Chapter 1

Confidence Intervals

The purpose of sampling or experimentation is to make *statistical inference* about the population. If the estimates of the parameters are given in the form of intervals, those intervals are called *confidence intervals*.

1.1 Basics of confidence intervals

The following example throughout this section.

To estimate the average height of all the high school students μ, we draw an SRS with $\overline{x} = 170$cm. What can we say about the parameter μ?

You can say that $\mu = 170$cm. In this case, the estimate only gives one value(170cm) for μ, this value is called **point estimate** and the statistic $\overline{x}$ is called a **point estimator**.

$\overline{x}$ varies for different samples and it is hardly true that the real value of μ is 170cm. In order to take the chance variance of the the sampling distribution into account, an interval can be constructed to estimate the parameter . This interval is called the **confidence interval**.

- **The idea of confidence interval**

  The confidence interval is constructed by the following formula:

  $$\textbf{statistic} \pm \textbf{margin of error}$$

  If the *margin of error* is given, for each sample, we can construct a confidence interval by invoking the above formula.

  If $\overline{\mathbf{x}} = 170$cm, the confidence interval for μ is $170\text{cm} \pm \textbf{margin of error}$
  If $\overline{\mathbf{x}} = 165$cm, the confidence interval for μ is $165\text{cm} \pm \textbf{margin of error}$

  . . . .

  Figure1.1 gives the sampling distribution of $\overline{x}$ and the confidence intervals constructed on basis of different samples.

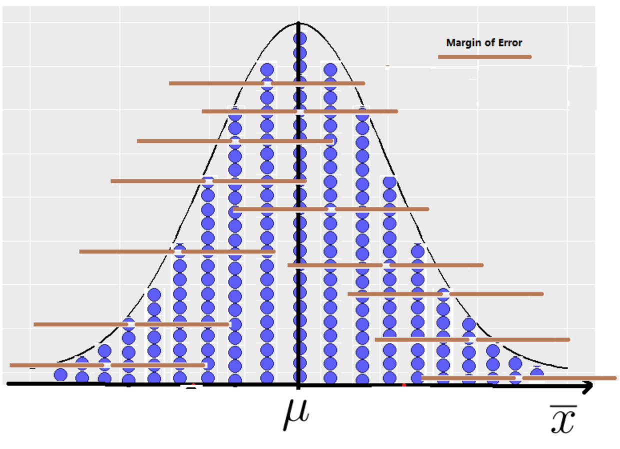

Figure 1.1: Confidence intervals

For samples with

$$\overline{x} \in [\mu - \textbf{ME},\ \mu + \textbf{ME}] \qquad (\textbf{ME} \text{ is short for } \textit{margin of error})$$

the confidence intervals can captures the μ, others can not. A proper margin of error can be chosen such that **C**% of all the possible confidence intervals can capture the parameter μ. This "**C**%" is called **confidence level**. 95% is the most widely used confidence level.

- **Interpreting confidence intervals and confidence levels**

Confidence interval

Suppose we have constructed a confidence interval $[a, b]$ with confidence level **C%**. It is interpreted as "We are **C%** confident that the interval from a to b captures the parameter(in context)'.'

For example, if [160, 180] is 95% confidence interval for the average height of all high school students, it is interpreted as "We are 95% confident the interval from 160cm to 180cm can capture the average height of all high school studets."

Remarks:

- The confident interval is constructed to estimate the *parameter* not the *statistic*. It's wrong to say "We are **95%** confident $\cdots\cdots$ can capture the average height of the sample."
- It is wrong to say "The probability the the interval [160, 180] can capture average height of all height school students is 95%", because the parameter μ is a fixed number not a random variable, and the interval [160, 180] is also fixed and the probability that [160, 180] can capture μ is either 0 or 1
- " the parameter(in context)" means you have to say it out clearly what is the parameter in the context.

Confidence level

C% confidence level is interpreted as "If we draw many samples and construct confidence intervals by using the same formula, **C%** of those intervals can capture the parameter(in context)."

