Tests for correlation and for the mean of a normal distribution

S5.1, S5.3 · Statistics · revision notes

Two further tests: whether a population correlation coefficient ρ is zero, using the sample product moment correlation coefficient r and a table of critical values, and a test for the mean of a normal distribution with known variance, using $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

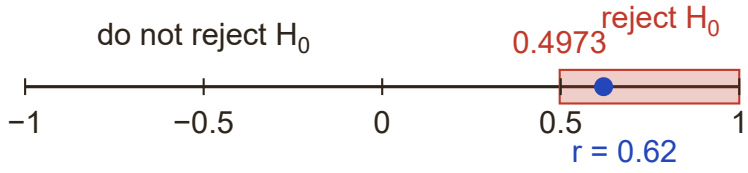
Key facts

- Correlation: $H_0 : \rho = 0$; $H_1 : \rho > 0$, $\rho < 0$ or $\rho \neq 0$
- $X \sim N(\mu, \sigma^2) \Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
- $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
- 1% level: one-tailed $z = 2.3263$; two-tailed $z = \pm 2.5758$
- Two-tailed correlation test at 5%: use the one-tailed 2.5% critical value
- 5% level: one-tailed $z = 1.6449$; two-tailed $z = \pm 1.9600$
- Upper-tail critical region: $\bar{X} > \mu_0 + z \times \frac{\sigma}{\sqrt{n}}$

Notes

Testing for correlation

- The sample PMCC, r, is used to test the population correlation coefficient, ρ
- $H_0 : \rho = 0$ (no correlation in the population). $H_1 : \rho > 0$ or $\rho < 0$ (one-tailed), or $\rho \neq 0$ (two-tailed).
- You are given a table of critical values (by sample size n and significance level) or a p-value. Read the table carefully: a two-tailed test at 5% uses the same critical value as a one-tailed test at 2.5%.
- Reject H_0 if r is beyond the critical value in the direction of H_1: r greater than the critical value for $\rho > 0$, r less than minus the critical value for $\rho < 0$, and $|r|$ greater than the critical value for $\rho \neq 0$

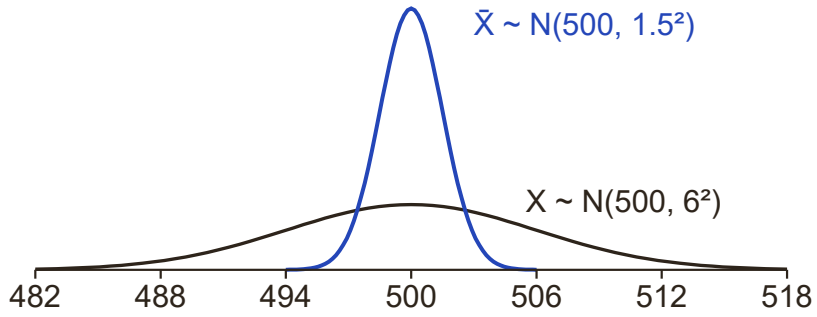


H_1: ρ > 0, n = 12, 5% level: r = 0.62 is in the critical region

- A significant result is evidence of linear correlation in the population. It does not show causation, and with a large sample even a weak correlation can be significant.

The sample mean

- If $X \sim N(\mu, \sigma^2)$, the mean of a random sample of size n has distribution $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
- The standard deviation of $\bar{X}$ is $\frac{\sigma}{\sqrt{n}}$: larger samples give sample means closer to μ



Means of samples of 16 vary far less than single values: σ/√n = 6/4 = 1.5

Testing a normal mean

- $H_0 : \mu = \mu_0$; $H_1 : \mu > \mu_0$, $\mu < \mu_0$ or $\mu \neq \mu_0$. The variance is known, given or assumed unchanged.
- Under H_0, $\bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$
- Either find the p-value, e.g. $P(\bar{X} \leq \bar{x})$, or standardise, $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$, and compare with the critical z value (e.g. 1.6449 for a 5% one-tailed test).
- Or find the critical region for $\bar{X}$, e.g. $\bar{X} > \mu_0 + 1.6449 \times \frac{\sigma}{\sqrt{n}}$, and see whether $\bar{x}$ lies in it.

How to answer each type of question

M = method mark, A = answer mark, B = a mark on its own.

Test for correlation using a critical value

3 to 4 marks | **Grade C**

1. Write the hypotheses in terms of ρ, defined in context.
2. Find the critical value for your n and significance level (half the level for a two-tailed test).
3. Compare r with it, decide, and conclude in context.

Example. A student takes a random sample of 12 days from the large data set for one UK station and finds that the product moment correlation coefficient between daily mean temperature and daily total sunshine is 0.62. Test, at the 5% level of significance, whether there is evidence of positive correlation between these variables. (For $n = 12$, the critical value for a one-tailed test at the 5% level is 0.4973.)

Model answer.

$H_0 : \rho = 0$, $H_1 : \rho > 0$, where ρ is the population correlation coefficient between daily mean temperature and daily total sunshine (B1)

Critical value 0.4973 (B1)

$0.62 > 0.4973$, so reject H_0 (M1)

There is evidence to suggest that there is positive correlation between daily mean temperature and daily total sunshine (A1)

Test a normal mean using a p-value

6 to 7 marks | **Grade B**

1. Write the hypotheses about μ
2. Under H_0, write $\bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$
3. Find the p-value (or z) and compare.
4. Decide and conclude in context.

