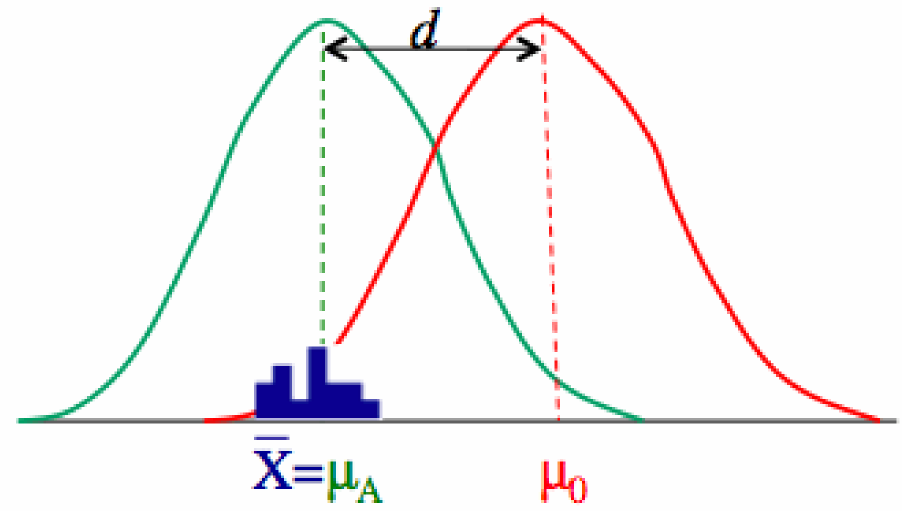
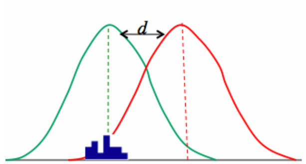
Inferential statistics

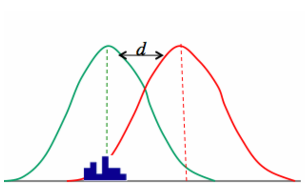


- Distribution of sample means
  - Standard error
  - Central limit theorem
- Hypothesis testing
  - 5 step program
    - Hypotheses
      - Null and Alternative
      - 1 or 2 tailed
    - Error types
      - p(Type I error) = 〈 alpha
      - p(Type II error) = ® beta
    - Reaching a conclusion
      - Reject the H_0
      - Fail to reject H_0
  - Test statistics
    - 1-sample z
    - 1-sample t
    - Related sample t
    - Independent sample t



Central Limit Theorem

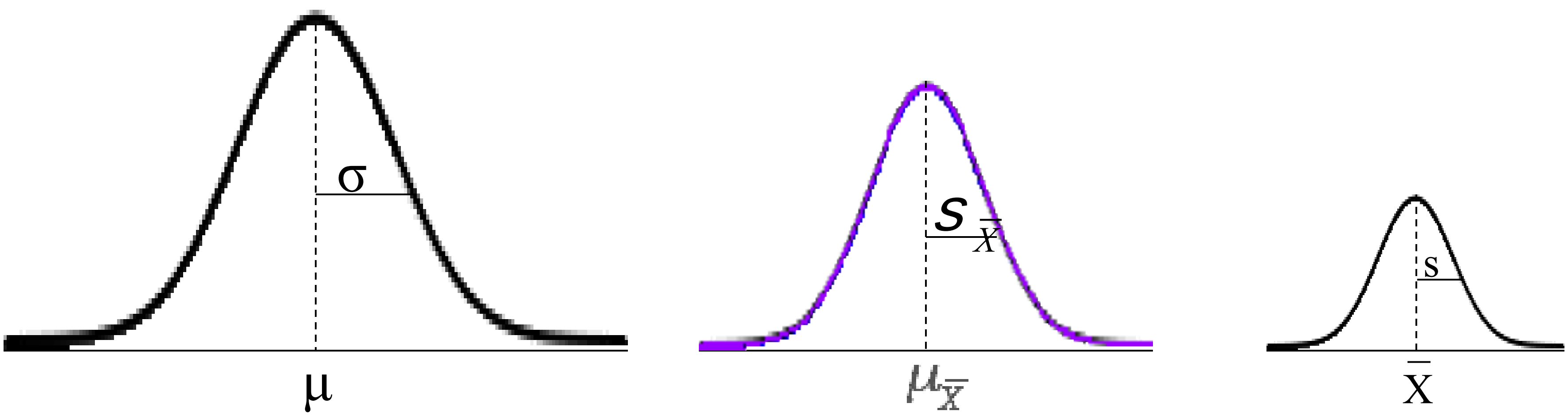
- "Sampling Distribution of the mean of any independent random variable will be normal"
- This applies to both discrete and continuous distributions.
- The random variable should have a well defined mean and variance (standard deviation).
- Applicable even when the original variable is not normally distributed.

