RATIOMERA

Complete Solutions

Hypothesis Testing and Confidence Intervals

Ratiomera Statistics

These complete solutions use the same identifiers and order as the Exercise Sheet. Intermediate values are retained until the stated rounding step, so small differences caused by earlier rounding are acceptable where noted. All settings, values, data, and software outputs are constructed teaching material; they are not empirical findings.

Part I: Theory

A01: Hypotheses and the Two Possible Errors

`T03-A01-V01`: Guided map practice

Identify the issue

A hypothesis concerns population means, not the two observed sample means.

A suitable formulation is $H_0 : \mu_T \leq \mu_C$ and $H_1 : \mu_T > \mu_C$.

Reason through the evidence

A Type I error would mean concluding that the population mean route-planning score is higher in the treatment population when the null hypothesis is true.

A Type II error would mean failing to detect that the population mean route-planning score is higher in the treatment population when that relationship is genuinely present.

State the conclusion and its limits

The sample provides uncertain evidence rather than direct access to both population means.

We therefore know the long-run error rates of the procedure under specified conditions, but the decision from one study does not come with a label telling us whether it was correct.

`T03-A01-V02`: Quiet reading spaces

Identify the issue

A hypothesis concerns population means, not the two observed sample means.

A suitable formulation is $H_0 : \mu_T = \mu_C$ and $H_1 : \mu_T \neq \mu_C$.

Reason through the evidence

A Type I error would mean concluding that the population mean reading-comprehension score is different between the two populations when the null hypothesis is true.

A Type II error would mean failing to detect that the population mean reading-comprehension score is different between the two populations when that relationship is genuinely present.

State the conclusion and its limits

The sample provides uncertain evidence rather than direct access to both population means.

We therefore know the long-run error rates of the procedure under specified conditions, but the decision from one study does not come with a label telling us whether it was correct.

`T03-A01-V03`: Archive-search checklist

Identify the issue

A hypothesis concerns population means, not the two observed sample means.

A suitable formulation is $H_0 : \mu_T \leq \mu_C$ and $H_1 : \mu_T > \mu_C$.

Reason through the evidence

A Type I error would mean concluding that the population mean number of correctly retrieved records is higher in the treatment population when the null hypothesis is true.

A Type II error would mean failing to detect that the population mean number of correctly retrieved records is higher in the treatment population when that relationship is genuinely present.

State the conclusion and its limits

The sample provides uncertain evidence rather than direct access to both population means.

We therefore know the long-run error rates of the procedure under specified conditions, but the decision from one study does not come with a label telling us whether it was correct.

`T03-A01-V04`: Short movement break

Identify the issue

A hypothesis concerns population means, not the two observed sample means.

A suitable formulation is $H_0 : \mu_T \leq \mu_C$ and $H_1 : \mu_T > \mu_C$.

Reason through the evidence

A Type I error would mean concluding that the population mean sustained-attention score is higher in the treatment population when the null hypothesis is true.

A Type II error would mean failing to detect that the population mean sustained-attention score is higher in the treatment population when that relationship is genuinely present.

State the conclusion and its limits

The sample provides uncertain evidence rather than direct access to both population means.

We therefore know the long-run error rates of the procedure under specified conditions, but the decision from one study does not come with a label telling us whether it was correct.

`T03-A01-V05`: Captioned tutorial

Identify the issue

A hypothesis concerns population means, not the two observed sample means.

A suitable formulation is $H_0 : \mu_T = \mu_C$ and $H_1 : \mu_T \neq \mu_C$.

Reason through the evidence

A Type I error would mean concluding that the population mean tutorial-comprehension score is different between the two populations when the null hypothesis is true.

A Type II error would mean failing to detect that the population mean tutorial-comprehension score is different between the two populations when that relationship is genuinely present.

State the conclusion and its limits

The sample provides uncertain evidence rather than direct access to both population means.

We therefore know the long-run error rates of the procedure under specified conditions, but the decision from one study does not come with a label telling us whether it was correct.

`T03-A01-V06`: Reduced notification setting

Identify the issue

A hypothesis concerns population means, not the two observed sample means.

A suitable formulation is $H_0 : \mu_T \geq \mu_C$ and $H_1 : \mu_T < \mu_C$.

Reason through the evidence

A Type I error would mean concluding that the population mean task-completion time is lower in the treatment population when the null hypothesis is true.

A Type II error would mean failing to detect that the population mean task-completion time is lower in the treatment population when that relationship is genuinely present.

State the conclusion and its limits

The sample provides uncertain evidence rather than direct access to both population means.

We therefore know the long-run error rates of the procedure under specified conditions, but the decision from one study does not come with a label telling us whether it was correct.

`T03-A01-V07`: Retrieval-practice cards

Identify the issue

A hypothesis concerns population means, not the two observed sample means.

A suitable formulation is $H_0 : \mu_T \leq \mu_C$ and $H_1 : \mu_T > \mu_C$.

Reason through the evidence

A Type I error would mean concluding that the population mean delayed-recall score is higher in the treatment population when the null hypothesis is true.

A Type II error would mean failing to detect that the population mean delayed-recall score is higher in the treatment population when that relationship is genuinely present.

State the conclusion and its limits

The sample provides uncertain evidence rather than direct access to both population means.

We therefore know the long-run error rates of the procedure under specified conditions, but the decision from one study does not come with a label telling us whether it was correct.

`T03-A01-V08`: Museum orientation map

Identify the issue

A hypothesis concerns population means, not the two observed sample means.

A suitable formulation is $H_0 : \mu_T \geq \mu_C$ and $H_1 : \mu_T < \mu_C$.

Reason through the evidence

A Type I error would mean concluding that the population mean number of navigation errors is lower in the treatment population when the null hypothesis is true.

A Type II error would mean failing to detect that the population mean number of navigation errors is lower in the treatment population when that relationship is genuinely present.

State the conclusion and its limits

The sample provides uncertain evidence rather than direct access to both population means.

We therefore know the long-run error rates of the procedure under specified conditions, but the decision from one study does not come with a label telling us whether it was correct.

`T03-A01-V09`: Peer explanation prompt

Identify the issue

A hypothesis concerns population means, not the two observed sample means.

A suitable formulation is $H_0 : \mu_T = \mu_C$ and $H_1 : \mu_T \neq \mu_C$.

Reason through the evidence

A Type I error would mean concluding that the population mean concept-explanation score is different between the two populations when the null hypothesis is true.

A Type II error would mean failing to detect that the population mean concept-explanation score is different between the two populations when that relationship is genuinely present.

State the conclusion and its limits

The sample provides uncertain evidence rather than direct access to both population means.

We therefore know the long-run error rates of the procedure under specified conditions, but the decision from one study does not come with a label telling us whether it was correct.

`T03-A01-V10`: Structured note template

Identify the issue

A hypothesis concerns population means, not the two observed sample means.

A suitable formulation is $H_0 : \mu_T \leq \mu_C$ and $H_1 : \mu_T > \mu_C$.

Reason through the evidence

A Type I error would mean concluding that the population mean number of correctly identified arguments is higher in the treatment population when the null hypothesis is true.

A Type II error would mean failing to detect that the population mean number of correctly identified arguments is higher in the treatment population when that relationship is genuinely present.

State the conclusion and its limits

The sample provides uncertain evidence rather than direct access to both population means.

We therefore know the long-run error rates of the procedure under specified conditions, but the decision from one study does not come with a label telling us whether it was correct.

A09: t Quantiles and Degrees of Freedom

`T03-A09-V01`: Quantiles with 5 degrees of freedom

Identify the issue

By symmetry, $t_{0.01}(5) = -t_{0.99}(5) = -3.3649$.

Reason through the evidence

The upper quantile satisfies $P(T \leq 3.3649) = 0.99$; the lower one satisfies $P(T \leq -3.3649) = 0.01$.

State the conclusion and its limits

Its magnitude 3.3649 is larger than 2.2363 because a t distribution with finite degrees of freedom has heavier tails than the standard normal.

As the degrees of freedom grow, the extra uncertainty from estimating σ diminishes and the t distribution approaches the standard normal distribution.

`T03-A09-V02`: Quantiles with 6 degrees of freedom

Identify the issue

By symmetry, $t_{0.01}(6) = -t_{0.99}(6) = -3.1427$.

Reason through the evidence

The upper quantile satisfies $P(T \leq 3.1427) = 0.99$; the lower one satisfies $P(T \leq -3.1427) = 0.01$.

State the conclusion and its limits

Its magnitude 3.1427 is larger than 2.3263 because a t distribution with finite degrees of freedom has heavier tails than the standard normal.

As the degrees of freedom grow, the extra uncertainty from estimating σ diminishes and the t distribution approaches the standard normal distribution.

`T03-A09-V03`: Quantiles with 7 degrees of freedom

Identify the issue

By symmetry, $t_{0.01}(7) = -t_{0.99}(7) = -2.9980$.

Reason through the evidence

The upper quantile satisfies $P(T \leq 2.9980) = 0.99$; the lower one satisfies $P(T \leq -2.9980) = 0.01$.

State the conclusion and its limits

Its magnitude 2.9980 is larger than 2.3263 because a t distribution with finite degrees of freedom has heavier tails than the standard normal.

As the degrees of freedom grow, the extra uncertainty from estimating σ diminishes and the t distribution approaches the standard normal distribution.

`T03-A09-V04`: Quantiles with 8 degrees of freedom

Identify the issue

By symmetry, $t_{0.01}(8) = -t_{0.99}(8) = -2.8965$.

Reason through the evidence

The upper quantile satisfies $P(T \leq 2.8965) = 0.99$; the lower one satisfies $P(T \leq -2.8965) = 0.01$.

State the conclusion and its limits

Its magnitude 2.8965 is larger than 2.3263 because a t distribution with finite degrees of freedom has heavier tails than the standard normal.

As the degrees of freedom grow, the extra uncertainty from estimating σ diminishes and the t distribution approaches the standard normal distribution.

`T03-A09-V05`: Quantiles with 9 degrees of freedom

Identify the issue

By symmetry, $t_{0.01}(9) = -t_{0.99}(9) = -2.8214$.

Reason through the evidence

The upper quantile satisfies $P(T \leq 2.8214) = 0.99$; the lower one satisfies $P(T \leq -2.8214) = 0.01$.

State the conclusion and its limits

Its magnitude 2.8214 is larger than 2.3263 because a t distribution with finite degrees of freedom has heavier tails than the standard normal.

As the degrees of freedom grow, the extra uncertainty from estimating σ diminishes and the t distribution approaches the standard normal distribution.

`T03-A09-V06`: Quantiles with 10 degrees of freedom

Identify the issue

By symmetry, $t_{0.01}(10) = -t_{0.99}(10) = -2.7638$.

Reason through the evidence

The upper quantile satisfies $P(T \leq 2.7638) = 0.99$; the lower one satisfies $P(T \leq -2.7638) = 0.01$.

State the conclusion and its limits

Its magnitude 2.7638 is larger than 2.3263 because a t distribution with finite degrees of freedom has heavier tails than the standard normal.

As the degrees of freedom grow, the extra uncertainty from estimating σ diminishes and the t distribution approaches the standard normal distribution.

`T03-A09-V07`: Quantiles with 11 degrees of freedom

Identify the issue

By symmetry, $t_{0.01}(11) = -t_{0.99}(11) = -2.7181$.

