2.4 Further Statistics: Hypothesis Testing, Type I and Type II Errors: Test Pack

AQA 7367 · Calculator allowed · about 65 minutes

Total Marks

Name: _____ Date: ____ / ____ / _____

Answer ALL questions. Show all your working.

Hypothesis Testing for a Poisson or Binomial Parameter

Revision Library

For a hypothesis test on a Poisson mean lambda or a binomial probability p, a null hypothesis H0 states a fixed value of the parameter, and an alternative hypothesis H1 states that the parameter has increased, has decreased (a one-tailed test), or is simply different (a two-tailed test). A test statistic X, assumed to follow the distribution stated in H0, is compared against a critical region: the set of values of X for which H0 is rejected in favour of H1.

Because X is discrete, the probability of X falling in a proposed critical region can rarely be made to equal a chosen significance level exactly. The usual convention is to choose the largest critical region whose probability does not exceed the significance level requested. This probability is called the actual significance level of the test, and it is also equal to P(Type I error), the probability of wrongly rejecting a true H0.

A Type II error occurs when H0 is not rejected even though H1 is actually true. Its probability depends on the true value of the parameter, and is found by calculating the probability that X falls outside the critical region, using the true (alternative) value of the parameter in place of the value stated in H0. The power of a test is defined as 1 - P(Type II error): the probability of correctly rejecting a false H0. Power increases as the true parameter value moves further from the value stated in H0, and is also affected by the significance level chosen and the sample size used.

© Revision Library | revisionlibrary.co.uk

1 A car insurance company models the number of claims made per month by a certain class of driver using a Poisson distribution, and regularly carries out hypothesis tests on the mean claim rate.

(a) State what is meant by the critical region (rejection region) of a hypothesis test. **(1)**

(b) Explain why, for a test statistic with a discrete distribution such as Poisson or binomial, the actual significance level of a test is not usually exactly equal to the significance level originally requested. **(2)**

(c) A test is carried out, at a chosen significance level, of H0: the mean monthly claim rate is unchanged, against H1: the mean monthly claim rate has increased. Define, in the context of this test, a Type I error and a Type II error. **(2)**

(d) State an equation linking the power of a hypothesis test to P(Type II error). **(1)**

(Total for Question 1 is 6 marks)

---

2 A call centre in Preston models the number of customer complaints received per day, under normal conditions, by $X \sim Po(4)$. Following a change to the automated phone system, the manager wants to test, at the 5% significance level, whether the mean number of complaints per day has decreased.

(a) State suitable null and alternative hypotheses, then find the critical region for the test and state the actual significance level. **(3)**

(b) On a particular day following the change, the number of complaints received is 0. State, with a reason, the conclusion of the test. **(2)**

(c) Suppose that the change has in fact reduced the true mean number of complaints per day to 2. Find the probability that this test results in a Type II error. **(3)**

(d) Hence state the power of the test in this case. **(1)**

(Total for Question 2 is 9 marks)

© Revision Library | revisionlibrary.co.uk

3 The random variable $X \sim B(12, p)$. A test of H0: $p = 0.4$ against H1: $p < 0.4$ is carried out at the 5% significance level.

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

(b) State the probability of a Type I error, to 3 significant figures. **(2)**

(c) Find the probability of a Type II error when the true value of p is 0.1, to 3 significant figures. **(2)**

(Total for Question 3 is 6 marks)

---

4 A fair coin-like device is claimed to land on 'success' with probability 0.1. A tester believes the probability of success is actually lower. Let X be the number of trials needed to get the first success, so under the claim $X \sim Geo(0.1)$. The first success occurs on trial 14.

(a) State suitable hypotheses for the test, in terms of p. **(2)**

(b) Find $P(X \geq 14)$ assuming H0 is true, and test the tester's belief at the 10% level. **(3)**

(c) Find the critical region for the test at the 10% level, and the probability of a Type II error if the true value of p is 0.05, to 3 significant figures. **(4)**

(Total for Question 4 is 9 marks)

© Revision Library | revisionlibrary.co.uk

5 A machine at a plastics factory in Telford produces components, 20% of which have historically been classed as premium grade. After the machine is recalibrated, an engineer wants to test, at the 5% significance level, whether the proportion of premium-grade components has increased. A random sample of 25 components is taken, and X is the number classed as premium grade.

