inkMaths

ANSWER BOOK

Hypothesis testing: correlation and the normal

Statistics · Year 12–13

8 questions · 31 marks

NAVY EDITION

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

SECTION A · Warm-up

A1 The sample mean is normal with the same centre and a tighter spread, so $\bar{X} \sim N(\mu, \sigma^2/n)$. B1 for the normal distribution with mean μ, B1 for the variance. Averaging cancels wobble, and the variance shrinks by a factor of n, so the standard deviation of the mean is $\sigma/\sqrt{n}$ rather than σ/n. [2]

A2 The standard error of the sample mean is $3/\sqrt{36} = 0.5$, so $z = (48.8 - 50)/0.5 = -2.4$. M1 for the standard error, M1 for the z formula, A1 for −2.4. The denominator is the standard error, not the population standard deviation. Dividing by 3 instead of 0.5 gives −0.4 and turns a decisive result into a limp one, and that is the single most common error on this topic. [3]

SECTION B · Standard

B1 H_0: $\mu = 100$ and H_1: $\mu < 100$. Under H_0, $\bar{X} \sim N(100, 4^2/25)$, so the standard error is 0.8 and $z = (98.6 - 100)/0.8 = -1.75$. Then $P(Z \le -1.75) = 0.0401$. Since $0.0401 < 0.05$, reject H_0. There is significant evidence at the 5% level that the mean mass of the bags is below 100 g. The five marks are B1 hypotheses, M1 distribution of $\bar{X}$, M1 probability, A1 0.0401, A1 comparison and conclusion in context. All three parts of the conclusion are needed: the comparison, the decision, and a sentence about flour rather than about z. [5]

B2 Under H_0, $\bar{X} \sim N(30, 16/16)$, so the standard error is 1. A one-tailed 5% test rejects when $z > 1.6449$, that is when $\bar{X} > 30 + 1.6449 \times 1 = 31.6449$ g. The critical region is $\bar{X} > 31.6449$. The observed mean 31.2 is well below that boundary, so it is not in the critical region. Do not reject H_0. There is insufficient evidence at the 5% level that the mean mass has risen. M1 for the standard error, B1 for $z = 1.6449$, M1 for the boundary, A1 for 31.6449, A1 for the conclusion. As a cross-check, $z = 1.2$ gives a tail probability of 0.115, comfortably above 0.05, and the two methods agree. [5]

Carry the boundary at full accuracy. Rounding it to 31.6 hands the critical region every mean between 31.6 and 31.6449, and a sample mean of 31.62 would then be declared significant when its tail probability is 0.0526, above 0.05. That is where the critical-region and p-value methods appear to disagree, and the rounding is always the culprit. Round the boundary only for reporting, and only when the observed mean is nowhere near it.

B3 H_0: $\rho = 0$ and H_1: $\rho > 0$, where ρ is the population correlation coefficient. The test is one-tailed, so the 0.01 column of the table is the right one and the critical value is 0.5155. Since $0.62 > 0.5155$, reject H_0. There is significant evidence at the 1% level of positive correlation between the two variables in the population. B1 for the hypotheses in terms of ρ, B1 for the critical value, M1 for the comparison, A1 for the conclusion in context. Hypotheses written about r rather than ρ lose the first mark every time, since r is only the sample's echo of the population quantity being tested. [4]

B4 The standard error $\sigma/\sqrt{n}$ shrinks as n grows, so the same difference is measured against a tighter yardstick and the z value rises. B1 for the standard error shrinking, B1 for the effect on z or on the tail probability. A drift of 1 gram on 4 bags is ordinary noise; on 400 bags the mean has no business straying that far, and the test says so. [2]

SECTION C · Stretch

C1 H_0: $\rho = 0$ and H_1: $\rho \neq 0$. The alternative hypothesis names no direction, so the test is two-tailed and the 5% is split between the two ends. The table is indexed by one-tail probability, so the row must be read at 0.025, giving critical values ± 0.3610. Since $0.42 > 0.3610$, reject H_0. There is significant evidence at the 5% level of correlation between the two variables. At the 2% level the relevant column is 0.01, where the critical value is 0.4226, and $0.42 < 0.4226$, so H_0 would not be rejected and the conclusion changes. B1 hypotheses, B1 identifying the test as two-tailed, B1 reading the 0.025 column, A1 first conclusion, A1 second conclusion. Reaching for the 0.05 column on a two-tailed 5% test is the mistake this question is built around, and it doubles the true significance level of the test. [5]

C2 Under H_0, $\bar{X} \sim N(50, 4/16)$, so the standard error is 0.5. A two-tailed 5% test puts 2.5% in each tail, so the critical z values are ± 1.96 and the boundaries are $50 \pm 1.96 \times 0.5$, that is 49.02 and 50.98. The critical region is $\bar{X} < 49.02$ or $\bar{X} > 50.98$. Since $51.1 > 50.98$ it lies inside the critical region, so reject H_0. There is significant evidence at the 5% level that the mean length is not 50 cm. M1 standard error, B1 $z = \pm 1.96$, M1 forming a boundary, A1 both boundaries, A1 conclusion. Using 1.6449 here would give a critical region starting at 50.82, and 51.1 would still be rejected, so the right answer arrives for the wrong reason and the accuracy marks are lost anyway. [5]