A Level · Edexcel · Maths

3 hours 29 questions

5.2 Hypothesis Testing (Binomial Distribution)

5.2.1 Binomial Hypothesis Testing

Easy (8 questions)	/51
Medium (7 questions)	/48
Hard (7 questions)	/55
Very Hard (7 questions)	/49
Total Marks	**/203**

Scan here to return to the course
or visit savemyexams.com

Easy Questions

1 (a) A random variable has distribution $B(n,p)$. Anna uses a single observation of the random variable to carry out a hypothesis test.

Write down the conditions that must be met in order to model a random variable using the binomial distribution.

(4 marks)

(b) Explain how the parameters n and p are used in the context of a hypothesis test.

(2 marks)

(c) Anna is carrying out a two-tailed hypothesis test. Explain how what is being tested for in a two-tailed test differs from what is being tested for in a one-tailed test.

(1 mark)

2 (a) For the random variable $X \sim B(20, 0.3)$ calculate

(i) $\mathrm{P}(X \leq 3)$

(ii) $\mathrm{P}(X < 3)$

(3 marks)

(b) A hypothesis test is to be carried out using an observation of the random variable X to test the hypotheses:

$$\mathrm{H}_0 : p = 0.3 \qquad \mathrm{H}_1 : p < 0.3$$

Before carrying out the test, a significance level of 5% is chosen. The critical region is defined as $X \leq 2$.

(i) Explain why the critical region is defined as $X \leq 2$.

(ii) State the actual significance level for the test.

(2 marks)

3 (a) Alina is running for student council president at her school. She claims she has the support of 60% of students in the school. A rival candidate, John, wants to test at the 10% level of significance whether Alina is overestimating her support.

In mathematical terms, the experiment John conducts uses an observation of the random variable $X \sim B(100, p)$ to test the hypotheses:

$$H_0 : p = 0.6 \qquad H_1 : p < 0.6$$

(i) State how many people are in John's sample and explain what the parameter p means in the context of this question.

(ii) Explain why John has chosen this alternative hypothesis for his test.

(3 marks)

(b) Assuming Alina's claim is true, John calculates the following cumulative probabilities:

x	51	52	53	54	55
$P(X \le x)$	0.04230	0.06379	0.09298	0.13109	0.17890

In John's survey, 55 people say that they will support Alina.

(i) Calculate $P(X = 55)$.

(ii) State whether John should compare the value of $P(X \le 55)$ or $P(X = 55)$ with his significance level of 10%.

(iii) In the context of this question, write a conclusion for John's test.

(4 marks)

(c) Write down the critical value and the critical region for John's test.

(2 marks)

4 (a) A hypothesis test at the 6% significance level is carried out on a coin using the following hypotheses:

$$\mathrm{H}_0 : p = \frac{1}{2} \qquad \mathrm{H}_1 : p \neq \frac{1}{2}$$

(i) Give an example of what the parameter, p, could represent.

(ii) In the context of your answer from part (i), explain what is meant by $p = \frac{1}{2}$.

(iii) This is a two-tailed test. Explain what should be done with the 6% significance level.

(3 marks)

(b) A single observation x is to be taken from a binomial distribution $X \sim B(100, p)$ to test the hypotheses for the coin. One tail of the critical region is found to be $X \leq 6$.

(i) Using your knowledge of the binomial distribution $B(100, 0.5)$, write down the other tail of the critical region.

(ii) Write down the set of values for x which would lead to the acceptance of the null hypothesis.

(2 marks)

5 In a school where all the teachers drink coffee in the staffroom, a headteacher wants to see if using a new brand of coffee in the staffroom improves teachers' report writing punctuality. Previously, 75% of teachers in her school would meet the deadline for writing reports. After the new brand of coffee is introduced the headteacher takes a random sample of 15 teachers, and conducts an experiment to test whether the proportion of teachers meeting the deadline on the next set of reports has improved. The test is conducted at the 5% significance level, using the following hypotheses:

$$H_0 : p = 0.75 \qquad H_1 : p > 0.75$$

(i) Write down a suitable distribution for the random variable X, the number of teachers in the headteacher's sample who meet the deadline.

(ii) Calculate the probability that all the teachers in the sample would meet the deadline if the null hypothesis were true.

(iii) By first calculating the probability that exactly 14 of the teachers would meet the deadline if the null hypothesis were true, find the critical value for the test.

(4 marks)

6 (a) A study of ladybirds in the UK found that 65% of all ladybirds are found to be of the seven-spot species. Alex believes that more than 65% of the ladybirds in his garden are of the seven-spot species. He conducts a hypothesis test at the 5% significance level by collecting a sample of 25 ladybirds from his garden and counting the number of them that are seven-spot ladybirds.