Reason through the evidence

The upper quantile satisfies $P(T \leq 2.7181) = 0.99$; the lower one satisfies $P(T \leq -2.7181) = 0.01$.

State the conclusion and its limits

Its magnitude 2.7181 is larger than 2.3263 because a t distribution with finite degrees of freedom has heavier tails than the standard normal.

As the degrees of freedom grow, the extra uncertainty from estimating σ diminishes and the t distribution approaches the standard normal distribution.

`T03-A09-V08`: Quantiles with 12 degrees of freedom

Identify the issue

By symmetry, $t_{0.01}(12) = -t_{0.99}(12) = -2.6810$.

Reason through the evidence

The upper quantile satisfies $P(T \leq 2.6810) = 0.99$; the lower one satisfies $P(T \leq -2.6810) = 0.01$.

State the conclusion and its limits

Its magnitude 2.6810 is larger than 2.3263 because a t distribution with finite degrees of freedom has heavier tails than the standard normal.

As the degrees of freedom grow, the extra uncertainty from estimating σ diminishes and the t distribution approaches the standard normal distribution.

`T03-A09-V09`: Quantiles with 15 degrees of freedom

Identify the issue

By symmetry, $t_{0.01}(15) = -t_{0.99}(15) = -2.6025$.

Reason through the evidence

The upper quantile satisfies $P(T \leq 2.6025) = 0.99$; the lower one satisfies $P(T \leq -2.6025) = 0.01$.

State the conclusion and its limits

Its magnitude 2.6025 is larger than 2.3263 because a t distribution with finite degrees of freedom has heavier tails than the standard normal.

As the degrees of freedom grow, the extra uncertainty from estimating σ diminishes and the t distribution approaches the standard normal distribution.

`T03-A09-V10`: Quantiles with 20 degrees of freedom

Identify the issue

By symmetry, $t_{0.01}(20) = -t_{0.99}(20) = -2.5280$.

Reason through the evidence

The upper quantile satisfies $P(T \leq 2.5280) = 0.99$; the lower one satisfies $P(T \leq -2.5280) = 0.01$.

State the conclusion and its limits

Its magnitude 2.5280 is larger than 2.3263 because a t distribution with finite degrees of freedom has heavier tails than the standard normal.

As the degrees of freedom grow, the extra uncertainty from estimating σ diminishes and the t distribution approaches the standard normal distribution.

Part II: Calculator Practice

A02: Sample Estimates and the Standard Error

`T03-A02-V01`: Daily reading minutes

Set up the calculation

There are $n = 9$ observations and their total is 333.00.

Thus $\bar{x} = 333.00/9 = 37.000$ minutes.

Work through the calculation

The squared deviations from $\bar{x}$ sum to 296.000, so $s^2 = 296.000/(9 - 1) = 37.000$ squared minutes, and $s = 6.083$ minutes.

The sample mean estimates the population mean μ; the sample variance estimates the population variance σ^2.

Interpret and check the result

The plug-in standard error is $s/\sqrt{n} = 6.083/\sqrt{9} = 2.028$ minutes.

It estimates the standard deviation of sample means across repeated samples of the same size.

It is not the spread of the individual observations.

`T03-A02-V02`: Archive requests completed

Set up the calculation

There are $n = 8$ observations and their total is 52.00.

Thus $\bar{x} = 52.00/8 = 6.500$ requests.

Work through the calculation

The squared deviations from $\bar{x}$ sum to 18.000, so $s^2 = 18.000/(8 - 1) = 2.571$ squared requests, and $s = 1.604$ requests.

The sample mean estimates the population mean μ; the sample variance estimates the population variance σ^2.

Interpret and check the result

The plug-in standard error is $s/\sqrt{n} = 1.604/\sqrt{8} = 0.567$ requests.

It estimates the standard deviation of sample means across repeated samples of the same size.

It is not the spread of the individual observations.

`T03-A02-V03`: Focus ratings

Set up the calculation

There are $n = 10$ observations and their total is 132.00.

Thus $\bar{x} = 132.00/10 = 13.200$ points.

Work through the calculation

The squared deviations from $\bar{x}$ sum to 33.600, so $s^2 = 33.600/(10 - 1) = 3.733$ squared points, and $s = 1.932$ points.

The sample mean estimates the population mean μ; the sample variance estimates the population variance σ^2.

Interpret and check the result

The plug-in standard error is $s/\sqrt{n} = 1.932/\sqrt{10} = 0.611$ points.

It estimates the standard deviation of sample means across repeated samples of the same size.

It is not the spread of the individual observations.

`T03-A02-V04`: Museum visit durations

Set up the calculation

There are $n = 8$ observations and their total is 459.00.

Thus $\bar{x} = 459.00/8 = 57.375$ minutes.

Work through the calculation

The squared deviations from $\bar{x}$ sum to 233.875, so $s^2 = 233.875/(8 - 1) = 33.411$ squared minutes, and $s = 5.780$ minutes.

The sample mean estimates the population mean μ; the sample variance estimates the population variance σ^2.

Interpret and check the result

The plug-in standard error is $s/\sqrt{n} = 5.780/\sqrt{8} = 2.044$ minutes.

It estimates the standard deviation of sample means across repeated samples of the same size.

It is not the spread of the individual observations.

`T03-A02-V05`: Correctly coded records

Set up the calculation

There are $n = 9$ observations and their total is 177.00.

Thus $\bar{x} = 177.00/9 = 19.667$ records.

Work through the calculation

The squared deviations from $\bar{x}$ sum to 44.000, so $s^2 = 44.000/(9 - 1) = 5.500$ squared records, and $s = 2.345$ records.

The sample mean estimates the population mean μ; the sample variance estimates the population variance σ^2.

Interpret and check the result

The plug-in standard error is $s/\sqrt{n} = 2.345/\sqrt{9} = 0.782$ records.

It estimates the standard deviation of sample means across repeated samples of the same size.

It is not the spread of the individual observations.

`T03-A02-V06`: Weekly practice hours
Set up the calculation

There are $n = 10$ observations and their total is 55.00.

Thus $\bar{x} = 55.00/10 = 5.500$ hours.

Work through the calculation

The squared deviations from $\bar{x}$ sum to 22.500, so $s^2 = 22.500/(10 - 1) = 2.500$ squared hours, and $s = 1.581$ hours.

The sample mean estimates the population mean μ; the sample variance estimates the population variance σ^2.

Interpret and check the result

The plug-in standard error is $s/\sqrt{n} = 1.581/\sqrt{10} = 0.500$ hours.

It estimates the standard deviation of sample means across repeated samples of the same size.

It is not the spread of the individual observations.

`T03-A02-V07`: Route-planning scores
Set up the calculation

There are $n = 9$ observations and their total is 646.00.

Thus $\bar{x} = 646.00/9 = 71.778$ points.

Work through the calculation

The squared deviations from $\bar{x}$ sum to 143.556, so $s^2 = 143.556/(9 - 1) = 17.944$ squared points, and $s = 4.236$ points.

The sample mean estimates the population mean μ; the sample variance estimates the population variance σ^2.

Interpret and check the result

The plug-in standard error is $s/\sqrt{n} = 4.236/\sqrt{9} = 1.412$ points.

It estimates the standard deviation of sample means across repeated samples of the same size.

It is not the spread of the individual observations.

`T03-A02-V08`: Response latencies
Set up the calculation

There are $n = 8$ observations and their total is 178.00.

Thus $\bar{x} = 178.00/8 = 22.250$ seconds.

Work through the calculation

The squared deviations from $\bar{x}$ sum to 55.500, so $s^2 = 55.500/(8 - 1) = 7.929$ squared seconds, and $s = 2.816$ seconds.

The sample mean estimates the population mean μ; the sample variance estimates the population variance σ^2.

Interpret and check the result

The plug-in standard error is $s/\sqrt{n} = 2.816/\sqrt{8} = 0.996$ seconds.

It estimates the standard deviation of sample means across repeated samples of the same size.

It is not the spread of the individual observations.

`T03-A02-V09`: Completed diary prompts
Set up the calculation

There are $n = 9$ observations and their total is 87.00.

Thus $\bar{x} = 87.00/9 = 9.667$ prompts.

Work through the calculation

The squared deviations from $\bar{x}$ sum to 20.000, so $s^2 = 20.000/(9 - 1) = 2.500$ squared prompts, and $s = 1.581$ prompts.

The sample mean estimates the population mean μ; the sample variance estimates the population variance σ^2.

Interpret and check the result

The plug-in standard error is $s/\sqrt{n} = 1.581/\sqrt{9} = 0.527$ prompts.

It estimates the standard deviation of sample means across repeated samples of the same size.

It is not the spread of the individual observations.

`T03-A02-V10`: Catalog accuracy scores
Set up the calculation

There are $n = 8$ observations and their total is 684.00.

Thus $\bar{x} = 684.00/8 = 85.500$ points.

Work through the calculation

The squared deviations from $\bar{x}$ sum to 58.000, so $s^2 = 58.000/(8 - 1) = 8.286$ squared points, and $s = 2.878$ points.

The sample mean estimates the population mean μ; the sample variance estimates the population variance σ^2.

Interpret and check the result

The plug-in standard error is $s/\sqrt{n} = 2.878/\sqrt{8} = 1.018$ points.

It estimates the standard deviation of sample means across repeated samples of the same size.

It is not the spread of the individual observations.

A03: Known-Sigma Confidence Intervals Across Sample Sizes

`T03-A03-V01`: Four samples of reading score
Set up the calculation

The four calculations are for $n = 25$, $SE = 12/\sqrt{25} = 2.400$ and the interval is $69 \pm 1.96(2.400) = [64.296, 73.704]$; for $n = 49$, $SE = 12/\sqrt{49} = 1.714$ and the interval is $74 \pm 1.96(1.714) = [70.640, 77.360]$; for $n = 100$, $SE = 12/\sqrt{100} = 1.200$ and the interval is $71 \pm 1.96(1.200) = [68.648, 73.352]$; for $n = 400$, $SE = 12/\sqrt{400} = 0.600$ and the interval is $73 \pm 1.96(0.600) = [71.824, 74.176]$.

Work through the calculation

Holding $\sigma = 12$ fixed, the standard error falls with $1/\sqrt{n}$, so larger samples produce narrower intervals.

Interpret and check the result

Each $\bar{x}$ is a sample statistic and can vary from sample to sample around the population mean.

A 95% confidence procedure has 95% long-run coverage under its assumptions; it does not require every sample mean to equal 72 points.

`T03-A03-V02`: Four samples of processing time

Set up the calculation

The four calculations are for $n = 16$, $SE = 10/\sqrt{16} = 2.500$ and the interval is $52 \pm 1.96(2.500) = [47.100, 56.900]$; for $n = 36$, $SE = 10/\sqrt{36} = 1.667$ and the interval is $46 \pm 1.96(1.667) = [42.733, 49.267]$; for $n = 64$, $SE = 10/\sqrt{64} = 1.250$ and the interval is $49 \pm 1.96(1.250) = [46.550, 51.450]$; for $n = 144$, $SE = 10/\sqrt{144} = 0.833$ and the interval is $47 \pm 1.96(0.833) = [45.367, 48.633]$.

