A Level Statistics

AQA Past Exam Questions

TOPIC: Hypothesis Testing
Wilcoxon Rank Sum Test

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 **17** questions in this question paper. The total mark for this paper is ***178***

• 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_JUNE_2015_3

(a) In carrying out a Wilcoxon Rank Sum test, random samples, each of size 6, are taken from two independent populations.

The complete set of 12 observations, no two of which are equal, is ranked in ascending order. The sum of the ranks for one of the samples is denoted by T.

Find the minimum possible value for T.

[2 marks]

(b) Two manufacturers of heat probes used in coal power stations are being compared by a company that supplies such probes to power stations.

Seven probes are obtained from manufacturer A and eight from manufacturer B.

The accuracy of each of the fifteen probes is measured and the probes are put into rank order for accuracy of measurement. Rank 1 indicates the least accurate probe.

The total, TA , of the ranks for manufacturer A is 46 and the total, TB , for manufacturer B is 74 .

Carry out a Wilcoxon Rank Sum test, at the 5% level of significance, to determine whether there is evidence of a difference between the accuracy of the probes from the two manufacturers.

[6 marks]

AQA_JAN_2012_6

A study of the effect of caffeine on muscle metabolism used 14 male volunteers who each underwent an arm-exercise test.

Seven of the men were randomly selected to take a capsule containing pure caffeine one hour before the test. The other seven men received a placebo capsule that had no active ingredient.

During each arm-exercise test, the subject's respiratory exchange ratio (RER) was measured. The question of interest to the experimenter was whether, on average, caffeine affects RER.

The results were as follows.

RER	
Men who have taken caffeine	Men who have not taken caffeine
106	119
99	105
96	101
94	100
93	97
89	95
88	94

Carry out the appropriate distribution-free test, using the 5% level of significance, to determine whether there is evidence of a difference, on average, between RER for men who have taken caffeine and that for men who have not.

(10 marks)

AQA_JAN_2013_5a

High levels of impurity in an alloy are a concern to the company that produces the alloy.

Two alternative processes, A and B, are used by the company for the production of the alloy. The production manager of the company, Jess, decided to compare the levels of impurity in ten randomly selected pieces of alloy made by process A and ten randomly selected pieces of alloy made by process B.

She wished to choose one of these two processes for the production of the alloy in the future. The results, in suitable units, of her investigation are given in Table 1. A higher reading indicates a higher level of impurity.

Table 1

Process A	3.65	3.50	3.45	3.15	3.00	2.95	2.90	2.85	2.80	2.65
Process B	3.35	3.25	3.05	2.75	2.65	2.60	2.55	2.20	2.15	2.05

(a) Carry out an appropriate distribution-free test, using the 5% level of significance, to determine whether there is a difference, on average, between the level of impurity in the alloy made by process A and that in the alloy made by process B.

(10 marks)

AQA_JUNE_2012_6

Jemma believed that the time taken for her to travel to work by bus differed according to the time of day. She worked different shifts on different days of the week, and travelled to work each day either for the morning shift, M, starting at 07.00 or for the afternoon shift, A, starting at 15.00 .
She recorded the bus journey times for her travel to work for seven randomly selected morning journeys and seven randomly selected afternoon journeys during a four-month period.
The results in the table are the times taken, x minutes, for the 14 journeys.

x	19.2	21.3	22.4	26.8	22.3	19.6	20.2	22.5	24.8	21.7	24.6	21.7	28.4	26.2
Shift	M	A	M	M	A	A	A	M	M	A	M	A	M	A

Carry out a non-parametric test, using the 5% level of significance, in order to investigate for a difference between Jemma's average journey time to work for the morning shift and that for the afternoon shift.

(11 marks)

AQA_JUNE_2013_3

A car manufacturer investigated a new device to see whether it would reduce CO2 emissions.
For this investigation, 12 new mid-sized cars, each with an engine capacity of 1.4 litres, were obtained and 6 of these cars were fitted with the new device. The other 6 cars were not fitted with the new device. The CO2 emission, in grams per kilometre, from each car was measured.
The results are given in the table.

Fitted with new device	139.1	134.6	128.9	139.8	129.5	140.9
Not fitted with new device	145.4	144.0	138.7	139.7	139.6	140.5

(a) (i) Carry out a distribution-free test, at the 5% level of significance, to investigate whether there is a reduction in the average CO2 emission for cars fitted with the new device***.***

(10 marks)

(ii) State the assumption that must be made regarding the 12 new cars in order for this test to be valid.

