A Level Statistics
AQA Past Exam Questions

TOPIC: HYPOTHEIS TESTING
Correlation Coefficients

Candidates may use any calculator allowed by Pearson regulations. Calculators must not have retrievable mathematical formulae stored in them.

Instructions

• Use **black** ink or ball-point pen.

• If pencil is used for diagrams/sketches/graphs it must be dark (HB or B). Coloured pencils and highlighter pens must not be used.

• **Fill in the boxes** at the top of this page with your name, centre number and candidate number.

• Answer **all** questions and ensure that your answers to parts of questions are clearly labelled.

• Answer the questions **on paper**

• You should show sufficient working to make your methods clear. Answers without working may not gain full credit.

• Unless otherwise stated, statistical tests should be carried out at the 5% significance level.

• When a calculator is used, the answer should be given to three significant figures unless otherwise stated.

Information

• **You may use the** booklet 'Statistical Formulae and Tables'

• There are **15** questions in this question paper. The total mark for this paper is **157**

• The marks for **each** question are shown in brackets – use this as a guide as to how much time to spend on each question.

Advice

• Read each question carefully before you start to answer it.

• Try to answer every question.

• Check your answers if you have time at the end.

• If you change your mind about an answer, cross it out and put your new answer and any working underneath.

• Check your answers if you have time at the end.

AQA_JAN_2012_1

For films released in the USA between 1998 and 2008, information is available on the body count, x, and on the box office total gross takings, $y million.

The table shows information for 10 of these films, selected at random from those which have a body count greater than 50.

Film	x	y
Titanic	307	1849
Return of the King	836	1133
The Two Towers	468	926
Troy	572	497
Saving Private Ryan	255	481
Gladiator	77	458
The Last Samurai	558	457
Bad Boys II	63	273
Rambo	271	117
We Were Soldiers	305	115

The box office total gross takings are given in $million adjusted to 2008 figures.

(a) Calculate the value of Spearman's rank correlation coefficient between x and y.

(6 marks)

(b) Carry out a hypothesis test, at the 10% level of significance, to determine whether the value that you calculated in part (a) indicates a positive association between x and y. Interpret your conclusion in context.

(4 marks)

AQA_JUNE_2013_6b

Nine randomly selected students with extrovert personalities were also asked to complete a subjective happiness assessment. For each selected student, the rank order of the score from this assessment and the rank order of the score from their personality test are given in Table 4.

A higher rank for a happiness score indicates a higher level of happiness and a higher rank for a personality score indicates a more extrovert personality.

Table 4

Student	A	B	C	D	E	F	G	H	I
Personality score rank	9	8	3½	6	7	5	2	1	3½
Happiness score rank	9	8	5	6	7	4	3	2	1

(i) Calculate the value of Spearman's rank correlation coefficient between personality score and happiness score.

(3 marks)

(ii) Carry out a hypothesis test, at the 1% level of significance, to determine whether the value that you calculated in part (b)(i) indicates an association between personality score and happiness score.

(5 marks)

AQA_JUNE_2012_2

Records of claims following hurricanes in the USA are kept by an insurance company. The number of injuries, x thousands, and the total approximate cost at today's prices, y million dollars, for each of 10 hurricanes that occurred between 1930 and 2010 are given in the table.

x	415	362	190	150	47	24	290	390	95	69
y	3200	1390	680	260	200	180	1050	2430	290	210

The 10 selected hurricanes may be regarded as a random sample.

(a) Calculate the value of Spearman's rank correlation coefficient between x and y.

(6 marks)

(b) Carry out a hypothesis test, at the 1% level of significance, to determine whether your calculated value in part (a) indicates a positive association between x and y. Interpret your conclusion in the context of the question.

(4 marks)

AQA_JAN_2013_6
Data were collected between 1995 and 2006 on the educational experience of children and their parents living in the UK.
Table 3 shows the number of years spent in full-time education beyond the age of 16 years for the mother, son and daughter of 10 families.

