Size and power of a test

Worked solutions

Further Maths topic practice

This topic online

1 Owen thinks a coin is biased towards heads. He tosses it 12 times and tests H_0: $p = 0.5$ against H_1: $p > 0.5$, where p is the probability of a head. He will reject H_0 if he gets 10 or more heads.

a. Find the size of the test.

b. Given that $p = 0.7$, find the power of the test.

(4 marks)

a) Size = $P(\text{reject } H_0 \mid H_0 \text{ true})$

$X \sim B(12, 0.5)$

$P(X \geq 10) = (C(12,10) + C(12,11) + C(12,12)) \times 0.5^{12}$

$P(X \geq 10) = \frac{(66 + 12 + 1)}{4096}$

$P(X \geq 10) = \frac{79}{4096}$

Size = 0.0193

b) Power = $P(\text{reject } H_0 \mid p = 0.7)$

$X \sim B(12, 0.7)$

$P(X \geq 10) = 0.167790 + 0.071184 + 0.013841$

Power = 0.2528

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

2 The number of goals scored by a football team in a match has a Poisson distribution with mean 3. A new striker joins the team and the manager tests whether the mean number of goals per match has increased, using the number of goals in the striker's first match.

a. Find the critical region for a test of H_0: $\lambda = 3$ against H_1: $\lambda > 3$ at the 5% level of significance.

b. Write down the size of the test.

c. Given that the mean is now 5 goals per match, find the power of the test.

(5 marks)

a) Under H_0, $X \sim Po(3)$

$P(X \geqslant 6) = 1 - P(X \leqslant 5) = 0.0839$

more than 0.05

$P(X \geqslant 7) = 1 - P(X \leqslant 6) = 0.0335$

$X \geqslant 7$

b) Size = $P(\text{in the critical region} \mid H_0 \text{ true})$

0.0335

c) Power = $P(X \geqslant 7 \mid \lambda = 5)$

$X \sim Po(5)$

$P(X \geqslant 7) = 1 - P(X \leqslant 6)$

$P(X \geqslant 7) = 1 - 0.762183...$

0.2378

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

3 A charity finds that 5% of people it phones agree to make a donation. A volunteer, Jamal, uses a new script and thinks this probability, p, has increased. He phones people one at a time until one agrees, and records the number of calls, X, including the successful one. He will reject H_0: $p = 0.05$ in favour of H_1: $p > 0.05$ if $X \leq c$.

a. Using a 10% level of significance, find the largest possible value of c and the size of the test.

b. Write down the power function of the test in terms of p.

c. Find the power of the test when $p = 0.2$

(5 marks)

a) Under H_0, $X \sim Geo(0.05)$

$P(X \leq c) = 1 - 0.95^c$

$P(X \leq 2) = 1 - 0.95^2 = 0.0975$

Less than 0.1

$P(X \leq 3) = 1 - 0.95^3 = 0.1426$

more than 0.1

$c = 2$, size $= 0.0975$

b) Power $= P(X \leq 2)$ for a general p

Power $= 1 - (1 - p)^2$

c)

Power $= 1 - 0.8^2$

0.36

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

4 A council says that 30% of households in a town recycle their food waste. A campaign group believes the proportion, p, is lower. They ask 20 randomly chosen households and test H_0: $p = 0.3$ against H_1: $p < 0.3$ at the 5% level of significance.

a. Find the critical region for the test.

b. Show that the power function of the test is $(1 - p)^{18}(1 + 18p + 171p^2)$.

c. Find the power of the test when $p = 0.1$

(6 marks)

a) Under H_0, $X \sim B(20, 0.3)$

$P(X \leq 2) = 0.0355$

$P(X \leq 3) = 0.1071$

$0.1071 > 0.05$, so stop at 2

$\boxed{X \leq 2}$

b) Power = $P(X \leq 2)$ when $X \sim B(20, p)$

$P(X \leq 2) = (1-p)^{20} + 20p(1-p)^{19} + 190p^2(1-p)^{18}$

Take out the factor $(1-p)^{18}$

$P(X \leq 2) = (1-p)^{18}[(1-p)^2 + 20p(1-p) + 190p^2]$

$(1-p)^2 + 20p(1-p) + 190p^2 = 1 - 2p + p^2 + 20p - 20p^2 + 190p^2$

$(1-p)^2 + 20p(1-p) + 190p^2 = 1 + 18p + 171p^2$

$\boxed{\text{Power} = (1-p)^{18}(1 + 18p + 171p^2)}$

c) $p = 0.1$

$\text{Power} = 0.9^{18}(1 + 1.8 + 1.71)$

$\text{Power} = 0.150095... \times 4.51$

$\boxed{0.6769}$

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

5 The number of breakdowns per month of a printing machine has a Poisson distribution with mean λ. Over many years λ = 5. After a service, the engineer tests H_0: λ = 5 against H_1: λ < 5 at the 5% level of significance, using the number of breakdowns in one month.

a. Find the critical region for the test.

