Confidence Intervals - Ideas: Normal population

Estimators for the normal parameters

Let $X_1, X_2, \ldots, X_n$ sample from N(μ, σ), μ, σ are unknown. We need to estimate both μ, and σ.

Point estimator of μ: $\hat{\mu} = \bar{X}$

Point estimator of variance σ^2:

$$\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Both estimators are very good. They are unbiased and have the smallest spread among all unbiased estimators.

They are called *best* (unbiased and minimum variance) point estimators of the corresponding parameters.

CONFIDENCE INTERVAL FOR MEAN WITH KNOWN σ.

Often we seek a *range* of values – an interval – for the unknown parameter.

We want to have some confidence that the interval covers the parameter – CONFIDENCE INTERVAL (CI)

We construct CIs based on sample data and theoretical knowledge about distributions, like the Central Limit Theorem.

DERIVATION OF THE 95% CONFIDENCE INTERVAL FOR THE MEAN, NORMAL POPULATION

Let $X_1, X_2, \ldots, X_n$ **sample from N(μ, σ), μ unknown, σ known. Then**

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$ **has the standard normal distribution.**

Note, P(-1.96 ≤ Z ≤ 1.96) = 0.95. So $$P(-1.96 \le \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \le 1.96) = 0.95.$$

Rearranging, $$P(\bar{X} - 1.96\sigma / \sqrt{n} \le \mu \le \bar{X} + 1.96\sigma / \sqrt{n}) = 0.95.$$

The interval $\bar{X} \pm 1.96\sigma / \sqrt{n}$ **covers/contains μ with 95% chance.**

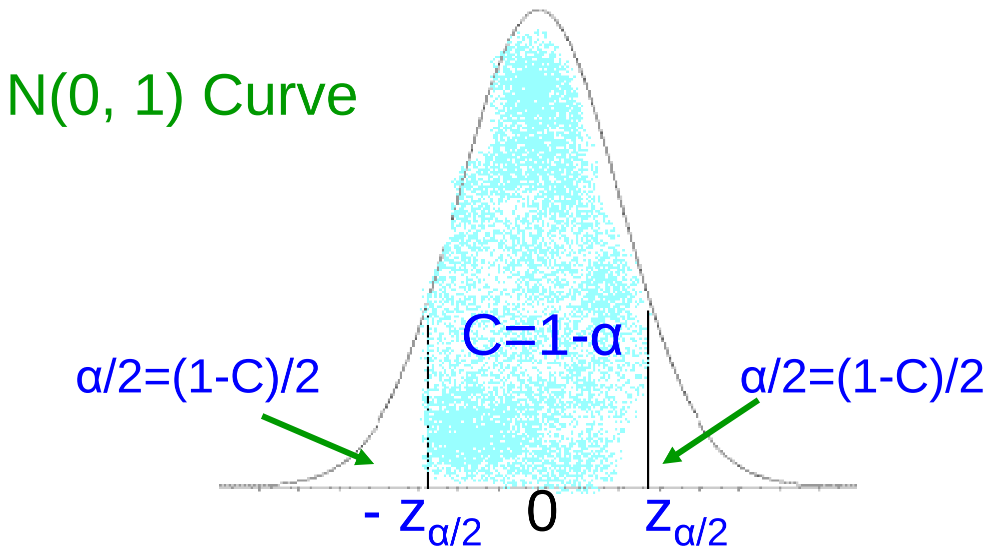
The interval ($\bar{X}$ -1.96σ√n, $\bar{X}$ +1.96σ√n) is called the 95% Confidence Interval for the mean.

A GENERAL C=(1-α) 100% CONFIDENCE INTERVAL FOR THE MEAN, NORMAL DATA, σ KNOWN.

What if we want to have a confidence level different from 95%?

We need to choose a different z-value.

Let confidence level C=1-α, the corresponding z-value is called $z_{\alpha/2}$.



A C=(1-α)100% confidence interval for μ is given by

$$\bar{X} \pm z_{\alpha/2}\sigma/\sqrt{n}$$

If the data is from a normal population with σ known.

