Assignment 11

Max Wagner

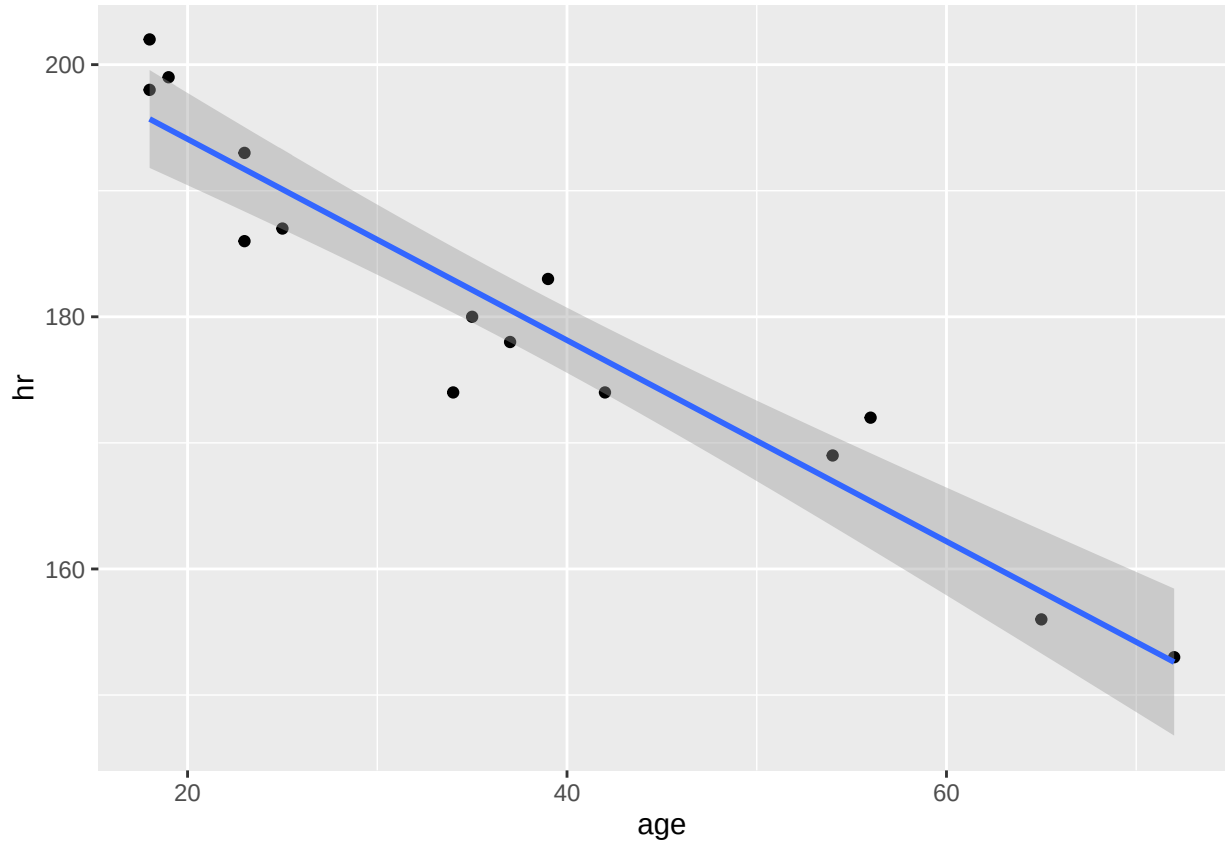
April 14, 2016

```
age <- c(18, 23, 25, 35, 65, 54, 34, 56, 72, 19, 23, 42, 18, 39, 37)
hr <- c(202, 186, 187, 180, 156, 169, 174, 172, 153, 199, 193,  174, 198, 183, 178)
df <- data.frame(age, hr)
summary(lm(hr ~ age, data = df))
```

```
##
## Call:
## lm(formula = hr ~ age, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.9258 -2.5383  0.3879  3.1867  6.6242
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 210.04846    2.86694   73.27  < 2e-16 ***
## age          -0.79773    0.06996  -11.40 3.85e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.578 on 13 degrees of freedom
## Multiple R-squared:  0.9091, Adjusted R-squared:  0.9021
## F-statistic:   130 on 1 and 13 DF,  p-value: 3.848e-08
```

