

# Unit 4: Inference for Quantitative Data: Means

Unit 4 takes everything you learned about proportions in Unit 3 and applies it to means. The structure is the same. You estimate a parameter with a confidence interval, or you test a claim with a hypothesis test. The difference is that means bring the t-distribution, because the population standard deviation is almost never known.

AP Statistics

Inference for Quantitative Data: Means

About 12 minutes to read

## How to use this guide

Read it in order the first time. The sampling distribution sets up the interval, the interval sets up the test, and the one-sample procedures set up the two-sample ones. Every procedure in this unit follows the same rhythm: state the parameter, check the conditions, do the calculation, and conclude in context.

After the first read, use the trap boxes and the comparison table to review the distinctions the exam tests most. The paired versus two-sample choice is the single most tested decision in this unit. Finish with the practice questions, then work through the recall check on the last page.

**What this unit is worth.** Inference is the heaviest part of the AP Statistics exam, and means show up in both the multiple-choice and free-response sections. The free-response questions almost always ask you to check conditions by name and to interpret your result in context. A correct calculation with no condition check and no contextual conclusion loses most of the points.

## 4.1 The Sampling Distribution of a Sample Mean

The **sampling distribution of a sample mean** is the distribution of  $\bar{x}$  values you would get if you took every possible sample of size  $n$  from the population. Its center is the population mean, so  $\mu(\bar{x}) = \mu$ . Its spread is smaller than the population spread:  $\sigma(\bar{x}) = \sigma/\sqrt{n}$ . That denominator is the whole point of taking larger samples. Quadrupling the sample size cuts the standard deviation of  $\bar{x}$  in half, which is why big samples give more precise estimates.

Whether that sampling distribution is normal depends on the **normality condition**. If the population itself can be modeled by a normal distribution, then  $\bar{x}$  is normal for any sample size. If not, the Central Limit Theorem says  $\bar{x}$  is approximately normal once  $n$  is at least 30. When the population is extremely skewed, you may need a sample much larger than 30 before the approximation is trustworthy.

Two more conditions protect the math underneath. The **randomization condition** requires the data to come from a random sample, or for two-sample work, two independent random samples or a randomized experiment. In an experiment, random assignment of treatments is what matters, and it is the only condition you need. The **10% condition** says that when sampling without replacement, the population must be at least 10 times the sample size. This keeps the independence assumption reasonable. You skip the 10% condition for data from a randomized experiment.

**Trap.** The Central Limit Theorem describes the sampling distribution of  $\bar{x}$ , not the population and not the sample. A question that shows a skewed population histogram and asks about the distribution of sample means wants the CLT answer. A question that asks about the population itself does not.

## 4.2 The t-Distribution and the One-Sample t-Interval

For proportions you used  $z$  because the standard error came straight from the sample proportion. For means, the population standard deviation  $\sigma$  is almost never known, so you estimate it with the sample standard deviation  $s$ . That extra estimation adds uncertainty, and the **t-distribution** accounts for it. It is a family of symmetric, bell-shaped curves with wider tails than the standard normal. Each one is identified by its **degrees of freedom**,  $df = n - 1$  for one-sample procedures. With small  $df$  the peak is narrower and the tails fatter. As  $df$  grows, the t-distribution looks more and more like the standard normal.

The confidence interval for a population mean has the familiar form: estimate  $\pm$  margin of error. The estimate is  $\bar{x}$ . The **standard error** is  $SE(\bar{x}) = s/\sqrt{n}$ . The **critical value**  $t^*$  is the value that encloses the middle  $C\%$  of the t-distribution with  $n - 1$  degrees of freedom. The **margin of error** is  $t^*$  times the standard error. Put together, the **one-sample t-interval** is  $\bar{x} \pm t^*(s/\sqrt{n})$ .

Before you calculate, check the **sample data condition**. One of these must hold: the population is indicated to be approximately normal, the sample size is at least 30, or, if  $n$  is below 30, the sample data show no strong skewness and no outliers. Also **state the parameter in context** before you start. Name the population mean, the response variable, and the population. For example:  $\mu$  is the true mean battery life, in hours, for all phones of this model.

**Trap.** Use  $t$ , not  $z$ , whenever  $\sigma$  is unknown and you are working with  $s$ . On the AP exam  $\sigma$  is essentially never given for means problems. If you reach for  $z^* = 1.96$  out of habit from the proportions unit, you are using the wrong critical value and your interval will be too narrow.

## 4.2 Continued: Matched Pairs

Some studies produce two measurements that belong together. The same subjects measured before and after a treatment, or twins split between two groups, give **paired** data. The two samples are dependent, so two-sample procedures do not apply. Instead, subtract within each pair to get one sample of differences, then run the one-sample  $t$  procedures on the mean difference. The interval becomes  $\bar{d} \pm t^*(s_d/\sqrt{n})$ , where  $n$  is the number of differences and  $df = n - 1$ .