Table 3

Family	Mother	Son	Daughter
A	8	7	7
B	7	4	8
C	6	8	6
D	5	4	6
E	4	6	6
F	3	2	5
G	3	3	4
H	2	0	3
I	1	1	2
J	0	5	0

You may regard the 10 families as a random sample. On the page below, Table 4 shows some of the rank values for the data in Table 3. Complete Table 4.

(4 marks)

(b) Calculate the value of Spearman's rank correlation coefficient between the number of years spent in full-time education beyond the age of 16 years for:
(i) mother and son;
(ii) mother and daughter.

(4 marks)

(c) Carry out hypothesis tests, at the 1% level of significance, to determine whether the values that you calculated in part (b)(i) and part (b)(ii) indicate a positive association between the number of years spent in full-time education beyond the age of 16 years for: (i) mother and son; (ii) mother and daughter.

(7 marks)

Table 4

Family	Mother	Son	Daughter
A	1	2	2
B	2	5½	1
C	3	1	
D		5½	
E		3	
F		8	
G		7	
H		10	
I	9	9	9
J	10	4	10

AQA_JUNE_2014_1

Kalia, a sports researcher, wished to investigate the correlation between an athlete's ability to sprint and ability to run longer distances. She selected, at random, 8 schoolboy county athletes from those who took part in both 400-metre sprint races and cross-country races during 2012. Table 1 shows, for these races, the rank order for the overall performance of each schoolboy in the 400-metre county sprint races together with the position of each schoolboy in the county cross-country race final.

Table 1

Schoolboy	A	B	C	D	E	F	G	H
400-metre (rank)	3	4	7	5	1	2	8	6
Cross-country (position)	28	1	12	9	68	30	15	19

(a) Explain why Spearman's rank correlation coefficient, rather than the product moment correlation coefficient, is the appropriate measure of correlation for these data.

[1 mark]

(b) Calculate the value of the Spearman's rank correlation coefficient, rs , between overall performance in 400-metre races and position in the cross-country race final.

[5 marks]

(c) Carry out a hypothesis test, at the 5% level of significance, to determine whether your value of rs calculated in part (b) indicates a correlation between overall performance in 400-metre races and position in the cross-country race final.

[4 marks]

(d) Explain, in the context of this question, the meaning of a Type II error.

[2 marks]

(e) Kalia decided to investigate further and so obtained, for each of the 8 schoolboys, his best time, in seconds, in a 400-metre race during 2012 and also his time taken, in minutes, to complete the 2012 county cross-country race final. These times are given in Table 2.

Table 2

Schoolboy	A	B	C	D	E	F	G	H
Race times 400-metre	53.3	54.7	55.3	54.9	52.3	53.2	55.6	55.1
Cross-country	21.18	17.02	18.41	18.18	24.94	22.64	18.73	18.97

(i) Calculate the value of the product moment correlation coefficient between the times taken for the two races.

(ii) Comment on your value in the context of this question.

[4 marks]

AQA_JUNE_2017_2

A wholesaler supplies many different types of 100-gram dark chocolate bars.
A customer wants to investigate whether the recommended retail price for a 100-gram dark chocolate bar is associated with the percentage cocoa content of the chocolate in that bar.
The customer decides to select nine types of bar at random from those supplied by the wholesaler.
For each type of bar, the recommended retail price, x pence, and the cocoa content, y per cent, of the chocolate was recorded.
The results are given in the table.

Type	A	B	C	D	E	F	G	H	I
x	189	209	295	220	175	295	255	195	225
y	57	70	72	80	65	90	95	62	70

(a) Calculate the value of Spearman's rank correlation coefficient between x and y .

[6 marks]

(b) (i) Carry out a hypothesis test, using the 1% level of significance, to investigate whether the correlation coefficient calculated in part (a) indicates a positive correlation between recommended retail price and percentage cocoa content for 100-gram dark chocolate bars.

[4 marks]

(b) (ii) Explain the meaning of a Type I error in the context of the test in part (b)(i).

