STAT 2607 – Assignment 3
SOLUTIONS

(Q.1) a) The MINITAB output is shown below. The fitted model is

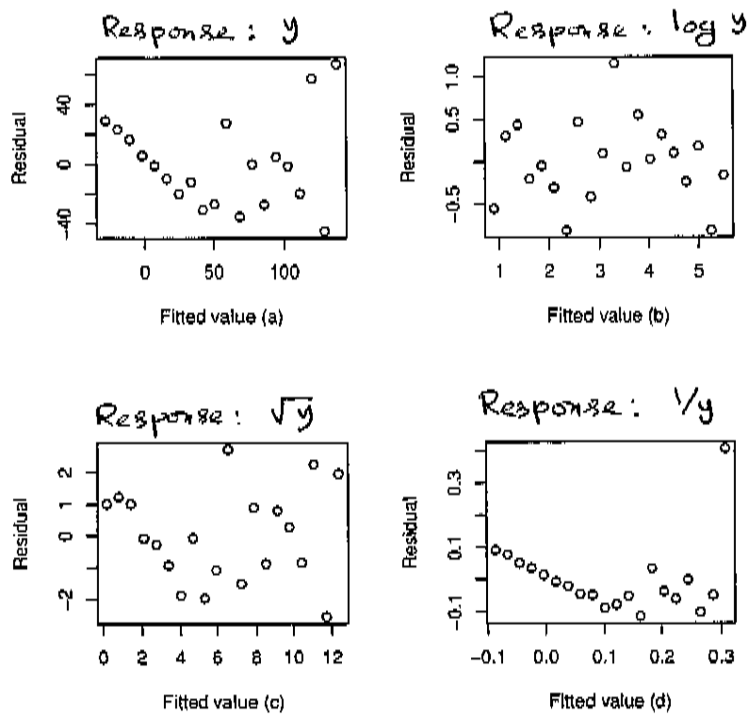
$$\hat{y} = -36.4 + 8.70x$$

```
Regression Analysis: y versus x

The regression equation is
y = - 36.4 + 8.70 x

Predictor    Coef  SE Coef      T      P
Constant   -36.42    14.26  -2.55  0.020
x           8.704    1.191   7.31  0.000

S = 30.7074   R-Sq = 74.8%   R-Sq(adj) = 73.4%
```

b) The scatter plots of residuals against the fitted values are shown on page 2. The plot for the model $y = \beta_0 + \beta_1 x + \epsilon$ clearly indicates a systematic pattern: the variability in residuals increases with increased fitted values. This indicates the violation of the assumption of common variance for the error term ϵ.

Here $R^2 = 0.748$. This means that 74.8% of the variation in y can be explained by the regression relationship with x.

Residual plots



Regression Analysis: LN.y versus x

The regression equation is
LN.y = 0.650 + 0.242 x

Predictor	Coef	SE Coef	T	P
Constant	0.6500	0.2276	2.86	0.011
x	0.24159	0.01900	12.71	0.000

S = 0.490012   R-Sq = 90.0%   R-Sq(adj) = 89.4%

Regression Analysis: SQRT.y versus x

The regression equation is
SQRT.y = - 0.463 + 0.641 x

Predictor	Coef	SE Coef	T	P
Constant	-0.4634	0.7035	-0.66	0.518
x	0.64082	0.05872	10.91	0.000

S = 1.51432   R-Sq = 86.9%   R-Sq(adj) = 86.1%

Regression Analysis: INV.y versus x

The regression equation is
INV.y = 0.329 - 0.0208 x

Predictor	Coef	SE Coef	T	P
Constant	0.32868	0.05338	6.16	0.000
x	-0.020761	0.004456	-4.66	0.000

S = 0.114913   R-Sq = 54.7%   R-Sq(adj) = 52.1%

(Q1) c) i) Model: $\log(y) = \beta_0 + \beta_1 x + \epsilon$

Residual plot (shown on page 2) indicates no violation of model assumptions.

Here $R^2 = 0.90$. That is, 90% of the variation in $\log y$ is explained by the regression relationship with x.

The fitted model is

$$\widehat{\log(y)} = 0.650 + 0.242\,x$$

ii) Model: $\sqrt{y} = \beta_0 + \beta_1 x + \epsilon$

The fitted model is

$$\widehat{\sqrt{y}} = -0.463 + 0.641\,x$$

Here the residual plot does not indicate any clear violation of model assumptions.

$R^2 = 0.869$. That is, 86.9% of the variation in $\sqrt{y}$ is explained by the regression relationship with x.

(Q.1) c) iii) <u>Model</u> : $(1/y) = \beta_0 + \beta_1 x + \epsilon$

The fitted model is

$$\widehat{(1/y)} = 0.329 - 0.0208\,x$$

The residual plot shows a clear pattern: the variability in residuals increases with the increased values of fitted response. This violates the assumption of <u>common variance</u> for ϵ. Also, the plot shows an extreme outlier.