(i) Alex uses the random variable $X \sim B(n,p)$ to represent the number of seven-spot ladybirds in his sample. Explain what n and p represent in the context of Alex' experiment.

(ii) State an assumption Alex has made in order to use the distribution in part (a)(i).

(iii) State suitable null and alternative hypotheses that Alex could use to test his belief that more than 65% of the ladybirds in his garden are seven-spot ladybirds.

(4 marks)

(b) Alex finds that 21 out of the 25 ladybirds in his sample are seven-spot ladybirds.

(i) The p-value for the observed test statistic is 0.03205. Write this in the form $P(X \geq a) = b$.

.

(ii) Alex has not yet calculated the critical value for his hypothesis test. Explain why he does not need to do this to come to a conclusion for his test.

(3 marks)

7 (a) The probability of a student in a primary school library returning his or her books on time had been found to be 0.35. Joanna, the school librarian, has started a new incentive scheme and believes that more students are now returning their books on time because of it. She conducts a hypothesis test using the null hypothesis $H_0: p = 0.35$ to test her belief

(i) State a suitable alternative hypothesis to test Joanna's belief that more students are now returning their books on time.

(ii) Write down the conditions under which Joanna could use a binomial probability distribution to model this problem.

(3 marks)

(b) Joanna takes a random sample of 30 students who have checked out books and finds that under the new incentive scheme 15 of them return their books on time. She calculates the following probabilities for the random variable $X \sim B(30, 0.35)$:

$$P(X = 14) = 0.06112$$

$$P(X = 15) = 0.03511$$

$$P(X \geq 16) = 0.03008$$

Write down the values of $P(X \geq 15)$ and $P(X \geq 14)$.

(2 marks)

(c) Write a conclusion for the hypothesis test, in context, if Joanna had chosen a significance level of:

(i) 5%

(ii) 10%

(3 marks)

8 (a) Wombats, some species of which are critically endangered, are Australian mammals with a lifespan of up to 15 years in the wild. In general, wombats in captivity tend to live longer, with the chance of a wombat in captivity living beyond 15 years being 50%.

Scientists in Australia report having found a population of wombats living in a nature reserve that have longer than average lifespans. In fact, the scientists claim that wombats in the reserve have a greater chance even than captive wombats of living beyond 15 years. A group of naturalists decide to conduct an experiment to test the scientists' claim, using a random sample of 30 wombats from the reserve.

State suitable null and alternative hypotheses to test the scientists' claim.

(2 marks)

(b) It is decided to conduct the test at a significance level of 5%.

It is given that, for $X \sim B(30,0.5)$,

$$P(X \geq 19) = 0.10024$$

$$P(X \geq 20) = 0.04937$$

$$P(X \geq 21) = 0.02139$$

(i) Write down the critical value for this test.

(ii) Write down the actual significance level for this test.

(2 marks)

(c) In the random sample of 30 wombats, it is found that 20 of them are over 15 years old.

Write a conclusion for the hypothesis test in the context of the question.

(2 marks)

Medium Questions

1 (a) A single observation is taken from a discrete random variable $X \sim B(35, 0.4)$ to test $\mathrm{H}_0: p = 0.4$ against $\mathrm{H}_1: p < 0.4$.

Using a 5% level of significance, find the critical region for this test.

(3 marks)

(b) The actual value for the observation was 10.

State a conclusion to the hypothesis test for this value, giving a reason for your answer.

(2 marks)

2 (a) Harry is using the random variable $X \sim B(40, 0.55)$ to test the hypotheses:

$$\mathrm{H}_0: p = 0.55$$

$$\mathrm{H}_1: p \neq 0.55$$

Harry states that the critical regions are $X \leq 17$ and $X \geq 27$.

(i) Calculate the probability of incorrectly rejecting the null hypothesis.

(ii) State, with a reason, the conclusion of Harry's test given that a value of $x = 18$ is observed for the test statistic.

(4 marks)

(b) Sally is using the random variable $Y \sim B(75, 0.3)$ to test the hypotheses:

$$\mathrm{H}_0: p = 0.3$$

$$\mathrm{H}_1: p > 0.3$$

Sally observes the value $y = 30$ for her test statistic.

(i) Calculate the p-value of the observed test statistic $y = 30$.

(ii) State, with a reason, the conclusion of Sally's test if a 5% level of significance is used.

