科目名稱	統計學	類組代碼	D38
		科目碼	D3801

※本項考試依簡章規定各考科均「不可以」使用計算機　　本科試題共計 9 頁

Choose the most appropriate ONE.

1. (4 points)

(A) The mean of a sample will be equal to the mean of the population.

(B) Outlier has undue effect on the sample mean, so does on the sample median.

(C) Mean absolute deviation is easier to understand than standard deviation, but people seldom apply it because its mathematical property is hard to derive.

(D) Standard deviation, not like sample mean, won't be greatly influenced by outlier(s).

(E) Two samples, one with range 10, the other with range 15, then the variation of the second sample is large than the first one.

2. (4 points)

(A) The hourly wages of a sample of 130 system analysts are mean = 60, median = 74, range = 20, variance = 324, then the coefficient of variation equals 30%.

(B) When data are negatively skewed, the mean will usually be greater than the median.

(C) Positive values of variance indicate positive relation between the independent and the dependent variable.

(D) The coefficient of correlation can be larger than 1.

(E) None of the above 4 questions.

3. (5 points)

(A) Suppose A_1, A_2 and A_3 are three sets, if $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$, then $P(A_i \cap A_j) = P(A_i)P(A_j), i \neq j$.

(B) Suppose sets A_1, A_2 and A_3 are three sets in the sample space S, and $A_i \cap A_j = \phi, i \neq j$. Let D be any set in S, then

$P(D) = \sum_{i=1}^{3} P(A_i)P(D|A_i)$ and $P(A_1|D) = \frac{P(A_1)P(D|A_1)}{\sum_{i=1}^{3} P(A_i)P(D|A_i)}$.

(C) Let A and B be two events with $P(A) = 0.4$, $P(B) = 0.3$, $P(A \cap B)$=0.2, then the probability of only one of A or B occurs is 0.5.

(D) X is a random variable taking values 0 and 1 respectively, also Y is a random variable taking values 10 and 20 only. If P(X=0, Y=10) = P(X=0)P(Y=10), then P(X=1, Y=10) = P(X=1)P(Y=10), P(X=0, Y=20) = P(X=0)P(Y=20) and P(X=1, Y=20) = P(X=1)P(Y=20).

(E) Two continuous random variables X, Y, and one discrete random variable Z taking values 1 and 2. If Y increases with X, then Y also increases with X for each value of Z.

　背面有題，請繼續作答。

4. (4 points) One box contains 1 red ball and 3 white balls. Three persons are to draw one ball in order. Let $X_i = 1$, i =1,2,3, for person i draws a red ball, $X_i = 0$, otherwise.
   (A) Each person has the same probability in drawing the red ball, E(X_i)=1/4, no matter his order in drawing ball from the box.
   (B) The variances of each X_i are equal, which is 3/16.
   (C) The X_i, i=1.2.3, are identically distributed.
   (D) The probability of drawing a red ball for the second person depends on the outcome of the first person.
   (E) True for all the above.
5. (4 points)
   (A) Let A, B be sets in sample space S, $\emptyset$ be empty set to S. Then sets A and $\emptyset$ are mutually exclusive,
   (B) If both A and B are not empty sets, then they cannot be independent and mutually exclusive simultaneously.
   (C) The skewness of a Poisson distribution is always positive. It cannot be negative.
   (D) True for all the above (A), (B) and (C).
   (E) None for the above (A), (B), (C) and (D).
6. (4 points) In a statistics class, the average grade on the final examination was 75 with a standard deviation of 5.
   (A) The value of the sum of the deviations from the mean, i.e., $\sum\left(x-\bar{x}\right)$ may not be zero, where $\bar{x}$ is the sample mean.
   (B) Using Chebyshev's theorem, at least 96 percentage of the students received grades between 50 and 100.
   (C) If the grades are normal, then 95% of the students will receive grades in between 60 and 90.
   (D) By central limit theorem, the distribution of the course grades will close to be a normal if the class size is large.
   (E) Wrong for all the above (A), (B), (C), and (D).
7. (5 points) Shown below is a portion of a computer output for regression analysis relating y (dependent variable) and x (independent variable).

