# Concepts: $t$-based Inference for One Sample or Paired Samples

Statistical Sleuth Sections 2.1, 2.2, 2.5

## Overview

Our goal is to review $t$-based tools for inference about either:

- a single mean $\mu$; or
- a difference $\delta$ between two means, with *paired* data.

Hypothesis tests/p-values and confidence intervals answer two related but different questions about a parameter:

1. How much evidence do the data provide that a parameter is *different from a specified value*? (answer with hypothesis tests/p-values)

2. What is a *range of plausible values* for a parameter? (answer with confidence intervals)

## Hypothesis Test: How much evidence that the parameter is not equal to a given value?

**Reminder of what we did for the creative writing study:**

- **Sample statistic** is difference in group means: 19.883 - 15.739 = 4.144

- $p$-value: if $H_0 : \delta = 0$ is true, what is the probability of obtaining a difference at least as extreme as 4.144?

    - Small $p$-value is evidence against $H_0$.

- Based on the **sampling distribution** of difference in group means if $\delta = 0$

    - Look at differences in group means that would be obtained from every possible assignment of people to groups
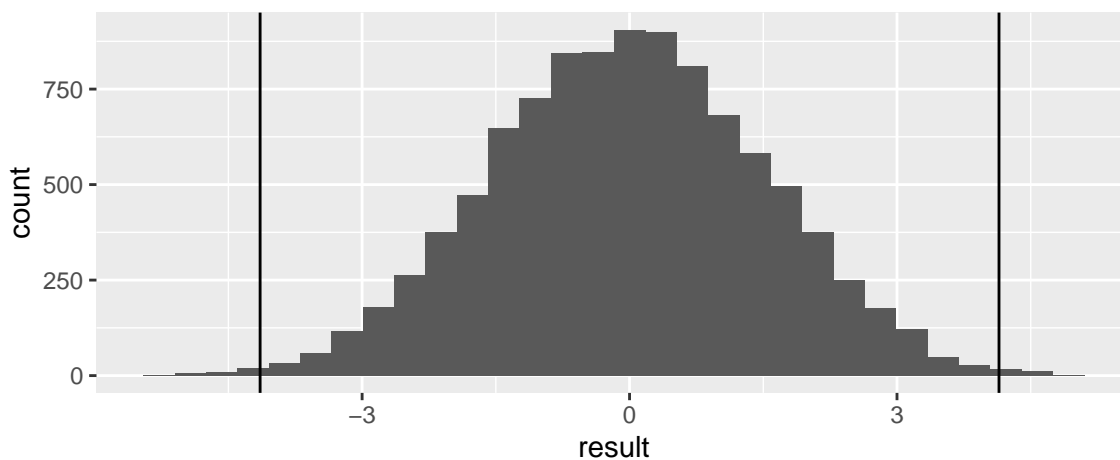


$p$-value is the proportion of these hypothetical samples that result in a difference in means that is greater than or equal to 4.144, or less than or equal to -4.144.

**$t$-based hypothesis tests - the short version**

Consider a test of $H_0 : \mu = \mu^{null}$ vs. $H_A : \mu \neq \mu^{null}$, where $\mu$ is a population mean.
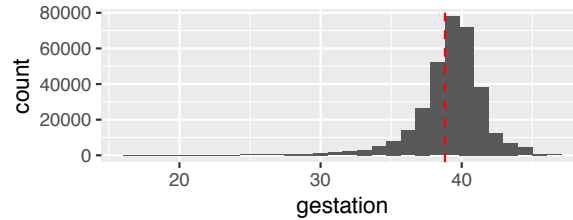
We need 3 ingredients for our test:

1. A **statistic**: $t = \frac{\bar{Y} - \mu^{null}}{SE(\bar{Y})} = \frac{\bar{Y} - \mu^{null}}{s/\sqrt{n}}$. How many standard errors away from $\mu^{null}$ is $\bar{Y}$?

2. The **sampling distribution** of the statistic, assuming $H_0$ is true: $t \sim t_{n-1}$

3. The **p-value** is the probability of getting a $t$ statistic at least as extreme as the value we calculate based on our observed sample, assuming $H_0$ is true.

