ANSWER BOOK

Non-parametric tests

Further Statistics 2 · Year FM

15 questions · 67 marks

NAVY EDITION

Every question in this book has a code beside its number.
Type inklearning.co.uk/q/ followed by the code to open that question online.

inklearning.co.uk/maths/ · youtube.com/@InkPhysics

SECTION A · Warm-up

A1 NP-A1 [2]

A non-parametric, or distribution-free, test assumes no family of distributions for the population it is testing; it works from the signs or the ranks of the readings, and its hypotheses concern a median, or the identity of two populations, rather than the parameters of a named distribution. B1.

Any one reason, B1: the population is plainly not normal, for instance strongly skewed incomes or waiting times; the sample is too small for normality to be worth checking; an outlier would drag a mean about; or the data arrive as ranks, preferences or ordered categories with no scale attached.

The trade is power: where the population really is normal, the t-test detects a smaller shift from the same sample size, because it uses the readings themselves and not their order.

A2 NP-A2 [3]

H_0: median = 25; H_1: median > 25. One-tailed at the 5% level. B1 for both hypotheses, written about the median and not about a mean.

Under H_0 each bag is above the median with probability ½, so the number above is B(11, ½).

$P(X \geq 9) = (55 + 11 + 1)/2048 = 67/2048 =$ **0.0327**. M1 for the binomial tail, including all three terms.

Since $0.0327 < 0.05$, **reject** H_0. There is evidence at the 5% level that the median mass exceeds 25 kg. A1 for the conclusion in context.

No table is needed anywhere in this test: the null distribution is binomial, which is why the sign test survives without a booklet of critical values.

A3 NP-A3 [2]

H_0: the population median takes the claimed value m_0. B1. Written about the median, not about a mean: the sign test says nothing about means, and for a skewed population the two are different numbers.

If m_0 really is the median, then any one reading falls above it with probability ½ by the definition of a median, and the readings of a random sample are independent of one another. A fixed number of independent trials, each with probability ½ of a success, is exactly B(n, ½). B1.

A4 NP-A4 [3]

The deviations, in the order given, are +1.2, −1.1, +2.6, −0.1, +0.4, +3.1, −2.3 and +1.9. M1 for subtracting 13.0 from each reading.

Ranking their **sizes**, smallest first, gives 4, 3, 7, 1, 2, 8, 6 and 5. A1.

The three negative deviations carry ranks 3, 1 and 6, so $T_- =$ **10** and $T_+ =$ **26**. A1.

Check: $10 + 26 = 36 = 8 \times 9/2$, which is what the eight ranks must total. Rank the sizes and then restore the signs; ranking the signed deviations themselves puts −2.3 at rank 1 and produces neither total.

SECTION B · Standard

B1 NP-B1 [3]

(a) **Wilcoxon** rank-sum test. Two independent groups with no partner for any reading, so the samples are unpaired; the two lists happening to be the same length settles nothing, and house prices are skewed enough that a test about means would be hard to justify. B1.
(b) **Sign** test on the twelve pairs. Each judge supplies a pair, so the data are paired, but the record is a direction with no size attached, so there are no deviations to rank and the signed-rank test has nothing to work with. B1.
(c) **Single-sample** Wilcoxon signed-rank test. One sample against a claimed median, with the deviations measured, so their sizes can be ranked and the test uses more of the data than a sign test would. B1.
Give the reason as well as the name; the reason follows from how the data were collected.

B2 NP-B2 [6]

H_0: median = 50; H_1: median > 50. One-tailed at the 5% level, $n = 9$. B1.
The deviations from 50 are +2.4, -0.7, +5.3, -2.1, +3.8, -0.8, +1.5, +6.4, +5.1. M1 for subtracting the claimed median from each reading.
Ranking their **sizes** from 1 for the smallest gives 5, 1, 8, 4, 6, 2, 3, 9, 7. M1 A1. Rank the absolute deviations, then restore their signs.
The three negative deviations carry ranks 1, 4 and 2, so $T_- =$ **7** and $T_+ = 38$. Check: $7 + 38 = 45 = 9 \times 10/2$.
H_1 says the deviations run positive, so it is T_- that should be small, and $T_- = 7$ is the statistic. Since $7 \le 8$, **reject** H_0. There is evidence at the 5% level that the median fill exceeds 50 g. M1 for the comparison, A1 for the conclusion in context.
A small value is the evidence here, which is the reverse of every other test in this unit. For the stated upper-tailed alternative, use the smaller negative-rank total and compare it with the supplied critical value.