Work through the calculation

Holding $\sigma = 10$ fixed, the standard error falls with $1/\sqrt{n}$, so larger samples produce narrower intervals.

Interpret and check the result

Each $\bar{x}$ is a sample statistic and can vary from sample to sample around the population mean.

A 95% confidence procedure has 95% long-run coverage under its assumptions; it does not require every sample mean to equal 48 minutes.

`T03-A03-V03`: Four samples of focus rating

Set up the calculation

The four calculations are for $n = 25$, $SE = 9/\sqrt{25} = 1.800$ and the interval is $57 \pm 1.96(1.800) = [53.472, 60.528]$; for $n = 81$, $SE = 9/\sqrt{81} = 1.000$ and the interval is $62 \pm 1.96(1.000) = [60.040, 63.960]$; for $n = 121$, $SE = 9/\sqrt{121} = 0.818$ and the interval is $61 \pm 1.96(0.818) = [59.396, 62.604]$; for $n = 225$, $SE = 9/\sqrt{225} = 0.600$ and the interval is $60 \pm 1.96(0.600) = [58.824, 61.176]$.

Work through the calculation

Holding $\sigma = 9$ fixed, the standard error falls with $1/\sqrt{n}$, so larger samples produce narrower intervals.

Interpret and check the result

Each $\bar{x}$ is a sample statistic and can vary from sample to sample around the population mean.

A 95% confidence procedure has 95% long-run coverage under its assumptions; it does not require every sample mean to equal 60 points.

`T03-A03-V04`: Four samples of visit duration

Set up the calculation

The four calculations are for $n = 36$, $SE = 18/\sqrt{36} = 3.000$ and the interval is $85 \pm 1.96(3.000) = [79.120, 90.880]$; for $n = 64$, $SE = 18/\sqrt{64} = 2.250$ and the interval is $94 \pm 1.96(2.250) = [89.590, 98.410]$; for $n = 100$, $SE = 18/\sqrt{100} = 1.800$ and the interval is $92 \pm 1.96(1.800) = [88.472, 95.528]$; for $n = 324$, $SE = 18/\sqrt{324} = 1.000$ and the interval is $89 \pm 1.96(1.000) = [87.040, 90.960]$.

Work through the calculation

Holding $\sigma = 18$ fixed, the standard error falls with $1/\sqrt{n}$, so larger samples produce narrower intervals.

Interpret and check the result

Each $\bar{x}$ is a sample statistic and can vary from sample to sample around the population mean.

A 95% confidence procedure has 95% long-run coverage under its assumptions; it does not require every sample mean to equal 90 minutes.

`T03-A03-V05`: Four samples of memory score

Set up the calculation

The four calculations are for $n = 25$, $SE = 15/\sqrt{25} = 3.000$ and the interval is $101 \pm 1.96(3.000) = [95.120, 106.880]$; for $n = 49$, $SE = 15/\sqrt{49} = 2.143$ and the interval is $108 \pm 1.96(2.143) = [103.800, 112.200]$; for $n = 100$, $SE = 15/\sqrt{100} = 1.500$ and the interval is $104 \pm 1.96(1.500) = [101.060, 106.940]$; for $n = 225$, $SE = 15/\sqrt{225} = 1.000$ and the interval is $106 \pm 1.96(1.000) = [104.040, 107.960]$.

Work through the calculation

Holding $\sigma = 15$ fixed, the standard error falls with $1/\sqrt{n}$, so larger samples produce narrower intervals.

Interpret and check the result

Each $\bar{x}$ is a sample statistic and can vary from sample to sample around the population mean.

A 95% confidence procedure has 95% long-run coverage under its assumptions; it does not require every sample mean to equal 105 points.

`T03-A03-V06`: Four samples of sound index

Set up the calculation

The four calculations are for $n = 16$, $SE = 8/\sqrt{16} = 2.000$ and the interval is $45 \pm 1.96(2.000) = [41.080, 48.920]$; for $n = 64$, $SE = 8/\sqrt{64} = 1.000$ and the interval is $40 \pm 1.96(1.000) = [38.040, 41.960]$; for $n = 144$, $SE = 8/\sqrt{144} = 0.667$ and the interval is $43 \pm 1.96(0.667) = [41.693, 44.307]$; for $n = 256$, $SE = 8/\sqrt{256} = 0.500$ and the interval is $42 \pm 1.96(0.500) = [41.020, 42.980]$.

Work through the calculation

Holding $\sigma = 8$ fixed, the standard error falls with $1/\sqrt{n}$, so larger samples produce narrower intervals.

Interpret and check the result

Each $\bar{x}$ is a sample statistic and can vary from sample to sample around the population mean.

A 95% confidence procedure has 95% long-run coverage under its assumptions; it does not require every sample mean to equal 42 points.

`T03-A03-V07`: Four samples of confidence score

Set up the calculation

The four calculations are for $n = 25$, $SE = 11/\sqrt{25} = 2.200$ and the interval is $51 \pm 1.96(2.200) = [46.688, 55.312]$; for $n = 100$, $SE = 11/\sqrt{100} = 1.100$ and the interval is $58 \pm 1.96(1.100) = [55.844, 60.156]$; for $n = 121$, $SE = 11/\sqrt{121} = 1.000$ and the interval is $56 \pm 1.96(1.000) = [54.040, 57.960]$; for $n = 400$, $SE = 11/\sqrt{400} = 0.550$ and the interval is $55 \pm 1.96(0.550) = [53.922, 56.078]$.

Work through the calculation

Holding $\sigma = 11$ fixed, the standard error falls with $1/\sqrt{n}$, so larger samples produce narrower intervals.

Interpret and check the result

Each $\bar{x}$ is a sample statistic and can vary from sample to sample around the population mean.

A 95% confidence procedure has 95% long-run coverage under its assumptions; it does not require every sample mean to equal 55 points.

`T03-A03-V08`: Four samples of response time

Set up the calculation

The four calculations are for $n = 16$, $SE = 80/\sqrt{16} = 20.000$ and the interval is $548 \pm 1.96(20.000) = [508.800, 587.200]$; for $n = 64$, $SE = 80/\sqrt{64} = 10.000$ and the interval is $505 \pm 1.96(10.000) = [485.400, 524.600]$; for $n = 100$, $SE = 80/\sqrt{100} = 8.000$ and the interval is $530 \pm 1.96(8.000) =$

$[514.320, 545.680]$; for $n = 256$, $SE = 80/\sqrt{256} = 5.000$ and the interval is $518 \pm 1.96(5.000) = [508.200, 527.800]$.

Work through the calculation

Holding $\sigma = 80$ fixed, the standard error falls with $1/\sqrt{n}$, so larger samples produce narrower intervals.

Interpret and check the result

Each $\bar{x}$ is a sample statistic and can vary from sample to sample around the population mean.

A 95% confidence procedure has 95% long-run coverage under its assumptions; it does not require every sample mean to equal 520 milliseconds.

`T03-A03-V09`: Four samples of trust rating

Set up the calculation

The four calculations are for $n = 25$, $SE = 7/\sqrt{25} = 1.400$ and the interval is $47 \pm 1.96(1.400) = [44.256, 49.744]$; for $n = 49$, $SE = 7/\sqrt{49} = 1.000$ and the interval is $52 \pm 1.96(1.000) = [50.040, 53.960]$; for $n = 196$, $SE = 7/\sqrt{196} = 0.500$ and the interval is $51 \pm 1.96(0.500) = [50.020, 51.980]$; for $n = 400$, $SE = 7/\sqrt{400} = 0.350$ and the interval is $50 \pm 1.96(0.350) = [49.314, 50.686]$.

Work through the calculation

Holding $\sigma = 7$ fixed, the standard error falls with $1/\sqrt{n}$, so larger samples produce narrower intervals.

Interpret and check the result

Each $\bar{x}$ is a sample statistic and can vary from sample to sample around the population mean.

A 95% confidence procedure has 95% long-run coverage under its assumptions; it does not require every sample mean to equal 50 points.

`T03-A03-V10`: Four samples of accuracy score

Set up the calculation

The four calculations are for $n = 36$, $SE = 6/\sqrt{36} = 1.000$ and the interval is $84 \pm 1.96(1.000) = [82.040, 85.960]$; for $n = 81$, $SE = 6/\sqrt{81} = 0.667$ and the interval is $88 \pm 1.96(0.667) = [86.693, 89.307]$; for $n = 144$, $SE = 6/\sqrt{144} = 0.500$ and the interval is $87 \pm 1.96(0.500) = [86.020, 87.980]$; for $n = 324$, $SE = 6/\sqrt{324} = 0.333$ and the interval is $86 \pm 1.96(0.333) = [85.347, 86.653]$.

Work through the calculation

Holding $\sigma = 6$ fixed, the standard error falls with $1/\sqrt{n}$, so larger samples produce narrower intervals.

Interpret and check the result

Each $\bar{x}$ is a sample statistic and can vary from sample to sample around the population mean.

A 95% confidence procedure has 95% long-run coverage under its assumptions; it does not require every sample mean to equal 86 points.

A04: Confidence Level, Sample Size, and Interval Width

`T03-A04-V01`: Interval width for reading fluency

Reason before calculating

The standard error at $n = 36$ is $12/\sqrt{36} = 2.000$ points.

The intervals are 80%: $[75.437, 80.563]$ with margin 2.563; 90%: $[74.710, 81.290]$ with margin 3.290; 95%: $[74.080, 81.920]$ with margin 3.920; 99%: $[72.848, 83.152]$ with margin 5.152.

Work through the calculation

At 95%, $n = 9$ gives width $2(1.96)(12)/\sqrt{9} = 15.680$; $n = 36$ gives width $2(1.96)(12)/\sqrt{36} = 7.840$; $n = 144$ gives width $2(1.96)(12)/\sqrt{144} = 3.920$.

A higher confidence level uses a larger critical value and widens the interval.

Interpret and check the result

A larger sample lowers the standard error and narrows it.

A larger population standard deviation increases the standard error and widens it.

The center remains $\bar{x}$ in every calculation.

`T03-A04-V02`: Interval width for archive processing

Reason before calculating

The standard error at $n = 25$ is $10/\sqrt{25} = 2.000$ minutes.

The intervals are 80%: $[41.437, 46.563]$ with margin 2.563; 90%: $[40.710, 47.290]$ with margin 3.290; 95%: $[40.080, 47.920]$ with margin 3.920; 99%: $[38.848, 49.152]$ with margin 5.152.

Work through the calculation

At 95%, $n = 9$ gives width $2(1.96)(10)/\sqrt{9} = 13.067$; $n = 25$ gives width $2(1.96)(10)/\sqrt{25} = 7.840$; $n = 100$ gives width $2(1.96)(10)/\sqrt{100} = 3.920$.

A higher confidence level uses a larger critical value and widens the interval.

Interpret and check the result

A larger sample lowers the standard error and narrows it.

A larger population standard deviation increases the standard error and widens it.

The center remains $\bar{x}$ in every calculation.

`T03-A04-V03`: Interval width for wellbeing index

Reason before calculating

The standard error at $n = 49$ is $15/\sqrt{49} = 2.143$ points.

The intervals are 80%: $[60.254, 65.746]$ with margin 2.746; 90%: $[59.475, 66.525]$ with margin 3.525; 95%: $[58.800, 67.200]$ with margin 4.200; 99%: $[57.480, 68.520]$ with margin 5.520.