Check Your Understanding

How much does the fat content of Brand X hot dogs vary? To find out, researchers measured the fat content (in grams) of a random sample of 10 Brand X hot dogs. A 95% confidence interval for the population standard deviation σ is 2.84 to 7.55.

(a) Interpret the confidence interval.

(b) Interpret the confidence level.

(c) True or false: The interval from 2.84 to 7.55 has a 95% chance of containing the actual population standard deviation s. Justify your answer.

- **The formula for confidence interval**

The confidence intervals for parameter p and parameter μ are

$$\hat{p} \pm \textit{margin of error} \qquad \bar{x} \pm \textit{margin of error}$$

$\hat{p}$ and $\bar{x}$ can be calculated from the sample directly. What about *margin of error*?

Refer to figure1.1. Suppose the conditions are met such that the sampling distribution of $\bar{x}$ is approximately normal: $\bar{x} \sim N(\mu, \sigma_{\bar{x}})$. Any sample with

$$\bar{x} \in [\mu - \mathbf{ME},\ \mu + \mathbf{ME}] \qquad (\mathbf{ME} \text{ is short for } \textit{margin of error})$$

can result in a confidence interval captures the parameter μ. If we want the *confidence level* to be **C%**, then

$$\mathbf{P}(\mu - \mathbf{ME} \le \bar{x} \le \mu + \mathbf{ME}) = \mathbf{C\%} \implies \mathbf{P}(\frac{-\mathbf{ME}}{\sigma_{\bar{x}}} \le \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \le \frac{\mathbf{ME}}{\sigma_{\bar{x}}}) = \mathbf{C\%}$$

$$\implies \mathbf{P}(\frac{-\mathbf{ME}}{\sigma_{\bar{x}}} \le z_{\bar{x}} \le \frac{\mathbf{ME}}{\sigma_{\bar{x}}}) = \mathbf{C\%}$$

In the above equation, $z_{\bar{x}} = \dfrac{\bar{x} - \mu}{\sigma_{\bar{x}}}$, where $z_{\bar{x}}$ is the *z-score* of $\bar{x}$. Since $\bar{x} \sim N(\mu, \sigma_{\bar{x}})$, $z_{\bar{x}}$ follows a *standard normal distribution*, $z_{\bar{x}} \sim N(0, 1)$.

Why the **critical value** for the **95%** confidence interval is 1.96?

What are the critical values for confidence interval with confidence levels **80%, 90%, 99%**

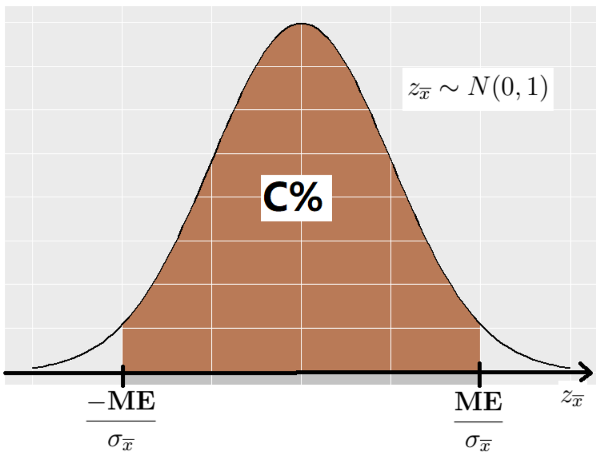

Figure 1.2: Calculate the **Margin of Error**

Referring to figure1.2, if **C%** $= 95\%$, then

$$\frac{\mathbf{ME}}{\sigma_{\bar{x}}} = 1.96 \implies \mathbf{ME} = 1.96 \times \sigma_{\bar{x}} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

The **95% confidence interval** for μ is $\quad \bar{\mathbf{x}} \pm \mathbf{1.96} \times \sigma_{\bar{\mathbf{x}}}$

The general formula for confidence interval is

$$\textbf{statistic} \pm (\textbf{critical value}) \times (\textbf{standard devaition of the statistic})$$

- **Something more about the confidence interval**

  And the *margin of error* is given by

  Margin of Error = (**critical value**) × (**standard devaition of the statistic**)

  The standard deviation of $\widehat{p}$ and $\overline{x}$ are given by

$$\sigma_{\widehat{p}} = \sqrt{\frac{p(1-p)}{n}} \qquad \sigma_{\overline{x}} = \frac{\sigma}{n}$$

  where p and σ are parameters, and n is the sample size.