ANOVA		
	df	*SS*
Regression	1	24.011
Residual	8	67.989

	Coefficients	*Standard Error*
Intercept	11.065	2.043
x	-0.511	0.304

(A) The sample size for the above regression analysis is 9.

(B) It would be significant if we perform a *t* test to determine whether or not *x* and *y* are related. Let $\alpha = .05$.

(C) Performing an *F* test to determine whether or not *x* and *y* are related would have the same results as the t test at $\alpha = .05$.

(D) The square root of the F statistic is the t statistic.

(E) The correlation coefficient of X and Y is 0.51.

8. (4 points) Let (Y_i, x_i), i= 1,2,…,n , be a random sample.

(A) Since the sample correlation coefficient r = 0.92 is large, simple linear regression model would be suitable in modelling the relationship for Y and x.

(B) The estimated regression coefficient would have the same value as the correlation coefficient if both the sample standard deviation of Y and x are 1's.

(C) Normal distribution assumption is a MUST for the error term if we want to find the least squares estimates.

(D) A significant result can be obtained if r = 0.92.

(E) None of the above.

9. (5 points) Part of an Excel output relating *x* (independent variable) and *y* (dependent variable) is shown below.

Summary Output:

Regression Statistics

R Square	0.5149
Root MSE	7.3413
Observations	11

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	?	(A)	?	(C)	0.0129
Residual	?	?	(B)		
Total	?	1000			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	?	29.4818	3.7946	0.0043
x	(D)	0.7000	-3.0911	0.0129

Then (A) 514.9　(B) 53.9　(C) 9.55　(D) -2.1638　(E) True for all the above four.

10. (4 points) Let p_1 and p_2 be the proportions for some characteristic in populations 1 and 2. Random samples with size n_1 and n_2 respectively are drawn from the two populations and found that the sample proportions are $\hat{p}_1, \hat{p}_2$. We are interested in testing $H_0: p_1 = p_2$

(A) The test statistic t should be taken to be

$$t = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

(B) The test statistic

$$t = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

where $\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1+n_2}$ is better than the one in (A).

(C) The test statistic in (B) is also good for the test $H_0: p_1 - p_2 = d_0$ vs $H_a: p_1 - p_2 \neq d_0$, where d_0 is some known value.

(D) The test statistic in (A) is also good if $n_1 + n_2$ is large enough.

(E) All the above are correct.

11. (4 points) Two independent random samples are drawn from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ with sizes $n_1 = 10$, $n_2 = 16$, respectively. It is found that $\overline{x_1} = 8$, $\overline{x_2} = 5$, $s_1^2 = 2$, $s_2^2 = 1$.

(A) The random variable $\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$ can be used to construct a 95% confidence interval for $\frac{\sigma_1^2}{\sigma_2^2}$ by using F random variable with 9 and 15 degrees of freedom.

(B) Based on the results in (A), $H_0: \sigma_1^2 = \sigma_2^2$ would be concluded if $\alpha = 0.05$.

(C) To test $H_0: \mu_1 = \mu_2$, the test statistic to be used would be a *t* with 24 degrees of freedom.

(D) True for all above (A),(B) and (C).

(E) None for the above (A),(B), (C) and (D).

12. (5 points) Consider the paired data: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ and we want to compare the means of X and Y, μ_x, μ_y. Suppose we have $\bar{x}, \bar{y}, S_x^2, S_y^2$ and *r*, the sample correlation coefficient, where S_x^2 and S_y^2 are the unbiased estimators for σ_x^2, σ_y^2, and r is positive.

(A) To test $H_0: \mu_x = \mu_y$, the test statistic $t = \frac{(\bar{x}-\bar{y})}{\sqrt{\frac{S_x^2}{n}+\frac{S_y^2}{n}}}$ is a good choice.

(B) Let $d_i = x_i - y_i$, the test statistic $t_d = \frac{(\bar{x}-\bar{y})}{S_d\sqrt{\frac{1}{n}}}$ is a better choice than the *t* in (A) because $S_d^2 < S_x^2 + S_y^2$, where S_d is the sample standard deviation of d_i.