(4 marks)

3 (a) Charlie, the owner of a chocolate shop, claims that more than 60% of people can tell the difference between two brands of chocolate. Charlie takes a random sample of 150 customers and asks them to taste both brands of chocolate. He records that 103 of them could successfully tell the difference between the two brands of chocolate.

State suitable null and alternative hypotheses to test Charlie's claim.

(2 marks)

(b) Test, at the 10% level of significance, whether Charlie's claim is justified.

(3 marks)

4 (a) Nationally it is reported that four out of five people are right-handed. Edward, an education researcher, takes a random sample of 30 children under the age of 18 years old and records the number of them, X , who write with their right hand.

If the national proportion applies to the sample, write down a suitable distribution for X .

(1 mark)

(b) Edward believes that the proportion of right-handed children differs from the national proportion for all people. To test his belief, he uses his sample of 30 children.

State suitable null and alternative hypotheses to test Edward's belief.

(2 marks)

(c) Using a 10% level of significance, find the critical values for a two-tailed test for Edwards' belief. You should state the probability of rejection for each tail, which should be less than 0.05 for each.

(3 marks)

(d) Find the actual level of significance of a test based on your critical values.

(1 mark)

(e) Out of the 30 children in the sample, Edward recorded that 20 of them write with their right hand.

Comment on Edward's belief based on this observation.

(1 mark)

5 **(a)** The existing treatment for a disease is known to be effective in 73% of cases. Dr Sabir develops a new treatment which she claims is more effective than the existing one. To test her claim she uses the new treatment on a sample of 60 patients with the disease and uses a binomial distribution to model the number of them who are cured.

Explain two assumptions that Dr Sabir has made when using a binomial distribution to model the number of patients cured by the vaccine.

(2 marks)

(b) Dr Sabir notes that her treatment was effective for 51 out of the 60 patients used in the sample.

Test, at the 1% level of significance, the validity of Dr Sabir's claim that her treatment is more effective than the existing one. State your hypotheses clearly.

(5 marks)

(c) State the conclusion you would have reached if a 5% level of significance had been used for this test.

(1 mark)

6 (a) A “double yolker” is an egg which contains two yolks. It is known that the probability of a chicken laying a double yolker is 0.1%. A chicken farmer, Paolo, claims that double yolkers are rarer than the stated 0.1%. To test his claim, Paolo records that his chickens lay 1217 eggs in a month and he uses these as his sample. He discovers that none of these eggs are double yolkers.

Test, at the 5% level of significance, whether there is evidence to support Paolo’s claim that double yolkers are rarer than 0.1%. State your hypotheses clearly.

(4 marks)

(b) Paolo decides to take a larger sample so extends his test to three months. During this time, a sample of 3425 eggs is formed and none of them are double yolkers.

Show that there is evidence, at the 5% level of significance, to support Paolo’s claim that double yolkers are rarer than 0.1%.

(2 marks)

(c) Paolo concludes that the probability of a double yolker is **definitely** less than 0.1%. Give a reason to explain whether Paolo’s conclusion is justified.

(1 mark)

7 (a) It is known that 61% of male dragons eat more than 20 sheep within a day. Bill, a dragon breeder, suspects that the proportion of female dragons that eat more than 20 sheep within a day is different to males.

State suitable null and alternative hypotheses for a two-tailed test for Bill's suspicion.

(2 marks)

(b) To test his suspicion, Bill observes a random sample of 80 female dragons during a full day and counts how many sheep they eat. He finds that 40 out of the 80 female dragons ate more than 20 sheep within the day.

Test, at the 5% level, whether there is evidence to support Bill's suspicion.

(3 marks)

(c) Determine the outcome of the test, at the 5% level, if Bill had used a one-tailed test to check whether the proportion of females eating more than 20 sheep within a day is lower than the proportion of males.

(2 marks)

Hard Questions

1 (a) Historical records show that England wins 45% of Ashes test cricket series. Rory believes that in the recent past that proportion has decreased. To test this, he looks at their results in the last 17 series.

Using a 5% level of significance identify the critical region to enable Rory to test his claim.

(5 marks)

(b) England have won five series in the last 17. Use your answer in part (a) to state, with a reason, whether this sample supports Rory's claim.

(2 marks)

(c) State an assumption made when using a binomial model.

(1 mark)

2 It is known that there is a 0.35% chance of picking up a virus when downloading a file from a particular website. After a security update, a security analyst believes that the chances of picking up a virus have been reduced. In a sample of 1542 files downloaded from the website, 1 virus is detected. The analyst wishes to use a hypothesis test with significance levels appropriately chosen to allow her to report with as much certainty as possible that her belief is correct.

