國立臺北大學107學年度碩士班一般入學考試試題

系（所）組別：都市計劃研究所甲、乙組

科　　目：統計學

☑可 ☐不可使用計算機

Part I (100%) 單選題

1. Suppose the city government would like to build a new metro route. Suppose there are three routes, A, B and C. They would like to know which route is the most popular route. 1000 residents are randomly selected to answer their choice. What is the most appropriate chart to summary the result for 1000 residents?
   (A) Bar chart
   (B) Box plot
   (C) Histogram
   (D) QQ plot
   (E) Scatter plot

2. Suppose you would like to know whether the household income is related to the price of their property. Suppose 100 households are randomly selected to collect data. Which of the following is the most appropriate chart to visualize the relationship between the household income and the price?
   (A) Bar chart
   (B) Box plot
   (C) Histogram
   (D) Scatter plot
   (E) Stem and leaf plot

3. Which of the following is the least appropriate descriptive statistic to summarize the household income?
   (A) First quartile
   (B) Frequency
   (C) Mean
   (D) Median
   (E) Mode

4. Suppose we would like to know the variability of the price of the property for a given sample? Which is the most appropriate measure?
   (A) Coefficient of variation
   (B) Frequency
   (C) Interquartile
   (D) Maximum
   (E) Median

5. Suppose we would like to know the distribution of the price of the properties for a given sample. Which of the following chart is not appropriate?
   (A) Bar chart
   (B) Box plot
   (C) Histogram
   (D) QQ plot
   (E) Stem and leaf plot

6. Which of the following is a true statement? (Assume that the first quartile and minimum for this dataset are not equal).
   (A) IQR is always larger than the range for a given data set.
   (B) IQR is always smaller than the range for a given data set.
   (C) IQR/2 is always equal to the median for a given data set.
   (D) Standard deviation is always larger than the range for a given data set.
   (E) Standard deviation is always smaller than the range for a given data set.

試題隨卷繳交　　接背面

7. Suppose the price of properties per square feet follows a normal distribution with mean $3000 and standard deviation $100. What is the probability that the price of a property is larger than 3200?
   (A) 0.0228
   (B) 0.0455
   (C) 0.5
   (D) 0.9545
   (E) 0.9772

8. Suppose you know the preference of residents regarding the choice of the metro routes. Suppose there are two choices, route A and route B. The probability of a resident selecting A equals 0.2. A random sample of size 30 is collected. On the average, how many residents select route B?
   (A) 0
   (B) 6
   (C) 15
   (D) 24
   (E) 30

9. The central limit theorem is an important theorem in statistic. Which of the following is not a necessary condition for the theorem? Let $X_1, X_2, \cdots, X_n$ denote the sample.
   (A) $X_1, X_2, \cdots, X_n$ are from the same normal distribution.
   (B) $X_1, X_2, \cdots, X_n$ are independent.
   (C) $X_1, X_2, \cdots, X_n$ have the finite variance.
   (D) $X_1, X_2, \cdots, X_n$ have the same mean.
   (E) $X_1, X_2, \cdots, X_n$ have the same variance.

10. Suppose you know the preference of residents regarding the choice of the metro routes. Assume there are two choices, route A and route B. Suppose the probability of selecting route A equals 0.2. Given a resident who selecting route A, the probability that this resident is older than or equal to 65 years old equals 0.4, and that is younger than 30 years old equals 0.35. Given a resident who selecting route B, the probability that this resident is older than and equal to 65 years old equals 0.3, and that is younger than 30 years old equals 0.3. Given the selecting resident is older than or equal to 65 years old, what is the probability that he/she would select route A?
    (A) 0.2
    (B) 0.25
    (C) 0.4
    (D) 0.7
    (E) 0.75

11. Suppose a geologist random selects 100 mountain areas and determines whether the selected area have occurred landslides. From the literatures, the geologist knows that the probability of landslide p should be 0.05. Out of 100 areas, there are 5 landslides. The 95% confidence interval is (0.0073, 0.0927). Which of the following statement is incorrect?
    (A) Among repeated construction, 95% confidence of all sample would produce confidence intervals that enclose 0.05.
    (B) An estimate of p is the sample proportion.
    (C) The sample proportion is an unbiased estimator of p.
    (D) We have 95% confidence that the sample proportion of the landslide is smaller than 0.05.
    (E) We have 95% confidence that the proportion of the landslide is 0.05.