**Example: Mean baby gestation time**

We have a data set with a record of every baby born in December 1998 in the United States, including the baby's gestation time in weeks (how many weeks pregnant the mother was when she gave birth), birth weight, and so on.
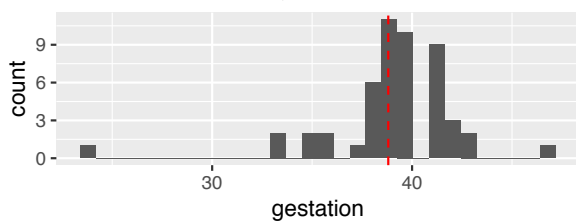
Let's explore what would happen if we used a sample from this population to estimate the population mean gestation time $\mu$.
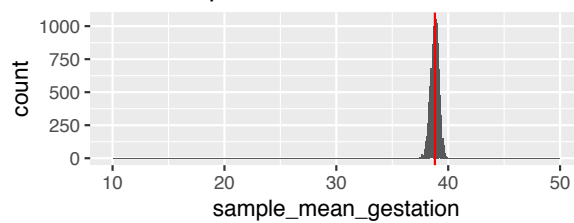
Population: 330,717 babies
mean 38.8 weeks, sd 2.61 weeks



**Draw 1 Sample**

**Draw Many Samples**

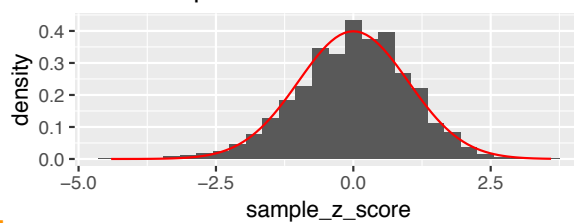...

Sample: 50 babies
mean 39.08 weeks, sd 3.28 weeks



Sample Means from 10000 samples
Each sample of size n = 30



$$z\text{-}\mathbf{score} = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} = \frac{39.08 - 38.8}{\boxed{2.61}/\sqrt{50}} \sim \mathbf{Normal}(0,1)$$

The standard deviation of values in the population. In real life we almost never know this!
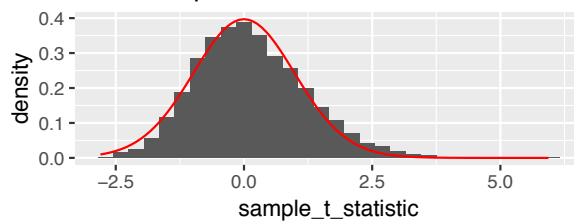
Sample Z–scores from 10000 samples
Each sample of size n = 50



$$t = \frac{\bar{Y} - \mu}{s/\sqrt{n}} = \frac{39.08 - 38.8}{\boxed{3.28}/\sqrt{50}} \sim t_{49}$$

The standard deviation of values in the sample. The best we can do without measuring everyone in the population!

Sample t statistics from 10000 samples
Each sample of size n = 50



2

**A few more details**

**Sample statistic** is a $t$ ratio.

- In general:
  - $t = \frac{\text{Estimate} - \text{Parameter}}{\text{SE(Estimate)}}$
  - SE(Estimate) measures the average distance of the estimate from the parameter
  - $t$ measures: "How many standard errors from the parameter was the estimate?"
  - The *degrees of freedom* is the sample size $n$ minus the number of parameters used to specify the mean
- In the case of one parameter, $\mu$:
  - $t = \frac{\bar{Y} - \mu^{null}}{\text{SE}(\bar{Y})}$
  - $\text{SE}(\bar{Y}) = \frac{s}{\sqrt{n}}$, where $s = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}}$ is the sample standard deviation
  - The *degrees of freedom* for $\text{SE}(\bar{Y})$ is $n-1$

$p$-value: if $H_0 : \mu = \mu^{null}$ is true, what is the probability of obtaining a $t$ statistic at least as extreme?

Based on the **sampling distribution** of $t$ if $\mu = \mu^{null}$

- Look at $t$ statistics that would be obtained from every possible sample of size $n$
- If the following conditions are satisfied, then $t \sim t_{n-1}$:
  - Distribution of values in the population is normally distributed
  - Observational units are independent (knowing one is above the mean doesn't give you information about whether or not another is above the mean)