**Two-sample  $t$  procedures require two independent samples.** That means two separate groups with no natural pairing between individual observations. The decision comes down to the study design, not the numbers. If each observation in one sample can be matched to exactly one observation in the other, the data are paired. If the samples were drawn separately with no such link, they are independent.

**Trap.** The paired versus two-sample choice is the most tested decision in this unit, and it is decided by the design. Twenty students take a test before and after a review course: paired, because each student is their own control. Twenty students in a morning class compared with twenty different students in an afternoon class: two-sample, because the groups are separate. Running a two-sample procedure on paired data throws away the pairing and weakens the analysis.

## 4.3 Reading a Confidence Interval

The **interpretation** follows a fixed sentence: we are  $C\%$  confident that the interval from  $a$  to  $b$  contains the true population mean. The confidence is about the method, not this particular interval. If you repeated the sampling process many times, about  $C\%$  of the intervals built this way would capture the true mean. This one either does or does not.

A confidence interval is also an answer to the question of which claims are plausible. **Justifying a claim** from an interval means checking whether the claimed value sits inside it. Values inside the interval are consistent with the data. Values outside are not. Two facts control the width. Raising the confidence level raises  $t^*$ , which widens the interval. Raising the sample size shrinks the standard error, which narrows it. Width shrinks roughly in proportion to  $1/\sqrt{n}$ , so quadrupling  $n$  cuts the width roughly in half.

**Trap.** Never say there is a  $C\%$  probability that the true mean is in this interval. The parameter is fixed. The randomness was in the sampling, and the confidence level describes the long-run success rate of the method. Write the interpretation sentence exactly and in context.

## 4.4–4.5 The One-Sample t-Test

A hypothesis test starts by writing the hypotheses. For a population mean, the **null hypothesis** is  $H_0: \mu = \mu_0$ , where  $\mu_0$  is the claimed value. The **alternative** states the direction of interest:  $H_a: \mu < \mu_0$  or  $H_a: \mu > \mu_0$  for a one-sided test,  $H_a: \mu \neq \mu_0$  for a two-sided test. For a matched-pairs test on the mean difference, the null is  $H_0: \mu_d = 0$ , with one-sided alternatives  $H_a: \mu_d < 0$  or  $H_a: \mu_d > 0$ , and two-sided  $H_a: \mu_d \neq 0$ .

The **test statistic** measures how far the sample mean is from the null value in standard errors:  $t = (\bar{x} - \mu_0)/(s/\sqrt{n})$ . When  $H_0$  is true, this statistic follows a t-distribution with  $n - 1$  degrees of freedom. The **p-value** is the probability of getting a test statistic as extreme as, or more extreme than, the one observed, in the direction of the alternative, assuming the null hypothesis is true. State it in context: if the true mean equaled the null value, the chance of a sample mean this far from it would be the p-value.

The conclusion has two allowed forms. If the p-value is below the significance level, reject  $H_0$  and say the data give convincing evidence for the alternative, in context. If not, fail to reject  $H_0$  and say the data do not give convincing evidence for the alternative. You never accept the null. The conditions are the same three as for the interval: random, 10%, and the sample data condition.

**Trap.** The p-value is not the probability that the null hypothesis is true, and it is not the probability that your conclusion is wrong. It is a conditional probability computed under the assumption that  $H_0$  is true. A p-value of 0.03 does not mean there is a 3% chance the null is true. It means that if the null were true, you would see data this extreme about 3% of the time.

## 4.6–4.7 Inference for Two Means

When the samples are independent, you work with the **sampling distribution of the difference in sample means**. Its center is the difference of the population means,  $\mu_1 - \mu_2$ . Its standard deviation is  $\sqrt{(\sigma_1^2/n_1 + \sigma_2^2/n_2)}$ . The distribution is normal if both populations are normal, and approximately normal when both sample sizes are at least 30. Since the population standard deviations are unknown, the **standard error** uses the sample versions:  $SE(\bar{x}_1 - \bar{x}_2) = \sqrt{(s_1^2/n_1 + s_2^2/n_2)}$ .

The **two-sample t-interval** is  $(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{(s_1^2/n_1 + s_2^2/n_2)}$ , where  $t^*$  comes from the t-distribution with degrees of freedom found by technology. The df lands somewhere between  $n_1 + n_2 - 2$  and the smaller of  $n_1 - 1$  and  $n_2 - 1$ . The **sample data condition** for two samples: both sample sizes at least 30, or both populations indicated approximately normal. If either sample is smaller than 30, both sample distributions must be free from strong skewness and outliers.