(a) State suitable hypotheses, then find the critical region for the test and state the actual significance level. **(4)**

(b) In the sample, 10 components are classed as premium grade. State, with a reason, the conclusion of the test. **(2)**

(c) Suppose that the recalibration has in fact increased the true proportion of premium-grade components to 0.35. Find the probability that the test results in a Type II error. **(4)**

(d) Hence state the power of the test in this case. **(1)**

(e) State, without further calculation, what would happen to the power of the test found in part (d) if the sample size were increased from 25 to 50, with the significance level kept at 5%. Justify your answer. **(2)**

(Total for Question 5 is 13 marks)

6 A hospital microbiology lab models the number of bacterial colonies per agar plate, under safe conditions, as a Poisson random variable with a known mean. Two proposed testing procedures are available to check whether the mean number of colonies per plate has risen above the safe threshold: Test A uses a 1% significance level, and Test B uses a 10% significance level. Failing to detect a genuine rise in bacterial colonies (a Type II error) could pose a public health risk, while wrongly signalling a rise when none has occurred (a Type I error) would trigger a costly, unnecessary shutdown of the affected ward for deep cleaning.

Discuss, with reference to Type I and Type II errors, which of the two significance levels the lab should adopt, and justify a recommendation.

(Total for Question 6 is 6 marks)

© Revision Library | revisionlibrary.co.uk

Mark scheme · 2.4 Further Statistics: Hypothesis Testing, Type I and Type II Errors: Test Pack

Question 1

(a) the set (range) of values of the test statistic that would lead to the rejection of H0 in favour of H1 (oe) **B1**

(a) Answer: The critical region is the set of values of the test statistic for which H0 is rejected in favour of H1.

(b) because the test statistic is discrete, its cumulative probabilities take only a limited set of values, so a critical region with probability exactly equal to the requested level usually cannot be found (oe) **B1**

(b) the convention is to choose the largest critical region whose probability does not exceed the requested level, so the actual significance level found is usually slightly less than (and never more than) the level requested (oe) **B1**

(b) Answer: Cumulative probabilities for a discrete distribution take only a limited set of values, so it is usually impossible to match the requested level exactly; the largest critical region not exceeding the requested level is chosen instead, giving an actual significance level that is usually a little below what was requested.

(c) Type I error: concluding that the mean claim rate has increased when it has not actually changed (i.e. rejecting a true H0) (oe) **B1**

(c) Type II error: concluding that there is not enough evidence that the mean claim rate has increased when it actually has increased (i.e. not rejecting a false H0) (oe) **B1**

(c) Answer: Type I error: wrongly concluding the mean claim rate has increased when it has not. Type II error: wrongly concluding there is insufficient evidence of an increase when the mean claim rate has in fact increased.

(d) power = 1 - P(Type II error), cao **B1**

(d) Answer: power = 1 - P(Type II error)

Question 2

(a) states H0: $\lambda = 4$, H1: $\lambda < 4$, and attempts to compare $P(X \leq 0)$ and $P(X \leq 1)$ with 0.05 **M1**

(a) $P(X \leq 0) = e^{-4}$ = awrt 0.0183 (< 0.05) but $P(X \leq 1)$ = awrt 0.0916 (> 0.05), so critical region is X = 0, cao **A1**

(a) actual significance level = awrt 0.0183 **A1**

(a) Answer: Critical region: X = 0. Actual significance level = awrt 0.0183.

(b) 0 lies in the critical region found in part (a) (ft) **B1**

(b) reject H0; there is evidence, at the 5% level, that the mean number of complaints per day has decreased (ft, in context) **B1**

(b) Answer: 0 is in the critical region {0}, so H0 is rejected; there is evidence at the 5% level that the mean daily complaint rate has decreased.

(c) recognises that a Type II error occurs here when X does not fall in the critical region {0}, i.e. when $X \geq 1$, using the true distribution Po(2) **M1**

(c) uses $P(X \geq 1 \mid \lambda = 2) = 1 - P(X = 0 \mid \lambda = 2) = 1 - e^{-2}$ **M1**