B3 NP-B3 [6]

The samples are **paired**: each operator supplies a before figure and an after figure, so each reading has a partner and the two lists are in the same order. B1. Treating them as two independent samples of 10 would throw the pairing away.
H_0: the median difference is 0; H_1: it is positive. One-tailed at the 5% level, n = 10 pairs.
The differences, after minus before, are +14, −3, +22, +7, −2, +11, +18, −5, +26, +9. M1.
Ranking their sizes gives 7, 2, 9, 4, 1, 6, 8, 3, 10, 5, so the three negative differences carry ranks 2, 1 and 3. M1 A1.
$T_- = 2 + 1 + 3 = \mathbf{6}$ and $T_+ = 49$; check $6 + 49 = 55 = 10 \times 11/2$.
Since $6 \le 10$, **reject** H_0. There is evidence at the 5% level that output has risen. M1 A1 for the comparison and the conclusion in context.
The sign test on the same data counts 7 rises and 3 falls, giving $P(X \ge 7) = 176/1024 = 0.172$ under $B(10, ½)$ and no significant result. The three operators who slipped slipped by 2, 3 and 5 units while the risers rose by up to 26, and only the test that ranks the sizes can use that.

B4 NP-B4 [4]

Each taster supplies a pair and records only a direction, so this is a **paired** sign test. Let p be the probability that a taster prefers brand A. H_0: $p = ½$; H_1: $p \ne ½$. B1 for the hypotheses.
Under H_0 the number preferring A is $B(12, ½)$, and 9 is above the mean of 6, so the upper tail is the one to measure:
$P(X \ge 9) = (220 + 66 + 12 + 1)/4096 = 299/4096 = \mathbf{0.0730}$. M1 for the binomial tail with all four terms, A1 for 0.0730.
The test is two-tailed at 5%, so the tail probability is compared with 0.025. Since $0.0730 > 0.025$, **do** not reject H_0. There is insufficient evidence at the 5% level that tasters prefer one brand to the other. A1 for the conclusion in context.
Compare the single tail with half the significance level, or double the tail and compare with the whole of it; doing neither turns a two-tailed test at 5% into a one-tailed test at 10%, which here would have rejected. Ten preferences rather than nine would have given a tail of 0.0193 and the opposite verdict.

B5 NP-B5

Each pair of twins supplies one difference, so the samples are **paired**, and the test that uses the sizes as well as the signs is the **Wilcoxon** matched-pairs signed-rank test. B1. [5]

H_0: the median difference is 0; H_1: it is negative, since a hint that helps makes the time with the hint the smaller. One-tailed at 5%, $n = 7$.

Ranking the sizes 3, 5, 7, 9, 12, 18, 21 from 1 for the smallest, the differences take ranks 5, 2, 6, 1, 4, 7 and 3. M1 for ranking the sizes, A1 for the ranks.

The one positive difference carries rank 1, so $T_+ = \mathbf{1}$ and $T_- = 27$; check $1 + 27 = 28 = 7 \times 8/2$. M1 for T_+ as the statistic.

H_1 says the differences run negative, so it is T_+ that should be small, and small values are the evidence. Since $1 \le 3$, **reject** H_0. There is evidence at the 5% level that the hint reduces the time taken. A1 for the conclusion in context.

Take the total the alternative hypothesis pushes downwards; using $T_- = 27$ here and comparing it with 3 reverses the test.

B6 NP-B6

The two groups have no partner for any reading and are of different sizes, so they are independent and the Wilcoxon rank-sum test applies. H_0: the heights come from the same population; H_1: they do not. B1. [5]

Rank all eleven heights together, from 1 for the shortest: 28(1), 34(2), 38(3), 39(4), 41(5), 44(6), 47(7), 49(8), 52(9), 55(10), 57(11). M1 for one ranking over the pooled data.

W is the rank total of the smaller sample, so for variety P, $W = 2 + 5 + 1 + 4 + 7 = \mathbf{19}$. A1. Check: variety Q's ranks total 47, and $19 + 47 = 66 = 11 \times 12/2$.

Since $18 < 19 < 42$, the statistic lies between the two critical values, so **do** not reject H_0. There is insufficient evidence at the 5% level that the two varieties differ in height. M1 for the comparison, A1 for the conclusion in context.

Under H_0 the mean of W is $5 \times 12/2 = 30$, so 19 is well below what chance alone would produce, and it misses the critical value by one. Report that as a near miss rather than as evidence that the two varieties are alike.

B7 NP-B7

(a) **Sign** test. One sample against a claimed median, and the record holds directions only, so there are no deviations whose sizes could be ranked. B1. [3]

(b) **Wilcoxon** matched-pairs signed-rank test. Each student supplies a pair, so the data are paired and reduce to twelve differences, and the marks themselves are known, so the sizes of those differences can be ranked. B1.

(c) **Wilcoxon** rank-sum test. No reading in one group has a partner in the other, and the unequal sizes settle that at once, so there is nothing to subtract and the pooled ranking is the only route. B1.

The reason comes from how the data were collected and what each reading carries, and it is settled before any arithmetic starts.