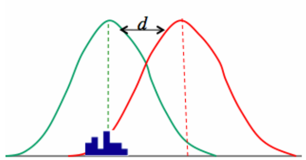


Central Limit Theorem

– For any population with mean μ and standard deviation σ, the distribution of sample means for sample size *n* will approach a *normal* distribution with a mean of μ and a standard deviation of $\frac{s}{\sqrt{n}}$ as *n* approaches infinity (good approximation if $n > 30$).

Population   Distribution of sample means   Sample





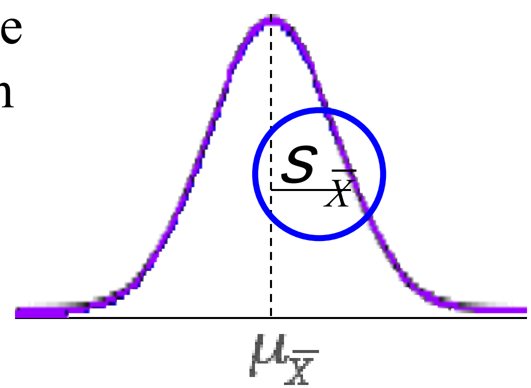
Properties of the distribution of sample means

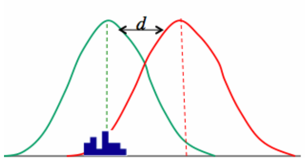
Central Limit Theorem

– For any population with mean μ and standard deviation σ, the distribution of sample means for sample size *n* will approach a *normal* distribution with a mean of μ and a standard deviation of $\frac{s}{\sqrt{n}}$ as *n* approaches infinity (good approximation if *n* > 30).

Distribution of sample means

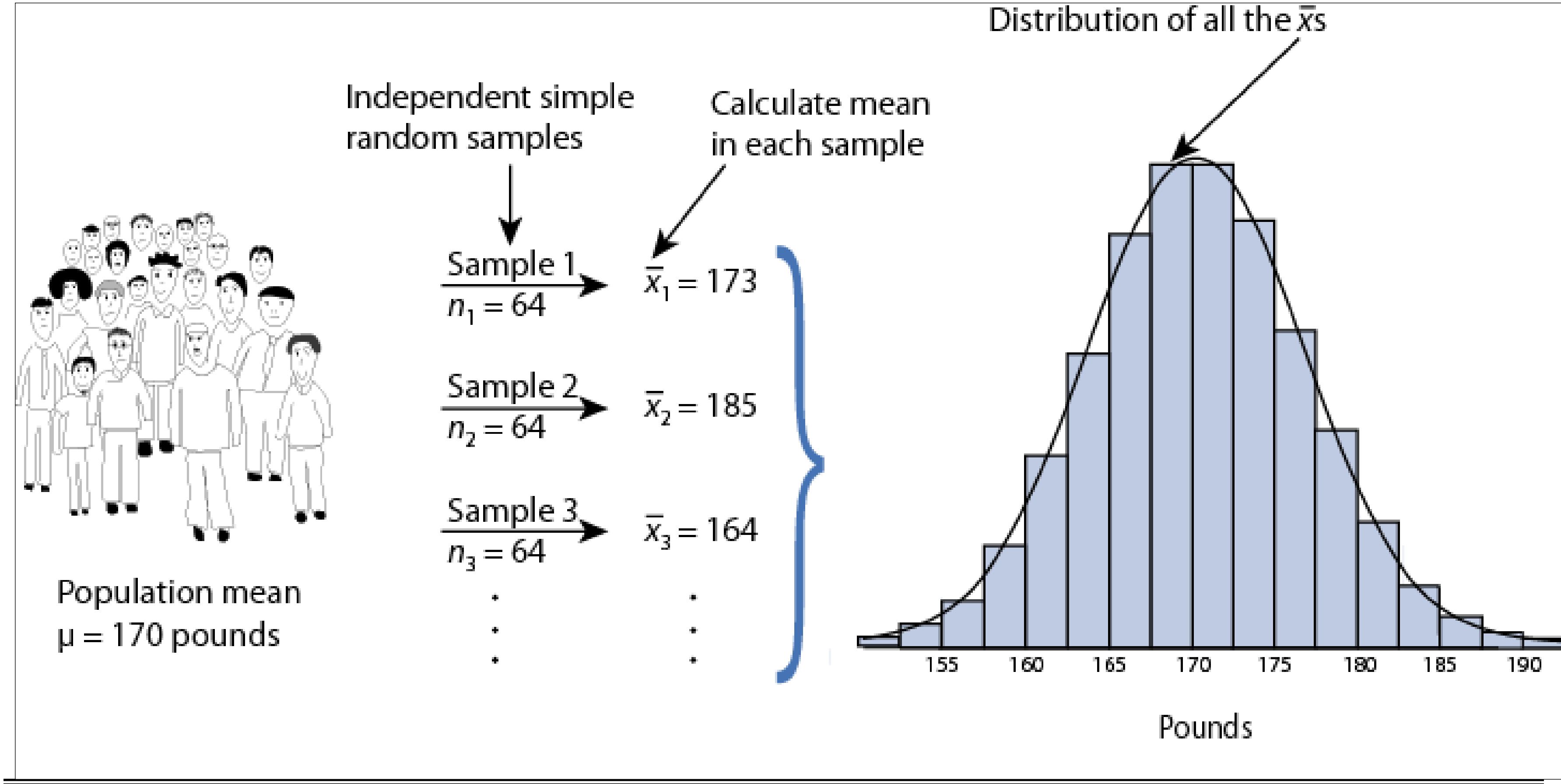
Standard error: the average amount a sample mean (of a particular sample size) will differ from the population mean.
Used as our *difference expected by chance* in our test statistics.