(C) Since $S_d^2 = S_x^2 + S_y^2 - 2S_{xy}$, S_{xy} has to be given so that t_d in (B) can be computed, where S_{xy} is the sample covariance of X and Y.

(D) True for the above (B) and (C).

(E) None for the above (A),(B), (C) and (D).

13. (5 points) Let $X_1, X_2, \ldots, X_n$, n≥4, be i.i.d. sample from some population with finite variance σ^2. Which of the following estimators is unbiased for σ^2 and has the smallest variance? ($\bar{X} = \sum_{i=1}^{n} X_i / n$, $\bar{X}_1 = \frac{\sum_{i=1}^{n_1} X_i}{n_1}$, $\bar{X}_2 = \frac{\sum_{i=n_1+1}^{n} X_i}{n_2}$, $n_1 + n_2 = n$; $n_1 \geq 2$, $n_2 \geq 2$)

(A) $X_1^2 - X_2X_3$

(B) $S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$

(C) $(X_1 - X_2)^2/2$

(D) $\hat{\sigma}^2=\sum_{i=1}^{n}(X_i-\bar{X})^2/n$

(E) $[\sum_{i=1}^{n_1}(X_i-\bar{X}_1)^2+\sum_{i=n_1+1}^{n}(X_i-\bar{X}_2)^2]/(n_1+n_2\text{-}2)$.

14. (4 points) A sample with size n=27 is obtained with $\hat{y}$=1.2-0.8x, SSE (sum of squares due to error)=150, SSR (sum of squares due to regression)=24. Then
   (A) R^2(coefficient of determination) = 0.863,
   (B) Correlation coefficient of Y_i and the predicted value $\hat{Y}_i$= 0.37,
   (C) Correlation coefficient of Y_i and X_i is also 0.37,
   (D) The statistics t(25) and F(1,25) can be applied to test $H_0: \beta_1 = 0$, but the conclusion would be different.
   (E) None for all the above (A),(B), (C) and (D)..
15. (4 points) Let $X_1, X_2, \dots, X_n$ be independent, identically distributed Bernoulli random variables with probability of success E(X) = p.
   (A) If $Y=\sum_{i=1}^{n} X_i$, then Y follows a binomial distribution with mean np and variance np(1-p).
   (B) (Continued) If sample size n large, but p small, n×p constant, then Y approximates to a Poisson distribution with mean np and variance np
   (C) If np is not small, say np ≥10, then the distribution can be,approximated by normal distribution with mean np and variance np(1-p).
   (D) For binomial, if p is not too extreme, say 0.2≤p≤0.9, then the probability distribution can be approximated by normal distribution with mean np and variance np(1-p).
   (E) True for all the above (A),(B), (C) and (D)..
16. (4 points) Let $X_1, X_2, \dots, X_n$, n=30, be a random sample from an uniform distribution

$$f(x;\theta) = 1/\theta,\ 0 < x < \theta.$$

Then (choose the most appropriate one)

(A) E(X) = θ,

(B) Var(X) = θ2/3,

(C) $\bar{X}$ approximately follows N(θ, θ2/(3n)),

(D) ($c\bar{X}$, ∞), c > 0, can be a lower confidence bound for suitable 100(1-α)% confidence level for θ.

(E) $(-\infty,\ c\bar{X})$, c>0, is a suitable upper confidence bound for some suitable 100(1-α)% confidence level for θ.

17. (4 points) In a survey sampling, what is the smallest sample size *n* required if the margin of error (suppose α is set to be 0.05), in estimating the population proportion p, is set to be less than 0.03?

(A) 1068 (B) 1000 (C) 1025 (D) 996 (E) None for all the above four.

18. (4 points) It is known that $X_1, X_2, \ldots, X_n$ is a random sample from $N(\mu, \sigma^2)$. The sample mean $\bar{X}$ and sample variance $S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$ is found to be 4 and 2.25, respectively. Suppose that n is large enough, then

(A) $P(4-1.645\times\frac{1.5}{\sqrt{n}} < \mu < 4+1.645\times\frac{1.5}{\sqrt{n}}) = 0.9$

(B) $P(4-1.96\times\frac{1.5}{\sqrt{n}} < \mu < 4+1.96\times\frac{1.5}{\sqrt{n}}) = 0.95$

(C) $P(4-2.33\times\frac{1.5}{\sqrt{n}} < \mu < 4+2.33\times\frac{1.5}{\sqrt{n}}) = 0.98$

(D) All the above (A), (B) and (C) are true.