Work through the calculation

At 95%, $n = 16$ gives width $2(1.96)(15)/\sqrt{16} = 14.700$; $n = 49$ gives width $2(1.96)(15)/\sqrt{49} = 8.400$; $n = 196$ gives width $2(1.96)(15)/\sqrt{196} = 4.200$.

A higher confidence level uses a larger critical value and widens the interval.

Interpret and check the result

A larger sample lowers the standard error and narrows it.

A larger population standard deviation increases the standard error and widens it.

The center remains $\bar{x}$ in every calculation.

`T03-A04-V04`: Interval width for museum duration

Reason before calculating

The standard error at $n = 64$ is $18/\sqrt{64} = 2.250$ minutes.

The intervals are 80%: $[93.116, 98.884]$ with margin 2.884; 90%: $[92.299, 99.701]$ with margin 3.701; 95%: $[91.590, 100.410]$ with margin 4.410; 99%: $[90.204, 101.796]$ with margin 5.796.

Work through the calculation

At 95%, $n = 16$ gives width $2(1.96)(18)/\sqrt{16} = 17.640$; $n = 64$ gives width $2(1.96)(18)/\sqrt{64} = 8.820$; $n = 256$ gives width $2(1.96)(18)/\sqrt{256} = 4.410$.

A higher confidence level uses a larger critical value and widens the interval.

Interpret and check the result

A larger sample lowers the standard error and narrows it.

A larger population standard deviation increases the standard error and widens it.

The center remains $\bar{x}$ in every calculation.

`T03-A04-V05`: Interval width for memory score

Reason before calculating

The standard error at $n = 36$ is $14/\sqrt{36} = 2.333$ points.

The intervals are 80%: $[105.010, 110.990]$ with margin 2.990; 90%: $[104.162, 111.838]$ with margin 3.838; 95%: $[103.427, 112.573]$ with margin 4.573; 99%: $[101.990, 114.010]$ with margin 6.010.

Work through the calculation

At 95%, $n = 9$ gives width $2(1.96)(14)/\sqrt{9} = 18.293$; $n = 36$ gives width $2(1.96)(14)/\sqrt{36} = 9.147$; $n = 144$ gives width $2(1.96)(14)/\sqrt{144} = 4.573$.

A higher confidence level uses a larger critical value and widens the interval.

Interpret and check the result

A larger sample lowers the standard error and narrows it.

A larger population standard deviation increases the standard error and widens it.

The center remains $\bar{x}$ in every calculation.

`T03-A04-V06`: Interval width for sound level

Reason before calculating

The standard error at $n = 25$ is $6/\sqrt{25} = 1.200$ decibels.

The intervals are 80%: $[37.462, 40.538]$ with margin 1.538; 90%: $[37.026, 40.974]$ with margin 1.974; 95%: $[36.648, 41.352]$ with margin 2.352; 99%: $[35.909, 42.091]$ with margin 3.091.

Work through the calculation

At 95%, $n = 9$ gives width $2(1.96)(6)/\sqrt{9} = 7.840$; $n = 25$ gives width $2(1.96)(6)/\sqrt{25} = 4.704$; $n = 100$ gives width $2(1.96)(6)/\sqrt{100} = 2.352$.

A higher confidence level uses a larger critical value and widens the interval.

Interpret and check the result

A larger sample lowers the standard error and narrows it.

A larger population standard deviation increases the standard error and widens it.

The center remains $\bar{x}$ in every calculation.

`T03-A04-V07`: Interval width for course confidence

Reason before calculating

The standard error at $n = 64$ is $10/\sqrt{64} = 1.250$ points.

The intervals are 80%: $[56.398, 59.602]$ with margin 1.602; 90%: $[55.944, 60.056]$ with margin 2.056; 95%: $[55.550, 60.450]$ with margin 2.450; 99%: $[54.780, 61.220]$ with margin 3.220.

Work through the calculation

At 95%, $n = 16$ gives width $2(1.96)(10)/\sqrt{16} = 9.800$; $n = 64$ gives width $2(1.96)(10)/\sqrt{64} = 4.900$; $n = 256$ gives width $2(1.96)(10)/\sqrt{256} = 2.450$.

A higher confidence level uses a larger critical value and widens the interval.

Interpret and check the result

A larger sample lowers the standard error and narrows it.

A larger population standard deviation increases the standard error and widens it.

The center remains $\bar{x}$ in every calculation.

`T03-A04-V08`: Interval width for response latency
Reason before calculating

The standard error at $n = 36$ is $72/\sqrt{36} = 12.000$ milliseconds.

The intervals are 80%: $[464.621, 495.379]$ with margin 15.379; 90%: $[460.261, 499.739]$ with margin 19.739; 95%: $[456.480, 503.520]$ with margin 23.520; 99%: $[449.090, 510.910]$ with margin 30.910.

Work through the calculation

At 95%, $n = 9$ gives width $2(1.96)(72)/\sqrt{9} = 94.080$; $n = 36$ gives width $2(1.96)(72)/\sqrt{36} = 47.040$; $n = 144$ gives width $2(1.96)(72)/\sqrt{144} = 23.520$.

A higher confidence level uses a larger critical value and widens the interval.

Interpret and check the result

A larger sample lowers the standard error and narrows it.

A larger population standard deviation increases the standard error and widens it.

The center remains $\bar{x}$ in every calculation.

`T03-A04-V09`: Interval width for community trust
Reason before calculating

The standard error at $n = 49$ is $8/\sqrt{49} = 1.143$ points.

The intervals are 80%: $[50.535, 53.465]$ with margin 1.465; 90%: $[50.120, 53.880]$ with margin 1.880; 95%: $[49.760, 54.240]$ with margin 2.240; 99%: $[49.056, 54.944]$ with margin 2.944.

Work through the calculation

At 95%, $n = 16$ gives width $2(1.96)(8)/\sqrt{16} = 7.840$; $n = 49$ gives width $2(1.96)(8)/\sqrt{49} = 4.480$; $n = 196$ gives width $2(1.96)(8)/\sqrt{196} = 2.240$.

A higher confidence level uses a larger critical value and widens the interval.

Interpret and check the result

A larger sample lowers the standard error and narrows it.

A larger population standard deviation increases the standard error and widens it.

The center remains $\bar{x}$ in every calculation.

`T03-A04-V10`: Interval width for catalog accuracy
Reason before calculating

The standard error at $n = 25$ is $7/\sqrt{25} = 1.400$ points.

The intervals are 80%: $[85.206, 88.794]$ with margin 1.794; 90%: $[84.697, 89.303]$ with margin 2.303; 95%: $[84.256, 89.744]$ with margin 2.744; 99%: $[83.394, 90.606]$ with margin 3.606.

Work through the calculation

At 95%, $n = 9$ gives width $2(1.96)(7)/\sqrt{9} = 9.147$; $n = 25$ gives width $2(1.96)(7)/\sqrt{25} = 5.488$; $n = 100$ gives width $2(1.96)(7)/\sqrt{100} = 2.744$.

A higher confidence level uses a larger critical value and widens the interval.

Interpret and check the result

A larger sample lowers the standard error and narrows it.

A larger population standard deviation increases the standard error and widens it.

The center remains $\bar{x}$ in every calculation.

A05: One-Sided z Tests and the Possible Decision Error

`T03-A05-V01`: Guided route rehearsal
Reason before calculating

The standard error is $SE = 10/\sqrt{25} = 2.0000$ points.

Therefore $z = (72 - 70)/2.0000 = 1.0000$.

Work through the calculation

The rejection boundary is $z > 1.6449$.

Because 1.0000 does not cross that boundary, we fail to reject H_0 at the 5% level.

Interpret and check the result

The evidence is not strong enough to establish the stated higher population route-planning score.

A Type II error remains possible: the procedure could miss a genuine population difference in the stated direction.

A test decision manages long-run error rates; it does not prove which hypothesis is true.

`T03-A05-V02`: Streamlined archive form
Reason before calculating

The standard error is $SE = 8/\sqrt{36} = 1.3333$ minutes.

Therefore $z = (42 - 45)/1.3333 = -2.2500$.

Work through the calculation

The rejection boundary is $z < -1.6449$.

Because −2.2500 crosses that boundary, we reject H_0 at the 5% level.

Interpret and check the result

The evidence is consistent with a lower population completion time.

A Type I error remains possible: the procedure could report the stated directional population difference even though the null value is correct.

A test decision manages long-run error rates; it does not prove which hypothesis is true.

`T03-A05-V03`: Retrieval-practice prompts

Reason before calculating

The standard error is $SE = 12/\sqrt{49} = 1.7143$ points.

Therefore $z = (64 - 62)/1.7143 = 1.1667$.

Work through the calculation

The rejection boundary is $z > 1.6449$.

Because 1.1667 does not cross that boundary, we fail to reject H_0 at the 5% level.

Interpret and check the result

The evidence is not strong enough to establish the stated higher population delayed-recall score.

A Type II error remains possible: the procedure could miss a genuine population difference in the stated direction.

A test decision manages long-run error rates; it does not prove which hypothesis is true.

`T03-A05-V04`: Orientation map

Reason before calculating

The standard error is $SE = 3/\sqrt{25} = 0.6000$ errors.

Therefore $z = (6.8 - 8)/0.6000 = -2.0000$.

Work through the calculation

The rejection boundary is $z < -1.6449$.

Because −2.0000 crosses that boundary, we reject H_0 at the 5% level.

Interpret and check the result

The evidence is consistent with a lower population navigation-error count.

A Type I error remains possible: the procedure could report the stated directional population difference even though the null value is correct.

A test decision manages long-run error rates; it does not prove which hypothesis is true.

`T03-A05-V05`: Structured note template

Reason before calculating

The standard error is $SE = 9/\sqrt{36} = 1.5000$ points.

Therefore $z = (56.5 - 55)/1.5000 = 1.0000$.

Work through the calculation

The rejection boundary is $z > 1.6449$.

Because 1.0000 does not cross that boundary, we fail to reject H_0 at the 5% level.

Interpret and check the result

The evidence is not strong enough to establish the stated higher population argument-identification score.

A Type II error remains possible: the procedure could miss a genuine population difference in the stated direction.

A test decision manages long-run error rates; it does not prove which hypothesis is true.

`T03-A05-V06`: Quiet work period
Reason before calculating

The standard error is $SE = 4/\sqrt{64} = 0.5000$ switches.

Therefore $z = (12.8 - 14)/0.5000 = -2.4000$.

Work through the calculation

The rejection boundary is $z < -1.6449$.

Because −2.4000 crosses that boundary, we reject H_0 at the 5% level.

Interpret and check the result

The evidence is consistent with a lower population task-switch count.

A Type I error remains possible: the procedure could report the stated directional population difference even though the null value is correct.

A test decision manages long-run error rates; it does not prove which hypothesis is true.

`T03-A05-V07`: Captioned demonstration
Reason before calculating

The standard error is $SE = 11/\sqrt{49} = 1.5714$ points.

Therefore $z = (78 - 76)/1.5714 = 1.2727$.

Work through the calculation

The rejection boundary is $z > 1.6449$.

Because 1.2727 does not cross that boundary, we fail to reject H_0 at the 5% level.

Interpret and check the result

The evidence is not strong enough to establish the stated higher population comprehension score.