Example. The masses of flour in bags filled by a machine are normally distributed with standard deviation 6 g. The machine is set to a mean of 500 g. An inspector suspects that the mean has decreased. A random sample of 16 bags has mean mass 497.2 g. Test the inspector's suspicion at the 5% level of significance.

Model answer.

$H_0 : \mu = 500$, $H_1 : \mu < 500$ (B1)

Under H_0, $\bar{X} \sim N\left(500, \frac{6^2}{16}\right)$, i.e. $N(500, 1.5^2)$ (M1)

$P(\bar{X} \leq 497.2) = P(Z \leq -1.867)$ (M1) $= 0.0310$ (A1)

$0.0310 < 0.05$, so reject H_0 (M1)

There is sufficient evidence to suggest that the mean mass of flour in the bags has decreased (A1)

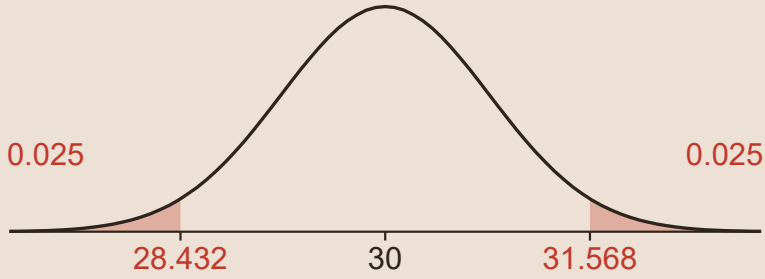
Find a critical region for the sample mean

4 marks | **Grade A**

1. Write the distribution of $\bar{X}$ under H_0
2. Find the z value(s) for the tail(s).
3. Critical values: $\mu_0 \pm z \times \dfrac{\sigma}{\sqrt{n}}$

Example. $X \sim N(\mu, 4^2)$. A test of $H_0 : \mu = 30$ against $H_1 : \mu \neq 30$ at the 5% level of significance uses a random sample of 25 observations. Find the critical region for $\bar{X}$

Model answer.



$\bar{X} \sim N(30, 0.8^2)$, since $\dfrac{4}{\sqrt{25}} = 0.8$ (M1)

$z = \pm 1.9600$ (B1)

$30 \pm 1.96 \times 0.8$ (M1)

$\bar{X} < 28.432$ or $\bar{X} > 31.568$ (A1)

Find the smallest sample size for a significant result

3 to 4 marks | **Grade A***

1. Write the standardised statistic in terms of n
2. Set it beyond the critical z value and rearrange for $\sqrt{n}$
3. Square, and round up to a whole number.

Example. For the flour bags above ($\sigma = 6$ g), find the smallest sample size for which a sample mean of 497.2 g would give evidence at the 1% level that the mean is less than 500 g.

Model answer.

Need $\dfrac{497.2 - 500}{6/\sqrt{n}} < -2.3263$ (M1 B1)

$\sqrt{n} > \dfrac{2.3263 \times 6}{2.8} = 4.985$, so $n > 24.8$ (M1)

$n = 25$ (A1)

Don’t lose marks

- Writing hypotheses in terms of r instead of ρ, or $\bar{x}$ instead of μ
- Using σ instead of $\frac{\sigma}{\sqrt{n}}$ for the sample mean.
- Entering the variance $\frac{\sigma^2}{n}$ into the calculator instead of the standard deviation $\frac{\sigma}{\sqrt{n}}$
- In a two-tailed test, using the critical value for the full significance level (e.g. the one-tailed 5% value for a two-tailed 5% test).
- Saying that a significant correlation proves causation, or means the correlation is strong.

More tips

Memory tricks

- Correlation tables: row = sample size, column = significance level. For a two-tailed test, use the one-tailed column for half the level (or the two-tailed heading if the table shows one).
- The sample mean varies less than single values: divide σ by $\sqrt{n}$
- Comparing z with the critical z value, or the p-value with the significance level, always gives the same decision.

Exam technique

- Define ρ as the population correlation coefficient between the two named variables.
- Write the distribution of $\bar{X}$ under H_0 in full: it usually earns a method mark.
- Conclude about the population (its mean or its correlation) in context, using ‘evidence’ or ‘suggest’.
- If asked for an assumption in a test for a mean, the usual one is that the standard deviation is unchanged.

What each grade needs

What you need to be able to do, from the first marks up to the top grade.

Grade E — **Write hypotheses for a correlation test**
$H_0 : \rho = 0$ and $H_1 : \rho > 0$, $\rho < 0$ or $\rho \neq 0$, with ρ defined.

Grade D — **Compare r with a critical value**
Read the table for n and the significance level, then decide whether r is beyond the critical value.

Grade C — **Carry out a two-tailed correlation test**
Use the one-tailed critical value for half the significance level and compare $|r|$ with it.

Grade C — **State the distribution of the sample mean**
If $X \sim N(\mu, \sigma^2)$, then $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

Grade B — **Test a normal mean using a p-value**
Find $P(\bar{X} \geq \bar{x})$ or $P(\bar{X} \leq \bar{x})$ under H_0 and compare it with the significance level.

Grade A — **Find a critical region for the sample mean**
Use $\mu_0 \pm z\frac{\sigma}{\sqrt{n}}$ with the right z value.

Grade A*

Find the sample size needed for significance

Rearrange the standardised inequality to find the smallest n

Free notes, Full packs, practice papers and mock exams at chhetriacademy.com. Written for this site; not taken from past papers. Stuck? 1-to-1 tutoring with a free 30-minute trial lesson.