The one-sample t test for a mean

Someone claims a value for a population mean, and a sample mean comes out different. Is the gap bigger than chance alone would plausibly make it, when the spread must be estimated from the same sample?

WORDS TO KNOW

μ_0: the null value, the claimed population mean.

Null and alternative hypotheses: $H_0 : \mu = \mu_0$, and H_a, what the investigator suspects: $\mu > \mu_0$, $\mu < \mu_0$ or $\mu \neq \mu_0$.

Significance level α: the cutoff for the p-value, fixed before the data are seen.

t statistic: how many standard errors $\bar{x}$ sits from μ_0.

p-value: the probability, if H_0 were true, of a t at least as far out as the observed one, in the direction H_a names.

KEY IDEAS

1. The steps of the test

Parameter and hypotheses: name μ in context; write $H_0 : \mu = \mu_0$ and H_a, chosen before the data are seen.

Significance level: fix α, often 0.05.

Conditions: random; $n \leq 0.10\,N$ when sampling without replacement; $n \geq 30$, or for a smaller sample a plot with no strong skew and no outliers.

Test statistic:

$$t = \frac{\bar{x} - \mu_0}{s_x / \sqrt{n}}, \qquad \text{df} = n - 1$$

P-value: the tail area of the t curve beyond t.

Decision and conclusion: if the p-value is at most α, reject H_0; otherwise fail to reject H_0. Say it in words about μ, in context, naming α and the p-value.

2. Which tail is the p-value?

The direction comes from H_a, not from the data: the upper tail for $\mu > \mu_0$, the lower tail for $\mu < \mu_0$, and both tails for $\mu \neq \mu_0$. Use tcdf(lower, upper, df) with 10^{99} for "no end". A t table only brackets the p-value between two columns of the row for your df.

3. Two-sided tests match intervals exactly

A two-sided test at level α rejects μ_0 exactly when μ_0 lies outside the $(1 - \alpha)$ t interval.

The match is exact because the test and the interval use the same standard error, $s_x / \sqrt{n}$, and the same t curve.

Free sample from math.wiki for classroom use. Do not post. math.wiki

WATCH OUT

- **Never "accept H_0".** Failing to reject means the data are consistent with μ_0, and with many other values too. It is never evidence that H_0 is true.
- **Hypotheses are about μ, never $\bar{x}$.** The value of $\bar{x}$ is already known.
- **Do not pick the side after seeing the data.** Switching to a one-sided H_a in the direction the data fell halves the p-value with no new evidence.
- **Rejecting is not a cause or a size.** It does not say what changed the mean, or that the change is large enough to matter.

WORKED EXAMPLES

Example 1. A city says buses on route 9 take a mean of 40 minutes end to end. A riders' group suspects longer. It times a random sample of 26 of the 2,000 trips this month: $\bar{x} = 42.6$ minutes, $s_x = 4.0$ minutes, and the dot plot shows mild skew and no outliers. Test at $\alpha = 0.05$. For df = 25 the t table gives 2.060, 2.485 and 2.787 at upper-tail areas 0.025, 0.01 and 0.005.

1. **Parameter and hypotheses.** μ is the mean time of all route 9 trips this month. $H_0\colon \mu = 40$, $H_a\colon \mu > 40$.
2. **Significance level.** $\alpha = 0.05$.
3. **Conditions.** Random sample. $26 \le 0.10(2{,}000) = 200$. With $n = 26 < 30$ the plot decides: mild skew, no outliers. Met.
4. **Test statistic.**
$$\mathrm{SE}_{\bar{x}} = \frac{4.0}{\sqrt{26}} \approx 0.7845 \qquad t = \frac{42.6 - 40}{0.7845} \approx 3.31, \qquad \mathrm{df} = 25$$
5. **P-value.** Upper tail. Since 3.31 > 2.787, the p-value is below 0.005. Technology: $\mathrm{tcdf}(3.31, 10^{99}, 25) \approx 0.0014$.
6. **Decision and conclusion.** 0.0014 ≤ 0.05, so reject H_0. If the mean were 40 minutes, a sample of 26 trips would average 42.6 or more only about 0.14% of the time.

Answer: at the $\alpha = 0.05$ level there is convincing evidence that the mean time of route 9 trips this month is more than 40 minutes.

Example 2. A machine is set to fill bags of rice with a mean of 250 grams. An inspector asks whether today's mean is off in either direction. A random sample of 20 of the day's thousands of bags gives $\bar{x} = 247.1$ grams and $s_x = 7.4$ grams, roughly symmetric with no outliers. Test at $\alpha = 0.05$, then compare with the 95% interval ($t^* = 2.093$ for df = 19).

1. **Parameter and hypotheses.** μ is the mean fill of all the bags filled today. $H_0\colon \mu = 250$, $H_a\colon \mu \ne 250$, at $\alpha = 0.05$.
2. **Conditions.** Random sample; 20 is far under a tenth of thousands; $n < 30$, but the plot shows no strong skew and no outliers.
3. **Test statistic.** $\mathrm{SE}_{\bar{x}} = 7.4/\sqrt{20} \approx 1.6547$, so $t = -2.9 \div 1.6547 \approx -1.75$ with df = 19.
4. **P-value.** Both tails count: with the unrounded $t \approx -1.7526$, $2 \times \mathrm{tcdf}(-10^{99}, -1.7526, 19) \approx 0.0958$.
5. **Decision and conclusion.** 0.0958 > 0.05, so fail to reject H_0. At the $\alpha = 0.05$ level the data do not give convincing evidence that today's mean fill differs from 250 grams. This does not show the mean is 250 grams.
6. **Interval.** $247.1 \pm 2.093(1.6547) = 247.1 \pm 3.463$, from 243.637 to 250.563 grams.

Answer: fail to reject H_0. **Check:** 250 lies inside the 95% interval, as a two-sided test at 0.05 that fails to reject requires.

Free sample from math.wiki for classroom use. Do not post. math.wiki