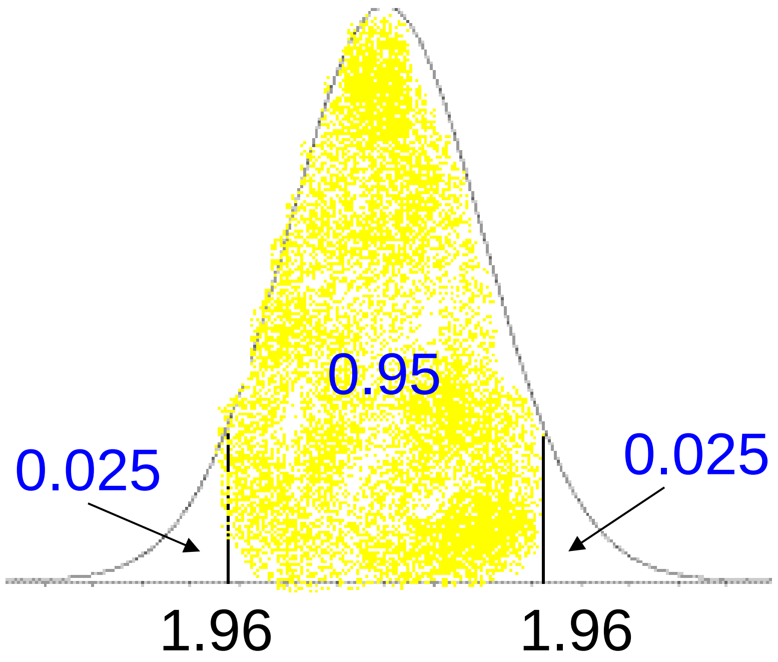
EXAMPLES OF $z_{\alpha/2}$.

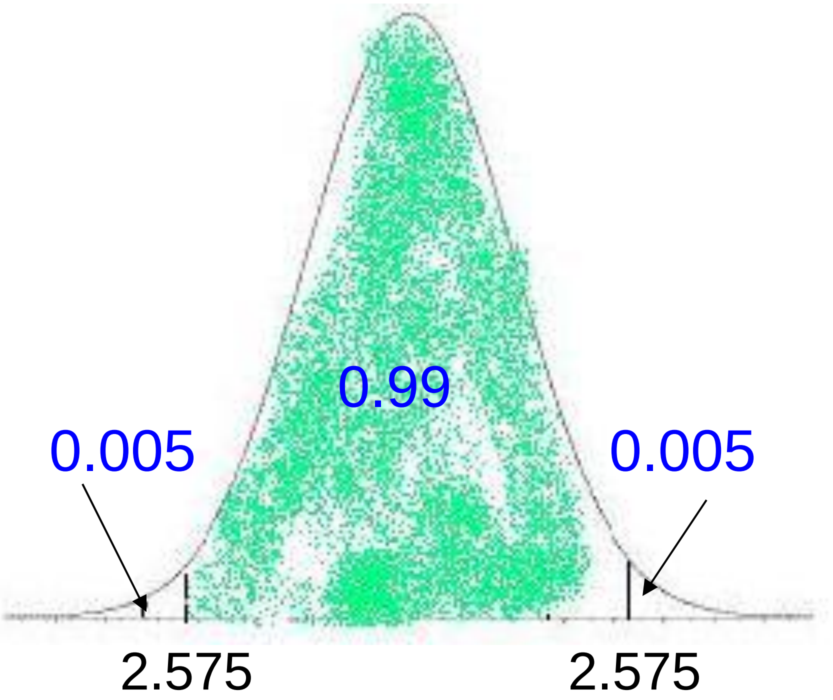
- **Take confidence level C=95%. Find $z_{\alpha/2}$.**

C=0.95, so α=0.05 → $\alpha/2$=0.025. From Z-table $z_{0.025}=1.96$.

- **Take confidence level 99%. Find $z_{\alpha/2}$.**

C=0.99, so α=0.01 → $\alpha/2$=0.005. From Z-table $z_{0.005}=2.575$.





Example

Suppose X, scores on a test, have a normal distribution with unknown mean and σ = 60. The sample values X1, X2, …, X900 give sample mean $\bar{X} = 272.$ Find a 95% as well as 98% confidence intervals for the true mean test score.

Solution. Point estimate of μ is $\bar{X} = 272,$ n=900, σ = 60.

For 95% CI, $z_{\alpha/2}$=1.96, so 95% CI is: $(\bar{X} - 1.96\sigma / \sqrt{n}, \bar{X} + 1.96\sigma / \sqrt{n})$

=(272 - 1.96(60/√900, 272 + 1.96(60/√900) = (268.08, 275.92).

