DFSC 5340.02 Assignment 7

Due: Thursday December 3rd@11:59PM
Total Points: 280

1. (40 pts) For students at Walden University, the relationship between y = college GPA (with range 0–4.0) and x_1 = high school GPA (range 0–4.0) and x_2 = verbal college board score (range 200–800) satisfies $E(y) = 0.20 + 0.50x_1 + 0.002x_2$.

   (a) Find the mean college GPA for students having (i) high school GPA = 4.0 and college board score = 800, (ii) $x_1 = 3.0$ and $x_2 = 300$.

Here, $x_1 = 4$
$x_2 = 800$

We know, $E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2$
$= .20 + 2 + 1.6$
$= 3.8$

Mean is 3.8

Here, $x_1 = 3$
$x_2 = 300$

$E(y) = .20 + .50 \times 3 + .002 \times 300$
$= .20 + 1.5 + 0.6$
$= 2.3$

Mean is 2.3

(b) Show that the relationship between y and x_1 for those students with $x_2 = 500$ is $E(y) = 1.2 + 0.5x_1$.

$x_2 = 500$

$E(y) = 0.20 + .50x_1 + .002x_2$
$= .20 + .50x_1 + .002 \times 500$
$= 1.2 + 0.50x_1$

(Showed)

(c) Show that when $x_2 = 600$, $E(y) = 1.4 + 0.5x_1$. Thus, increasing x_2 by 100 shifts the line relating y to x_1 upward by $100\beta_2 = 0.2$ units.

$$E(y) = .20 + .50x_1 + .002x_2$$

$$= .20 + .50x_1 + .002 \times 600$$

$$= 1.4 + .50x_1$$

(Showed)

If we increase β_2 by one unit the x_2 will shift the line relating y to x_1 upward by β_2.

By this we can say x_2 increase by 100 shifts the line relating y to x_1 shift upward by $100\beta = .2$ units

(d) Show that setting x_1 at a variety of values yields a collection of parallel lines, each having slope 0.002, relating the mean of y to x_2.

If we set x_1 at a certain value, $.50x_1$ will have a specific value which will make $.20 + .50x_1$ as constant.

$$E(y) = .20 + .50x_1 + .002x_2$$

The equation has the intercept of $.20 + .50x_1$ and it has slope .002, relating the mean of y to x_2

2. (50 pts) The following Table shows Stata output from fitting the multiple regression model to recent statewide data, excluding D.C., on y = violent crime rate (per 100,000 people), x_1 = poverty rate (percentage with income below the poverty level), and x_2 = percentage living in urban areas.

```
. regress violent poverty urban

      Source |       SS       df       MS      Number of obs =      50
-------------+------------------------------   F(2, 47)      =  31.249
       Model |  2448368.07     2  1224184.     Prob > F      =  0.0001
    Residual |  1841257.15    47  39175.68     R-squared     =  0.5708
-------------+------------------------------   Adj R-squared =  0.5525
       Total |  4289625.22    49  87543.37     Root MSE      =  197.928

------------------------------------------------------------------------------
     violent |      Coef.   Std. Err.      t    P>|t|
-------------+----------------------------------------------------------------
     poverty |      32.622     6.677     4.885   0.0001
       urban |       9.112     1.321     6.900   0.0001
       _cons |    -498.683   140.988    -3.537   0.0009

. corr violent poverty urban
(obs=50)
             |  violent  poverty    urban
-------------+---------------------------
     violent |   1.0000
     poverty |   0.3688   1.0000
       urban |   0.5940  -0.1556   1.0000
```

(a) Report the prediction equation.

prediction equation $\hat{y} = \alpha + \beta_1 x_1 + \beta_2 x_2$

Here, $\alpha = -498.683$, $\beta_1 = 32.622$, $\beta_2 = 9.112$

$\hat{y} = -498.683 + 32.622 x_1 + 9.112 x_2$

(b) Massachusetts had y = 805, x_1 = 10.7, and x_2 = 96.2. Find its predicted violent crime rate. Find the residual and interpret.

$\hat{y} = -498.683 + 32.622 x_1 + 9.112 x_2$

$= -498.683 + 32.622(10.7) + 9.112(96.2)$

$= 727$

Residual $= y - \hat{y} = 805 - 727 = 78$

Here, Obsessed violent rate is greater than the violent rate that model predict.

(c) Interpret the fit by showing the prediction equation relating ŷ and x1 for states with (i) $x_2 = 0$, (ii) $x_2 = 100$. Interpret.

(i) $\hat{y} = -498.683 + 32.622x_1 + 9.112x_2$

$= -498.683 + 32.622x_1 + 9.112(0)$ [Here $x_2 = 0$]

$\hat{y} = -498.683 + 32.622x_1$

(ii) $\hat{y} = -498.683 + 32.622x_1 + 9112(100)$ [Here $x_2 = 100$]

$= 412.517 + 32.622x_1$

The violent crime rate is higher for the states having higher percentage living in urban areas.

(d) Interpret the correlation matrix.

