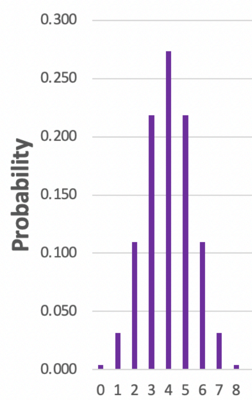
The single-sample sign test

E.g. Leona plays golf where the aim is to take as few shots as possible. Her median number of shots for a round is 85. After some lessons from the golf pro, she plays eight rounds in the following number of shots:

$$82 \quad 78 \quad 89 \quad 84 \quad 74 \quad 91 \quad 80 \quad 77$$

(a) Test at the $10\,\%$ level to see if Leona has improved her game.

(b) If in (a) H_0 is not rejected, state how many rounds would have had to be below 85 in order for the the test to show significance.



Working: (a) H_0 : Median is 85

H_1 : Median is less than 85

Let X be the number of values below 85

so $X \sim \mathrm{B}(8, 0.5)$

The signs of deviations from the median are:

$- \quad - \quad + \quad - \quad - \quad + \quad - \quad -$

$p = P(X \geq 6) = P(X \leq 2) = 0.145$

Since $p = 0.145 \nleq 0.1$, we do not reject H_0.

There is evidence to suggest that Leona has not improved as a golfer.

(b) $P(X \geq 7) = P(X \leq 1) = 0.0352 \leq 0.1$

So if 7 of Leona's rounds of golf were below 85, there would have been evidence at the $10\,\%$ (and, in fact, the $5\,\%$ level) to suggest she had improved.

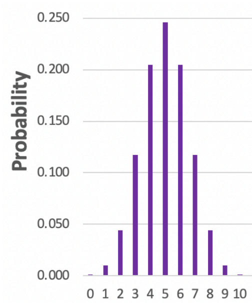
N.B. $X \geq 7$ is the critical region of the test.

E.g. 1 An airline ran a course for ten people who have a fear of flying. They were asked to rate the course on a scale of 1 (very dissatisfied) to 10 (very satisfied). The results were:

$$1 \quad 2 \quad 4 \quad 8 \quad 8 \quad 8 \quad 9 \quad 9 \quad 10 \quad 10$$

(a) Use a single-sample sign test to test, at the $10\,\%$ significance level, the null hypothesis $M = 6$, against the alternative hypothesis $M > 6$, where M is the median rating given for such courses.

(b) What is the minimum number of values that need to be higher than 6 so that the null hypothesis is rejected at the $5\,\%$



Working: (a) H_0 : Median is 6

H_1 : Median is more than 6

Let X be the number of values below 6

so $X \sim \mathrm{B}(10, 0.5)$

The signs of deviations from the median are:

$- \quad - \quad - \quad + \quad + \quad + \quad + \quad + \quad + \quad +$

$p = P(X \geq 7) = P(X \leq 3) = 0.172$

Since $p = 0.172 \nleq 0.1$, we do not reject H_0.

There is no evidence to suggest the median is above 6.

(b) $P(X \geq 8) = P(X \leq 2) = 1 - 0.9453 = 0.0547 \nleq 0.05$

$P(X \geq 9) = P(X \leq 1) = 0.0107 \leq 0.05$.

So there would have to be 9 ratings above 6 for the median of 6 to be rejected at the $5\,\%$ level.

E.g. 2 The times, in minutes, spent travelling to school by the 15 members of a statistics class are as follows:

$13 \quad 23 \quad 15 \quad 21 \quad 22 \quad 18 \quad 30 \quad 25 \quad 45 \quad 12 \quad 17 \quad 24 \quad 32 \quad 28 \quad 29$

The school prospectus claims that 'half of the pupils live less than 20 minutes from the school'.

(a) Use a one-tailed sign test, at the $5\ \%$ level, to test whether the median is higher than that stated by the school.

(b) Find the critical region of this test.

Working: (a) H_0 : Median is 20 minutes

H_1 : Median is above 20 minutes

Let X be the number of values above 20 so $X \sim \text{B}(15, 0.5)$.

The signs of deviations from the median are:

$- \quad + \quad - \quad + \quad + \quad - \quad + \quad + \quad + \quad - \quad - \quad + \quad + \quad + \quad +$

$p = P(X \geq 10) = P(X \leq 5) = 0.151$

Since $p = 0.151 \nleq 0.1$, we do not reject H_0.

There is evidence to suggest that the school's claim is correct.

(b) $P(X \geq 11) = P(X \leq 4) = 0.592 \nleq 0.05$

$P(X \geq 12) = P(X \leq 3) = 0.0176 \leq 0.05$

The critical region is 12 or more values more than the median.

E.g. 3 A paper company plans to buy a wood from a farmer who claims that the trees have an average of 15 cm³ of usable wood. The company employs a statistician to check this claim, as they believe the median is less than 15 cm³. With agreement of the farmer the statistician has a random sample of 13 trees felled and their usable wood collected. Here are the results:

$11.3 \quad 8.8 \quad 6.4 \quad 8.7 \quad 14.6 \quad 16.2 \quad 12.4 \quad 7.1 \quad 15.2 \quad 10.3 \quad 9.5 \quad 17.2 \quad 7.8$

What is the researcher's conclusion after conducting a test at the $5\ \%$ significance level?

Working: H_0 : Median is 15 cm³ of usable wood

H_1 : Median is less than 15

Let X be the number of values more than 15 so $X \sim \text{B}(13, 0.5)$

The signs of deviations from the median are:

$- \quad - \quad - \quad - \quad - \quad + \quad - \quad - \quad + \quad - \quad - \quad + \quad -$

$p = P(X \geq 10) = P(X \leq 3) = 0.0461$

Since $p = 0.0461 \leq 0.05$, the researcher rejects H_0.

There is evidence to suggest that the farmer's claim is not correct.

If data values equal the median, they are simply ignored and the sample size is consequently reduced (remember, this is not in the OCR syllabus).

E.g. 4* Applicants to university may have to wait before being invited to interviews. One university claims that the median waiting time for its applicants is 31 days. A random sample of 18 successful applicants waited for these times:

$$\begin{array}{ccccccccc} 33 & 32 & 35 & 30 & 33 & 37 & 30 & 31 & 37 \\ 29 & 31 & 31 & 32 & 34 & 30 & 31 & 33 & 38 \end{array}$$

Test, at the 10% level, the claim that the median is 31 days against the alternative that it is not 31 days.

Working:

H_0 : Median is 31 days
H_1 : Median is not 31 days
Removing the values that equal 31

$$\begin{array}{ccccccccc} 33 & 32 & 35 & 30 & 33 & 37 & 30 & & 37 \\ 29 & & & 32 & 34 & 30 & & 33 & 38 \end{array}$$

Let X be the number of values more than 31 so $X \sim \text{B}(14, 0.5)$
The signs of deviations from the median are:

$+ \quad + \quad + \quad - \quad + \quad + \quad - \quad + \quad - \quad + \quad + \quad - \quad + \quad +$

$p = P(X \geq 10) = P(X \leq 4) = 0.0898$
Since $p = 0.0898 \nleq 0.05$, we do not reject H_0.
There is evidence to suggest that the median is 31 days.

Video (password needed): Introduction to non-parametric tests
Video (password needed): The single-sample sign test
Video (password needed): Sign test with a single sample

Solutions to Starter and E.g.s

Exercise
p49 4A Qu 1-5, (6 red)