The resulting equation is $210.0458 - .7977age$. The significance is significant and can be seen in the above as nearly 0. The plot can be seen below.

```
library(ggplot2)
ggplot(df, aes(x = age, y = hr)) + geom_point() + stat_smooth(method = lm)
```



First lets do all 392:

```
df <- scan("auto-mpg.data")
df <- data.frame(matrix(df, ncol = 5, byrow = TRUE))
colnames(df) <- c("dp", "hp", "wt", "acc", "mpg")
summary(lm(mpg ~ dp + hp + wt + acc, data = df))
```

```
##
## Call:
## lm(formula = mpg ~ dp + hp + wt + acc, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -11.378  -2.793  -0.333   2.193  16.256
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 45.2511397  2.4560447  18.424  < 2e-16 ***
## dp          -0.0060009  0.0067093  -0.894  0.37166
## hp          -0.0436077  0.0165735  -2.631  0.00885 **
## wt          -0.0052805  0.0008109  -6.512  2.3e-10 ***
## acc         -0.0231480  0.1256012  -0.184  0.85388
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.247 on 387 degrees of freedom
## Multiple R-squared:  0.707,  Adjusted R-squared:  0.704
## F-statistic: 233.4 on 4 and 387 DF,  p-value: < 2.2e-16
```

Horsepower is significant with a pvalue below .001, and weight is more significant with a pvalue of essentially 0. Displacement has a pvalue of .37 and acceleration has a pvalue of .85, which makes both variables insignificant to the model.

The standard errors for all 392 points are: dp = .0067093, hp = .0165735, wt = .0008109, and accel = .1256012.

$$mpg = 45.2511397 - 0.0060009dp - 0.0436077dp - 0.0052805wt - 0.0231480acc$$

And now picking a random 40 sample from the data.

```
df_samp <- df[sample(nrow(df), 40, replace = F),]
summary(lm(mpg ~ dp + hp + wt + acc, data = df_samp))
```

```
##
## Call:
## lm(formula = mpg ~ dp + hp + wt + acc, data = df_samp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.8429 -3.1512 -0.2789  3.5026 10.2936
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 35.795765   9.608203   3.726 0.000685 ***
## dp           0.014099   0.021297   0.662 0.512290
## hp          -0.036258   0.072094  -0.503 0.618168
## wt          -0.006532   0.003274  -1.995 0.053850 .
## acc          0.564578   0.495615   1.139 0.262379
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.982 on 35 degrees of freedom
## Multiple R-squared:  0.6371, Adjusted R-squared:  0.5957
## F-statistic: 15.36 on 4 and 35 DF,  p-value: 2.399e-07
```

The pvalues can be seen above, and 5 or so iterations I ran, none were ever significant. The standard errors can be read above as well, but are irrelavant as none of the variables are significant. The equation, although pointless, is below.

$$mpg = 42.205939 - .016186dp - .062396hp - .003724wt + .193301acc$$

If we view all 392 points, with only hp and wt included, the best solution can be found to be:

```
summary(lm(mpg ~ hp + wt, data = df))
```

```
##
## Call:
## lm(formula = mpg ~ hp + wt, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.0762  -2.7340  -0.3312   2.1752  16.2601
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 45.6402108  0.7931958  57.540  < 2e-16 ***
## hp          -0.0473029  0.0110851  -4.267 2.49e-05 ***
## wt          -0.0057942  0.0005023 -11.535  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.24 on 389 degrees of freedom
## Multiple R-squared:  0.7064, Adjusted R-squared:  0.7049
## F-statistic: 467.9 on 2 and 389 DF,  p-value: < 2.2e-16
```

$$mpg = 45.6402108 - .0473029hp - .0057942wt$$