The moderate weak positive association between violent crime rate and poverty rate .3688.

The moderate weak negative association between poverty rate percentage living in urban areas −0.1556

(e) Report R^2 and the multiple correlation and interpret.

$R^2 = .5708$

$R = .755$

Here $R^2 = 57.08\%$ reduction in error in predicting violent crime using poverty rate and percentage living in urban areas instead of using $\bar{y}$

$R = .755$ is the correlation predicted and obsessed violent crime rate.

3. (60 pts) For a random sample of 66 state precincts, data are available on y = percentage of adult residents who are registered to vote, x_1 =percentage of adult residents owning homes, x_2 = percentage of adult residents who are nonwhite, x_3 = median family income (thousands of dol-lars), x_4 =median age of residents, x_5 =percentage of res-idents who have lived in the precinct for at least 10 years. The following table shows some output used to analyze the data.

	Sum of Squares	DF	Mean Square	F	Sig
Regression	813.3	5	162.66	3.3	.009
Residual	2940.0	60	49		
Total	3753.3	65			

R-Square ---- .217

Root MSE ---- 7

Variable	Parameter Estimate	Standard Error	t	Sig
Intercept	70.0000		2.2222	.030
x1	0.1000	0.0450	2.0	.050
x2	-0.1500	0.0750	-.5	.6189
x3	0.1000	0.2000	.8	.4268
x4	-0.0400	0.0500	-2.4	.0195
x5	0.1200	0.0500		

(a) (30 pts) Fill in all the missing values.

R Filled upward

R square calculation = $1 - \left(\frac{TSS - SSR}{TSS}\right)$

$= 1 - .783 = 0.217$

Root MSE = $\sqrt{\text{mean square}}$

$= \sqrt{49} = 7$

(b) (10 pts) Do you think it is necessary to include all five explanatory variables in the model? Explain.

It is not necessary to include all five explanatory variables in model. We can ignore x_3 or x_4 or both of them. Their p-values are large for their partial tests.

(c) (10 pts) To what test does "F value" refer? Interpret the result of that test.

Lets assume the level of significance .05
The x_1 variable, the p value is .003 which is less than level of significance. So can reject hypothesis.
The evidence is strong for overall regression is significant.

(d) (10 pts) To what test does the t-value opposite x_1 refer? Interpret the result of that test.

Ho : $B_1 = 0$
Ho : $B_1 \neq 0$
significance level = .05
for X_1, the p value is .03
We can reject null Hypothesis as the p-value is smaller than .05.
X_1 has an effect on y which control other xs.

4. (130 pts) The following Table shows results of regressing y = birth rate (number of births per 1000 population) on x_1 = women's economic activity and x_2 = literacy rate, using UN data for 23 nations.

	Mean	Std Deviation	N
BIRTHS	22.117	10.469	23
ECON	47.826	19.872	23
LITERACY	77.696	17.665	23

Correlations

		BIRTHS	ECON	LITER
Correlation	BIRTHS	1.00000	-0.61181	-0.81872
	ECON	-0.61181	1.00000	0.42056
	LITERACY	-0.81872	0.42056	1.00000

	Sum of Squares	DF	Mean Square	F	Sig
Regression	1825.969	2	912.985	31.191	0.0001
Residual	585.424	20	29.271		
Total	2411.393	22			

Root MSE (Std. Error of the Estimate) 5.410 R Square 0.7572

	Unstandardized Coeff. B	Std. Error	Standardized Coeff. (Beta)	t	Sig
(Constant)	61.713	5.2453		11.765	0.0001
ECON	-0.171	0.0640	-0.325	-2.676	0.0145
LITERACY	-0.404	0.0720	-0.682	-5.616	0.0001

(a) (70 pts) Report the value of each of the following: (i) r_{yx1}, (ii) r_{yx2}, (iii) R^2, (iv) TSS, (v) SSE, (vi) mean square error, (vii) s, (viii) s_y, (ix) *se* for b_1, (x) t for H_0: $\beta_1 = 0$, (xi) P for H_0: $\beta_1 = 0$ against H_a: $\beta_1 = 0$, (xii) P for H_0: $\beta_1 = 0$ against H_a: $\beta_1 < 0$, (xiii) F for H_0: $\beta_1 = \beta_2 = 0$, (xiv) P for H_0: $\beta_1 = \beta_2 = 0$.

(1) $r_{yx1} = .61181$
(2) $r_{yx_2} = -0.81872$
(3) $R^2 = .7572$
(4) TSS = 2411.39
(5) SSE = 585.424
(6) mean square error = 29.271
(7) $s = 5.41$
(8) $s_y = 10.469$
(9) se for $b_1 = .0640$
(10) t for $H_0 = -2.676$
(11) P for $H_0 = .0145$
(12) P for $H_0 = .0145$
(13) .0001
(14) .007

(b) (10 pts) Report the prediction equation and interpret the signs of the estimated regression coefficients.

$\hat{y} = 61.713 - 0.171x_1 - .404x_2$

Controlling the women's economic activity as one unit increase is literacy rate, the predicted birth rate decrease by .404 units.