12. Suppose a geologist random selects 50 mountain areas and 50 riversides that have occurred landslides and measures the range of landslides. Assume that the range of landslides of mountain areas and riversides are independent and follow normal distributions with means μ_1 and μ_2 and equal standard deviation 8 m^2. The sample mean of the range of landslides of mountain areas and riversides are 55 and 45. The sample standard deviations are 9 m^2 and 10 m^2. What is the 95% confidence interval of μ_1-μ_2?
    (A) (4.733,15.267)
    (B) (5.565,14.435)
    (C) (6.271,13.729)
    (D) (6.864,13.136)
    (E) (9.255,10.745)

13. Suppose a geologist random selects 50 mountain areas and 50 riversides that have occurred landslides and measures the range of landslides. Assume that the range of landslides of mountain areas and riversides are independent and follow a normal distribution with mean μ_1 and μ_2. Based on the data, we find that the 95% confidence interval of μ_1-μ_2 equals (3.756, 6.467)?
    (A) There is sufficient evidence that mountain areas have a larger mean range of landslides than that of riversides.
    (B) There is sufficient evidence that mountain areas have a smaller mean range of landslides than that of riversides.
    (C) There is sufficient evidence that mountain areas have a larger sample mean ranges of landslides than that of riversides.
    (D) There is sufficient evidence that mountain areas have a smaller sample mean ranges of landslides than that of riversides.
    (E) There is no evidence that there is a difference in the mean range of landslides.

14. Suppose a geologist random selects 50 mountain areas and 50 riversides that have occurred landslides. The 100 random samples then classify into Type A and Type B of localities (地質種類). Suppose this geologist would like to know whether mountain areas and riversides have similar localities. What method should be used to analyze this data?
    (A) Chi-square test
    (B) One sample T test
    (C) Paired T test
    (D) Two independent sample proportions test
    (E) Two independent samples T test

15. Suppose a geologist random selects 50 mountain areas and 50 riversides that have occurred landslides and measure the range of landslides of each area. Assume that the range of landslides of mountain areas and riversides are independent and follow normal distributions with mean μ_1 and μ_2, respectively. Let the sample mean of the range of landslides of mountain areas and riversides be denoted as $\bar{x}_1$ and $\bar{x}_2$. The geologist would like to test whether the mean range of landslides from the mountain area and riverside is the same. What are the appropriate hypotheses H_0 and H_a?
    (A) $H_0: \bar{x}_1 = \bar{x}_2$ versus $H_a: \bar{x}_1 \neq \bar{x}_2$
    (B) $H_0: \bar{x}_1 \geq \bar{x}_2$ versus $H_a: \bar{x}_1 < \bar{x}_2$
    (C) $H_0: \bar{x}_1 \leq \bar{x}_2$ versus $H_a: \bar{x}_1 > \bar{x}_2$
    (D) $H_0: \mu_1 \leq \mu_2$ versus $H_a: \mu_1 > \mu_2$
    (E) $H_0: \mu_1 = \mu_2$ versus $H_a: \mu_1 \neq \mu_2$

16. Continued 15., what method should be used to analyze this data?
    (A) Chi-square test
    (B) One sample T test
    (C) Paired T test
    (D) Two independent sample proportions test
    (E) Two independent samples T test

17. Continued 16., suppose we have collected the data and use an appropriate test to find a p-value of 0.009. Which of the following is our conclusion at a given level $\alpha = 0.05$?
    (A) We do not have sufficient evidence to show that the mean range of landslide for the mountain area and riverside is the same.
    (B) We do not have sufficient evidence to show that the sample mean range of landslide for the mountain area and riverside is the same.
    (C) We have sufficient evidence to show that the mean ranges of landslide for the mountain area and riverside are not the same.
    (D) We have sufficient evidence to show that the sample mean range of landslide for the mountain area and riverside is not the same.
    (E) We cannot determine without more information.

