UNIVERSITY OF DUBLIN

**TRINITY COLLEGE**

**Faculty of Engineering, Mathematics and Science**

**School of Computer Science and Statistics**

**JF MSISS**

**JF Business and Computing**

**SF Engineering and Management**

**JS Computer Science Trinity Term 2012**

**ST1002 Statistical Analysis**

**????? ????? ?????**

**Dr. O’ Regan**

**Answer all questions. Useful formulae and statistical tables are in Appendix A. Nonprogrammable calculators may be used.**

1. a) Why is it important to know the type of data (categorical or quantitative)? **(5 marks)**

b) The diagram below shows a boxplot of affluence by those who buy organic vs those who not buy organic foods.



b) Describe the main features in the above plot. **(5 marks)**

c) What is the difference between the mean and median? **(5 marks)**

d) Compare the range and the standard deviation as measures of variability. . **(5 marks)**

e) The Diagonal length of the banknotes is normally distributed with a mean of 141 mm and a standard deviation of 2.5 mm,. What percentage of notes would you expect in the interval 137 to 144mm? **(5 marks)**

2. A company was interested in who purchased organic foods. They carried out a survey of 651 shoppers. The company was interesting to know if gender is related to whether a person purchased organic products. The table below shows a breakdown of gender and whether a customer purchased organic food and is taken from Minitab. The numbers in bold type are expected values.

**Tabulated statistics: Purchased Organic, Type of Loyalty Card**

**Tabulated statistics: Buy Organic, GENDER**

Rows: Buy Organic Columns: GENDER

F M All

No 296 172 468

***322.8 145.2* 468.0**

yes 153 30 183

*126.2 56.8* **183.0**

All 449 202 651

449.0 202.0 651.0

Cell Contents: Count

Expected count

Pearson Chi-Square = 25.479, DF = 1, P-Value = 0.000

Likelihood Ratio Chi-Square = 27.551, DF = 1, P-Value <= 0.001

1. Give a 95% confidence interval for the proportion of people who purchased organic food. Explain what it means. **( 8 marks)**
2. What are the null and alternate hypotheses underlying the Chi-Square result? **( 4 marks)**
3. What are the Expected values (italicised numbers) in the table?. **( 6 marks)**
4. Interpret the Chi-Square result and give an appropriate confidence interval. **( 7 marks)**

3) A company was interested in investigating the differences in two groups of banknotes( fraudulent and non- fraudulent). They selected a sample of a hundred of each type and measured the length of the diagonal of each note. Using Minitab they generated the following diagram:



1. Explain the above diagram. **( 5 marks)**

They generated the following Minitab output

**Two-Sample T-Test and CI: Diagonal, Status**

Two-sample T for Diagonal

Status N Mean StDev SE Mean

Not Fraud 100 141.517 0.447 0.045

Fraudulent 100 139.450 0.558 0.056

Difference = mu (0) - mu (1)

Estimate for difference: 2.0670

95% CI for difference: (1.9260, 2.2080)

T-Test of difference = 0 (vs not =): T-Value = 28.91 P-Value = 0.000 DF = 198

Both use Pooled StDev = 0.5055

1. What does SE mean measure? (5 marks)
2. .Explain the above output from Minitab in detail**(10 marks)**
3. What is a systematic sample? Describe how you would select a systematic sample of 35 people from a population of 1136 people. **(5 marks)**

4. A company is interested in investigating the relationship between yield(in grams) of a process and the temperature measured in degree centigrade. They collected a set of data for this purpose and produced the following output:



1. Describe the main features in the above plot. **( 3 marks)**

A colleague suggested fitting a simple linear regression to the above data obtaining the following output:

**Regression Analysis: Yield gms versus Temp C**

The regression equation is

Yield gms = 17.0 + 2.00 Temp C

Predictor Coef SE Coef T P

Constant 17.002 4.072 4.18 0.000

Temp C 1.99517 0.05334 37.41 0.000

S = 4.01967 R-Sq = 98.4% R-Sq(adj) = 98.3%

1. Explain the results of the above regression analysis. **(17 marks)**
2. Using the above results show how to predict the yield for a temperature of 75.. (**5 marks)**