(c) (10 pts) Interpret the correlations r_{yx1} and r_{yx2}

The correlation between birth rate and women is economic activity -.61181. Moderate negative association with rate and women's economic activity

The correlation between birth rate and literacy -.81872. There is strong negative associated between birth rate and literacy rate.

(d) (10 pts) Report R^2 and interpret its value.

$R^2 = .7572$

There is 75.72% reduction in error is using the regression equation to predict birth rate instead of using $\bar{y}$

(e) (10 pts) Find the correlation and interpret its value.

$R = \sqrt{R^2} = \sqrt{.7572}$

$= .8701$

The correlation is obsessed y and predicted y is .8701

(b) (10 pts) Test the null hypothesis that mean number of children is independent of mother's educational level, and report and interpret the P-value.

Ho : $B_1 = B_2 = 0$ vs 31.191

P value is .0001

The literacy rate and womens economic activity has significant effect as P value is less than .05. So we can reject the null hypothesis.

(c) (10 pts) Sketch a potential scatterplot such that the analyses you conducted in (a) and (b) would be inappropriate.

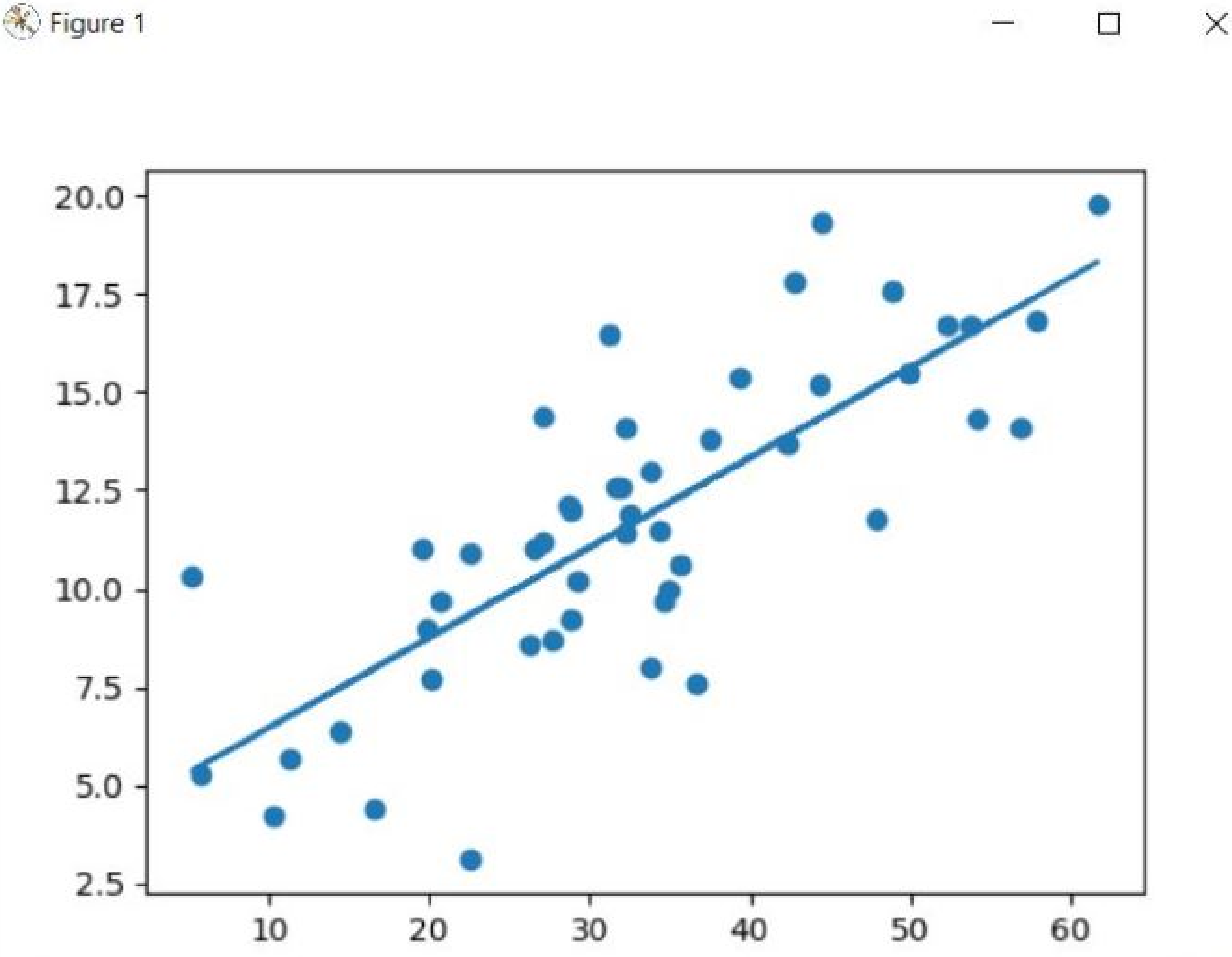
Figure 1
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
10
20
30
40
50
60