For 98% CI, $z_{\alpha/2}$=2.33, so 98% CI is: $(\bar{X} - 2.33\sigma / \sqrt{n}, \bar{X} + 2.33\sigma / \sqrt{n})$

=(272 – 2.33(60/√900, 272 + 2.33(60/√900) = (267.35, 276.65).

Note that the 98% CI is longer than the 95% CI. We gained confidence, but we lost accuracy.

Illustration of Capturing True Mean

- Here is a normal curve, which represents the distribution of $\overline{X}$.
- The middle 95% of the curve, extending a distance of $1.96\sigma_{\overline{X}}$ on either side of the population mean μ, is shaded.
- The following illustrates a 95% CI if $\overline{X}$ lies within the middle 95% of the distribution:

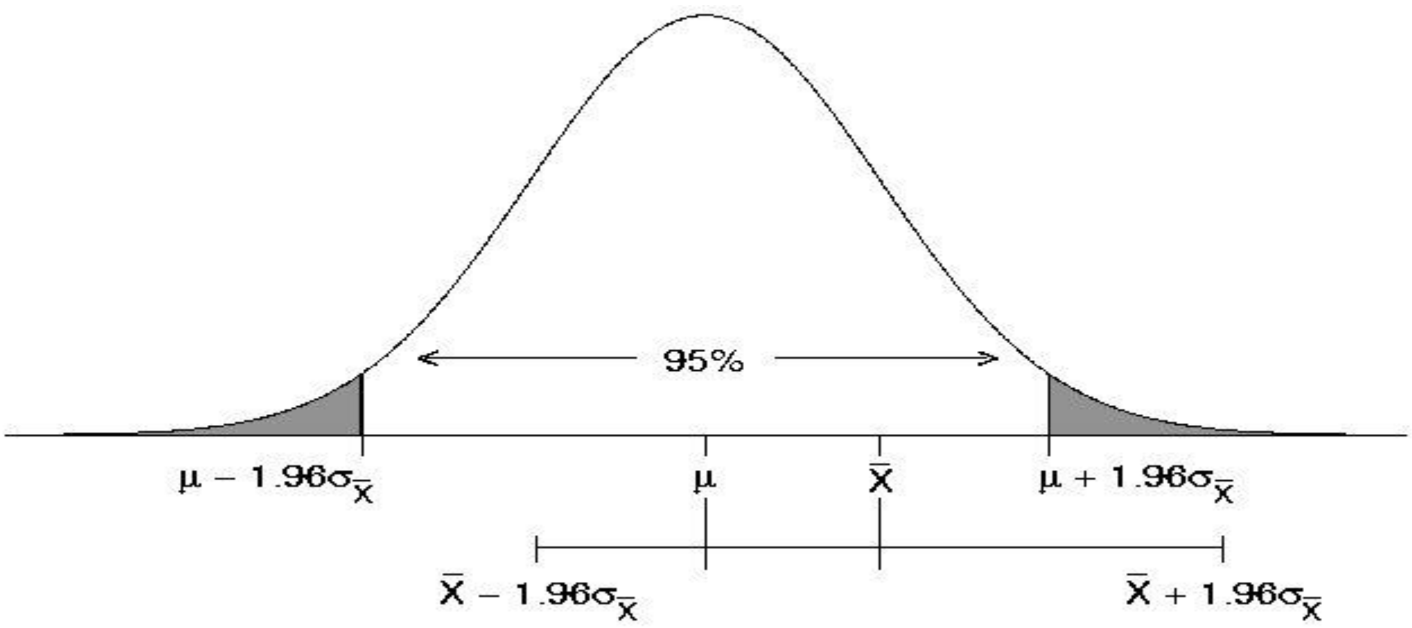
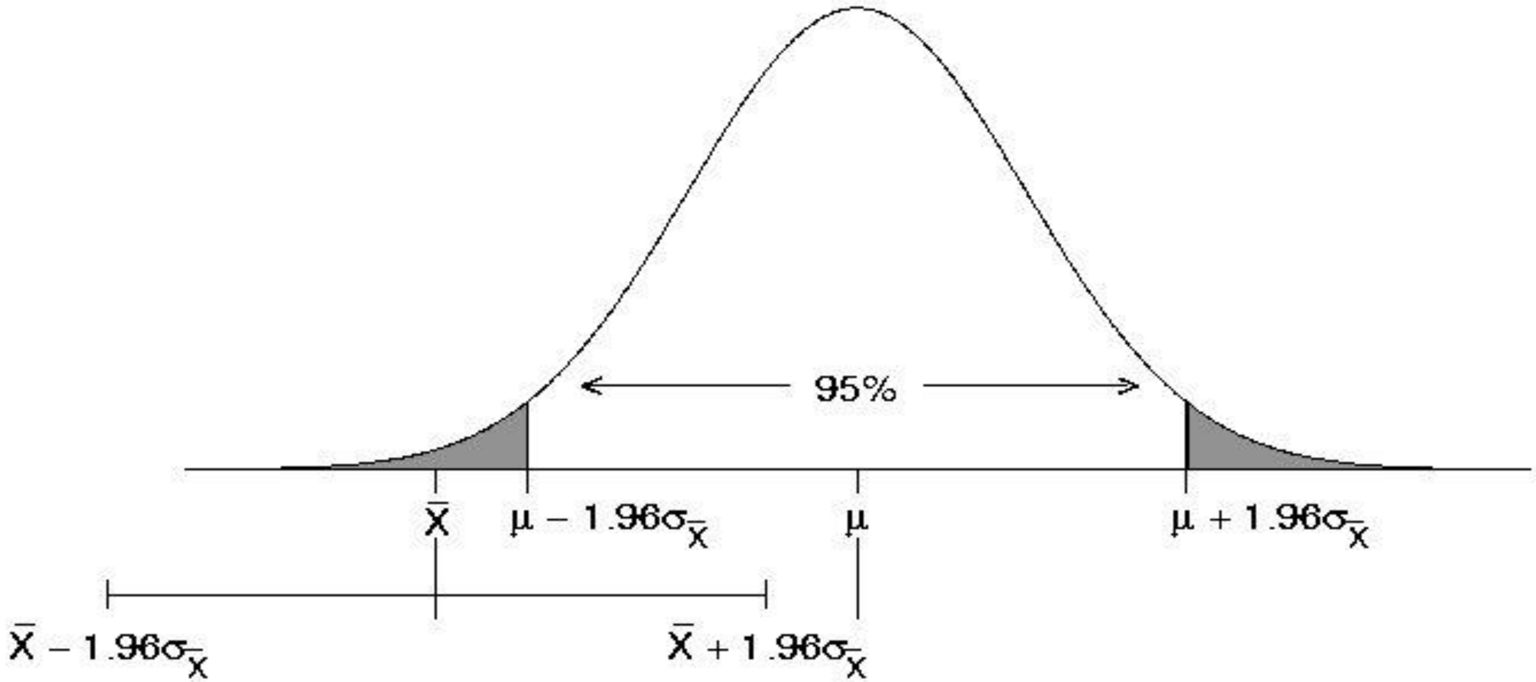


