國立成功大學

114學年度碩士班招生考試試題

編　號：167

系　所：統計學系

科　目：統計學

日　期：0211

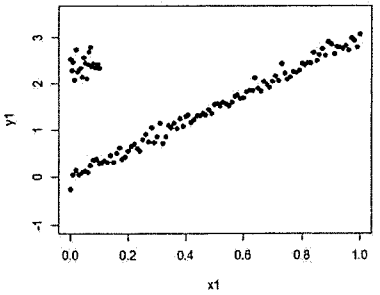
節　次：第 3 節

注　意：1.不可使用計算機
2.請於答案卷(卡)作答，於試題上作答，不予計分。

A. Multiple Choice (3.5%×20=70%)

1. Decide which of the following statements is false (a) There are no limited number of normal distributions (b) The inflection points for any normal distribution are one standard deviation on either side of the mean (c) The x-axis is a horizontal asymptote for all normal distributions (d) The line of symmetry for all normal distributions is $x = 0$
2. Assume that the population standard deviation is unknown and the population distribution is approximately normal. How will decreasing the level of confidence without changing the sample size affect the width of a confidence interval for a population mean? Decide which of the following statements is false (a) The margin of error will increase because the critical value will decrease. The increased margin of error will cause the confidence interval to be wider. (b) The margin of error will decrease because the critical value will decrease. The decreased margin of error will cause the confidence interval to be narrower. (c) The margin of error will decrease because the critical value will increase. The decreased margin of error will cause the confidence interval to be narrower. (d) The margin of error will increase because the critical value will increase. The increased margin of error will cause the confidence interval to be wider of error will cause the confidence interval to be wider.
3. A parent interest group is looking at whether birth order affects scores on the ACT test. It was suggested that, on average, first-born children earn lower ACT scores than second-born children. Let first-born children be Population 1 and let second-born children be Population 2. Is there sufficient evidence at the 10% significance level to say that the mean ACT score of first-born children is lower than that of second-born children? The null and alternative hypotheses for the test are $H_0: \mu_1 = \mu_2$ v.s. $H_a: \mu_1$ ___ μ_2. Fill in the blank: (a) > (b) $\neq$ (c) < (d) =
4. Refer to 4; after surveying a random sample of 100 first-born children, the parents' group found that they had a mean score of 24.3 on the ACT. A survey of 200 second-born children resulted in a mean ACT score of 24.7. Assume that the population variance for first-born children is 4 and the population variance for second-born children is 10. What is your decision for the hypothesis testing? (a) 0.01 < p-value < 0.025, reject H_0 (b) 0.025 < p-value < 0.05, reject H_0 (c) 0.05 < p-value < 0.1, reject H_0 (d) 0.1 < p-value < 0.2, do not reject H_0
5. A study was done on a new drug designed to lower blood pressure. There are individuals involved in the study who have high blood pressure but are not given the drug. Select the term that best describes the above scenario. (a) single-blind experiment (b) placebo effect (c) control group (d) treatment group
6. A research divided subjects into two groups according to gender and then selected members from each group. What sampling method was the researcher using? (a) Cluster (b) Stratified (c) Random (d) Systematic
7. In linear regression, a good estimator of the standard deviation of the error term is called the (a) standard deviation (b) variance (c) standard error of the estimate (d) sample standard deviation
8. If the two regression lines are given by $Y = a + bX$ and $X = c + dY$. What is the correlation coefficient between variables X and Y? (a) $\sqrt{bc}$ (b) $\sqrt{ac}$ (c) $\sqrt{ad}$ (d) $\sqrt{bd}$
9. When the level of significance of a hypothesis test is increased, the probability of committing a Type I error (a) is decreased (b) remains the same (c) is increased (d) cannot be determined

10. In a large city, 50% of the people choose a movie, 30% choose a play, and 20% choose shopping as a leisure activity. If a sample of 5 people is randomly selected, find the probability that 3 are planning to go to a movie, 1 to a play, and 1 to the shopping mall. (a) 0.04 (b) 0.15 (c) 0.01 (d) 0.30
11. A campaign team plans to conduct a poll to estimate the percentage of Americans who would vote for a presidential candidate in the USA. What sample size is needed to ensure the estimate is accurate within 5 percentage points at a 90% confidence level? (a) 1068 (b) 423 (c) 271 (d) None of the above
12. What is true? (I) Random scatter in the residuals indicates a linear model (II) If two variables are very strongly associated, then the correlation between them will be near +1 or -1 (III) Changing the units of measurement changes the correlation coefficient (a) I only (b) II only (c) I and II only (d) II and III only.
13. In a multiple regression analysis, if the model provides a poor fit, this indicates that: (a) the sum of squares for error will be large. (b) the standard error of estimate will be large. (c) the coefficient of determination will be close to zero. (d) All of these choices are true.
14. ANOVA was used to test the outcomes of three drug treatments. Each drug was given to 4 individuals. The total sum of the squares is 244, and the mean squares between treatments was 50. What is the standard deviation for all 60 individuals sampled for this study? (a) 48 (b) 16 (c) 4 (d) None of the above
15. Dimension reduction methods have the goal of using the correlation structure among the predictor variables to accomplish which of the following: (I) To reduce the number of predictor components. (II) To help ensure that these components are dependent. (III) To provide a framework for interpretability of the results. (IV) To help ensure that these components are independent. (V) To increase the number of predictor components. Choose the correct answer from the options given below: (a) I, II, IV and V only (b) I, III and IV only (c) I, II, III and V only (d) II, III, IV and V only
16. The data in the right scatterplot would have a correlation coefficient that is close to: (a) 0.9 (b) 0.5 (c) 0 (d) -0.5