**Appendix A**

t=

s2=

SE(diff)=

Chi-square

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | **Proportion of Area to left of point for Standardised Normal Distribution** | | | | | | | |
|  |  |  |  |  |  |  |  |  |  |  |
|  | **0.0** | **0.01** | **0.02** | **0.03** | **0.04** | **0.05** | **0.06** | **0.07** | **0.08** | **0.09** |
| **0.0** | 0.50 | 0.50 | 0.51 | 0.51 | 0.52 | 0.52 | 0.52 | 0.53 | 0.53 | 0.54 |
| **0.1** | 0.54 | 0.54 | 0.55 | 0.55 | 0.56 | 0.56 | 0.56 | 0.57 | 0.57 | 0.58 |
| **0.2** | 0.58 | 0.58 | 0.59 | 0.59 | 0.59 | 0.60 | 0.60 | 0.61 | 0.61 | 0.61 |
| **0.3** | 0.62 | 0.62 | 0.63 | 0.63 | 0.63 | 0.64 | 0.64 | 0.64 | 0.65 | 0.65 |
| **0.4** | 0.66 | 0.66 | 0.66 | 0.67 | 0.67 | 0.67 | 0.68 | 0.68 | 0.68 | 0.69 |
| **0.5** | 0.69 | 0.69 | 0.70 | 0.70 | 0.71 | 0.71 | 0.71 | 0.72 | 0.72 | 0.72 |
| **0.6** | 0.73 | 0.73 | 0.73 | 0.74 | 0.74 | 0.74 | 0.75 | 0.75 | 0.75 | 0.75 |
| **0.7** | 0.76 | 0.76 | 0.76 | 0.77 | 0.77 | 0.77 | 0.78 | 0.78 | 0.78 | 0.79 |
| **0.8** | 0.79 | 0.79 | 0.79 | 0.80 | 0.80 | 0.80 | 0.81 | 0.81 | 0.81 | 0.81 |
| **0.9** | 0.82 | 0.82 | 0.82 | 0.82 | 0.83 | 0.83 | 0.83 | 0.83 | 0.84 | 0.84 |
| **1.0** | 0.84 | 0.84 | 0.85 | 0.85 | 0.85 | 0.85 | 0.86 | 0.86 | 0.86 | 0.86 |
| **1.1** | 0.86 | 0.87 | 0.87 | 0.87 | 0.87 | 0.87 | 0.88 | 0.88 | 0.88 | 0.88 |
| **1.2** | 0.88 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.90 | 0.90 | 0.90 | 0.90 |
| **1.3** | 0.90 | 0.90 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.92 | 0.92 |
| **1.4** | 0.92 | 0.92 | 0.92 | 0.92 | 0.93 | 0.93 | 0.93 | 0.93 | 0.93 | 0.93 |
| **1.5** | 0.93 | 0.93 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 |
| **1.6** | 0.945 | 0.946 | 0.947 | 0.948 | 0.949 | 0.951 | 0.952 | 0.953 | 0.954 | 0.954 |
| **1.7** | 0.955 | 0.956 | 0.957 | 0.958 | 0.959 | 0.960 | 0.961 | 0.962 | 0.962 | 0.963 |
| **1.8** | 0.964 | 0.965 | 0.966 | 0.966 | 0.967 | 0.968 | 0.969 | 0.969 | 0.970 | 0.971 |
| **1.9** | 0.971 | 0.972 | 0.973 | 0.973 | 0.974 | 0.974 | 0.975 | 0.976 | 0.976 | 0.977 |
| **2.0** | 0.977 | 0.978 | 0.978 | 0.979 | 0.979 | 0.980 | 0.980 | 0.981 | 0.981 | 0.982 |
| **2.1** | 0.982 | 0.983 | 0.983 | 0.983 | 0.984 | 0.984 | 0.985 | 0.985 | 0.985 | 0.986 |
| **2.2** | 0.986 | 0.986 | 0.987 | 0.987 | 0.987 | 0.988 | 0.988 | 0.988 | 0.989 | 0.989 |
| **2.3** | 0.989 | 0.990 | 0.990 | 0.990 | 0.990 | 0.991 | 0.991 | 0.991 | 0.991 | 0.992 |
| **2.4** | 0.992 | 0.992 | 0.992 | 0.992 | 0.993 | 0.993 | 0.993 | 0.993 | 0.993 | 0.994 |
| **2.5** | 0.994 | 0.994 | 0.994 | 0.994 | 0.994 | 0.995 | 0.995 | 0.995 | 0.995 | 0.995 |
| **2.6** | 0.995 | 0.995 | 0.996 | 0.996 | 0.996 | 0.996 | 0.996 | 0.996 | 0.996 | 0.996 |
| **2.7** | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 |
| **2.8** | 0.997 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 |
| **2.9** | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.999 | 0.999 | 0.999 |
| **3.0** | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 |
|  |  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Critical values for 2-tailed tests for t-distributions for different ** | | | | | |
