A. Multiple Choice Questions: (50%, The 10 questions are weighted equally.)
Instruction: There may be more than one correct responses to each question.

1. Which of the following statements, if any, is (are) not the necessary condition(s) for an integrable function $f(x)$ to be the probability density function (p.d.f.) of a random variable X of the continuous type, with space R.
   (1) $f(x) > 0,\ x \in R$.
   (2) $0 < f(x) < 1,\ x \in R$.
   (3) $\int_R f(x)dx = 1$.
   (4) The probability of the event $X \in A$ is $P(X \in A) = \int_A f(x)dx$.
   (5) E(X) exist and $0 < Var(x) < \infty$.
2. If A and B are two mutually exclusive events, P(A)=0.5 and P(A∪B)=0.6, then
   (1) P(B) = 0.1.
   (2) A and B are dependent.
   (3) For any two mutually exclusive events, they must be independent to each other.
   (4) For any two mutually exclusive events, they must be dependent to each other.
3. Which of the following statements correctly apply to Chi-square distributions (χ^2):
   (1) If random variable X is $N(\mu,\sigma^2)$, $\sigma^2>0$, then the random variable $V=(X-\mu)^2/\sigma^2$ is $\chi^2(1)$.
   (2) If X_1 and X_2 are two random variables which are chi-square distributed with degrees of freedom r_1 and r_2, respectively. Then $Y=X_1+X_2$ would also be chi-square distributed with degree of freedom (r_1+r_2).
   (3) If $r_1>r_2$, then $\chi^2_{0.05}(r_1) > \chi^2_{0.05}(r_2)$, where r_1 and r_2 are degree of freedom.
   (4) none of the above.

4. An urn contains 10 red and 6 white balls. Draw one ball at random from the urn. Suppose the ball is drawn with replacement. A random variable X is defined as the trial number on which the first white ball occurs. Then
   (1) X has a geometric distribution.
   (2) $f(x)=(5/8)^{x-1}(3/8), \quad x = 0,1,2,\ldots$
   (3) E(X)=8/3.
   (4) Suppose the ball is drawn without replacement, then X has a negative binomial distribution.
5. Suppose Y has a binomial distribution, i.e. Y ~ b(36, 0.5), where sample size is 36 and the population success rate is 0.5.
   (1) E(Y)=18, Var(Y)=9.
   (2) $P(12<Y\leq 18) \cong \phi(0.167) - \phi(-1.833)$, where $\phi(.)$ is the distribution function of the standard normal distribution.
   (3) $P(11<Y\leq 14) = P(12\leq Y\leq 14)$.
   (4) All of the above.

The next three questions (from no.6 to no.8) refer to the following setting: A random sample of size n is drawing from a normal distribution with mean μ and variance 100. H_0: $\mu=60$, and H_1: $\mu>60$. The rejection area C is defined as $C = \{\bar{x}: \bar{x} \geq 62\}$.
[Note: Some values of z_α are shown as follows. $z_1=0.8413$, $z_{0.80}=0.7881$, $z_{1.50}=0.9332$, $z_{1.58}=0.9429$, $z_{2.00}=0.9772$, $z_{1.983}=0.9763$, $z_{1.973}=0.9759$]

6. What of the following statement, if any, is (are) correct when the sample size n=25:
   (1) The power at $\mu=60$ is 0.2119,
   (2) The significant level of the test is 0.1587,
   (3) The type two error when $\mu=65$ is 0.0571,
   (4) The type two error when $\mu=65$ is 0.0668.
7. Furthermore, suppose type one error α is set to 0.025 and Type two error is set to 0.05 when the true mean $\mu=65$. What should the critical point c and the sample size n be?
   (1) c=62.718,
   (2) c=63.718,
   (3) n=52,
   (4) n=53.

8. Which of the following statements, if any, does (do) not correctly apply to R^2 or $\bar{R}^2$ (the adjusted R^2):
   (1) Suppose F statistic tests the joint hypothesis that all the coefficients except the intercept equal zero. Then the higher the value of R^2 is, the higher the value of the F statistic will be.
   (2) If all the coefficients of the linear regression model have high t values, then R_2 must be high as well and vise versa.
   (3) If the number of explanatory variable equals one, then $R^2=\bar{R}^2$.
   (4) $\bar{R}^2$ which can be less than zero is never larger then R^2.