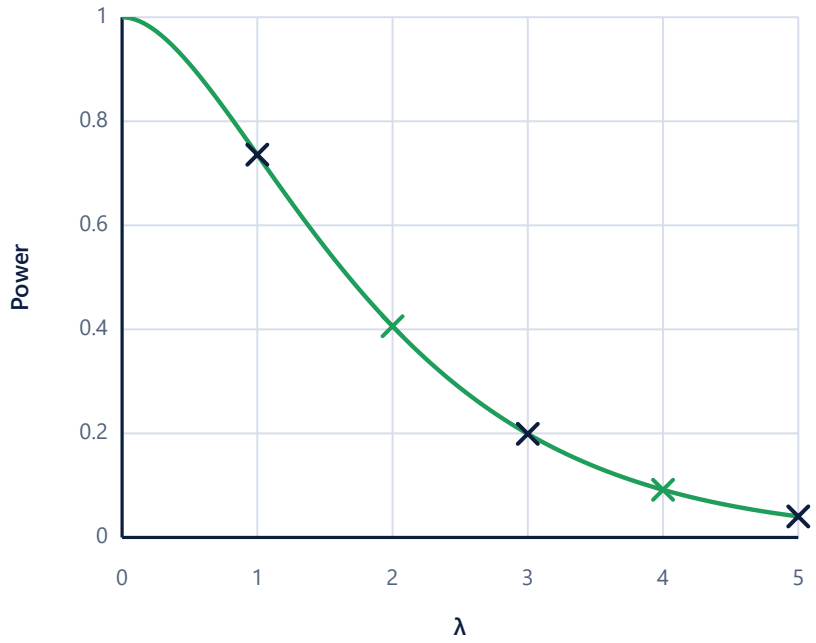
b. Show that the power function of the test is $e^{-\lambda}(1 + \lambda)$.

c. Copy and complete the table, giving values to 4 decimal places.

λ	1	2	3	4	5
Power	0.7358		0.1991		0.0404

d. Sketch the power function on the axes, for 0 ≤ λ ≤ 5

e. State the size of the test.



(7 marks)

a) Under H_0, $X \sim Po(5)$

$P(X \leq 1) = e^{-5}(1 + 5) = 0.0404$

$P(X \leq 2) = e^{-5}(1 + 5 + 12.5) = 0.1247$

0.1247 > 0.05, so stop at 1

$X \leq 1$

b) Power = $P(X \leq 1)$ for any λ

$P(X \leq 1) = P(X = 0) + P(X = 1)$

$P(X \leq 1) = e^{-\lambda} + \lambda e^{-\lambda}$

Power $= e^{-\lambda}(1 + \lambda)$

c)

$\lambda = 2: 3e^{-2} = 0.4060$

$\lambda = 4: 5e^{-4} = 0.0916$

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

0.4060 and 0.0916

d) Plot the five points and join with a smooth curve

At $\lambda = 0$ the power is $e^{0}(1) = 1$

Decreasing curve from (0, 1) to (5, 0.0404)

e) Size = power when H_0 is true ($\lambda = 5$)

0.0404

6 The time, in minutes, that customers wait on a phone line is normally distributed with mean μ and standard deviation 4. The company claims that μ = 50. A customer group thinks that μ is greater than 50. They record the waiting times of a random sample of 16 calls and test H_0: μ = 50 against H_1: μ > 50 at the 5% level of significance.

a. Find the critical region for the sample mean $\bar{X}$.

b. Write down the size of the test.

c. Find the power of the test when μ = 52.5

d. State, with a reason, how the power at μ = 52.5 would change if a larger sample were used with the same significance level.

