

Name	Date started	Target end date
------	--------------	-----------------

GCE AS / A LEVEL – FURTHER STATISTICS A QUESTION PACK

0985-01 (Legacy S3) · New spec Unit 2 Topic 4

REVISE

.wales

FURTHER MATHS – FS A · CORRELATION, REGRESSION & CHI-SQUARED

Linear Regression, Correlation & Chi-Squared Tests

Every linear-regression question from the legacy WJEC S3 papers (June 2006 – June 2017 + Specimen). Spearman / Pearson / chi-squared content was not assessed by S2/S3 so the pack focuses on the regression strand of new-spec Topic 4.

LEGACY 2008 SPECIFICATION

Estimated time for entire question pack: ~4 hours 58 minutes

Derived from the legacy S3 paper's pace of ~1.5 min/mark (199 marks over 13 questions).

You are advised to **not** attempt to complete all of this in one sitting.

ABOUT THIS QUESTION PACK

This is a **comprehensive practice question pack**, not a single mock paper. It contains every correlation, regression & chi-squared question from the legacy WJEC S3 papers (2008 modular spec) that maps onto new-spec AS Unit 2 Topic 4 (2.2.2).

Questions are ordered roughly by difficulty.

INSTRUCTIONS

Use black ink or black ball-point pen. Show all working – method marks are awarded for clear setup.

A calculator is allowed. The WJEC Formula Booklet and statistical tables (Murdoch & Barnes) may be referred to.

All question content is © WJEC CBAC Ltd. and reproduced for revision purposes.

For Examiner's use only

Q	Source	Max	Mark
1	Jun 06 Q5	20	
2	Jun 07 Q7	16	
3	Jun 08 Q7	12	
4	Jun 09 Q7	13	
5	Jun 10 Q7	13	
6	Jun 11 Q6	13	
7	Jun 12 Q5	13	
8	Jun 13 Q6	13	
9	Jun 14 Q5	18	
10	Jun 15 Q5	20	
11	Jun 16 Q5	15	
12	Jun 17 Q6	15	
13	Spec. Q5	18	
Total		199	

Linear Regression, Correlation & Chi-Squared Tests – what the new spec asks

WJEC GCE AS / A Level Further Mathematics (from 2017) · Unit 2: Further Statistics A · Topic 2.2.2.

Linear regression 2.2.2

- Linear model $y = \alpha + \beta x + \varepsilon$ with $\varepsilon \sim N(0, \sigma^2)$ independent.
- Least-squares estimates: $\hat{\beta} = \frac{S_{xy}}{S_{xx}}$, $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$.
- $S_{xy} = \sum xy - n\bar{x}\bar{y}$, $S_{xx} = \sum x^2 - n\bar{x}^2$.
- Use line for interpolation only – not extrapolation beyond the data range.

Correlation coefficients 2.2.2

- Pearson's product-moment $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$ measures linear association ($-1 \leq r \leq 1$).
- Spearman's rank $\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$ where d_i are differences of ranks.
- Tied ranks: assign the mean of the tied positions.
- Use Spearman for monotonic (not necessarily linear) relationships and ordinal data.

Significance testing & CIs 2.2.2

- For $\hat{\beta}$: standard error $\text{SE}(\hat{\beta}) = \sigma / \sqrt{S_{xx}}$ when σ is known.
- Test $H_0 : \beta = \beta_0$ using $Z = (\hat{\beta} - \beta_0) / \text{SE}(\hat{\beta}) \sim N(0, 1)$.
- Confidence interval: $\hat{\beta} \pm z_{\alpha/2} \text{SE}(\hat{\beta})$.
- For correlation test: critical values from Pearson / Spearman tables (sample size n).

Chi-squared tests 2.2.2

- Goodness of fit: $\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$ with $\nu = k - 1 - p$ d.o.f. (p estimated params).
- Contingency table: $\nu = (r - 1)(c - 1)$ for an $r \times c$ table.
- Pool adjacent cells if any $E_i < 5$.
- Compute E_i for each row/column from the marginal totals.

Correlation, Regression & Chi-Squared in one page

Quick-reference notes – revisit before each section. Don't use during questions.

Linear regression setup

Model: $y_i = \alpha + \beta x_i + \varepsilon_i$ with $\varepsilon_i \sim N(0, \sigma^2)$ i.i.d.

x values *exact*; y values measured with normal error.