**Trap.** Do not pool the sample variances. The AP Statistics two-sample t procedures always use the unpooled standard error  $\sqrt{(s_1^2/n_1 + s_2^2/n_2)}$ . Pooling belongs to a different procedure that this course does not use. Also, do not invent the degrees of freedom. On the exam the df come from technology or are given, and the conservative option is the smaller of  $n_1 - 1$  and  $n_2 - 1$ .

## 4.8–4.10 The Two-Sample t-Test

The hypotheses compare the two population means. The null states no difference:  $H_0: \mu_1 - \mu_2 = 0$ , equivalently  $H_0: \mu_1 = \mu_2$ . The alternative takes the usual three forms, for example  $H_a: \mu_1 - \mu_2 > 0$  for a one-sided test that the first mean is larger. The **test statistic** is  $t = ((\bar{x}_1 - \bar{x}_2) - 0) / \sqrt{(s_1^2/n_1 + s_2^2/n_2)}$ . The p-value and conclusion work exactly as in the one-sample test, with the same three conditions checked for both samples.

The confidence interval for the difference doubles as a test. If the interval for  $\mu_1 - \mu_2$  contains 0, the data do not give convincing evidence of a difference between the population means. If 0 falls outside the interval, they do. The sign of the interval tells the direction: an interval entirely above 0 supports  $\mu_1 > \mu_2$ , and an interval entirely below 0 supports  $\mu_1 < \mu_2$ .

## Confusions That Cost Points

| PAIR                             | HOW TO KEEP THEM STRAIGHT                                                                                                                                                                  |
|----------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Paired vs two-sample             | Paired means the data come in linked pairs, so subtract within pairs and run a one-sample t on the differences. Two-sample means two separate groups. The design decides, not the numbers. |
| t vs z for means                 | Use t whenever $\sigma$ is unknown, which is nearly always. The z procedures from the proportions unit do not transfer to means.                                                           |
| df for one-sample vs paired      | One-sample: $n - 1$ . Paired: number of differences $- 1$ . It is the count of differences, not the total number of measurements.                                                          |
| CI interpretation vs probability | We are C% confident the interval captures the parameter. The confidence describes the method over many repetitions, not the chance for this one interval.                                  |
| Reject vs accept $H_0$           | You reject $H_0$ or fail to reject it. Failing to reject is not proof the null is true. It means the data were not convincing.                                                             |
| Unpooled vs pooled SE            | Always unpooled in this course: $\sqrt{(s_1^2/n_1 + s_2^2/n_2)}$ . Pooling is a different procedure the AP exam does not ask for.                                                          |

# Practice Questions

Original questions written for this guide in the style of the AP exam. Answers and explanations are on the next page, so complete the questions before checking them.

1. A researcher takes a random sample of 40 batteries and finds a mean life of 12.4 hours with a standard deviation of 1.8 hours. Which of the following is the correct 95% confidence interval for the true mean battery life?
  - A.  $12.4 \pm 1.96(1.8/\sqrt{40})$
  - B.  $12.4 \pm t^*(1.8/\sqrt{40})$ , where  $t^*$  is the critical value for the middle 95% of the t-distribution with 39 degrees of freedom
  - C.  $12.4 \pm 1.96(1.8/40)$
  - D.  $12.4 \pm t^*(1.8/40)$ , where  $t^*$  is the critical value for the middle 95% of the t-distribution with 40 degrees of freedom
2. Fifteen runners complete a time trial, then follow a six-week training program, then complete a second time trial. Which procedure is appropriate for testing whether the program reduced mean times?
  - A. Two-sample t-test, because there are two sets of times
  - B. One-sample t-test on the differences (before minus after) for each runner
  - C. Two-sample z-test, because  $n = 15$  in each group
  - D. One-sample z-test on the after times only
3. A 95% confidence interval for the difference in mean reaction times (caffeine minus no caffeine) is (0.04, 0.31) seconds. Which conclusion is justified?
  - A. There is a 95% probability that caffeine increases reaction time by between 0.04 and 0.31 seconds.
  - B. We are 95% confident the true mean difference lies between 0.04 and 0.31 seconds, which gives convincing evidence that caffeine increases mean reaction time.
  - C. The interval proves that every person reacts 0.04 to 0.31 seconds slower with caffeine.
  - D. Because 0 is close to the interval, there is no evidence of a difference.
4. In a one-sample t-test of  $H_0: \mu = 50$  against  $H_a: \mu > 50$ , the p-value is 0.021. Which statement is correct?
  - A. There is a 2.1% probability that the null hypothesis is true.
  - B. If the true mean were 50, the probability of getting a sample mean this far above 50, or farther, is 0.021.
  - C. There is a 97.9% probability that the alternative hypothesis is true.
  - D. The sample mean must have been below 50.

