A Level Statistics

AQA Past Exam Questions

Solutions

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.

1 (a)

Film	Ti	Ret	2T	Tr	Ry
x rank	1	2	3	4	5
y rank	5	1	4	2	8
Film	Gl	Sam	BB	Ra	Sol
x rank	6	7	8	9	10
y rank	9	3	10	7	6

M1 Attempt at ranks

M1 14 correct (can be reversed)

A1

Alternative

$d = 4, 1, 1, 2, 3, 3, 4, 2, 2, 4$

$\sum d^2 = 80$ B1

$r_s = 0.515$ or 17/33 (3 sf from calc) B3 6

$r_s = 1 - \frac{6 \times 80}{10 \times 99} = 0.515$ M1, A1

sc2: 0.51/2

sc1: 0.5 awrt

(b)

H_o Rank orders of gross takings and body counts are independent.

H_1 Rank orders of gross takings and body counts are not independent: a positive association exists

1 tail 10% B1 or equivalent in words/symbols

cv = 0.4424 B1

test stat $r_s = 0.515$

$r_s > cv$ M1 comparison ts/cv

comparison ts/cv

Reject H_o Significant evidence at 10% level to suggest a positive association between rank orders of gross takings and body counts. For films with a body count greater than 50, those with higher body counts tend to have higher gross takings. E1 4 Correct conclusion in context

AQA_JUNE_2013_6b

6(b)(i)	d 0 0 1.5 0 0 1 1 1 2.5 $\sum d^2 = 11.5$	M1		Differences
	SRCC $r_s = 1 - \frac{6\times\sum d^2}{9\times 80} = 0.904$	M1 A1		Formula correct
	or SRCC $r_s = 0.904$ (from calc)	(B3)	3	SC1 0.9 SC2 0.90 if no method shown
(ii)	H_0 : Rank orders of personality score and happiness score are independent. H_1 : Rank orders of personality score and happiness score are not independent. 2 tail 1%	B1		Hypothesis
	cv = 0.8167 test stat r_s = 0.904	B1		cv cao
	r_s > cv	M1		comparison ft seen or implied
	Reject H_0 Significant evidence at 1% level to suggest an association (positive) between rank orders of personality score and happiness score. Students with a higher extrovert personality score tend to have a higher happiness score.	A1 E1	5	ts/cv correct in context – vice versa OK

AQA_JUNE_2012_2

2				
(a)	ranks x 1, 3, 5, 6, 9, 10, 4, 2, 7, 8, 10, 8, 6, 5, 2, 1, 7, 9, 4, 3 y 1, 3, 5, 7, 9, 10, 4, 2, 6, 8 10, 8, 6, 4, 2, 1, 7, 9, 5, 3	M1 M1 A1		for any ranks 2 separate sets of ranks All correct
	r_s (from calculator) = 0.988 or 0.987 0.98/0.99 allow B2 if no method seen	B3	6	alternatively differences, d : 0, 0, 0, 1, 0, 0, 0, 0, 1, 0 $\sum d^2 = 2$ M1 diffs $r_s = 1 - \frac{6\times 2}{10\times 99} = 0.988$ or 0.987 M1, A1
(b)	H_0 no association H_1 positive association 1 tail 1%	B1		Allow $p/\rho = 0$ or words Must be 1 tail
	test stat r_s = 0.988 critical value = 0.7333 tests stat > 0.7333 so significant evidence exists to reject H_o and conclude that a positive association exists.	B1 M1		for cv comparison ts/cv; ft r_s in (a) 0.7667,0.7818/0.6485/0.700 B0 M1 E0
	This suggests that hurricanes in which there are higher numbers of injuries also result in a greater cost in property damage (or positive assoc in context)	E1	4	explanation in context

6(a) Ranks

	Rank mother	*Rank son*	*Rank d'ter*
1	1	2	2
2	2	5½	1
3	3	1	**4**
4	**4**	5½	**4**
5	**5**	3	**4**
6	**6½**	8	**6**
7	**6½**	7	**7**
8	**8**	10	**8**
9	9	9	9
10	10	4	10