  - *Confidence level C%* does not change, *sample size n* increases. $\Longrightarrow$

    Margin of Error decreases. $\Longrightarrow$ More precise estimate.

  - *Confidence level C%* increases, *sample size n* dose not change. $\Longrightarrow$

    Margin of Error increases. $\Longrightarrow$ Less precise estimate.

  If the sample size is kept the same, there is a trade-off between the confidence level and the precision of the estimate. More confident means low precision. This is intuitive just by considering the extreme case when the confidence level is 100%, the *margin of error* will be infinitely large.

  If the confidence level is kept the same, and the sample size increase, which means the sample is more representative, and the estimation is more precise, thus a smaller margin of error.

  Check the conditions!!!

  What conditions must be met for all the above formulas to hold?

1.2 One sample z-interval for population proportion

The setting is: Draw an SRS of size n and sample proportion $\hat{p}$, from a population of size N. Construct a C% confidence interval for population proportion p.

- **Check conditions**

  In order to use apply the formulas in previous, we have to check conditions to make sure the sampling distribution of $\hat{p}$ can be approximated by normal distribution.

  - **Random:** The data should either come from a *simple random sample* or an experiment with *random assignments*. Otherwise, the formulas can not be used.

    * **10% condition**. Make sure the sample size is no larger than 10% of the population size. Because the individuals need to be independent to calculate the *standard deviation of the statistics*

    $$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

  - **Large counts condition**

    According to Chapter 6, if we want to approximate the sampling distribution of $\hat{p}$ with a normal distribution, we have to make sure the **10% condition** is met, that is

    $$np \geq 10 \qquad n(1-p) \geq 10$$

    However, we don't know the parameter p. We will use $\hat{p}$ to replace p. For *10% condition* we check

    $$n\hat{p} \geq 10 \qquad n(1-\hat{p}) \geq 10$$

Can you explain in detail why we have to check all those conditions ?

- **The formula**

  $$\hat{p} \pm z^* \sqrt{\frac{p(1-p)}{n}}$$

  - $\mathbf{z}^*$ is the critical value. Here we use z^* means it is closely related to the *z-score*. For 95% confidence level, $z^* = 1.96$
  - $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$. Just like in the *10% condition*, p is not known, and it is replaced by $\hat{p}$.

  $$\mathbf{SE}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

  $\mathbf{SE}_{\hat{p}}$ is the **standard error of** $\hat{p}$.

  When the standard deviation of a statistic is estimated from the statistic, the results is called the **standard error** of the statistic.

- **Put it all together**

  When you are asked to construct a confidence interval for population proportion p, you are supposed to do the following:

  - **Identify appropriate confidence interval.**
  - **Check conditions.**
  - **Construct confidence interval by the formula**

    $$\widehat{p} \pm z^* \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

  - **Conclude and interpret.**

Example

Alcohol abuse has been described by college presidents as the number one problem on campus, and it is an important cause of death in young adults. How common is it? A survey of 10,904 randomly selected U.S. college students collected information on drinking behavior and alcohol-related problems. The researchers defined frequent binge drinking as having five or more drinks in a row three or more times in the past two weeks. According to this definition, 2486 students were classified as frequent binge drinkers. Construct and interpret a 99% confidence interval for the proportion of binge drinkers among college students.

Solutions:

We should construct a one-sample z interval for the proportion p of the frequent binge drinkers among college students if conditions are met.

Conditions:

- **Random**. According to the problem, the 10,904 students were randomly selected.
  * **10% condition**, the sample size 10,904 is less than 10% of all college students
- **Large counts condition**

$$n\widehat{p} = 2,486 \geq 10, \qquad n()1 - \widehat{p} = 10,904 - 2,486 \geq 10$$

All the above conditions are met.

