STAT 512 HW #5 Solutions

1.

a. $$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{11}X_{12} \\ 1 & X_{21} & X_{21}X_{22} \\ 1 & X_{31} & X_{31}X_{32} \\ 1 & X_{41} & X_{41}X_{42} \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

b. $$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{12} \\ 1 & X_{21} & X_{22} \\ 1 & X_{31} & X_{32} \\ 1 & X_{41} & X_{42} \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

2.

The output is shown below. The estimated regression equation is $Y = 37.65 + 4.425X_1 + 4.375X_2$. The $R^2 = .9521$, and the P-value of the F-test is $< .0001$. This suggests that the combination of moisture and sweetness is helpful in predicting the brandness liking. This P-value is associated with the null hypothesis that $\beta_1 = \beta_2 = 0$ versus the alternative that at least one of them is not equal to zero. The test statistic is 129.08 and the degrees of freedom are 2 and 13.

3.

The 95% CIs are shown below. Because 0 is not included in either confidence interval, this implies that we would reject the hypothesis that $\beta_i = 0$ for the alternative $\beta_i \neq 0$. This can also be seen from the P-values. The test statistics are 14.70 and 6.50, the degrees of freedom are 13. The conclusion would be that either variable **after fitting/adjusting for the other variable** significantly helps predict brandness liking.

Dependent Variable: liking

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1872.70000	936.35000	129.08	<.0001
Error	13	94.30000	7.25385		
Corrected Total	15	1967.00000			

Root MSE	2.69330	R-Square	0.9521
Dependent Mean	81.75000	Adj R-Sq	0.9447
Coeff Var	3.29455		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	37.65000	2.99610	12.57	<.0001
moisture	1	4.42500	0.30112	14.70	<.0001
sweetness	1	4.37500	0.67332	6.50	<.0001

Parameter Estimates

Variable	DF	95% Confidence Limits	
Intercept	1	31.17731	44.12269
moisture	1	3.77447	5.07553
sweetness	1	2.92037	5.82963

4.

a Compute the pairwise correlations between the X's and between each X and Y. Which X variable appears to be the best individual predictor?

There do not appear to be any really strong correlations among the explanatory variables (all < 0.70). Age ($x1$) appears to be the best individual predictor since it has the largest correlation ($r = -0.79$).

Pearson Correlation Coefficients, N = 46
Prob > |r| under H0: Rho=0

	x1	x2	x3	y
x1	1.00000	0.56795	0.56968	-0.78676
		<.0001	<.0001	<.0001
x2	0.56795	1.00000	0.67053	-0.60294
	<.0001		<.0001	<.0001
x3	0.56968	0.67053	1.00000	-0.64459
	<.0001	<.0001		<.0001
y	-0.78676	-0.60294	-0.64459	1.00000
	<.0001	<.0001	<.0001	

b Run the linear regression with age, severity of illness and anxiety level as the explanatory variables and satisfaction as the response variable. Summarize the regression results.

The fitted regression line is $\widehat{Y} = 158.49 - 1.14X_1 - 0.44X_2 - 13.47X_3$ where X_1 is age, X_2 is severity level, and X_3 is anxiety level. The coefficient of determination is $R^2 = 0.6822$ and the P-value of the F test is $< .0001$ suggesting that this set of explanatory variables is helpful in explaining satisfaction. The test is testing the hypotheses

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$$
$$H_a : \text{at least one } \beta_i \neq 0$$

The test statistic is 30.05 and the degrees of freedom are 3 and 42.

Only one individual parameter t-tests is significant (Age) suggesting that perhaps not all variables are needed in predicting satisfaction. All of the coefficients are negative suggests an increase in any one of them results in an decrease in satisfaction.

c Plot the residuals versus the predicted satisfaction and each of the explanatory variables. Are there any unusual patterns?

I do not see any unusual patterns in any of the residual plots.

Residual
Predicted Value of y

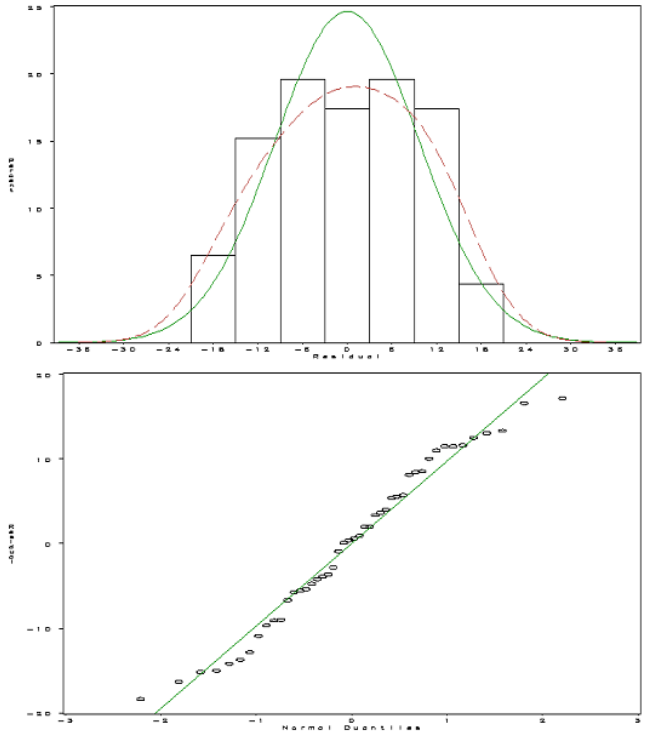
Residual
x1

Residual
x2

Residual
x3

d Examine the assumption of normality for the residuals using a qqplot or histogram. State your conclusions.

The data appears to be reasonably normal. No real deviations can be found in these plots.



e Predict the satisfaction for a 55 year old patient with illness severity 50 and anxiety level 2.8. Provide a 95% prediction interval with your prediction.

The following is from SAS using the CLI option.

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Predict	
47	.	35.8859	4.6332	13.5381	58.2338