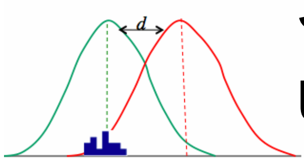




Properties of the distribution of sample means

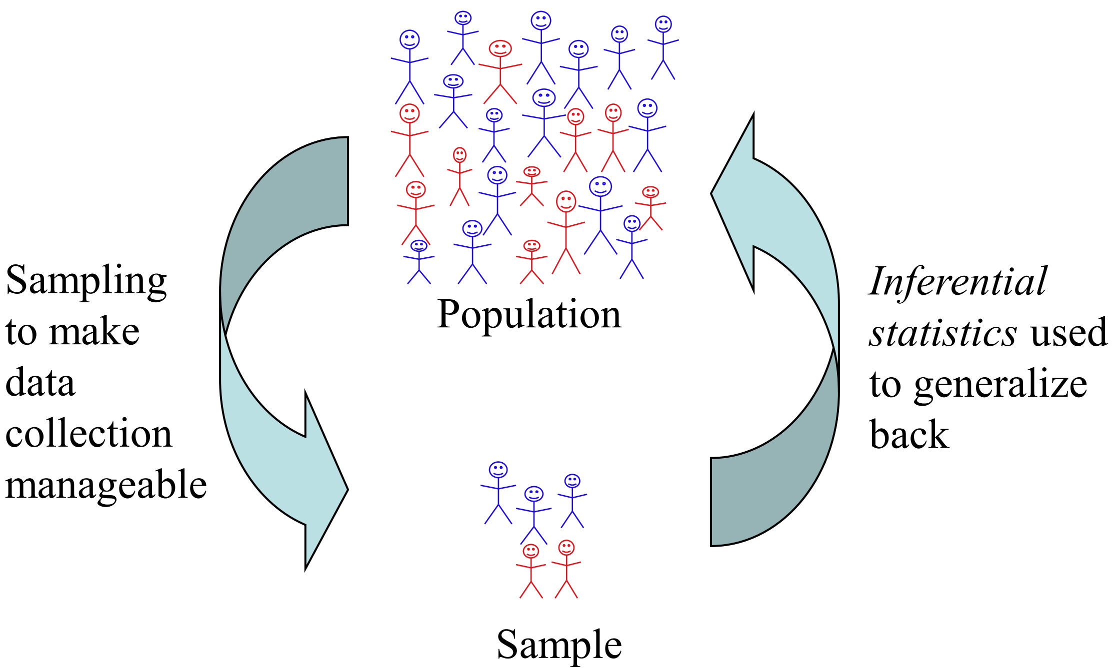
Central Limit Theorem

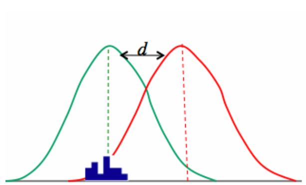




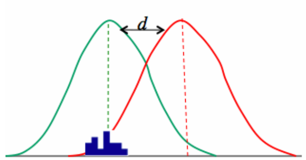
Sampling distribution of xbar under H_0: $\mu = 170$ for $n = 64 \Rightarrow$ $\bar{x} \sim N(170,5)$

- Hypothesis testing
  - Testing claims about populations (and the effect of variables) based on data collected from samples
    - Using estimates of the **sampling error** (expected difference between the sample statistics and the population parameters)



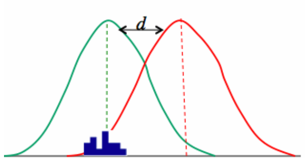


- **Population** ≡ all possible values
- **Sample** ≡ a portion of the population
- **Statistical inference** ≡ generalizing from a sample to a population with calculated degree of certainty
- Two forms of statistical inference
  - **Hypothesis testing**
  - **Estimation**
- **Parameter** ≡ a characteristic of population, e.g., population mean μ
- **Statistic** ≡ calculated from data in the sample, e.g., sample mean $(\bar{x})$



Distinctions Between Parameters and Statistics

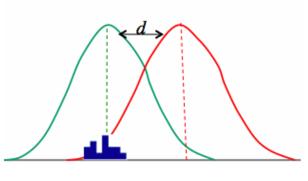
	Parameters	Statistics
Source	Population	Sample
Notation	Greek (e.g., μ)	Roman (e.g., *xbar*)
Vary	No	Yes
Calculated	No	Yes



- Hypothesis testing
  - Testing claims about populations (and the effect of variables) based on data collected from samples