Illustration of Not Capturing True Mean

- In this example, the sample mean lies outside the middle 95% of the curve.
- Only 5% of all the samples that could have been drawn fall into this category.
- For those more unusual samples, the 95% confidence interval $\bar{X} \pm 1.96\sigma_{\bar{X}}$ fails to cover the true population mean μ.



What if the population is not normal?

If the population is not normal, but the sample size is large, then by the Central Limit Theorem

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

has an approximately standard normal distribution no matter what the population is.

So, provided that the sample size is large, we can use the same confidence intervals for data sets from any population.

This is one of many practical applications of the CLT!

CONFIDENCE INTERVAL FOR MEAN WITH KNOWN σ – choice of the sample size for a given margin of error.

Let $X_1, X_2, \ldots, X_n$ **sample from N(μ, σ), μ unknown, σ known.**

The margin of error m= half-length of the CI for μ, m $= z_{\alpha/2}\sigma/\sqrt{n}$ **depends on:**

- **the confidence level via** $z_{\alpha/2}$ **(as conf. level** $\uparrow$**, m** $\uparrow$ **);**
- **the variability in the population σ (as σ** $\uparrow$**, m** $\uparrow$**);**
- **the sample size (as n** $\uparrow$**, m** $\downarrow$ **).**

To decrease the error, but keep the confidence level unchanged, we need to increase the sample size.

$$m = z_{\alpha/2}\sigma/\sqrt{n} \text{ or } n = \frac{z_{\alpha/2}^2\sigma^2}{m^2}.$$

For a (1-α) CI for μ (σ known) to have margin of error m, we need sample size

$$n = \frac{z_{\alpha/2}^2\sigma^2}{m^2}.$$

Example

Suppose $X_1, X_2, \ldots, X_n$ are heights from a normal distribution with σ=10''. In a sample of size 16 we obtained $\bar{X} = 60"$.

a) Find a 95% confidence interval for μ. What is its margin of error?
b) Find the sample size needed for the margin of error to be 3 inches.

Solution. σ =10'', n=16, m=3.

a) C = 0.95, so α = 0.05, so $z_{\alpha/2}$=1.96. 95% CI for the mean (sampling from normal distribution) is:

$$(\bar{X} - 1.96\sigma/\sqrt{n}, \bar{X} + 1.96\sigma/\sqrt{n})$$ **=(60 +/-1.96(10)/√16) = (55.1, 64.9).**

Margin of error=(64.9 - 55.1)/2 = 4.9.

b) $$n = \frac{z_{\alpha/2}^2 \sigma^2}{m^2} = \frac{1.96^2 10^2}{3^2} = \left(\frac{1.96 \times 10}{3}\right)^2 = 42.68.$$ **Finally, n = 43.**

We need a larger sample size to get smaller error, with the same confidence level.

Question re last example

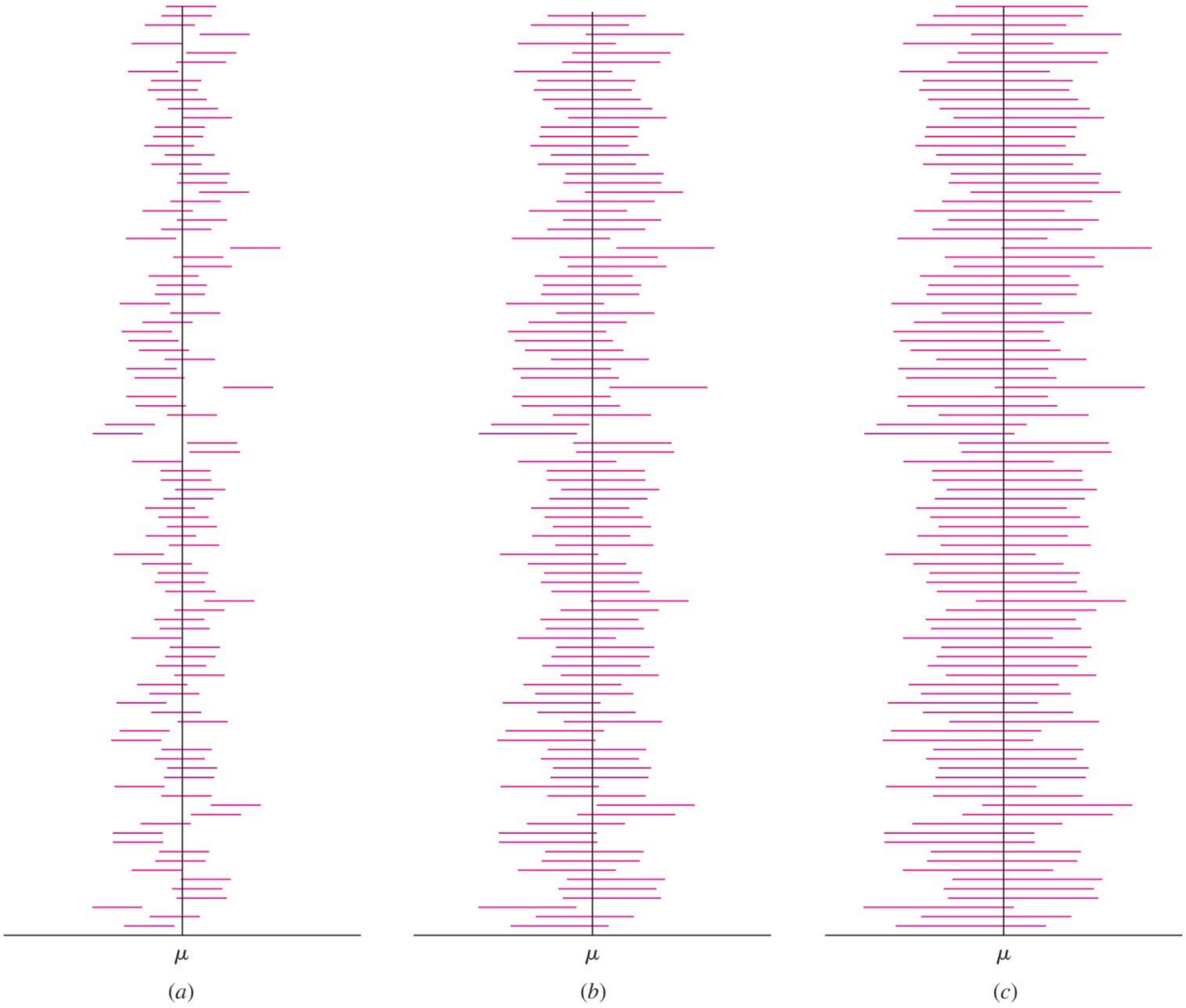
Does this 95% confidence interval (55.1, 64.9) actually cover the population mean μ?

