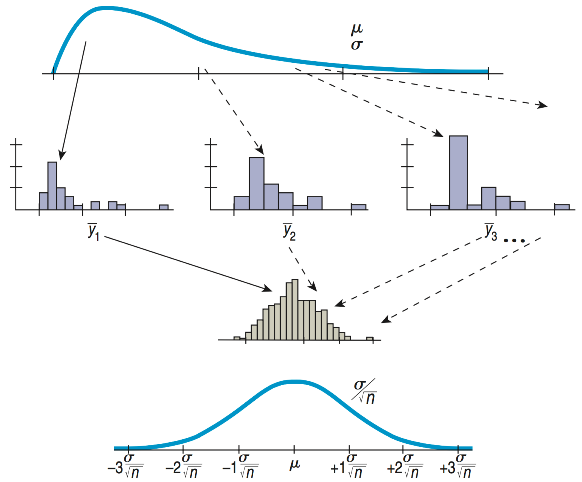
M. Singh

Recall the Central Limit Theorem



When a random sample is drawn from any population with mean (μ) and standard deviation (σ), its sample mean , has a sampling distribution with the same mean μ but the standard deviation is $\frac{\sigma}{\sqrt{n}}$

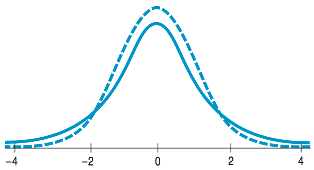
** no matter what population the random sample comes from, the shape of the sampling distribution is approximately normal as long as the sample size is large enough. The larger the sample used, the more closely the normal model approximates the sampling distribution for the mean.

** When we don't know (population SD), we approximate it with $SE(\bar{y}) = \frac{s}{\sqrt{n}}$ ← sample standard deviation

This leads to problems. Model doesn't give correct results.

- extra uncertaintly
- bigger margin of error

To resolve the issue, we have a different model for different degrees of freedom (df) ⟶ called student's *t-models*



The t-model (solid curve) on 2 degrees of freedom has fatter tails than the Normal model (dashed curve). So the 68–95–99.7 Rule doesn't work for t-models with only a few degrees of freedom.

** Student's t models are unimodal and symmetric like normal models but for small sample size (df) the tails are fatter.

As df increases, t model ⟶Normal

A PRACTICAL SAMPLING DISTRIBUTION MODEL FOR MEANS

When certain assumptions and conditions[2] are met, the standardized sample mean,

$$t = \frac{\bar{y} - \mu}{SE(\bar{y})},$$

follows a Student's *t*-model with $n - 1$ degrees of freedom. We estimate the standard deviation with

$$SE(\bar{y}) = \frac{s}{\sqrt{n}}.$$

Conditions: Independence, Randomization, 10%, nearly normal condition (data gives unimodal dist. Some skew of if sample large >= 40

ONE-SAMPLE *t*-INTERVAL FOR THE MEAN

When the assumptions and conditions[3] are met, we are ready to find the confidence interval for the population mean, μ. The confidence interval is

$$\bar{y} \pm t^*_{n-1} \times SE(\bar{y}),$$

where the standard error of the mean is $SE(\bar{y}) = \frac{s}{\sqrt{n}}$.

The critical value t^*_{n-1} depends on the particular confidence level, C, that you specify and on the number of degrees of freedom, $n - 1$, which we get from the sample size.

ONE-SAMPLE *t*-TEST FOR THE MEAN

The assumptions and conditions for the one-sample t-test for the mean are the same as for the one-sample t-interval. We test the hypothesis $H_0: \mu = \mu_0$ using the statistic

$$t_{n-1} = \frac{\bar{y} - \mu_0}{SE(\bar{y})}.$$

The standard error of $\bar{y}$ is $SE(\bar{y}) = \frac{s}{\sqrt{n}}$.

When the conditions are met and the null hypothesis is true, this statistic follows a Student's t-model with $n - 1$ degrees of freedom. We use that model to obtain a P-value.

Scenario: Motor Vehicle crashes → Leading cause of death for people between 4 - 33 yr old

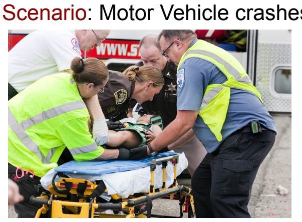
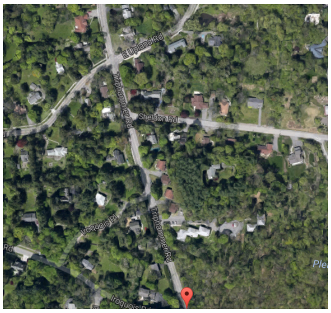


Photo of child on stretcher after crash

2006 --- 43,300 died in US

119 deaths each day, 1 death every 12 min on avg.

