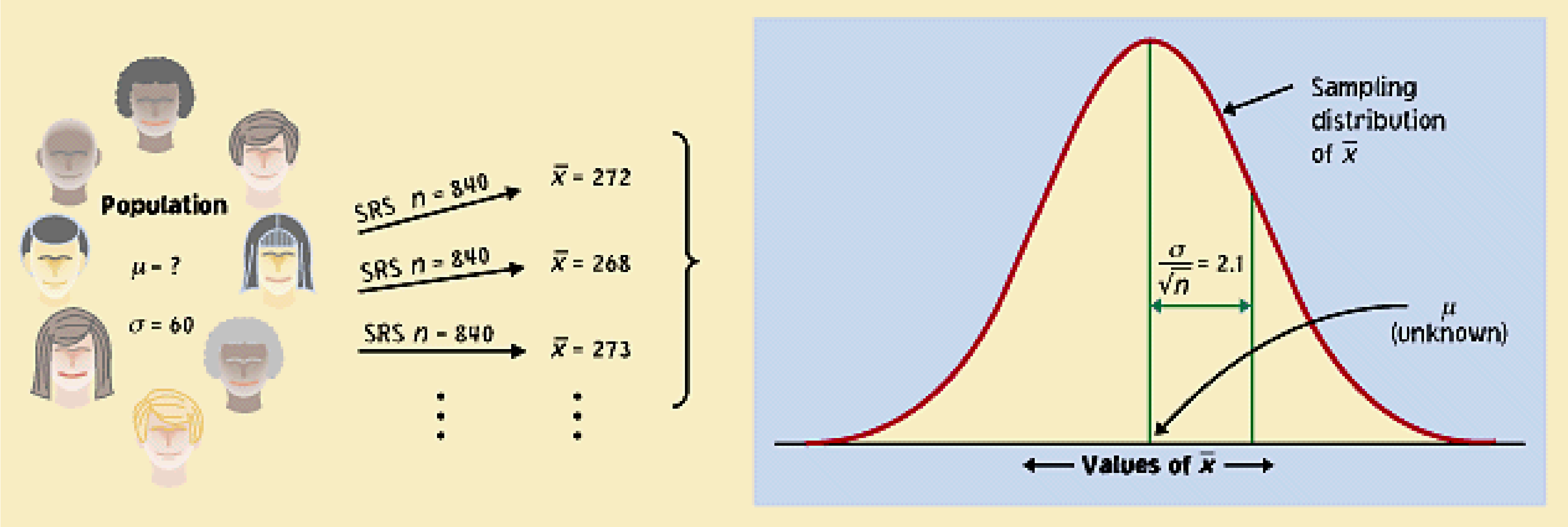
Statistical Lecture 2

Statistical Inference. Inference based on a single population sample

Uncertainty and confidence

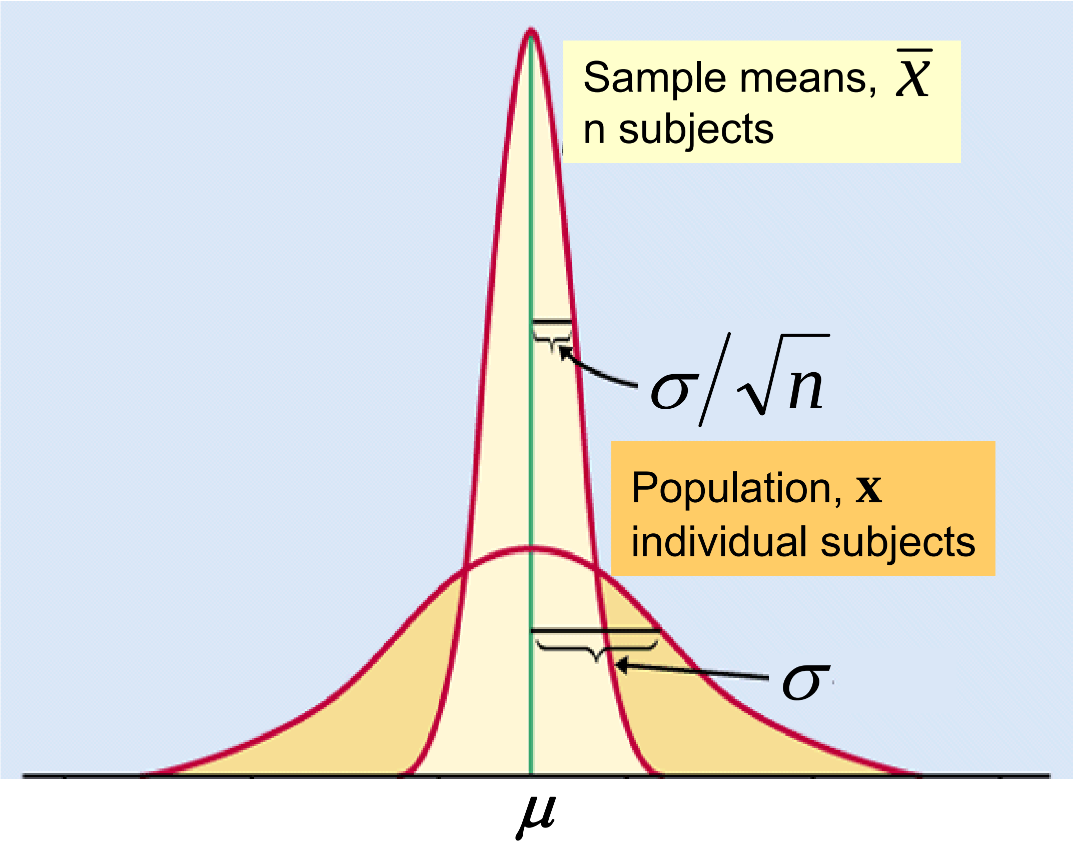
Although the sample mean, $\bar{x}$, is a unique number for any particular sample, if you pick a different sample you will probably get a different sample mean.

In fact, you could get many different values for the sample mean, and virtually none of them would actually equal the true population mean, μ.



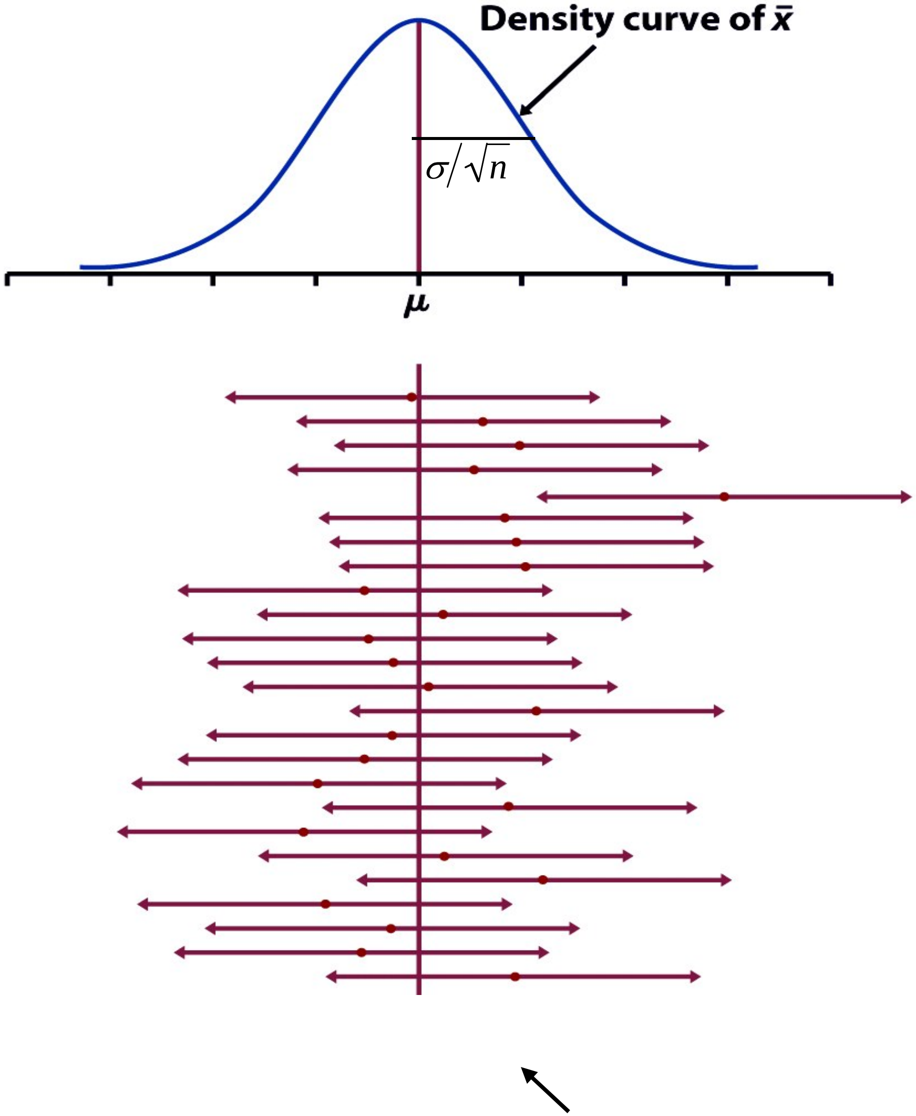
But the sample distribution is narrower than the population distribution, by a factor of $\sqrt{n}$ (by CLT)

Thus, the estimates $\bar{x}$ gained from our samples are always relatively close to the population parameter μ.



If the population is normally distributed $N(\mu,\sigma)$, so will the sampling distribution $N(\mu,\sigma/\sqrt{n})$, 95% of all sample means will be within roughly 2 standard deviations ($2*\sigma/\sqrt{n}$) of the population parameter μ.

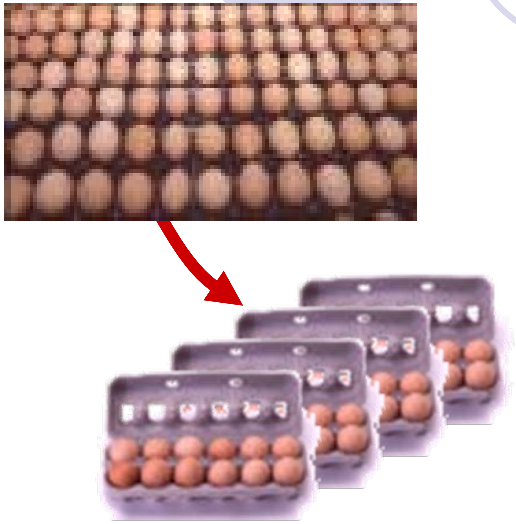
Because distances are symmetrical, this implies that **the population parameter μ must be within roughly 2 standard deviations from the sample average $\bar{x}$, in 95% of all samples.**



Red dot: mean value of individual sample

This reasoning is the essence of statistical inference.

The weight of single eggs of the brown variety is normally distributed *N*(65 g,5 g). Think of a carton of 12 brown eggs as a random sample of size 12.

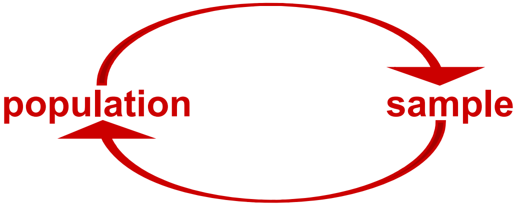


- What is the distribution of the sample mean $\bar{x}$?

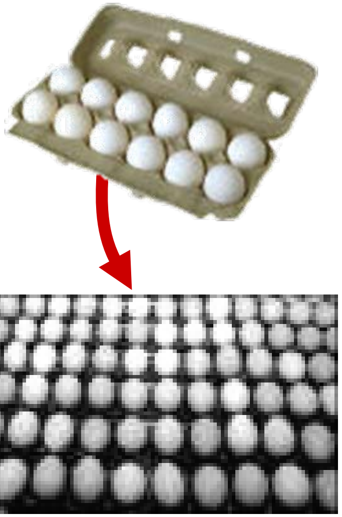
Normal (mean μ, standard deviation $\sigma/\sqrt{n}$) = *N*(65 g,1.44 g).

- Find the middle 95% of the sample means distribution.

Roughly ± 2 standard deviations from the mean, or 65g ± 2.88g.



You buy a carton of 12 white eggs instead. The box weighs 770 g. The average egg weight from that SRS is thus $\bar{x}$ = 64.2 g.

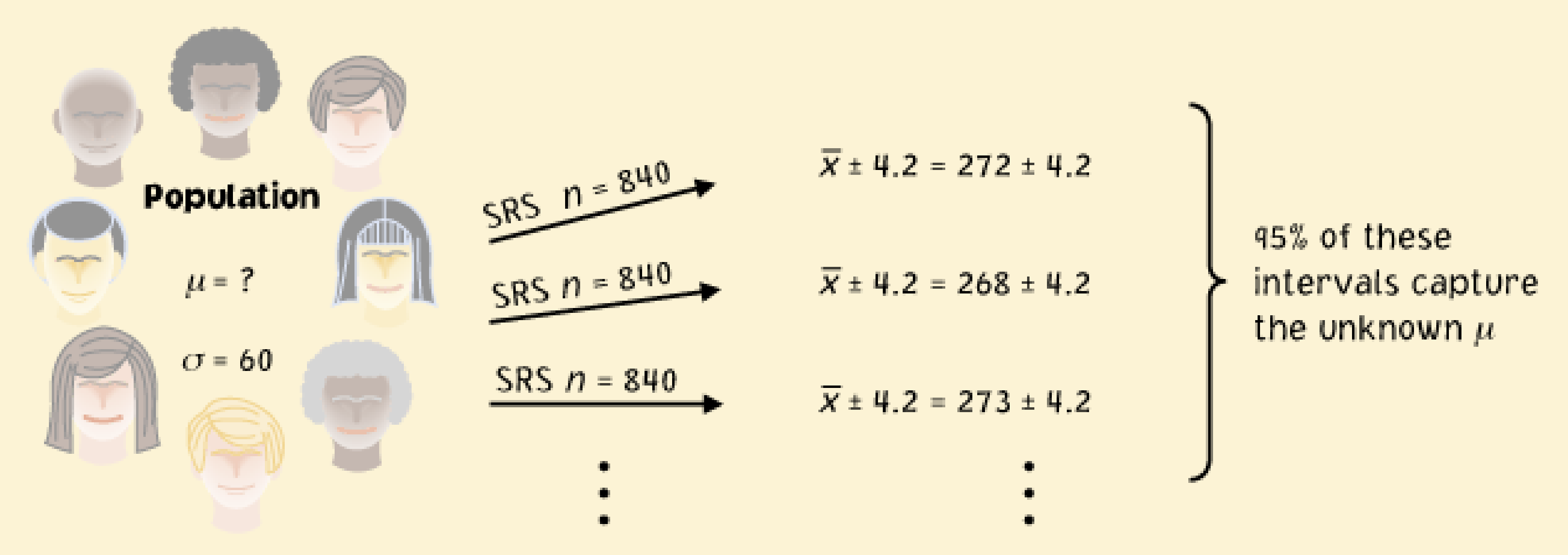


- Knowing that the standard deviation of egg weight is 5 g, what can you infer about the mean μ of the white egg population?