[2 marks]

AQA_JUNE_2016_5
Fatima, a film reviewer, was asked to investigate the level of gender bias in films.
A random sample of 12 films, released during 2010, was obtained.
Fatima watched the 12 films and then assigned to each of them a gender bias score based on the importance of the female characters in the film and on how well-rounded their roles were.
The score assigned was measured on a scale of 0 to 20, where 20 indicated the least gender bias.
For each film, the gender bias score assigned and the rank of the gross box office takings, where rank 1 indicates the highest takings, are given in Table 1.

Table 1

Film	Gender bias score	Rank of gross box office takings
A	15	2
B	18	10 ½
C	19	9
D	8	7
E	10	10 ½
F	17	3
G	13	4
H	20	12
I	13	1
J	16	8
K	12	6
L	14	5

Fatima decided to quantify the correlation between gender bias score and gross box office takings.

(a)
(i) Calculate an appropriate measure of correlation.

[5 marks]

(ii) Carry out a hypothesis test, using the 10% level of significance, to investigate whether the correlation coefficient calculated in part (a)(i) indicates an association between gender bias score and gross box office takings.

[4 marks]

(b) Information was also available about the running time, in minutes, and the budget, in $ million, for making the film for each of the 12 films in Fatima's sample.
This information is given in Table 2.

Table 2

Film	Running time	Budget
A	135	125
B	118	110
C	91	5
D	120	40
E	117	110
F	95	15
G	108	13
H	101	20
I	95	69
J	100	60
K	113	80
L	140	155

Arthur, Fatima's assistant, believed that films that had a longer running time also had a larger budget.

(i) Find the value of the product moment correlation coefficient between running time and budget.

[3 marks]

(ii) Hence carry out a hypothesis test, using the 1% level of significance, to investigate Arthur's belief.

[4 marks]

AQA_JUNE_2015_1
A paediatric doctor measures the height, x cm, and the systolic blood pressure, y mmHg, of 12
randomly selected healthy boys aged between 5 years and 10 years.
The results are given in the following table.

Child	1	2	3	4	5	6	7	8	9	10	11	12
x	96	146	151	126	112	132	107	115	121	111	142	136
y	98	108	112	105	106	109	101	108	103	106	109	106

(a) Find the value of the product moment correlation coefficient between height and systolic blood
pressure.

[3 marks]

(b) The paediatric doctor believes that there is a positive correlation between height and systolic
blood pressure in healthy boys aged between 5 years and 10 years. Carry out a hypothesis test, at
the 1% significance level, to investigate this belief.

[4 marks]

(c) Give one reason why your conclusion in part (b) might not apply to all children aged between 5
years and 10 years.

[1 mark]

AQA_JAN_2007_6a
Percentage marks were obtained for a random sample of 10 university economics students for
their second year examinations in micro-economics and in macro-economics. The results are
given in the table.

Student	Micro-economics	Macro-economics
1	38	28
2	41	45
3	47	53
4	51	57
5	54	52
6	56	47
7	59	54
8	61	58
9	63	63
10	70	69

(a) (i) Calculate the value of Spearman's rank correlation coefficient between the marks for micro-
economics and macro-economics.

(5 marks)

(ii) Carry out a hypothesis test, at the 2% level of significance, to determine whether the value that
you calculated in part (a)(i) indicates an association between the marks for micro-economics and
macro-economics. Interpret your conclusion in context.

(5 marks)

AQA_JAN_2008_5
The gate receipts, x, and the player costs, y, during 1990 for a random sample of eleven US baseball teams are given in the table.
All values for x and y are given in millions of US dollars.

Team	x	y
A	25.3	22.7
B	25.2	22.2
C	24.6	22.3
D	22.5	20.4
E	20.8	19.6
F	19.7	13.8
G	19.0	22.5
H	18.1	8.1
I	16.0	14.2
J	16.0	23.6
K	11.1	16.8

(a) Calculate the value of Spearman's rank correlation coefficient between x and y.
(6 marks)

(b) Carry out a hypothesis test, at the 10% level of significance, to determine whether the value that you calculated in part (a) indicates a positive association between x and y. Interpret your conclusion in context.
(5 marks)

AQA_JAN_2010_2
An investigation was carried out into the fat content and the energy content of 'medium-sized' chocolate milkshakes sold in fast-food outlets in the USA.
A 'medium-sized' chocolate milkshake was purchased from each of seven fast-food outlets. The number of calories, x, the fat content, y grams, and the volume, z fluid ounces, were measured for each milkshake. The results are given in the table.