A Type II error remains possible: the procedure could miss a genuine population difference in the stated direction.

A test decision manages long-run error rates; it does not prove which hypothesis is true.

`T03-A05-V08`: Reminder checklist
Reason before calculating

The standard error is $SE = 2/\sqrt{25} = 0.4000$ steps.

Therefore $z = (4.1 - 5)/0.4000 = -2.2500$.

Work through the calculation

The rejection boundary is $z < -1.6449$.

Because −2.2500 crosses that boundary, we reject H_0 at the 5% level.

Interpret and check the result

The evidence is consistent with a lower population missed-step count.

A Type I error remains possible: the procedure could report the stated directional population difference even though the null value is correct.

A test decision manages long-run error rates; it does not prove which hypothesis is true.

`T03-A05-V09`: Peer explanation

Reason before calculating

The standard error is $SE = 10/\sqrt{100} = 1.0000$ points.

Therefore $z = (69.2 - 68)/1.0000 = 1.2000$.

Work through the calculation

The rejection boundary is $z > 1.6449$.

Because 1.2000 does not cross that boundary, we fail to reject H_0 at the 5% level.

Interpret and check the result

The evidence is not strong enough to establish the stated higher population reasoning score.

A Type II error remains possible: the procedure could miss a genuine population difference in the stated direction.

A test decision manages long-run error rates; it does not prove which hypothesis is true.

`T03-A05-V10`: Reduced notifications

Reason before calculating

The standard error is $SE = 9/\sqrt{81} = 1.0000$ minutes.

Therefore $z = (50.0 - 52)/1.0000 = -2.0000$.

Work through the calculation

The rejection boundary is $z < -1.6449$.

Because −2.0000 crosses that boundary, we reject H_0 at the 5% level.

Interpret and check the result

The evidence is consistent with a lower population completion time.

A Type I error remains possible: the procedure could report the stated directional population difference even though the null value is correct.

A test decision manages long-run error rates; it does not prove which hypothesis is true.

A06: Two-Sided z Tests

`T03-A06-V01`: Memory-support schedule

Set up the calculation

The hypotheses are $H_0 : \mu = 65$ and $H_1 : \mu \neq 65$.

The standard error is $8/\sqrt{64} = 1.0000$ points, so $z = (67.5 - 65)/1.0000 = 2.5000$.

Work through the calculation

The two-sided p-value is $2[1 - \Phi(|2.5000|)] = 0.0124$.

Because 0.0124 is below 0.05, we reject H_0.

Interpret and check the result

The sample provides evidence that the population mean differs from the reference value.

A failure to reject would reflect limited evidence, not proof of exact equality.

`T03-A06-V02`: Revised search protocol
Set up the calculation

The hypotheses are $H_0 : \mu = 50$ and $H_1 : \mu \neq 50$.

The standard error is $10/\sqrt{100} = 1.0000$ seconds, so $z = (48 - 50)/1.0000 = -2.0000$.

Work through the calculation

The two-sided p-value is $2[1 - \Phi(|-2.0000|)] = 0.0455$.

Because 0.0455 is below 0.05, we reject H_0.

Interpret and check the result

The sample provides evidence that the population mean differs from the reference value.

A failure to reject would reflect limited evidence, not proof of exact equality.

`T03-A06-V03`: New reading prompt
Set up the calculation

The hypotheses are $H_0 : \mu = 72$ and $H_1 : \mu \neq 72$.

The standard error is $12/\sqrt{36} = 2.0000$ points, so $z = (74 - 72)/2.0000 = 1.0000$.

Work through the calculation

The two-sided p-value is $2[1 - \Phi(|1.0000|)] = 0.3173$.

Because 0.3173 is not below 0.05, we fail to reject H_0.

Interpret and check the result

The sample does not provide sufficiently strong evidence of a population-mean difference at this significance level.

A failure to reject would reflect limited evidence, not proof of exact equality.

`T03-A06-V04`: Gallery route signs
Set up the calculation

The hypotheses are $H_0 : \mu = 40$ and $H_1 : \mu \neq 40$.

The standard error is $9/\sqrt{81} = 1.0000$ minutes, so $z = (37.5 - 40)/1.0000 = -2.5000$.

Work through the calculation

The two-sided p-value is $2[1 - \Phi(|-2.5000|)] = 0.0124$.

Because 0.0124 is below 0.05, we reject H_0.

Interpret and check the result

The sample provides evidence that the population mean differs from the reference value.

A failure to reject would reflect limited evidence, not proof of exact equality.

`T03-A06-V05`: Concept-map format
Set up the calculation

The hypotheses are $H_0 : \mu = 60$ and $H_1 : \mu \neq 60$.

The standard error is $10/\sqrt{49} = 1.4286$ points, so $z = (61.5 - 60)/1.4286 = 1.0500$.

Work through the calculation

The two-sided p-value is $2[1 - \Phi(|1.0500|)] = 0.2937$.

Because 0.2937 is not below 0.05, we fail to reject H_0.

Interpret and check the result

The sample does not provide sufficiently strong evidence of a population-mean difference at this significance level.

A failure to reject would reflect limited evidence, not proof of exact equality.

`T03-A06-V06`: Appointment reminder

Set up the calculation

The hypotheses are $H_0 : \mu = 24$ and $H_1 : \mu \neq 24$.

The standard error is $6/\sqrt{64} = 0.7500$ hours, so $z = (22 - 24)/0.7500 = -2.6667$.

Work through the calculation

The two-sided p-value is $2[1 - \Phi(|-2.6667|)] = 0.0077$.

Because 0.0077 is below 0.05, we reject H_0.

Interpret and check the result

The sample provides evidence that the population mean differs from the reference value.

A failure to reject would reflect limited evidence, not proof of exact equality.

`T03-A06-V07`: Peer-review structure

Set up the calculation

The hypotheses are $H_0 : \mu = 55$ and $H_1 : \mu \neq 55$.

The standard error is $8/\sqrt{100} = 0.8000$ points, so $z = (56 - 55)/0.8000 = 1.2500$.

Work through the calculation

The two-sided p-value is $2[1 - \Phi(|1.2500|)] = 0.2113$.

Because 0.2113 is not below 0.05, we fail to reject H_0.

Interpret and check the result

The sample does not provide sufficiently strong evidence of a population-mean difference at this significance level.

A failure to reject would reflect limited evidence, not proof of exact equality.

`T03-A06-V08`: Document preview tool

Set up the calculation

The hypotheses are $H_0 : \mu = 35$ and $H_1 : \mu \neq 35$.

The standard error is $7/\sqrt{49} = 1.0000$ seconds, so $z = (32.5 - 35)/1.0000 = -2.5000$.

Work through the calculation

The two-sided p-value is $2[1 - \Phi(|-2.5000|)] = 0.0124$.

Because 0.0124 is below 0.05, we reject H_0.

Interpret and check the result

The sample provides evidence that the population mean differs from the reference value.

A failure to reject would reflect limited evidence, not proof of exact equality.

`T03-A06-V09`: Route rehearsal card
Set up the calculation

The hypotheses are $H_0 : \mu = 70$ and $H_1 : \mu \neq 70$.

The standard error is $9/\sqrt{144} = 0.7500$ points, so $z = (71 - 70)/0.7500 = 1.3333$.

Work through the calculation

The two-sided p-value is $2[1 - \Phi(|1.3333|)] = 0.1824$.

Because 0.1824 is not below 0.05, we fail to reject H_0.

Interpret and check the result

The sample does not provide sufficiently strong evidence of a population-mean difference at this significance level.

A failure to reject would reflect limited evidence, not proof of exact equality.

`T03-A06-V10`: Annotation display
Set up the calculation

The hypotheses are $H_0 : \mu = 48$ and $H_1 : \mu \neq 48$.

The standard error is $6/\sqrt{100} = 0.6000$ points, so $z = (46.5 - 48)/0.6000 = -2.5000$.

Work through the calculation

The two-sided p-value is $2[1 - \Phi(|-2.5000|)] = 0.0124$.

Because 0.0124 is below 0.05, we reject H_0.

Interpret and check the result

The sample provides evidence that the population mean differs from the reference value.

A failure to reject would reflect limited evidence, not proof of exact equality.

A07: p-Values, Power, and Planned Sample Size

`T03-A07-V01`: Planning a study of guided map practice
Reason before calculating

The observed statistic is $z = (71.5 - 70)/\left(10/\sqrt{36}\right) = 0.9000$.

Its upper-tail p-value is $1 - \Phi(0.9000) = 0.1841$, so we fail to reject H_0 at $\alpha = 0.05$.

Work through the calculation

This p-value describes results at least this large under the null model; it is not the probability that H_0 is true.

If the true improvement is $\delta = 4$, power is $1 - \Phi\left(1.6449 - 4\sqrt{36}/10\right) = 0.7749$.

Interpret and check the result

Power is the long-run probability of rejection under that specified alternative.

For 90% planned power, $n = \lceil [(1.6449 + 1.2816)10/4]^2 \rceil = 54$.

Rounding up is necessary because a fraction of an observation cannot meet the target.

`T03-A07-V02`: Planning a study of an archive-search checklist
Reason before calculating

The observed statistic is $z = (20 - 18)/\left(5/\sqrt{49}\right) = 2.8000$.

Its upper-tail p-value is $1 - \Phi(2.8000) = 0.0026$, so we reject H_0 at $\alpha = 0.05$.

Work through the calculation

This p-value describes results at least this large under the null model; it is not the probability that H_0 is true.

If the true improvement is $\delta = 2.5$, power is $1 - \Phi\left(1.6449 - 2.5\sqrt{49}/5\right) = 0.9682$.

Interpret and check the result

Power is the long-run probability of rejection under that specified alternative.

For 90% planned power, $n = \lceil[(1.6449 + 1.2816)5/2.5]^2\rceil = 35$.

Rounding up is necessary because a fraction of an observation cannot meet the target.

`T03-A07-V03`: Planning a study of retrieval-practice cards

Reason before calculating

The observed statistic is $z = (61.5 - 60)/\left(12/\sqrt{64}\right) = 1.0000$.

Its upper-tail p-value is $1 - \Phi(1.0000) = 0.1587$, so we fail to reject H_0 at $\alpha = 0.05$.

Work through the calculation

This p-value describes results at least this large under the null model; it is not the probability that H_0 is true.

If the true improvement is $\delta = 4$, power is $1 - \Phi\left(1.6449 - 4\sqrt{64}/12\right) = 0.8466$.

Interpret and check the result

Power is the long-run probability of rejection under that specified alternative.

For 90% planned power, $n = \lceil[(1.6449 + 1.2816)12/4]^2\rceil = 78$.

Rounding up is necessary because a fraction of an observation cannot meet the target.

`T03-A07-V04`: Planning a study of a quiet reading room

Reason before calculating

The observed statistic is $z = (78.5 - 75)/\left(14/\sqrt{49}\right) = 1.7500$.

Its upper-tail p-value is $1 - \Phi(1.7500) = 0.0401$, so we reject H_0 at $\alpha = 0.05$.

Work through the calculation

This p-value describes results at least this large under the null model; it is not the probability that H_0 is true.

If the true improvement is $\delta = 5$, power is $1 - \Phi\left(1.6449 - 5\sqrt{49}/14\right) = 0.8038$.

