UNIT 1: Regression and correlation

Return to overview

SPECIFICATION REFERENCES

2.2 Change of variable may be required e.g. using knowledge of logarithms to reduce a relationship of the form $y = ax^n$ or $y = kb^x$ into linear form to estimate a and n or k and b

5.1 Understand and apply the language of statistical hypothesis testing, …., extend to correlation coefficients as measures of how close data points lie to a straight line and be able to interpret a given correlation coefficient using a given p-value or critical value (calculation of correlation coefficients is excluded)

PRIOR KNOWLEDGE

AS Mathematics – Statistics content

2.2 Understanding of regression (See Unit 2b of the SoW)
Understanding of correlation (See Unit 2b of the SoW)

AS Mathematics – Statistics content

5.1 Use appropriate language of statistical hypothesis testing (See Unit 5.1 of the SoW)

5.2 Be able to apply a hypothesis test to the binomial distribution (See Unit 5.2 of the SoW)

AS Mathematics – Pure Mathematics content

6.3, 6.4 Knowledge of logarithms (See Unit 8 of the SoW)

KEYWORDS

Hypotheses, significance level, one-tailed test, two-tailed test, test statistic, null hypothesis, alternative hypothesis, critical value, critical region, acceptance region, p-value, binomial model, correlation coefficients, product moment correlation coefficient, population coefficient, sample, inference, mean, normal distribution, variance, assumed variance, linear regression, interpolation, extrapolation, coded data

1a. Change of variable (2.2)

Teaching time
2 hours

OBJECTIVES

By the end of the sub-unit, students should:

- be able to change the variable in a regression line;
- be able to estimate values from regression line.

TEACHING POINTS

Start the revision of topics from year one by recapping regression.

This needs to be extended to working with changing variables (coding) within regression lines. This relies on logarithms from the pure content and students should be able to work with equations of the form $y = ax^n$ and $y = kb^x$. Students will need to know how to put these into linear form and be able to estimate a and n or k and b. An understanding of reliability when extrapolating will also need to be recapped.

OPPORTUNITIES FOR PROBLEM SOLVING/MODELLING

Relate to real-world problems and discuss the reality of extrapolation.

1b. Correlation coefficients; Statistical hypothesis testing for correlation coefficients (5.1)

Teaching time
5 hours

OBJECTIVES

By the end of the sub-unit, students should:

- understand correlation coefficients;
- be able to calculate the PMCC (calculator only);
- be able to interpret a correlation coefficient;
- be able to conduct a hypothesis test for a correlation coefficient.

TEACHING POINTS

Recap scatter diagrams and the terminology used in year one to describe correlation. Students should understand that measures of correlation can be calculated to identify the strength of correlation. They need to understand that one of these, the product moment correlation coefficient (PMCC) is denoted by r, and that $|r| \leq 1$. If $r = \pm 1$ then the data points lie on a perfect straight line on a graph.

Students are expected to be able to calculate r using their calculators, but are not required to know or use the formula. They should be able to interpret their value for the PMCC in the context of the question.

Students are required to perform hypotheses tests for correlation coefficients. The hypotheses need to be stated in terms of ρ where ρ represents the population correlation coefficient. All tests should have the null hypothesis $\mathrm{H}_0: \rho = 0$. Tables of critical values or a p-value will be given to students.

OPPORTUNITIES FOR PROBLEM SOLVING/MODELLING

This is a good opportunity here to bring together the AS and A level content relating to regression and correlation.

COMMON MISCONCEPTIONS/EXAMINER REPORT QUOTES

Notation and stating a conclusion are the most common errors: 'some students failed to state their hypotheses in terms of ρ. Common errors include failing to ensure that critical values match the alternative hypothesis and giving conclusions that do not include a reference to the context.

NOTES

A small p-value (≤ 0.05) shows strong evidence against the null hypothesis, therefore reject the null hypothesis (at the 5% significance level).

A large p-value (> 0.05) shows weak evidence against the null hypothesis, therefore accept the null hypothesis (at the 5% significance level).