18. Suppose a geologist random selects 50 areas of the range of landslides for 3 different types (A, B, C) of localities (地質種類). Assume that the range of landslides for 3 different types are independent and follow normal distributions with mean μ_1, μ_2 and μ_3, respectively. What method should be used to analyze this data?
    (A) Analysis of variance
    (B) Chi-square test
    (C) Paired T test
    (D) Two independent samples T test
    (E) Wilcoxon test

19. Suppose you conduct a significance test for the population proportion and your p-value is 0.09. Given a 0.10 level of significance, which of the following should be your conclusion?
    (A) Accept H_0
    (B) Accept H_a
    (C) Fail to reject H_0
    (D) Fail to reject H_a
    (E) Reject H_0

20. Global warming has gotten a lot of media attention over the past few years, but researchers in Antarctica are keeping track of temperatures at the South Pole to see if temperatures really are rising. They know that from 1900-1999 (last century) the average temperature at the South Pole was -6 degrees Celsius. They have randomly sampled 25 days from the past 2 years (this century). The sample mean for this century is -2.1. Let μ_1 and μ_2 denote the average temperature of South Pole for last and this century. What are the appropriate H_0 and H_a?
    (A) $H_0: \mu_2 \leq -6$ verus $H_0: \mu_2 > -6$
    (B) $H_0: \mu_2 \leq -2.1$ verus $H_0: \mu_2 > -2.1$
    (C) $H_0: \mu_2 = -6$ verus $H_0: \mu_2 \neq -6$
    (D) $H_0: \mu_2 = -2.1$ verus $H_0: \mu_2 \neq -2.1$
    (E) None of the above

21. Continued 20., the 98% confidence interval obtained from this century is (-5.65, 1.75). Based on this interval, which of the following statements are true? We are 98% confident that:
    (A) The true mean temperature of this century is between 5.65 degrees lower and 1.75 degrees higher than the true mean temperature of last century at the South Pole.
    (B) The true mean temperature of this century is between 5.65 degrees higher and 1.75 degrees lower than the true mean temperature of last century at the South Pole.
    (C) The true mean temperature of this century is different than the true mean temperature of last century at the South Pole.
    (D) The true mean temperature of this century is the same as the true mean temperature of last century at the South Pole.
    (E) None of the above statements are true.

22. Suppose you would like to know whether the household income is related to the price of their property. Suppose 100 households are randomly selected to collect data. Which of the following is the most appropriate method to evaluate the relationship between the household income and the price?
    (A) Chi-square test
    (B) One-way analysis of variance
    (C) Simple linear regression
    (D) Two independent sample T test
    (E) Wilcoxon test.

23. When using ANOVA, which assumption is not necessary?
    (A) The residual is from the same normal distribution.
    (B) The sample data is collected independently.
    (C) The sample data is from the same normal distribution.
    (D) The sample data from each population have the same variance.
    (E) The sample data have to be collected randomly.

24. Suppose the city government would like to build a new metro route. They would like to know whether the local resident agrees such a policy. They would like to use the survey to understand the resident's opinion. Let the probability of approving such a policy be denoted as p. What is the least sample size that the government needs to have the marginal error of E=0.045 with 95% confidence?
    (A) 335
    (B) 475
    (C) 669
    (D) 949
    (E) None of the above

25. Regarding the test statistic for testing the null hypothesis (H_0) versus the alternative (H_a), which of the following is incorrect?
    (A) Power is related to the probability of Type II error.
    (B) Power is defined as the probability that the test rejects H_0 given H_a is true.
    (C) The test that can provide the smallest Type I and Type II errors simultaneously is the best.
    (D) Type I error is defined as the the test statistic rejects H_0 given H_0 is true.
    (E) Type II error is defined as the the test statistic does not reject H_0 given H_a is true.