Interpret and check the result

Power is the long-run probability of rejection under that specified alternative.

For 90% planned power, $n = \lceil[(1.6449 + 1.2816)14/5]^2\rceil = 68$.

Rounding up is necessary because a fraction of an observation cannot meet the target.

`T03-A07-V05`: Planning a study of a structured note template

Reason before calculating

The observed statistic is $z = (12.5 - 12)/\left(3/\sqrt{36}\right) = 1.0000$.

Its upper-tail p-value is $1 - \Phi(1.0000) = 0.1587$, so we fail to reject H_0 at $\alpha = 0.05$.

Work through the calculation

This p-value describes results at least this large under the null model; it is not the probability that H_0 is true.

If the true improvement is $\delta = 1.5$, power is $1 - \Phi\left(1.6449 - 1.5\sqrt{36}/3\right) = 0.9123$.

Interpret and check the result

Power is the long-run probability of rejection under that specified alternative.

For 90% planned power, $n = \lceil[(1.6449 + 1.2816)3/1.5]^2\rceil = 35$.

Rounding up is necessary because a fraction of an observation cannot meet the target.

`T03-A07-V06`: Planning a study of a captioned tutorial

Reason before calculating

The observed statistic is $z = (70 - 68)/\left(9/\sqrt{81}\right) = 2.0000$.

Its upper-tail p-value is $1 - \Phi(2.0000) = 0.0228$, so we reject H_0 at $\alpha = 0.05$.

Work through the calculation

This p-value describes results at least this large under the null model; it is not the probability that H_0 is true.

If the true improvement is $\delta = 2.5$, power is $1 - \Phi\left(1.6449 - 2.5\sqrt{81}/9\right) = 0.8038$.

Interpret and check the result

Power is the long-run probability of rejection under that specified alternative.

For 90% planned power, $n = \lceil[(1.6449 + 1.2816)9/2.5]^2\rceil = 111$.

Rounding up is necessary because a fraction of an observation cannot meet the target.

`T03-A07-V07`: Planning a study of a reminder checklist

Reason before calculating

The observed statistic is $z = (20.6 - 20)/\left(4/\sqrt{49}\right) = 1.0500$.

Its upper-tail p-value is $1 - \Phi(1.0500) = 0.1469$, so we fail to reject H_0 at $\alpha = 0.05$.

Work through the calculation

This p-value describes results at least this large under the null model; it is not the probability that H_0 is true.

If the true improvement is $\delta = 2$, power is $1 - \Phi\left(1.6449 - 2\sqrt{49}/4\right) = 0.9682$.

Interpret and check the result

Power is the long-run probability of rejection under that specified alternative.

For 90% planned power, $n = \lceil[(1.6449 + 1.2816)4/2]^2\rceil = 35$.

Rounding up is necessary because a fraction of an observation cannot meet the target.

`T03-A07-V08`: Planning a study of peer explanation

Reason before calculating

The observed statistic is $z = (74.1 - 72)/\left(11/\sqrt{100}\right) = 1.9091$.

Its upper-tail p-value is $1 - \Phi(1.9091) = 0.0281$, so we reject H_0 at $\alpha = 0.05$.

Work through the calculation

This p-value describes results at least this large under the null model; it is not the probability that H_0 is true.

If the true improvement is $\delta = 3$, power is $1 - \Phi\left(1.6449 - 3\sqrt{100}/11\right) = 0.8605$.

Interpret and check the result

Power is the long-run probability of rejection under that specified alternative.

For 90% planned power, $n = \lceil [(1.6449 + 1.2816)11/3]^2 \rceil = 116$.

Rounding up is necessary because a fraction of an observation cannot meet the target.

`T03-A07-V09`: Planning a study of a museum orientation map

Reason before calculating

The observed statistic is $z = (51 - 50)/\left(8/\sqrt{64}\right) = 1.0000$.

Its upper-tail p-value is $1 - \Phi(1.0000) = 0.1587$, so we fail to reject H_0 at $\alpha = 0.05$.

Work through the calculation

This p-value describes results at least this large under the null model; it is not the probability that H_0 is true.

If the true improvement is $\delta = 2.5$, power is $1 - \Phi\left(1.6449 - 2.5\sqrt{64}/8\right) = 0.8038$.

Interpret and check the result

Power is the long-run probability of rejection under that specified alternative.

For 90% planned power, $n = \lceil [(1.6449 + 1.2816)8/2.5]^2 \rceil = 88$.

Rounding up is necessary because a fraction of an observation cannot meet the target.

`T03-A07-V10`: Planning a study of a reduced-notification setting

Reason before calculating

The observed statistic is $z = (57.0 - 55)/\left(10/\sqrt{100}\right) = 2.0000$.

Its upper-tail p-value is $1 - \Phi(2.0000) = 0.0228$, so we reject H_0 at $\alpha = 0.05$.

Work through the calculation

This p-value describes results at least this large under the null model; it is not the probability that H_0 is true.

If the true improvement is $\delta = 3$, power is $1 - \Phi\left(1.6449 - 3\sqrt{100}/10\right) = 0.9123$.

Interpret and check the result

Power is the long-run probability of rejection under that specified alternative.

For 90% planned power, $n = \lceil [(1.6449 + 1.2816)10/3]^2 \rceil = 96$.

Rounding up is necessary because a fraction of an observation cannot meet the target.

A08: Two-Sided Testing, Confidence Intervals, and Practical Size

`T03-A08-V01`: Full inference for alternative reading layout

Reason before calculating

The standard error is $30/\sqrt{64} = 3.7500$ words per minute, and the margin is $1.96(3.7500) = 7.3500$.

The 95% interval is $215 \pm 7.3500 = [207.6500, 222.3500]$.

Work through the calculation

The null value 210 is inside this interval, so the matching two-sided test fails to reject H_0.

Directly, $z = 1.3333$ and $p = 0.1824$, which gives the same decision.

Interpret and check the result

The observed difference is $215 - 210 = 5.000$ words per minute.

Its practical importance depends on subject-matter consequences and the measurement scale, not on the p-value alone.

`T03-A08-V02`: Full inference for new catalog interface

Reason before calculating

The standard error is $12/\sqrt{49} = 1.7143$ seconds, and the margin is $1.96(1.7143) = 3.3600$.

The 95% interval is $70 \pm 3.3600 = [66.6400, 73.3600]$.

Work through the calculation

The null value 75 is outside this interval, so the matching two-sided test rejects H_0.

Directly, $z = -2.9167$ and $p = 0.0035$, which gives the same decision.

Interpret and check the result

The observed difference is $70 - 75 = -5.000$ seconds.

Its practical importance depends on subject-matter consequences and the measurement scale, not on the p-value alone.

`T03-A08-V03`: Full inference for ambient sound setting

Reason before calculating

The standard error is $10/\sqrt{36} = 1.6667$ points, and the margin is $1.96(1.6667) = 3.2667$.

The 95% interval is $60 \pm 3.2667 = [56.7333, 63.2667]$.

Work through the calculation

The null value 58 is inside this interval, so the matching two-sided test fails to reject H_0.

Directly, $z = 1.2000$ and $p = 0.2301$, which gives the same decision.

Interpret and check the result

The observed difference is $60 - 58 = 2.000$ points.

Its practical importance depends on subject-matter consequences and the measurement scale, not on the p-value alone.

`T03-A08-V04`: Full inference for museum route markers

Reason before calculating

The standard error is $18/\sqrt{81} = 2.0000$ minutes, and the margin is $1.96(2.0000) = 3.9200$.

The 95% interval is $89 \pm 3.9200 = [85.0800, 92.9200]$.

Work through the calculation

The null value 95 is outside this interval, so the matching two-sided test rejects H_0.

Directly, $z = -3.0000$ and $p = 0.0027$, which gives the same decision.

Interpret and check the result

The observed difference is $89 - 95 = -6.000$ minutes.

Its practical importance depends on subject-matter consequences and the measurement scale, not on the p-value alone.

`T03-A08-V05`: Full inference for memory cue format
Reason before calculating

The standard error is $14/\sqrt{49} = 2.0000$ points, and the margin is $1.96(2.0000) = 3.9200$.

The 95% interval is $75 \pm 3.9200 = [71.0800, 78.9200]$.

Work through the calculation

The null value 72 is inside this interval, so the matching two-sided test fails to reject H_0.

Directly, $z = 1.5000$ and $p = 0.1336$, which gives the same decision.

Interpret and check the result

The observed difference is $75 - 72 = 3.000$ points.

Its practical importance depends on subject-matter consequences and the measurement scale, not on the p-value alone.

`T03-A08-V06`: Full inference for survey reminder wording
Reason before calculating

The standard error is $8/\sqrt{64} = 1.0000$ hours, and the margin is $1.96(1.0000) = 1.9600$.

The 95% interval is $27 \pm 1.9600 = [25.0400, 28.9600]$.

Work through the calculation

The null value 30 is outside this interval, so the matching two-sided test rejects H_0.

Directly, $z = -3.0000$ and $p = 0.0027$, which gives the same decision.

Interpret and check the result

The observed difference is $27 - 30 = -3.000$ hours.

Its practical importance depends on subject-matter consequences and the measurement scale, not on the p-value alone.

`T03-A08-V07`: Full inference for workshop seating plan
Reason before calculating

The standard error is $9/\sqrt{100} = 0.9000$ points, and the margin is $1.96(0.9000) = 1.7640$.

The 95% interval is $51.2 \pm 1.7640 = [49.4360, 52.9640]$.

Work through the calculation

The null value 50 is inside this interval, so the matching two-sided test fails to reject H_0.

Directly, $z = 1.3333$ and $p = 0.1824$, which gives the same decision.

Interpret and check the result

The observed difference is $51.2 - 50 = 1.200$ points.

Its practical importance depends on subject-matter consequences and the measurement scale, not on the p-value alone.

`T03-A08-V08`: Full inference for archive image contrast
Reason before calculating

The standard error is $7/\sqrt{49} = 1.0000$ seconds, and the margin is $1.96(1.0000) = 1.9600$.

The 95% interval is $39.5 \pm 1.9600 = [37.5400, 41.4600]$.

Work through the calculation

The null value 42 is outside this interval, so the matching two-sided test rejects H_0.

Directly, $z = -2.5000$ and $p = 0.0124$, which gives the same decision.

Interpret and check the result

The observed difference is $39.5 - 42 = -2.500$ seconds.

Its practical importance depends on subject-matter consequences and the measurement scale, not on the p-value alone.

`T03-A08-V09`: Full inference for route-description style

Reason before calculating

The standard error is $12/\sqrt{144} = 1.0000$ points, and the margin is $1.96(1.0000) = 1.9600$.

The 95% interval is $67.5 \pm 1.9600 = [65.5400, 69.4600]$.

Work through the calculation

The null value 66 is inside this interval, so the matching two-sided test fails to reject H_0.

Directly, $z = 1.5000$ and $p = 0.1336$, which gives the same decision.

Interpret and check the result

The observed difference is $67.5 - 66 = 1.500$ points.

Its practical importance depends on subject-matter consequences and the measurement scale, not on the p-value alone.

`T03-A08-V10`: Full inference for note-taking display

Reason before calculating

The standard error is $15/\sqrt{100} = 1.5000$ points, and the margin is $1.96(1.5000) = 2.9400$.