## Answer Key

- 1. B.** The population standard deviation is unknown, so use the t-distribution with  $n - 1 = 39$  degrees of freedom. The standard error is  $s/\sqrt{n} = 1.8/\sqrt{40}$ . A uses  $z^*$ , which is wrong when  $\sigma$  is unknown. C divides by  $n$  instead of  $\sqrt{n}$ . D makes both errors: wrong denominator and wrong degrees of freedom.
- 2. B.** Each runner is measured twice, so the data are paired. Subtract within each runner to get 15 differences, then run a one-sample t-test on the mean difference. A ignores the pairing and treats the two time trials as independent groups. C and D use z procedures, which do not apply when  $\sigma$  is unknown, and z also requires larger samples or known parameters.
- 3. B.** The interpretation follows the fixed form, stated in context, and because the entire interval sits above 0, it gives convincing evidence that the true mean difference is positive. A turns the confidence level into a probability about this specific interval. C claims the interval describes every individual, but it describes the mean difference. D misreads the interval: 0 is outside it, not close enough to matter.
- 4. B.** The p-value is computed assuming  $H_0$  is true: if  $\mu$  really were 50, a sample mean this far above 50 would occur with probability 0.021. A and C treat the p-value as the probability that a hypothesis is true, which it is not. D contradicts the setup: a right-tailed test with a small p-value means the sample mean was above 50, not below.

# One-Page Recall Check

- State the mean and standard deviation of the sampling distribution of  $\bar{x}$ , and explain why larger samples give a smaller standard deviation.
- State the normality condition for the sampling distribution of  $\bar{x}$ , including the  $n \geq 30$  rule and the extreme-skewness exception.
- State the randomization condition and the 10% condition, and say when each one is unnecessary.
- Explain why the t-distribution is used for means instead of the standard normal, and give the degrees of freedom for a one-sample procedure.
- Write the one-sample t-interval formula and identify each piece: estimate, critical value, standard error.
- State the sample data condition for one-sample t procedures.
- Explain how to handle matched-pairs data and why two-sample procedures are wrong for it.
- Interpret a confidence interval for a mean in context, using the correct wording.
- Explain how confidence level and sample size each affect interval width.
- Write the null and alternative hypotheses for a one-sample t-test and for a matched-pairs test.
- Write the one-sample t test statistic and interpret a p-value in context.
- State the mean and standard deviation of the sampling distribution of  $\bar{x}_1 - \bar{x}_2$ .
- Write the two-sample t-interval and the two-sample t test statistic, using the unpooled standard error.
- Explain how a confidence interval for  $\mu_1 - \mu_2$  can serve as a hypothesis test.

**Where to go next.** Turn every missed item above into flashcards and drill them spaced out over several days rather than in one sitting. In Rycal, open the Inference for Quantitative Data deck under AP Statistics. The deck covers the terms in this guide, and its practice questions target the same traps named here. If you have a test date, add it in the Test Planner. You can also start your next review with a Brain Dump, then check what you missed against this guide.

## Key terms for this unit

Sampling distribution of a sample mean, Randomization condition (means), 10% condition (means), Normality condition (sampling distribution of a sample mean), t-distribution (Student's t), Degrees of freedom (t), One-sample t-interval for a population mean, One-sample t-interval for a population mean difference (matched pairs), Paired vs. independent samples, Sample data condition (one-sample t procedures), Standard error (sample mean), Margin of error (t-interval), Critical value ( $t^*$ ), Stating the parameter in context (mean / mean difference), Confidence interval interpretation (population mean or mean difference), Relationships among confidence level, margin of error, sample size, and interval width (means), Justifying a claim based on a confidence interval (population mean or mean difference), One-sample t-test for a population mean, One-sample t-test for a population mean difference (matched pairs), Null and alternative hypotheses (one-sample t-test for a mean), Null and alternative hypotheses (one-sample t-test for a mean difference), Test statistic (one-sample t-test), p-value interpretation (means t-test), Sampling distribution of the difference in sample means, Normality condition (sampling distribution of the difference in sample means), Two-sample t-interval for a difference between population means, Sample data condition (two-sample t procedures), Standard error (difference in means), Margin of error (two-sample t-interval), Confidence interval for a difference in means (interpretation and claim), Two-sample t-test for a difference between two population means, Null and alternative hypotheses (two-sample t-test for means), Test statistic (two-sample t-test for means).

**KEEP STUDYING**

## **Study this unit in Rycal**

Drill the flashcards, run the practice questions, and build a study plan for your test date.

**<https://rycal.web.app/apstats>**