M1 Any 3 correct ranks mother

M1 Any 3 correct ranks daughter

M1 Ties correct in any column

A1 All correct

4

(b)(i) $r_s = 0.598$ (3 sig figs)

B2

Alt diffs $|d|$
1, 3½, 2, 1½, 2, 1½, ½, 2, 0,6
$\sum d^2 = 66$ M1

$r_s = 1 - \frac{6 \times 66}{10 \times 99} = 0.6$ A1

(ii) $r_s = 0.972$ (3 sig figs)

B2

Alt diffs $|d|$
1, 1, 1, 0, 1, ½, ½, 0, 0, 0,
$\sum d^2 = 4.5$ M1

$r_s = 1 - \frac{6 \times 4.5}{10 \times 99} = 0.973$ A1

4

(c)(i)(ii)

H_O no assoc in ranks in population between mother and son/daughter
H_1 positive assoc in ranks in population between mother and son/daughter

or equivalent

B1 for both

1 tail test 1 % level
cv = 0.7333

B1 for cv

Mother/son ts $r_s = 0.598 < 0.7333$ M1
Accept H_O A1 conclusion correct
Mother/daughter ts $r_s = 0.972 > 0.7333$ M1
Reject H_O A1 conclusion correct
There is significant evidence of a positive correlation between number of years spent in full-time education for mother and daughter but no significant evidence of a positive correlation for mother and son. E1

7

1a Spearman's rank correlation coefficient is the appropriate measure of correlation for these data because there are no measured values given.

E1

1

E1 Ranks only for 400m & position only for cross country or ref to orders given

1b Ranks

	Rank 400m	*Rank cross country*	$\|d\|$
A	3	6	3
B	4	1	3
C	7	3	4
D	5	2	3
E	1	8	7
F	2	7	5
G	8	4	4
H	6	5	4

M1
A1

M1 for ranks attempt cross country
A1 all correct (can be reversed) can be implied by d

m1dep for $\sum d^2 = 134$

$r_s = 1 - \frac{6\times134}{8\times63} = -0.595$ M1 A1

$r_s = -0.595$ (3 sig figs)

B3

B1 r_s negative
B2 $0.590 \leq |r_s| \leq 0.599$

5

1c H_0: $\rho_s = 0$
H_1: $\rho_s \neq 0$ 2 tail 5%

B1

Hypotheses oe

test stat $|r_s| = 0.595$
|critical value| = 0.7381
–0.595 > –0.7381 so no significant evidence exists to reject H_0

B1
M1

Correct abs value for cv 0.738(1)
Correct comparison both –ve/ +ve

This suggests that there is no correlation between rank/ position in 400m races and position in county cross country final race.

E1dep

4

Conclusion correct in context

1d H_0 accepted in error as H_0 actually untrue
Conclusion made that there is no correlation between rank/ position in 400m races and position in county cross country final race when, in reality, there is a correlation between them.

B1

E1

2

Correct explanation of Type II error

In context

1e(i) PMCC $r = -0.904$ (3 sf) (from calculator)

B3

(–0.905, –0.903)
or $r = \frac{8671.488 - \frac{434.4\times160.07}{8}}{3.17\times7.08} = \frac{-20.3}{22.4}$
$= -0.904$ (3 sf)
M1 (num), M1(denom), A1

(ii) sc
–0.90 allow M1 M1 A0 (or B2)
–0.9 sc allow B1

PMCC indicates a strong negative correlation between best time taken to run 400m and time taken to run cross country race final. This indicates that we would expect faster 400m runners to be slower at running the cross country race.