    - Using estimates of the **sampling error** (expected difference between the sample statistics and the population parameters)

A five step program

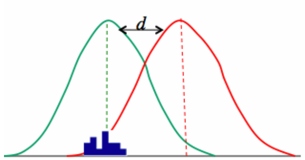
- Step 1: State your hypotheses
- Step 2: Set your decision criteria
- Step 3: Collect your data from your sample
- Step 4: Compute your test statistics
- Step 5: Make a decision about your null hypothesis



Testing Hypotheses

Null and Alternative Hypotheses

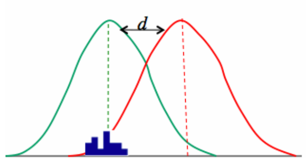
- Convert the research question to null and alternative hypotheses
- The **null hypothesis (H_0)** is a claim of "no difference in the population"
- The **alternative hypothesis (H_a)** claims "H_0 is false"
- Collect data and seek evidence against H_0 as a way of bolstering H_a (deduction)



Concluding that there is a difference between groups (“an effect”) when there really isn’t

		Real world (‘truth’)	
		H_0 is correct	H_0 is wrong
Experimenter’s conclusions	Reject H_0	Type I error *a*	
	Fail to Reject H_0		Type II error *b*

Concluding that there isn’t an effect, when there really is



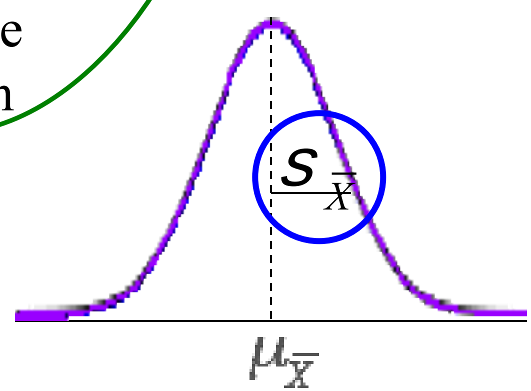
Error types

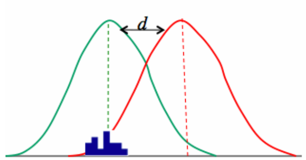
Could be difference between a sample and a population, or between different samples

"Generic" test statistic

$$\text{test statistic} = \frac{\text{observed difference}}{\text{difference expected by chance}}$$