Construct the confidence interval:

The 99% confidence interval for the proportion p is given by

$$\widehat{p} \pm z^* \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

where $\widehat{p} = \frac{2486}{10904} = 0.228$ and $z^* = 2.58$.
The confidence interval for p is

$$0.228 \pm 2.58\sqrt{\frac{0.228(1-0.228)}{10904}} = 0.228 \pm 0.010$$

which is the interval (0.218, 0.238).

Interpret:

We are 99% confident that the interval from 0.218 to 0.238 can capture the proportion of frequent binge drinkers among college students.

Teens' Texting

A Pew Internet and American Life Project survey found that 392 of 799 randomly selected teens reported texting with their friends every day.

(a) Calculate and interpret a 95% confidence interval for the population proportion p that would report texting with their friends every day.

(b) Is it plausible that the true proportion of American teens who text with their friends every day is 0.45? Use your result from part (a) to support your answer.

- **Choosing proper sample size**

  In planning a study, we may want the margin of error restricted within certain range, that is

  $$z^* \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \leq \mathbf{C}, \qquad \mathbf{C} \textit{ is a given value.}$$

  How to choose proper sample size n such that the above inequality holds?

  Analyzing the inequality, we find that if we know the value of z^* and $\widehat{p}$, we can solve n. z^* is decided by the *confidence level*. The only trouble is $\hat{p}$. There are two ways to find $\widehat{p}$.

  - Give $\widehat{p}$ a reasonable guess according to pilot studies or past experiences.
  - Solve the inequality

    $$n \geq (\frac{z^*}{\mathbf{C}})^2 \, \widehat{p}(1-\widehat{p})$$

    $$n \geq (\frac{z^*}{\mathbf{C}})^2 \, 0.5(1-0.5) \implies n \geq (\frac{z^*}{\mathbf{C}})^2 \, \widehat{p}(1-\widehat{p}) \qquad \text{for any } \widehat{p}$$

    This means, if we let $\widehat{p} = 0.5$ and solve n, then n can make sure $\mathbf{ME} \leq \mathbf{C}$. This estimation of n is a **conservative estimation**.

Can you taste PTC?

PTC is a substance that has a strong bitter taste for some people and is tasteless for others. The ability to taste PTC is inherited. About 75% of Italians can taste PTC, for example. You want to estimate the proportion of Americans who have at least one Italian grandparent and who can taste PTC.

(a) How large a sample must you test to estimate the proportion of PTC tasters within 0.04 with 90% confidence? Answer this question using the 75% estimate as the guessed value for $\widehat{p}$.

(b) Answer the question in part (a) again, but this time use the conservative guess $\widehat{p} = 0.5$. By how much do the two sample sizes differ?

1.3 Two-sample z interval for $\mathbf{P_1 - P_2}$

The setting is: draw two SRS from two populations. The sample sizes are n_1 and n_2, and sample proportion $\widehat{p}_1$ and $\widehat{p}_2$. Construct a C% confidence interval for the difference between the two population proportions: $p_1 - p_2$.

According the general formula for confidence interval

statistic $\pm$ **critical valuse** $\times$ **standard deviation of the statistic**,

the confidence interval for $p_1 - p_2$ is given by

$$\widehat{p}_1 - \widehat{p}_2 \pm z^* \, \sigma_{\widehat{p}_1 - \widehat{p}_2}$$

- **Check conditions**

  In order to apply the above formulas, we have to make sure the sampling distribution of $\widehat{p}_1 - \widehat{p}_2$ is approximately normal. We only have to make sure both the sampling distribution of $\widehat{p}_1$ and $\widehat{p}_2$ are normal, and they are independent.

  Why the sampling distributions of $\widehat{p}_1$ and $\widehat{p}_2$ are normal and independent?

  - **Independence**

    Make sure the two set of data are independent.

  - **Random**

    Both of the samples are random or both of the experiments are designed with *random assignments*.

