Stat 5120: Assignment 1

Due on September 12, 2018 at 4 pm

Professor Chao DU Section 001

Ryan Sanders

Problem 1

Part (a) Based in Figure 3, a simple linear regression appears to be reasonable for the data.

Figure 3 shows a graph.

Table 1: Regression Results

	Dependent variable:
	Fare
Distance	0.220^{***}
	(0.004)
Constant	48.972^{***}
	(4.405)
Observations	17
R^2	0.994
Adjusted R^2	0.994
Residual Std. Error	10.412 (df = 15)
F Statistic	$2,469.296^{***}$ (df = 1; 15)
Note:	*p<0.1; **p<0.05; ***p<0.01

Part (b) The estimated equation for the linear regression

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\epsilon}_i \tag{1}$$

is

$$\hat{y} = 48.972 + 0.220x$$

with estimated slope $\hat{\beta}_1 = 0.220$ and constant$\hat{\beta}_0 = 0.48.972$.

Part (c) The coefficient of determination, $R^2 = 0.994$, suggests a high degree of correlation against Cost of Fare and Distance Traveled.
Part (d) The estimated correlation coefficient, $R^2 = 0.997$, suggests a high degree of positvie colinearity against Cost of Fare and Distance Traveled.
Part (e) Based on the Standard Error shown in Table **??**, $\hat{\sigma} = 10.42$, gives $\hat{\sigma^2} = 108.58$.