The 95% interval is $76.5 \pm 2.9400 = [73.5600, 79.4400]$.

Work through the calculation

The null value 80 is outside this interval, so the matching two-sided test rejects H_0.

Directly, $z = -2.3333$ and $p = 0.0196$, which gives the same decision.

Interpret and check the result

The observed difference is $76.5 - 80 = -3.500$ points.

Its practical importance depends on subject-matter consequences and the measurement scale, not on the p-value alone.

A10: One-Sample t Tests

`T03-A10-V01`: Testing guided search training

Set up the calculation

The population standard deviation is unknown and is replaced by $s = 6.5$, so the source's one-sample procedure uses $T = (\bar{X} - \mu_0)/(S/\sqrt{n})$ with $df = n - 1 = 7$.

Work through the calculation

Here $SE = 6.5/\sqrt{8} = 2.2981$ points and $t = (76 - 70)/2.2981 = 2.6109$.

The rejection rule is $t > 1.8946$.

Interpret and check the result

The statistic crosses that boundary, so we reject H_0.

The sample supports the stated directional change in search-accuracy score, subject to independent observations and an approximately normal population model for this small-sample procedure.

`T03-A10-V02`: Testing a shorter intake form
Set up the calculation

The population standard deviation is unknown and is replaced by $s = 5.4$, so the source's one-sample procedure uses $T = \left(\bar{X} - \mu_0\right)/\left(S/\sqrt{n}\right)$ with $df = n - 1 = 8$.

Work through the calculation

Here $SE = 5.4/\sqrt{9} = 1.8000$ minutes and $t = (41.5 - 45)/1.8000 = -1.9444$.

The rejection rule is $t < -1.8595$.

Interpret and check the result

The statistic crosses that boundary, so we reject H_0.

The sample supports the stated directional change in completion time, subject to independent observations and an approximately normal population model for this small-sample procedure.

`T03-A10-V03`: Testing a retrieval cue
Set up the calculation

The population standard deviation is unknown and is replaced by $s = 6.2$, so the source's one-sample procedure uses $T = \left(\bar{X} - \mu_0\right)/\left(S/\sqrt{n}\right)$ with $df = n - 1 = 9$.

Work through the calculation

Here $SE = 6.2/\sqrt{10} = 1.9606$ points and $t = (66 - 62)/1.9606 = 2.0402$.

The rejection rule is $t > 1.8331$.

Interpret and check the result

The statistic crosses that boundary, so we reject H_0.

The sample supports the stated directional change in recall score, subject to independent observations and an approximately normal population model for this small-sample procedure.

`T03-A10-V04`: Testing an orientation card
Set up the calculation

The population standard deviation is unknown and is replaced by $s = 3.8$, so the source's one-sample procedure uses $T = \left(\bar{X} - \mu_0\right)/\left(S/\sqrt{n}\right)$ with $df = n - 1 = 10$.

Work through the calculation

Here $SE = 3.8/\sqrt{11} = 1.1457$ errors and $t = (7.1 - 9)/1.1457 = -1.6583$.

The rejection rule is $t < -1.8125$.

Interpret and check the result

The statistic does not cross that boundary, so we fail to reject H_0.

The sample does not provide sufficiently strong evidence for the stated directional change in navigation-error count, subject to independent observations and an approximately normal population model for this small-sample procedure.

`T03-A10-V05`: Testing a structured note page

Set up the calculation

The population standard deviation is unknown and is replaced by $s = 5.5$, so the source's one-sample procedure uses $T = (\bar{X} - \mu_0)/(S/\sqrt{n})$ with $df = n - 1 = 11$.

Work through the calculation

Here $SE = 5.5/\sqrt{12} = 1.5877$ points and $t = (57.5 - 54)/1.5877 = 2.2044$.

The rejection rule is $t > 1.7959$.

Interpret and check the result

The statistic crosses that boundary, so we reject H_0.

The sample supports the stated directional change in argument score, subject to independent observations and an approximately normal population model for this small-sample procedure.

`T03-A10-V06`: Testing a quiet work block

Set up the calculation

The population standard deviation is unknown and is replaced by $s = 4.0$, so the source's one-sample procedure uses $T = (\bar{X} - \mu_0)/(S/\sqrt{n})$ with $df = n - 1 = 12$.

Work through the calculation

Here $SE = 4.0/\sqrt{13} = 1.1094$ switches and $t = (13.0 - 15)/1.1094 = -1.8028$.

The rejection rule is $t < -1.7823$.

Interpret and check the result

The statistic crosses that boundary, so we reject H_0.

The sample supports the stated directional change in task-switch count, subject to independent observations and an approximately normal population model for this small-sample procedure.

`T03-A10-V07`: Testing a captioned demonstration

Set up the calculation

The population standard deviation is unknown and is replaced by $s = 6.0$, so the source's one-sample procedure uses $T = (\bar{X} - \mu_0)/(S/\sqrt{n})$ with $df = n - 1 = 13$.

Work through the calculation

Here $SE = 6.0/\sqrt{14} = 1.6036$ points and $t = (77.2 - 74)/1.6036 = 1.9956$.

The rejection rule is $t > 1.7709$.

Interpret and check the result

The statistic crosses that boundary, so we reject H_0.

The sample supports the stated directional change in comprehension score, subject to independent observations and an approximately normal population model for this small-sample procedure.

`T03-A10-V08`: Testing a reminder checklist

Set up the calculation

The population standard deviation is unknown and is replaced by $s = 2.1$, so the source's one-sample procedure uses $T = (\bar{X} - \mu_0)/(S/\sqrt{n})$ with $df = n - 1 = 14$.

Work through the calculation

Here $SE = 2.1/\sqrt{15} = 0.5422$ steps and $t = (4.9 - 6)/0.5422 = -2.0287$.

The rejection rule is $t < -1.7613$.

Interpret and check the result

The statistic crosses that boundary, so we reject H_0.

The sample supports the stated directional change in missed-step count, subject to independent observations and an approximately normal population model for this small-sample procedure.

`T03-A10-V09`: Testing peer explanation

Set up the calculation

The population standard deviation is unknown and is replaced by $s = 5.3$, so the source's one-sample procedure uses $T = \left(\bar{X} - \mu_0\right)/(S/\sqrt{n})$ with $df = n - 1 = 15$.

Work through the calculation

Here $SE = 5.3/\sqrt{16} = 1.3250$ points and $t = (69.6 - 67)/1.3250 = 1.9623$.

The rejection rule is $t > 1.7531$.

Interpret and check the result

The statistic crosses that boundary, so we reject H_0.

The sample supports the stated directional change in reasoning score, subject to independent observations and an approximately normal population model for this small-sample procedure.

`T03-A10-V10`: Testing a reduced-notification setting

Set up the calculation

The population standard deviation is unknown and is replaced by $s = 4.9$, so the source's one-sample procedure uses $T = \left(\bar{X} - \mu_0\right)/(S/\sqrt{n})$ with $df = n - 1 = 16$.

Work through the calculation

Here $SE = 4.9/\sqrt{17} = 1.1884$ minutes and $t = (47.8 - 50)/1.1884 = -1.8512$.

The rejection rule is $t < -1.7459$.

Interpret and check the result

The statistic crosses that boundary, so we reject H_0.

The sample supports the stated directional change in completion time, subject to independent observations and an approximately normal population model for this small-sample procedure.

A11: One-Sample t Confidence Intervals and Test Duality

`T03-A11-V01`: Estimating mean reading fluency

Reason before calculating

The estimated standard error is $s/\sqrt{n} = 7/\sqrt{8} = 2.4749$ points.

Work through the calculation

The margin is $2.3646(2.4749) = 5.8521$, so the interval is $76 \pm 5.8521 = [70.1479, 81.8521]$.

The null value 70 falls outside the interval, so the matching two-sided 5% test rejects H_0.

Interpret and check the result

Both procedures use the same standard error and symmetric 2.5% tail boundaries, which makes their decisions agree.

In repeated sampling, 95% of intervals produced by this method cover the fixed population mean under the model assumptions; the completed interval does not assign a 95% probability to that fixed parameter.

`T03-A11-V02`: Estimating mean archive processing

Reason before calculating

The estimated standard error is $s/\sqrt{n} = 6/\sqrt{9} = 2.0000$ minutes.

Work through the calculation

The margin is $2.3060(2.0000) = 4.6120$, so the interval is $43 \pm 4.6120 = [38.3880, 47.6120]$.

The null value 47 falls inside the interval, so the matching two-sided 5% test fails to reject H_0.

Interpret and check the result

Both procedures use the same standard error and symmetric 2.5% tail boundaries, which makes their decisions agree.

In repeated sampling, 95% of intervals produced by this method cover the fixed population mean under the model assumptions; the completed interval does not assign a 95% probability to that fixed parameter.

`T03-A11-V03`: Estimating mean focus rating

Reason before calculating

The estimated standard error is $s/\sqrt{n} = 8/\sqrt{10} = 2.5298$ points.

Work through the calculation

The margin is $2.2622(2.5298) = 5.7230$, so the interval is $61 \pm 5.7230 = [55.2770, 66.7230]$.

The null value 55 falls outside the interval, so the matching two-sided 5% test rejects H_0.

Interpret and check the result

Both procedures use the same standard error and symmetric 2.5% tail boundaries, which makes their decisions agree.

In repeated sampling, 95% of intervals produced by this method cover the fixed population mean under the model assumptions; the completed interval does not assign a 95% probability to that fixed parameter.

`T03-A11-V04`: Estimating mean museum duration

Reason before calculating

The estimated standard error is $s/\sqrt{n} = 15/\sqrt{11} = 4.5227$ minutes.

Work through the calculation

The margin is $2.2281(4.5227) = 10.0770$, so the interval is $92 \pm 10.0770 = [81.9230, 102.0770]$.

The null value 100 falls inside the interval, so the matching two-sided 5% test fails to reject H_0.

Interpret and check the result

Both procedures use the same standard error and symmetric 2.5% tail boundaries, which makes their decisions agree.

In repeated sampling, 95% of intervals produced by this method cover the fixed population mean under the model assumptions; the completed interval does not assign a 95% probability to that fixed parameter.

`T03-A11-V05`: Estimating mean memory score
Reason before calculating

The estimated standard error is $s/\sqrt{n} = 10/\sqrt{12} = 2.8868$ points.

Work through the calculation

The margin is $2.2010(2.8868) = 6.3537$, so the interval is $106 \pm 6.3537 = [99.6463, 112.3537]$.

The null value 100 falls inside the interval, so the matching two-sided 5% test fails to reject H_0.

Interpret and check the result

Both procedures use the same standard error and symmetric 2.5% tail boundaries, which makes their decisions agree.

In repeated sampling, 95% of intervals produced by this method cover the fixed population mean under the model assumptions; the completed interval does not assign a 95% probability to that fixed parameter.

`T03-A11-V06`: Estimating mean sound index
Reason before calculating

The estimated standard error is $s/\sqrt{n} = 5/\sqrt{13} = 1.3868$ points.

Work through the calculation

The margin is $2.1788(1.3868) = 3.0215$, so the interval is $40 \pm 3.0215 = [36.9785, 43.0215]$.

The null value 43 falls inside the interval, so the matching two-sided 5% test fails to reject H_0.

Interpret and check the result

Both procedures use the same standard error and symmetric 2.5% tail boundaries, which makes their decisions agree.