    * **10% condition**: Both of the two samples satisfy the 10% condition.

  - **Large counts condition**

    Both of the samples satisfy the large counts condition.

  Describe the sampling distribution of $\widehat{p}_1 - \widehat{p}_2$ if all the conditions are met.

- **The formula**

$$\widehat{p}_1 - \widehat{p}_2 \pm z^* \sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}$$

$$\mathbf{SE}_{\widehat{p}_1 - \widehat{p}_2} = \sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}$$

Can you prove the formula of $\mathbf{SE}_{\widehat{p}_1 - \widehat{p}_2}$

When constructing the confidence interval in **AP exam**, the general steps are the same as constructing one-sample z interval for p.

Teens and Adults on Social Networking Sites

As part of the Pew Internet and American Life Project, researchers conducted two surveys in 2012. The first survey asked a random sample of 799 U.S. teens about their use of social media and the Internet. A second survey posed similar questions to a random sample of 2253 U.S. adults. In these two studies, 80% of teens and 69% of adults used social-networking sites.

Construct and interpret a 95% confidence interval for the difference between the proportion of all U.S. teens and adults who use social-networking sites.

1.4 One-sample z interval for population mean

The setting is: draw an SRS of size n from a population of size N and standard deviation σ. The sample mean is $\overline{x}$. Construct a C% confidence interval for population mean μ.

The formula for confidence interval is: $\overline{x} \pm z^* \, \sigma_{\overline{x}} \qquad \sigma_{\overline{x}} = \dfrac{\sigma}{\sqrt{n}}.$

- **Check conditions:**

  The above formula is built on basis that the sampling distribution of $\overline{x}$ is a normal distribution. We have to check related conditions to guarantee this.

  - **Random**

    The sample is random or the experiment is designed with *random assignment*.

    * **10% condition:** the sample size is less than 10 percent of the population size.

  - **Normal/Large sample**

    Check either of the following:

    * The population has a normal distribution.
    * The sample size $n \geq 30$.
    * The distribution of the sample data is roughly symmetric and no obvious outliers, and thus it is reasonable to assume the population distribution is approximately normal.

    Those three conditions are check in the order as listed until one of them is satisfied.

- **The formula**

  $$\overline{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

  Here, we know σ. We don't use $\mathbf{SE}_{\overline{x}}$, for .

Describe the sampling distribution of $\overline{x}$ if all the conditions are met.

1.5 Two-sample z interval for $\mu_1 - \mu_2$

The setting is: draw two samples of sizes n_1 and n_2, sample means $\overline{x}_1$ and $\overline{x}_2$, from two populations with population standard deviations σ_1 and σ_2. Construct a C% confidence interval for $\mu_1 - \mu_2$

The formula for this confidence interval is:

$$\overline{x}_1 - \overline{x}_2 \pm z^* \ \sigma_{\overline{x}_1 - \overline{x}_2}$$

- **Check conditions**

  In order to apply the formula given above, we have to make sure the sampling distribution of $\overline{x}_1 - \overline{x}_2$ is approximately normal by checking certain conditions.

  – **Independence**

    The two simple random sample are independent

  – **Random**

    The samples are simple random samples or the experiments are designed by random assignments

    ∗ **10% condition.** Both the two samples satisfy the *10% condition*.

  – **Normal/Large sample**

    Check either of the following for both of the samples.

    ∗ The population has a normal distribution.
    ∗ The sample size $n \geq 30$.
    ∗ The distribution of the sample data is roughly symmetric and no obvious outliers, and thus it is reasonable to assume the population distribution is

Describe the sampling distribution of $\overline{x}_1 - \overline{x}_2$ if all the conditions are met.

- **The formula**

Can you deduce the formulas?

Why don't we use $\mathbf{SE}_{\overline{x}_1 - \overline{x}_2}$?

$$\overline{x}_1 - \overline{x}_2 \pm z^* \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Who is taller?

