Exercise 2

1 Answer 1: Linear regression

Let there be n linear equations with n unknowns i.e. $y_i = f(x_i)$, where β_i are coefficients of the independent variable x_i and ε_i is the noise. It is assumed that noise is distributed as iid and follows a normal Gaussian distribution i.e. $\varepsilon = N(0,\sigma^2)$

$Y_1 = \beta_0 + \beta_1 X_1 + \varepsilon_1$

$Y_2 = \beta_0 + \beta_1 X_2 + \varepsilon_2$

……………………………………………

$Y_n = \beta_0 + \beta_1 x_n + \varepsilon_n$

These linear equations can be written in matrix form as

$$\begin{pmatrix} Y1 \\ Y2 \\ .. \\ Yn \end{pmatrix} = \begin{pmatrix} \beta0 + \beta1\, X1 \\ \beta0 + \beta1\, X2 \\ .. \\ \beta0 + \beta1\, Xn \end{pmatrix} + \begin{pmatrix} \varepsilon1 \\ \varepsilon2 \\ .. \\ \varepsilon n \end{pmatrix}$$

$$= \begin{pmatrix} 1 & X1 \\ 1 & X2 \\ .. & .. \\ 1 & Xn \end{pmatrix} \cdot \begin{pmatrix} \beta0 \\ \beta1 \end{pmatrix} + \begin{pmatrix} \varepsilon1 \\ \varepsilon2 \\ .. \\ \varepsilon n \end{pmatrix}$$

This can be written as $\mathbf{Y = \beta X + \varepsilon}$ **;** note bold characters refer to matrixes

Or $\mathbf{\varepsilon = Y - X\beta}$

$\mathbf{\varepsilon^2 = \varepsilon^T \varepsilon}$

We want to find elements of $\boldsymbol{\beta}$ such that $\boldsymbol{\varepsilon}^2$ is minimized.

$\mathbf{\varepsilon^2 = \varepsilon^T \varepsilon = (Y - X\beta)^T(Y - X\beta)}$ **….. (1)**

To minimize $\boldsymbol{\varepsilon}^2$ **,** we will take its partial derivative w.r.t. $\boldsymbol{\beta}$ and set it equal to zero.

$$\frac{\partial \boldsymbol{\varepsilon}^2}{\partial \beta} = \frac{\partial \mathbf{(Y - X\beta)^T(Y - X\beta)}}{\partial \beta} \quad \text{….. (2)}$$

Let $\mathbf{S} = \mathbf{(Y - X\beta)^T(Y - X\beta)}$ **….. (3)**

$= \mathbf{(Y^T - \beta^T X^T)\,(Y - X\beta)}$

$$= Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X\beta \qquad \text{......(4)}$$

$(Y^T X\beta) = Y^T_{1xn} X_{nx2}\beta_{2x1}$ = 1x1 matrix

Also, $\beta^T X^T Y = \beta^T_{1x2} X^T_{2xn} Y_{nx1}$ = 1 x 1 matrix

For a 1x1 matrix, $A = A^T$

Therefore, $(Y^T X\beta) = (Y^T X\beta)^T = \beta^T X^T Y$

Therefore, (4) reduces to

$$S = Y^T Y - 2\,\beta^T X^T Y + \beta^T X^T X\beta \qquad \text{.......(5)}$$

Now, according to theorem of matrix calculus where x is n × 1, A is n × n, and A does not depend on x, then

i) If $\alpha = y^T Ax$, then $\frac{\partial \alpha}{\partial y} = x^T y$

ii) If $\alpha = x^T Ax$, then $\frac{\partial \alpha}{\partial x} = x^T(A^T + A)$

$\frac{\partial S}{\partial \beta} = -2\,Y^T X + (\beta^T X^T X\beta)$” ; using “ to denote partial derivative

$(\beta^T X^T X\beta)$” $= 2\beta^T X^T X$; using second point of above theorem

Putting $\frac{\partial S}{\partial \beta} = 0$,

We get,

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

2 Answer 2

The 50 items in the data set for training is split in 75%:25% ratio so that 75% (38 entries) of entries are used for training and remaining 12 entries are used for validation.

Regression model of upto 5 orders are fitted on the 38 entries of the training set and mean of squared errors is computed.

