Aldo Sunjoto || STAT 425 Section 1 UG || Steven A. Culpepper || #HW1

1 a) $RSS(\beta_1) = \sum_{i=1}^{n} (y_i - \beta_1 x_i)^2$

$RSS(\beta_1)' = \sum_{i=1}^{n} (\cancel{y_i^2}^{\,0} - 2\beta_1 x_i y_i + \beta_1^2 x_i^2)$

$= \sum_{i=1}^{n} (-2x_i y_i + 2\beta_1 x_i^2)$

$= -2\sum_{i=1}^{n} (x_i y_i - \beta_1 x_i^2) = 0$ //

b) $-2\sum_{i=1}^{n} (x_i y_i - \beta_1 x_i^2) = -2\sum_{i=1}^{n} x_i y_i + 2\sum_{i=1}^{n} \beta_1 x_i^2 = 0$

$= \cancel{2}\sum_{i=1}^{n} \beta_1 x_i^2 = \cancel{2}\sum_{i=1}^{n} x_i y_i$

$= \beta_1 = \dfrac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$ //

c) Assume that $X_1, X_2 \ldots X_n$ is not all zero. Because if the value for all x_i's are all zero, then it can lead to infinite solutions.

d) $RSS(\beta_1)' = -2\sum_{i=1}^{n} (\cancel{x_i y_i}^{\,0} - \beta_1 x_i^2) = 0$

$RSS(\beta_1)'' = +2\sum_{i=1}^{n} (+x_i^2)$

$= 2\sum_{i=1}^{n} x_i^2$

We assume that $2\sum_{i=1}^{n} x_i^2$ is greater than zero $(2\sum_{i=1}^{n} x_i^2 > 0)$

2a) $Y = X\beta + e$ (Gauss - Markov Conditions)

$E(e) = 0$ // $Var(e) = \sigma^2 I$ //

b) $E(Y) = E(X\beta + e)$ // substitute

$= X\beta + E(e)$ // we know $E(e)=0$

$= X\beta$ //

$Var(Y) = Var(X\beta + e)$ // substitute

$= Var(e)$

$= \sigma^2 I$ //

c) defn $\hat{\beta} = (X^T X)^{-1} X^T Y$

$E(\hat{\beta}) = E((X^T X)^{-1} X^T Y) = (X^T X)^{-1} X^T \overset{X\beta}{\cancel{E(Y)}} = \cancel{(X^T X)^{-1} X^T} X \beta = \beta$ //

$Var(\hat{\beta}) = Var((X^T X)^{-1} X^T Y) = \cancel{(X^T X)^{-1} X^T} \overset{\sigma^2 I}{\cancel{Var(Y)}} \cancel{X} (X^T X)^{-1}$

$= \sigma^2 (X^T X)^{-1}$ //

d) $E(\hat{Y}) = E(X\hat{\beta}) = X \overset{\beta}{\cancel{E(\hat{\beta})}} = X\beta$ //

$Var(\hat{Y}) = Var(X\hat{\beta}) = X\, Var(\hat{\beta})\, X^T = X \sigma^2 (X^T X)^{-1} X^T = \sigma^2 X (X^T X)^{-1} X^T$ //

hw 01

Aldo Sanjoto

```
#install.packages("alr4")
library("alr4")
```

```
## Loading required package: car
```

```
## Loading required package: effects
```

```
##
## Attaching package: 'effects'
```

```
## The following object is masked from 'package:car':
##
##     Prestige
```

```
data("ais")
head(ais)
```

```
##   Sex    Ht   Wt   LBM  RCC WCC   Hc   Hg Ferr   BMI   SSF  Bfat    Label
## 1   1 195.9 78.9 63.32 3.96 7.5 37.5 12.3   60 20.56 109.1 19.75 f-b_ball
## 2   1 189.7 74.4 58.55 4.41 8.3 38.2 12.7   68 20.67 102.8 21.30 f-b_ball
## 3   1 177.8 69.1 55.36 4.14 5.0 36.4 11.6   21 21.86 104.6 19.88 f-b_ball
## 4   1 185.0 74.9 57.18 4.11 5.3 37.3 12.6   69 21.88 126.4 23.66 f-b_ball
## 5   1 184.6 64.6 53.20 4.45 6.8 41.5 14.0   29 18.96  80.3 17.64 f-b_ball
## 6   1 174.0 63.7 53.77 4.10 4.4 37.4 12.5   42 21.04  75.2 15.58 f-b_ball
##    Sport
## 1 b_ball
## 2 b_ball
## 3 b_ball
## 4 b_ball
## 5 b_ball
## 6 b_ball
```

3a)

```
model = lm(formula = Wt ~ Sex + Ht + SSF + Bfat, data = ais)
summary(model)
```

```
##
## Call:
## lm(formula = Wt ~ Sex + Ht + SSF + Bfat, data = ais)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.8944  -4.7384  -0.7099   3.5957  31.6139
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -68.693175  11.579901  -5.932 1.32e-08 ***
## Sex         -14.017058   2.008283  -6.980 4.41e-11 ***
## Ht            0.759674   0.063985  11.873  < 2e-16 ***
## SSF           0.199369   0.071631   2.783  0.00591 **
## Bfat          0.004401   0.439623   0.010  0.99202
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.858 on 197 degrees of freedom
## Multiple R-squared:  0.7623, Adjusted R-squared:  0.7574
## F-statistic: 157.9 on 4 and 197 DF,  p-value: < 2.2e-16
```