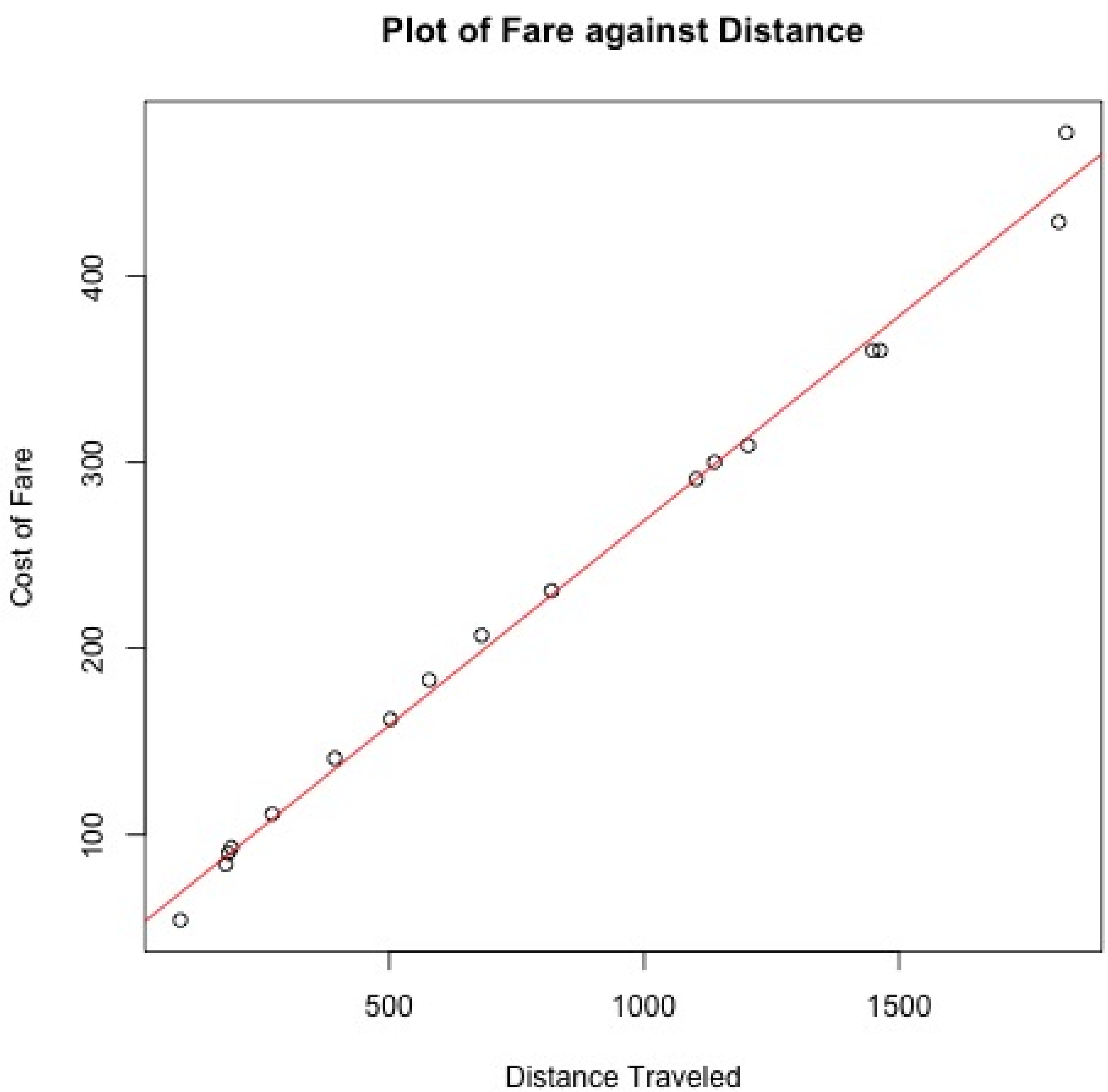

Figure 1: A graph.

Problem 2

Let $\Sigma = \{0, 1\}$. Construct a DFA A that recognizes the language that consists of all binary numbers that can be divided by 5.

Let the state q_k indicate the remainder of k divided by 5. For example, the remainder of 2 would correlate to state q_2 because $7 \mod 5 = 2$.

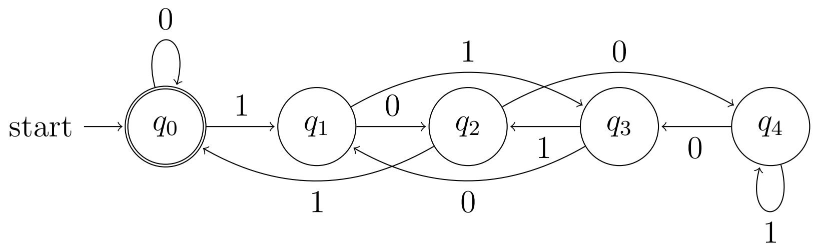

Figure 2: DFA, A, this is really beautiful, ya know?

Justification

Take a given binary number, x. Since there are only two inputs to our state machine, x can either become $x0$ or $x1$. When a 0 comes into the state machine, it is the same as taking the binary number and multiplying it by two. When a 1 comes into the machine, it is the same as multipying by two and adding one.

Using this knowledge, we can construct a transition table that tell us where to go:

	$x \mod 5 = 0$	$x \mod 5 = 1$	$x \mod 5 = 2$	$x \mod 5 = 3$	$x \mod 5 = 4$
$x0$	0	2	4	1	3
$x1$	1	3	0	2	4

Therefore on state q_0 or ($x \mod 5 = 0$), a transition line should go to state q_0 for the input 0 and a line should go to state q_1 for input 1. Continuing this gives us the Figure 2.

Problem 3

Write part of **Quick-Sort**($list, start, end$)

```
function Quick-Sort(list, start, end)
    if start ≥ end then
        return
    end if
    mid ← Partition(list, start, end)
    Quick-Sort(list, start, mid − 1)
    Quick-Sort(list, mid + 1, end)
end function
```

Algorithm 1: Start of QuickSort

Problem 4

Suppose we would like to fit a straight line through the origin, i.e., $Y_i = \beta_1 x_i + e_i$ with $i = 1, \ldots, n$, $\mathrm{E}[e_i] = 0$, and $\mathrm{Var}[e_i] = \sigma_e^2$ and $\mathrm{Cov}[e_i, e_j] = 0, \forall i \neq j$.

Part A

Find the least squares esimator for $\hat{\beta}_1$ for the slope β_1.

Solution

To find the least squares estimator, we should minimize our Residual Sum of Squares, RSS:

$$\begin{aligned} RSS &= \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 \\ &= \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_1 x_i\right)^2 \end{aligned}$$

By taking the partial derivative in respect to $\hat{\beta}_1$, we get:

$$\frac{\partial}{\partial \hat{\beta}_1}(RSS) = -2 \sum_{i=1}^{n} x_i (Y_i - \hat{\beta}_1 x_i) = 0$$

This gives us:

$$\begin{aligned} \sum_{i=1}^{n} x_i (Y_i - \hat{\beta}_1 x_i) &= \sum_{i=1}^{n} x_i Y_i - \sum_{i=1}^{n} \hat{\beta}_1 x_i^2 \\ &= \sum_{i=1}^{n} x_i Y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 \end{aligned}$$

Solving for $\hat{\beta}_1$ gives the final estimator for β_1:

Problem 4 continued on next page...

$$\hat{\beta}_1 = \frac{\sum x_i Y_i}{\sum x_i^2}$$

Problem 4 continued on next page. . .

