Math312 **Spring 2020**

HW07

1. For simple Linear Regression model ($y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ \forall\ 1 \le i \le n, \epsilon_i \overset{\text{uncorr.}}{\sim} (0, \sigma^2)$.

$$
\begin{aligned}
S_{xx} &= \sum_i (x_i - \bar{x})^2, \\
S_{yy} &= \sum_i (y_i - \bar{y})^2, \\
SS_{\mathrm{R}} &= \sum_i (\hat{y}_i - \bar{y})^2 = \hat{\beta}_1^2 S_{xx},\ MS_{\mathrm{R}} = \frac{SS_{\mathrm{R}}}{1} \\
SS_{\mathrm{Res}} &= \sum_i (y_i - \hat{y}_i)^2,\ MS_{\mathrm{Res}} = \frac{SS_{\mathrm{Res}}}{n-2} = \hat{\sigma}^2 \\
SS_{\mathrm{T}} = S_{yy} &= SS_{\mathrm{R}} + SS_{\mathrm{Res}} \\
\mathrm{R}^2 &= \frac{SS_{\mathrm{R}}}{SS_{\mathrm{T}}} \\
s.e.(\hat{\beta}_0) &= \sqrt{MS_{\mathrm{Res}} \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)} \\
s.e.(\hat{\beta}_1) &= \sqrt{\frac{MS_{\mathrm{Res}}}{S_{xx}}} \\
t^*(\mathrm{H}_0 : \beta_0 = 0) &= \frac{\hat{\beta}_0}{s.e.(\hat{\beta}_0)} \\
t^*(\mathrm{H}_0 : \beta_1 = 0) &= \frac{\hat{\beta}_1}{s.e.(\hat{\beta}_1)}
\end{aligned}
$$

The following shows a result from regressing a response variable (y) on a predictor (x) based on 20 pairs of observations, fill in the missing entries (1)-(11):

ANOVA table				
Source	S.S.	d.f.	M.S.	F-test
Regression	1848.76	(1)	(2)	(3)
Error	(4)	(5)	(6)	

Parameter Estimates				
Variable	Coefficient	s.e.	t-test	p-value
Intercept	-23.4325	12.74	(7)	0.0824
x	(8)	0.1528	8.32	< 0.0001

n	R^2	$\hat{\sigma}$
(9)	(10)	(11)

2. (SLR with SAS) One may wonder if people of similar heights tend to marry each other. For this purpose, a sample of newly married couples was selected. Data can be found in "HW07.sas"on coursesite. Use the "SLR.sas"as the template, prepare annotated report that answers the following questions (you will use ODS command to output the result and MS/WORD to edit your report)

   a. Generate a scatter plot for the heights of husband (x-axis) versus the heights of wife (y-axis). Comment on the possible association.

   b. Perform correlation analysis based on both Pearson and Spearman rank correlation coefficient, i.e. the numerical value, confidence interval and hypothesis testing.

   c. (Assumed significantly correlated) Perform Simple Linear Regression (SLR) analysis for regressing the heights wives on the heights of husbands.

      (1) Test on the regression effect at 5% level of significance.

      (2) Report the value for the coefficient of determination.

      (3) Perform model diagnostics and report any serious violations.

      (4) If no serious violation, report the fitted SLR line and interpret the meaning of the slope estimate.

Note that prepare your turn-in in a single PDF file, i.e. combine and convert the scan of your write-up for problem #1 and report for problem #2 together as a PDF file for submission.