

Preliminary Exam in Distribution Theory & Regression Analysis: Summer 2011

1.(i) If X, Y are mutually *independent* non-degenerate random variables (r.v.) with finite variances, and if η_1 and η_2 are the c.v. (*coefficient of variation*) of X and Y respectively; show that the c.v. η_* of their product XY strictly exceeds the c.v. of both X and Y , by proving

$$(1 + \eta_*^2) = (1 + \eta_1^2)(1 + \eta_2^2).$$

(The c.v. of any r.v. X with $EX^2 < \infty$, is defined as $\text{c.v.} = \frac{\sqrt{\text{var}(X)}}{|EX|}$, a dimensionless measure of variability.)

(ii) Suppose X is a strictly positive r.v. ($P(X > 0) = 1$). To answer a) - c) below; assume that appropriate moments of X and $(\frac{1}{X})$ are finite.

- a) Using Cauchy-Schwarz inequality, show that X and $\frac{1}{X}$ cannot be strictly positively correlated.
- b) What additional condition can imply a strict negative correlation between X and $\frac{1}{X}$?
- c) If Y is another r.v. such that $\text{cov}(X, Y + \frac{1}{X}) > 0$; then show that we must have $\text{cov}(X, Y) > 0$.

2. A random variable (r.v.) Y is said to be *Skew-Normally* distributed with parameter α (and denoted by $Y \sim \text{SN}(\alpha)$; $-\infty < \alpha < \infty$), if it has a probability density function (p.d.f.)

$$g(y; \alpha) := 2\phi(y)\Phi(\alpha y); \quad -\infty < y < \infty$$

where $\phi(y)$ and $\Phi(y)$ are respectively the p.d.f. and the c.d.f.(\equiv cumulative distribution function) of the standard Normal distribution $N(0, 1)$.

(i) Show that $g(y; \alpha)$ is indeed a p.d.f. for each $\alpha \in (-\infty, \infty)$. For what value of α is $Y \sim N(0, 1)$?

(ii) If $Z \sim N(0, 1)$ and $Y \sim \text{SN}(\alpha)$; prove that $|Y|$ and $|Z|$ are identically distributed. What is the distribution of Y^2 ?

(iii) Suppose r.v.s Z and W are i.i.d. $N(0, 1)$, and let $Y \stackrel{d}{=} Z | \alpha Z > W$; i.e., Y has the same distribution as the *conditional distribution* of Z , given $\alpha Z > W$. Then, show that $Y \sim \text{SN}(\alpha)$.

3. The joint distribution of (X, Y) is specified by the conditional distribution of X given Y , together with the marginal distribution of Y , as follows :

$$X | y \sim N(0, y), \quad Y \sim \text{Exp} \left(\frac{1}{2} \right);$$

where $N(0, y)$ denotes the Normal distribution with mean zero and *variance* y , and $\text{Exp}(\lambda)$ denotes the exponential distribution with the density $\lambda e^{-\lambda x}$, on $(0, \infty)$.

Using the method of moment generating functions or, otherwise; find the marginal distribution of X .

4. Observations Y_1, \dots, Y_n are described by the relationship $Y_i = \theta x_i^2 + \epsilon_i$, where x_1, \dots, x_n are fixed constants and $\epsilon_1, \dots, \epsilon_n$ are independent with common distribution $N(0, \sigma^2)$.

- (i) Find the least squares estimator of θ .
- (ii) Find the maximum likelihood estimator of θ .
- (iii) Give an unbiased estimator of θ that has minimum variance among all unbiased linear estimators of θ . What is its variance?

5. This question has two independent parts.

- (i) For the simple linear regression model, develop the Bonferroni procedure for obtaining joint confidence intervals of β_0 and β_1 with family confidence coefficient at least $1 - \alpha$. Derive any formulas that you may need to develop the procedure.
- (ii) Dwaine Studios, Inc., operates portrait studios for children in 21 cities of moderate size. The company is considering an expansion into other cities of medium size and wishes to investigate whether sales in thousands of dollars (Y) in a community can be predicted from the number of persons aged 16 or younger (expressed in thousands of persons) in the community (X_1) and the per capita disposable personal income (expressed in thousands of dollars) in the community (X_2). A multiple regression partial Minitab output is given below:

The regression equation is
 $Y = -68.9 + 1.45 X_1 + 9.37 X_2$

Predictor	Coef	SE Coef	T	P
Constant	-68.86	60.02	-1.15	0.266
X1	1.4546	0.2118	6.87	0.000
X2	9.366	4.064	2.30	0.033

S = 11.0074 R-Sq = 91.7% R-Sq(adj) = 90.7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	24015	12008	99.10	0.000
Residual Error	18	2181	121		

The matrix $(\mathbf{X}'\mathbf{X})^{-1}$ is given below:

	1	2	3
1	29.7289		
2	0.0722	0.00037	
3	-1.9926	-0.0056	0.1363

Predict the sales for two new cities with the following characteristics with a 90% family confidence coefficient.

	City A	City B
Number of persons aged 16 or younger	65.4	53.1
Per capita disposable income	17.6	17.7

6. This question has two independent parts.

- (i) Derive the sampling distribution of the slope estimate in the simple linear regression model.
- (ii) A company operates two production lines for making soap bars. For each line, the relation between the speed of the line and the amount of scrap for the day was studied. A scatter plot of the data for the two production lines suggests that the regression relation between production line speed (X_1) and amount of scrap (Y) is linear but not the same for the two production lines. An analyst decided to fit the regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \epsilon_i, \quad (1)$$

where $\epsilon \sim N(0, \sigma^2)$, $i = 1, \dots, 27$; $X_2 = 1$ for production line 1 and $X_2 = 0$ for production line 2. A partial output is given below (values indicated by ** may be completed by exam-taker). Answer parts (a)–(c) below.

Regression Coefficients			
Regression Coefficient	Estimated Regression Coefficient	Estimated Standard Deviation	t^*
β_0	7.57	20.87	0.36
β_1	1.322	0.09262	14.27
β_2	90.39	28.35	3.19
β_3	-0.1767	0.1288	-1.37

Analysis of Variance			
Source of Variation	SS	df	MS
Regression	169,165	**	**
X_1	149,661	**	**
$X_2 X_1$	18,694	**	**
$X_1X_2 X_1, X_2$	810	**	**
Error	9,904	**	**
Total	179,069	**	

- (a) Explain the characteristics of the regression lines permitted by the model given by Eq. (1).
- (b) Test whether or not the regression functions for the two production lines are identical.
- (c) Test whether or not the slopes of the two regression lines are the same.