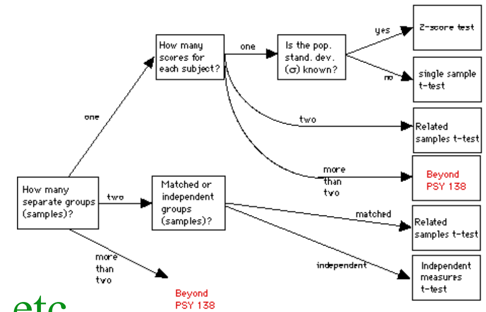
Standard error: the average amount a sample mean (of a particular sample size) will differ from the population mean.
Used as our *difference expected by chance* in our test statistics.





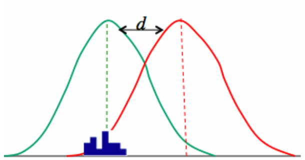
Properties of the distribution of sample means

- Analyze the question/problem.
  - The design of the research: how many groups, how many scores per person, is the population σ known, etc.
    - Use the decision tree
      - Will you use z's or t's?
      - If two samples, are they independent or related?
  - Write out what information is given
    - Means, standard deviations, number of subjects, α-level, etc.
  - Is it asking you to *test a difference* or *make an estimate*?
    - If hypothesis test:
      - What are the H_0 and H_A?
      - 1-tailed or 2-tailed hypotheses?
  - What is your critical value of your test statistic (z or t from table, you will need your α-level, and *df*, and 1-or-2 tailed)



	Proportion in one tail				
	0.10	0.05	0.025	0.01	0.005
	Proportion in two tails				
df	0.20	0.10	0.05	0.02	0.01
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1,476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
:	:	:	:	:	:
15	1.341	1.753	2.131	2.602	2.947
:	:	:	:	:	:

|z|	.00	.01
0	0.5000	0.4960
:	:	:
:	:	:
1.0	0.1587	0.1562
:	:	:
2.0	0.0228	0.0222
:	:	:
3.0	0.0013	0.0013



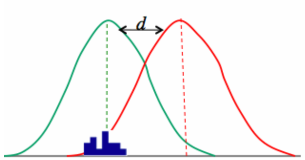
- Analyze the question/problem.

- Now you are ready to do some computations
  - Write out all of the formulas that you will need
    - Test statistic, (estimated) standard error, standard deviation, SS, mean, degrees of freedom
  - Then fill in the numbers as you know them

$$t = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{s_{\bar{X}_A - \bar{X}_B}} \qquad df = n_A + n_B - 2$$

$$s_{\bar{X}_A - \bar{X}_B} = \sqrt{\frac{s_P^2}{n_A} + \frac{s_P^2}{n_B}} \qquad t = \frac{(50 - 44.5) - 0}{5.81} = 0.95$$

$$s_p^2 = \frac{\left(s_A^2 df_A\right) + \left(s_B^2 df_B\right)}{df_A + df_B}$$



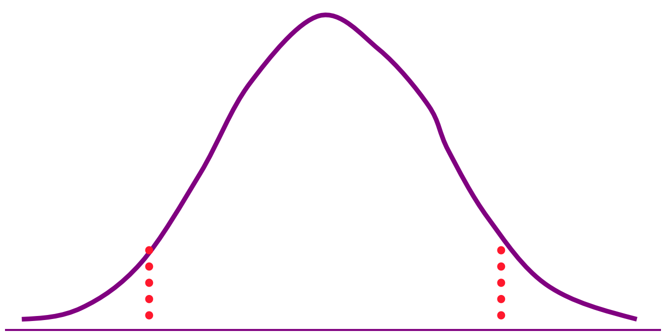
Performing your inferential statistics

- Analyze the question/problem.

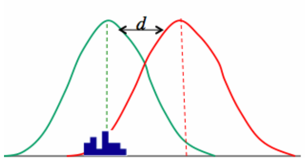
- Now you are ready to do some computations

- Draw a Conclusion and Interpret your final answer
  - Reject or fail to reject the null hypothesis?