(1 mark)

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

(2 marks)

AQA_JUNE_2016_2

Twelve adults, all of whom suffered from moderate lower back pain, agreed to take part in an investigation into the effectiveness of a commonly recommended painkiller used for the relief of back pain. They were each given tablets and instructed to take one tablet four times a day.
They were also instructed to keep moving as much as possible and asked to record how many days it took for the pain to go away completely. The adults were randomly assigned to take either tablets containing the recommended painkiller or placebo tablets that contained no active ingredient.
The numbers of days taken by each patient for the back pain to go away completely are given in the table.

Recommended painkiller tablets	13	20	18	16	15	21
Placebo tablets	16	19	17	14	23	22

(a) Carry out a distribution-free test, using the 21 2 % level of significance, to investigate whether the patients taking the recommended painkiller had, on average, fewer days to wait for their back pain to go away completely.

[9 marks]

(b) Give a reason why all the adults were given tablets and the same instructions.

[1 mark]

AQA_JUNE_2017_3

A researcher investigated the reading ability of children aged seven years. The 16 children involved in the study were randomly selected from those aged seven years attending primary schools in a large town and with exactly one younger sibling but no older siblings.

The children were categorised according to whether their sibling was ‘less than two years younger’ or ‘at least two years younger’. The number of words that each child could successfully read from a list of 60 words was recorded.

The results are given in the table.

Sibling age									
Less than two years younger	21	23	28	20	19	16	18		
At least two years younger	22	26	29	17	25	28	24	27	24

(a) Carry out a Wilcoxon Rank Sum test, using the 5% level of significance, to investigate whether a child whose sibling is ‘less than two years younger’ can successfully read, on average, fewer words than a child whose sibling is ‘at least two years younger’

[9 marks]

(b) Explain how the researcher could obtain information to investigate whether a child aged seven years with more than one sibling can, on average, successfully read more words, from the same list of 60 words, than a child aged seven years with only one sibling.

[4 marks]

AQA_JUNE_2018_6a

(a) Twenty healthy female international flight attendants, aged between 26 years and 32 years were selected for an investigation into the effects of jet-lag. Each attendant had more than five years’ service with an airline company.

Ten of the flight attendants were randomly selected from those working for airlines with short recovery periods of 5 days or fewer between outbound long haul flights. The other ten flight attendants were randomly selected from those working for airlines with long recovery periods of more than 14 days between outbound long haul flights. Each flight attendant was asked to complete a task that assessed her reaction time two days into her recovery period.

The reaction time, in milliseconds, for each flight attendant is given in Table 1.

Table 1

Recovery period	Short	Long
	735	635
	783	638
	787	688
	802	712
	819	719
	849	769
	860	774
	902	786
	905	843
	927	903

Carry out a distribution-free test, using the 2.5% level of significance, to investigate whether female flight attendants with short recovery periods have, on average, slower reaction times than those with long recovery periods.

[9 marks]

AQA_JAN_2007_5

A supermarket fruit manager wishes to investigate the quality of bananas. Bananas are either delivered 'chilled' from source or 'stored at 10 ^{o}C' from source. The times to over-ripening of a random sample of 'chilled' bananas and of a random sample of 'stored at 10 ^{o}C' bananas are obtained and the rank values are given in the table.

(Rank 1 indicates the banana that was the fastest to become over-ripe.)

Chilled	2	6	5	8	1	
Stored at 10 ^{o}C	3	4	9	10	7	11

The supermarket fruit manager believes that 'chilled' bananas become over-ripe faster than bananas 'stored at 10 ^{o}C'. Carry out a Wilcoxon Rank Sum test, at the 5% level of significance, to investigate this belief. Interpret your conclusion in context.

(9 marks)

AQA_JAN_2008_2

As part of a Psychology research project, a student carried out a personality test on eight golfers and seven rugby players. All players involved in the project were selected at random.

The scores achieved are given in the table, with a higher score indicating a more outgoing personality.

Golfers	10	11	12	14	17	18	20	21
Rugby players	13	16	17	19	22	23	24	

Carry out a distribution-free test, at the 5% level of significance, to investigate whether rugby players have more outgoing personalities than golfers. Interpret your conclusion in context.

(10 marks)

AQA_JAN_2011_4

A company is investigating a new method of preparing coffee to package into pods for a coffee-making machine. Samples of pods are produced using the current method and also the new method. Cups of coffee are then made by the machine from each pod.

A coffee taster ranks the cups of coffee, with rank 1 indicating the best tasting coffee.

The results are given in the table.

Cup	Method	Rank
A	Current	10
B	Current	12
C	New	3
D	Current	6
E	New	2
F	New	7
G	New	1
H	Current	8
I	New	4
J	Current	11
K	Current	5
L	New	9

It is claimed by the company that the taste of cups of coffee made from pods produced using the new method is better than that using the current method. Carry out a Wilcoxon Rank Sum test, at the 5% level of significance, to investigate this claim.