There is a 95% chance that the population mean μ is roughly within ± $2\sigma/\sqrt{n}$ of $\bar{x}$, or 64.2 g ± 2.88 g.

Confidence interval

The **confidence interval** is a range of values with an associated probability or **confidence level *C***. The probability quantifies the chance that the interval contains the true population parameter.

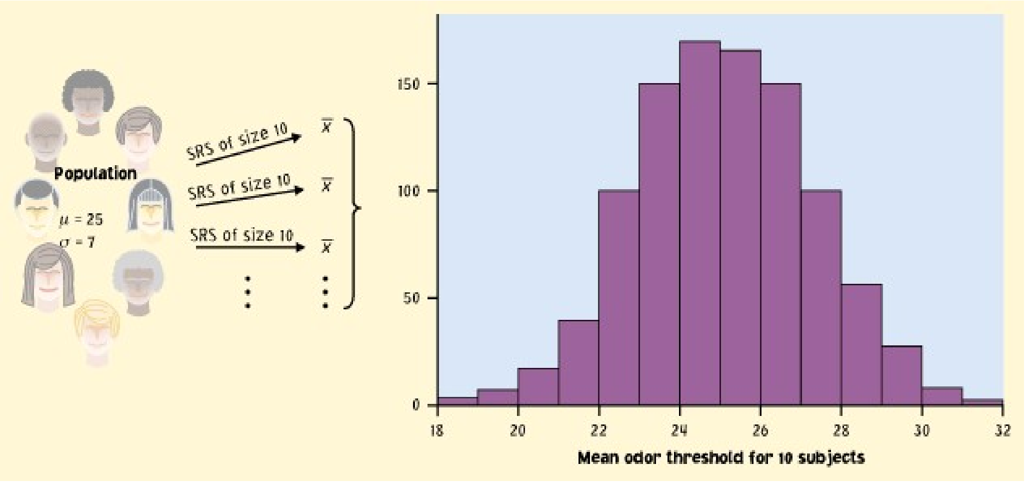


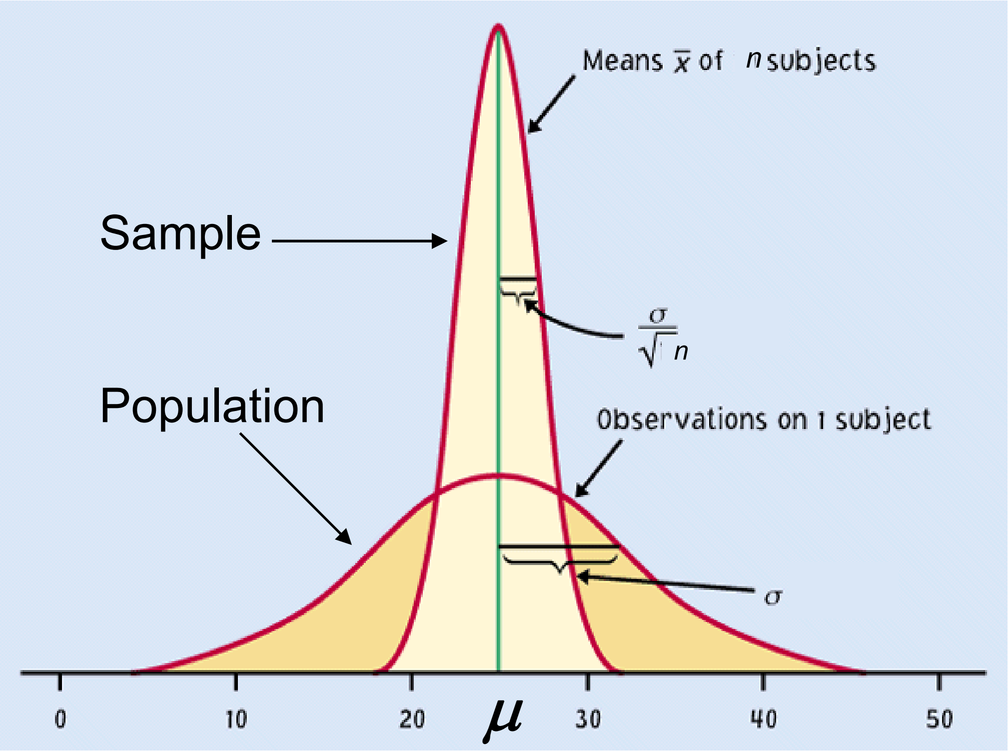
$\bar{x} \pm 4.2$ is a 95% confidence interval for the population parameter μ.

This equation says that in 95% of the cases, the actual value of μ will be within 4.2 units of the value of $\bar{x}$.

Implications

We don’t need to take a lot of random samples to “rebuild” the sampling distribution and find μ at its center.



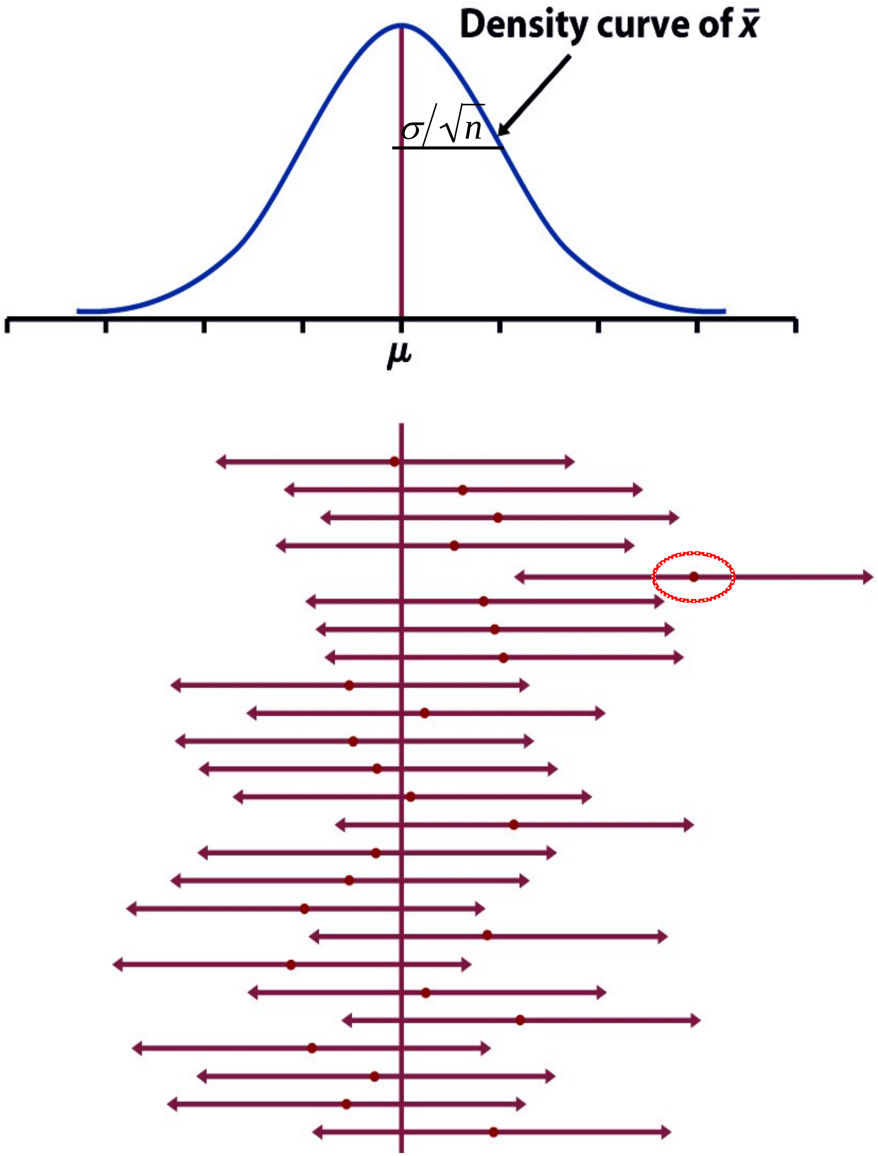


All we need is one SRS of size *n* and relying on the properties of the sample means distribution to infer the population mean μ.

Reworded

With 95% confidence, we can say that μ should be within roughly 2 standard deviations ($2^*\sigma/\sqrt{n}$) from our sample mean $\bar{x}$ bar.

- In 95% of all possible samples of this size n, μ will indeed fall in our confidence interval.
- In only 5% of samples would $\bar{x}$ be farther from μ.



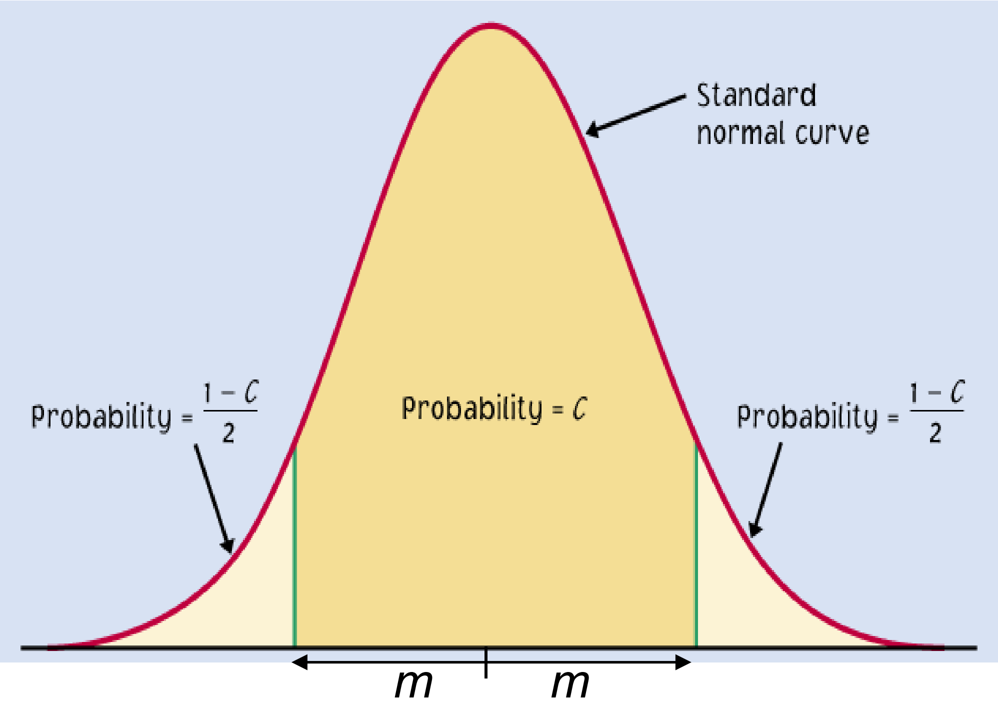
A **confidence interval** can be expressed as:

- Mean ± *m*
  m is called the **margin of error**
  μ within $\bar{x} \pm m$
  Example: 120 ± 6

- Two endpoints of an interval
  μ within $(\bar{x} - m)$ to $(\bar{x} + m)$
  ex. 114 to 126

A **confidence level *C*** (in %) indicates the probability that the μ falls within the interval.

It represents the area under the normal curve within ± *m* of the center of the curve.



Varying confidence levels

Confidence intervals contain the population mean μ in *C*% of samples.

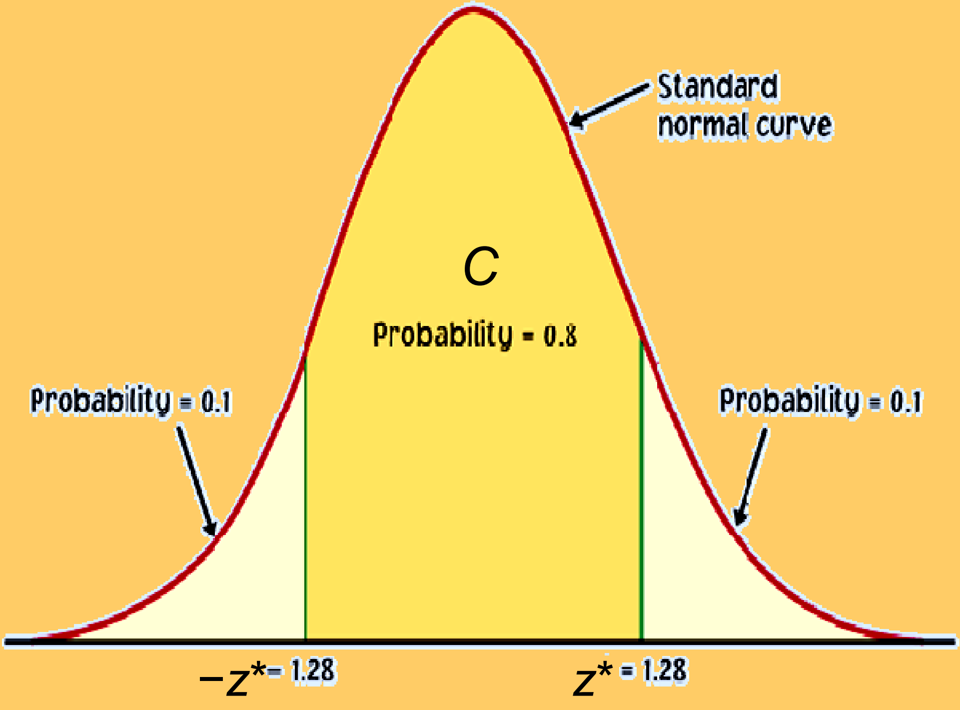
Different areas under the curve give different confidence levels *C*.