Distribution of the t-statistic or z-statistic

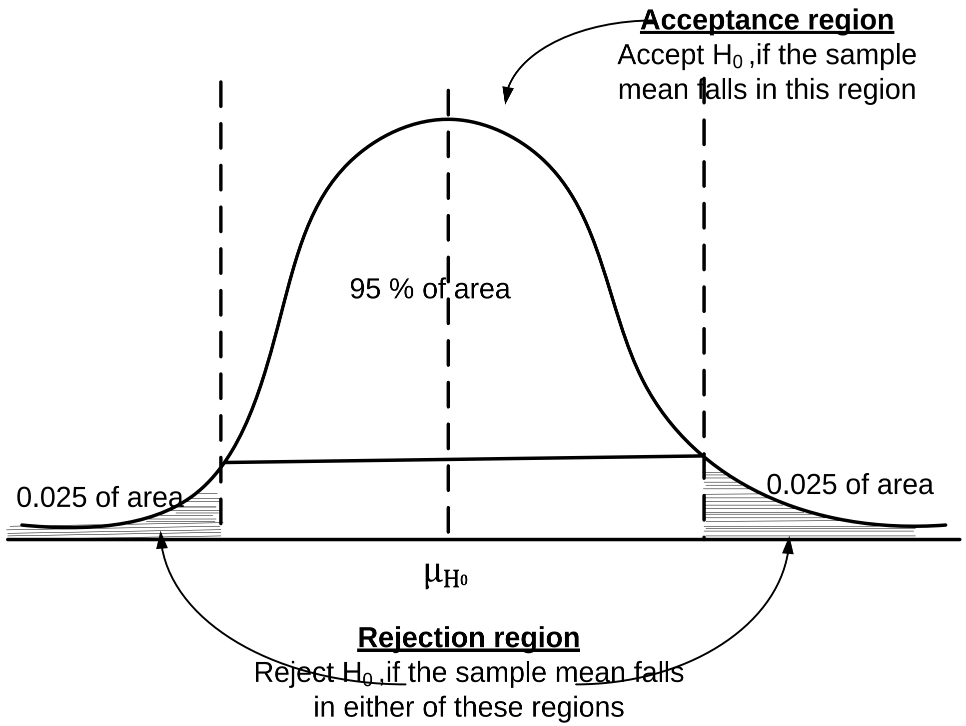


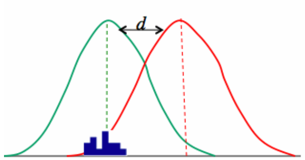
The α-level defines
the critical regions



Performing your inferential statistics

Two-Tailed Test



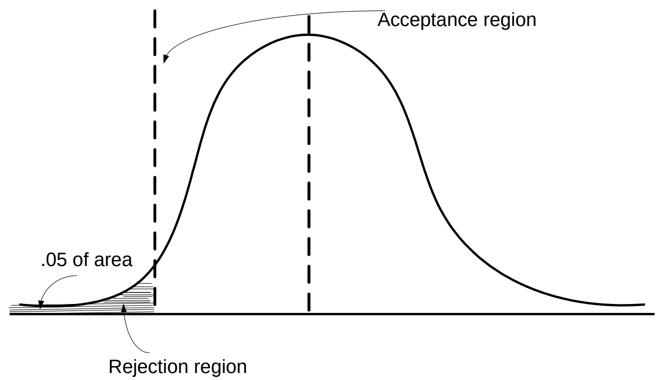


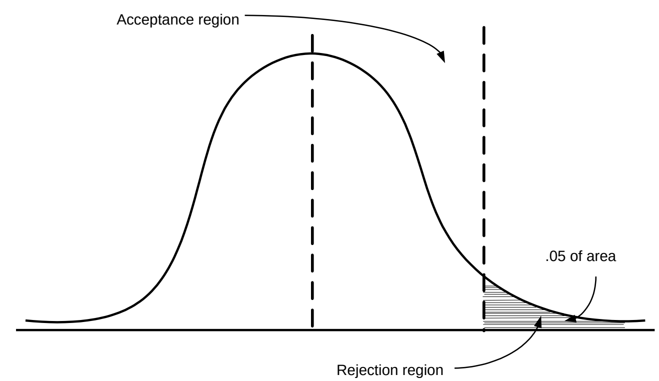
Acceptance and rejection regions in case of a two-tailed test with 5% significance level.

One-Tailed Test

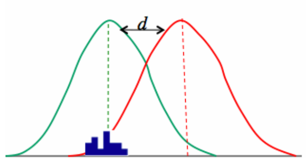
A one-tailed test would be used when we are to test, say, whether the population mean is either lower or higher than the hypothesis test value.

