題號： 267
科目： 統計學(E)
節次： 2

國立臺灣大學 112 學年度碩士班招生考試試題

題號：267
共 3 頁之第 1 頁

注意：本試卷共有三頁。請考生答題前，務必閱讀每大項的注意事項說明！

第一大題：計算問答說明題（50 分）

注意：（1）第一大題有二題計算問答說明題。

（2）請標示清楚，並將所有過程、步驟交代清楚；沒有說明過程者，甚者只給簡單回答如 Yes、No 等，不給分。每大題之下的小題分數，如括號內所示。

Note: You should carefully state the reasons or calculations in the following questions in order to get the points. A short answer, such as "Yes" or "No" will NOT receive any point.

1. **(40 points)** A researcher is interested in the relationship between maternal smoking and infant health (measured by birth weight). To study this question, he collected a sample of 189 birth records that contain information on the birth weight of infants as well as health status and demographic characteristics of mothers. The table below presents the definitions and sample means of the selected variables for the full sample, the smoking mother subsample, and the non-smoking mother subsample, respectively. Using this table to answer the following questions.

Variable	Definition	Full sample	Smoking mothers	Non-smoking mothers
bwt	birth weight (g)	2944	2772	3055
smoke	=1 if smoking during pregnancy, 0 otherwise	0.392	1	0
age	Age of mother	23.238	22.946	23.426
white	=1 if white race, 0 other	0.508	0.703	0.383
black	=1 if black race, 0 other	0.138	0.135	0.139
other	=1 if other race, 0 other	0.354	0.162	0.478
ht	=1 if having history of hypertension, 0 otherwise	0.063	0.068	0.061
ftv	Number of prenatal visits	0.794	0.757	0.817
Observations		189	74	115

A. To begin with, the researcher specifies a simple linear regression model:
$$bwt_i = \beta_0 + \beta_1\, smoke_i + \varepsilon_i.$$
Write down the <u>estimated</u> regression equation. (6 points)

B. Interpret the estimated coefficients of β_0 and β_1. (6 points)

C. At the 5% significance level, what is your conclusion regarding the relationship between maternal smoking and birth weight? The standard error of the coefficient estimate of β_1 is 107 and the critical value of the test statistic at the 5% significance level is 1.96. Be sure to write down each step of the hypothesis testing procedure. (8 points)

D. (i) Calculate the 95% confidence interval for the estimated coefficient of *smoke* and (ii) interpret it. (6 points)

E. The researcher next extends the simple linear regression model to a multiple linear regression model:
$$bwt_i = \beta_0 + \beta_1\, smoke_i + \beta_2\, age_i + \beta_3\, white_i + \beta_4\, black_i + \beta_3\, ht_i + \beta_4\, ftv_i + \varepsilon_i.$$
What is the point of including additional explanatory variables in the model? (2 points)

見背面

題號： 267
科目： 統計學(E)
節次： 2

國立臺灣大學 112 學年度碩士班招生考試試題

題號：267
共 3 頁之第 2 頁

F. Is it always good to include more and more explanatory variables in the model? Explain. (4 points)

G. Why does the researcher only include two variables of race (*white* and *black*) in the model, even though there are three race variables in the table (*white, black,* and *other*)? (4 points)

H. Suppose the researcher additionally suspects that the effect of maternal smoking depends on the age of mother. What model specification can you use to test such a relationship? Write down the regression model. (4 points)

2. **(10 points)** Suppose in a county 50%, 30%, and 20% of residents live in urban, suburban, and rural areas, respectively. Additionally, 60% of urban residents have a bachelor's degree, 50% of suburban residents have a bachelor's degree, and 40% of rural residents have a bachelor's degree.

A. What is the probability that a randomly selected resident has a bachelor's degree? (5 points)

B. For a randomly selected resident who has a bachelor's degree, what is the probability that the resident lives in an urban area? (5 points)

第二大題：填空題（每格 5 分，共 50 分）

（1）第二大題為填空題，共有 10 個空格，每一空格 5 分；此部分<u>不須</u>計算過程。

（2）答題不要求任何計算過程，只依答案格內的答案對錯給分。

（3）如果沒有特別指示，請將答案約分至「<u>最簡分數</u>」表示，否則不予計分。

（4）請自行製作<u>答題區</u>。範例如下：

第 1 格	第 2 格	第 3 格	第 4 格	第 5 格
第 6 格	第 7 格	第 8 格	第 9 格	第 10 格

Part II：填空題

A. Let the bivariate normal random variables X and Y have the linear conditional means:

$E(Y|x) = 1 - x$ and $E(X|y) = 5 - y/9$, as well as the unconditional variance of X, $V(X) = 4$.

Then, $E[X] + E[Y] = \underline{\quad (1) \quad}$; $V[Y] = \underline{\quad (2) \quad}$; the correlation coefficient between X and Y is $\underline{\quad (3) \quad}$.

題號： 267
科目： 統計學(E)
節次： 2

國立臺灣大學 112 學年度碩士班招生考試試題

題號：267

共 3 頁之第 3 頁

B. The length of life is a random variable Y with a density function given by: $f(y|\alpha, m) = \frac{my^{m-1}e^{-y^m/\alpha}}{\alpha}, 0 \leq y < \infty, \ \alpha > 0, m > 0$ and 0, elsewhere is said to have a **Weibull** distribution. Its distribution function $F(y|\alpha, m) = \underline{\quad (4) \quad}$. When $m = 1$, the moment generating function (MGF) of Y is $\underline{\quad (5) \quad}$; $E[Y^2|\alpha, m = 1] = \underline{\quad (6) \quad}$; $E[Y|Y \geq c, \alpha, m = 1] = \underline{\quad (7) \quad}$.

C. The duration X of long-distance telephone calls (in minutes) monitored by a station is a random variable with the properties that $P[X = 3] = 1/3$ and $P[X = 6] = 1/6$; otherwise, X has a **chi-squared** distribution with 4 degrees of freedom. $E[X] = \underline{\quad (8) \quad}$; $E[X^2] = \underline{\quad (9) \quad}$; $V[X] = \underline{\quad (10) \quad}$.

試題隨卷繳回