(6 marks)

a) Under H_0

$$\text{X-bar} \sim N\left(50, \frac{4^2}{16}\right)$$

$$\text{sd} = \frac{4}{\sqrt{16}} = 1$$

Upper 5% point: 1.6449

$$c = 50 + 1.6449 \times 1 = 51.645$$

X-bar > 51.64

b) Continuous: the size is exactly the significance level

0.05

c) Power = $P(\text{X-bar} > 51.645 \mid \mu = 52.5)$

$$\text{X-bar} \sim N(52.5, 1^2)$$

$$P(\text{X-bar} > 51.645) = P(Z > 51.645 - 52.5)$$

$$P(\text{X-bar} > 51.645) = P(Z > -0.8551)$$

0.8038

d) sd of X-bar = $\frac{4}{\sqrt{n}}$ falls as n grows

Power increases: X-bar more likely to be above the critical value

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

7 The probability that a seedling of a type of tomato plant survives being planted out is 0.3. A gardener, Priya, thinks this probability, p, is lower in her garden. She considers two tests of H_0: $p = 0.3$ against H_1: $p < 0.3$, each at the 5% level of significance.

Test A: plant 25 seedlings and count the number, X, that survive.
Test B: plant seedlings one at a time until one survives, and record the number planted, Y.

a. Find the critical region and the size of Test A.

b. Find the critical region and the size of Test B.

c. Find the power of each test when $p = 0.1$

d. State, giving your reasons, which test you would recommend when $p = 0.1$

(9 marks)

a) Test A: under H_0, $X \sim B(25, 0.3)$

$P(X \leq 3) = 0.0332$

$P(X \leq 4) = 0.0905$

$X \leq 3$, size 0.0332

b) Test B: under H_0, $Y \sim Geo(0.3)$

p smaller means more seedlings: upper tail

$P(Y \geq c) = 0.7^{c-1}$

$0.7^8 = 0.0576$

more than 0.05

$0.7^9 = 0.0404$

$Y \geq 10$, size 0.0404

c) Power = $P(\text{reject } H_0 \mid p = 0.1)$

Test A: $X \sim B(25, 0.1)$

$P(X \leq 3) = 0.7636$

Test B: $Y \sim Geo(0.1)$

$P(Y \geq 10) = 0.9^9 = 0.3874$

A: 0.7636, B: 0.3874

d) Compare size and power

Size: A 0.0332 < B 0.0404

Power at $p = 0.1$: A 0.7636 > B 0.3874

Test A: smaller size and much higher power

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

8 The volume of shampoo in a bottle is normally distributed with mean μ ml and standard deviation 12 ml. A quality manager will test H_0: μ = 200 against H_1: μ < 200 at the 1% level of significance, using the mean $\bar{X}$ of a random sample of n bottles. The critical region is $\bar{X} < c$.

a. Find an expression for c in terms of n.

b. The manager wants the power of the test to be at least 0.9 when μ = 195. Find the smallest value of n that achieves this. Show your working clearly.

c. For this value of n, find the value of c.

(8 marks)

a) Under H_0

$$\bar{X} \sim N\left(200, \frac{12^2}{n}\right)$$

Lower 1% point: $z = -2.3263$

$$\frac{(c - 200)}{(12/\sqrt{n})} = -2.3263$$

$$c = 200 - \frac{27.916}{\sqrt{n}}$$

$$\boxed{c = 200 - \frac{27.916}{\sqrt{n}}}$$

b) Power $= P(\bar{X} < c \mid \mu = 195) \geq 0.9$

$$\bar{X} \sim N\left(195, \frac{12^2}{n}\right)$$

Upper 10% point: $P(Z < 1.2816) = 0.9$

$$\frac{(c - 195)}{(12/\sqrt{n})} \geq 1.2816$$

$$\frac{(5 - 27.916/\sqrt{n})}{(12/\sqrt{n})} \geq 1.2816$$

$$\frac{5\sqrt{n}}{12} - 2.3263 \geq 1.2816$$

$$\frac{5\sqrt{n}}{12} \geq 3.6079$$

$$\sqrt{n} \geq 8.6590$$

$$n \geq 74.98$$

Check: $n = 74$ gives power 0.8958, $n = 75$ gives 0.9001

$$\boxed{n = 75}$$

c)

$$c = 200 - \frac{27.916}{\sqrt{75}}$$

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

$c = 196.776...$

$\boxed{c = 196.78}$

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

9 A manufacturer requires that no more than 10% of the light bulbs it makes are faulty. An inspector, Tomasz, tests H_0: $p = 0.1$ against H_1: $p > 0.1$, where p is the proportion of faulty bulbs. D is the number of faulty bulbs in a random sample of 20

Test P: Tomasz takes a sample of 20.

- If $D \geq 5$ he rejects H_0.
- If $D \leq 3$ he does not reject H_0.
- If $D = 4$ he takes a second sample of 20, and rejects H_0 only if this second sample contains 3 or more faulty bulbs.

a. Find the size of Test P.

Given that the proportion of faulty bulbs is actually 0.25

b. find the power of Test P,

c. find the expected number of bulbs inspected using Test P.