E1

4

Interpretation in context

2(a) Ranks

	x		y	
A	2	8	1	9
B	4	6	4½	5½
C	8½	1½	6	4
D	5	5	7	3
E	1	9	3	7
F	8½	1½	8	2
G	7	3	9	1
H	3	7	2	8
I	6	4	4½	5½

M1 Attempt to rank one column/category

m1 Correct use of ties at least once

A1 All correct and consistent

sc Ranked as one group gains M1 m1 only

r_S = **0.807** from calculator

B3 (condone **0.806**) AWRT

or

d = 1, -½, 2½, -2, -2, ½, -2, 1, 1½

$\sum d^2$ = 23

(m1) Differences (ignore sign)

(m1) Formula correct

SRCC $r_s = 1 - \frac{6 \times 23}{9 \times 80} = \mathbf{0.808}$

(A1)

sc2 no method r_S = 0.81
sc1 no method r_S = 0.8
(-B1 if negative)
sc 4 -0.806/7/8 nms
Note PMCC 0.691 gains 0

6

(b)(i)

H_o Rank orders of price and cocoa content are not associated
H_1 Rank orders of price and cocoa content have a positive association

B1 Correct hypotheses oe
H_o no correlation
H_1 positive correlation

1 tail 1% |cv| = 0.7667

B1 cv correct cao awrt 0.767

$|r_S|$ = 0.808 or 0.807 or 0.806 > 0.7667

M1 consistent comparison with correct cv +/+ or -/-
(allow small slip in part (a))

Reject H_o
Significant evidence to suggest that rank orders of recommended **retail price** and percentage cocoa/ **cocoa content** have a positive association.

E1 Correct conclusion in context need price and cocoa percentage or cocoa content
Dep ts/cv correct

4

(ii)

H_o rejected in error.

Type I error defined

E1

Conclusion that there is a positive association between recommended retail price and percentage cocoa content, when, in fact, there is no positive association.

E1 In context need price and cocoa percentage or cocoa content

2

(a) (i) Ranks

	Score rank	Takings ranks
A	6 7	2
B	3 10	10½
C	2 11	9
D	12 1	7
E	11 2	10½
F	4 9	3
G	8½ 4½	4
H	1 12	12
I	8½ 4½	1
J	5 8	8
K	10 3	6
L	7 6	5

M1 — Attempt to rank **score** (can be reversed)

M1 — Correct use of ties

Note PMCC 0.2749 B0

$r_S = \pm 0.312$ from calculator — B3 — scB2 no method $r_S = \pm 0.31$

or

$|d| =$ 4, 7½, 7, 5, ½, 1, 4½, 11, 7½, 3, 4, 2

or (5, ½, 2, 6, 8½, 6, ½, 0, 3½, 0, 3, 1) — or B1 — Differences and $\sum d^2$ effort

$\sum d^2 =$ 374 or $\sum d^2 =$ 196 — M1 — formula correct

$r_s = 1 - \frac{6\times374}{12\times143} =$ - 0.308 or $= 1 - \frac{6\times196}{12\times143} =$ 0.315 — A1 — - 0.308, 0.315

5

(ii) B1

H_o Rank orders of gender bias score and takings are independent.

H_1 Rank orders of gender bias score and takings are not independent.

Allow H_o No association

H_1 Association(or ref to r/ρ)

Correct hypotheses stated in (i) or (ii)

2 tail 10% |cv| = 0.5035 — B1 — cv correct cao

$|r_S| = 0.312$ or 0.308 or 0.315 < 0.5035 — M1 — consistent comparison correct SRCC and correct cv +/+ or -/- PI

Accept H_o

No significant evidence at 10% level to doubt that rank orders of gender bias score and takings are independent — E1 — Conclusion in context

4

(b) (i) PMCC = 0.813 from calculator — B3

SRCC 0.754 B0 sc 0.81 allow M1 M1 A0 (or B2)

or

$r_s = \frac{96124 - \frac{1333\times802}{12}}{\sqrt{2728.9\times27429.7}} = \frac{87253.2}{52.2\times165.6}$

$= \frac{7035.2}{8651.8} = 0.813$ (3 sf)

3

M1, M1, A1 awfw (0.811, 0.815)