Practical use of z: z*

- *z** is the critical value related to the chosen confidence level *C*.
- *C* is the area under the standard normal curve between −*z** and *z**.

The confidence interval is thus:

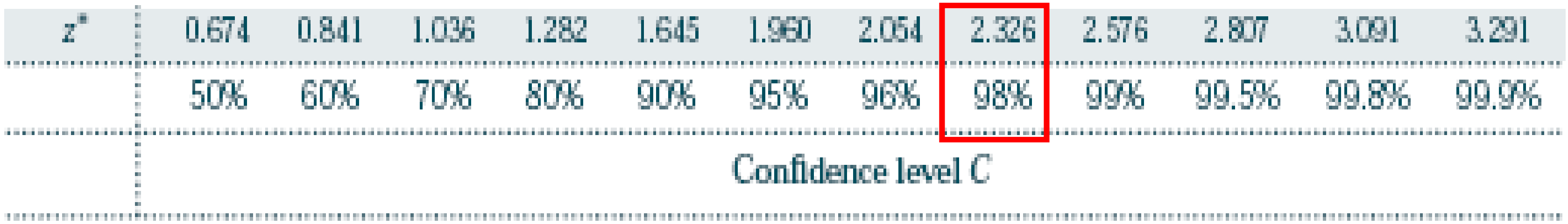
$$\bar{x} \pm z^* \sigma/\sqrt{n}$$



Example: For an 80% confidence level *C*, 80% of the normal curve's area is contained in the interval.

How do we find specific *z** values?

We can use a table of *z/t* values or software. For a particular confidence level, *C*, the appropriate *z** value is just above it.



z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Example: For a 98% confidence level, *z**=2.326

We can use software. In **R:**

qnorm(probability,mean,standard_dev)

gives z quantile for a given probability.

Since we want the middle *C* probability, the probability we need to input is (1 - C)/2

Example: For a 98% confidence level, qnorm(0.01,0,1) = −2.326348 (= neg. *z**)

Link between confidence level and margin of error

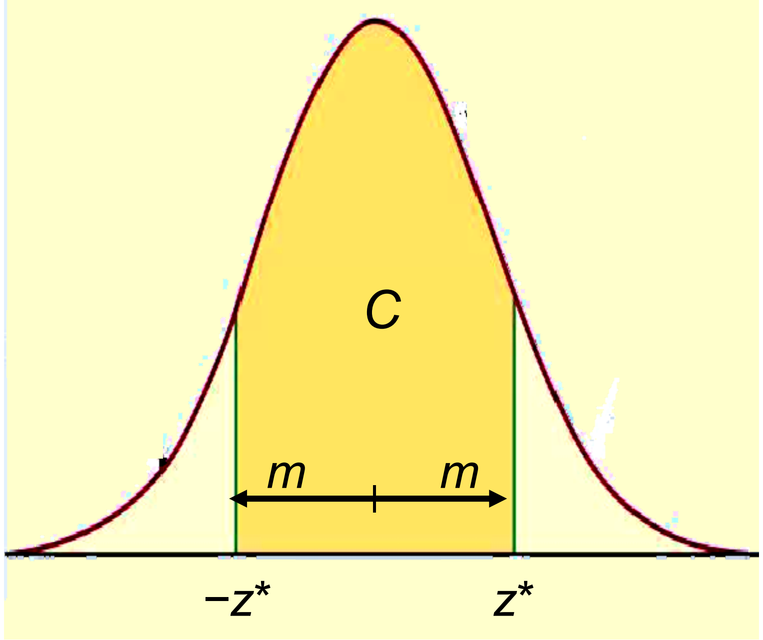
The confidence level *C* determines the value of *z** (in table C).

The margin of error also depends on *z**.

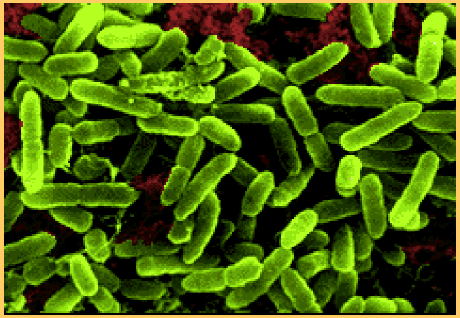
$$m = z^* \sigma / \sqrt{n}$$

Higher confidence ***C*** implies a larger margin of error ***m*** (thus less precision in our estimates).

A lower confidence level ***C*** produces a smaller margin of error ***m*** (thus better precision in our estimates).



Different confidence intervals for the same set of measurements



Density of bacteria in solution:

Measurement equipment has normal errors with deviation $\sigma = 1 * 10^6$ bacteria/ml fluid.

Three measurements: 24, 29, and 31 * 10^6 bacteria/ml fluid

Mean: $\bar{x} = 28 * 10^6$ bacteria/ml. Find the 96% and 70% CI.

- 96% confidence interval for the true density, $z^* = 2.054$, and write

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 28 \pm 2.054(1/\sqrt{3})$$

$$= 28 \pm 1.19 \times 10^6$$

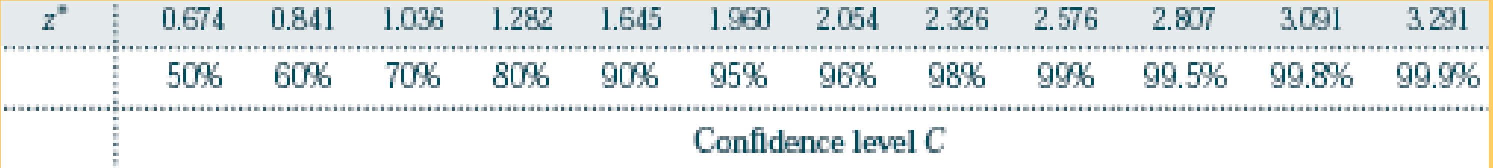
bacteria/ml

- 70% confidence interval for the true density, $z^* = 1.036$, and write

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 28 \pm 1.036(1/\sqrt{3})$$

$$= 28 \pm 0.60 \times 10^6$$

bacteria/ml

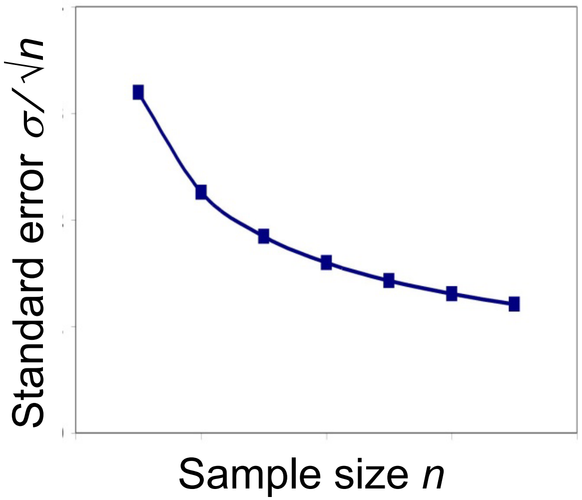


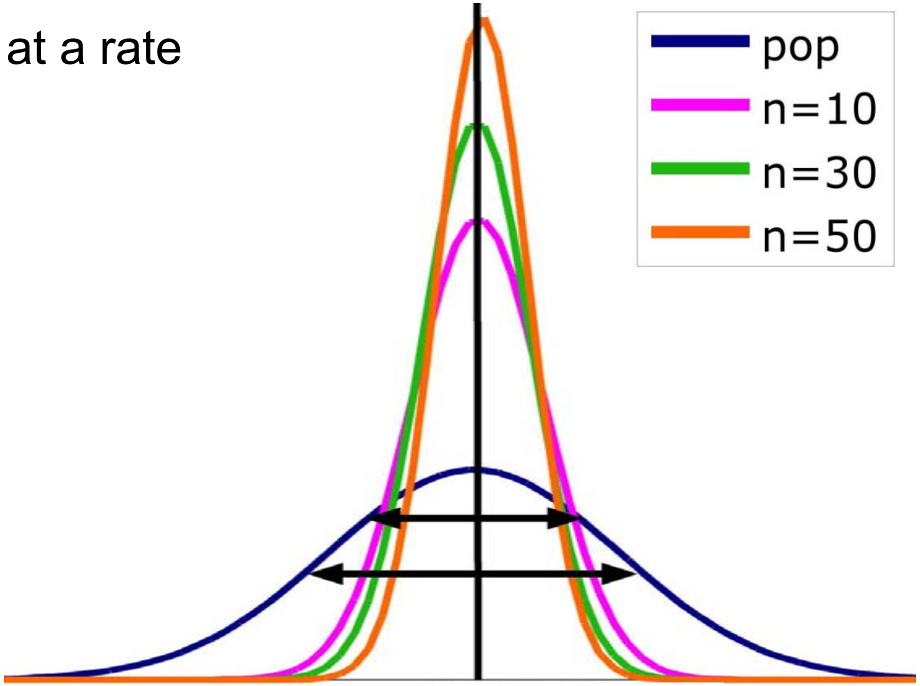
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Impact of sample size

The spread in the sampling distribution of the mean is a function of the number of individuals per sample.

- The larger the sample size, the smaller the standard deviation (spread) of the sample mean distribution.
- But the spread only decreases at a rate equal to √*n*.





Sample size and experimental design

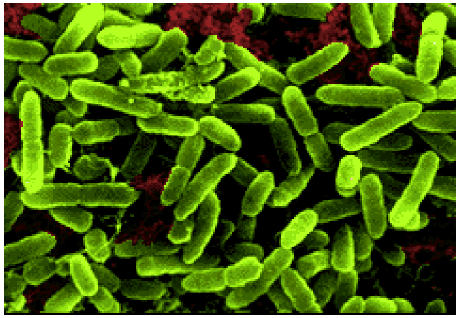
You may need a certain margin of error (e.g., drug trial, manufacturing specs). In many cases, the population variability (σ) is fixed, but we can choose the number of measurements (*n*).

So plan ahead what sample size to use to achieve that margin of error.

$$m = z^* \frac{\sigma}{\sqrt{n}} \quad \Leftrightarrow \quad n = \left(\frac{z^* \sigma}{m} \right)^2$$

Remember, though, that sample size is not always stretchable at will. There are typically costs and constraints associated with large samples. The best approach is to use the smallest sample size that can give you useful results.

What sample size for a given margin of error?



Density of bacteria in solution:

Measurement equipment has standard deviation

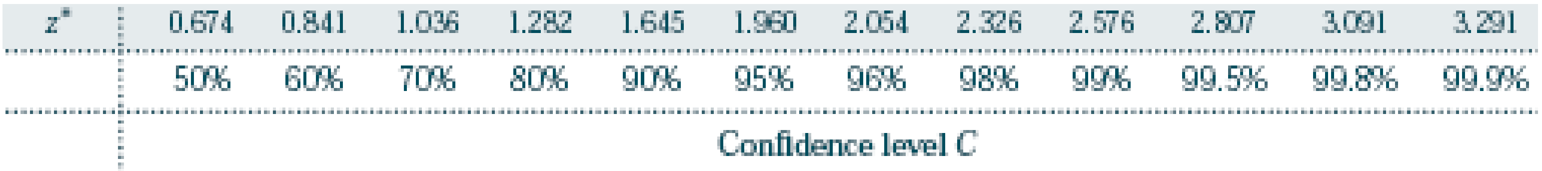
$\sigma = 1 * 10^6$ bacteria/ml fluid.

How many measurements should you make to obtain a margin of error of at most $0.5 * 10^6$ bacteria/ml with a confidence level of 90%?

For a 90% confidence interval, $z^* = 1.645$.

$$n = \left(\frac{z^* \sigma}{m}\right)^2 \Rightarrow n = \left(\frac{1.645 * 1}{0.5}\right)^2 = 3.29^2 = 10.8241$$

Using only 10 measurements will not be enough to ensure that *m* is no more than 0.5 * 106. Therefore, we need at least 11 measurements.



z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%

Confidence level C

Cautions:

- Data: a Simple Random Sample
- Formulas for other randomized designs available
- Haphazard data = unreliable conf. int.
- Population need not be normal and the outliers pose a threat to the validity of conclusions.
- We need to know σ the population variability. We will recall how to deal with the usually unknown σ later in this lecture.

Tests of Significance

- The scheme of reasoning
- Stating hypotheses
- Test statistics
- P-values
- Statistical significance
- Test for population mean
- Two-sided test and confidence intervals

We have seen that the properties of the sampling distribution of *x* bar help us estimate a range of likely values for population mean μ.

We can also rely on the properties of the sample distribution to test hypotheses.

Example: You are in charge of quality control in your food company. You sample randomly four packs of cherry tomatoes, each labeled 1/2 lb. (227 g).

The average weight from your four boxes is 222 g. Obviously, we cannot expect boxes filled with whole tomatoes to all weigh exactly half a pound. Thus,

- Is the somewhat smaller weight simply due to chance variation?
- Is it evidence that the calibrating machine that sorts cherry tomatoes into packs needs revision?

Null and alternative hypotheses

A **test of statistical significance** tests a specific hypothesis using sample data to decide on the validity of the hypothesis.

In statistics, a **hypothesis** is an assumption or a theory about the characteristics of one or more variables in one or more populations.

What you want to know: Does the calibrating machine that sorts cherry tomatoes into packs need revision?

The same question reframed statistically: Is the population mean μ for the distribution of weights of cherry tomato packages equal to 227 g (i.e., half a pound)?

The **null hypothesis** is a very specific statement about a parameter of the population(s). It is labeled H_0.

The **alternative hypothesis** is a more general statement about a parameter of the population(s) that is exclusive of the null hypothesis. It is labeled H_a.

