

# Tutorial Assignment 2 - SOLUTIONS

Instructor: Kevin McGregor

MATH 4330

## Question 1:

```
# After reading in data using dat <- read.csv("...")
fit <- lm(FEV~Age+Height+Sex+Smoker, data=dat)
```

## Question 2:

```
summary(fit)

##
## Call:
## lm(formula = FEV ~ Age + Height + Sex + Smoker, data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.37656 -0.25033  0.00894  0.25588  1.92047
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.544220   0.232046 -19.583  < 2e-16 ***
## Age          0.065509   0.009489   6.904 1.21e-11 ***
## Height       0.104199   0.004758  21.901 < 2e-16 ***
## SexMale      0.157103   0.033207   4.731 2.74e-06 ***
## SmokerNon    0.087246   0.059254   1.472   0.141
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4122 on 649 degrees of freedom
## Multiple R-squared:  0.7754, Adjusted R-squared:  0.774
## F-statistic: 560 on 4 and 649 DF, p-value: < 2.2e-16
```

Here are the  $\beta$  parameters and their interpretations:

- $\beta_0 = -4.544220$ . Note that FEV (litres) cannot be negative. Also, it's impossible to have someone with height=0. Therefore, we say that  $\beta_0$  has no interpretation.
- $\beta_1 = 0.065509$ . When age increases by one year, the mean of FEV increases by 0.065509 litres, assuming all other predictors are held constant.
- $\beta_2 = 0.104199$ . When height increases by one inch, the mean of FEV increases by 0.104199 litres, assuming all other predictors are held constant.
- $\beta_3 = 0.157103$ . This measures the mean FEV difference between males and females. Males have FEV of 0.157103 litres higher than females, on average, assuming all other predictors are held constant.
- $\beta_4 = 0.087246$ . This measures the mean FEV difference between non-smokers and smokers. Be careful, nonsmoker=1 for nonsmokers. Non-smokers have FEV of 0.087246 litres higher than smokers, on average, assuming all other predictors are held constant.

## Question 3:

Suppose we choose significance level  $\alpha = 0.05$ . Examining the  $p$ -values for each of the predictors, we can see that age, height, sex, are all significant at this level, while smoking status is not. Thus, we have evidence that age, height, and sex are linearly related to FEV, and we do not have evidence that smoking status is linearly related to FEV.