$p$-value is the proportion of samples that result in $t$ statistics that are more extreme than the $t$ statistic obtained from our actual sample.
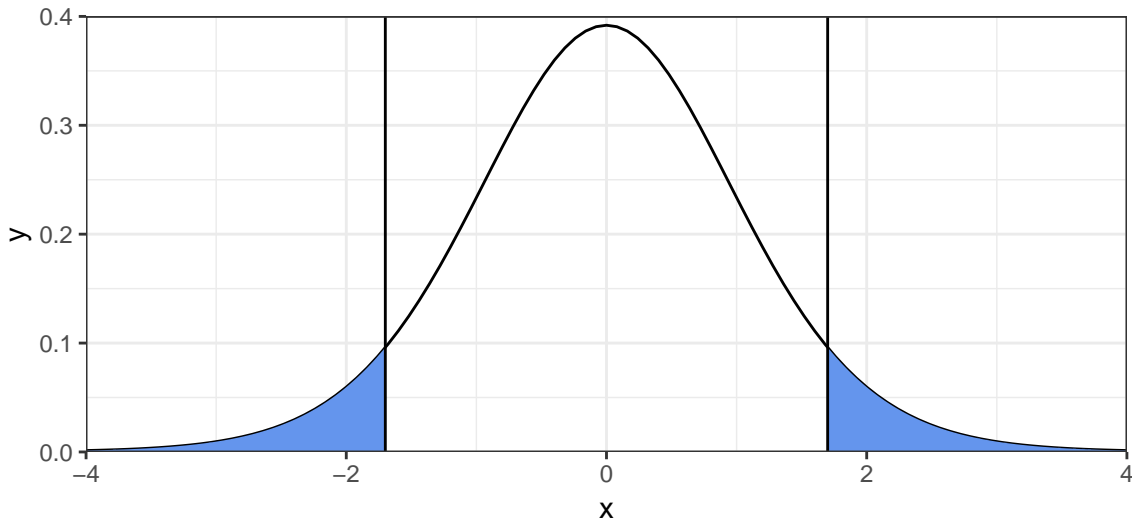


A small $p$-value means one of two things:

1. $H_0$ is correct, and you got a "weird"/unusual sample; or
2. $H_0$ is incorrect

The smaller the $p$-value, the stronger the evidence against the null hypothesis. Use this scale (See Section 2.5):

| $p$-value | Strength of evidence against $H_0$ | Compare to... |
| --- | --- | --- |
| 0.10 or less | Some evidence; not conclusive | Probability of 4 heads in a row is 0.0625 |
| 0.05 or less | Moderate amount of evidence | Probability of 5 heads in a row is 0.03125 |
| 0.01 or less | Strong evidence | Probability of 7 heads in a row is 0.007813 |
| 0.001 or less | Very strong evidence | Probability of 10 heads in a row is 0.0009766 |

## Confidence Interval: What is a range of plausible values for the parameter?

- General formula for a confidence interval: [Estimate - $t^*$ SE(Estimate), Estimate + $t^*$ SE(Estimate)]

- In the case of estimating a mean $\mu$:

$$[\bar{Y} - t^* SE(\text{Estimate}), \bar{Y} + t^* SE(\text{Estimate})]$$

For a 95% confidence interval, $t^*$ is the 97.5th percentile of the $t_{n-1}$ distribution. Why?

Interpretation: For 95% of samples, an interval calculated using this procedure will contain the parameter we are estimating.

Here is a picture of confidence intervals for the mean gestation time based on 100 different samples from the population, color coded by whether or not the interval contains the mean.

Out of these 100 samples, 92 produced confidence intervals that actually contained the population mean.
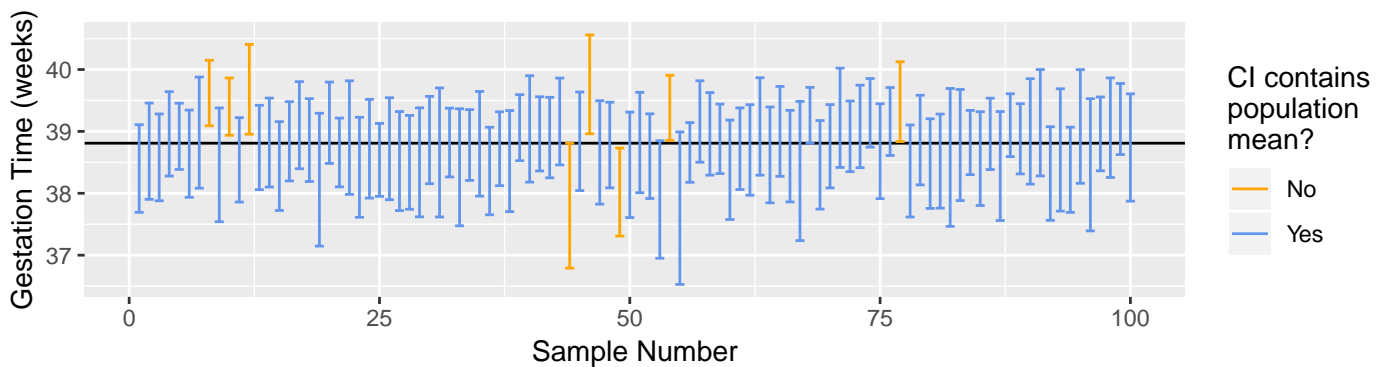
**Here's an example of confidence intervals from Section 2.5.2 of Sleuth 3.**

One of the papers Albert Einstein published in 1915 about relativity made predictions about the arc of deflection of light around the sun. The predictions were that the deflection of light would be given by the equation

$\frac{1}{2}(1 + \gamma)\frac{1.75}{d}$

Here $d$ is the distance of the closest approach of the light ray to the sun. The parameter $\gamma$ captures the effect on the deflection due to the sun's gravity curving space-time. According to Newtonian physics, $\gamma$ would have the value 0; according to Einstein's theory, $\gamma$ would have the value 1.

