Hypothesis Testing (Binomial)

SCAN ME

REVISE THIS TOPIC

1 A game at a school fair states that 30% of players will win a prize.

Winnie claims that the statement is false and that in fact the true percentage of players who will win a prize is less than 30%.

Winnie observes the next 40 players of the game and records that 7 of those players win.

(a) Stating your hypotheses clearly, test Winnie's claim at a 5% level of significance. **(4)**

(b) For the test in part (a)

(i) write down the p-value **(1)**

(ii) find the critical region **(2)**

(iii) find the actual level of significance **(1)**

(a) $H_0 : p = 0.3$

$H_1 : p < 0.3$

Let X be the number of players who win a prize.

Under H_0, $X \sim B(40, 0.3)$

$P(X \leq 7) = 0.05528288716$

$0.05528288716 > 0.05$ so do not reject H_0

There is insufficient evidence to suggest that the true percentage of players who win is less than 30%

(b) (i) 0.05528288716

(ii) $P(X \leq 7) = 0.05528288716 > 0.05$ [leads to do not reject H_0]

$P(X \leq 6) = 0.0237609464 < 0.05$ [leads to reject H_0]

Critical region: $X \leq 6$

(iii) $P(X \leq 6) = 0.0237609464$

Actual level of significance = 0.02376 (or 2.376%)

(Total for Question 1 is 8 marks)

2 A fair 6-sided dice can land on the numbers 1, 2, 3, 4, 5, or 6.

Gemma rolls the dice once and records the number it lands on.

(a) Write down the name of the distribution that can be used to model the number that the dice lands on. **(1)**

Harpreet rolls the dice 10 times and records the number of times she rolls the number 6.

(b) Find the probability that Harpreet rolls the number 6

(i) once **(1)**

(ii) more than once **(2)**

Rauri suspects that the dice may not be fair and in fact be biased as the number 6

Rauri tests the dice by rolling it 90 times and he records the number 6 a total of 22 times.

(c) Carry out a suitable test to determine whether the dice is biased towards the number 6. You should state your hypothesis clearly and use a 5% level of significance. **(4)**

(a) Discrete uniform distribution

(b) (i) Let X be the number of times Harpreet rolls a number 6, then $X \sim B(10, \frac{1}{6})$

$P(X = 1) = 0.3230$

(ii) $P(X > 1) = 1 - P(X \leq 1) = 1 - 0.4845 = 0.5155$

(c) $H_0 : p = \frac{1}{6}$

$H_1 : p > \frac{1}{6}$

Let Y be the number of times Rauri rolls a number 6

Under H_0, $Y \sim B(90, \frac{1}{6})$

$P(Y \geq 22) = 1 - P(Y \leq 21) = 1 - 0.96229\ldots = 0.03770452053$

$0.03770452053 < 0.05$ so reject H_0

There is sufficient evidence to suggest that the dice is biased towards the number 6

(Total for Question 2 is 8 marks)

3 A revision website claims that 60% of the students who pay to use the site will achieve a grade A*

Sofia believes that the true proportion of students achieving an A* is less than 60%.

She collects data from a random sample of 30 students who used the website and finds that 13 of them achieved a grade A*.

(a) State suitable hypotheses to test Sofia's claim. **(1)**

(b) Carry out a hypothesis test at a 10% level of significance to determine whether there is evidence that the proportion of students achieving a grade A* is less than 60%. **(3)**

(c) Write down the p-value for the test in part (b) **(1)**

(d) Find the acceptance region for the test in part (b) **(2)**

(a) $H_0 : p = 0.6$

$H_1 : p < 0.6$

(b) Let X be the number of students in the sample who got an A*

Under H_0, $X \sim B(30, 0.6)$

$P(X \leq 13) = 0.04811171242$

$0.04811171242 < 0.1$ so reject H_0

There is sufficient evidence to suggest that the proportion of students achieving A* is less than 60%

(c) 0.04811171242

(d) $P(X \leq 14) = 0.09705684391 < 0.1$ [leads to reject H_0]

$P(X \leq 15) = 0.1753690545 > 0.1$ [leads to do not reject H_0]

Acceptance region: $X \geq 15$

(Total for Question 3 is 7 marks)

4 It is known that 16% of visitors to a company's website will purchase a product.

A website designer claims that if the company changes the website layout, the proportion of visitors who purchase a product will increase.

The company changes the website layout as suggested and tracks whether the next 200 visitors purchase a product or not.

The company carries out a hypothesis test, at a 5% level of significance, to test the website designer's claim.

(a) State suitable hypotheses for the test. **(1)**

(b) Find the critical region for the test. **(2)**

The company observes that 44 of the 200 visitors purchased a product.

(c) State the conclusion of the hypothesis test. **(1)**

(d) Find the actual find the actual level of significance for the test. **(1)**

(a) $H_0 : p = 0.16$

$H_1 : p > 0.16$

(b) Let X be the number of visitors who purchased a product.

