inkMaths

ANSWER BOOK

The quality of tests

Further Statistics 1 · Year FM

7 questions · 26 marks

NAVY EDITION

inkphysics.co.uk/maths/ · youtube.com/@InkPhysics

SECTION A · Warm-up

A1 A Type I error rejects H_0 when H_0 is true: a false alarm. A Type II error fails to reject H_0 when a particular alternative is true: a missed change. B1 B1 for the two definitions. The first needs no alternative specified; the second cannot be computed without one. [2]

A2 Size = $P(X \le 1 \mid p = 0.5) = (1 + 10)/2^{10} = 11/1024 =$ **0.0107**. M1 for the lower tail under H_0, A1 for 11/1024, A1 for 0.0107. That sits well below a nominal 5%. The distribution is discrete, so the jump from $X \le 1$ to $X \le 2$ overshoots the level and no region has size exactly 0.05. [3]

SECTION B · Standard

B1 $P(X \le 1 \mid p = 0.2) = 0.8^{10} + 10(0.2)(0.8^9) = 0.1074 + 0.2684 = 0.3758$, so the test rejects with probability 0.3758: that is the power. The Type II probability is $1 - 0.3758 = 0.6242$, so the test misses a drop to 0.2 nearly two times in three. M1 for the binomial terms, A1 for 0.3758, A1 for naming that as the power, A1 for 0.6242. [4]

B2 $P(X \le 1 \mid p = 0.1) = 0.9^{10} + 10(0.1)(0.9^9) = 0.3487 + 0.3874 = 0.7361$. M1 for the two binomial terms, A1 for 0.7361, B1 for the comparison. The power has almost doubled: the further the truth lies from H_0, the more easily the test detects it. [3]

B3 Size = $P(X \ge 10 \mid \lambda = 5) = 1 - P(X \le 9) = 1 - 0.9682 =$ **0.0318**. Power at $\lambda = 8$ is $P(X \ge 10 \mid \lambda = 8) = 1 - P(X \le 9) = 1 - 0.7166 =$ **0.2834**. M1 for the one minus form, A1 for 0.0318, M1 for the same at $\lambda = 8$, A1 for 0.2834. A rise from 5 to 8 is caught barely more than a quarter of the time by a single observation. Both figures come from the same cumulative table, read at different values of λ. [4]

B4 A wider region rejects H_0 more readily, so it catches more true departures and the power rises at every alternative. It also rejects more often when H_0 is true, so the size rises with it and there are more false alarms. With the sample size fixed the two move together, and the right balance depends on which error costs more. B1 for the power rising, B1 for the size rising, B1 for the trade-off. Only a larger sample improves both at once. [3]

SECTION C · Stretch

C1 The alternative is one-tailed and low, so the critical region sits in the lower tail. [7]
$P(X \leq 3) = 0.1071$ exceeds 0.05, so 3 cannot be included, while $P(X \leq 2) = 0.0355$ does not.
The critical region is $\mathbf{X} \leq 2$.
The size is the probability of rejecting H_0 when it is true, that is **0.0355**, not the nominal 0.05. A discrete distribution rarely offers a region of size exactly 0.05, and the actual size is what must be quoted.
Power at $p = 0.15 = P(X \leq 2 \mid p = 0.15) = 0.85^{20} + 20(0.15)(0.85^{19}) + 190(0.15^2)(0.85^{18}) = 0.0388 + 0.1368 + 0.2293 =$ **0.405**.
P(Type II error) = 1 - 0.405 = **0.595**. The test misses a fall from 0.3 to 0.15 in nearly six cases out of ten, so a non-significant result here is weak evidence for H_0 rather than support for it.
M1 for comparing the tail probabilities with 0.05, A1 for $X \leq 2$, B1 for the size 0.0355, M1 for the binomial terms at $p = 0.15$, A1 for 0.405, A1 for 0.595, B1 for the closing comment.
Power is always computed with the critical region fixed and the parameter changed.
Recomputing the region at $p = 0.15$ is the standard error.