Answers - Hypothesis testing

Year: Year 12

Answer to Question 1

a. The null hypothesis always states the value that is assumed (a fair coin); the alternative states what Ravi suspects. $H_0 : p = 0.5$ (the coin is fair) and $H_1 : p > 0.5$ (biased towards heads). The test is one-tailed because he only suspects a bias towards heads.
b. Each toss is independent with the same probability of heads, and there is a fixed number of trials. Under H_0, $X \sim B(20, 0.5)$. This distribution is used to calculate the probabilities.
c. The test statistic is $X = 15$; the probability of a result at least this extreme in the direction of H_1 is $P(X \geq 15)$. $P(X \geq 15) = 1 - P(X \leq 14) = 1 - 0.9793 =$ **0.0207**. This is the p-value of the test.
d. Compare the p-value with the significance level 0.05. 0.0207, so reject H_0. There is sufficient evidence at the 5% level to suggest the coin is biased towards heads.

Answer to Question 2

a. The new drug is claimed to be better, so the alternative hypothesis is one-tailed (upper tail). $H_0 : p = 0.4$, $H_1 : p > 0.4$. Here p is the probability that a patient is cured by the new drug.
b. Use the complement: $P(X \geq 12) = 1 - P(X \leq 11)$. $X \sim B(20, 0.4)$. $P(X \geq 12) = 1 - P(X \leq 11) = 1 - 0.9435 =$ **0.0565**. This is the probability of a result at least as extreme as 12 cures if the new drug is no better.
c. Compare the p-value with 0.05. $0.0565 > 0.05$, so we do **not reject** H_0. There is insufficient evidence at the 5% level to suggest the new drug is more effective than the old one. Note that we do not say that the drugs are equally effective; we just lack evidence.

Answer to Question 3

a. $H_0 : p = 0.25$, $H_1 : p \neq 0.25$. The question asks whether the probability has changed, in either direction, so the test is two-tailed.
b. Under H_0 the number of reds is binomial; 17 is above the expected value $40 \times 0.25 = 10$, so we look at the upper tail. $X \sim B(40, 0.25)$. $P(X \geq 17) = 1 - P(X \leq 16) =$ **0.0116**. This is the probability of 17 or more reds if $p = 0.25$.
c. In a two-tailed test at 5% the significance level is split equally between the two tails. Each tail has 2.5% at the 5% level. 0.0116, so reject H_0. There is sufficient evidence that the probability of landing on red is not 0.25 (it appears to be

higher).

Answer to Question 4

a. For a lower-tail test, the critical region is $X \leq c$ where c is the largest value with $P(X \leq c) \leq 0.05$. We need the largest r with $P(X \leq r) \leq 0.05$: $P(X \leq 2) = 0.0355 \leq 0.05$ but $P(X \leq 3) = 0.1071 > 0.05$. Critical region: $X \leq 2$. Any value of X in this region would lead us to reject H_0.
b. The actual significance level is the probability of the critical region when H_0 is true. Actual significance level $=$ **3.55%**. It is less than 5% because the binomial distribution is discrete.
c. Check whether the observed value lies in the critical region. 3 is not in the critical region, so do not reject H_0: insufficient evidence that p. So the proportion may still be 0.3.

Answer to Question 5

a. Look for the largest r with $P(X \leq r) \leq 0.05$, using cumulative probabilities for $B(20, 0.5)$. $P(X \leq 5) = 0.0207 \leq 0.05$ and $P(X \leq 6) = 0.0577 > 0.05$, so the lower tail is $X \leq 5$. The lower tail of the critical region is $X \leq 5$.
b. $B(20, 0.5)$ is symmetrical about 10, so $P(X \geq 20 - r) = P(X \leq r)$. By symmetry $P(X \geq 15) = 0.0207 \leq 0.05$ and $P(X \geq 14) = 0.0577 > 0.05$. The upper tail is $X \geq 15$. The upper tail is $X \geq 15$.
c. Add the probabilities of the two tails. Actual level $= 0.0207 + 0.0207 =$ **0.0414** (4.14%). The actual significance level is about 4.14%, which is within the 10% allowed.