(ii) H_0: ρ = 0

H_1: ρ > 0 1 tail 1% — B1 — Hypotheses oe

test stat $|r| = 0.813$

|cv| = 0.6581 — B1 — cao cv

0.813 > 0.6581 — M1 — consistent comparison correct PMCC with cv

Reject H_0

Significant evidence to suggest that there is a positive correlation between running time and budget

or Arthur's belief is correct/supported by the conclusion. — E1 — correct conclusion in context

4

AQA_JUNE_2015_1

1(a)	$r = 0.809$ from calculator or $\sum xy = 158928$ B1 and $r = \dfrac{\frac{158928}{12} - \left(\frac{1495}{12} \times \frac{1271}{12}\right)}{\sqrt{\left(\frac{189473}{12} - \left(\frac{1495}{12}\right)^2\right) \times \left(\frac{134781}{12} - \left(\frac{1271}{12}\right)^2\right)}}$ oe $= \dfrac{48.549}{16.383 \times 3.662} = 0.809$ M1 A1	B3	**3**	sc 0.81 no workings B2 sc 0.8 no working B1 (0.799 – 0.815)
1(b)	H_0: $\rho = 0$ H_1: $\rho > 0$ 1 tail 1% test stat $r = 0.809$ critical value = 0.658 0.809 > 0.658 so significant evidence exists to reject H_0 This suggests that there is a **positive correlation** between **height** and systolic **blood pressure** for healthy boys aged between 5 years and 10 years.	B1 B1 M1 dep cv E1	**4**	Hypotheses oe Correct value for cv Comparison 'ts'/cv [or Reject H_0] Conclusion correct in context Dep ts and cv correct
1(c)	Conclusion can only refer to **healthy/boys no girls**, not all children		**1**	

6(a)(i)

student	1	2	3	4	5
micro rank	1	2	3	4	5
macro rank	1	2	5	7	4

student	6	7	8	9	10
micro rank	6	7	8	9	10
macro rank	3	6	8	9	10

M1 A1 attempt at ranks

$r_s = 0.854(5)$ (3 sf from calc) B3 5

Alternative
$d = 0, 0, 2, 3, 1, 3, 1, 0, 0, 0$
$\sum d^2 = 24$ B1 $r_s = 1 - \frac{6\times 24}{10\times 99}$
$= 0.854(5)$ M1, A1
(2 sf and no working SC4, A0)

(a)(ii)

H_0 Rank orders of micro-economics marks and macro-economics marks are independent.
H_1 Rank orders of micro-economics marks and macro-economics marks are not independent. 2 tail 2% B1

H_0 no association
H_1 association

cv $= \pm 0.7333$ B1 for cv

test stat $r_s = 0.854(5)$
$r_s > 0.7333$ or $|r_s| > |\text{cv}|$ M1

for comparison ts/cv
$r_s = 0.854(5)$

Reject H_0
Significant evidence at 2% level to suggest an association between rank orders of micro-economics marks and macro-economics marks.
[Student with higher rank mark in micro-economics also has higher rank mark in macro-economics] A1 E1 5

allow A1 if r 'close' and marks lost in (a)(i)
in context

5(a)

Team	A	B	C	D	E
x rank	1	2	3	4	5
y rank	2	5	4	6	7

Team	F	G	H	I	J	K
x rank	6	7	8	$9\frac{1}{2}$	$9\frac{1}{2}$	11
y rank	10	3	11	9	1	8

M1 Attempt at ranks (can be reversed)

M1 For 8 correct

A1

$r_s = 0.355$ (3sf from calc) B3 6

Or

$d = 1, 3, 1, 2, 2, 4, 4, 3, \frac{1}{2}, 8\frac{1}{2}, 3$

$\sum d^2 = 141\frac{1}{2}$ B1

$r_s = 1 - \frac{6 \times 141\frac{1}{2}}{11 \times 120} = 0.357$ M1A1

SC4 0.36

SC4 0.318

(b) H_0 Rank orders of gate receipts and player costs are independent B1

H_1 Rank orders of gate receipts and player costs are not independent – there is a positive association