Weight of cherry tomato packs:

$H_0 : \mu = 227$ g (μ is the average weight of the population of packs)

$H_a : \mu \neq 227$ g (μ is either larger or smaller)

One-sided and two-sided tests

- A **two-tail** or **two-sided test** of the population mean has these null and alternative hypotheses:

$$H_0:\ \mu = \text{[a specific number]} \quad H_a:\ \mu \neq \text{[a specific number]}$$

- A **one-tail** or **one-sided test** of a population mean has these null and alternative hypotheses:

$$H_0:\ \mu = \text{[a specific number]} \quad H_a:\ \mu < \text{[a specific number]} \qquad \text{OR}$$

$$H_0:\ \mu = \text{[a specific number]} \quad H_a:\ \mu > \text{[a specific number]}$$

The FDA tests whether a generic drug has an absorption extent similar to the known absorption extent of the brand-name drug it is copying. Higher or lower absorption would both be problematic, thus we test:

$$H_0 : \mu_{\text{generic}} = \mu_{\text{brand}} \qquad H_a : \mu_{\text{generic}} \neq \mu_{\text{brand}} \qquad \text{two-sided}$$

How to choose?

What determines the choice of a one-sided versus a two-sided test is what we know about the problem before we perform a test of statistical significance.

A health advocacy group tests whether the mean nicotine content of a brand of cigarettes is greater than the advertised value of 1.4 mg.

Here, the health advocacy group suspects that cigarette manufacturers sell cigarettes with a nicotine content higher than what they advertise in order to better addict consumers to their products and maintain revenues.

Thus, this is a one-sided test: $H_0 : \mu = 1.4$ mg $H_a : \mu > 1.4$ mg

It is important to make that choice before performing the test or else you could make a choice of “convenience” or fall in circular logic.

The P-value

The packaging process has a known standard deviation σ = 5 g.

$H_0 : \mu = 227$ g versus $H_a : \mu \neq 227$ g

The average weight from your four random boxes is 222 g.

What is the probability of drawing a random sample such as yours if H_0 is true?

Tests of statistical significance quantify the chance of obtaining a particular random sample result if the null hypothesis were true. This quantity is the **P-value.**

This is a way of assessing the "believability" of the null hypothesis given the evidence provided by a random sample.

Interpreting a P-value

Could random variation alone account for the difference between the null hypothesis and observations from a random sample?

- A small P-value implies that random variation because of the sampling process alone is not likely to account for the observed difference.

- With a small p-value we **reject H_0**. The true property of the population is **significantly** different from what was stated in H_0.

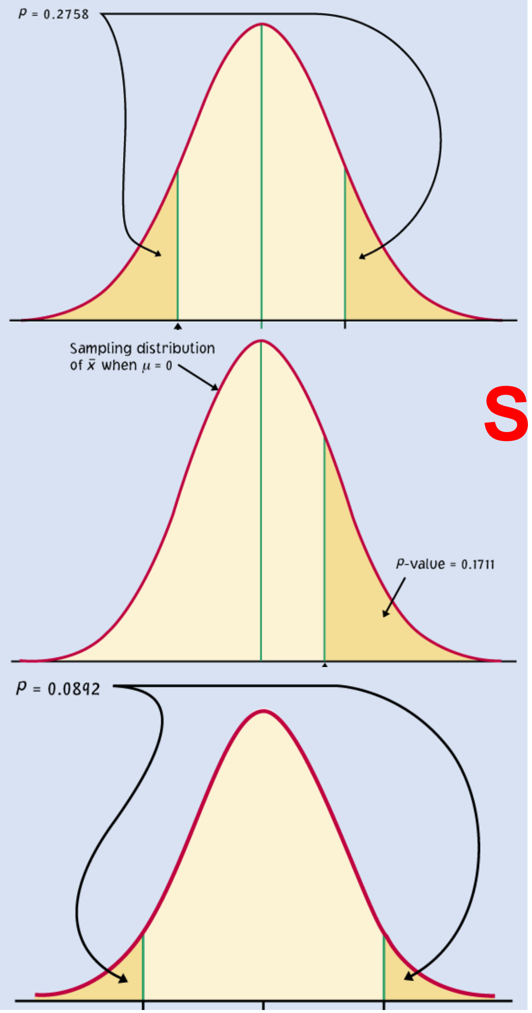
Thus, small P-values are strong evidence AGAINST H_0.

But how small is small…?

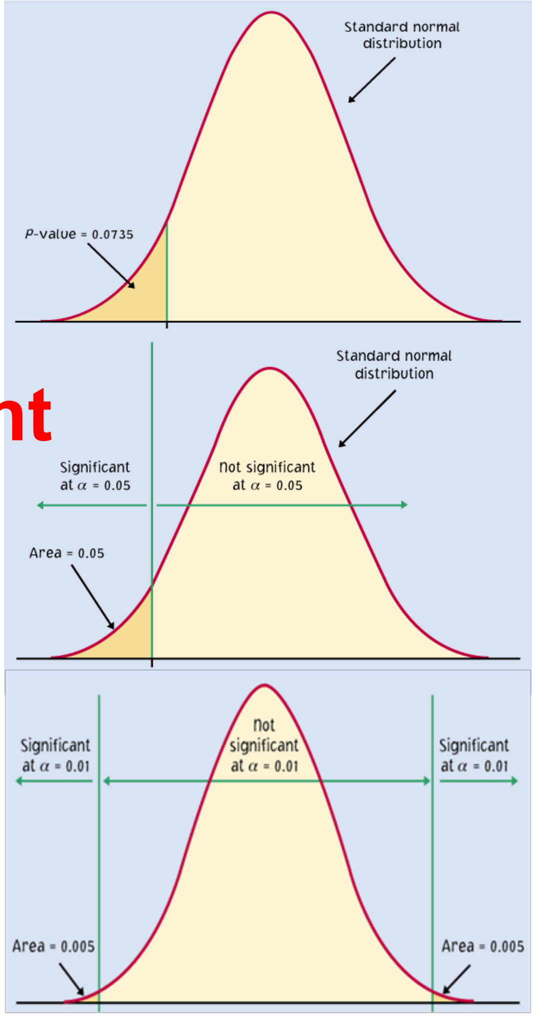
P = 0.2758

P = 0.1711

P = 0.0892



Significant P-value ???



P = 0.0735

P = 0.05

P = 0.01

When the shaded area becomes very small, the probability of drawing such a sample at random gets very slim. Oftentimes, a P-value of 0.05 or less is considered **significant:** The phenomenon observed is unlikely to be entirely due to chance event from the random sampling.

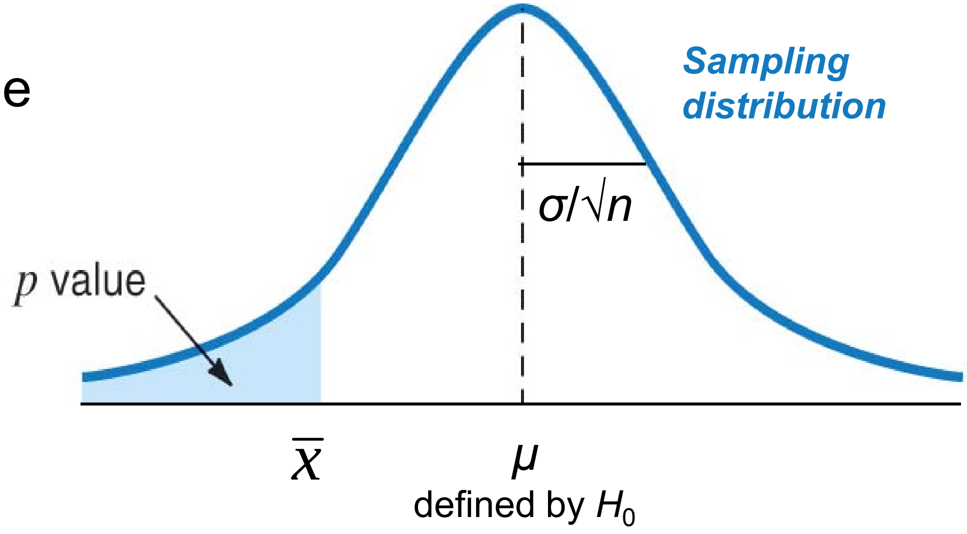
Tests for a population mean

To test the hypothesis $H_0 : \mu = \mu_0$ based on an SRS of size n from a Normal population with unknown mean μ and known standard deviation σ, we rely on the properties of the sampling distribution $N(\mu, \sigma\sqrt{n})$.

The P-value is the area under the sampling distribution for values at least as extreme, in the direction of H_a, as that of our random sample.

Again, we first calculate a z-value and then use Table A.

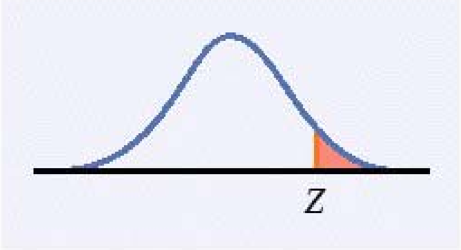
$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$



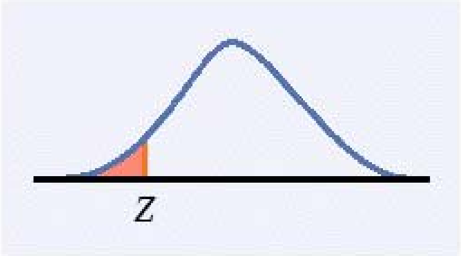
P-value in one-sided and two-sided tests

One-sided
(one-tailed) test

$H_a: \mu > \mu_0$ is $P(Z \geq z)$

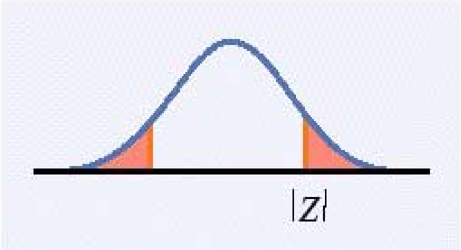


$H_a: \mu < \mu_0$ is $P(Z \leq z)$



Two-sided
(two-tailed) test

$H_a: \mu \neq \mu_0$ is $2P(Z \geq |z|)$



To calculate the P-value for a two-sided test, use the symmetry of the normal curve. Find the P-value for a one-sided test, and double it.

Does the packaging machine need revision?

- $H_0 : \mu = 227$ g versus $H_a : \mu \neq 227$ g
- What is the probability of drawing a random sample such as yours if H_0 is true?

$$\bar{x} = 222\text{g} \qquad \sigma = 5\text{g} \qquad n = 4$$

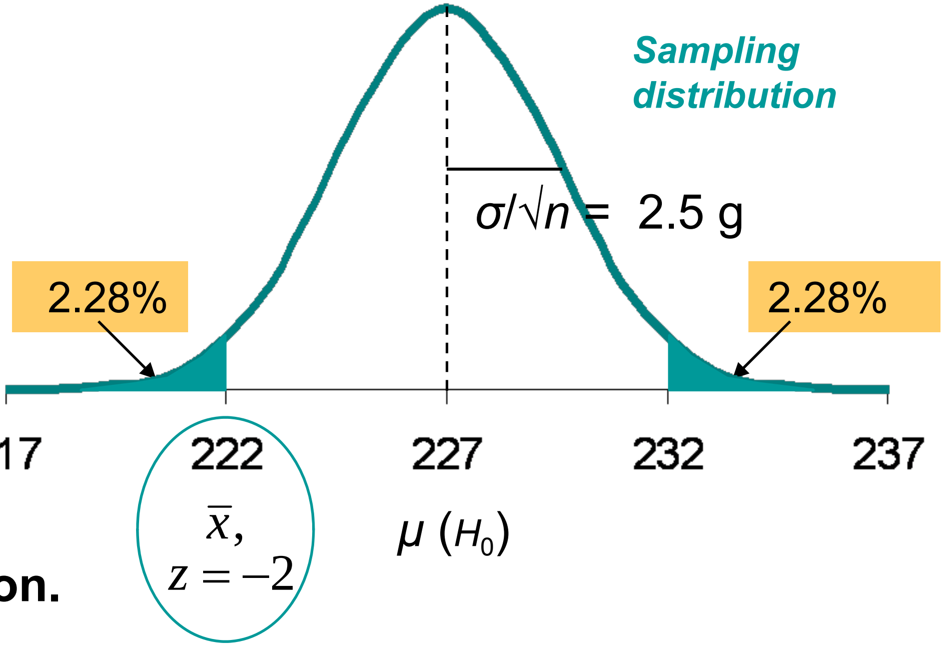
$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{222 - 227}{5 / \sqrt{4}} = -2$$

From table A, the area under the standard normal curve to the left of z is 0.0228.

Thus, P-value = 2*0.0228 = 4.56%.

The probability of getting a random sample average so different from μ is so low that we reject H_0.

➔**The machine does need recalibration.**



The significance level α

The significance level, α, is the largest P-value tolerated for rejecting a true null hypothesis (how much evidence against H_0 we require). This value is decided arbitrarily <u>before</u> conducting the test.

- If the P-value is equal to or less than α (**$P \leq \alpha$**), then we **reject H_0**.
- If the P-value is greater than α (**$P > \alpha$**), then we **fail to reject H_0**.

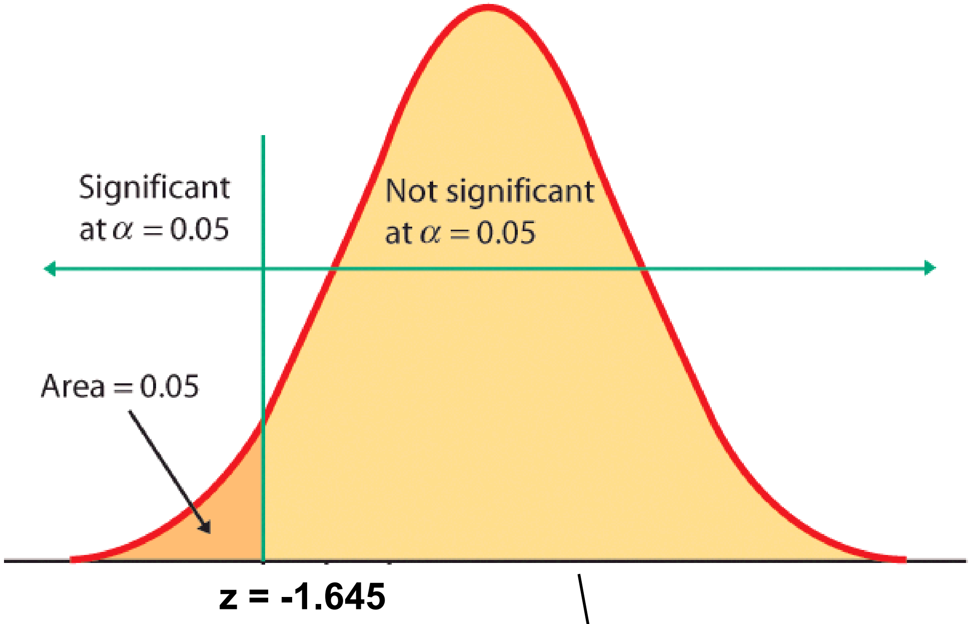
Does the packaging machine need revision?