Answer to Question 6

a. There is sufficient evidence, at the stated significance level, to suggest that the proportion of plants that flower with the new fertiliser has increased (or changed).
b. There is insufficient evidence, at the stated level, to suggest that the proportion of buses on time differs from 90%. (We do not "prove" the claim; we fail to find evidence against it.)
c. A sample can never prove a hypothesis; it can only provide evidence against it. A result in the non-critical region is consistent with H_0 but may also be consistent with other values of p. This is why we say that we do not reject H_0, rather than that we accept it.

Answer to Question 7

a. The shopper suspects there are more prizes than claimed, so the test is one-

tailed. $H_0 : p = 0.25$, $H_1 : p > 0.25$.

b. Under H_0 the number of prize boxes in 15 is binomial; use the complement of $P(X \leq 6)$. $X \sim B(15, 0.25)$. $P(X \geq 7) = 1 - P(X \leq 6) =$ **0.0566**. This is the p-value.

c. Compare the p-value with 0.05. $0.0566 > 0.05$: do not reject H_0. Insufficient evidence that more than 25% of boxes contain a prize.

d. Find the smallest c with $P(X \geq c) \leq 0.05$. $P(X \geq 8) = 0.0173 \leq 0.05$ and $P(X \geq 7) = 0.0566 > 0.05$, so the critical region is $X \geq 8$, with actual significance level **1.73%**. The observed value 7 is just outside the critical region, which agrees with part (c).

Answer to Question 8

a. The gardener suspects a lower germination rate, so the test is one-tailed (lower tail). $H_0 : p = 0.8$, $H_1 : p$. Here p is the probability that a seed germinates.

b. Under H_0 the number that germinate is binomial; we need the lower tail $P(X \leq 16)$. $X \sim B(25, 0.8)$: $P(X \leq 16) =$ **0.0468**. This is the p-value of the test.

c. Compare the p-value with 0.05. 0.0468: reject H_0. There is sufficient evidence that fewer than 80% of seeds germinate. So the company's claim is not supported by her result.

d. Repeat the calculation with $P(X \leq 17)$. $P(X \leq 17) = 0.1091 > 0.05$: do not reject H_0; the result is no longer significant. There is now insufficient evidence that the germination rate is lower than 80%.

Answer to Question 9

a. The student suspects fewer sixes than expected, so the test is one-tailed (lower tail). $H_0 : p = \frac{1}{6}$, $H_1 : p$. Here p is the probability that the die shows a six.

b. Under H_0, $X \sim B\left(30, \frac{1}{6}\right)$; add the probabilities of 0 and 1 six. $P(X = 0) + P(X = 1) = \left(\frac{5}{6}\right)^{30} + 30 \times \frac{1}{6} \times \left(\frac{5}{6}\right)^{29} = 0.0042 + 0.0253 =$ **0.0295**. This is the p-value.

c. Compare the p-value with 0.05. 0.0295: reject H_0. There is sufficient evidence that the die gives sixes less often than a fair die.

d. $P(X = 0) = 0.0042$, which is far smaller, making a chance explanation even less plausible. The more extreme the result, the smaller the p-value and the stronger the evidence against H_0.

Answer to Question 10

a. The new treatment is hoped to be better, so the test is one-tailed (upper tail). $H_0 : p = 0.65$, $H_1 : p > 0.65$. Here p is the probability of recovery with the new treatment.

b. Under H_0, $X \sim B(12, 0.65)$ and every patient must recover.

$P(X = 12) = 0.65^{12} =$ **0.00569**. So this result would happen by chance less than 1% of the time.

c. Compare the p-value with 0.01. 0.0057: reject H_0. There is sufficient evidence at the 1% level that the new treatment has a higher recovery rate.

d. Now the p-value is $P(X \geq 11) = P(X = 11) + P(X = 12)$. $P(X \geq 11) = 0.0424$: reject H_0 at the 5% level (but $0.0424 > 0.01$, so it is not significant at 1%). So the result is significant at 5% but not at 1%.

