Stats 519: Homework 1 - Introduction to R

Due on January 26, 2009

Dr. Stephen Lee 1:30

Monte J. Shaffer

Problem 1

Basic Calculations Using R

- $27(38 - 17)$

Program 1 Simple Calculation

```
> 27*(38-17);
[1] 567
```

- $ln(14^7)$

Program 2 Natural Log of a Power

```
> log(14^7);
[1] 18.4734
```

- $\sqrt{\frac{436}{12}}$

Program 3 Square Root of a Fraction

```
> sqrt(436/12);
[1] 6.027714
```

Problem 2

Vector Arithmetic

Program 4 Defining a, b, and d as product

```
> a=seq(5,160,by=5);
> a;
 [1]   5  10  15  20  25  30  35  40  45  50  55  60  65  70  75  80  85  90 ... 160
> b=seq(87,56,by=-1);
> b;
 [1] 87 86 85 84 83 82 81 80 79 78 77 76 75 74 73 72 71 70 69 68 67 66 65 64 ... 56
> d=a*b;
> d;
 [1]  435  860 1275 1680 2075 2460 2835 3200 3555 3900 4235 4560 4875 5180 ... 8960
```

- What are the 19^{th}, 20^{th}, 21^{st} elements of d?

Program 5 Sequential Subset

```
> d[19:21]
[1] 6555 6800 7035
```

- What are the elements of d less than 2000?

Program 6 Logical Subset

```
> d[d<2000]
[1]  435  860 1275 1680
```

- How many elements of d are greater than 6000?

Program 7 Count of Logical Subset

```
> length(d[d>6000]);
[1] 16
```

Problem 3

Summary Stats

- sum?

Program 8 Sum

```
> sum(d);
[1] 175120
```

- median?

Program 9 Median

```
> median(d);
[1] 5897.5
```

- standard deviation?

Program 10 Standard Deviation

```
> sd(d);
[1] 2608.563
```

Problem 4

Matrix Multiplication

$$\begin{bmatrix} 7 & 9 & 12 \\ 2 & 4 & 13 \end{bmatrix} \times \begin{bmatrix} 1 & 7 & 12 & 19 \\ 2 & 8 & 13 & 20 \\ 3 & 9 & 14 & 21 \end{bmatrix}$$

Program 11 Multiplying Two Matrices in R

```
> A=matrix(c(7,9,12,2,4,13),nrow=2,ncol=3,byrow=TRUE);
          [,1] [,2] [,3]
     [1,]    7    9   12
     [2,]    2    4   13
> B=matrix(c(1,7,12,19,2,8,13,20,3,9,14,21),nrow=3,ncol=4,byrow=TRUE);
          [,1] [,2] [,3] [,4]
     [1,]    1    7   12   19
     [2,]    2    8   13   20
     [3,]    3    9   14   21
> A%*%B;
          [,1] [,2] [,3] [,4]
     [1,]   61  229  369  565
     [2,]   49  163  258  391
```

Result: $\begin{bmatrix} 61 & 229 & 369 & 565 \\ 49 & 163 & 258 & 391 \end{bmatrix}$

Problem 5

Read Data From File

Program 12 Multiplying Two Matrices in R

```
## COPY FROM http://www.webpages.uidaho.edu/~stevel/519/Data/radiology.txt
> radiology = read.table('clipboard');
## OR save to C:\
> radiology = read.table("C:/radiology.txt");
```

- mean?

Program 13 Mean

```
> mean(radiology);
 Rad.Vis   P.Days   ER.Vis   Cl.Vis
3151.710 6209.258 1383.323 1135.548
```

- standard deviation?