Under H_0, $X \sim B(200, 0.16)$

$P(X \geq 41) = 1 - P(X \leq 40) = 1 - 0.9460302064 = 0.05396979363 > 0.05$ [leads to do not reject H_0]

$P(X \geq 42) = 1 - P(X \leq 41) = 1 - 0.963108686 = 0.03689131399 < 0.05$ [leads to reject H_0]

Critical region: $X \geq 42$

(c) Reject H_0

There is sufficient evidence to suggest that the proportion of visitors purchasing a product increased

(d) $P(X \geq 42) = 0.03689131399$

Actual level of significance = 0.03689 (or 3.689%)

(Total for Question 4 is 5 marks)

5 Data from a school shows that each week, 8% of the students receive a detention.

One week, the headteacher selects a sample of 25 students.

(a) State a suitable distribution to model the number of students in the sample who receive a detention. **(1)**

(b) Find the probability that from the sample
(i) no students receive a detention **(1)**
(ii) less than 3 students receive a detention. **(1)**

The headteacher believes that with a new behaviour policy the proportion of students receiving a detention in a week will decrease.

After the introduction of the new policy, a new sample of 50 students is taken and the headteacher finds that 1 student received a detention.

(c) Carry out a suitable test to determine whether the proportion of students receiving a detention has decreased. You should state your hypothesis clearly and use a 10% level of significance. **(4)**

(a) Let X be the number students who receive a detention then $X \sim B(25, 0.08)$

(b) $P(X = 0) = 0.1244$

$P(X < 3) = P(X \leq 2) = 0.6768$

(c) $H_0 : p = 0.08$

$H_1 : p < 0.08$

Let Y be the number of students who receive a detention.

Under H_0, $Y \sim B(50, 0.08)$

$P(Y \leq 1) = 0.08271202293$

$0.08271202293 < 0.1$ so reject H_0

There is sufficient evidence to suggest that the proportion of students receiving a detention has decreased

(Total for Question 5 is 7 marks)

6 A train operator claims that 95% of their trains arrive on time.

Thomas carries believes that the true proportion of trains that are on time is lower than 95%

He decides to take a random sample of 20 trains to see if they are on time or not.

(a) Suggest a suitable distribution to model the number of trains from the sample that are on time. **(1)**

(b) State an assumption that you have made when modelling the distribution in part (a) **(1)**

Using his sample, Thomas carries out a hypothesis test, at a 1% level of significance, to test the train operator's claim.

(c) State suitable hypotheses that Thomas could use. **(1)**

(d) For the hypothesis test, find

(i) the critical region **(2)**

(ii) the acceptance region **(1)**

(iii) the actual significance level **(1)**

Thomas observes that 17 of the trains from the sample are on time.

(e) State the conclusion of the hypothesis test. **(1)**

(a) Let X be the number trains that arrive on time then $X \sim B(20, 0.95)$

(b) The train arrivals are independent of each other e.g. one train being on time will not affect the probability of another train being on time.

(c) $H_0 : p = 0.95$

$H_1 : p < 0.95$

(d) (i) $P(X \leq 15) = 0.002573940335 < 0.01$ [leads to reject H_0]

$P(X \leq 16) = 0.01590152602 > 0.01$ [leads to do not reject H_0]

Critical region: $X \leq 15$

(ii) Acceptance region: $X \geq 16$

(iii) $P(X \leq 15) = 0.002573940335$

Actual level of significance = 0.002574 (or 0.2574%)

(e) Do not reject H_0

There is insufficient evidence to suggest that the proportion of trains that are on time is less than 95%

(Total for Question 6 is 8 marks)

7 During training sessions, a tennis player faults on their first serve 28% of the time.

During a match, the player takes 85 first serves and faults on 32 of them.

(a) Carry out a suitable test to determine if the proportion of first serves that are faults in matches is different to that in training sessions.
You should state your hypothesis clearly and use a 5% level of significance. **(4)**

(b) Write down the *p*-value for the test in part (a) **(1)**

During another match, the tennis player serves 72 first serves.

Another hypothesis test is carried out, at the 5% level of significance, to determine if the proportion of first serves that are faults in matches is different to that in training sessions.

(c) Find the acceptance region for this new hypothesis test. **(3)**

(a) $H_0 : p = 0.28$

$H_1 : p \neq 0.28$

Let X be the number of times the tennis player faults on their first serve.

Under H_0, $X \sim B(85, 0.28)$

$P(X \geq 32) = 1 - P(X \leq 31) = 1 - 0.9659407325 = 0.03405926749$

$0.03405926749 > 0.025$ so do not reject H_0

There is insufficient evidence to suggest that the proportion faults on the first serve in matches is different to that in training.

(b) $0.03405926749 \times 2 = 0.06811853497$

Let Y be the number of times the tennis player faults on their first serve.

Under H_0, $Y \sim B(72, 0.28)$

$P(X \leq 12) = 0.01846541633 < 0.025$ [leads to reject H_0]

$P(X \leq 13) = 0.03609993761 > 0.025$ [leads to do not reject H_0]

$P(X \geq 28) = 1 - P(X \leq 27) = 1 - 0.9702505863 = 0.02974941371 > 0.025$ [leads to do not reject H_0]

$P(X \geq 29) = 1 - P(X \leq 28) = 1 - 0.9834729908 = 0.01652700917 < 0.025$ [leads to reject H_0]

Acceptance region: $13 \leq X \leq 28$

(Total for Question 7 is 8 marks)

8 Data shows that 11% of people in the UK are left-handed.

A random sample of 1000 people from the UK is taken and they are asked if they are left-handed.