Goal: estimate α, β from the data $\{(x_i, y_i)\}$.

Least-squares formulae

$$S_{xx} = \sum x^2 - n\bar{x}^2$$

$$S_{xy} = \sum xy - n\bar{x}\bar{y}$$

$$\hat{\beta} = b = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\alpha} = a = \bar{y} - b\bar{x}$$

Standard errors

If σ is known and x_i exact:

$$SE(b) = \frac{\sigma}{\sqrt{S_{xx}}}$$

$$SE(\hat{y}_0) = \sigma \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

Minimum SE at $x_0 = \bar{x}$.

Hypothesis test for β

$H_0: \beta = \beta_0$ vs $H_1: \beta \neq \beta_0$ (two-sided).

$Z = \frac{b - \beta_0}{SE(b)} \sim N(0, 1)$ under H_0 .

p -value = $2P(Z > |z_{\text{obs}}|)$. Reject if $p < \alpha$.

Confidence intervals

For β : $b \pm z_{\alpha/2} SE(b)$.

For \hat{y}_0 at $x = x_0$: $(a + bx_0) \pm z_{\alpha/2} SE(\hat{y}_0)$.

For 95% CI use $z_{0.025} = 1.96$. For 99%: $z_{0.005} = 2.576$.

Pearson's correlation

$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$ where $S_{yy} = \sum y^2 - n\bar{y}^2$.

$-1 \leq r \leq 1$; $r = \pm 1$ means perfect linear.

Test $H_0: \rho = 0$ via Pearson tables (critical r for sample size n).

Spearman's rank

Rank the x values and the y values separately.

$\rho_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$ where d_i is the rank difference.

Tied ranks: assign mean of tied positions. Compare to Spearman critical-value tables.

Chi-squared test

$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$, compared with χ^2_ν critical values.

Goodness of fit: $\nu = k - 1 - p$ (k cells, p params estimated).

$r \times c$ contingency: $\nu = (r - 1)(c - 1)$.

Pool cells with $E_i < 5$.

Strategy

1. Compute $\bar{x}, \bar{y}, S_{xx}, S_{xy}$ from summary stats.
2. $b = S_{xy}/S_{xx}$, $a = \bar{y} - b\bar{x}$.
3. Use $SE(b) = \sigma/\sqrt{S_{xx}}$ for tests and CIs.
4. Only interpolate – never extrapolate beyond data range.

SECTION T4

Correlation, Regression & Chi-Squared

Questions 1-13 · 199 marks

5. Alan is investigating the relationship between the resistance, y ohms, of a new type of electrical component and the temperature, $x^\circ\text{C}$. He obtains the following results.

x	0	5	10	15	20	25
y	10.3	11.8	14.2	16.6	17.4	18.9

- (a) Evaluate $\sum x$, $\sum y$, $\sum xy$ and $\sum x^2$. [2]
- (b) Assuming a linear relationship $y = \alpha + \beta x$, calculate a and b , the least squares estimates of α and β . [6]

The values of x are exact whereas the values of y are subject to independent normally distributed errors with zero mean and standard deviation 0.4.

- (c) (i) Use your values of a and b to estimate the true value of the resistance at 20°C . Determine the standard error of your estimate.
- (ii) Hence find a 95% confidence interval for the true value of the resistance at 20°C .
- (d) Alan predicted beforehand that the value of β would be 0.4. Determine, at the 1% significance level, whether or not his results are consistent with this prediction. [12]

7. The following table shows the temperature, $y^{\circ}\text{C}$, of the water in a boiler at various times, x minutes, after switching on.

x	0	5	10	15	20	25	30
y	20	25	31	38	44	49	55

[You may assume that $\sum xy = 4760$, $\sum x^2 = 2275$]

- (a) Assuming a linear relationship $y = \alpha + \beta x$, calculate least squares estimates a , b for α , β . Give your answers correct to four decimal places. [7]
- (b) The boiler is being tested to determine whether or not the value of β is 1.2. Assuming that the values of x are exact whereas the values of y are subject to independent normally distributed errors with zero mean and standard deviation 0.25,
- state suitable hypotheses,
 - calculate the p -value of your value of b and state your conclusion. [9]

7. The length, y cm, of a wire is related to its temperature, $x^\circ\text{C}$, by the equation $y = \alpha + \beta x$. The values of x can be controlled exactly whereas the measured values of y are subject to independent normally distributed errors with mean zero and standard deviation 0.15 cm. The following results were obtained for a particular wire.