Answer to Question 11

a. The developers hope for a decrease, so the test is one-tailed (lower tail). $H_0 : p = 0.1$, $H_1 : p$. Here p is the probability that a user reports a crash.

b. Under H_0 the number of crashes in 50 users is $B(50, 0.1)$; the expected number is 5. $X \sim B(50, 0.1)$: $P(X \leq 3) =$ **0.2503**. This is the p-value.

c. Compare the p-value with 0.05. $0.2503 > 0.05$: do not reject H_0. Insufficient evidence that the crash rate has fallen.

d. Find the largest c with $P(X \leq c) \leq 0.05$. $P(X \leq 1) = 0.0338 \leq 0.05$ and $P(X \leq 2) = 0.1117 > 0.05$, so the critical region is $X \leq 1$; actual significance level **3.38%**. Only 0 or 1 crash would have been significant.

Answer to Question 12

a. The council asks whether support has changed in either direction, so the test is two-tailed. $H_0 : p = 0.5$, $H_1 : p \neq 0.5$ (two-tailed).

b. 8 is below the expected value $30 \times 0.5 = 15$, so we use the lower tail. $X \sim B(30, 0.5)$: $P(X \leq 8) =$ **0.0081**. This is the p-value for the lower tail.

c. Compare with 2.5% in each tail: 0.0081, so reject H_0.

d. There is sufficient evidence at the 5% level that support for the plan has changed (it has decreased). The direction of the change comes from the data: 8 out of 30 is about 27%.

Answer to Question 13

a. The buyer suspects more defects than claimed, so the test is one-tailed (upper tail). $H_0 : p = 0.02$, $H_1 : p > 0.02$. Here p is the probability that a component is defective.

b. Under H_0 the expected number of defects is only $40 \times 0.02 = 0.8$; use the complement of $P(X \leq 3)$. $X \sim B(40, 0.02)$: $P(X \geq 4) = 1 - P(X \leq 3) =$ **0.0082**. This is the p-value.

c. Compare the p-value with each significance level. 0.0082 and : reject H_0 at both levels. Strong evidence the defect rate exceeds 2%.

d. Challenge the manufacturer\'s claim or test further before accepting the delivery. The result is very unlikely if the defect rate were really 2%.

Answer to Question 14

a. The question asks whether she has improved, so the test is one-tailed (upper tail). $H_0 : p = 0.6$, $H_1 : p > 0.6$. Here p is the probability that she scores with a shot.
b. Under H_0, $X \sim B(25, 0.6)$; use $P(X \geq 19) = 1 - P(X \leq 18)$. $X \sim B(25, 0.6)$: $P(X \geq 19) =$ **0.0736**. This is the p-value.
c. Compare the p-value with 0.05. $0.0736 > 0.05$: do not reject H_0; no significant evidence of improvement. There is insufficient evidence at the 5% level that the training has improved her scoring rate.
d. The critical region is $X \geq c$ with c the smallest value such that $P(X \geq c) \leq 0.05$. $P(X \geq 20) = 0.0294 \leq 0.05$ while $P(X \geq 19) = 0.0736 > 0.05$. She would need **20 or more**. So the critical region is $X \geq 20$.

Answer to Question 15

a. The test is lower-tailed, so read down the table for the largest probability that is at most 0.05. 5%: largest r with $P(X \leq r) \leq 0.05$ is 2. Critical region $X \leq 2$, actual level **3.55%**.
b. Repeat with the 10% threshold. 10%: $P(X \leq 2) = 0.0355 \leq 0.10$ but $P(X \leq 3) = 0.1071 > 0.10$. Critical region $X \leq 2$ again. Here the critical region is the same at both levels because of the jump from 0.0355 to 0.1071.
c. Check whether 3 lies in each critical region. 3 is outside both critical regions: do not reject H_0 at either level. There is insufficient evidence that fewer than 30% of students walk to school.
d. The binomial distribution is discrete, so the probabilities jump; the critical region is the largest one that does not exceed 5%, and the next value would exceed it.

