being used,  $x_n = \frac{x_1 + x_2}{2}$ . This could also be formatted as  $Y_n = \beta_1 + \frac{\beta_2}{2}x_1 + \frac{\beta_2}{2}x_2 + U_n$ . To test differing effects of mother and father heights on the models, we need to allow those coefficients to vary, so we will consider the slightly different model of:  $Y_n = \beta_1 + \beta_2x_1 + \beta_3x_2$ . In this framework, the hypothesis that sons are affected differently by their mother's height than by their father's height is:  $\beta_2 \neq \beta_3$

**b**

The Null hypothesis that we wish to test is that:  $\beta_2 = \beta_3$ .

```
X <- data.matrix( sons[,c(6,2,3)] )
Y <- data.matrix( sons$V5 )

R <- matrix( c(0,1,-1), ncol = 3 )
r <- c(0)

data <- testFStat( X, Y, R, r )

#P-Value
print( data[1] )

## [1] 0.1908053

#F-Stat
print( data[2] )

## [1] 1.716253

#Beta Hat
print( data[3:length(data)] )

## [1] 19.4929408 0.4156405 0.3279953

testReg <- lm( sons$V5 ~ sons$V2 + sons$V3 )
nullReg <- lm( sons$V5 ~ I(sons$V2 + sons$V3))

testStat <- 2*( logLik( testReg )[1] - logLik( nullReg )[1] )
pValue <- pchisq( testStat, df = 1, lower.tail = FALSE )
pValue

## [1] 0.1891877
```

As we can see by the high p-value on both tests, we fail to reject the null hypothesis unless we are at a very high threshold. We are left to conclude that it is possible that the height of mother's has the same effect on the height of a son as the height of father's.