The regression model for upto 5 orders is as follows:

	Model
1st Order	0.9726429 + 0.3200254 X
2nd Order	0.8012937 + 1.0156365 X - -0.4644960 X^2
3rd Order	0.9632519 + -0.3195257 X + 1.8010990 X ^2 - -1.0146796 X^3
4th Order	1.05416420 - -1.51462702 X + 5.45151544 X^2 - 4.90249923 X^3 + 1.32960573 X^4
5th Order	0.98867379 - 0.31487545 X - 0.09292972 X^2 + 5.05491793 X^3 - 6.27506523 X^4 + 2.07199640 X^5

The mean of standard error per regression model is also computed as follows on the training and validation set

	Mean Standard Error – Training Set	Mean Standard Error – Validation Set
1st order	0.1232344	0.08690818
2nd order	0.0939629	0.05096113
3rd order	0.0733392	0.07940147
4th order	0.06825019	0.06375078
5th order	0.06596547	0.06640959

The regression model of 2nd order is chosen as that model gives the minimum Mean of Standard Error and tested on the test set. The Mean Standard Error on the test set is computed as 0.1153557.

The R code for the program is attached.

3 ANSWER 3

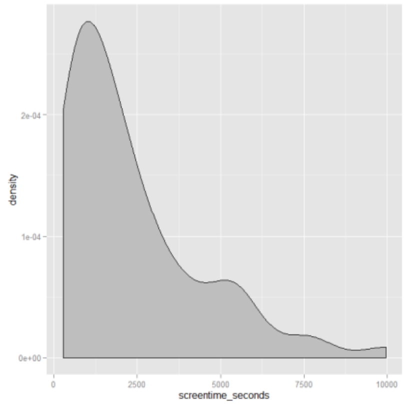


Figure 1

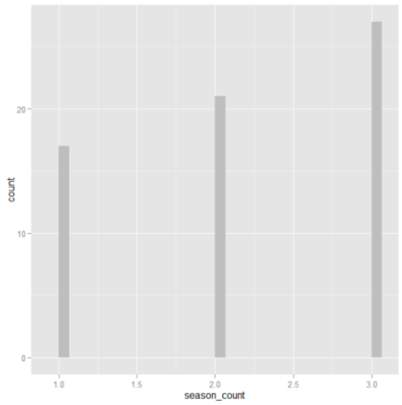


Figure 2

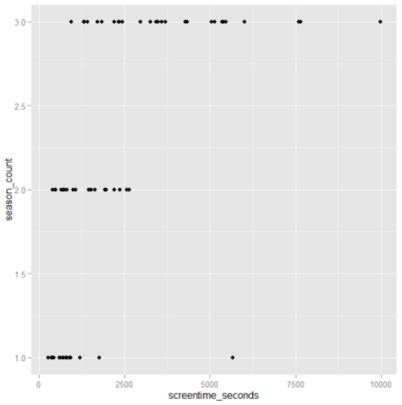


Figure 3

The summary statistics of screen_time_seconds shows

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   277     788    1640    2321    3262    9975
```

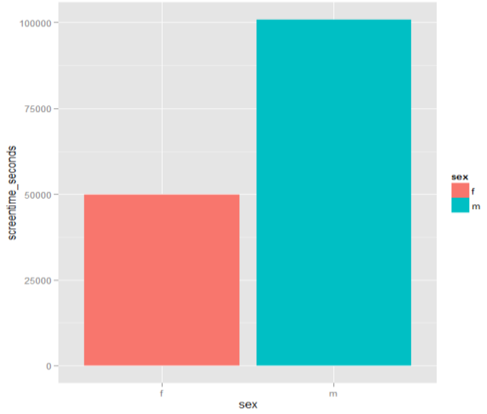
Since the mean > median , therefore the distribution of screen_time_seconds is right skewed.

Figure 1 is the density distribution of screen_time_seconds

Figure 2 shows that most of the participants have done 3 shows.

Figure 3 shows the relationship between screen_time_seconds and screen_count. The following table shows that generally people who have higher screen_count have higher screen_time.

```
  season_count screentime_seconds.mean screentime_seconds.median
1          1            1004.647              699.000
2          2            1384.524             1462.000
3          3            3877.556             3476.000
```



This figure shows that males have had more screentime than females.