Midterm Test for ST3131

(please answer all the questions for full marks)

(All the notations are defined according to the lecture notes)

1. Consider the simple linear regression model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ with all the assumptions (L-I-N-E). Based on 12 observations with

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = 5.2193, \quad \bar{X} = -0.4356,$$

someone obtained the following estimated model

$\hat{Y}$	=	0.75	-	2.22 X,
(SE)		(0.185)		(0.234)

$$R^2 = 89.96\%, \quad \hat{\sigma} = 0.5347, \quad F - Statistic = 89.58.$$

(a) If the previous experience claimed that the intercept is zero, test the claim with significance level $\alpha = 0.05$.

(b) If the previous experience suggested that when X increases by 1 unit, the expected response Y will increase by 2 unit. Give a statistical test for the suggestion with $\alpha = 0.05$.

(c) Suppose two individuals have 2 units difference in X. What is the expected difference in their response? give a 95% CI for the difference.

(d) Suppose a new individual has $X = 1.5$, find the 95% confidence prediction interval for the possible response from this individual.

(a) $H_0 : \beta_0 = 0$,

$$|t^*| = |\frac{0.75 - 0}{0.185}| = 4.054 > t(0.975, 10) = 2.228$$

We reject H_0.

(b) $H_0 : \beta_1 = -2$,

$$|t^*| = |\frac{-2.22 + 2}{0.234}| = 0.94 < t(0.975, 10) = 2.228$$

We accept H_0.

(c) we need to find the CI for $2\beta_1$, which is

$$2(b_1 \pm t(0.975, 10) * s(b_1)) = 2 * (-2.22 \pm 2.228 * 0.234) = [-5.4827, -3.3973]$$

(d) $\hat{Y} = 0.75 - 2.22 * 1.5 = -2.5800$ and

$$s(pred) = \hat{\sigma}\{1 + 1/12 + (-0.4356 - 1.5)^2/5.2193\}^{1/2} = 0.7176$$

Thus the PI is

$$-2.5800 \pm 0.7176 * 2.228 = [-4.1788, -0.9812]$$

2. Someone fits a regression model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ and get the following ANOVA table

Response: Y

	Df	Sum Sq	Mean Sq	F-value
X	1	18.4873	18.4873	11.355
Residuals	13	21.1660	1.6282	

(a) Find the value for the least estimation squares

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} \{Y_i - \beta_0 - \beta_1 X_i\}^2$$

(b) Test $H_0 : \beta_1 = 0$ with $\alpha = 0.05$ under the assumptions (L,I, N, E).

(a)

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} \{Y_i - \beta_0 - \beta_1 X_i\}^2 = SSE = 21.1660$$

(b) $H_0 : \beta_1 = 0$

$$F = 11.355 > F(0.95, 1, 13) = 4.67$$

Thus, we reject H_0.

3. For data $(X_i, Y_i), i = 1, ..., 10$ and its model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, some parts, marked by ???, of the calculation are missing.

Coefficients:

	Estimate	Std. Error	t value	p-value
(Intercept)	0.4409	0.1352	**3.2611**	0.0115
x	-1.0344	**0.1161**	-8.913	1.99e-05

Residual standard error: 0.4258 on 8 degrees of freedom
Multiple R-squared: **0.9085**
F-statistic: **79.4416** on 1 and 8 DF, p-value: 1.99e-05

(a) Fill in the missed values.

(b) Set up the ANOVA table.

(a) See the bold values

(b)

$$SSE = MSE * 8 = 0.4258^2 * 8 = 1.4504 \quad \text{with d.f } 8$$

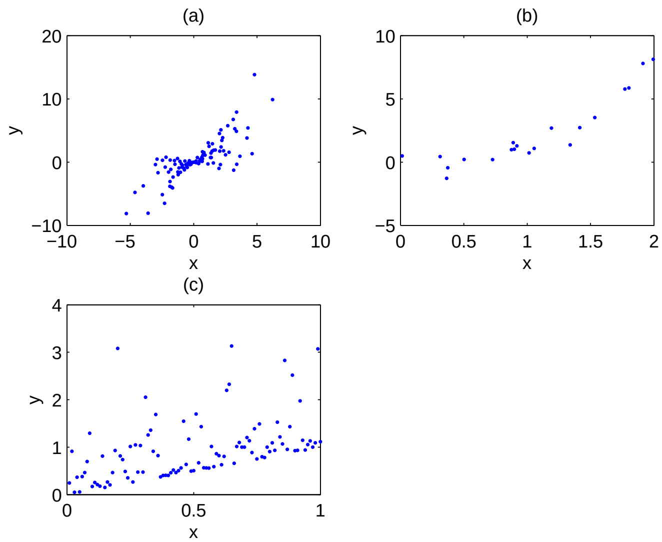
and

$$(t^*)^2 = 79.4416 = F^* = \frac{SSR/1}{MSE}$$

Thus

$$SSR = 79.4416 * 0.4258^2 = 14.4032 \quad \text{with d.f } 1$$

$$SST = 14.4032 + 1.4504 = 15.8536 \quad \text{with d.f } 9$$



4. Three sets of data are plotted below and marked by (a), (b) and (c) respectively, do you think which assumptions in the simple linear regression model are violated for each of the data?

   (a): "equal variance" is violated

   (b): "linearity assumption" is violated

   (c): "normal random error" is violated.

5. For model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,\ i = 1, ..., n$, we have $\sum_{i=1}^{n} Y_i = 0,\ SSR = 15,\ SSE = 5$. Let e_i be the fitted residuals of the least squares estimation. Find $\sum_{i=1}^{n}(Y_i + 2e_i)^2$.

$$
\begin{aligned}
\sum_{i=1}^{n}(Y_i + 2e_i)^2 &= \sum_{i=1}^{n}(\hat{Y}_i + e_i + 2e_i)^2 \\
&= \sum_{i=1}^{n}(\hat{Y}_i + 3e_i)^2 \\
&= \sum_{i=1}^{n}\hat{Y}_i^2 + 6\sum_{i=1}^{n}\hat{Y}_i e_i + 9\sum_{i=1}^{n} e_i^2 \\
&= \sum_{i=1}^{n}\hat{Y}_i^2 + 9\sum_{i=1}^{n} e_i^2 \\
&= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + 9\sum_{i=1}^{n} e_i^2 \\
&= SSR + 9 * SSE = 60
\end{aligned}
$$