Here $R^2 = 0.547$. This means that only 54.7% of the variability in $(1/y)$ can be explained by the regression relationship with x.

d) The best model is

$$\widehat{\log(y)} = 0.650 + 0.242\,x$$

This model provides the largest R-squared value ($R^2 = 0.90$). Also the residual plot for this model does not indicate any violation of model assumptions.

Q.2 a) $n = 12$ observations

b) Total SS $= 5182.189$

$DF = n-1 = 11$

c) $MSE = \dfrac{SSE}{DF}$

$= \dfrac{SSE}{8} = \dfrac{3459.6803}{8} = 432.46$

d) Explained variation

= Regression SS

$= 535.9569 + 1167.5634 + 18.9886$

$= 1722.509$

e) Proportion of variation explained by the model

$= R^2$

$= \dfrac{\text{Regression SS}}{\text{Total SS}}$

$= \dfrac{1722.509}{5182.189}$

$= 0.3324$

Q.2 f) Here we wish to test

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0$$

(overall regression is not useful)

vs H_1: at least one β_j $(j = 1, 2, 3)$ is non-zero

(overall regression is useful)

Test statistic (ANOVA F test) is

$$F = \frac{MSR}{MSE} = \frac{SSR/3}{SSE/8}$$

$$= \frac{1722.509/3}{432.46} = \frac{574.1697}{432.46}$$

$$= 1.33$$

At $\alpha = .05$, rejection point is $F_{.05,3,8} = 4.07$.

So we fail to reject H_0, and conclude that the overall regression is not useful for prediction.

Q.3

a) The MINITAB output is shown below.

```
Regression Analysis: Sales versus Advertising, Price, Stores

The regression equation is
Sales = 31.0 + 0.820 Advertising - 0.325 Price + 1.84 Stores

Predictor        Coef   SE Coef      T      P
Constant       30.992     7.728   4.01  0.007
Advertising    0.8202    0.5023   1.63  0.154
Price        -0.32502   0.08935  -3.64  0.011
Stores          1.841     3.855   0.48  0.650

S = 5.46455   R-Sq = 96.7%   R-Sq(adj) = 95.0%

Analysis of Variance

Source          DF      SS      MS      F      P
Regression       3  5179.2  1726.4  57.81  0.000
Residual Error   6   179.2    29.9
Total            9  5358.4

Source       DF  Seq SS
Advertising   1  4692.8
Price         1   479.6
Stores        1     6.8
```

(Q.3) b) i) The regression coefficient for "advertising" is $\hat{\beta}_1 = 0.820$. So for each additional hundred dollars of advertising, the light fixture sales are estimated to <u>increase</u> by an average of 820 units when other covariates remain fixed

ii) The regression coefficient for "price" is $\hat{\beta}_2 = -0.325$. That is, for each additional dollar of increase in the price of the fixture, the number of fixtures sold is estimated to <u>decrease</u> by an average of 325 units when other two covariates remain fixed.

iii) The regression coefficient for "stores" is $\hat{\beta}_3 = 1.84$. So for each additional retail store used to sell the light fixture, sales are estimated to <u>increase</u> by an average of 1840 units when other two covariates are held constant.

Q.3 c) The estimated monthly light fixture sale is

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$$

$$= 31.0 + 0.820\,(40) - 0.325\,(60) + 1.84\,(3)$$

$$= 49.82$$

The corresponding residual is

$$y - \hat{y} = 42 - 49.82 = -7.82$$

d) 95% confidence interval for β_1 is

$$\{\hat{\beta}_1 \pm t_{\alpha/2,\, n-p}\ s.e.(\hat{\beta}_1)\}$$

or $\{0.8202 \pm 2.447\,(0.5023)\}$ or $\{0.8202 \pm 1.2291\}$

or $(-0.4089,\ 2.0493)$

We are 95% confident that if advertising expenditure increases by \$100, while the price of the product and the number of stores selling the product remain constant, then the average monthly unit sales of this fixture will decrease by at most 408 units and will increase by at most 2049 units.

Q.3 e)

We wish to test

$H_0: \beta_2 = 0$ ("price" is not useful for regression)

vs $H_1: \beta_2 \neq 0$ (price is useful)

Test statistic,

$$t = \frac{\hat{\beta}_2}{s.e.(\hat{\beta}_2)} = \frac{-0.32502}{0.08935} = -3.64$$

p-value = 0.011

So at $\alpha = .05$, we reject H_0 and conclude that the independent variable "price" is useful for regression.

f) Here $H_0: \beta_1 = \beta_2 = \beta_3 = 0$, H_1: at least one β_j is non-zero.

$$F = \frac{MSR}{MSE} = \frac{1726.4}{29.9} = 57.81$$

The p-value is close to 0. So there is a strong evidence to conclude that the overall regression is useful for prediction purposes.

STAT 2607 — Assignment 3

Marks distribution

Total marks : 50

[14]

Q.1
- a) 1
- b) 2
- c) 3 + 3 + 3
- d) 2

[14]

Q.2
- a) 1
- b) 2
- c) 2
- d) 2
- e) 2
- f) 5

[22]

Q.3
- a) 4
- b) 3
- c) 3
- d) 4
- e) 4
- f) 4