Two-sided test. The P-value is 4.56%.

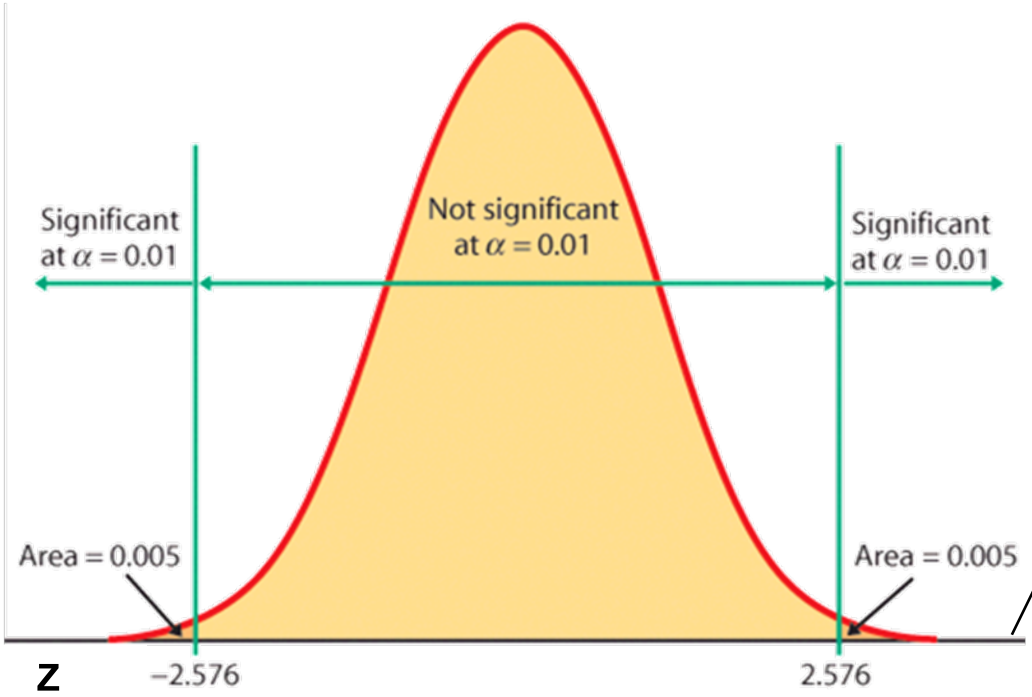
* If α had been set to 5%, then the P-value would be significant.

* If α had been set to 1%, then the P-value would <u>not</u> be significant.

When the z score falls within the rejection region (shaded area on the tail-side), the p-value is smaller than α and you have shown statistical significance.



One-sided test, α = 5%



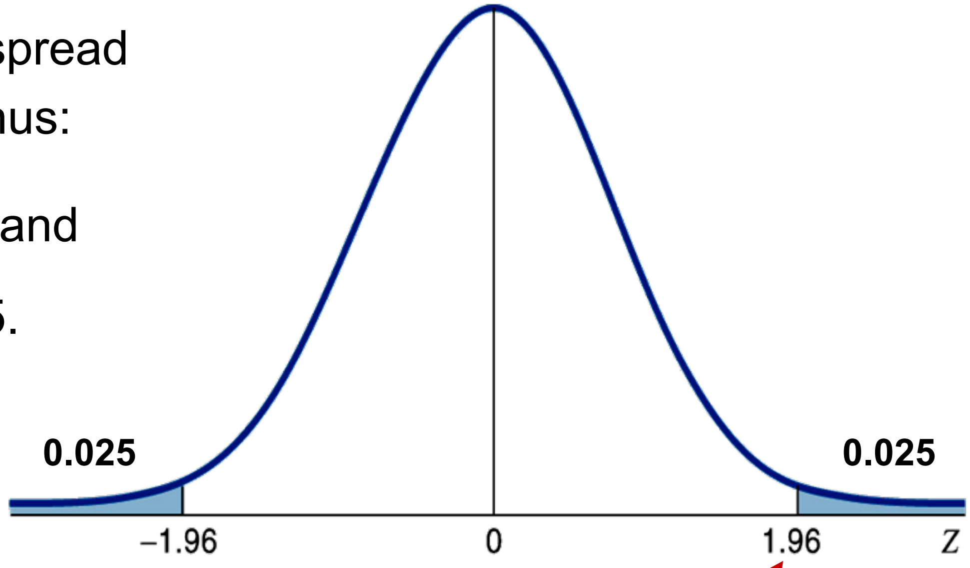
Two-sided test, α = 1%

Rejection region for a two-tail test of μ with $\alpha = 0.05$ (5%)

A two-sided test means that α is spread between both tails of the curve, thus:

-A middle area C of $1 - \alpha = 95\%$, and

-An upper tail area of $\alpha/2 = 0.025$.



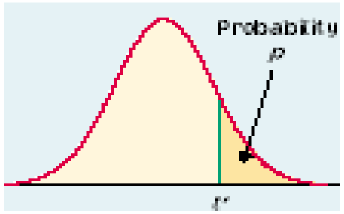


Table C

upper tail probability p	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
(…)												
z*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
Confidence interval C	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%

Confidence intervals to test hypotheses

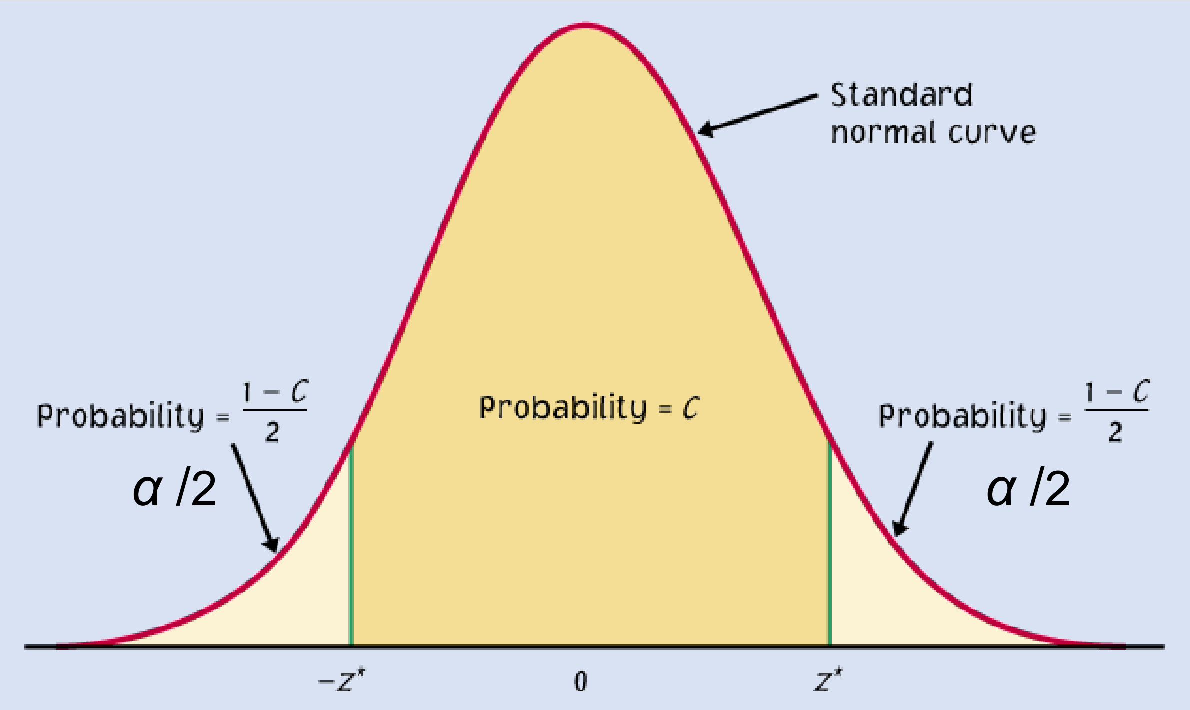
Because a two-sided test is symmetrical, you can also use a confidence interval to test a two-sided hypothesis.

In a two-sided test,

$C = 1 - \alpha$.

C confidence level

α significance level



Packs of cherry tomatoes ($\sigma = 5$ g): $H_0 : \mu = 227$ g versus $H_a : \mu \neq 227$ g

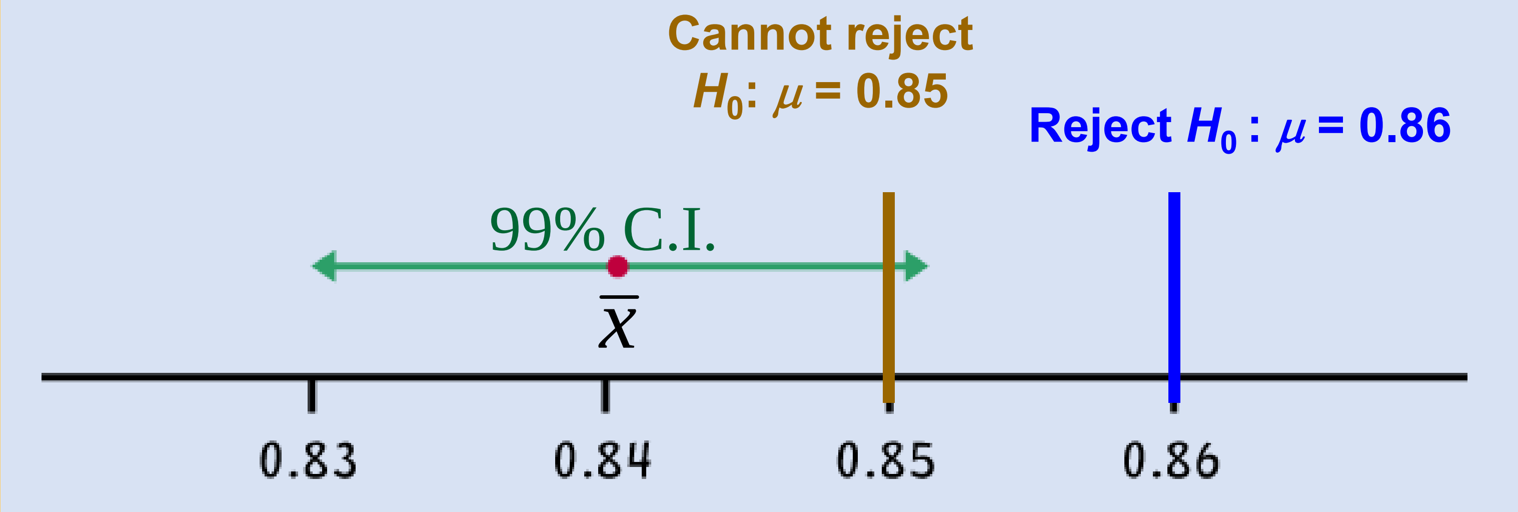
Sample average 222 g. 95% CI for $\mu = 222 \pm 1.96 * 5/\sqrt{4} = 222$ g $\pm$ 4.9 g

227 g does not belong to the 95% CI (217.1 to 226.9 g). Thus, we reject H_0.

Logic of confidence interval test

Ex: Your sample gives a 99% confidence interval of $\bar{x} \pm m = 0.84 \pm 0.0101$.

With 99% confidence, could samples be from populations with $\mu = 0.86$? $\mu = 0.85$?



A confidence interval gives a black and white answer: Reject or don't reject H_0.

But it also estimates a range of likely values for the true population mean μ.

A P-value quantifies how strong the evidence is against the H_0. But if you reject H_0, it doesn’t provide any information about the true population mean μ.

Section 6.3: Use and abuse of tests

- Choosing the level of significance
- Significance vs. practical significance
- Lack of significance may be informative
- Dangers of searching for significance
- Assumptions about the data

Caution about significance tests

Choosing the significance level *α*

- α=0.05 is accepted standard, but…
  - if the conclusion that H_a is true has "costly" implications, smaller α may be appropriate

e.g.,

- What are the consequences of rejecting the null hypothesis (e.g., global warming, convicting a person for life with DNA evidence)?
- Are you conducting a preliminary study? If so, you may want a larger *α* so that you will be less likely to miss an interesting result.

Some conventions:

- We typically use the standards of our field of work.
- There are no "sharp" cutoffs: e.g., 4.9% versus 5.1 %. Oftentimes, describing the evidence using the P-value itself may be enough
- It is the order of magnitude of the P-value that matters: "somewhat significant," "significant," or "very significant."

Practical significance

Statistical significance only says whether the effect observed is likely to be due to chance alone because of random sampling.

Statistical significance may not be practically important. That's because statistical significance doesn't tell you about the **magnitude** of the effect, only that there is one.

An effect could be too small to be relevant. And with a large enough sample size, significance can be reached even for the tiniest effect.

- A drug to lower temperature is found to reproducibly lower patient temperature by 0.4°Celsius (P-value < 0.01). But clinical benefits of temperature reduction only appear for a 1° decrease or larger.

Interpreting lack of significance

- Consider this provocative title from the British Medical Journal: “Absence of evidence is not evidence of absence”.
- Having no proof of whom committed a murder does not imply that the murder was not committed.

Indeed, failing to find statistical significance in results means that we do not reject the null hypothesis. This is very different from actually accepting it. The sample size, for instance, could be too small to overcome large variability in the population.

When comparing two populations, lack of significance does not imply that the two samples come from the same population. They could represent two very distinct populations with the similar mathematical properties.

Interpreting effect size: It’s all about context

There is no consensus on how big an effect has to be in order to be considered meaningful. In some cases, effects that may appear to be trivial can in reality be very important.

