Topic 4: Statistical applications

Topic number	Content
4.1	The normal distribution The concept of a random variable; of the parameters μ and σ;of the bell shape; the symmetry about $x = \mu$. Diagrammatic representation (using a sketch of a normal curve and shading the indicated area)'Normal probability calculations. Expected value Inverse normal calculations
4.2	Bivariate data (two variables being compared); the concept of correlation. Scatter diagrams; line of best fit, by eye, passing through the mean point Pearson's product-moment correlation coefficient, r Interpretation of positive, zero and negative, strong or weak correlations.
4.3	The regression line for y on x Use of the regression line for prediction purposes.
4.4	The χ^2 test for independence; formulation of null and alternative hypotheses; significance levels; contingency tables; expected frequencies; degrees of freedom; p-values

Brain Dump

Use your calculator to find the values for Chi-squared, p-value and degrees of freedom

To enter your data in a matrix: 2nd Matrix, go over to EDIT, choose 1:[A], enter the size of the matrix and the data

Then go to STAT key, go over to TESTS menu, go down to C: χ^2, Enter, then go down to CALCULATE and ENTER again

To see your expected values, go to 2nd Matrix, choose 2:[B] and Enter, the matrix that displays is the expected value matrix

By Chi-Squared

If your calculated $\chi^2 > critical\ value$, reject the H_o, conclude that the variables are dependent

If your calculated $\chi^2 < critical\ value$, accept the H_o, conclude that the variables are independent

By p-value: Compare your p-value to the significance level (it is a percent)

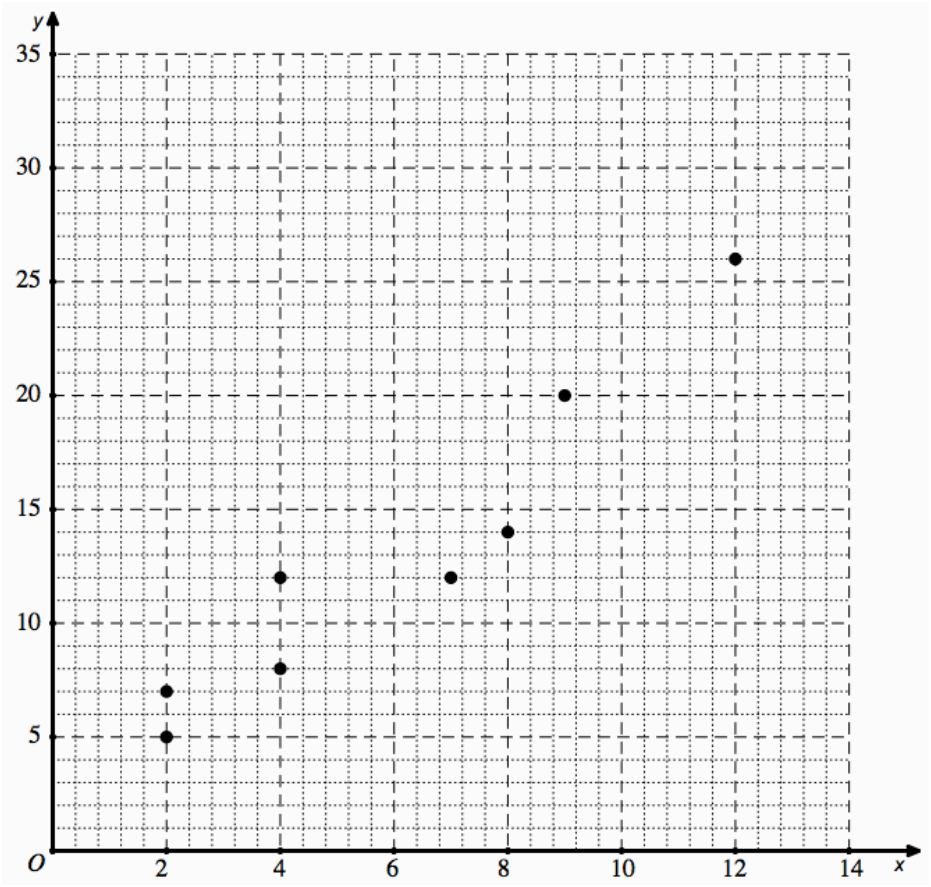
If $p < significance\ level$, reject the H_o, conclude that the variables are dependent

If $p > significance\ level$, , accept the H_o, conclude that the variables are independent

Line of best fit: to get the correlation coefficient (r) you must turn on the diagnostics- 2nd Zero, scroll down the catalog until you find DiagnosticON and push ENTER twice until it says DONE

1. Consider the following set of data which is plotted on the scatter diagram below.

x	2	4	7	12	4	8	9	2
y	5	8	12	26	12	14	20	7



a) Write down the coordinates of the mean point $(\overline{x},\ \overline{y})$ (2 marks)

b) Write down the value of r, the Pearson's product-moment correlation coefficient for this set of data. (2 marks)

c) Draw the regression line for y on x on the set of axes above. (2 marks)

2. Members of a certain club are required to register for one of three sports, badminton, volleyball or table tennis. The number of club members of each gender choosing each sport in a particular year is shown in the table below.

A χ^2 (Chi-squared) test at the 5% significance level is used to determine whether the choice of sport is independent of gender.

	Badminton	**Volleyball**	**Table tennis**
Male	40	20	10
Female	20	15	15

a) Find the expected number of female volleyball players under this hypothesis. (2 marks)

b) Write down the p-value for the test. (2 marks)

c) State, with a reason, the conclusion of the test. (2 marks)

3. A study was carried out to determine whether the country chosen by students for their university studies was influenced by a person's gender. A random sample was taken. The results are shown in the following table.

	Country Chosen		
	USA	**Australia**	**UK**
Male	55	26	40
Female	25	31	41

A χ^2 test was performed at the 1% significance level.
The critical value for this test is 9.210

a) State the null hypothesis. (1 mark)

b) Write down the number of degrees of freedom. (1 mark)

c) Write down (2 marks)

i. The χ^2 statistic;

ii. The associated p −value.

d) State, giving a reason, whether the null hypothesis should be accepted.
(2 marks)

4. A group of 100 students gave the following responses to the question of how they get to school.

	Walk	**Public Transport**	**Car**	**Bicycle**	**Total**
Female	18	13	14	3	48
Male	9	17	10	16	52
Total	27	30	24	19	100

A χ^2 test for independence was conducted at the 5% significance level. The null hypothesis was defined as

H_0: Method of getting to school is independent of gender.

a) Find the expected frequency for the females who use public transport to get to school. (2 marks)

b) Find the χ^2 statistic. (2 marks)

The χ^2 critical value is 7.815 at the 5% significance level.

c) State whether or not the null hypothesis is accepted. Give a reason for your answer. (2 marks)

5. A survey investigated the relationship between the number of cleaners, n, and the amount of time, t, it takes them to clean a school.

Number of Cleaners, n	**Time, t (minutes)**
1	193
2	172
3	118
5	112
6	87

a) Use your graphic display calculator to write down the equation of the regression line t on n. (2 marks)

b) Write down the value of the Pearson's product-moment correlation coefficient, r. (2 marks)

c) Use your regression equation to find the amount of time 4 cleaners take to clean the school. (2 marks)