Test at the 1%, 5% and 10% significance levels and explain which of those significance levels the analyst should use for the test in her report.

(6 marks)

3 (a) At a school it is known that 80% of students obtain 3 or more A* to C grades in their A levels. The principal claims that another local school has a different proportion of A* to C grades at A level.

Last year the other school had 57 out of 80 students achieve 3 or more A* to C grades at A level.

Using a two-tailed hypothesis test, test the principal's claim at the 5% significance level stating suitable null and alternative hypotheses.

(4 marks)

(b) Suggest a hypothesis test with an alternative significance level that would lead to a different conclusion.

(2 marks)

4 (a) On a particular Pokémon game, it is possible to create offspring through breeding. Sometimes when two Pokémon are bred together, their offspring possess a hidden ability.

One Pokémon fan website, A, claims that the probability of producing offspring with a hidden ability is 0.6. Another website, B, claims the probability is higher.

Maya conducts an experiment and finds that 38 out of 50 of her Pokémon's offspring possess a hidden ability.

Use Maya's sample to test the claim of website A against website B at the 5% significance level, stating your null and alternative hypotheses clearly.

(4 marks)

(b) Website B claims that the probability is 0.8.

Using a two-tailed hypothesis test, test the claim at the 5% significance level and make a statement about which website is more likely to be correct.

(4 marks)

(c) The critical regions of a two-tailed hypothesis test to test B's claim are $X \leq 34$ and $X \geq 45$. Find the actual level of significance.

(3 marks)

5 (a) A single observation is taken from a discrete random variable $X \sim B(25, p)$ to test the hypotheses

$$H_0 : p = \frac{a}{10} \qquad H_1 : p \neq \frac{a}{10}$$

where a is a positive integer to be found.

Given that at the 10% significance level one of the critical regions is $X \leq 5$, find the value of a.

(3 marks)

(b) Find the other critical region for the test at the 10% significance level.

(2 marks)

6 (a) It is known that customers have an 83% success rate when attempting to order bundles of 5 or more festival tickets from a particular website. A second website claims that its customers have a greater probability of success.

State suitable null and alternative hypotheses to test the second website's claim.

(2 marks)

(b) In one day, 1159 out of 1358 bundles of 5 or more tickets are successfully ordered from the second website.

Test the second website's claim using a 1% significance level.

(3 marks)

(c) It is suggested that instead of using a 1% significance level for the test, the critical region $X \geq 1150$ could have been used for testing the second website's claim against observed data for the number of successes out of 1358 attempted orders.

Find the probability of incorrectly rejecting the null hypothesis if that critical region is used.

(3 marks)

7 (a) Nguyen is using the large data set to investigate claims by a local newspaper that the chance of rain on a given day in Leuchars is 18%. She believes that this is incorrect.

Nguyen finds that out of the 184 days in 1987 for which she has data, 43 were rainy.

(i) Using a 5% level of significance, test the newspaper's claim. State your null and alternative hypotheses clearly.

(ii) With reference to the large data set, state one limitation to your conclusion about the newspaper's claim.

(5 marks)

(b) Nguyen also finds that in 2015 there were 37 rainy days out of the 184 for which she has data.

State, giving a reason for your answer and without calculating any additional probabilities, whether the chances of rejecting the newspaper's claim are higher or lower with this sample.

(2 marks)

(c) Nguyen chooses to combine the data from the two years and test at the 5% significance level. Nguyen believes that the chance of rain is higher than 18%.

Calculate the -value for this test and comment on Nguyen's belief.

(4 marks)

Very Hard Questions

1 A single observation, x, is taken from a discrete random variable $X \sim B(25, p)$ to test $\mathrm{H}_0 : p = 0.2$ against $\mathrm{H}_1 : p \neq 0.2$.

For the purposes of this test, the critical regions are specified as being $X \leq 1$ and $X \geq 9$.

Calculate the probability of incorrectly rejecting the null hypothesis when conducting this test.

(4 marks)

2 (a) The probability of a wild Asian elephant living past 40 years old is 45%.

Rosie, a zoologist, obtains data on 25 elephants in captivity and records their ages at death. She suspects that the proportion of Asian elephants that live past 40 years old is smaller for those in captivity than those in the wild.

Using a 1% level of significance, find the critical value of a one-tailed hypothesis test to enable Rosie to test her suspicion. Clearly state your hypotheses.

(3 marks)

(b) Rosie finds that only one elephant, out of the 25 in captivity, lived past 40 years old.