SECTION C · Stretch

C1 NP-C1 [7]

The samples are of different sizes, so they are **not** paired and the Wilcoxon rank-sum test is the one that fits. B1.
H_0: the yields come from the same population; H_1: they do not. Two-tailed at the 5% level, $m = 6$ and $n = 8$. B1. The hypotheses are about the populations being identical, not about two means.
Ranking all fourteen yields together from 1 for the smallest gives fertiliser A the ranks 1, 2, 3, 5, 6 and 10, and fertiliser B the ranks 4, 7, 8, 9, 11, 12, 13 and 14. M1 A1 for one ranking over the pooled data. Pool the observations before assigning ranks; separate rankings do not produce the Wilcoxon rank-sum statistic.
W is taken from the smaller sample, so $W = 1 + 2 + 3 + 5 + 6 + 10 = \mathbf{27}$. Check: the other total is 78, and $27 + 78 = 105 = 14 \times 15/2$. M1.
Under H_0 the mean of W is $6 \times 15/2 = 45$, so 27 lies well below what chance alone would produce.
Since $27 \leq 29$, **reject** H_0. There is evidence at the 5% level that the two fertilisers give different yield distributions, with A the lower. A1 A1 for the comparison and the conclusion in context.
The two critical values add to $m(m + n + 1) = 6 \times 15 = 90$, since $29 + 61 = 90$; the distribution of W is symmetric about 45, so a table printing one end gives the other.

C2 NP-C2 [6]

H_0: the median difference is 0; H_1: it is positive. One-tailed at the 5% level, so the critical value is -1.645. B1.
For $n = 25$ the signed-rank statistic has mean $n(n + 1)/4 = 25 \times 26/4 = \mathbf{162.5}$ and variance $n(n + 1)(2n + 1)/24 = 25 \times 26 \times 51/24 = 33150/24 = \mathbf{1381.25}$, so the standard deviation is $\sqrt{1381.25} = 37.17$. M1 A1 for the mean and the variance.
H_1 says the differences run positive, so $T_- = 95$ is the statistic and it is the lower tail that matters. T takes whole values only, so $P(T \leq 95)$ becomes the area below 95.5.
$z = (95.5 - 162.5)/37.17 = -67/37.17 = \mathbf{-1.803}$. M1 A1 for the standardising with the half included.
Since $-1.803 < -1.645$, **reject** H_0. There is evidence at the 5% level that the median difference is positive. A1.
Without the correction z would be $-67.5/37.17 = -1.816$, the same verdict from a slightly overstated statistic. The correction moves z towards zero, so leaving it out tilts the test towards rejecting; here it changes the third decimal place, not the conclusion. Note also that a two-tailed test at 5% would compare 1.803 with 1.96 and would **not** reject, so the tail has to be settled before the critical value is quoted.

C3
NP-C3

H_0: the two samples come from the same population; H_1: they do not. [6]
Two-tailed at 5%, so the critical values are ±1.96. B1.
Take the smaller sample first, so $m = 15$ and $n = 18$. Under H_0, W has mean $m(m + n + 1)/2 = 15 \times 34/2 =$ **255** and variance $mn(m + n + 1)/12 = 15 \times 18 \times 34/12 = 9180/12 =$ **765**, so the standard deviation is $\sqrt{765} = 27.66$. M1 A1 for the mean and the variance.
$W = 200$ lies below the mean, so it is the lower tail that matters. W takes whole values only, so $P(W \le 200)$ becomes the area below 200.5.
$z = (200.5 - 255)/27.66 = -54.5/27.66 =$ **−1.970**. M1 A1 for standardising with the half included.
Since $-1.970 < -1.96$, **reject** H_0. There is evidence at the 5% level that the two populations differ, with the sample of 15 taking the lower ranks. A1.
Without the correction z would be $-55/27.66 = -1.989$: the same verdict from a statistic that overstates the case. The correction always moves z towards zero, so leaving it out tilts a test towards rejecting, and here the margin over 1.96 is thin enough for that to matter to how the result is reported.

C4
NP-C4

The smallest total the sample of 7 could take is $1 + 2 + \ldots + 7 =$ **28**, when its readings are the seven lowest, and the largest is $10 + 11 + \ldots + 16 =$ **91**, when they are the seven highest. B1 for both. [6]
Those two extremes add to $119 = m(m + n + 1) = 7 \times 17$, and the distribution of W is symmetric about its mean of $119/2 = 59.5$, so the two critical values add to 119 as well. M1 for $m(m + n + 1)$.
The upper critical value is therefore $119 - 40 =$ **79**. A1.
H_0: the two samples come from the same population; H_1: they do not. Since $40 < 42 < 79$, the statistic lies between the critical values, so **do** not reject H_0. There is insufficient evidence at the 5% level that the two populations differ. A1 for the conclusion in context.
$U = 42 - \tfrac{1}{2}(7)(8) = 42 - 28 =$ **14**. M1 for the subtraction, A1 for 14 with what it counts: taking the smallest possible total off W shifts the origin to zero, and what is left counts the pairs, one reading from each sample, in which the reading from the sample of 7 is the larger. Out of $7 \times 9 = 63$ such pairs, 14 go that way.
U and W are the same test written from different origins, so a table for one converts to a table for the other by the same shift of 28.

THIS BOOK ONLINE

Scan the code, or type the address, to open the questions in this book online:

inklearning.co.uk/maths/practise/questions/non-parametric-tests/