Part B

Calculate the bias and the variance for the estimated slope $\hat{\beta}_1$.

Solution

For the bias, we need to calculate the expected value $\mathrm{E}[\hat{\beta}_1]$:

$$\begin{aligned}
\mathrm{E}[\hat{\beta}_1] &= \mathrm{E}\left[\frac{\sum x_i Y_i}{\sum x_i^2}\right] \\
&= \frac{\sum x_i \mathrm{E}[Y_i]}{\sum x_i^2} \\
&= \frac{\sum x_i(\beta_1 x_i)}{\sum x_i^2} \\
&= \frac{\sum x_i^2 \beta_1}{\sum x_i^2} \\
&= \beta_1 \frac{\sum x_i^2 \beta_1}{\sum x_i^2} \\
&= \beta_1
\end{aligned}$$

Thus since our estimator's expected value is β_1, we can conclude that the bias of our estimator is 0.

For the variance:

$$\begin{aligned}
\mathrm{Var}[\hat{\beta}_1] &= \mathrm{Var}\left[\frac{\sum x_i Y_i}{\sum x_i^2}\right] \\
&= \frac{\sum x_i^2}{\sum x_i^2 \sum x_i^2}\mathrm{Var}[Y_i] \\
&= \frac{\sum x_i^2}{\sum x_i^2 \sum x_i^2}\mathrm{Var}[Y_i] \\
&= \frac{1}{\sum x_i^2}\mathrm{Var}[Y_i] \\
&= \frac{1}{\sum x_i^2}\sigma^2 \\
&= \frac{\sigma^2}{\sum x_i^2}
\end{aligned}$$

Problem 5

Prove a polynomial of degree k, $a_k n^k + a_{k-1}n^{k-1} + \ldots + a_1 n^1 + a_0 n^0$ is a member of $\Theta(n^k)$ where $a_k \ldots a_0$ are nonnegative constants.

Proof. To prove that $a_k n^k + a_{k-1}n^{k-1} + \ldots + a_1 n^1 + a_0 n^0$, we must show the following:

$$\exists c_1 \exists c_2 \forall n \geq n_0,\ c_1 \cdot g(n) \leq f(n) \leq c_2 \cdot g(n)$$

For the first inequality, it is easy to see that it holds because no matter what the constants are, $n^k \leq a_k n^k + a_{k-1}n^{k-1} + \ldots + a_1 n^1 + a_0 n^0$ even if $c_1 = 1$ and $n_0 = 1$. This is because $n^k \leq c_1 \cdot a_k n^k$ for any nonnegative constant, c_1 and a_k.

Taking the second inequality, we prove it in the following way. By summation, $\sum\limits_{i=0}^{k} a_i$ will give us a new constant, A. By taking this value of A, we can then do the following:

$$\begin{aligned} a_k n^k + a_{k-1}n^{k-1} + \ldots + a_1 n^1 + a_0 n^0 = & \\ & \leq (a_k + a_{k-1} \ldots a_1 + a_0) \cdot n^k \\ & = A \cdot n^k \\ & \leq c_2 \cdot n^k \end{aligned}$$

where $n_0 = 1$ and $c_2 = A$. c_2 is just a constant. Thus the proof is complete. □

Problem 18

Evaluate $\sum_{k=1}^{5} k^2$ and $\sum_{k=1}^{5}(k-1)^2$.

Problem 19

Find the derivative of $f(x) = x^4 + 3x^2 - 2$

Problem 6

Evaluate the integrals $\int_0^1 (1-x^2)\mathrm{d}x$ and $\int_1^{\infty} \frac{1}{x^2}\mathrm{d}x$.
Figure 3 shows a graph.