- Example: Improving the format of a computerized test reduces the average response time by about 2 seconds. Although this effect is small, it is important since this is done millions of times a year. The *cumulative* time savings of using the better format is gigantic.

Always think about the context. Try to plot your results, and compare them with a baseline or results from similar studies.

Type I and II errors

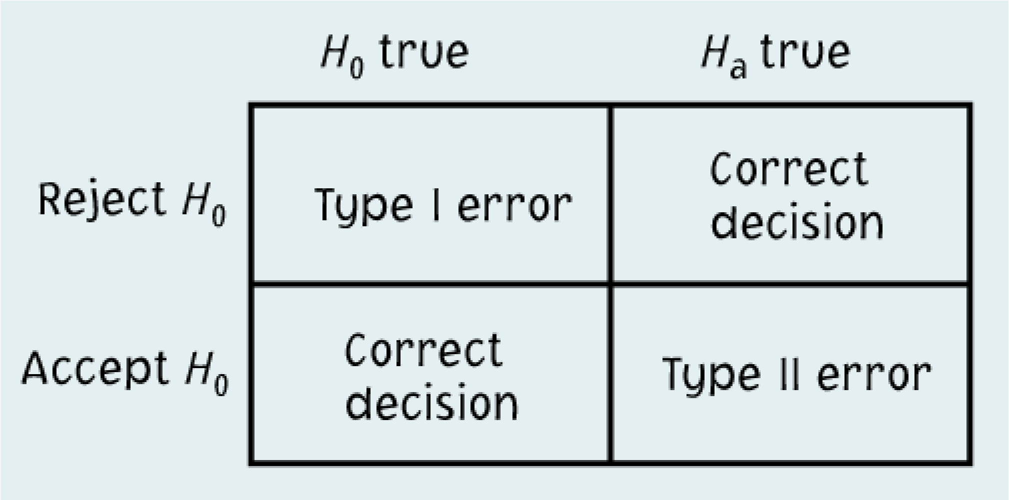
- A **Type I error** is made when we reject the null hypothesis and the null hypothesis is actually true (incorrectly reject a true H_0).

  The probability of making a Type I error is the significance level α

- A **Type II error** is made when we fail to reject the null hypothesis and the null hypothesis is false (incorrectly keep a false H_0).

  The probability of making a Type II error is labeled β.

Running a test of significance is a balancing act between the chance α of making a **Type I error** and the chance β of making a **Type II error.** Reducing α increases β. And vice versa.



	H_0 true	H_a true
Reject H_0	Type I error	Correct decision
Accept H_0	Correct decision	Type II error

A type II error is not definitive since a failure to reject the null hypothesis does not imply that the null hypothesis is wrong.

Inference for the mean of a population

Change: Population s.d. sigma unknown.

- The ***t*** distribution
- One-sample ***t*** confidence interval
- One-sample ***t*** test
- **Matched pairs *t*** procedures
- Robustness of ***t*** procedures

Sweetening colas

Cola manufacturers want to test how much the sweetness of a new cola drink is affected by storage. The sweetness loss due to storage was evaluated by 10 professional tasters (by comparing the sweetness before and after storage):

Taster	Sweetness loss
1	2.0
2	0.4
3	0.7
4	2.0
5	−0.4
6	2.2
7	−1.3
8	1.2
9	1.1
10	2.3

Obviously, we want to test if storage results in a loss of sweetness, thus:

H_0: $\mu = 0$ versus H_a: $\mu > 0$

This looks familiar. However, here we do not know the population parameter σ.

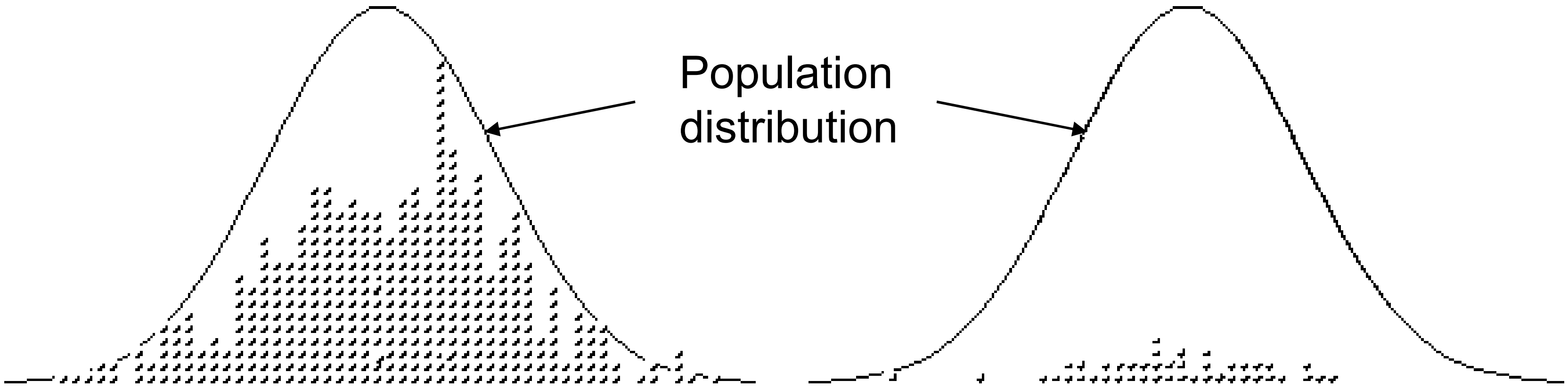
- The population of all cola drinkers is too large.
- Since this is a new cola recipe, we have no population data.

This situation is very common with real data.

When σ is unknown

The sample standard deviation *s* provides an estimate of the population standard deviation σ.

- When the sample size is large, the sample is likely to contain elements representative of the whole population. Then *s* is a good estimate of σ.
- But when the sample size is small, the sample contains only a few individuals. Then *s* is a more mediocre estimate of σ.



Large sample

Small sample

Standard deviation s – standard error $s/\sqrt{n}$

For a sample of size *n*,

the sample standard deviation *s* is:

$$s = \sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2}$$

n − 1 is the "degrees of freedom."

The value ***s*/√*n*** is called the standard error of the sample mean or simply standard error of the mean (**SEM**).

Scientists often present sample results as mean ± SEM.

A study examined the effect of a new medication on the seated systolic blood pressure. The results, presented as mean ± SEM for 25 patients, are 113.5 ± 8.9.

What is the standard deviation *s* of the sample data?

$$\text{SEM} = s/\sqrt{n} \iff s = \text{SEM}*\sqrt{n}$$

$$s = 8.9*\sqrt{25} = 44.5$$

The *t* distribution:

The goal is to estimate or test for unknown ***μ*** in situation when ***σ*** is also unknown .

Solution: estimate ***σ*** by ***s*** and use intelligently in formulas.

Challenge: the distribution of the test statistic will change and will no longer be z-distribution.

Suppose that an SRS of size *n* is drawn from an *N*(μ, σ) population.

- When σ is known, the sampling distribution is *N*(μ, $\sigma/\sqrt{n}$).

- When σ is estimated from the sample standard deviation *s*, the sampling distribution

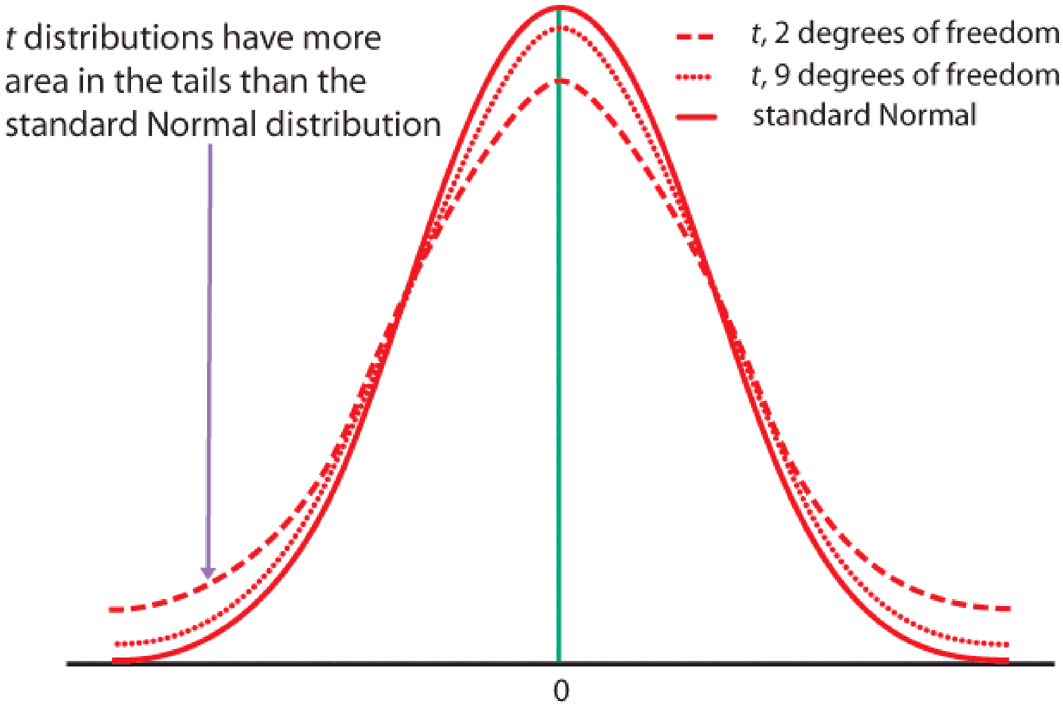
follows a ***t* distribution *t*(μ, $s/\sqrt{n}$) with degrees of freedom *n* − 1.**

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

is the **one-sample *t* statistic.**

When *n* is very large, *s* is a very good estimate of σ and the corresponding *t* distributions are very close to the normal distribution.

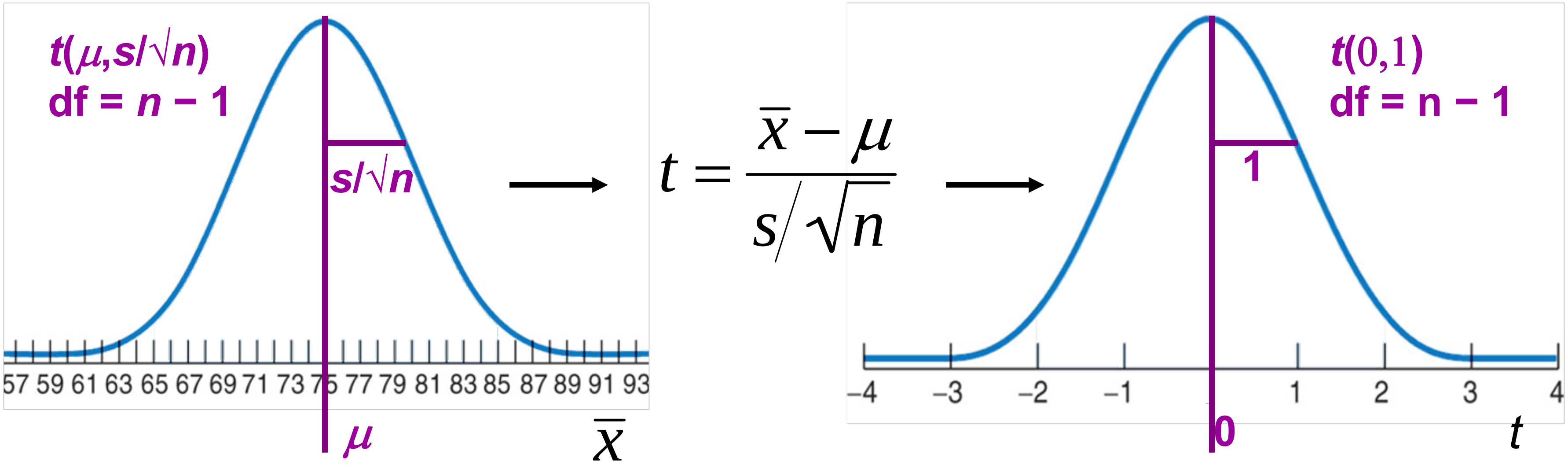
The *t* distributions become wider for smaller sample sizes, reflecting the lack of precision in estimating σ from *s*.



Standardizing the data before using Table A.5

As with the normal distribution, the first step is to standardize the data.

Then we can use **Table A.5** to obtain the area under the curve.