The heights of young men follow a Normal distribution with standard deviation 2.8 inches. The heights of young women follow a Normal distribution with standard deviation 2.5 inches. Suppose we select independent SRSs of 16 young men and 9 young women. The sample mean heights of young men is: $\overline{x}_M = 69.3$ inches. The sample young women heights is: $\overline{x}_W = 64.5$ inches

Construct a 90% confidence interval for the difference of the mean heights be young men and young women.

1.6 t intervals for population mean

In previous section we learned how to construct confidence intervals for population means with population standard deviation known. However, most of the time if we don't know the population we don't know the population standard deviation. People may say, we can replace the population standard deviation with the sample standard deviation and apply the above formulas. The situation is different here. If population standard deviation is replaced by the sample standard deviation, the sampling distribution is the *t-distribution*.

Suppose the C% confidence interval for population mean μ is

$$\overline{x} \pm \mathbf{ME}$$

where **ME** is short for the *margin of error*. Suppose the sampling distribution of $\overline{x}$ is normal. Referring to section 1, we have the following:

$$\begin{aligned}
\mathbf{P}(\mu - \mathbf{ME} \leq \overline{x} \leq \mu + \mathbf{ME}) &= \mathbf{C\%} \\
\mathbf{P}(\frac{-\mathbf{ME}}{s_x/\sqrt{n}} \leq \frac{\overline{x}-\mu}{s_x/\sqrt{n}} \leq \frac{\mathbf{ME}}{s_x/\sqrt{n}}) &= \mathbf{C\%} \\
\mathbf{P}(\frac{-\mathbf{ME}}{\mathbf{SE}_{\overline{x}}} \leq \frac{\overline{x}-\mu}{\mathbf{SE}_{\overline{x}}} \leq \frac{\mathbf{ME}}{\mathbf{SE}_{\overline{x}}}) &= \mathbf{C\%} \\
\mathbf{P}(\frac{-\mathbf{ME}}{\mathbf{SE}_{\overline{x}}} \leq t_{\overline{x}} \leq \frac{\mathbf{ME}}{\mathbf{SE}_{\overline{x}}}) &= \mathbf{C\%}
\end{aligned}$$

In the above equation, $t_{\overline{x}}$ follows the *t-distribution* with *degree of freedom* $dg = n-1$, where n is the sample size.

Find t^* for the following cases

- $n = 5, \quad \mathbf{C\%} = 95\%$
- $n = 10, \quad \mathbf{C\%} = 95\%$
- $n = 20, \quad \mathbf{C\%} = 95\%$

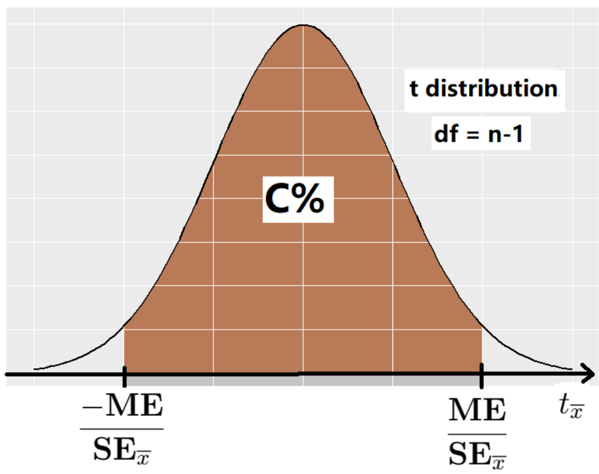

Figure 1.3: t interval

As shown in figure1.3, suppose $t^* = \dfrac{\mathbf{ME}}{\mathbf{SE}_{\overline{x}}}$, then $\mathbf{ME} = t^* \times \mathbf{SE}_{\overline{x}}$. The t-interval is

$$\bar{x} \pm t^* \times \mathbf{SE}_{\bar{x}}$$

t^* is the **critical value**, and the formula is consistent with

$$\textbf{statistic} \pm (\textbf{critical value}) \times (\textbf{standard error of the statistic})$$

- **One-sample t interval for population mean**

  The setting of this type of interval is exactly the same as *one-sample z interval for population mean*, except that we don't know the population standard deviation σ.