Program 14 Standard Deviation

```
> sd(radiology);
  Rad.Vis    P.Days    ER.Vis    Cl.Vis
376.14015 656.29434 176.64114  98.84899
```

Problem 6

Histogram and Box Plots Clinic Visits (Cl.Vis) and Radiology Visits (Rad.Vis)

Program 15 Plots

```
> par(mfrow=c(1,2));
> hist(radiology$Cl.Vis,main="");
> boxplot(radiology$Cl.Vis);
> hist(radiology$Rad.Vis,main="");
> boxplot(radiology$Rad.Vis);
```

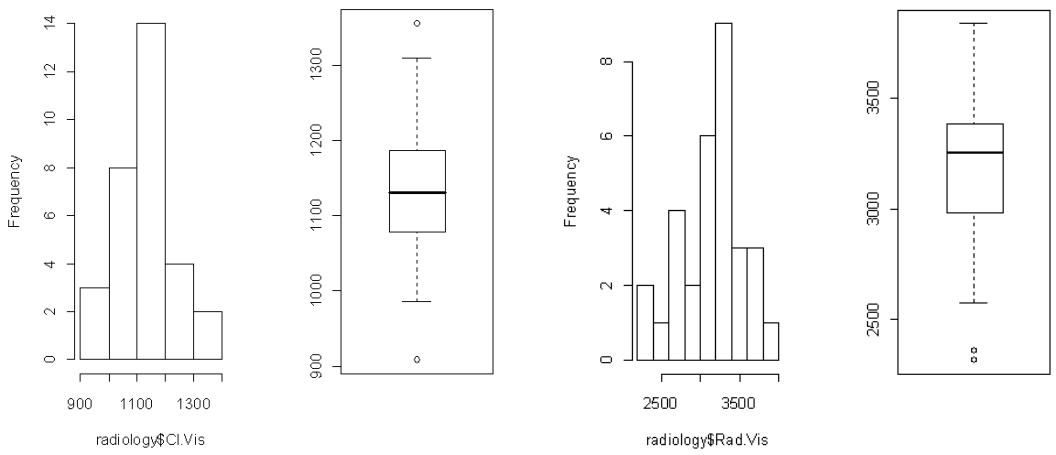

Figure 1: Clinic Visits (Cl.Vis) and Radiology Visits (Rad.Vis)

Problem 7

Scatterplot and Simple Regression

Program 16 Regression

```
> x=radiology$Cl.Vis
> y=radiology$Rad.Vis
> reg=lm(y~x);
> par(mfrow=c(1,1));
> plot(x,y);
> abline(reg);
> summary(reg);
    Call:
    lm(formula = y ~ x)
    Residuals:
         Min       1Q   Median       3Q      Max
    -469.969 -226.496    1.506  200.670  475.644
    Coefficients:
                Estimate Std. Error t value Pr(>|t|)
    (Intercept) 299.0915   604.8869   0.494    0.625
    x             2.5121     0.5307   4.733 5.32e-05 ***
    ---
    Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
    Residual standard error: 287.4 on 29 degrees of freedom
    Multiple R-squared: 0.4358,     Adjusted R-squared: 0.4164
    F-statistic:  22.4 on 1 and 29 DF,  p-value: 5.322e-05
```

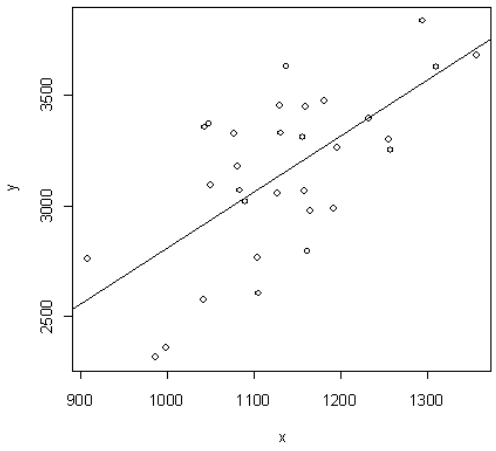

Figure 2: Scatterplot and Fitted Line

The model is significant, as is the β coefficient for x. This suggests that the number or radiology visits (y) is positively correlated with the number of clinic visits (x). The fit is "okay" with an adjusted R^2 of 0.4164 - this means x accounts for 41.6% of the variance of y.