|  |  |  |  |  |  |  |
|  | **df** | **0.05** | **0.025** | **0.01** | **0.001** |  |
|  | 10 | 2.23 | 2.63 | 3.17 | 4.59 |  |
|  | 11 | 2.20 | 2.59 | 3.11 | 4.44 |  |
|  | 12 | 2.18 | 2.56 | 3.05 | 4.32 |  |
|  | 13 | 2.16 | 2.53 | 3.01 | 4.22 |  |
|  | 14 | 2.14 | 2.51 | 2.98 | 4.14 |  |
|  | 15 | 2.13 | 2.49 | 2.95 | 4.07 |  |
|  | 16 | 2.12 | 2.47 | 2.92 | 4.01 |  |
|  | 17 | 2.11 | 2.46 | 2.90 | 3.97 |  |
|  | 18 | 2.10 | 2.45 | 2.88 | 3.92 |  |
|  | 19 | 2.09 | 2.43 | 2.86 | 3.88 |  |
|  | 20 | 2.09 | 2.42 | 2.85 | 3.85 |  |
|  | 21 | 2.08 | 2.41 | 2.83 | 3.82 |  |
|  | 22 | 2.07 | 2.41 | 2.82 | 3.79 |  |
|  | 23 | 2.07 | 2.40 | 2.81 | 3.77 |  |
|  | 24 | 2.06 | 2.39 | 2.80 | 3.75 |  |
|  | 25 | 2.06 | 2.38 | 2.79 | 3.73 |  |
|  | 26 | 2.06 | 2.38 | 2.78 | 3.71 |  |
|  | 27 | 2.05 | 2.37 | 2.77 | 3.69 |  |
|  | 28 | 2.05 | 2.37 | 2.76 | 3.67 |  |
|  | 29 | 2.05 | 2.36 | 2.76 | 3.66 |  |
|  | 30 | 2.04 | 2.36 | 2.75 | 3.65 |  |
|  | 31 | 2.04 | 2.36 | 2.74 | 3.63 |  |
|  | 32 | 2.04 | 2.35 | 2.74 | 3.62 |  |
|  | 33 | 2.03 | 2.35 | 2.73 | 3.61 |  |
|  | 34 | 2.03 | 2.35 | 2.73 | 3.60 |  |
|  | 35 | 2.03 | 2.34 | 2.72 | 3.59 |  |
|  | 36 | 2.03 | 2.34 | 2.72 | 3.58 |  |
|  | 37 | 2.03 | 2.34 | 2.72 | 3.57 |  |
|  | 38 | 2.02 | 2.33 | 2.71 | 3.57 |  |
|  | 39 | 2.02 | 2.33 | 2.71 | 3.56 |  |
|  | 40 | 2.02 | 2.33 | 2.70 | 3.55 |  |
|  | 41 | 2.02 | 2.33 | 2.70 | 3.54 |  |
|  | 42 | 2.02 | 2.32 | 2.70 | 3.54 |  |
|  | 43 | 2.02 | 2.32 | 2.70 | 3.53 |  |
|  | 44 | 2.02 | 2.32 | 2.69 | 3.53 |  |
|  | 45 | 2.01 | 2.32 | 2.69 | 3.52 |  |
|  | 50 | 2.01 | 2.31 | 2.68 | 3.50 |  |
|  | 55 | 2.00 | 2.30 | 2.67 | 3.48 |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  | **Critical values for 2-tailed tests for t-distributions for different ** | | | | | |
|  |  |  |  |  |  |  |
|  | df | 0.05 | 0.025 | 0.01 | 0.001 |  |
|  | 60 | 2.00 | 2.30 | 2.66 | 3.46 |  |
|  | 80 | 1.99 | 2.28 | 2.64 | 3.42 |  |
|  | 100 | 1.98 | 2.28 | 2.63 | 3.39 |  |
|  | 120 | 1.98 | 2.27 | 2.62 | 3.37 |  |
|  | 140 | 1.98 | 2.27 | 2.61 | 3.36 |  |
|  | 160 | 1.97 | 2.26 | 2.61 | 3.35 |  |
|  | 180 | 1.97 | 2.26 | 2.60 | 3.35 |  |
|  | 200 | 1.97 | 2.26 | 2.60 | 3.34 |  |
|  | 220 | 1.97 | 2.26 | 2.60 | 3.34 |  |
|  | 240 | 1.97 | 2.26 | 2.60 | 3.33 |  |
|  | 260 | 1.97 | 2.25 | 2.59 | 3.33 |  |
|  | 280 | 1.97 | 2.25 | 2.59 | 3.33 |  |
|  | 300 | 1.97 | 2.25 | 2.59 | 3.32 |  |
|  | 320 | 1.97 | 2.25 | 2.59 | 3.32 |  |
|  | 340 | 1.97 | 2.25 | 2.59 | 3.32 |  |
|  | 360 | 1.97 | 2.25 | 2.59 | 3.32 |  |
|  | 380 | 1.97 | 2.25 | 2.59 | 3.32 |  |
|  |  | 1.96 | 2.24 | 2.58 | 3.29 |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Chi-Square table** | | |
|  | **α -significance level** | | |
|  |  |  |  |
| **df** | **0.05** | **0.025** | **0.001** |
| **1** | 3.84 | 5.02 | 10.83 |
| **2** | 5.99 | 7.38 | 13.82 |
| **3** | 7.81 | 9.35 | 16.27 |
| **4** | 9.49 | 11.14 | 18.47 |
| **5** | 11.07 | 12.83 | 20.52 |
| **6** | 12.59 | 14.45 | 22.46 |
| **7** | 14.07 | 16.01 | 24.32 |
| **8** | 15.51 | 17.53 | 26.12 |
| **9** | 16.92 | 19.02 | 27.88 |
| **10** | 18.31 | 20.48 | 29.59 |