Speeding is a factor in 31% of all fatal accidents NHTSA



Triphammer Road, Ithaca, NY

Residents concerned...

WHO	Vehicles on Triphammer Road
WHAT	Speed >30mph
UNITS	Miles per hour
WHEN	April 11, 2000, 1 p.m.
WHERE	A small town in the northeastern United States
WHY	Concern over impact on residential neighborhood

Policemen put a radar to let drivers know their speed

↓

Residents don't think a sign is effective

↓

A resident stood by radar and recorded data

Speed		
29	29	24
34	34	34
34	32	36
28	31	31
30	27	34
29	37	36
38	29	21
31	26	

What is the true mean speed of all vehicles on Triphammer Road? Does the mean speed of all cars exceed the posted speed limit?

M. Singh

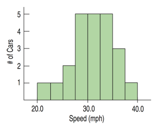
What is the true mean speed of all vehicles on Triphammer Road?

One sample t-interval for the mean

I want to find a 90% confidence interval for the mean speed, μ, of vehicles driving on Triphammer Road. I have data on the speeds of 23 cars there, sampled on April 11, 2000.

- ✔ **Independence Assumption:** This is a convenience sample, but care was taken select cars that were not driving near each other, so their speeds are plausibly independent.
- ✔ **Randomization Condition:** Not really me This is a convenience sample, but I have reason to believe that it is representati
- ✔ **10% Condition:** The cars I observed were fewer than 10% of all cars that travel Triphammer Road.
- ✔ **Nearly Normal Condition:** The histogram of the speeds is unimodal and symmetric.

Here's a histogram of the 23 observed speeds



The conditions are satisfied, so I will use a Student's t-model with

$(n - 1) = 22$ degrees of freedom

and find a **one-sample t-interval for the mean.**

Calculating from the data (see page 530):

$n = 23$ cars
$\bar{y} = 31.0$ mph
$s = 4.25$ mph.

The standard error of $\bar{y}$ is

$$SE(\bar{y}) = \frac{s}{\sqrt{n}} = \frac{4.25}{\sqrt{23}} = 0.886 \text{ mph.}$$

The 90% critical value is $t^*_{22} = 1.717$, so the margin of error is

$$\begin{aligned} ME &= t^*_{22} \times SE(\bar{y}) \\ &= 1.717(0.886) \\ &= 1.521 \text{ mph.} \end{aligned}$$

The 90% confidence interval for the mean speed is 31.0 ± 1.5 mph.

I am 90% confident that the interval from 29.5 mph to 32.5 mph contains the true mean speed of all vehicles on Triphammer Road.

Caveat: This was not a random sample of vehicles. It was a convenience sample taken at one time on one day. And the participants were not blinded. Drivers could see the police device, and some may have slowed down. I'm reluctant to extend this inference to other situations.

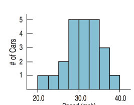
Does the mean speed of all cars exceed the posted speed limit?

One sample t-test for the mean

I want to know whether the mean speed of vehicles on Triphammer Road exceeds the posted speed limit of 30 mph. I have a sample of 23 car speeds on April 11, 2000.

H_0: Mean speed, $\mu = 30$ mph
H_A: Mean speed, $\mu > 30$ mph

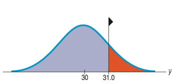
- ✔ **Independence Assumption:** These cars are a convenience sample, but they were selected so no two cars were driving near each other, so I am justified in believing that their speeds are independent.
- ✔ **Randomization Condition:** Although I have a convenience sample, I have reason to believe that it is a representative sample.
- ✔ **Nearly Normal Condition:** The histogram of the speeds is unimodal and reasonably symmetric.



The conditions are satisfied, so I'll use a Student's t-model with $(n - 1) = 22$ degrees of freedom to do a **one-sample t-test for the mean.**

$n = 23$ cars
$\bar{y} = 31.0$ mph
$s = 4.25$ mph

$$SE(\bar{y}) = \frac{s}{\sqrt{n}} = \frac{4.25}{\sqrt{23}} = 0.886 \text{ mph.}$$



$$t = \frac{\bar{y} - \mu_0}{SE(\bar{y})} = \frac{31.0 - 30.0}{0.886} = 1.13$$

$$P\text{-value} = P(t_{22} > 1.13) = 0.136$$

The P-value of 0.136 says that if the true mean speed of vehicles on Triphammer Road were 30 mph, samples of 23 vehicles can be expected to have an observed mean of at least 31.0 mph 13.6% of the time. That P-value is not small enough for me to reject the hypothesis that the true mean is 30 mph at any reasonable alpha level. I conclude that there is not enough evidence to say the average speed is too high.