Handout 1

STAT3530 Applied Linear Models

Work through Part 1 on your own or with a neighbor. Parts 2 and 3 follow the lecture slides, so you can fill them in as we go. Nothing here requires a computer.

For a sample $x_1, \ldots, x_n$, recall:

Quantity	Definition
sample mean	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
sample standard deviation	$s_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

Sums like $\sum_i (x_i - \bar{x})^2$ appear constantly, so it is worth remembering that we can write

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = (n-1)s_x^2$$

Part 1: Warm-up

Use the following data for questions 1–5.

i	1	2	3	4	5
x_i	1	2	3	4	5
$x_i - \bar{x}$					
y_i	3	5	4	8	10
$y_i - \bar{y}$					

1. Compute $\bar{x}$ and $\bar{y}$, then fill in the deviations $x_i - \bar{x}$ and $y_i - \bar{y}$ in the table above.

2. Compute $\sum_i (x_i - \bar{x})$.

3. Compute $\sum_i (x_i - \bar{x})^2$ and $\sum_i (x_i - \bar{x})(y_i - \bar{y})$.

4. Now compute $\sum_i (x_i - \bar{x}) y_i$ — note that the y_i here are *not* centered. Compare with your answer to question 3. Can you explain what you see?

5. *Optional, if time allows.* The least squares estimates for these data are $\hat{\beta}_0 = 0.9$ and $\hat{\beta}_1 = 1.7$. Verify that $\hat{\beta}_1 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}$ and that $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$.

Part 2: Simple linear regression examples

Cognitive function and age

The RFFT measures nonverbal fluency and executive cognitive function on a 0–175 scale. Below are $n = 208$ observations of RFFT score and age.

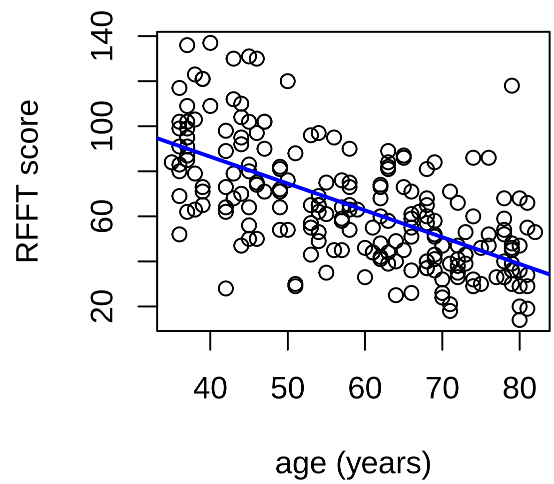
```
fit <- lm(rfft ~ age, data = prevend)
print(fit)
```

```
Call:
lm(formula = rfft ~ age, data = prevend)

Coefficients:
(Intercept)          age
    134.098       -1.191
```

```
sigma(fit)
```

```
[1] 20.51788
```



6. Using the output above:
   (a) $\hat{\beta}_0 =$ ____________ , $\hat{\beta}_1 =$ ____________ , $\hat{\sigma} =$ ____________
   (b) Write out the fitted model equation.
   (c) Interpret each of the three estimates in context.

Stopping distance and speed

The `cars` data record the speed of $n = 50$ cars and the distance each took to come to a stop.

```r
fit_cars <- lm(dist ~ speed, data = cars)
print(fit_cars)
```

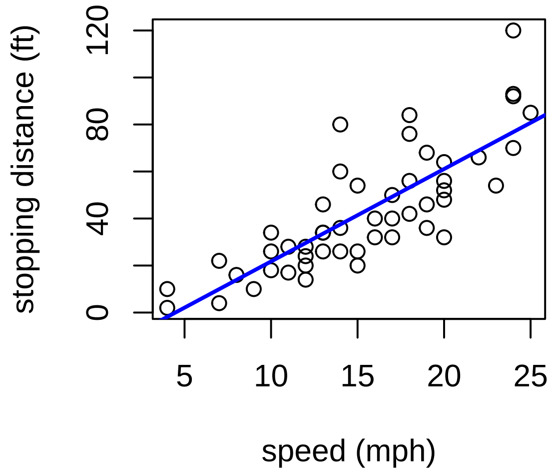
```

Call:
lm(formula = dist ~ speed, data = cars)

Coefficients:
(Intercept)        speed
    -17.579        3.932
```

```r
sigma(fit_cars)
```

```
[1] 15.37959
```



7. Using the output above:

   (a) $\hat{\beta}_0 =$ ____________ , $\hat{\beta}_1 =$ ____________ , $\hat{\sigma} =$ ____________

   (b) Write out the fitted model equation.

   (c) Interpret each of the three estimates in context.

Part 3: Vectors and matrices

Recall that an n-dimensional vector $a \in \mathbb{R}^n$ is a column of n numbers, that an $n \times m$ matrix $A \in \mathbb{R}^{n\times m}$ has n rows and m columns, and that the inner product of $a, b \in \mathbb{R}^n$ is $\langle a, b\rangle = a^T b = \sum_{i=1}^n a_i b_i$.

The vectors and matrices below are the same ones used in the slides, so you can fill in the blanks as we go.

$$a = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}, \qquad b = \begin{bmatrix} 3 \\ 5 \\ 6 \end{bmatrix}, \qquad A = \begin{bmatrix} 1 & 0 & 3 \\ 4 & -2 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 1 \\ -1 & 2 \\ 1 & 0 \end{bmatrix}$$

8. $a \in \mathbb{R}^n$ for $n =$ _________ . Now compute the inner product $\langle a, b\rangle = a^T b =$ _________ .
9. $A \in \mathbb{R}^{n\times m}$ with $n =$ _________ and $m =$ _________ , so A is $A_{\_\_\times\_\_}$.

10. Which of the following are defined? For those that are, give the dimensions *without* computing the product.

$$AB, \qquad BA, \qquad A^T B, \qquad A^T A$$

In the slides we had $AB = \begin{bmatrix} 3 & 1 \\ 2 & 0 \end{bmatrix}$. What is the $(1, 1)$ entry of BA?

11. Explain why $A^T A$ must be symmetric without computing it. Does your argument work for *any* matrix A, whatever its dimensions?

12. Using the column representation of A, verify that $a_3 = 3a_1 + 6a_2$. Are a_1 and a_3 linearly independent? What is $\text{rank}(A)$?

Part 4: Random variables and vectors

Throughout, Y, Y_1, and Y_2 are random variables.

13. If $\mathbb{E}[Y] = 2$ and $\text{Var}[Y] = 5$, then $\mathbb{E}[3Y + 5] =$ ________ and $\text{Var}[3Y + 5] =$ ________ .

14. If $Y \sim N(\mu, \sigma^2)$, then $Y - \mu \sim N($ ________ , ________ $)$.

15. Suppose $\mathbb{E}[Y_1] = 1$, $\mathbb{E}[Y_2] = 4$, $\text{Var}[Y_1] = 2$, $\text{Var}[Y_2] = 3$, and Y_1, Y_2 are independent. Compute $\mathbb{E}[2Y_1 - Y_2]$ and $\text{Var}[2Y_1 - Y_2]$.

16. Now suppose instead that $\text{Cov}[Y_1, Y_2] = 1$. Which of your two answers to question 15 changes, and what does it become?

17. Let Y be an $n \times 1$ random vector with $\text{Var}[Y] = \sigma^2 I$. Fill in the dimensions: $\mathbb{E}[Y]$ is _____×_____ and $\text{Var}[Y]$ is _____×_____ . If A is $2 \times n$, then AY is _____×_____ and $\text{Var}[AY]$ is _____×_____ .