- **It depends on whether this particular sample happened to have mean in the middle 95% of the distribution, or whether it was a sample with an unusually large or small mean, in the outer 5% of the population.**

- **There is no way to know for sure into which category this particular sample falls.**

- **In the long run, if we repeatedly constructed these confidence intervals, then 95% of the samples will have means in the middle 95% of the population and 95% of the confidence intervals will cover the population mean.**

Coverage of the mean for Cis with increasing level of confidenc



Probability vs. Confidence

- **In computing a CI, it is tempting to say that the probability that μ lies in this interval is 95%.**

- **The term *probability* refers to random events, which can come out differently when experiments are repeated.**

- **The numbers 55.1 and 64.9 which are ends of the CI (55.1, 64.9) are fixed, not random. The population mean is also fixed. The mean height is either in the interval or not.**

- **There is no randomness involved.**

- **So, we say that we have 95% *confidence* (not probability) that the population mean is in this interval.**

CONFIDENCE INTERVAL FOR MEAN WITH UNKNOWN σ

If the population variance σ is unknown, estimate is using the sample variance S:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2.$$

Then, replace σ with S in the Z-statistic used for the confidence interval:

Get t-statistic:

$$t \equiv \frac{\overline{X} - \mu}{S/\sqrt{n}}.$$

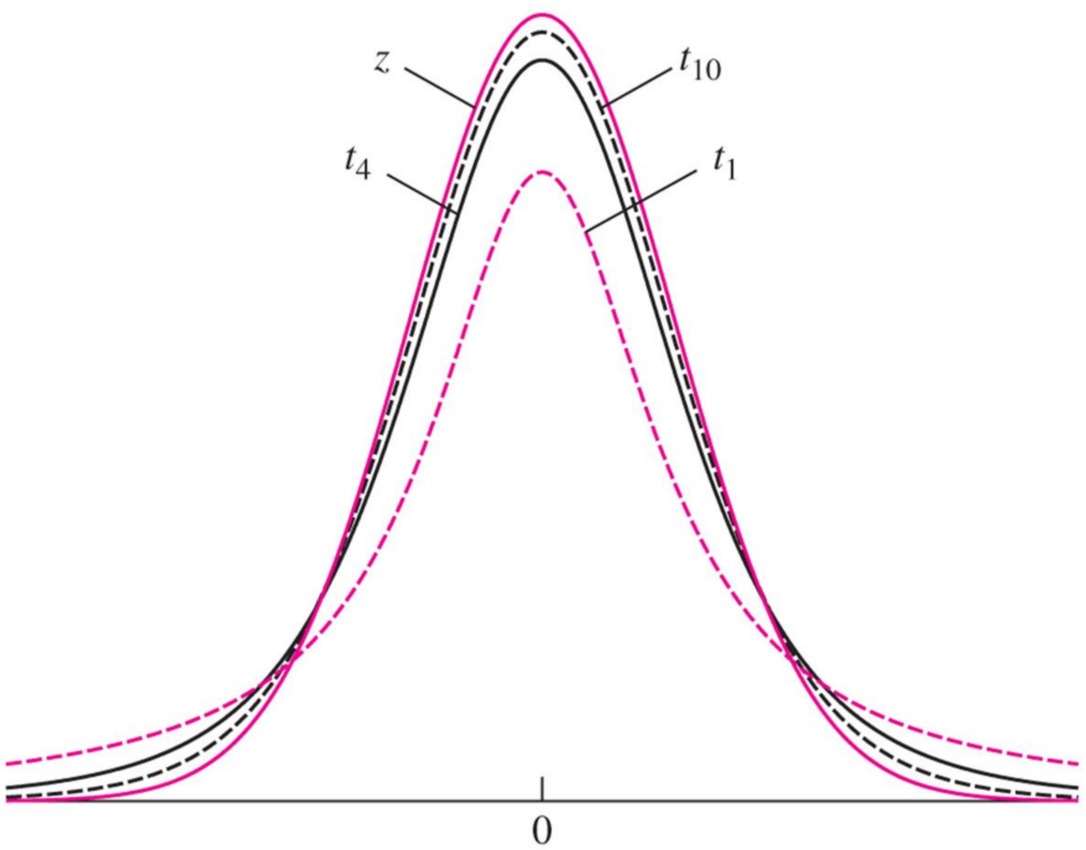
The t-statistic does not have standard normal distribution.

