Statistics Qualifying Exam

August 2014

1. Let $X_1, X_2, \ldots\ldots, X_n$ be a random sample from the Uniform distribution in (0, θ), where $\theta > 0$ is an unknown parameter.

(a) Find the most powerful test for testing $H_0: \theta = 1$ against $H_1: \theta = 2$, at a given level of significance α . Show that the test is based on a simple statistic Y.

(b) Find explicitly the critical region when $\alpha = 0.05$.

2. Let $X_1, X_2, \ldots\ldots, X_n$ be a random sample from the normal distribution $N(\theta, 1$, where θ is unknown.

(**a**) Show that there are infinitely many confidence intervals for θ at a fixed level of confidence of 100 $(1-\alpha)\%$. Show how to find them explicitly.

(**b**) Show that among all of those in (a), the confidence interval of shortest length is the (classical) confidence interval symmetric about X.

3. Let $X_1, X_2, \ldots\ldots, X_n$ be a random sample from the Uniform distribution in $(0, \theta)$. Let $P_n(\theta)$ be a polynomial of degree n in θ.

(**a**) Find the UMVUE (uniformly minimum variance unbiased estimator) of $P_n(\theta)$.

(**b**) Find the UMVUE of $g(\theta \;= \cos\theta$.

4. Let X and Y be independent random variables, where X is Uniform in $(0,1)$, and Y is Uniform in $(-0.5, +0.5)$.

(a) Find the joint p.d.f. of X and Y .

(b) Calculate $P\,(X > Y)$.

5. The normal error regression model is specified as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

where β_0 is the intercept, β_1 is the slope, and $\epsilon_i's$ are identically and independently distributed as $N(0, \sigma^2$, $i = 1, \ldots, n$.

(a) Derive the maximum likelihood estimator for $\beta_0, \beta_1,$ and σ^2.

(b) Derive the best linear unbiased estimator for β_0 and β_1.

6. A multiple linear regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$ with $\epsilon \sim N(0, \sigma^2\mathbf{I}$ is fitted to a dataset ($\mathbf{I}$ is an identity matrix).

(a) The figures below show results of residual analysis for the model (resid: residuals). Interpret Figure 1 and Figure 2 and discuss the model fit based on these figures.

figure 1

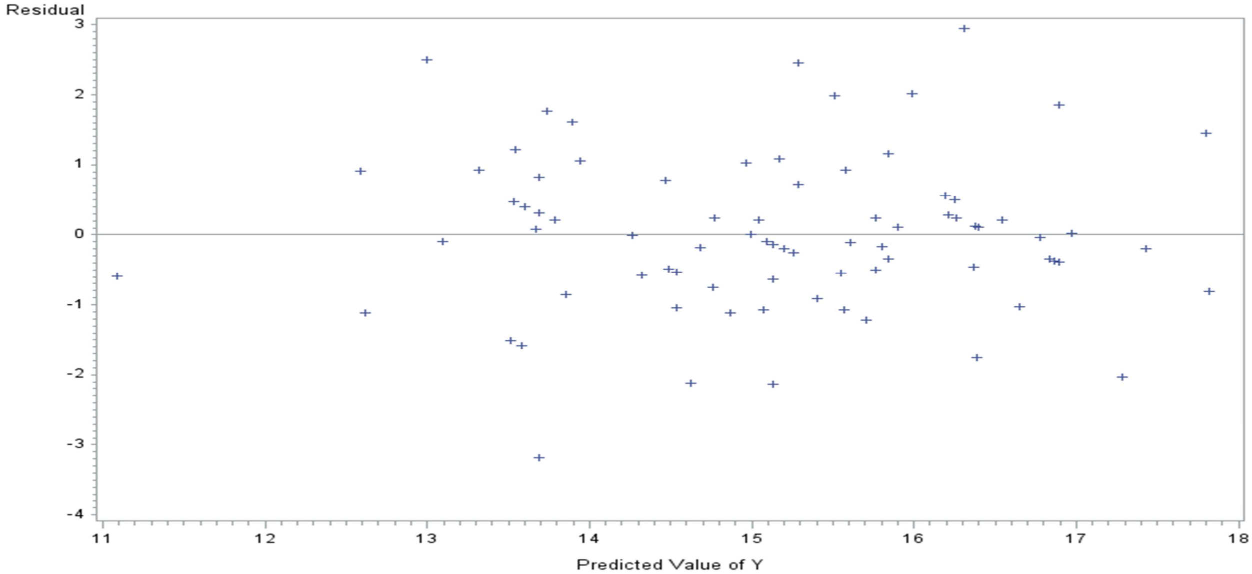
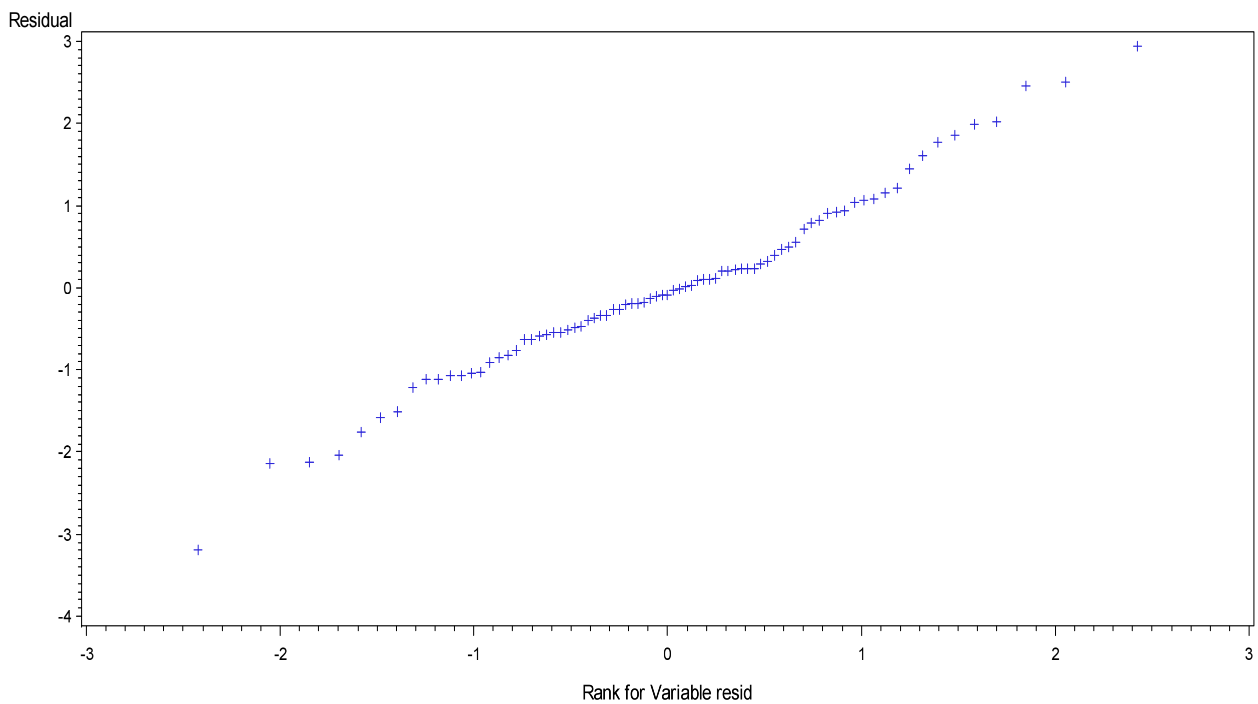