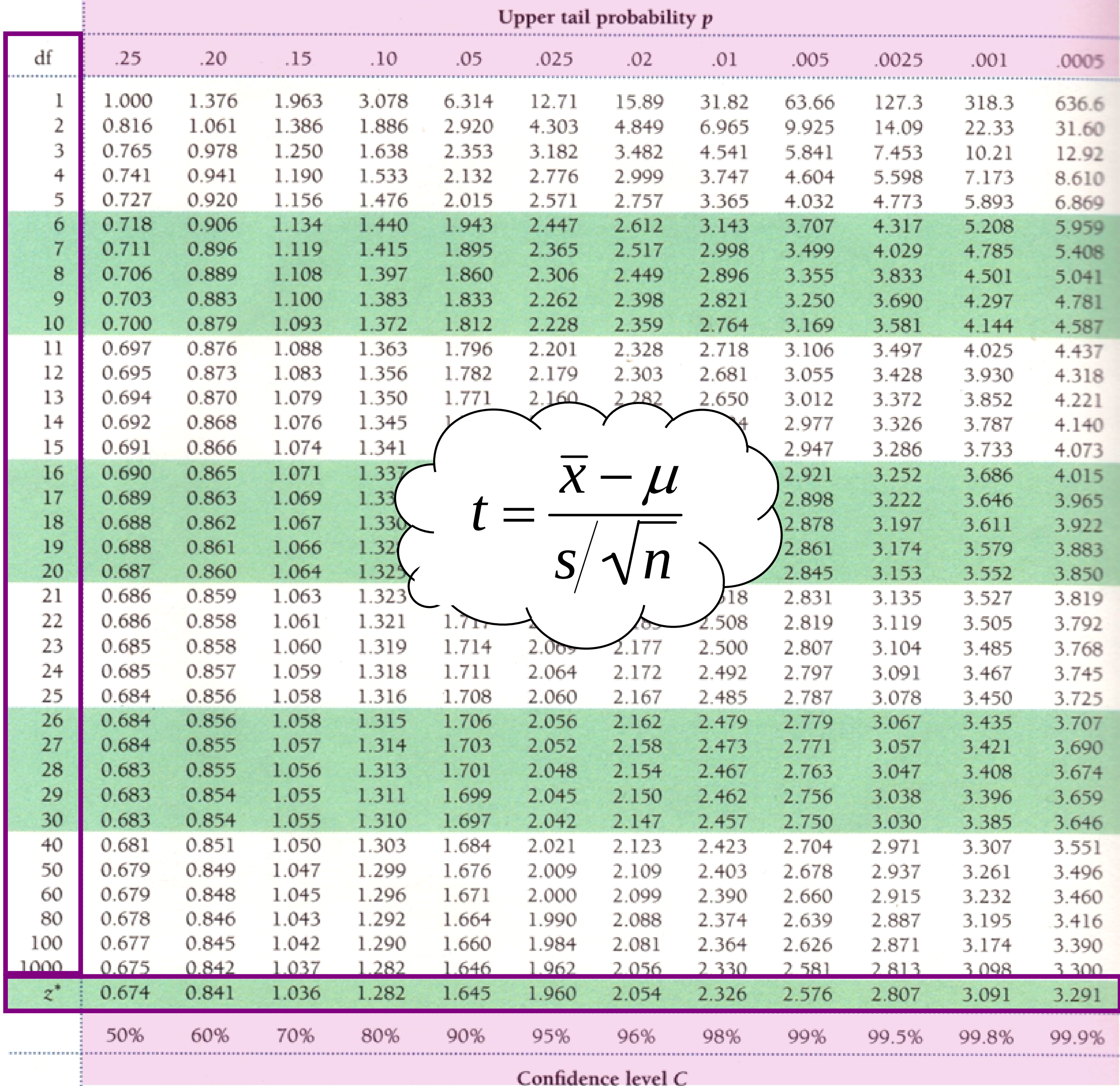
$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Here, μ is the mean (center) of the sampling distribution,
and the standard error of the mean $s/\sqrt{n}$ is its standard deviation (width).
You obtain s, the standard deviation of the sample, with your calculator.

Table A.5

When σ is unknown, we use a *t* distribution with "*n*−1" degrees of freedom (df).

Table A.5 shows the *z*-values and *t*-values corresponding to landmark P-values/ confidence levels.



Upper tail probability *p*

df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	[illegible]	[illegible]	[illegible]	[illegible]	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	[illegible]	[illegible]	[illegible]	[illegible]	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	[illegible]	[illegible]	[illegible]	[illegible]	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	[illegible]	[illegible]	[illegible]	[illegible]	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	[illegible]	[illegible]	[illegible]	[illegible]	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	[illegible]	[illegible]	[illegible]	[illegible]	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	[illegible]	[illegible]	[illegible]	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%

Confidence level *C*

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

When σ is known, we use the normal distribution and the standardized z-value.

Table A.3 vs. Table A.5

Table A.3 gives the area to the LEFT of hundreds of *z*-values.

It should only be used for Normal distributions.

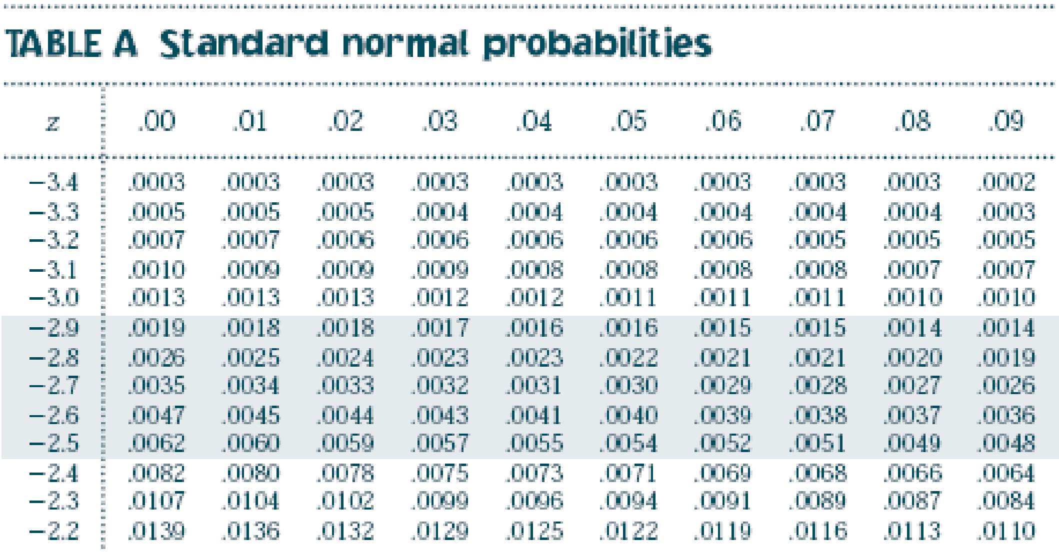
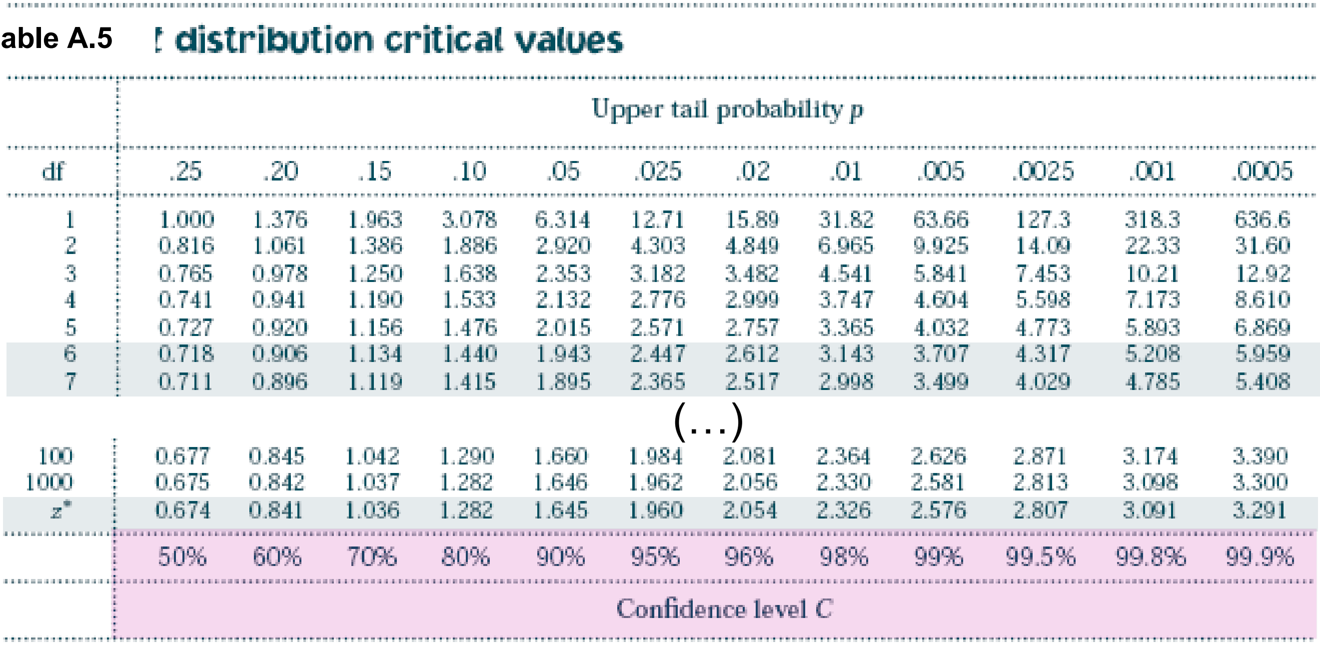


TABLE A Standard normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110

(…)

Table A.5



t distribution critical values

	Upper tail probability *p*											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
(…)												
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level *C*											

Table A.5 gives the area to the RIGHT of a dozen *t* or *z*-values.

It can be used for *t* distributions of a given df, and for the Normal distribution.

Table A.5 also gives the middle area under a *t* or normal distribution comprised between the negative and positive value of *t* or *z*.

The one-sample *t*-confidence interval

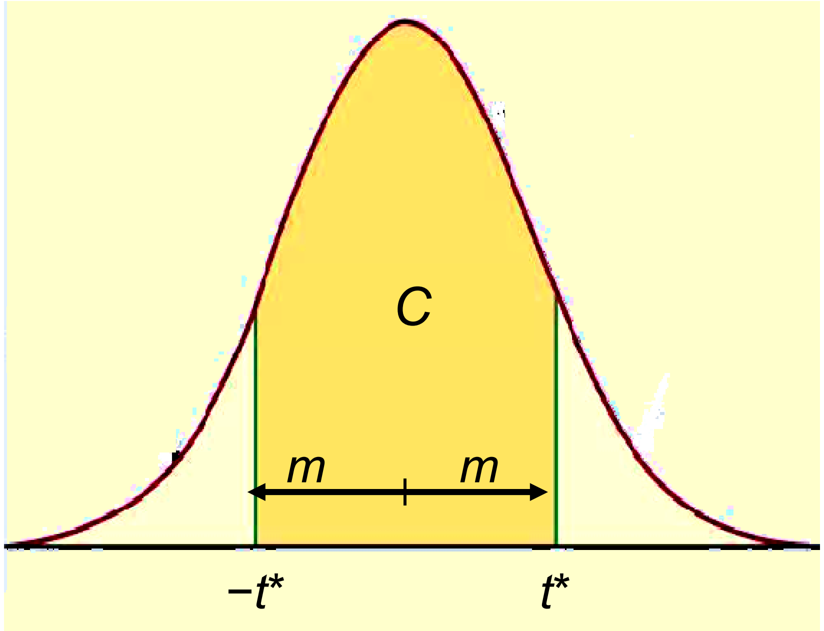
The **level *C* confidence interval** is an interval with probability *C* of containing the true population parameter.

We have a data set from a population with both μ and σ unknown. We use $\bar{x}$ to estimate μ, and *s* to estimate σ, using a *t* distribution (df n−1).

Practical use of *t* : *t**

- *C* is the area between −*t** and *t**.
- We find *t** in the line of Table D for df = n−1 and confidence level *C*.
- The margin of error *m* is:

$$\boxed{m = t^* s / \sqrt{n}}$$

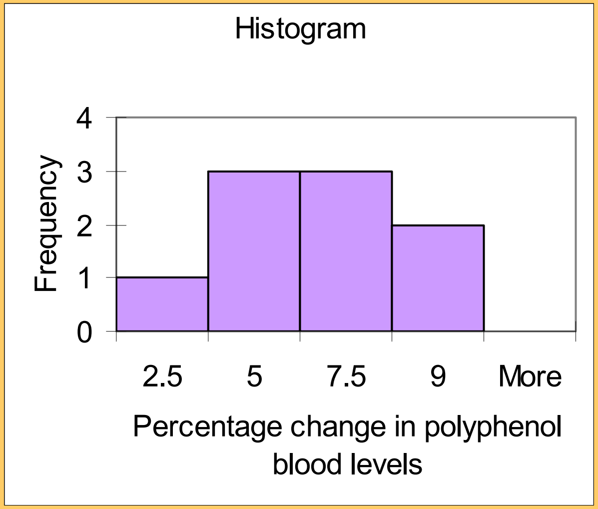


Red wine, in moderation

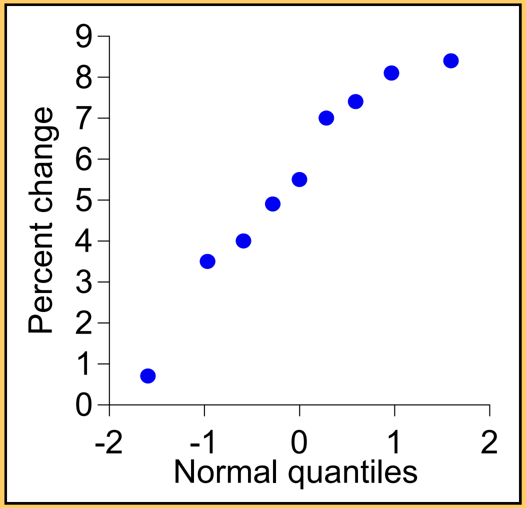
Drinking red wine in moderation may protect against heart attacks. The polyphenols it contains act on blood cholesterol and thus are a likely cause.

To see if moderate red wine consumption increases the average blood level of polyphenols, a group of nine randomly selected healthy men were assigned to drink half a bottle of red wine daily for two weeks. Their blood polyphenol levels were assessed before and after the study, and the percent change is presented here: **0.7 3.5 4 4.9 5.5 7 7.4 8.1 8.4**

Firstly: Are the data approximately normal?



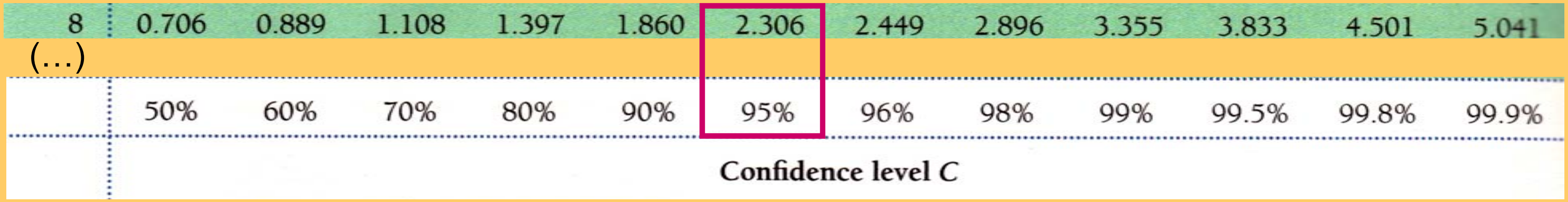
0	7
1	
2	
3	5
4	09
5	5
6	
7	04
8	14



There is a low value, but overall the data can be considered reasonably normal.

What is the 95% confidence interval for the average percent change?

Sample average = 5.5; $s = 2.517$; df = $n - 1 = 8$



8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
(…)												
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

The sampling distribution is a t distribution with $n - 1$ degrees of freedom.