Temperature ($x^\circ\text{C}$)	20	30	40	50	60	70
Measured length (y cm)	82.3	83.9	85.3	86.8	88.6	90.1

[You are given that $\sum x = 270$, $\sum y = 517$, $\sum x^2 = 13900$, $\sum xy = 23538$]

- (a) Calculate least squares estimates for α and β . [6]
- (b) Calculate a 99% confidence interval for the actual length of the wire when its temperature is 60°C . [6]

7. The variables x and y are known to be related by an equation of the form $y = \alpha + \beta x$. In order to estimate the values of α and β , the values of y were measured for six different values of x . The following results were obtained.

x	5	10	15	20	25	30
y	15.5	27.2	37.4	49.1	60.8	72.6

[You are given that $\sum x = 105$, $\sum y = 262.6$, $\sum x^2 = 2275$, $\sum xy = 5590.5$]

The values of x are exact but the values of y are subject to independent normally distributed measurement errors with mean zero and standard deviation 0.5.

- (a) Calculate least squares estimates for α and β . [6]
- (b) The value of β is thought to be 2.34. The following hypotheses are therefore defined:

$$H_0: \beta = 2.34 \text{ versus } H_1: \beta < 2.34$$

Calculate the p -value of your result and interpret it. [6]

- (c) Alun is given the same data and he evaluates the least squares estimate of β as 0.52. Explain briefly why this answer is obviously incorrect. [1]

7. The length, y metres, of an elastic string and its tension, x Newtons, are related by an equation of the form $y = \alpha + \beta x$. In order to estimate the values of α and β , the values of y were measured for six different values of x . The following results were obtained.

x	10	20	30	40	50	60
y	2.02	2.23	2.39	2.56	2.77	2.95

The values of x are exact but the values of y are subject to independent normally distributed measurement errors with mean zero and standard deviation 0.02 metres.

- (a) Calculate least squares estimates for α and β . [8]
- (b) Determine a 90% confidence interval for α . [5]

6. The solubility y , in appropriate units, of a certain chemical in water is related to the temperature, $x^\circ\text{C}$, by an equation of the form $y = \alpha + \beta x$. In order to estimate α and β , the following measurements were made.

x	10	12	14	16	18	20
y	21.7	24.4	27.3	29.6	31.7	34.5

[You are given that $\sum x = 90$, $\sum x^2 = 1420$, $\sum y = 169.2$, $\sum xy = 2626.2$]

- (a) Calculate least squares estimates for α and β . [6]
- (b) The values of x are exact but the values of y are subject to independent normally distributed measurement errors with mean zero and standard deviation 0.15. Determine a 99% confidence interval for the solubility of the chemical in water at 17°C . [7]

TURN OVER

5. The temperature $y^{\circ}\text{C}$ in an oven x minutes after switching on the oven can be assumed to satisfy the equation $y = \alpha + \beta x$. In order to estimate α and β , the following measurements were made.

x	0	1	2	3	4	5
y	20.0	34.4	49.3	65.6	79.7	96.5

- (a) Calculate least squares estimates for α and β . [8]
- (b) The values of x are exact but the values of y are subject to independent normally distributed measurement errors with mean zero and standard deviation 0.75. Determine a 99% confidence interval for β . [5]

TURN OVER

6. The resistance, y ohms, of an electrical component is related to the temperature, $x^\circ\text{C}$, by an equation of the form $y = \alpha + \beta x$. In order to estimate the unknown constants α and β , the following measurements were made.

x	10	15	20	25	30	35	40
y	12.3	13.9	15.1	16.6	18.6	20.1	21.5

- (a) Calculate least squares estimates for α and β . [8]
- (b) The values of x are exact but the values of y are subject to independent normally distributed measurement errors with mean zero and standard deviation 0.1. Determine a 95% confidence interval for α . [5]

5. The variables x and y are related by an equation of the form $y = \alpha + \beta x$. In order to estimate the unknown constants α and β , the following measurements were made.