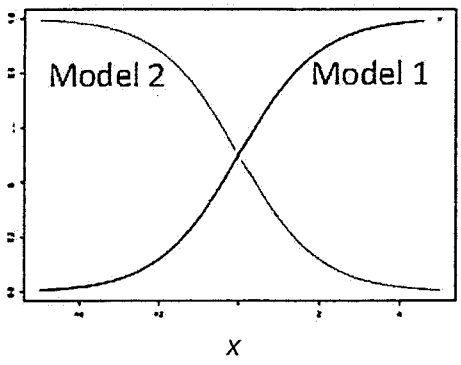
17. Let's build a regression model with mpg (miles per gallon) as the outcome and wt (weight) as the predictor for the mtcars dataset. Which value from the following output doesn't make sense?

```
## lm(formula = mpg ~ wt, data = mtcars)
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.2851     1.8776  19.858  < 2e -16 ***
## wt           -5.3445     0.5591  -9.559  1.29 e-10 ***
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 3.046 on 30 degrees of freedom
## Multiple R-squared: 3.598, Adjusted R-squared: 3.385
## F-statistic: 91.38 on 1 and 30 DF, p-value: 1.294e-10
```

(a) The coefficient of wt (b) The p-value of wt (c) The Multiple R-squared (d) The minimum value of residuals

18. Refer to 17; The sample size of the mtcars dataset dataset is (a) 30 (b) 31 (c) 32 (d) None of the above

19. Consider two simple logistic regression models based on the logit link. The logit link is defined as $logit(p(x)) = \beta_0 + \beta_1 x$ where $p(x)$ is the probability corresponding x. The predicted models are shown in the figure on the right. Which statement is true? (a) β_1 for model 2 is lower than model 1 (b) β_1 for model 2 is higher than model 1 (c) β_1 for both models are the same (d) None of the above



20. Adding additional independent variables in a regression equation will (a) increase the value of the coefficient of multiple determination (b) improve the predictive power of the equation (c) reduce the predictive power of the equation (d) Additional independent variables may not be added to a regression equation

Problems: (30%)

1. Given a dataset $\{X_i, Y_i\}_{i=1}^n$, where the means, standard deviations, and correlation coefficient are $\bar{X} = 1.0, \bar{Y} = 2.0, S_X = 3.0, S_Y = 9.0, r = 0.8$. Determine the regression line of Y on X. (7%)

2. Consider the linear model for one-way ANOVA

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, i = 1, \dots, k, j = 1, \dots, n.$$

Show that if a linear function of the treatment parameters $\sum_{i=1}^k c_i \alpha_i$ is estimable, then $\sum_{i=0}^k c_i = 0$. (8%)

3. A researcher designed an experiment to investigate the effects of three different dietary plans (A, B, C) and two levels of exercise intensity (low and high) on weight loss. Assume that weight loss follows a normal distribution and exhibits homogeneity of variance. The experiment involved random assignment of samples to each combination (with equal sample sizes for each group, total sample size $N = 60$). The researcher conducted a Two-Way ANOVA, and part of the results is as follows:

Source	Degree of Freedom	Sum of Square	Mean Square	F-Statistic
Dietary Plan			200	
Exercise Intensity			180	
Diet × Exercise				
Error		540		
Total		1420		

The researcher further conducted a detailed analysis of the interaction effect and designed a linear contrast to test the interaction between dietary plans and exercise intensity:

$$L = \frac{1}{3}(A, low\ intensity) + \frac{1}{3}(B, low\ intensity) + \frac{1}{3}(C, low\ intensity) - (A, high\ intensity)$$

which aims to test whether there is a significant linear trend difference between low-intensity and high-intensity levels for the dietary plans. Suppose that the contrast Sum of Squares is calculated as $SS_L = 35$. The significance level is set at $\alpha = 0.1$.

(a) Is the interaction between dietary plans and exercise intensity significant? (5%)

(b) Test the significance of the linear contrast L. (10%)

附表：Cumulative Probabilities for the Standard Normal Distribution

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

參考值：

Let $\Pr\{X < t_\alpha(df)\} = \alpha$, where X be a random variable following a t distribution with df degrees of freedom.

$t_{0.01}(53) = -2.3988,\ \ t_{0.05}(53) = -1.6741,\ \ t_{0.1}(53) = -1.2977$

$t_{0.01}(54) = -2.3974,\ \ t_{0.05}(54) = -1.6736,\ \ t_{0.1}(54) = -2.2974$

Let $\Pr\{F > F_\alpha(df_1, df_2)\} = \alpha$, where Y be a random variable following a F distribution with degrees of freedom df_1 and df_2.

$F_{0.1}(df_1 = 2, df_2 = 53) = 2.4056,\ \ F_{0.05}(df_1 = 2, df_2 = 53) = 3.1716$

$F_{0.1}(df_1 = 2, df_2 = 54) = 2.4036,\ \ F_{0.05}(df_1 = 2, df_2 = 54) = 3.1682$