(a) Suggest a suitable distribution to model the number of left-handed people in the sample. **(1)**

(b) Find the probability that from the sample
(i) less than 100 are left-handed **(1)**
(ii) at least 110 are left-handed **(2)**

Jeremy wants to test if the proportion of people in his city that are left-handed is different to that of the UK.

Jeremy randomly samples 90 people and finds that 4 of them are left-handed.

(b) Carry out a suitable test to determine if the proportion of people in Jeremy's city that are left-handed if different to that of the UK.
You should state your hypothesis clearly and use a 5% level of significance. **(4)**

(a) Let X be the number of left-handed people in the sample then $X \sim B(1000, 0.11)$

(b) $P(X < 100) = P(X \leq 99) = 0.14384372$

$P(X \geq 110) = 1 - P(X \leq 109) = 1 - 0.4850831223 = 0.5149168777$

(c) $H_0 : p = 0.11$

$H_1 : p \neq 0.11$

Let Y be the number of left-handed people in the sample

Under H_0, $Y \sim B(90, 0.11)$

$P(X \leq 4) = 0.02484029779$

$0.02484029779 < 0.025$ so reject H_0

There is sufficient evidence to suggest that the proportion of left-handed people is different to 11%

(Total for Question 8 is 8 marks)

9 A pharmaceutical company states that when taking a particular medicine, 1 in 10 patients will experience side effects.

A doctor claims that the proportion of patients experiencing side effects when taking this medicine is different to that stated by the pharmaceutical company.

The doctor samples their next 68 patients who take this medicine and records how many experience side effects.

(a) State the type of sampling method used by the doctor. **(1)**

To test their claim the doctor carries out a hypothesis, at the 5% level of significance.

(b) State suitable hypotheses that the doctor could use. **(1)**

(c) Find the critical region for this test.
You should state the probability associated with each tail, which should be as close to 2.5% as possible. **(4)**

(d) Find the actual find the actual level of significance for the test. **(1)**

The doctor observed that 5 of the 68 patients experienced side effects.

(e) State the conclusion of the hypothesis test. **(1)**

(a) Opportunity/convenience sampling

(b) $H_0 : p = 0.1$

$H_1 : p \neq 0.1$

(c) Let X be the number of patients who experience side effects Under H_0, $X \sim B(68, 0.1)$

$P(X \leq 1) = 0.006618184304$

$P(X \leq 2) = 0.02837319705$ [closest to 0.025]

$P(X \geq 12) = 1 - P(X \leq 11) = 1 - 0.9637611464 = 0.03623885363$

$P(X \geq 13) = 1 - P(X \leq 12) = 1 - 0.9837060857 = 0.01629391426$ [closest to 0.025]

Critical region: $X \leq 2$ or $X \geq 13$ with probabilities 0.02837319705 and 0.01629391426

(d) $0.02837319705 + 0.01629391426 = 0.04466711131$

Actual level of significance = 0.04467 or 4.467%

(e) Do not reject H_0

There is insufficient evidence to suggest that the proportion of patients with side affects differs from 1 in 10

(Total for Question 9 is 8 marks)

10 A skincare company claims that 4 in 5 users of their product reported "visibly better skin".

An advertising standards organisation decided to conduct a hypothesis test, at the 1% level of significance, to investigate if the true proportion differed from 4 in 5 users.

A random sample of 45 users of the product was taken and they were asked if they had "visibly better skin".

(a) State suitable hypotheses that the advertising standards organisation could use for this test. **(1)**

(b) Find the critical region for this test.
You should state the probability associated with each tail, which should be as close to 0.5% as possible. **(4)**

(c) Find the actual find the actual level of significance for the test. **(1)**

The advertising standards organisation observed that 27 of the 45 users of the product reported "visibly better skin".

(d) State the conclusion of the hypothesis test. **(1)**

(a) $H_0 : p = 0.8$

$H_1 : p \neq 0.8$

(b) Let X be the number of users who reported "visibly better skin" Under H_0, $X \sim B(45, 0.8)$

$P(X \leq 28) = 0.004446975818$ [closest to 0.005]

$P(X \leq 29) = 0.01100454675$

$P(X \geq 42) = 1 - P(X \leq 41) = 1 - 0.9871141873 = 0.01288581267$

$P(X \geq 43) = 1 - P(X \leq 42) = 1 - 0.9967714009 = 0.003228599097$ [closest to 0.025]

Critical region: $X \leq 28$ or $X \geq 43$ with probabilities 0.004446975818 and 0.003228599097

(c) $0.004446975818 + 0.003228599097 = 0.00767557491$

Actual level of significance = 0.007676 or 0.7676%

(d) Reject H_0

There is sufficient evidence to suggest that the proportion of users who report "visibly better skin" differs from 4 in 5

(Total for Question 10 is 7 marks)