It has a t distribution with (n-1) degrees of freedom. The number of degrees of freedom is a parameter of the t distribution.

NOTE. T-distribution is also called “Student’s t distribution”.

T-distribution

T-distribution has similarities and differences with standard Normal distribution: symmetric around zero, but it has fatter (heavier) tails.



As degrees of freedom increase, t distribution becomes indistinguishable from standard normal.

CONFIDENCE INTERVAL FOR MEAN WITH UNKNOWN σ

Following a procedure similar to the one for constructing CI for μ when σ was known, when data was from a normal population, we can find a CI for μ when σ is not known. We replace σ with S and standard normal percentiles with percentiles from t distribution.

Let $X_1, X_2, \ldots, X_n$ be observations from a normal distribution with mean μ and standard deviation σ, both μ and σ unknown.

A C=(1-α) confidence interval for μ is given by

$$\bar{X} \pm t_{\alpha/2} S / \sqrt{n}$$

when the data is from a normal population with σ unknown.

Here $t_{\alpha/2}$ satisfies $P(t(n-1) > t_{\alpha/2}) = \alpha/2$, i.e. $t_{\alpha/2}$ is a percentile from t(n-1) distribution.

EXAMPLE

A biologist studying brain weights of tigers took a random sample of 16 animals and measured their brain weights in ounces. This data gave a sample mean of 10 and standard deviation of 3.2 ounces. Assuming that these weights follow a Normal distribution, find a 95% confidence interval for the true mean tiger brain weight μ.

Solution. $\overline{X}$ =10, s=3.2, n=16, need 95%CI for μ.

C=0.95, so α=0.05, so α/2=0.025. Since n=16, then n-1=15.

We need $t(15)_{0.025}$. From the table $t(15)_{0.025}=2.131$, so the 95% CI for μ is:

10+/- 2.131 (3.2)/√16 = (8.295, 11.705) oz.

We are 95% confident that the true mean weight of a tiger brain is between 8.295 and 11.705 oz.

MINITAB EXAMPLES. EXAMPLE 1

A graduate class in probability theory was running in a university for several years. The professor wanted to estimate the average final exam results. She took a sample of 25 exams. Find a 90% CI for the true mean final score for this course.

Solution: I used MINITAB (data available on class web site probs.xls). We do not know population st. deviation, so we need to use t-distribution to get CI.

MINITAB output in the "Session" window

T Confidence Intervals

Variable	N	Mean	StDev	SE Mean	90.0 % CI
score	25	84.68	5.72	1.14	(82.72, 86.64)

In the output: SE Mean = estimate of the Standard Error of the Mean

that is $\widehat{\sigma_x} = \frac{s}{\sqrt{n}} = 5.\frac{72}{5} = 1.144.$

MINITAB EXAMPLES. EXAMPLE 2

Use the data on "freshmen's 15"

Find the 90% CI for the mean BMI of a freshman student in September.

Results for: FRESH15v16.MTW

One-Sample T: BMISP

Variable	N	Mean	StDev	SE Mean	90% CI
BMISP	67	22.030	3.309	0.404	(21.356, 22.704)