Statistics Qualifying Exam

August 19, 2013

1. Consider an experiment in which a person chooses at random a point (X, Y) from the unit square S = { (x,y) : 0< x< 1, 0<y<1}. Assume that the distribution of probability over the unit square is uniform, i.e., $f_{X,Y}(x,y) = 1$, 0<x<1, 0<y<1; =0, elsewhere. Let U=X+Y and V=X-Y.

(i) Find the joint probability distribution of (U,V).

(ii) Find the marginal distribution of U.

2. If the correlation coefficient ρ of two random variables, X and Y, show that $-1 \leq \rho \leq 1$.

3. Let $Y_1 < Y_2 < Y_3 < Y_4$ be the order statistics of a random sample of size n=4 from a distribution with its pdf f(x) = 2x, 0<x<1, zero elsewhere.

(i) Find the joint pdf of Y_3 and Y_4.

(ii) Find the conditional pdf of Y_3, given $Y_4 = y_4$.

(iii) Evaluate $E(Y_3|y_4)$.

4. Let $X_1, \ldots, X_n$ be a random sample from the Poisson distribution with mean $\theta > 0$.

(i) Find a sufficient statistics (SS) for θ.

(ii) Find the (uniformly) minimum variance unbiased estimator (MVUE) of $a\theta^2 + b\theta + c$, where a, b, c are given constants.

(iii) Does there exist an unbiased estimator of $1/\theta$? Please JUSTIFY your answer!

5. Two microprocessors are compared on a sample of six benchmark codes to determine whether there is a difference in speed. The times (in seconds) used by each processors on each code are given in the following table.

	Code					
	1	2	3	4	5	6
Processor A	27.2	18.1	27.2	19.7	24.5	22.1
Processor B	24.1	19.3	26.8	20.1	27.6	29.8

(i) Find a 95% confidence interval for the mean difference of speeds between two processors.

(ii) Can you conclude that the mean speeds of the two processors differ? Use an appropriate statistical test at the $\alpha = 0.05$ level.

6. A firm has two possible sources for its computer hardware. It is thought that supplier X tends to charge more than supplier Y for comparable items. The following are the price data collected on 10 items supplied by both X and Y.

Item	Price(X), $	Price(Y), $	Item	Price(X), $	Price(Y), $
1	6,000	5,900	6	5,650	5,600
2	575	580	7	10,000	9,975
3	15,000	15,000	8	850	870
4	150,000	145,000	9	900	890
5	76,000	75,000	10	3,000	2,900

(i) Use the Signed Rank test to test if the data support the above contention at the $\alpha = 0.05$ level. (For one-sided Signed Rank test with $n = 10$ and $\alpha = 0.05$, the critical value for the test statistic is 11.)
(ii) Use the Sign test to test if the data support the above contention at the $\alpha = 0.05$ level. Does the Sign test yield the same results as the Signed Rank test? If not, what are the possible reasons for the discrepancy?

7. (You may need percentage points of the F-distribution given at the end of this problem to answer the following questions.)

Part I: Giving the following information from SAS PROC REG, answer questions (i) – (iii).

Dependent Variable: Y

R-Square Selection Method

Number of Observations Read 46
Number of Observations Used 46

Number in Model	R-Square	Adjusted R-Square	MSE	SSE	Variables in Model
1	0.6190	0.6103	115.77081	5093.91550	X1
1	0.4155	0.4022	177.59980	7814.39120	X3
1	0.3635	0.3491	193.38737	8509.04435	X2
2	0.6761	0.6610	100.70930	4330.49973	X1 X3
2	0.6550	0.6389	107.27907	4613.00020	X1 X2
2	0.4685	0.4437	165.26498	7106.39406	X2 X3
3	0.6822	0.6595	101.16287	4248.84068	X1 X2 X3

(i) Using the adjusted R^2 model selection criterion find the best model. Justify your choice.
(ii) For the full model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2 I)$, test the following hypothesis. Use $\alpha = 0.05$.
$H_0 : \beta_1 = \beta_3 = 0$ v.s. H_1 : not both β_1 and β_3 are equal to 0
(iii) Use the forward regression to select the best model. Use $\alpha = 0.05$ as the significance level for entry.

