National Qualifications 2026

X803/77/11

Statistics Paper 1

FRIDAY, 22 MAY

9:00 AM – 10:00 AM

Total marks — 30

Attempt ALL questions.

You may use a calculator.

To earn full marks you must show your working in your answers.

State the units for your answer where appropriate.

Write your answers clearly in the spaces provided in the answer booklet. The size of the space provided for an answer is not an indication of how much to write. You do not need to use all the space.

Additional space for answers is provided at the end of the answer booklet. If you use this space you must clearly identify the question number you are attempting.

Use **blue** or **black** ink.

You must leave your answer booklet on your desk; if you do not, you could lose all the marks for this paper.

You may refer to the Statistics Advanced Higher Statistical Formulae and Tables.

Qualifications Scotland
Teisteanasan Alba

X8037711

A/HTP

[BLANK PAGE]

DO NOT WRITE ON THIS PAGE

MARKS

Total marks — 30

Attempt ALL questions

1. During the 2015 Rugby Union Men's World Cup, data was collected on each player from each country's squad. The data included:

   - the country the player represented
   - the number of times they had represented that country in international competitions
   - the player's height (in cm) and mass (in kg)
   - whether they played in a 'forward' position or a 'back' position.

   Output 1 shows the 132 players who played for four countries' squads and the number of times they have represented that country internationally. For example, there were 7 English players who had played internationally between 1 and 15 times.

   Output 1

	1–15	16–30	31–45	46–60	>60
England	7	12	8	4	1
Ireland	10	6	6	5	6
Scotland	14	9	3	4	2
Wales	13	4	6	5	7

   (a) State the null and alternative hypotheses of a chi-squared test of association applied to the table in **Output 1**. 1

   The expected values under the null hypothesis were calculated and are displayed in **Output 2**. One value has been deleted and replaced with *****.

   Output 2

	1–15	16–30	31–45	46–60	>60
England	10.67	7.52	*****	4.36	3.88
Ireland	11.00	7.75	5.75	4.50	4.00
Scotland	10.67	7.52	5.58	4.36	3.88
Wales	11.67	8.22	6.10	4.77	4.24

   (b) Calculate the expected value marked by ***** in **Output 2**. 1

   (c) Give a reason why it is not appropriate to perform a chi-squared test of association on this data. 1

[Turn over

1. **(continued)** MARKS

Output 3 shows the same data reported in **Output 1** but with one less grouping of the number of times players had played internationally.

Output 3

	1-15	16-30	31-45	>45
England	7	12	8	5
Ireland	10	6	6	11
Scotland	14	9	3	6
Wales	13	4	6	12

The output from a chi-squared test applied to the table in **Output 3** is shown in **Output 4.**

Output 4

```
Chi-squared test of association:
X-squared = 13.942, df = 9
```

(d) Using the information in **Output 4**, write down an appropriate decision and conclusion for this test. **3**

It was also of interest to examine the relationship between the masses and heights of players playing in 'forward' and 'back' positions separately. **Figure 1** was produced for this purpose.

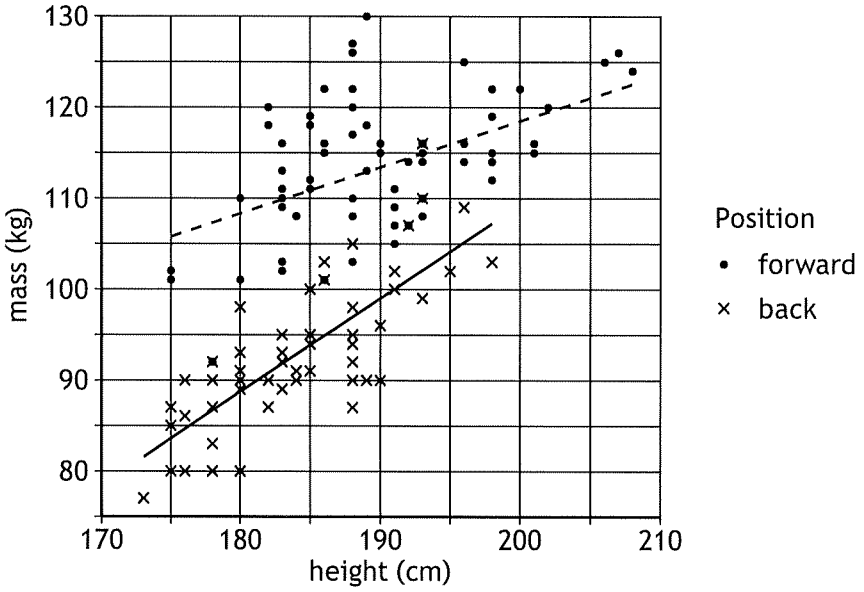


Figure 1 Scatterplot of masses and heights of players, their playing positions identified, and the associated least squares regression models