Fast-food outlet	x	y	z
A	320	2	10.2
B	400	9	11.4
C	430	8	10.9
D	455	13	12.3
E	390	11	9.7
F	465	14	13.2
G	425	12	14.1

(a) Find the value of the product moment correlation coefficient between: (i) x and y; (ii) x and z.
(5 marks)

(b) Carry out hypothesis tests, at the 5% level of significance, to determine whether the values of the two product moment correlation coefficients found in part (a) indicate a positive association between each of the pairs of variables.
(4 marks)

(c) Summarise, in context, your findings in parts (a) and (b).
(2 marks)

AQA_JAN_2011_1
During a particular month, the average number of hours slept per night, x, and the average
diastolic blood pressure, y, were recorded for each of a sample of 12 adult males.
The results are given in the table.
You may regard the sample as a random sample of all adult males.

Male	x	y
A	10.6	70
B	10.2	74
C	9.8	78
D	9.2	68
E	8.4	72
F	8.0	86
G	7.8	76
H	7.4	80
I	7.4	85
J	7.2	88
K	6.2	92
L	4.1	91

(a) Calculate the value of Spearman's rank correlation coefficient between x and y.

(6 marks)

(b) Carry out a hypothesis test, at the 1% level of significance, to determine whether the value that
you calculated in part (a) indicates an association between x and y. Interpret your conclusion in
context.

(4 marks)

AQA_JUNE_2007_2
An American study investigated the weight gains, x kg, of mothers during pregnancy and the
weights, y kg, of their children at 3 years of age.
The table gives the results for a random sample of 10 mothers and their children.

Mother	A	B	C	D	E	F	G	H	I	J
x	18.3	10.8	16.9	8.2	9.8	10.1	12.4	14.6	15.3	18.8
y	19.2	11.5	18.3	10.5	12.4	13.7	14.6	13.5	14.1	20.1

(a) Calculate the value of the product moment correlation coefficient between the weight gains of
mothers during pregnancy and the weights of their children at 3 years of age.

(3 marks)

(b) Carry out a hypothesis test, at the 1% level of significance, to determine whether the value that
you calculated in part (a) indicates a positive association between the weight gains of mothers
during pregnancy and the weights of their children at 3 years of age. Interpret your conclusion in
context.

(5 marks)

AQA_JUNE_2008_2

A road safety organisation obtained the annual number of road deaths, x per 100 000 of the population, and the number of motor vehicles, y per 1000 of the population, for countries in the EU.

The table gives the results for a random sample of 10 countries in the EU.

Country	A	B	C	D	E	F	G	H	I	J
x	5.9	6.1	6.3	8.0	8.4	10.2	10.5	12.8	14.8	19.3
y	559	528	518	650	487	607	754	597	496	480

(a) Calculate the value of Spearman's rank correlation coefficient between x and y.

(6 marks)

(b) Carry out a hypothesis test, at the 10% level of significance, to determine whether the value that you calculated in part (a) indicates an association between the annual number of road deaths per 100 000 of the population and the number of motor vehicles per 1000 of the population for countries in the EU.

(5 marks)

AQA_JUNE_2009_1

A clinical nutrition department at a large hospital carried out research into the levels of body fat in females.

The age, x years, and the body fat, y per cent, for each of 10 randomly selected females are given in the table.

Female	x	y
A	23	27.9
B	39	31.4
C	41	25.9
D	49	25.2
E	53	34.7
F	56	32.5
G	57	30.3
H	58	33.0
I	60	41.1
J	61	34.5

(a) Calculate the value of Spearman's rank correlation coefficient between x and y.

(6 marks)

(b) Carry out a hypothesis test, at the 10% level of significance, to determine whether the value that you calculated in part (a) indicates an association between x and y. Interpret your conclusion in context.

(4 marks)