# 2014

## Question 1

### a)

1. $\beta_0 = 47.989, \beta_1 = 2.686$
2. $\hat{\sigma}^2 = 6.833^2 = 46.69$

### b)

$$R^2 = 0.9192 = 91.92\%$$

### c)

$F$-statistic's value is 79.64. With a $p$-value of $4.507 \times 10^{-5}$, and at a significance level of 5%, we have enough evidence to reject the null hypothesis. This indicates that there is a statistically significant linear relationship - and furthermore that the intercept explains a statistically significant amount of variation.

### d)

The $t$-statistic's value is 8.924. To test a one-sided alternative, you would divide the $p$-value in half, as the $t$-distribution is symmetric.

### e)

You can derive these yourselves, or it's contained in the lecture slides for week 2. To derive it yourself, CI for mean interval is about $\hat{y}(x_0)$, and the prediction interval is about $y(x_0) - \hat{y}(x_0)$.

$$\text{Prediction interval} = \hat{y}(x_0) \pm t_{\alpha/2,n-p}\hat{\sigma}\left(1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}\right)$$

$$\text{Confidence interval} = \hat{y}(x_0) \pm t_{\alpha/2,n-p}\hat{\sigma}\left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}\right)$$

Therefore, we have:

1. $101.79 \pm 2.364 \cdot 6.833 \cdot \left(1 + \frac{1}{9} + \frac{(20-21.37)^2}{463.24}\right) = (83.70, 119.72)$
2. $101.79 \pm 2.364 \cdot 6.833 \cdot \left(\frac{1}{9} + \frac{(20-21.37)^2}{463.24}\right) = (99.85, 103.57)$

# Question 2

## a)

We'd stop at `rings ~ length`, as at this step, theres no more predictors with a $p$-value $< 0.05$.

## b)

1. $\gamma$ is 46.401. You can find this by considering that `<none>` is just the base model of that step, and then look in the previous step, for what the `RSS` was there.

   $\delta$ is 23.676. Reminder that sequential sum of squares is the difference between $SS^{(2)}_{reg/res} - SS^{(1)}_{reg/res}$. You can find this by minusing `<none>`'s RSS to `data$weight`'s RSS in the null step.

2. A reminder that
$$F = \frac{R(\beta^{(2)}|R(\beta^{1}))}{SS_{reg}} \sim F_{1,n-r}$$
   where $r$ is the total number of parameters estimated in $\beta^{(2)}|\beta^{(1)}$. We have 14 samples, and in the first step for $\alpha$, 2 parameters for each line being estimated (new parameter + intercept).

   So, we have:
$$\frac{\alpha}{51.538/(14-2)} = 5.5127$$
   and thus that $\alpha = 23.68$.

   In a similar vein, $\beta$ can be found by:
$$\beta = \frac{0.17801}{46.223/(14-4)} = 0.385$$
   Just don't forget $p = k + 1$, where there are $k$ predictors (+ intercept!)

3. So the models here are:
$$M_0 : y = \beta_0 + \beta_{length}$$
   and
$$M_1 : y = \beta_0 + \beta_{length} + \beta_{diameter} + \beta_{weight}$$
   This means we want to check whether $\beta_{diameter} = \beta_{weight} = 0$. We can check this by doing:
$$F = \frac{R(\beta_{diameter}, \beta_{weight}|\beta_{length})/2}{SS_{reg}/(14-4)} \sim F_{2,10}$$
   The $F$-statistic equals:
$$\frac{R(\beta_{diameter}|\beta_{length}) + R(\beta_{weight}|\beta_{diameter}, \beta_{length})}{46.223/10} = \frac{2.309505}{4.6623} = 0.495$$
   Since $0.495 < 4.102821$, **we cannot reject the null hypothesis**.

# Question 3

## a)

PRESS considers out of sample prediction, $C_p$ considers subset selection bias-variance tradeoff, Adj. $R^2$ and $R^2$ both consider the total variation the regression component captures; where *adjusted* scales the regression component by the number of parameters used.

## b)

$$R^2 = 1 - \frac{SS_{res}}{SS_{total}}$$

$$\bar{R}^2 = 1 - \frac{(n-1)SS_{res}}{(n-p)SS_{total}} = 1 - \frac{n-1}{n-p}(1 - R^2)$$

Now whenever $\frac{n-1}{n-p}(1-R^2) > 1$, we have that $\bar{R}^2 < 0$.

# Question 4

## a)

$$\begin{aligned}
E(y^T Ay) &= \sum_{i=1}^{n}\sum_{j=1}^{n} E(y_j A_{ji} y_i) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ji} E(y_j y_i) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ji}(V_{ij} + \mu\mu) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ji}V_{ij} + \mu A_{ji}\mu \\
&= tr(AV) + \mu^T a\mu
\end{aligned}$$

## b)

$$\begin{aligned}
E\left(\sum_{i=1}^{k} Z_i^2\right) &= \sum_{i=1}^{k} E(Z_i^2) \\
&= \sum_{i=1}^{k} \text{Var}(Z_i) + E(Z_i)^2 \\
&= \sum_{i=1}^{k} \sigma^2 + \mu^2 \\
&= k(\sigma^2 + \mu^2)
\end{aligned}$$

$\sum_{i=1}^{k} Z_i^2 = \chi_k^2$, so the expectations are the same.

# Question 5

I'm almost certain questions like c) and d) aren't asked anymore. If they ask us this... I mean, we've learnt 0 MGF stuff throughout the term, been asked almost 0 questions (besides 1 tute question), and have not used it in lectures.