(c) awrt 0.865 **A1**

(c) Answer: P(Type II error) = $1 - e^{-2}$ = awrt 0.865

(d) power = 1 - 0.865 = awrt 0.135 (ft from part (c)) **B1**

(d) Answer: power = awrt 0.135

Question 3

(a) uses B(12, 0.4) cumulative probabilities **M1**

(a) critical region $X \leq 1$ **A1**

(a) Answer: critical region $X \leq 1$

(b) P(X in critical region | p = 0.4) **M1**

(b) 0.0196 **A1**

(b) Answer: 0.0196

(c) P(X not in critical region | p = 0.1) **M1**

(c) 0.341 **A1**

(c) Answer: 0.341

Question 4

(a) H0: p = 0.1 **B1**

(a) H1: p < 0.1 **B1**

(a) Answer: H0: p = 0.1, H1: p < 0.1

(b) $P(X \geq 14) = (1 - 0.1)^{14 - 1}$ **M1**

(b) 0.254 **A1**

(b) do not reject H0 **A1**

(b) Answer: 0.254; do not reject H0

(c) finds the smallest d with $(1 - 0.1)^{d - 1} \leq 0.1$ **M1**

(c) critical region $X \geq 23$ **A1**

(c) $P(X \leq 22 \mid p = 0.05) = 1 - (1 - 0.05)^{22}$ **M1**

(c) 0.676 **A1**

(c) Answer: $X \geq 23$; Type II error 0.676

© Revision Library | revisionlibrary.co.uk

Question 5

(a) states H0: p = 0.2, H1: p > 0.2, with X ~ B(25, 0.2) under H0, and attempts an upper-tail search comparing P(X≥8) and P(X≥9) with 0.05 **M1**

(a) critical region is X ≥ 9 (i.e. {9, 10, ..., 25}), cao **A1**

(a) actual significance level = awrt 0.0468 **A1**

(a) correctly notes P(X≥8) = awrt 0.109 exceeds 0.05, confirming X=8 cannot be included **B1**

(a) Answer: Critical region: X ≥ 9 (i.e. {9, 10, ..., 25}). Actual significance level = awrt 0.0468.

(b) 10 lies in the critical region {9, 10, ..., 25} found in part (a) (ft) **B1**

(b) reject H0; there is evidence, at the 5% level, that the proportion of premium-grade components has increased (ft, in context) **B1**

(b) Answer: 10 is in the critical region {9, ..., 25}, so H0 is rejected; there is evidence at the 5% level that the proportion of premium-grade components has increased.

(c) identifies the Type II error region as the acceptance region X ≤ 8, using the true distribution B(25, 0.35) **M1**

(c) attempts to sum P(X=0) to P(X=8) with p = 0.35 **M1**

(c) awrt 0.467 **A1**

(c) correctly identifies this cumulative value itself as P(Type II error), distinguishing it from the power **B1**

(c) Answer: P(Type II error) = awrt 0.467

(d) power = 1 - 0.467 = awrt 0.533 (ft from part (c)) **B1**

(d) Answer: power = awrt 0.533

(e) the power of the test would increase (oe) **B1**

(e) a larger sample size reduces the variability (standard error) of the sample proportion, so the sampling distributions under H0 (p=0.2) and under the true value (p=0.35) overlap less; a greater proportion of the alternative distribution then falls in the (relatively better-targeted) critical region, so P(Type II error) falls and power rises (oe) **B1**

(e) Answer: The power would increase, because with a larger sample the distributions of X under H0 and under the true p = 0.35 overlap less, so the critical region captures a larger proportion of the true distribution, reducing P(Type II error).

Question 6

Level 1 Makes a simple, largely one-sided statement, e.g. states that a smaller significance level reduces the probability of a Type I error (or that a larger significance level reduces the probability of a Type II error), with little development or link to the context given. **1-2**

Level 2 Gives some balanced reasoning referring to both Type I and Type II errors and links this to at least one consequence described in the context (health risk or cost of shutdown), but the discussion is not fully developed or does not reach a clearly justified recommendation. **3-4**

Level 3 Gives a fully balanced, well-reasoned discussion that correctly explains the Type I/Type II error trade-off as the significance level changes, links this clearly to both consequences described in the context, and reaches a justified recommendation supported by that reasoning. **5-6**

Indicative content:

- Reducing the significance level (e.g. from 10% to 1%) makes the critical region smaller, so P(Type I error) decreases.
- A smaller critical region means a smaller proportion of the alternative (raised-mean) distribution falls inside it, so P(Type II error) increases (power decreases) for a fixed true mean above the threshold.
- In this context, a Type II error means failing to detect a genuine rise in bacterial colonies, risking patient and staff health.
- A Type I error means wrongly concluding contamination has risen when it has not, triggering an unnecessary and costly ward shutdown.
- Since the consequence of a Type II error (health risk) is generally judged more serious than a Type I error (cost and disruption), the larger significance level (Test B, 10%), which gives higher power and a lower P(Type II error), may be preferred despite its higher false-alarm rate.
- A well-justified alternative conclusion favouring Test A could still be credited, provided it correctly links significance level to the Type I/Type II error probabilities and gives sound context-based reasoning (e.g. if shutdowns are judged to carry very severe cost or disruption).

© Revision Library | revisionlibrary.co.uk