1 tail 10%

cv = 0.4182 B1 For cv

ts $r_s = 0.355$ M1 For comparison ts/cv

$r_s < 0.4182$ A1 $r_s = 0.355$ or 0.357

Accept H_0 No significant evidence at 10% level to suggest a positive association between rank orders of gate receipts and player costs E1 5 In context

Q	Solution	Marks	Total	Comments
2(a)(i)	From calculator $r = 0.891$			Alternative $n = 7$
	or $r = \dfrac{29495 - \dfrac{2885 \times 69}{7}}{\sqrt{14242.86} \times \sqrt{98.86}}$			$\sum y = 69 \quad \sum x = 2885$
				$\sum y^2 = 779$
	$= \dfrac{1057.14}{119.34 \times 9.94}$			$\sum x^2 = 1203275$
	$= 0.891$			$\sum xy = 29495$ M1
		M1		
		m1		m1 formula in (i) or (ii)
		A1		0.885 to 0.905 A1 (3sf)
(ii)	From calculator $r = 0.658$			
	or $r = \dfrac{34021 - \dfrac{2885 \times 81.8}{7}}{\sqrt{14242.86} \times \sqrt{15.35}}$			Alternative $n = 7$
				$\sum z = 81.8 \quad \sum z^2 = 971.24$
	$= \dfrac{307.71}{119.34 \times 3.92}$			$\sum xz = 34021$ M1
	$= 0.658$	M1A1	5	0.650 to 0.665 A1
(b)	$r_{xy} = 0.891 \quad r_{xz} = 0.658$			
	$H_0 \ \rho = 0$			
	$H_1 \ \rho > 0$ 1 tail 5 % sig level	B1		For hypotheses stated correctly once
	Need only be stated once			
	test stat $r_{xy} = 0.891$			
	cv = 0.6694 $n = 7$			For cv and comparison
	since $t > 0.6694$			
	Reject H_0	M1		For Reject H_0; ft
	test stat $r_{xz} = 0.658$	A1√		
	cv = 0.6694 $n = 7$			
	since $t < 0.6694$			
	Accept H_0	A1	4	For Accept H_0
(c)	There is significant evidence to suggest a positive correlation between the calories and the fat content of milkshakes: the higher the fat content, the higher the calories.	E1		
	There is no significant evidence to suggest a positive correlation between the calories and the volume of the milkshakes.	E1	2	Need to refer to part (b)

1(a)

Male	A	B	C	D	E
x rank	1	2	3	4	5
y rank	11	9	7	12	10
Male	F	G	H	I	J
x rank	6	7	8½	8½	10
y rank	4	8	6	5	3
Male	K	L			
x rank	11	12			
y rank	1	2			

M1 attempt at ranks
M1 18 correct (can be reversed)
A1 all correct — can be reversed

alternative
d = 10, 7, 4, 8, 5, 2, 1, 2½, 3½, 7, 10, 10
$\sum d^2$ =526½ B1

r_s = –0.844 (3 sf from calc) B3 6

$r_s = 1 - \frac{6\times 526.5}{12\times 143} = -0.841$ M1, A1

SC4 –0.84 (no method seen)
SC2 –0.8 (no method seen)
SC4 0.846 (no ties)
SC4 +0.844 (inconsistent ranks)

(b)

H_0 Rank orders number of average number of hours slept and diastolic blood pressure are independent. B1

H_1 Rank orders are not independent.
2 tail 1%

cv = ± 0.7273 B1 for cv — ignore sign
test stat r_s = – 0.844 (or – 0.841)
r_s < – 0.7273 or $|r_s| > |cv|$ M1 for comparison ts/cv Must be consistent = or –
Reject H_0. Significant evidence at 1% level to suggest an association between rank orders number of average number of hours slept and diastolic blood pressure.
Results suggests that adult males who sleep less on average tend to have a higher diastolic blood pressure. E1 4 in context

