注意事項： 1. 答案一律寫在試卷上，否則不予計分。
2. 請標明題號依序作答，不必抄題。
3. 試題應隨同試卷繳回，不得攜出試場。

1. (28 points) An urn contains 3 white balls and 4 red balls and balls are to be drawn from the urn.

(a) If the balls are drawn with replacement n times and we count the number of white balls (denoted as Y_1) in the sample. What is the distribution of Y_1 ?

(b) For (a), find $E(Y_1)$ and $Var(Y_1)$.

(c) If the balls are drawn without replacement n times and we let Y_2 be the number of white balls in the experiment. What is the distribution of Y_2?

(d). For (c), find $E(Y_2)$ and $Var(Y_2)$.

(e) Compare the results in (b) and (d). Explain intuitively the results.

(f) If the experiment now is set to be drawing balls with replacement from the urn until one white ball is observed. Let Y_3 be the number of times needed to have one white ball. What is the distribution of Y_3?

(g) What is the distribution of Y_3 if now the condition of the experiment in (f) "**with replacement**" is replaced by "**without replacement**"?

2.(12 points) State True or False for the following statements. You <u>must</u> briefly describe the reasons why they are true or false.

(a) If X and Y are two random variables and Cov(X,Y) = 0, then X and Y are independent.

(b) If X and Y are two normal random variables, then X+Y and X-Y are normal and are independently distributed.

(c) If (X,Y) is bivariate normal, then X+Y and X-Y are normal and are independently distributed.

(d) If (X,Y) is bivariate normal and Cov(X,Y) = 0, then X and Y are independent normal.

3.(12 points) Let $X_1, X_2, \ldots, X_6$ be an i.i.d. random sample from N(2,1). Let $Y_i = 1$ if $X_i \le 2$; $= 0$, otherwise. Define $T = \sum_{i=1}^{6} Y_i$.

(a) Find E(T) and Var(T).

(b) What is the distribution of T? Compute $P(T = j)$, the probability of $T = j$, where j ranges over all the possible values of T.

(c) Compute $P(T = j)$ by normal approximation with continuity correction.

(d) Comment on the results in (b) and (c).

4. (8 points) We usually use sample mean, median and trimmed mean to represent the concentration of the collected data. State the advantages and disadvantages of the three location measures.

Note: A $p\%$ trimmed mean for a random sample $X_1, X_2, \ldots, X_n$ is defined as the average of the data left after $(1/2)p\%$ of the smallest and $(1/2)p\%$ of the largest ordered values of $X_1, X_2, \ldots, X_n$ are trimmed.

(背面仍有題目,請繼續作答)

5. (5 points) Suppose $X_1, X_2, \ldots, X_{n_1}$ is an i.i.d. random sample from $N(\mu_1, \sigma_1^2)$ and $Y_1, Y_2, \ldots, Y_{n_2}$ is an i.i.d. random sample from $N(\mu_2, \sigma_2^2)$. From these two sets of data, we found range of $X_i = 2$ and range of $Y_i = 10$. Can we claim that the variance of Y is larger than the variance of X? Why?

6. (5 points) Suppose we have two sets of data coming from two normal distributions. The sample means and sample sizes are $\bar{X} = 5$, $n_1 = 100$ and $\bar{Y} = 100$, $n_2 = 100$, respectively. Then in testing $H_0: \mu_1 = \mu_2$ *vs* $H_1: \mu_1 \neq \mu_2$, we will conclude H_1 since $\bar{Y} - \bar{X} = 95$ is sufficiently large.

For Problems 7 and 8, suppose we have normal populations.

7. (5 points) In testing $H_0: \mu_1 = \mu_2$ *vs* $H_1: \mu_1 \neq \mu_2$, why we usually construct the test statistic based on $\hat{\mu}_2 - \hat{\mu}_1 = \bar{Y} - \bar{X}$, but $\bar{Y}/\bar{X}$?

8. (5 points) In testing $H_0: \sigma_1^2 = \sigma_2^2$ *vs* $H_1: \sigma_1^2 \neq \sigma_2^2$, why we usually construct the test statistic based on $\hat{\sigma}_1^2/\hat{\sigma}_2^2 = \hat{S}_1^2/\hat{S}_2^2$, but not $\hat{\sigma}_1^2 - \hat{\sigma}_2^2 = \hat{S}_1^2 - \hat{S}_2^2$? Here $\hat{S}_i^2$ is unbiased sample variance of σ_i^2, $i = 1,2$.

9. (5 points) Let "X and Y be two random variables. If $Cov(X,Y) = 0$, then X and Y do not have any relationship." Is this statement correct? Why?

10. (15 points) The following data are collected

Y_i:	1	2	3	4	5.1
X_i:	-0.2	-0.1	0	0.1	0.2

(a) Give the scatterplot. Does the plot reveal a strong linear association between Y and X?
(b) Suppose simple linear regression model is suitable for the data, write down the model and the necessary assumptions so that inference based on small sample can be made. Suppose $Var(Y_i) = 10$.
(c) Find the estimated least squares line.
(d) Test $H_0: \beta_1 = 0$ *vs* $H_1: \beta_1 \neq 0$; i.e. whether there exists linear association between Y and X at $\alpha = 0.05$.
(e) Does the results in (d) contradict your observation in (a)? Why?