In repeated sampling, 95% of intervals produced by this method cover the fixed population mean under the model assumptions; the completed interval does not assign a 95% probability to that fixed parameter.

`T03-A11-V07`: Estimating mean confidence score
Reason before calculating

The estimated standard error is $s/\sqrt{n} = 7/\sqrt{14} = 1.8708$ points.

Work through the calculation

The margin is $2.1604(1.8708) = 4.0417$, so the interval is $57 \pm 4.0417 = [52.9583, 61.0417]$.

The null value 52 falls outside the interval, so the matching two-sided 5% test rejects H_0.

Interpret and check the result

Both procedures use the same standard error and symmetric 2.5% tail boundaries, which makes their decisions agree.

In repeated sampling, 95% of intervals produced by this method cover the fixed population mean under the model assumptions; the completed interval does not assign a 95% probability to that fixed parameter.

`T03-A11-V08`: Estimating mean response latency
Reason before calculating

The estimated standard error is $s/\sqrt{n} = 54/\sqrt{15} = 13.9427$ milliseconds.

Work through the calculation

The margin is $2.1448(13.9427) = 29.9044$, so the interval is $490 \pm 29.9044 = [460.0956, 519.9044]$.

The null value 520 falls outside the interval, so the matching two-sided 5% test rejects H_0.

Interpret and check the result

Both procedures use the same standard error and symmetric 2.5% tail boundaries, which makes their decisions agree.

In repeated sampling, 95% of intervals produced by this method cover the fixed population mean under the model assumptions; the completed interval does not assign a 95% probability to that fixed parameter.

`T03-A11-V09`: Estimating mean trust rating

Reason before calculating

The estimated standard error is $s/\sqrt{n} = 6/\sqrt{16} = 1.5000$ points.

Work through the calculation

The margin is $2.1314(1.5000) = 3.1971$, so the interval is $51 \pm 3.1971 = [47.8029, 54.1971]$.

The null value 48 falls inside the interval, so the matching two-sided 5% test fails to reject H_0.

Interpret and check the result

Both procedures use the same standard error and symmetric 2.5% tail boundaries, which makes their decisions agree.

In repeated sampling, 95% of intervals produced by this method cover the fixed population mean under the model assumptions; the completed interval does not assign a 95% probability to that fixed parameter.

`T03-A11-V10`: Estimating mean catalog accuracy

Reason before calculating

The estimated standard error is $s/\sqrt{n} = 5/\sqrt{17} = 1.2127$ points.

Work through the calculation

The margin is $2.1199(1.2127) = 2.5708$, so the interval is $86 \pm 2.5708 = [83.4292, 88.5708]$.

The null value 83 falls outside the interval, so the matching two-sided 5% test rejects H_0.

Interpret and check the result

Both procedures use the same standard error and symmetric 2.5% tail boundaries, which makes their decisions agree.

In repeated sampling, 95% of intervals produced by this method cover the fixed population mean under the model assumptions; the completed interval does not assign a 95% probability to that fixed parameter.

A12: Pooled Independent-Samples t Tests

`T03-A12-V01`: Comparing guided versus self-directed map practice

Reason before calculating

Different observational units belong to the two groups, so no case-level pairing is available.

Under the equal-population-variance model, $s_p^2 = [(10-1)36 + (10-1)25]/(10+10-2) = 30.5000$.

Work through the calculation

The standard error is $\sqrt{30.5000(1/10 + 1/10)} = 2.4698$.

With $df = 18$, $t = (74 - 68)/2.4698 = 2.4293$.

Because 2.4293 exceeds 1.7341, we reject the equal-mean null in favor of $\mu_1 > \mu_2$.

Interpret and check the result

The result supports a higher population mean in Group 1.

The procedure requires independent observations, independently sampled groups, an approximately normal outcome within each population for these sample sizes, and equal population variances.

Random assignment is additionally needed for a causal interpretation.

`T03-A12-V02`: Comparing captioned versus uncaptioned tutorials

Reason before calculating

Different observational units belong to the two groups, so no case-level pairing is available.

Under the equal-population-variance model, $s_p^2 = [(11-1)30 + (11-1)28]/(11+11-2) = 29.0000$.

Work through the calculation

The standard error is $\sqrt{29.0000(1/11 + 1/11)} = 2.2962$.

With $df = 20$, $t = (79 - 74)/2.2962 = 2.1775$.

Because 2.1775 exceeds 1.7247, we reject the equal-mean null in favor of $\mu_1 > \mu_2$.

Interpret and check the result

The result supports a higher population mean in Group 1.

The procedure requires independent observations, independently sampled groups, an approximately normal outcome within each population for these sample sizes, and equal population variances.

Random assignment is additionally needed for a causal interpretation.

`T03-A12-V03`: Comparing quiet versus usual reading rooms

Reason before calculating

Different observational units belong to the two groups, so no case-level pairing is available.

Under the equal-population-variance model, $s_p^2 = [(12-1)42 + (12-1)35]/(12+12-2) = 38.5000$.

Work through the calculation

The standard error is $\sqrt{38.5000(1/12 + 1/12)} = 2.5331$.

With $df = 22$, $t = (82 - 76)/2.5331 = 2.3686$.

Because 2.3686 exceeds 1.7171, we reject the equal-mean null in favor of $\mu_1 > \mu_2$.

Interpret and check the result

The result supports a higher population mean in Group 1.

The procedure requires independent observations, independently sampled groups, an approximately normal outcome within each population for these sample sizes, and equal population variances.

Random assignment is additionally needed for a causal interpretation.

`T03-A12-V04`: Comparing checklist versus usual archive searches

Reason before calculating

Different observational units belong to the two groups, so no case-level pairing is available.

Under the equal-population-variance model, $s_p^2 = [(13-1)9 + (13-1)12]/(13+13-2) = 10.5000$.

Work through the calculation

The standard error is $\sqrt{10.5000(1/13 + 1/13)} = 1.2710$.

With $df = 24$, $t = (22 - 19)/1.2710 = 2.3604$.

Because 2.3604 exceeds 1.7109, we reject the equal-mean null in favor of $\mu_1 > \mu_2$.

Interpret and check the result

The result supports a higher population mean in Group 1.

The procedure requires independent observations, independently sampled groups, an approximately normal outcome within each population for these sample sizes, and equal population variances.

Random assignment is additionally needed for a causal interpretation.

`T03-A12-V05`: Comparing structured versus free-form notes

Reason before calculating

Different observational units belong to the two groups, so no case-level pairing is available.

Under the equal-population-variance model, $s_p^2 = [(14-1)28 + (14-1)32]/(14+14-2) = 30.0000$.

Work through the calculation

The standard error is $\sqrt{30.0000(1/14 + 1/14)} = 2.0702$.

With $df = 26$, $t = (61 - 56)/2.0702 = 2.4152$.

Because 2.4152 exceeds 1.7056, we reject the equal-mean null in favor of $\mu_1 > \mu_2$.

Interpret and check the result

The result supports a higher population mean in Group 1.

The procedure requires independent observations, independently sampled groups, an approximately normal outcome within each population for these sample sizes, and equal population variances.

Random assignment is additionally needed for a causal interpretation.

`T03-A12-V06`: Comparing orientation map versus standard leaflet

Reason before calculating

Different observational units belong to the two groups, so no case-level pairing is available.

Under the equal-population-variance model, $s_p^2 = [(15-1)40 + (15-1)36]/(15+15-2) = 38.0000$.

Work through the calculation

The standard error is $\sqrt{38.0000(1/15 + 1/15)} = 2.2509$.

With $df = 28$, $t = (70 - 65)/2.2509 = 2.2213$.

Because 2.2213 exceeds 1.7011, we reject the equal-mean null in favor of $\mu_1 > \mu_2$.

Interpret and check the result

The result supports a higher population mean in Group 1.

The procedure requires independent observations, independently sampled groups, an approximately normal outcome within each population for these sample sizes, and equal population variances.

Random assignment is additionally needed for a causal interpretation.

`T03-A12-V07`: Comparing peer explanation versus private rereading

Reason before calculating

Different observational units belong to the two groups, so no case-level pairing is available.

Under the equal-population-variance model, $s_p^2 = [(16-1)25 + (16-1)30]/(16+16-2) = 27.5000$.

Work through the calculation

The standard error is $\sqrt{27.5000(1/16 + 1/16)} = 1.8540$.

With $df = 30$, $t = (73 - 69)/1.8540 = 2.1574$.

Because 2.1574 exceeds 1.6973, we reject the equal-mean null in favor of $\mu_1 > \mu_2$.

Interpret and check the result

The result supports a higher population mean in Group 1.

The procedure requires independent observations, independently sampled groups, an approximately normal outcome within each population for these sample sizes, and equal population variances.

Random assignment is additionally needed for a causal interpretation.

`T03-A12-V08`: Comparing reduced versus usual notifications
Reason before calculating

Different observational units belong to the two groups, so no case-level pairing is available.

Under the equal-population-variance model, $s_p^2 = [(17 - 1)34 + (17 - 1)38]/(17 + 17 - 2) = 36.0000$.

Work through the calculation

The standard error is $\sqrt{36.0000(1/17 + 1/17)} = 2.0580$.

With $df = 32$, $t = (64 - 60)/2.0580 = 1.9437$.

Because 1.9437 exceeds 1.6939, we reject the equal-mean null in favor of $\mu_1 > \mu_2$.

Interpret and check the result

The result supports a higher population mean in Group 1.

The procedure requires independent observations, independently sampled groups, an approximately normal outcome within each population for these sample sizes, and equal population variances.

Random assignment is additionally needed for a causal interpretation.

`T03-A12-V09`: Comparing retrieval cards versus summary rereading
Reason before calculating

Different observational units belong to the two groups, so no case-level pairing is available.

Under the equal-population-variance model, $s_p^2 = [(18 - 1)45 + (18 - 1)41]/(18 + 18 - 2) = 43.0000$.

Work through the calculation

The standard error is $\sqrt{43.0000(1/18 + 1/18)} = 2.1858$.

With $df = 34$, $t = (77 - 72)/2.1858 = 2.2875$.

Because 2.2875 exceeds 1.6909, we reject the equal-mean null in favor of $\mu_1 > \mu_2$.

Interpret and check the result

The result supports a higher population mean in Group 1.

The procedure requires independent observations, independently sampled groups, an approximately normal outcome within each population for these sample sizes, and equal population variances.

Random assignment is additionally needed for a causal interpretation.

`T03-A12-V10`: Comparing visual versus text-only route guidance
Reason before calculating

Different observational units belong to the two groups, so no case-level pairing is available.

Under the equal-population-variance model, $s_p^2 = [(19-1)50 + (19-1)44]/(19+19-2) = 47.0000$.

Work through the calculation

The standard error is $\sqrt{47.0000(1/19 + 1/19)} = 2.2243$.

With $df = 36$, $t = (81-76)/2.2243 = 2.2479$.

Because 2.2479 exceeds 1.6883, we reject the equal-mean null in favor of $\mu_1 > \mu_2$.

Interpret and check the result

The result supports a higher population mean in Group 1.

The procedure requires independent observations, independently sampled groups, an approximately normal outcome within each population for these sample sizes, and equal population variances.

Random assignment is additionally needed for a causal interpretation.