(e) (i) Looking at the points for 'back' position players in **Figure 1**, comment on the relationship shown between mass and height. **1**

(ii) By comparing the **slopes** of the two least squares regression lines, interpret the relationships between the masses and heights of the 'forward' and 'back' position players. **1**

MARKS

1. (continued)

Output 5 summarises the two least squares regression models from the scatterplot in **Figure 1**. The results from conducting tests on the null hypothesis that the slope parameter is zero are also included.

Output 5

```
Position = Forward
Mass = 40.75685 + 0.37567 Height
            Estimate   Std. Error    t-value     p-value
Intercept    40.75685
Slope         0.37567     0.05787      6.492     <0.00001

Position = Back
Mass = -87.7474 + 0.9821 Height
            Estimate   Std. Error    t-value     p-value
Intercept   -87.7474
Slope         0.9821      0.05929     16.564     <0.00001
```

(f) Using the 'forward' information in **Output 5**, interpret the p-value at the 5% significance level and conclude whether there is evidence of the model being useful for predicting a 'forward' position player's mass from their height. **2**

A 'forward' position player with height 202 cm was being considered to join one of the squads. The appropriate fitted model for the player's mass in **Output 5** was used to calculate the 95% prediction interval and the 95% confidence interval relevant to this player. Both intervals are shown in **Output 6**. Two equal values have been deleted and replaced with *****.

Output 6

```
Prediction Interval: Position=Forward & Height=202cm
       fit      lower      upper
     *****   101.2221   132.0617

Confidence Interval: Position=Forward & Height=202cm
       fit      lower      upper
     *****   114.9216   118.3622
```

(g) Calculate the fitted value marked by ***** in **Output 6**. **1**

(h) Interpret the prediction interval and the confidence interval shown in **Output 6** in this context. **2**

(i) Using the information in **Figure 1**, comment on the suitability of using the 'back' position regression model to estimate the mass of the player with height 202 cm. **1**

[Turn over

2. An extract of a draft report by a researcher is given below.

It is known to contain some flaws and questionable methodology.

Read it and then answer the questions that follow.

Introduction

The Ministry of Transport (MOT) test is one that every road vehicle older than 3 years has to pass every year. Since April 2010, the maximum cost allowed to be charged for a car MOT test has been set by the UK Government at £54.85, but many garages charge less than this. In general, there are two types of public garage that can conduct a MOT test: an independent garage, and one that is part of a national chain of garages. This report will research whether national chain garages charge less for a MOT test than the independent garages.

Method

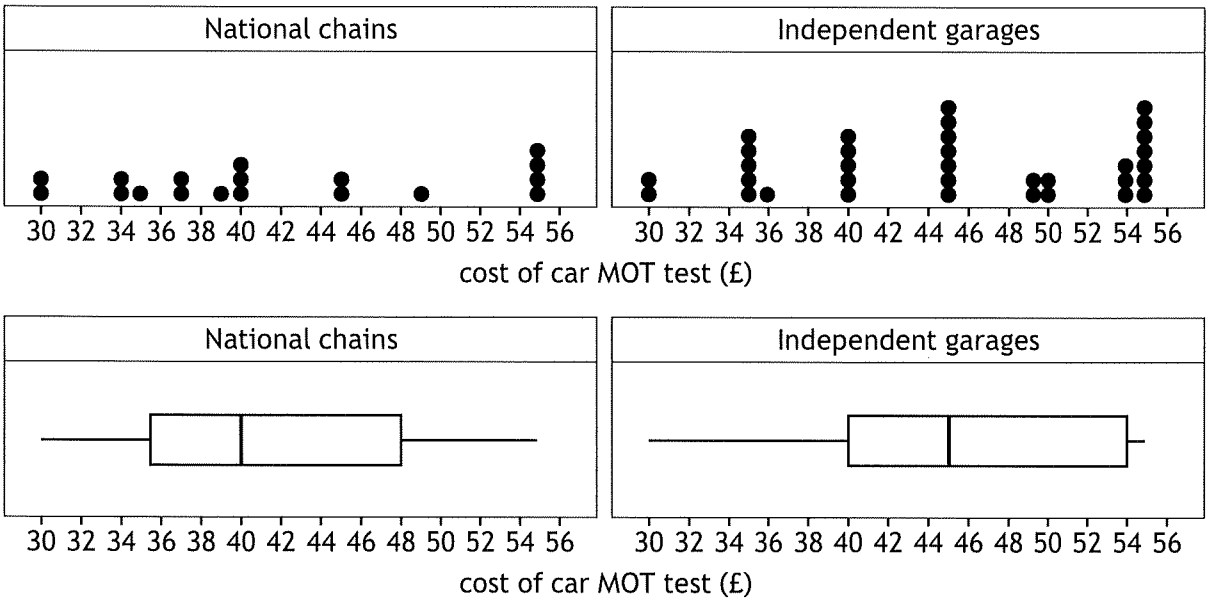
A full list of all 23 158 active MOT testing centres was downloaded from the UK Government website, that included the garage name, postal address, postcode, phone number and type of vehicle they tested.