Part II:

(iv) The following is the SAS PROC REG output for the full model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2 I)$. Fill in the numbered blanks (1) and (2). Show your steps.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	9120.46367	3040.15456	30.05	<.0001
Error	42	4248.84068	101.16287		
Corrected Total	45	13369			

Root MSE	10.05798	R-Square	0.6822
Dependent Mean	61.56522	Adj R-Sq	0.6595
Coeff Var	16.33711		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Type I SS	Type II SS
Intercept	1	158.49125	18.12589	8.74	<.0001	174353	7734.52
X1	1	-1.14161	0.21480	-5.31	<.0001	**(1)**	**(2)**
X2	1	-0.44200	0.49197	-0.90	0.3741	480.92	81.66
X3	1	-13.47016	7.09966	-1.90	0.0647	364.16	364.16

Percentage Points for the F-distribution $F_{v_1,v_2,0.05} = F^*$ implies $P(F_{v_1,v_2} > F^*) = 0.05$

$F_{1,41,0.05} = 4.08$	$F_{1,42,0.05} = 4.07$	$F_{1,43,0.05} = 4.07$	$F_{1,44,0.05} = 4.06$
$F_{2,41,0.05} = 3.23$	$F_{2,42,0.05} = 3.22$	$F_{2,43,0.05} = 3.21$	$F_{2,44,0.05} = 3.21$

8. A chemical production process consists of a first reaction with an alcohol and a second reaction with a base. A factorial experiment with three alcohols and two bases was conducted with three replicate reactions conducted in a completely randomized design. The collected data were percent yield.

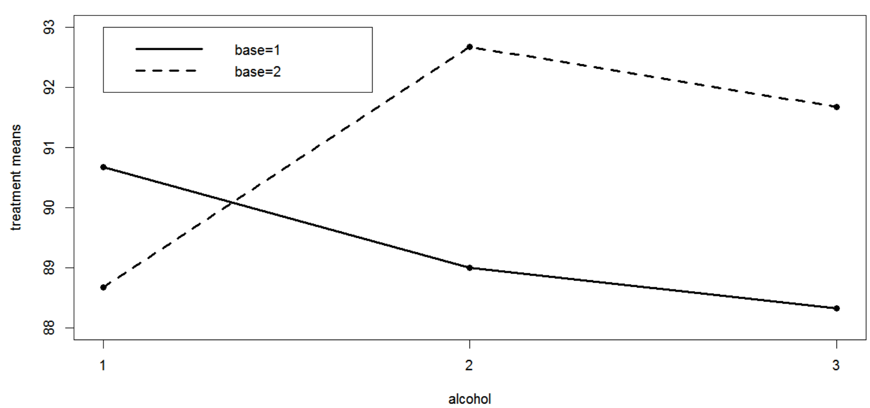
Base	**Alcohol**			
	1	2	3	
1	91, 90, 91	89, 88, 90	87, 88, 90	$\bar{Y}_{1..} = 89.33$
2	87, 88, 91	91, 92, 95	90, 92, 93	$\bar{Y}_{2..} = 91$
	$\bar{Y}_{.1.} = 89.67$	$\bar{Y}_{.2.} = 90.83$	$\bar{Y}_{.3.} = 90$	$\bar{Y}_{...} = 90.17$

(i) Write an ANOVA model for this experiment. Explain the terms and specify assumptions.
(ii) What are the constraints need to be satisfied?
(iii) What are the estimates of effects for base =1, and for the following combination of base and alcohol: (base, alcohol)=(2,3)?
(iv) The following is the ANOVA table from SAS. Calculate the missing values in the numbered blanks (1)-(6) and test if the main effects and interactions are significant at $\alpha = 0.05$.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	47.16666667	9.43333333	3.86	0.0257
Error	12	29.33333333	2.44444444		
Corrected Total	17	76.50000000			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
base	**(1)**	12.50000000	12.50000000	**(4)**	0.0431
alcohol	**(2)**	4.33333333	2.16666667	**(5)**	0.4375
base*alcohol	**(3)**	30.33333333	15.16666667	**(6)**	0.0141

(v) Interpret the following interaction plot between base and alcohol.



9. The surface finish of metal parts made on four machines is being studied. An experiment is conducted in which each machine is run by three different operators and two specimens from each operator are collected and tested. Because of the location of the machines, different operators are used on each machine, and the operators are chosen at random.

(i) Write down the appropriate model for this experiment along with the assumptions.

(ii) Clearly specify the expected mean squares for each component in the ANOVA table for this experiment and construct the appropriate F-test based on the expected mean squares.