國立臺北大學107學年度碩士班一般入學考試試題

系（所）組別：都市計劃研究所甲、乙組

科　　目：統計學

☑可 ☐不可使用計算機

Table 1: Area under the standard normal curve

	$P[Z \le z]$									
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.500	0.504	0.508	0.512	0.516	0.519	0.523	0.527	0.531	0.535
0.10	0.539	0.543	0.547	0.551	0.555	0.559	0.563	0.567	0.571	0.575
0.20	0.579	0.583	0.587	0.591	0.594	0.598	0.602	0.606	0.610	0.614
0.30	0.617	0.621	0.625	0.629	0.633	0.636	0.640	0.644	0.648	0.651
0.40	0.655	0.659	0.662	0.666	0.670	0.673	0.677	0.680	0.684	0.687
0.50	0.691	0.695	0.698	0.701	0.705	0.708	0.712	0.715	0.719	0.722
0.60	0.725	0.729	0.732	0.735	0.738	0.742	0.745	0.748	0.751	0.754
0.70	0.758	0.761	0.764	0.767	0.770	0.773	0.776	0.779	0.782	0.785
0.80	0.788	0.791	0.793	0.796	0.799	0.802	0.805	0.807	0.810	0.813
0.90	0.815	0.818	0.821	0.823	0.826	0.828	0.831	0.834	0.836	0.838
1.00	0.841	0.843	0.846	0.848	0.850	0.853	0.855	0.857	0.859	0.862
1.10	0.864	0.866	0.868	0.870	0.872	0.874	0.877	0.879	0.881	0.883
1.20	0.884	0.886	0.888	0.890	0.892	0.894	0.896	0.898	0.899	0.901
1.30	0.903	0.904	0.906	0.908	0.909	0.911	0.913	0.914	0.916	0.917
1.40	0.919	0.920	0.922	0.923	0.925	0.926	0.927	0.929	0.930	0.931
1.50	0.933	0.934	0.935	0.937	0.938	0.939	0.940	0.941	0.942	0.944
1.60	0.945	0.946	0.947	0.948	0.949	0.950	0.951	0.952	0.953	0.954
1.70	0.955	0.956	0.957	0.958	0.959	0.959	0.960	0.961	0.962	0.963
1.80	0.964	0.964	0.965	0.966	0.967	0.967	0.968	0.969	0.969	0.970
1.90	0.971	0.971	0.972	0.973	0.973	0.974	0.975	0.975	0.976	0.976
2.00	0.977	0.977	0.978	0.978	0.979	0.979	0.980	0.980	0.981	0.981
2.10	0.982	0.982	0.983	0.983	0.983	0.984	0.984	0.985	0.985	0.985
2.20	0.986	0.986	0.986	0.987	0.987	0.987	0.988	0.988	0.988	0.989
2.30	0.989	0.989	0.989	0.990	0.990	0.990	0.990	0.991	0.991	0.991
2.40	0.991	0.992	0.992	0.992	0.992	0.992	0.993	0.993	0.993	0.993
2.50	0.993	0.994	0.994	0.994	0.994	0.994	0.994	0.994	0.995	0.995
2.60	0.995	0.995	0.995	0.995	0.995	0.996	0.996	0.996	0.996	0.996
2.70	0.996	0.996	0.996	0.996	0.996	0.997	0.997	0.997	0.997	0.997
2.80	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.998	0.998
2.90	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998
3.00	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.999	0.999

接下頁

國立臺北大學107學年度碩士班一般入學考試試題

系（所）組別：都市計劃研究所甲、乙組

科　　目：統計學

☑可 ☐不可使用計算機

Table 2 Selected quantiles of the *t* distribution

	$P[T \le t]$				
r	0.900	0.950	0.975	0.99	0.995
1	3.08	6.31	12.71	31.82	63.66
2	1.89	2.92	4.30	6.96	9.92
3	1.64	2.35	3.18	4.54	5.84
4	1.53	2.13	2.78	3.75	4.60
5	1.48	2.02	2.57	3.36	4.03
6	1.44	1.94	2.45	3.14	3.71
7	1.41	1.89	2.36	3.00	3.50
8	1.40	1.86	2.31	2.90	3.36
9	1.38	1.83	2.26	2.82	3.25
10	1.37	1.81	2.23	2.76	3.17
11	1.36	1.80	2.20	2.72	3.11
12	1.36	1.78	2.18	2.68	3.05
13	1.35	1.77	2.16	2.65	3.01
14	1.35	1.76	2.14	2.62	2.98
15	1.34	1.75	2.13	2.60	2.95
16	1.34	1.75	2.12	2.58	2.92
17	1.33	1.74	2.11	2.57	2.90
18	1.33	1.73	2.10	2.55	2.88
19	1.33	1.73	2.09	2.54	2.86
20	1.33	1.72	2.09	2.53	2.85
21	1.32	1.72	2.08	2.52	2.83
22	1.32	1.72	2.07	2.51	2.82
23	1.32	1.71	2.07	2.50	2.81
24	1.32	1.71	2.06	2.49	2.80
25	1.32	1.71	2.06	2.49	2.79
26	1.31	1.71	2.06	2.48	2.78
27	1.31	1.70	2.05	2.47	2.77
28	1.31	1.70	2.05	2.47	2.76
29	1.31	1.70	2.05	2.46	2.76
30	1.31	1.70	2.04	2.46	2.75

試題隨卷繳交