(E) All the above (A), (B), (C) and (D) are wrong.

19. (4 points) Consider a normal random variable X with μ = 0 and standard deviation σ = 1. Which of the following is true?

(A) $P(X > 1.645) = 0.1$

(B) $P(X < -1.96) = 0.05$

(C) $P(X < 3) > 1 - P(X > -3)$

(D) $P(X < 0.5) = P(X > -0.5)$

(E) $P(X = 0) \neq P(X = 1)$.

20. (5 points) Random variable X follows exponential distribution with density

$$f(x;\lambda) = \lambda e^{-\lambda x}, x > 0, \lambda > 0$$

(A) $P(X > x_0) = 1 - e^{-\lambda x_0}$, some positive value x_0.

(B) The exponential random variable X has the property

$$P(X > x_0 + \Delta | X > x_0) = P(X > \Delta), \forall \Delta > 0,$$

科目名稱	統計學	類組代碼	D38
		科目碼	D3801

※本項考試依簡章規定各考科均「不可以」使用計算機　　本科試題共計 9 頁

i.e. a used one likes a new one.

(C) The skewness of X, like normal distribution, can be zero, positive, or negative.

(D) $E(X) = \lambda, and\ Var(X) = \lambda$.

(E) None for the above four.

21. (5 points) Assume that $X_1, X_2, \ldots, X_{n_1}$ is a random sample from some population with mean μ_1, variance σ^2; $Y_1, Y_2, \ldots, Y_{n_2}$ is another sample from population with mean μ_2, variance σ^2. We are interested in estimating the difference of the two population means.

(A) One point estimator of $\mu_1 - \mu_2$ is the difference of the sample means $\overline{X} - \overline{Y}$;

(B) We had better apply $S_p^2 = \frac{(n_1-1)S_X^2+(n_2-1)S_Y^2}{n_1+n_2-2}$ to estimate σ^2, where S_X^2 and S_Y^2 are sample variances for X-sample and Y-sample.

(C) The standard error of $\overline{X} - \overline{Y}$ is $S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$, where S_P is the root of S_p^2.

(D) Suppose both n_1 and n_2 are large, a 95% confidence interval for $\mu_1 - \mu_2$ is, approximately, $(\overline{X} - \overline{Y} - 1.96 \times S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}, \overline{X} - \overline{Y} + 1.96 \times S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}})$.

(E) True for all the above four.

22. (4 points) In testing the hypothesis $H_0: \mu = \mu_0$ vs $H_a: \mu > \mu_0$, where μ_0 is some known value.

(A) As the sample size n gets larger, the sample mean $\bar{x}$ will be closer to μ, so p-value is getting smaller.

(B) As the sample size n gets larger , then the probability of rejecting H_0 is larger because the p-value is tending to be smaller than α, the significant level.

(C) Two group of persons are collected to test $H_0: \mu = \mu_0$, one obtained $\overline{x_1} - \mu_0 = 10$, the other got $\overline{x_2} - \mu_0 = 5$. If the one with $\overline{x_2} - \mu_0 = 5$ is found to be significant, one with $\overline{x_1} - \mu_0 = 10$ would be more significant.

(D) True for all the above three.

(E) False for the above four.

23. (5 points) The following data are collected to examine the existence of treatment effect:

Treatment		
1	2	3
20	22	40
30	26	30
25	20	28
33	28	22

(A) The mean square due to treatments (MSTR) equals to 36.
(B) The mean square due to error (MSE) equals to 34.
(C) The test statistic to test the null hypothesis equals to1.06.
(D) The null hypothesis is to be tested at the 1% level of significance. Then *p*-value is greater than 0.1.
(E) True for all the above four.

$F_{0.975}(9,15) = 0.265$, $F_{0.95}(9,15) = 0.327$, $F_{0.05}(9,15) = 2.59$, $F_{0.025}(9,15) = 3.12$.