For df = 8 and C = 95%, $t^* = 2.306$.

The margin of error m is: $m = t^* s/\sqrt{n} = 2.306 * 2.517/\sqrt{9} \approx 1.93$.

With 95% confidence, the population average percent increase in polyphenol blood levels of healthy men drinking half a bottle of red wine daily is between 3.6% and 7.6%. *Important: The confidence interval shows how large the increase is, but not if it can have an impact on men's health.*

The one-sample *t*-test

As in the previous chapter, a test of hypotheses requires a few steps:

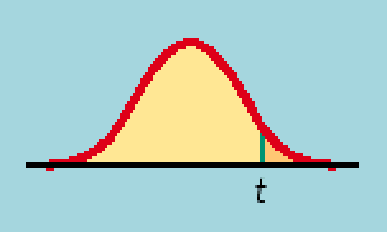
1. Stating the null and alternative hypotheses (H_0 versus H_a)
2. Deciding on a one-sided or two-sided test
3. Choosing a significance level α
4. Calculating t and its degrees of freedom
5. Finding the area under the curve *with Table D*
6. Stating the P-value and interpreting the result

The ***P-value*** is the probability, if H_0 is true, of randomly drawing a sample like the one obtained or more extreme, in the direction of H_a.

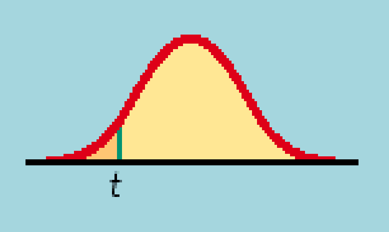
The P-value is calculated as the corresponding area under the curve, one-tailed or two-tailed depending on H_a:

One-sided (one-tailed)

$H_a: \mu > \mu_0 \Rightarrow P(T \geq t)$



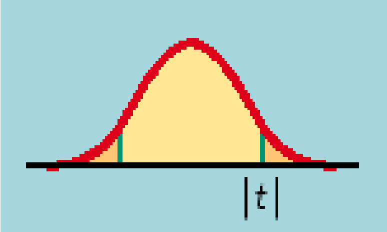
$H_a: \mu < \mu_0 \Rightarrow P(T \leq t)$



$$\boxed{t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}}$$

Two-sided (two-tailed)

$H_a: \mu \neq \mu_0 \Rightarrow 2P(T \geq |t|)$



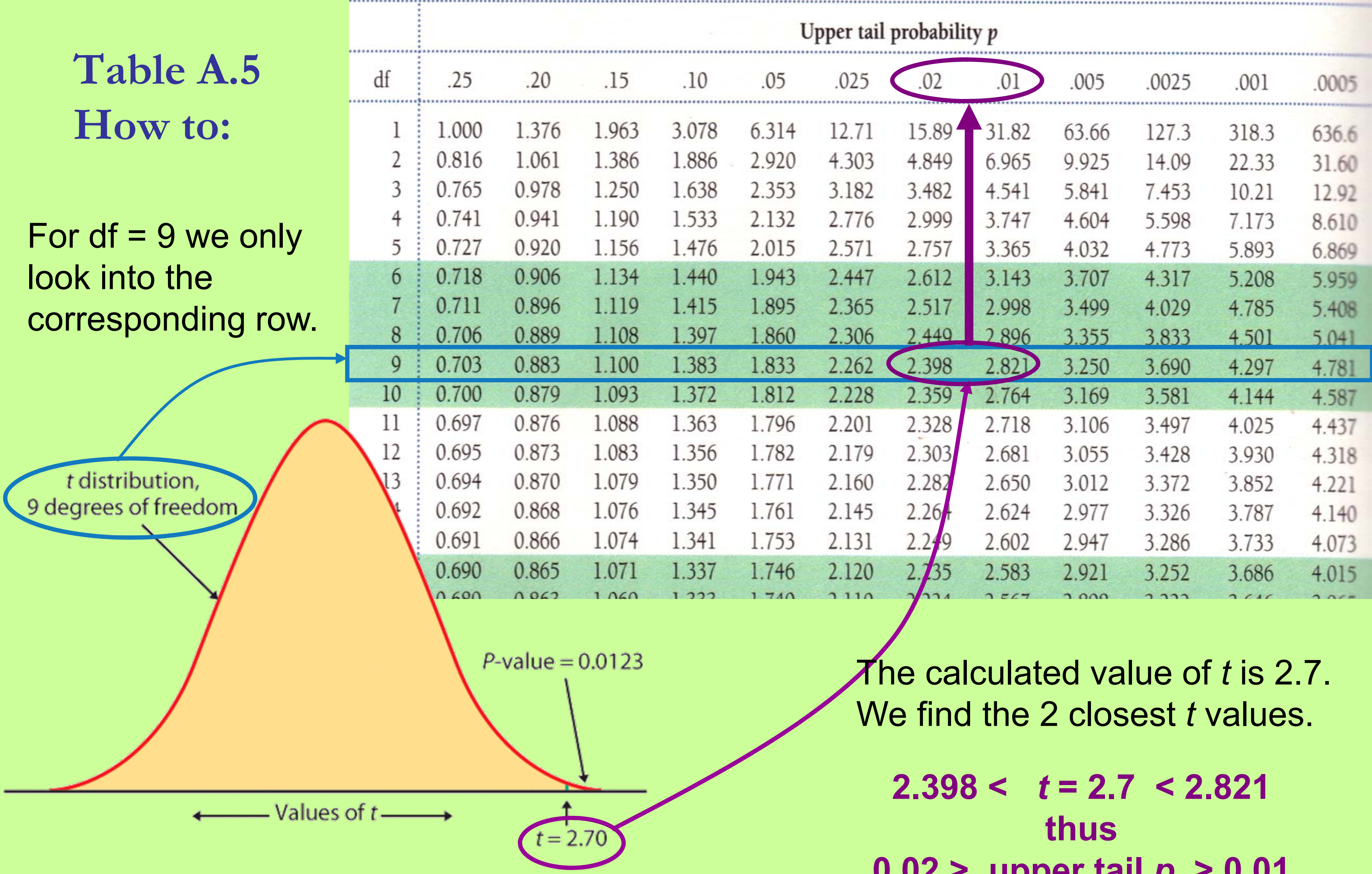


Table A.5
How to:

For df = 9 we only look into the corresponding row.

df	Upper tail probability p											
	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
4	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015

The calculated value of *t* is 2.7. We find the 2 closest *t* values.

2.398 < *t* = 2.7 < 2.821
thus
0.02 > upper tail *p* > 0.01

For a one-sided H_a, this is the P-value (between 0.01 and 0.02); for a two-sided H_a, the P-value is doubled (between 0.02 and 0.04).

- These are just approximate values.
- To find exact p-values use R
- The function to be used is:
  - pt(quantile, df)

Sweetening colas (*continued*)

Is there evidence that storage results in sweetness loss for the new cola recipe at the 0.05 level of significance (α = 5%)?

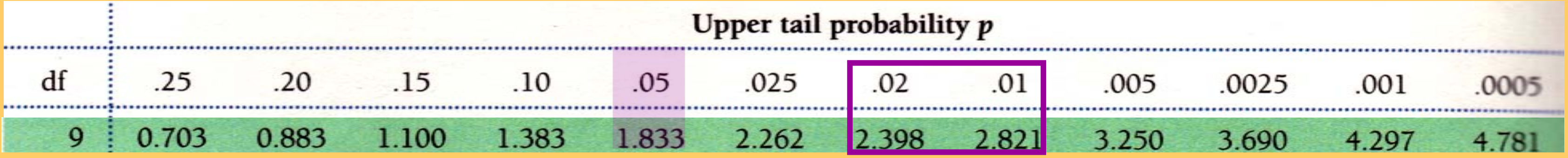
H_0: $\mu = 0$ versus H_a: $\mu > 0$ (one-sided test)

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{1.02 - 0}{1.196/\sqrt{10}} = 2.70$$

- The critical value t_α = 1.833.
  $t > t_\alpha$ thus the result is significant.

- $2.398 < t = 2.70 < 2.821$ thus $0.02 > p > 0.01$.
  $p < \alpha$ thus the result is significant.

Taster	**Sweetness loss**
1	2.0
2	0.4
3	0.7
4	2.0
5	-0.4
6	2.2
7	-1.3
8	1.2
9	1.1
10	2.3
Average	**1.02**
Standard deviation	**1.196**
Degrees of freedom	***n* − 1 = 9**

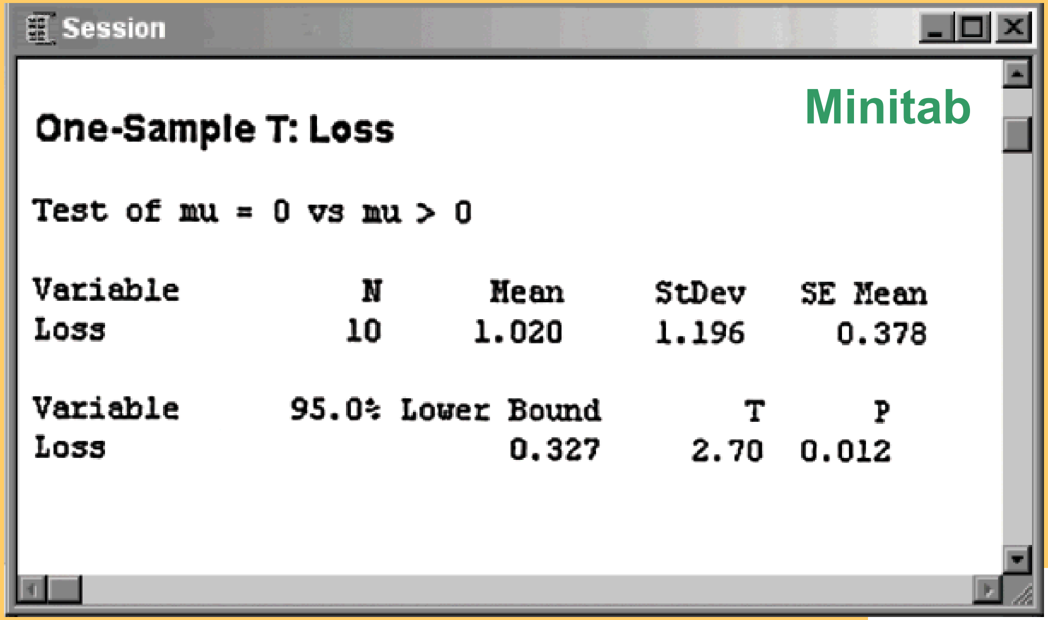


	Upper tail probability *p*											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781

The *t*-test has a significant p-value. We reject H_0.

There is a significant loss of sweetness, on average, following storage.

Sweetening colas (*continued*)



Minitab

```
One-Sample T: Loss

Test of mu = 0 vs mu > 0

Variable          N      Mean     StDev    SE Mean
Loss             10     1.020     1.196      0.378

Variable      95.0% Lower Bound        T        P
Loss                      0.327     2.70    0.012
```

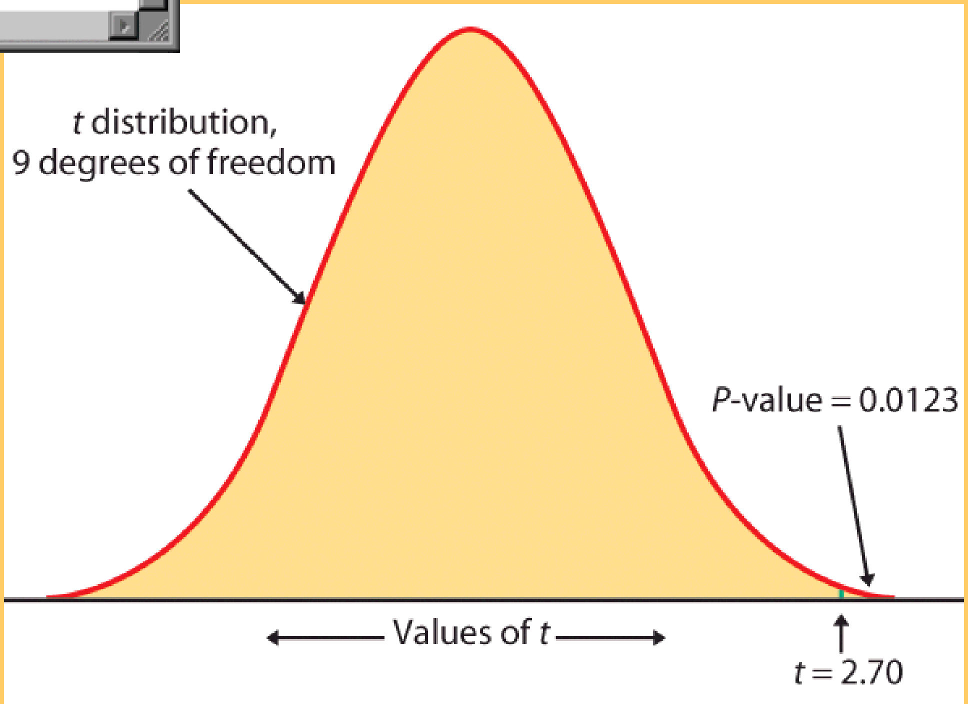
$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{1.02 - 0}{1.196/\sqrt{10}} = 2.70$$

$$df = n - 1 = 9$$

In R, you can obtain the precise p-value once you have calculated *t*:

Using the function pt(2.7, 9)

which gives 0.9878032 and taking 1- this value (WHY?) we obtain 0.01219685



Robustness

The *t* procedures are exactly correct when the population is distributed exactly normally. However, most real data are not exactly normal.

The *t* procedures are **robust** to small deviations from normality – the results will not be affected too much. Factors that strongly matter:

- **Random sampling.** The sample **must** be an SRS from the population.
- **Outliers and skewness.** They strongly influence the mean and therefore the *t* procedures. However, their impact diminishes as the sample size gets larger because of the Central Limit Theorem.

Specifically:

- When $n < 15$, the data must be close to normal and without outliers.
- When $15 > n > 40$, mild skewness is acceptable but not outliers.
- When $n > 40$, the t-statistic will be valid even with strong skewness.