figure 2



Refer to the SAS output from PROC REG in **Table 1** and answer the following questions.

(b) Fill in the numbered blanks **(1)** and **(2).**

(c) For the full model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$, test the following hypotheses use a given significance level α. **Clearly** specify the test statistic, the sampling distribution under the null hypothesis and the decision rule.

$$H_0: \beta_2 = \beta_3 = \beta_4 = 0 \text{ versus } H_1: \text{not all } \beta_2, \beta_3, \text{and } \beta_4 \text{equals } 0.$$

Table 1 SAS OUTPUT FOR PROBLEM 6

The REG Procedure

Model: MODEL1

Dependent Variable: Y

Number of Observations Read	81
Number of Observations Used	81

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	138.32691	34.58173	26.76	<.0001
Error	76	98.23059	1.29251		
Corrected Total	80	236.55750			

Root MSE	1.13689	R-Square	0.5847
Dependent Mean	15.13889	Adj R-Sq	0.5629
Coeff Var	7.50970		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Type I SS	Type II SS
Intercept	1	12.20059	0.57796	21.11	<.0001	18564	575.97646
X1	1	-0.14203	0.02134	-6.65	<.0001	(1)	(2)
X2	1	0.28202	0.06317	4.46	<.0001	72.80201	25.75896
X3	1	0.61934	1.08681	0.57	0.5704	8.38142	0.41975
X4	1	0.00000792	0.00000138	5.72	<.0001	42.32496	42.32496

7. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a $N(\mu_1, \sigma^2)$ population and $Y_1, Y_2, \ldots, Y$ denote a random sample from a $N(\mu_2, \sigma^2$ population, σ^2 is unknown. For testing

$$H_0: \mu_1 = \mu_2 \quad versus \quad H_1: \mu_1 \neq \mu_2 \ ,$$

the same conclusions should be obtained by the two-sample t-test and the fixed effect model. Clearly construct test procedures by these two methods and show their equivalence.

8. An engineer is interested in the effects of cutting speed (A), tool geometry (B), and cutting angle (C) on the life (in hours) of a machine tool. Two levels of each factor are chosen. Suppose that a full replicate of the experiment cannot all be run using the same bar stock.

(a) Set up an experimental design to run the treatment combinations in two blocks of four treatment combinations each, with ABC confounded.

(b) Assume the data obtained as below. Analyze the data.

Treatment Combination	(1)	a	b	ab	c	ac	bc	abc
	22	32	35	55	44	40	60	39