3b) $\beta_{int} = -68.693175$ $\beta_{Ht} = 0.759674$ $\beta_{Bfat} = 0.004401$

$\therefore \beta_{sex} = -14.017058$ $\beta_{SSF} = 0.199369$

3c)

```
summary(model)$r.squared
```

```
## [1] 0.7622643
```

3d)

```
which.max(model$residuals)
```

```
## 163
## 163
```

```
model$fitted.values[which.max(model$residuals)]
```

```
##      163
## 91.58611
```

3 e) According to the model $\beta_{sex} = -14.0171$. Thus, the weight for a male will be larger than female by 14.0171 kg.

3 f) Sex, Ht and SSF.

3 g)

```
confint(model)
```

```
##                    2.5 %      97.5 %
## (Intercept) -91.52965492 -45.8566953
## Sex         -17.97755074 -10.0565662
## Ht            0.63348925   0.8858580
## SSF           0.05810749   0.3406307
## Bfat         -0.86257156   0.8713728
```

3 h)

```
#default is 5% siglevel
new_data = data.frame(Sex = 1, Ht = 170, SSF = 60, Bfat = 12)
predict.lm(model, newdata = new_data, interval = 'prediction')
```

```
##        fit      lwr     upr
## 1 58.44924 44.75327 72.1452
```

Predicted value: 58.45 Interval: (44.75, 72.15)

3 i)

```
model_new = lm(Wt ~ Sex + Ht, data = ais)
anova(model, model_new)
```

```
## Analysis of Variance Table
##
## Model 1: Wt ~ Sex + Ht + SSF + Bfat
## Model 2: Wt ~ Sex + Ht
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1    197  9266.5
## 2    199 14548.6 -2     -5282 56.146 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

HW 0~~2~~ 1

Aldo Sanjoto

September 15, 2017

```
library("alr4")
```

```
## Loading required package: car
```

```
## Loading required package: effects
```

```
##
## Attaching package: 'effects'
```

```
## The following object is masked from 'package:car':
##
##     Prestige
```

4a)

```
data("fuel2001")
model = lm(formula = FuelC ~ Income + Pop + Tax, data = fuel2001)
summary(model)
```

```
##
## Call:
## lm(formula = FuelC ~ Income + Pop + Tax, data = fuel2001)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2696785  -113067    -7653   141991  1668076
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.698e+06  5.820e+05   2.918  0.00539 **
## Income      -3.083e+01  1.723e+01  -1.789  0.07998 .
## Pop          5.638e-01  1.646e-02  34.258  < 2e-16 ***
## Tax         -3.373e+04  1.645e+04  -2.051  0.04587 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 522600 on 47 degrees of freedom
## Multiple R-squared:  0.9647, Adjusted R-squared:  0.9625
## F-statistic: 428.3 on 3 and 47 DF,  p-value: < 2.2e-16
```

4b) $H_0: \beta_{income} = 0$ $H_A: \beta_{income} \neq 0$

- T-value = -1.789 with DF = 47.
- P value = 0.07998, so we fail to reject the null

4c)

- we know that for $H_0: \beta_{income} = \beta_{pop} = \beta_{Tax} = 0$ have
- DF of 3 and 47 w/ p-value < 2.2e-16. Thus, we'll
- reject the null hypothesis.

4d)

```
model_dri = lm(FuelC ~ Income + Pop + Tax + Drivers, data = fuel2001)
summary(model_dri)
```

```
##
## Call:
## lm(formula = FuelC ~ Income + Pop + Tax + Drivers, data = fuel2001)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1697519  -118108   -17214   122668  1849795
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.231e+06  5.231e+05   2.353 0.022954 *
## Income      -2.644e+01  1.512e+01  -1.749 0.086930 .
## Pop          1.666e-02  1.403e-01   0.119 0.906006
## Tax         -2.204e+04  1.470e+04  -1.499 0.140648
## Drivers      6.496e-01  1.657e-01   3.920 0.000292 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 457400 on 46 degrees of freedom
## Multiple R-squared:  0.9735, Adjusted R-squared:  0.9712
## F-statistic: 423.2 on 4 and 46 DF,  p-value: < 2.2e-16
```

(e)

```
anova(model, model_dri)
```

```
## Analysis of Variance Table
##
## Model 1: FuelC ~ Income + Pop + Tax
## Model 2: FuelC ~ Income + Pop + Tax + Drivers
##   Res.Df        RSS Df Sum of Sq      F    Pr(>F)
## 1     47 1.2838e+13
## 2     46 9.6235e+12  1 3.215e+12 15.368 0.0002923 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the model of analysis above, the p-value is 0.0002923. So there's a strong evidence that $\beta_{drivers} = 0$

(f)

Yes they are the same. In d) p-value for the t-test of H_0: $\beta_{drivers} = 0$ is 0.000292 which is the same for part (e).