  The conditions to check are exact the same as the conditions to check in *one-sample z interval for population mean.*

  The formula:

  $$\bar{x} \pm t^* \frac{s_X}{\sqrt{n}}$$

  The degree of freedom of t^* is $n-1$

- **Two-sample t interval for $\mu_1 - \mu_2$**

  What is true for *one-sample t interval* is also true here.

  The formula:

  $$\bar{x}_1 - \bar{x}_2 \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

  $$\mathbf{SE}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

  The degree of freedom are given by one of the two methods.

  - $df = \min(n_1 - 1, n_2 - 1)$, this is a **conservative** way.
  - $df = \dfrac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{1}{n_1-1}(\frac{s_1^2}{n_1})^2 + \frac{1}{n_2-1}(\frac{s_2^2}{n_2})^2}$ This will be given by the calculator.

  In practice, we just report the *df* given by the calculator.

Can you deduce the formula of $\mathbf{SE}_{\bar{x}_1}$?

The t statistic of $\bar{x}$ is given by

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

Can you find the formula for the t statistic of $\bar{x}_1 - \bar{x}_2$

Remark:

- Confidence intervals are for parameters.
- There is no t intervals for populations proportions.
- For population means, t intervals are more frequently used than z intervals
- The steps of constructing confidence t intervals are the same as of z intervals. **The only difference is you have to report the *degree of freedom* of t^* when constructing t intervals.**

- **Choosing proper sample size**

  Sometimes we have to choose proper sample size such that the margin of error is less than a specified value, say C.

  $$\textit{Margin of error } = t^* \frac{s_x}{\sqrt{n}} \leq \mathbf{C}$$

  The problem of solving n is t^* depends on *the degree of freedom*, which is $n-1$, and s_x is related to the sample size n as well. We do the following two things to simplify:

  – Replace t^* by z^* which is totally decided by the confidence level C%.
  – Replace s_x by some value, say σ, from a pilot study or past experience, or a reasonable guess.

  Thus the inequality becomes $z^* \dfrac{\sigma}{\sqrt{n}} \leq \mathbf{C}$. Easy to solve for n.

How many monkeys

Researchers would like to estimate the mean cholesterol level m of a particular variety of monkey that is often used in laboratory experiments. They would like their estimate to be within 1 milligram per deciliter (mg/dl) of the true value of m at a 95% confidence level. A previous study involving this variety of monkey suggests that the standard deviation of cholesterol level is about 5 mg/dl.

Obtaining monkeys for research is time-consuming, expensive, and controversial. What is the minimum number of monkeys the researchers will need to get a satisfactory estimate?

Self-healing

Biologists studying the healing of skin wounds measured the rate at which new cells closed a cut made in the skin of an anesthetized newt. Here are data from a random sample of 18 newts, measured in micrometers (millionths of a meter) per hour:

29	27	34	40	22	28	14	35	26	35	12	30	23	18	11	22	23	33

Calculate and interpret a 95% confidence interval for the mean healing rate.

Tropical Flowers

Different varieties of the tropical flower *Heliconia* are fertilized by different species of hummingbirds. Researchers believe that over time, the lengths of the flowers and the forms of the hummingbirds beaks have evolved to match each other. Here are data on the lengths in millimeters for random samples of two color varieties of the same species of flower on the island of Dominica:

H. *caribaea* red											
41.90	42.01	41.93	43.09	41.17	41.69	39.78	40.57	39.63	42.18	40.66	37.87
39.16	37.40	38.20	38.07	38.10	37.97	38.79	38.23	38.87	37.78	38.01	

H. *caribaea* yellow											
36.78	37.02	36.52	36.11	36.03	35.45	38.13	37.10	35.17	36.82	36.66	35.68
36.03	34.57	34.63									

(a) Draw two boxplots to compare the distribution of the two samples.

(b) Construct and interpret a 95% confidence interval for the difference in the mean lengths of these two varieties of flowers.

(c) Does the interval support the researchers belief that the two flower varieties have different average lengths? Explain.