x	2	4	6	8	10	12
y	19.8	33.9	49.9	64.1	77.9	95.0

- (a) Calculate least squares estimates for α and β . [8]
- (b) The values of x are exact but the values of y are subject to independent normally distributed measurement errors with mean zero and standard deviation 0.5.
- (i) Calculate an unbiased estimate of the value of y when $x = 5$.
- (ii) Determine a 95% confidence interval for the value of y when $x = 5$.
- (iii) It was thought beforehand that the value of β was 7.6. Determine whether or not, at the 5% significance level, the values in the table above are consistent with this value of β . [10]

5. The speed of sound in air, y ms⁻¹, and the air temperature, x °C, may be assumed to be related by an equation of the form $y = \alpha + \beta x$. In order to estimate the unknown constants α and β , the following measurements were made.

x	10	15	20	25	30
y	337.1	340.7	343.0	346.1	349.7

- (a) Calculate least squares estimates for α and β . [8]
- (b) The values of x are exact but the values of y are subject to independent normally distributed measurement errors with mean zero and standard deviation 0.25.
- (i) Determine a 99% confidence interval for α , giving your answer correct to one decimal place. [5]
- (ii) Test, at the 5% significance level, the null hypothesis $H_0 : \beta = 0.65$ against a two-sided alternative. [7]

5. The amount, y grams, of chemical that dissolves in 1 litre of water at a temperature of $x^\circ\text{C}$ satisfies the relationship $y = \alpha + \beta x$. In order to estimate the unknown constants α and β , the following measurements were made.

x	10	20	30	40	50	60
y	162	183	201	225	248	267

- (a) Calculate least squares estimates for α and β . [8]
- (b) The values of x are exact but the values of y are subject to independent normally distributed measurement errors with mean zero and standard deviation 1.5.
- (i) Determine a 95% confidence interval for β , giving your limits correct to three significant figures.
- (ii) A 95% confidence interval is to be determined for the value of y when $x = x_0$. Giving a reason, state the value of x_0 for which the confidence interval has minimum width. [7]

6. The length, y cm, of a spring subjected to a tension of x Newtons satisfies the relationship $y = \alpha + \beta x$, where α and β are unknown constants. In order to estimate α and β , the following measurements were made.

x	10	15	20	25	30	40
y	12.4	14.3	16.4	18.9	20.7	24.6

You are given that $\sum x = 140$, $\sum y = 107.3$, $\sum x^2 = 3850$, $\sum xy = 2744$.

- (a) Calculate least squares estimates for α and β , giving your answers correct to three significant figures. [6]
- (b) The values of x are exact but the values of y are subject to independent normally distributed measurement errors with mean zero and standard deviation 0.2. Before the measurements were made, Emlyn believed that the value of β was 0.4.
- (i) State suitable hypotheses to carry out a two-sided test of Emlyn's belief.
- (ii) Calculate the p -value of the above results.
- (iii) State whether or not the data support Emlyn's belief. [9]

TURN OVER

5. Jim is investigating the relationship between the length of a wire, y cm, and the temperature, $x^\circ\text{C}$, of the wire. He obtains the following experimental results.

x	10	15	20	25	30	35	40	45
y	74.3	76.1	77.2	78.6	80.4	82.1	83.9	85.5

[You are given that $\Sigma y = 638.1$; $\Sigma xy = 17882.5$]

- (a) Assuming a linear relationship $y = \alpha + \beta x$, calculate a and b , the least squares estimates of α and β . [5]

The values of x are exact whereas the measured values of y are subject to independent normally distributed errors with zero mean and standard deviation 0.2.

- (b) Test the null hypothesis that $\beta = 0.3$ against a two-sided alternative, using a significance level of 5%. [7]
- (c) (i) Use your values of a and b to estimate the true length of the wire when the temperature is 35°C . Determine the standard error of your estimate.
- (ii) Hence find a 95% confidence interval for the true length of the wire when the temperature is 35°C . [6]

END OF CORRELATION, REGRESSION & CHI-SQUARED PACK

Source: WJEC S3 (2008 modular spec) · 2005–2017
Curated for WJEC FM 2017 spec AS Unit 2 – Topic 4 (2.2.2)

© WJEC CBAC Ltd. Pack layout © revise.wales for revision purposes only.