(i) Write down the conclusion of the hypothesis test based on this information.

(ii) Calculate the actual probability that this observed result has led to the incorrect conclusion.

(3 marks)

3 (a) A newspaper article claims that 60% of animal species have become extinct since 1970. Iggy, a taxonomic biologist, has been studying animal extinction statistics and believes that the proportion is different to the newspaper's claim.

Iggy collects a random sample of 500 species that were recorded as being in existence in 1970 and records the number that are now extinct. He uses multiple sources to verify the species' taxonomic status and finds that the number of extinct species could be as low as 319 or as high as 329.

Iggy wants to use his findings to test his belief, but he also wants to make sure that the test will support his belief for all possible values of the number of extinct species. What is the smallest possible integer percentage value he could use for a significance level to make sure that the test supports his belief?

(4 marks)

(b) Iggy repeats this process with a different random sample of 1000 species for which the taxonomic status is more certain, and in this second sample he finds the total number of extinct species to be 650.

Show that the evidence from this second sample offers better support for Iggy's hypothesis than the evidence from his first sample.

(3 marks)

4 (a) Historic company data shows that the proportion of customers ordering a large vegetable box from Lakebridge Organics is 0.4.

Last week a random sample of 35 customers' orders was taken and it was found that 21 of them had ordered a large vegetable box.

Test, at the 5% significance level, whether this suggests that the proportion of customers ordering large vegetable boxes has increased. State your hypotheses clearly.

(4 marks)

(b) Historic company data shows that the proportion of customers ordering a medium vegetable box is also 0.4. Lucinda, a delivery driver, suspects that the proportion of current customers ordering medium boxes is different from 0.4. She takes a random sample of 40 customers' orders.

(i) Stating your hypotheses clearly, identify the critical regions for a hypothesis test at the 5% significance level.

(ii) State the actual level of significance of the test.

(iii) Of the 40 orders, 21 of them were for medium vegetable boxes. State whether this supports Lucinda's suspicion at the 5% level of significance.

(6 marks)

5 (a) A football team, Dinamo Galacticos, are trying to decide on a colour for their new kit. They are told by a local sports commentator that the proportion of games won by teams wearing red is 0.5.

The manager takes a random sample of 20 games where one team wears red and finds that of them are won by the team in red. Assuming that the proportion of games won by teams in red really is 0.5, then the probability that n or fewer games out of 20 would be won is equal to 0.1316 correct to 4 decimal places.

Find the value of n.

(1 mark)

(b) The manager decides to increase her random sample to include 50 games. She suspects that the actual proportion of games won by teams in red is lower than 0.5.

(i) Using a 5% level of significance, find the critical region for a one-tailed hypothesis test. State your hypotheses clearly.

(ii) Using a 10% level of significance, find the largest number of games that could be won by teams wearing red without the test suggesting that the manager's suspicions are unfounded.

(3 marks)

(c) The sports commentator later says that he may have gotten mixed up, and that he thinks it is actually teams wearing green that win 50% of their games.

Given that teams wearing green have won 12 games from the last 35 in which they played, test at the 5% significance level whether the commentator's new claim is justified. You must clearly state your null and alternative hypotheses.

(4 marks)

6 (a) A catering company's past records show that the proportion of their customers who are vegetarian has previously been 0.2. The company decides to take a random sample of 30 current customers to test whether that figure 0.2 is still valid for their current customer base.

Let X denote the number of vegetarians in the sample. Find the critical region for a two-tailed test using a 10% significance level, stating the actual probability in each tail.

(3 marks)

(b) The number of vegetarian customers in the sample is 11. One of the company's employees used the same sample to suggest that this shows that the proportion of vegetarian customers has not just changed but has in fact increased.

Use this test statistic to test, at the 5% level of significance, whether the proportion of vegetarian customers has increased. State your hypotheses clearly.

(2 marks)

7 (a) A random variable has distribution $X \sim B(20, p)$.

In order to test the hypotheses

$$H_0 : p = 0.45 \qquad H_1 : p \neq 0.45$$

a random observation of X is taken and found to be 5.

Carry out a hypothesis test at the 10% significance level.

(3 marks)

(b) Another random variable has distribution $Y \sim B(m, p)$.

A random observation of is taken and found to be 1. Using the same hypotheses as above, find the maximum value of m for which H_0 would not be rejected at the 10% significance level.

(3 marks)

(c) Another random observation of Y is taken and found to be y. For the same hypotheses as above, this provides evidence to reject H_0 at the 5% significance level.

Using your answer to part (b) as the value for m, find the possible values of y.

(3 marks)