(9 marks)

AQA_JUNE_2014_4

The marks of 6 randomly selected students from School A and of 6 randomly selected students from School B, in a Statistics module examination in June 2013, were ranked.

The rankings are given in the table.

Rank	1	2	3	4	5	6	7	8	9	10	11	12
School	B	B	B	A	B	B	A	B	A	A	A	A

Carry out a distribution-free test, using the 5% significance level, to investigate whether there is any difference in the average marks in the Statistics module examination for the two schools.

[9 marks]

AQA_JUNE_2007_5

Ms Testum wishes to investigate whether students will score differently in a test depending on whether the test is taken in a morning session or an afternoon session. She selects a group of 19 students of similar ability and randomly assigns some of them to take the test in the morning and the remainder to take the same test in the afternoon.

The students taking the test in the morning are kept apart from the students taking the test in the afternoon until all the students have taken the test.

The ordered scores are given in the table.

Session	Score
morning	44
afternoon	46
afternoon	47
afternoon	51
morning	53
morning	54
morning	56
afternoon	58
afternoon	59
afternoon	61
afternoon	62
morning	63
morning	63
morning	65
afternoon	67
afternoon	68
morning	72
morning	74
morning	81

(a) Carry out a Wilcoxon Rank Sum test, at the 5% level of significance, to investigate whether there is any difference in the average test score between mornings and afternoons. Interpret your conclusion in context***.***

(10 marks)

(b) A matched-pairs design was suggested for this investigation.

(i) Explain why a matched-pairs design might be preferred when comparing two groups.

(2 marks)

(ii) Explain how Ms Testum tried to ensure that her test was not biased.

(2 marks)

AQA_JUNE_2008_5a

The LDL cholesterol level was measured for each of 16 males living in the USA in 2006.
Of these, 8 had been randomly selected from males aged under 30 years and 8 had been randomly selected from males aged over 50 years.
The age and the LDL cholesterol level, in mg/dl, for each male are given in the table.

Male	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Age	29	18	29	28	23	19	21	27	56	54	51	52	71	65	54	76
LDL	121	137	140	159	177	189	191	201	181	196	225	228	234	249	259	339

(a) Carry out a Mann-Whitney U test, at the 5% level of significance, to investigate whether, in the USA, males aged under 30 years have, on average, a lower LDL cholesterol level than those aged over 50 years.

(10 marks)

AQA_JUNE_2009_3

A coin expert carries out an analysis to determine the percentage of silver in coins taken from two separate coin mintings during the reign of King Manuel I.
The percentages for the coins in a sample from each minting are given in the table.

First Minting	5.8	6.6	6.3	6.9	7.5	7.0	6.7	6.1
Second Minting	6.7	8.8	6.5	8.2	9.4	9.1	8.4	

Carry out a distribution-free test to investigate the claim that coins from the second minting contain a higher percentage of silver than those from the first minting. Use the 5% level of significance and assume each sample to be random.

(10 marks)

AQA_JUNE_2010_4

(a) The Wilcoxon Rank Sum test is to be used to compare two populations.
A random sample of size 8 is taken from each of the two populations. The complete set of sixteen sample values is ranked in ascending order.
It is known that no two sample values are equal. The sum of the ranks for one of the samples is denoted by T.
Find the minimum and the maximum possible values for T.

(4 marks)

(b) A sample of 18 children, born during 1999, was obtained. Six of these children, Group A, had older siblings only and so were the youngest in their family. The other 12 children, Group B, were either an only child or not the youngest in their family. The heights of all 18 children were measured and were then ranked from 1 to 18, with the smallest given rank 1. The total, TA , of the ranks for Group A was 31 and the total, TB , of the ranks for Group B was 140.
(i) Carry out a Wilcoxon Rank Sum test, at the 5% level of significance, to determine whether there is evidence of a difference in heights between children who are the youngest in their family and those who are either an only child or not the youngest in their family. The sample of children may be regarded as random.

(6 marks)

(ii) Interpret your conclusion in context.

(1 mark)

AQA_JUNE_2011_1

Two drugs, A and B, can each be used to reduce levels of a particular substance in the blood of adults. Fifteen adults all had a similar high level of the substance in their blood.

Of these adults, 8 took drug A for six weeks and 7 took drug B for six weeks. At the end of the six weeks, the level of the substance in a fixed volume of each adult's blood was measured.

The results are given in the table.

Drug A	Drug B
0.43	0.51
0.53	0.55
0.60	0.65
0.69	0.71
0.44	0.49
0.59	0.62
0.65	0.67
0.68	

(a) Carry out a Wilcoxon Rank Sum test, at the 10% level of significance, to investigate whether there is any difference in the average level of the substance in the blood after six weeks of taking either drug A or drug B.

(10 marks)

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

(2 marks)