AQA_JUNE_2007_2

2(a)	From calculator $r = 0.915$ (0.91456)	B3	3	AWRT B2 for 0.914 or 0.91 – 0.92 B1 for 0.9
	or $r = \dfrac{2102.57 - \left(\dfrac{135.2 \times 147.9}{10}\right)}{\sqrt{128.976} \times \sqrt{98.269}}$ $= \dfrac{102.962}{11.35 \times 9.913}$ $= 0.915$			**Alternative:** $n = 10$ $\sum x = 135.2$ $\sum y = 147.9$ $\sum x^2 = 1956.88$ $\sum y^2 = 2285.71$ $\sum xy = 2102.57$ (M1) sub in formula (m1) (A1)
(b)	H_0 $\rho = 0$ H_1 $\rho > 0$ 1 tail 1% sig level	B1		Or words
	test statistic $r = 0.915$ cv = 0.7155 n = 10 since ts > 0.7155	B1 M1		For cv For comparison ts/cv
	Reject H_0	A1		
	Significant evidence at 1% level to suggest a positive linear association between the weight gain of mothers during pregnancy and the weight of their children at 3 years of age	E1	5	In context (ft)

2(a)

Country	A	B	C	D	E
x rank	1	2	3	4	5
y rank	6	5	4	9	2
Country	F	G	H	I	J
x rank	6	7	8	9	10
y rank	8	10	7	3	1

M1 attempt at ranks (can be reversed)

M1 for 16 correct

A1

$r_s = -0.212$ (3 sf from calc) B3 6 Award B2 for –0.22 ~ –0.20, B1 for –0.2, but B0 for –0.189 (PMCC)

Alternative

$d = 5, 3, 1, 5, 3, 2, 3, 1, 6, 9$

$\sum d^2 = 200$ (B1)

$r_s = 1 - \frac{6\times 200}{10\times 99}$ (M1)

$= 1 - 1.212 = -0.212$ (A1)

(b) H_0 Rank orders of annual road deaths and number of motor vehicles are independent. H_0 no association

H_1 Rank orders of annual road deaths and number of motor vehicles are not independent – there is an association B1 H_1 some association

2 tail 10%

cv = ±0.5636 $n = 10$ 2 tail 10% B1 for cv

test stat $r_s = -0.212$

$r_s > -0.5636$ M1 for comparison ts/cv; needs r_s correct 2sf Allow $r_s = 0.212$, cv = 0.5636 but not if signs are different

Accept H_0 No significant evidence at 10% level to suggest an association between rank orders of annual road deaths and number of motor vehicles for countries in the EU. A1

E1 5 **SC** –0.189 used can earn max B1B1M1

AQA_JUNE_2009_1

1(a)

Rank x	Rank y
10	8
9	6
8	9
7	10
6	2
5	5
4	7
3	4
2	1
1	3

M1 attempt at ranks inconsistent (can be reversed) SC M1M1 B2

M1 for 16 correct

A1

$r_s = 0.673$ (3 sf from calc) B3 6

AWRT B2 0.67
B1 0.7
ft B2 from wrong ranks (small slip)

No ranks seen, **SC** 0.67 B4
0.7 B3

alternative
$d = 2, 3, 1, 3, 4, . , 3, 1, 1, 2$
$\sum d^2 = 54$ B1

$r_s = 1 - \frac{6\times54}{10\times99} = 0.673$ M1, A1

(b)

H_o Rank orders of age and percentage body fat in females are independent. B1 or equivalent

H_1 Rank orders of age and percentage body fat in females are not independent – there is an association

2 tail 10%

cv = ±0.5636 $n = 10$ B1 for cv

test stat $r_s = 0.673$
$r_s > 0.5636$ M1 for comparison ts/cv
SC Allow M1 0.593/0.5494 (pmcc)

Reject H_o . Significant evidence at 10% level to suggest an association between rank orders of age and percentage body fat in females.(or positive association) E1 4 correct and in context