The next two questions (no.9 and no.10) refer to the following setting: Suppose that a firm uses two types of production processes to obtain its output. Upon the assumption that the output obtained from each process is normally distributed with different expected values but identical variances, we can represent the production process as a regression equation:

$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$,
where Y_i is the output associated with the i^{th} input process and X_i is a dummy variable
where $X_i = 1$ if output associated from machine A,
$= 0$ if output associated from machine B.

9. Which of the following statements, if any, is (are) incorrect?
   (1) The intercept of the regression line measures the expected output associated with machine B.
   (2) The insignificant t value of the estimate of β_1 represents that there is no difference in the output associated with machines A and B.
   (3) The failure of rejecting the F test from this regression means that there is no difference in the output associated with machines A and B.
   (4) none of the above.

10. If the researcher set up the regression model, in stead, as follows:
$Y_i = \alpha_1 X_i + \alpha_2 Z_i + \epsilon_i$,
where $X_i = 1$ if output associated from machine A,
$= 0$ if output associated from machine B.
$Z_i = 0$ if output associated from machine A,
$= 1$ if output associated from machine B.
Then which of the following statements related to the relationship between these two regression models, if any, is (are) accurate:
   (1) There exists a problem of multicolinearity in this linear regression model.
   (2) $\alpha_1 = \beta_0 - \beta_1$ for machine A.
   (3) $\alpha_2 = \beta_0$ for machine B.
   (4) None of the above.

B. Problem Solving (50%)

1. Bob "Mr. Baseball" Uecker, journeyman catcher and movie/commercial star, will celebrate his 56^{th} birthday. In honor of this day, his career should be reevaluated. His career statistics include

Year	At Bat (打數)	Hits (安打)	Average	Home Run	RBI (打點)	Strike-out (三振)	Walks (保送)
1980	64	16	.250	1	8	15	7
1981	16	4	.250	0	0	5	2
1982	106	21	.198	1	6	24	17
1983	145	33	.228	2	10	27	24
1984	207	43	.208	7	30	36	22
1985	193	29	.150	3	20	60	24

[Note: $t_{0.05}(5)=2.015$, $t_{0.05}(6)=1.943$, $t_{0.025}(5)=2.571$, $t_{0.025}(6)=2.447$]

(1) Calculate Uecker's career batting average and strikeouts per season. What is the median number of walks per season? (3%)
(2) Calculate the variance and standard deviation of RBI's per season. (3%)
(3) Test the following hypothesis at the 10% significance level
H_0: Bob Uecker was a career .250 hitter (3%)
(4) If Uecker would have played in the 1986 season. In what range would you be 95% certain that his hits this season would have fallen? (3%)
(5) Test the following hypothesis at the 5% significance level
H_0: Uecker averaged the same number of hits as walks each season. (3%)

2. Is it true that the estimated slope of the regression of Y on X will always equal the reciprocal of the estimated slope of the regression of X on Y? Please prove it and explain intuitively why is that. (15%)

3. Suppose a researcher use a set of "wage data" to estimate a model with both the intercept dummy and a term which allows the coefficient on experience differ between sexes. i.e.
$$Earn = \beta_0 + \beta_1 Exp + \beta_2 Educ + \beta_3 Sex + \beta_4 (Sex*Exp) + \epsilon,$$
where Earn=earnings, Exp=experience, Educ=education, and ϵ is the error term of the normal assumptions. Suppose the F-test on the wage regressions suggest that earnings of men and earning of women have similar intercepts and returns to education, but men receive higher returns to experience. However, it turns out the experience is not measured directly, but is calculated as Age−Educ−6: i.e. it is assumed that all individuals work full time starting from when they leave school. However, it might be the case in reality that some people spend a number of years "out of the labor force" (voluntarily no employed at a paying job)
(1) What would happen if you tried to include Age along with Educ and Exp as a regressor? Why? (10%)
(2) Suppose the return to experience, properly measured, is the same for men as for women. Can the differing estimated returns be explained if women spend more time out of the labor force than men, on average? Why? (10%)

參考用