Answer to Question 16

a. For the lower tail we need the largest r with $P(X \leq r) \leq 0.025$. $P(X \leq 12) = 0.0133 \leq 0.025$ and $P(X \leq 13) = 0.0280 > 0.025$. Lower tail $X \leq 12$.
b. For the upper tail we need the smallest r with $P(X \geq r) \leq 0.025$. $P(X \geq 28) = 0.0160 \leq 0.025$ and $P(X \geq 27) = 0.0314 > 0.025$. Upper tail $X \geq 28$.
c. Critical region: $X \leq 12$ **or** $X \geq 28$. Actual significance level $= 0.0133 + 0.0160 =$ **0.0293** (2.93%).
d. 28 lies in the critical region: reject H_0. There is sufficient evidence that the proportion of patients for whom the drug is effective differs from 40% (it appears to be higher).

Answer to Question 17

a. The actual significance level is the probability of the critical region when H_0 is

true. $P(X \geq 8 \mid p = 0.2) =$ **0.0321** (3.21%). This is the probability of wrongly rejecting a true H_0.

b. Now calculate the probability of the same critical region using the true value $p = 0.4$. $P(X \geq 8 \mid p = 0.4) =$ **0.5841**. So the test detects the change only about 58% of the time.

c. This is the complement of part (b). $1 - 0.5841 =$ **0.4159**: the test misses the change over 40% of the time.

d. Increase the sample size, or raise the significance level (making the critical region larger). Both changes make it easier to reject H_0 when it is false; the second also makes wrong rejections more likely.

Answer to Question 18

a. If every toss is a head, $P(X = n) = 0.5^n$ under H_0. $P(X = n) = 0.5^n$ needs $n \geq 5$ ($0.5^5 = 0.03125$; $0.5^4 = 0.0625$). So at least 5 tosses are needed for even the most extreme result to be significant.

b. At least $n - 1$ heads means exactly $n - 1$ or exactly n heads. $P(X = n - 1) = \binom{n}{n-1} 0.5^n = n \cdot 2^{-n}$ and $P(X = n) = 2^{-n}$. The sum is $\dfrac{n+1}{2^n}$. This is the required formula.

c. Read the table to find the first n with $\dfrac{n+1}{2^n}$. From the table, $\dfrac{n+1}{2^n}$ first at $n = 8$ (0.0352) since $n = 7$ gives 0.0625. So $n = 8$.

d. With small n only an extreme outcome can be significant, so a test is unlikely to detect a bias unless the coin is very unfair; larger samples are more powerful.

Answer to Question 19

a. Under H_0 the number in class D who prefer the menu is $B(10, 0.7)$; 4 is well below the expected 7. $H_0 : p = 0.7$, $H_1 : p$. $X \sim B(10, 0.7)$: $P(X \leq 4) =$ **0.0473** , reject H_0 for class D alone. (Check: $P(X \leq 4) = 0.0473$.) So class D alone gives evidence against the claim.

b. Add the five classes, assuming the students answer independently with the same probability. Total preferring $= 8 + 6 + 9 + 4 + 7 = 34$ out of 50. $X \sim B(50, 0.7)$: $P(X \leq 34) =$ **0.4308**, which is far above 0.05, so do not reject H_0. The overall data are consistent with 70%.

c. A larger sample gives a more reliable estimate and reduces the effect of one unusual class. Testing every class separately would also give a high chance of at least one false rejection (about $1 - 0.95^5 = 23\%$). So the overall conclusion is more trustworthy than the result for class D.

Answer to Question 20

a. State $H_0 : p = 0.5$ and $H_1 : p > 0.5$; the p-value is $P(X \geq 20)$. $X \sim B(30, 0.5)$. $P(X \geq 20) = 0.0494$: reject H_0; evidence that the coin favours heads. The result is just significant at 5%.
b. Now $H_1 : p \neq 0.5$, so each tail is allowed only 2.5%. Two-tailed: compare $P(X \geq 20) = 0.0494$ with 0.025 in each tail. $0.0494 > 0.025$: do not reject H_0 (the equivalent two-tailed probability is $2 \times 0.0494 = 0.0987 > 0.05$). There is insufficient evidence that the coin is biased.
c. The two-tailed test splits the 5% into two tails, so it needs a more extreme result. If you chose the direction after seeing 20 heads you would effectively be giving yourself two chances to find a "significant" result, which inflates the false rejection rate. Therefore the hypothesis must be set in advance.