Test Q uses a single sample of 30 bulbs and rejects H_0 if it contains 6 or more faulty bulbs. Test Q has size 0.0732 and, when $p = 0.25$, power 0.7974

d. Suggest, giving your reasons, which test Tomasz should use.

(9 marks)

a) Size = $P(\text{reject } H_0 \mid p = 0.1)$, $0 \sim B(20, 0.1)$

$P(0 \geq 5) = 1 - P(0 \leq 4) = 0.043174...$

$P(0 = 4) = 0.089779...$

$P(\text{second} \geq 3) = 1 - P(0 \leq 2) = 0.323073...$

$\text{Size} = 0.043174 + 0.089779 \times 0.323073$

$\boxed{0.0722}$

b) Power = $P(\text{reject } H_0 \mid p = 0.25)$, $0 \sim B(20, 0.25)$

$P(0 \geq 5) = 0.585158...$

$P(0 = 4) = 0.189685...$

$P(\text{second} \geq 3) = 0.908740...$

$\text{Power} = 0.585158 + 0.189685 \times 0.908740$

$\boxed{0.7575}$

c) 20 bulbs always, 20 more only if $0 = 4$

$E(\text{number}) = 20 + 20 \times 0.189685$

$E(\text{number}) = 23.79...$

$\boxed{23.8 \text{ bulbs}}$

d) Compare

Size: P 0.0722, Q 0.0732 (almost equal)

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

Power at 0.25: P 0.7575, Q 0.7974

Bulbs: P 23.8 on average, Q 30

Test P: similar size and power but fewer bulbs inspected

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

10 The number of faults reported per day on a telephone network has a Poisson distribution with mean 2. An engineer wants to test whether the mean number of faults per day has increased, at the 5% level of significance.

Test A uses the number of faults, X, on one day.
Test B uses the number of faults, Y, over two days.

a. Find the critical region and the size of each test.

b. Write down the power function of Test A in terms of λ, the mean number of faults per day.

c. Find the power of each test when λ = 3.5

d. State, giving a reason, which test you would recommend.

(9 marks)

a) Test A: under H_0, $X \sim Po(2)$

$P(X \geq 5) = 0.0527$

$P(X \geq 6) = 0.0166$

A: $X \geq 6$, size 0.0166

Test B: two days, $Y \sim Po(4)$

$P(Y \geq 8) = 0.0511$

$P(Y \geq 9) = 0.0214$

B: $Y \geq 9$, size 0.0214

b) Power of A = $P(X \geq 6)$ with $X \sim Po(\lambda)$

$P(X \geq 6) = 1 - P(X \leq 5)$

$$1 - e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{6} + \frac{\lambda^4}{24} + \frac{\lambda^5}{120}\right)$$

c) $\lambda = 3.5$ per day

Test A: $X \sim Po(3.5)$

$P(X \geq 6) = 1 - 0.857614...$

$P(X \geq 6) = 0.1424$

Test B: $Y \sim Po(7)$

$P(Y \geq 9) = 1 - 0.729091...$

$P(Y \geq 9) = 0.2709$

A: 0.1424, B: 0.2709

d) Both sizes are below 0.05

Test B: much higher power (fewer Type II errors)

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

11 A spinner has five equal sections numbered 1, 2, 3, 4 and 5. Callum spins it 60 times and calculates the mean score, $\bar{X}$. He tests the hypotheses

H_0: the spinner is fair H_1: the spinner is not fair

and will reject H_0 if $\bar{X} < 2.7$ or $\bar{X} > 3.3$

a. Show that, if the spinner is fair, the score on one spin has mean 3 and variance 2

b. Using a suitable approximation, show that the size of the test is 0.100 to 3 significant figures.

c. Given that the spinner is actually biased so that the score on one spin has mean 3.4 and variance 1.9, find the power of the test.

d. Describe what would happen to the power of this test if Callum spun the spinner more times, using the same critical region. Give a reason for your answer.