Over the years, many different experiments have been conducted to test the predictions by the theory, resulting in different estimates and confidence intervals for $\gamma$; these are displayed in the figure (which comes from Sleuth3).



DISPLAY 2.13  Estimates and confidence intervals for $\gamma$, the deflection of light around the sun, from 20 experiments
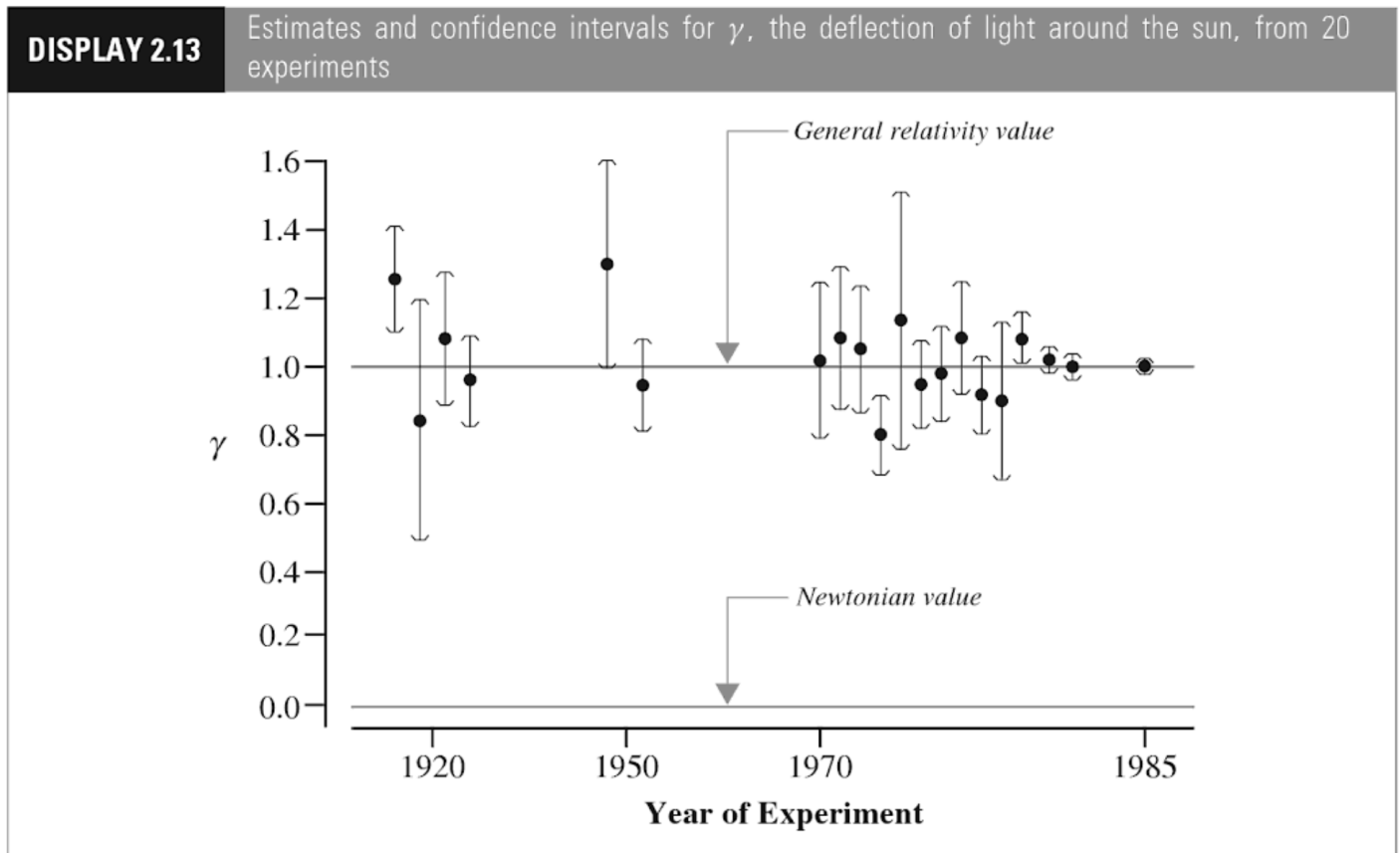
## References

1. Ramsey, F., and Schafer, D. (2012). The statistical sleuth: a course in methods of data analysis. Cengage Learning.