## a)

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-2}\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)^2 \\
&= \ldots (y_i - \bar{y} + b_1\bar{x} - b_1 x_i)^2 \\
&= \ldots (y_i - \bar{y} - b_1(x_i - \bar{x}))^2 \\
&= \ldots (y_i - \bar{y})^2 - 2b_1(x_i - \bar{x})(y_i - \bar{y}) + b_1^2(x_i - \bar{x})^2 \\
&= \frac{S_{yy} - 2b_1 S_{xy} + b_1^2 S_{xx}}{n-2} \\
&= \frac{S_{yy} - b_1^2 S_{xx}}{n-2} \qquad \text{Since } S_{xy} = b_1^2 S_{xx}
\end{aligned}
$$

## b)

$$
\begin{aligned}
E(S_{yy}) &= \sum_{i=1}^{n} E[(y_i - \bar{y})^2] \\
&= \sum_{i=1}^{n} \text{Var}(y_i - \bar{y}) + \sum_{i=1}^{n} E(y_i - \bar{y})^2 \\
E(y_i - \bar{y}) &= E(\beta_0 + \beta_1 x_i + \epsilon_i - \beta_0 - \beta_1 \bar{x}) \\
&= \beta_1(x_i - \bar{x}) \\
E(S_{yy}) &= \sum_{i=1}^{n} \text{Var}(y_i - \bar{y}) + \beta_1^2 S_{xx} \\
\text{Var}(y_i - \bar{y}) &= \text{Var}(y_i) + \text{Var}(\bar{y}) - 2\text{Cov}(y_i, \bar{y}) \\
&= \sigma^2 + \sigma^2/n - 2\sigma^2/n \\
&= \sigma^2 - \sigma^2/n \\
E(S_{yy}) &= n\sigma^2 - \sigma^2 + \beta_1^2 S_{xx}
\end{aligned}
$$

Now considering the other component:

$$
\begin{aligned}
E(b_1^2 S_{xx}) &= S_{xx} E(b_1^2) \\
&= S_{xx}(\text{Var}(b_1) + [E(b_1)]^2) \\
&= S_{xx}(\sigma^2/S_{xx} + \beta_1^2) \\
&= \sigma^2 + \beta_1^2 S_{xx}
\end{aligned}
$$

This combines to $(n-2)\sigma^2$.

## Question 6

a, b) are fairly uninteresting. These are pretty standard MLE derivations - this specific question can be found in one of the week's tutorials.

### c)

Best, Linear and Unbiased Estimator. Best, alluding to lowest variance. Linear, and unbiased, which means it's expectation must be $\beta$.

### d)

Consider an alternative estimator $\hat{b} = ((X^TX)^{-1}X^T + C)y$. It must unbiased so:

$$
\begin{aligned}
E(b + Cy) &= \beta + CX\beta \\
\beta + CX\beta &= \beta \qquad \text{[Since it must be unbiased]} \\
CX\beta &= 0 \\
CX &= 0
\end{aligned}
$$

Now considering the variance:

$$
\begin{aligned}
\text{Var}(((X^TX)^{-1}X^T + C)y) &= ((X^TX)^{-1}X^T + C)\text{Var}(y)(X(X^TX)^{-1} + C^T) \\
&= \sigma^2((X^TX)^{-1} + (X^TX)^{-1}(CX)^T + CX(X^TX)^{-1} + CC^T) \\
&= \sigma^2((X^TX)^{-1} + CC^T) \qquad \text{[From the above derived identities]}
\end{aligned}
$$

Note $CC^T \sum_{i=1}^n \sum_{j=1}^n C_{ij}^2$, so $\text{Var}(\hat{b}) \geq \text{Var}(b)$. Thereby, BLUE.

## Question 7

### a)

$$
\begin{pmatrix} y_1 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ \dots & \dots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \dots \\ \epsilon_n \end{pmatrix}
$$

### b)

$$
\begin{aligned}
\text{Var}(b_1) &= \text{Var}\left(\frac{S_{xy}}{S_{xx}}\right) \\
&= \frac{1}{S_{xx}^2}\text{Var}\left(\sum_{i=1}(x_i - \bar{x})y_i\right) \\
&= \frac{\sigma^2}{S_{xx}} \\
\text{Var}(b_0) &= \frac{\sigma^2}{n} + \frac{\sigma^2}{S_{xx}}
\end{aligned}
$$

For the covariances, watch out for $\sum_{i=1}^n (x_i - \bar{x}) = 0$.

### c)

$$
\begin{aligned}
\text{Var}(\hat{y}(x_0)) &= \text{Var}(b_0 + b_1x_0) \\
&= \text{Var}(b_0) + x_0^2\text{Var}(b_1) + 2x_0\text{Cov}(b_0, b_1) \\
&= \frac{\sigma^2}{n} + \frac{\sigma^2}{S_{xx}} + \frac{x_0^2\sigma^2}{S_{xx}} - \frac{2x_0\bar{x}\sigma^2}{S_{xx}} \\
&= \sigma^2\left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}\right)
\end{aligned}
$$

## d)

Trick for this question is to recognise the variance formula from the last question. We want to use:

$$\mathrm{Var}(x_i^T b) = \sigma^2 \left( \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right)$$

Now note the $i$-th entry for $H$ can be represented as:

$$h_{ii} = x_i^T (X^T X)^{-1} x_i$$

where $x_i = (x_{i1}, x_{i2}, \dots, x_{in})^T$. Note the flip of the transposes, due to this $i$-th row definition. Okay, now just:

$$\begin{aligned}
\mathrm{Var}(x_i^T b) &= x_i^T \mathrm{Var}(b) x_i \\
&= \sigma^2 x_i^T (X^T X)^{-1} x_i \\
&= \sigma^2 h_{ii}
\end{aligned}$$

Thereby completing the proof (compare coefficients).