Table 2: Regression Results

	Dependent variable:
	Fare
Distance	0.220***
	(0.004)
Constant	48.972***
	(4.405)
Observations	17
R^2	0.994
Adjusted R^2	0.994
Residual Std. Error	10.412 (df = 15)
F Statistic	2,469.296*** (df = 1; 15)
Note:	*p<0.1; **p<0.05; ***p<0.01

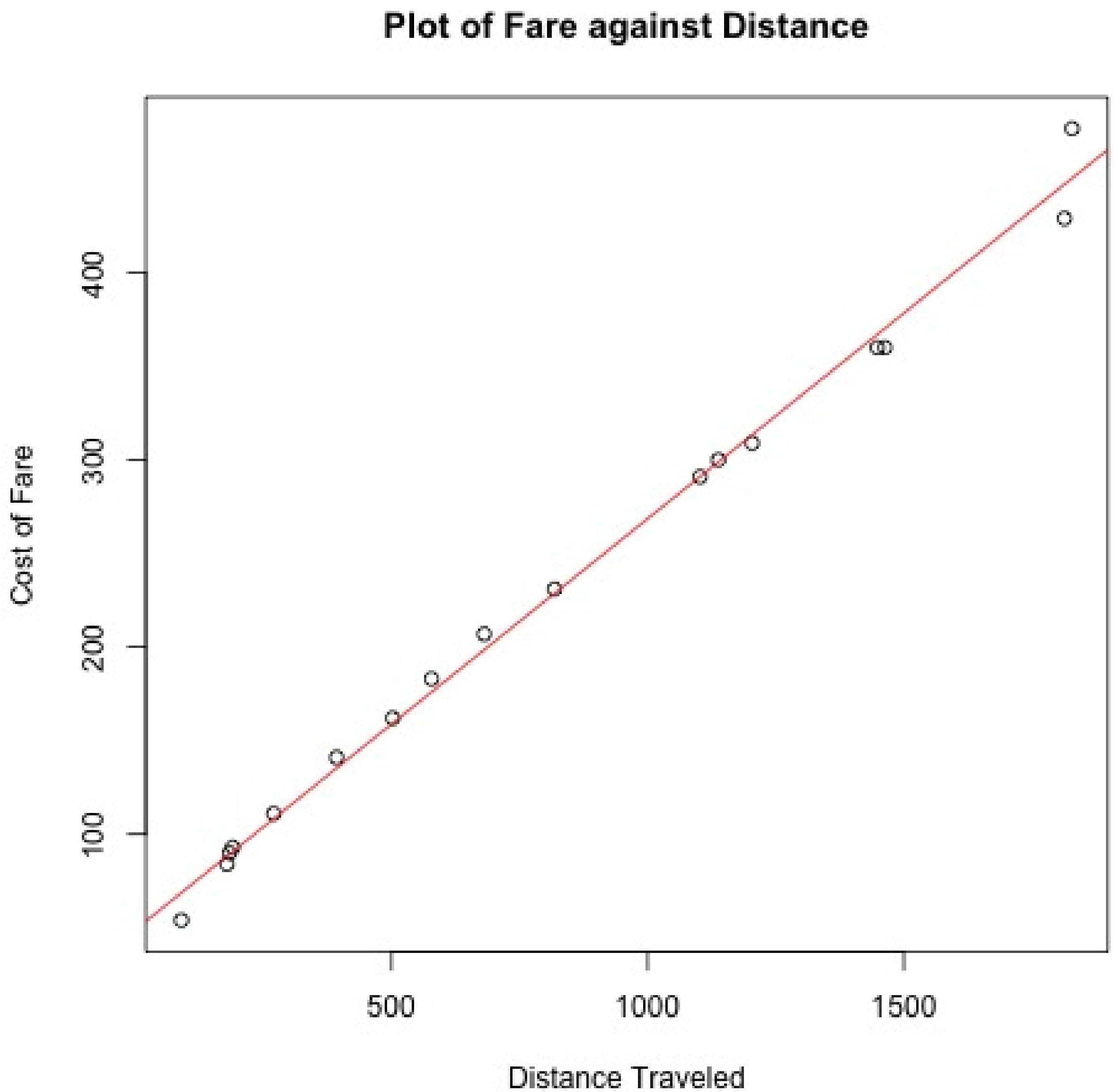

Figure 3: A graph.