Symbolically,





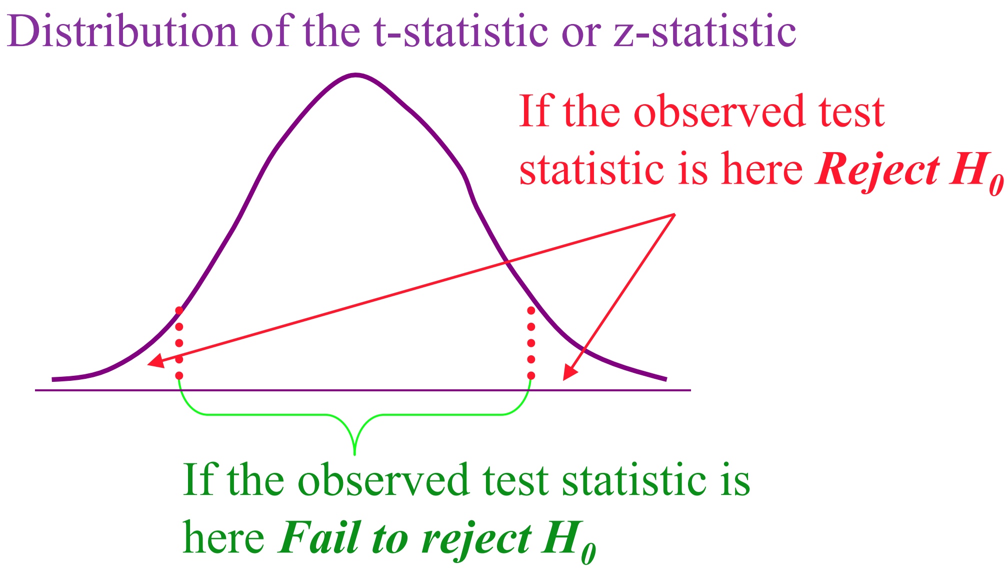
Wherein there is one rejection region only on the left-tail (or right-tail).



- Analyze the question/problem.

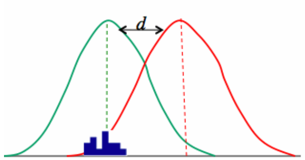
- Now you are ready to do some computations

- Draw a Conclusion and Interpret your final answer
  - Reject or fail to reject the null hypothesis?



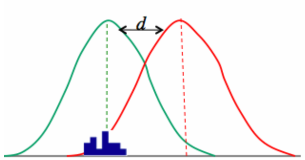
"Evidence suggests that the treatment has an effect"

"Evidence suggests that the treatment has no effect"



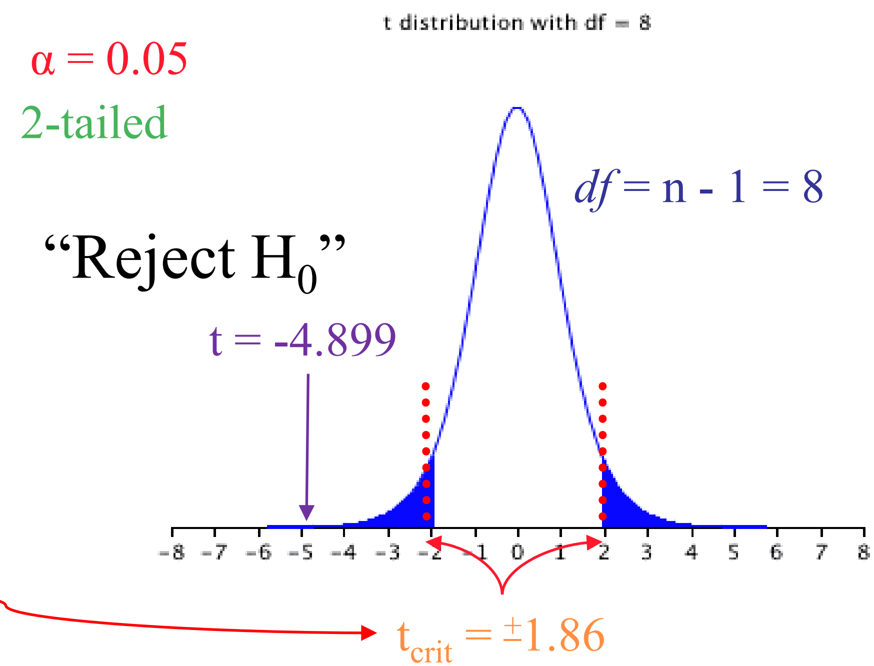
Performing your inferential statistics

- P-value answer the question: What is the probability of the observed test statistic … **when H_0 *is true*?**
- Thus, smaller and smaller P-values provide stronger and stronger evidence against H_0
- Small P-value $\Rightarrow$ strong evidence