The 18 939 garages that appeared only once on the list were taken to be independent garages. A simple random sample was taken by randomly selecting 34 garages from the list of independent garages and their postcodes noted. Using a MOT cost search website, the independent garage that was closest to the noted postcode had its cost for an MOT test recorded.

This searching process meant that the recorded MOT cost may not have been for the intended randomly selected garage, because not all independent garages were listed on the MOT cost search website. However, it was still effectively a randomly selected independent garage based on a postcode.

It was decided that garage names that appeared more than 10 times on the list were part of national chains, so this identified 22 national chains. The cost of a MOT test was only able to be found for 18 of these chains, without actually booking in a car for a real MOT test.

Data



2. (continued)

Analysis

For each sample, computer software was used to generate a 95% confidence t-interval for the mean cost of a car MOT test:

```
data: national chains
95 percent confidence interval: (37.61143, 46.19190)

data: independent garages
95 percent confidence interval: (42.18428, 47.83631)
```

Interestingly, neither of these intervals captured the Government's maximum cost of a car MOT test of £54.85, but the two intervals overlapped between the costs of £42.18 and £46.19.

It was thought that not enough information was available to perform a two-sample z-test for the difference in population means, so computer software was used to perform both a two-sample t-test and a Mann-Whitney test using a normal approximation, and the results are shown below.

```
data: national chains and independent garages
t = -1.2876, df = 50, p-value = 0.1019
alternative hypothesis: true difference in means is less than 0
sample estimates:
mean of national chains      mean of independent garages
 41.90167                     45.01029
stdev of national chains     stdev of independent garages
  8.627261                     8.099399

data: national and independent
m = 18, n = 34, W = 404, p-value = 0.0823
alternative hypothesis: true difference in medians is less than 0
```

Conclusion

At the 10% significance level, the Mann-Whitney test suggests that the mean cost for a car MOT test is different depending on whether an independent or national chain garage is used.

[Turn over

2. **(continued)** MARKS

Refer to lines 9 to 21.

The researcher recognises a potential shortcoming in their sampling process.

(a) Describe which of the 18 939 independent garages could not be included in the random sample, and suggest what the researcher should have done to find out the cost that these garages would charge for a car MOT test. 2

Read lines 29 to 31.

(b) Assuming that the required conditions for the calculation of confidence intervals are valid, explain why a maximum value would not be expected to fall within a 95% confidence interval. 1

Read lines 32 to 35.

(c) Describe the numerical information that was indeed missing that prevented a z-test from being performed. 1

The two sample t-test requires the pooling of the samples' standard deviations to obtain the best estimate for the population standard deviation of cost.

(d) Calculate this pooled standard deviation, and state the assumption which is being made when performing this calculation. 2

(e) Write down the calculations for the z-value that would give the p-value of 0.0823 for the Mann-Whitney test. 3

Examine the computer output after line 35.

(f) Using a 10% level of significance will give a different conclusion for each of the tests. Describe the information in the output that leads to this difference. 1

Read the Conclusion, lines 37 to 39.

The Conclusion contains two mistakes.

(g) Identify and describe what the mistakes are. 2

A reviewer of the report has concerns about the validity of the required assumption that the samples must have each come from a normally distributed population.

(h) (i) Describe what the reviewer must have identified from the dotplots and/or the boxplots that concerned them. 1

(ii) Write down which parts of the draft report's analysis would no longer be valid, without the normality assumption of costs being assumed. 2

Read lines 22 to 23.

(i) Describe a feature of the gathered sample of national chains that potentially further undermines the statistical inference procedures conducted in the Analysis section of the draft report. 1

[END OF QUESTION PAPER]