(10 marks)

a) Discrete uniform on 1, ..., n with n = 5

$$\text{mean} = \frac{(5+1)}{2} = 3$$

$$\text{variance} = \frac{(5^2-1)}{12} = \frac{24}{12} = 2$$

b) CLT under H_0, n = 60

$$\bar{X} \approx N\left(3, \frac{2}{60}\right)$$

$$\text{sd} = \sqrt{\frac{2}{60}} = 0.182574...$$

$$P(\bar{X} < 2.7) = P\left(Z < \frac{(2.7-3)}{0.182574}\right) = P(Z < -1.6432)$$

$$P(\bar{X} < 2.7) = 0.050171...$$

Symmetrical, so double

$$\text{Size} = 2 \times 0.050171 = 0.10034...$$

Size = 0.100 (3 s.f.)

c) Power = P(reject H_0) when mean 3.4, variance 1.9

$$\bar{X} \approx N\left(3.4, \frac{1.9}{60}\right)$$

$$\text{sd} = \sqrt{\frac{1.9}{60}} = 0.177951...$$

$$P(\bar{X} < 2.7) = P(Z < -3.9337) = 0.000042$$

$$P(\bar{X} > 3.3) = P(Z > -0.5620) = 0.712942$$

$$\text{Power} = 0.000042 + 0.712942$$

0.7130

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

d) Var(X-bar) = $\frac{\sigma^2}{n}$ falls as n grows

Power increases: X-bar more concentrated about 3.4, so more likely to be above 3.3

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

12 A machine makes ceramic tiles. The manager claims that the probability, p, that a tile is cracked is 0.1. A buyer believes that p is greater than 0.1 and wants to test H_0: $p = 0.1$ against H_1: $p > 0.1$ at the 5% level of significance.

a. The buyer first considers inspecting tiles until she finds one cracked tile, rejecting H_0 if this happens soon enough. Show that no critical region exists for this test at the 5% level.

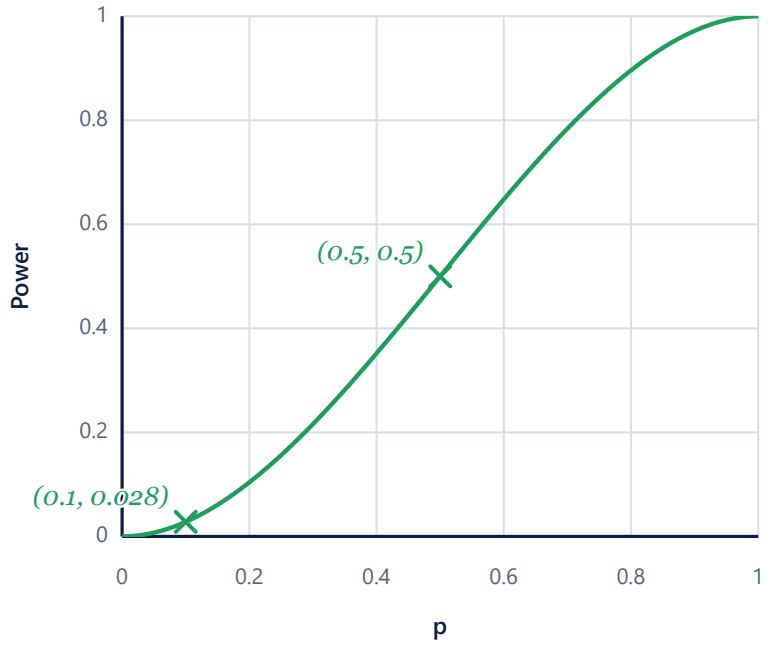
Instead, she inspects tiles until she has found 2 cracked tiles. Y is the number of tiles inspected.

b. Find the critical region and the size of this test.

c. Show that the power function of the test is $p^2(3 - 2p)$.

d. Sketch the power function for $0 \leq p \leq 1$ on the axes.

e. Show that the power of the test is 0.5 when $p = 0.5$



(10 marks)

a) Geometric: $X \sim Geo(0.1)$; reject for small X

$P(X \leq 1) = 0.1$

$0.1 > 0.05$, so even $X = 1$ is not significant

No critical region exists at 5%

b) $Y \sim NB(2, 0.1)$

$P(Y = y) = (y - 1)(0.1)^2(0.9)^{y-2}$

$P(Y = 2) = 0.01$

$P(Y = 3) = 2(0.01)(0.9) = 0.018$

$P(Y = 4) = 3(0.01)(0.81) = 0.0243$

$P(Y \leq 3) = 0.028$

$P(Y \leq 4) = 0.0523$

$0.0523 > 0.05$, so stop at 3

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

$Y \leq 3$, size 0.028

c) Power = $P(Y \leq 3)$ for a general p

$$P(Y \leq 3) = p^2 + 2p^2(1 - p)$$

$$P(Y \leq 3) = p^2(1 + 2 - 2p)$$

Power $= p^2(3 - 2p)$

d) Power is 0 at $p = 0$ and 1 at $p = 1$, increasing

S-shaped increasing curve from (0, 0) to (1, 1)

e)

$$0.5^2(3 - 2 \times 0.5) = 0.25 \times 2$$

$= 0.5$

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.