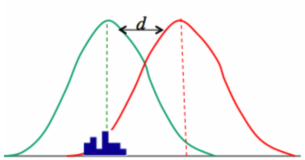


Interpretation

- Drawing your conclusions: Using the t-table
  - Suppose that you conduct a one sample t-test and get a computed t = -4.899 (with a df = 8). Using a 2-tailed test with an α-level of 0.05, what would you conclude?



	Proportion in one tail				
	0.10	0.05	0.025	0.01	0.005
	Proportion in two tails				
df	0.20	0.10	0.05	0.02	0.01
:	:	:	:	:	:
8	1.397	1.753	1.860	2.896	3.355
:	:	:	:	:	:



The t-distribution

- Drawing your conclusions: Using the SPSS output
  - Suppose that you conduct a one sample t-test and get a computed t = -4.899 (with a df = 8). Using a 2-tailed test with an α-level of 0.05, what would you conclude?

T-Test

One-Sample Statistics

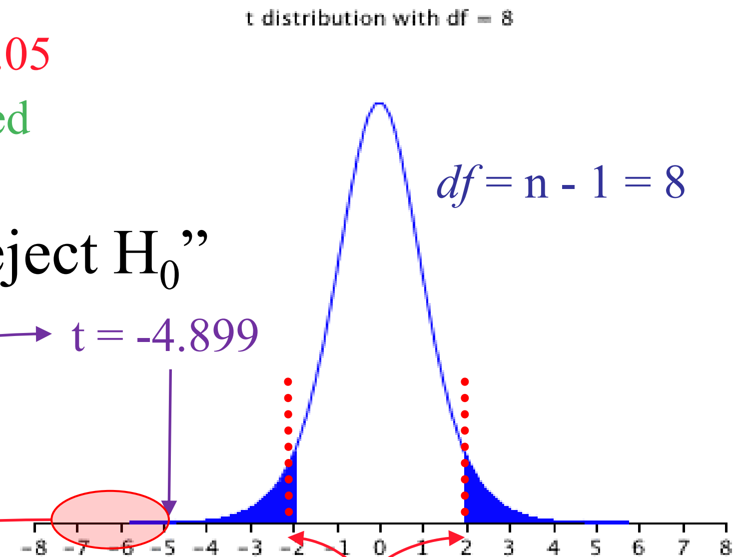
	N	Mean	Std. Deviation	Std. Error Mean
SCORES	9	8.0000	1.2247	.4082

One-Sample Test

	Test Value = 10					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference Lower	Upper
SCORES	-4.899	8	.001	-2.0000	-2.9414	-1.0586

$\alpha = 0.05$

2-tailed



$df = n - 1 = 8$

"Reject H_0"

$t = -4.899$

$t_{crit} = \pm 1.86$

0.001 area in the tail, which is within the .05 defined by α

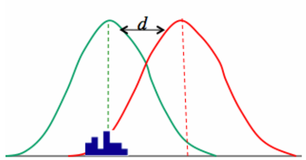
Can compare the p-value and the alpha level

t-value large enough, $p \leq \alpha$ (e.g., .05)

Reject H_0

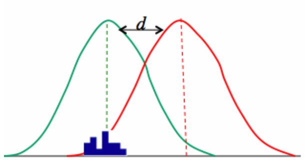
t-value too small, $p > \alpha$ (e.g., .05)

Fail to reject H_0



The t-distribution

Design	Test statistic	(Estimated) Standard error	*df*
One sample, σ known	$z_{\bar{X}} = \dfrac{\bar{X} - m_{\bar{X}}}{s_{\bar{X}}}$	$s_{\bar{X}} = \dfrac{s}{\sqrt{n}}$	
One sample, σ unknown	$t = \dfrac{\bar{X} - m_{\bar{X}}}{s_{\bar{X}}}$	$s_{\bar{X}} = \dfrac{s}{\sqrt{n}}$	$n - 1$
Two related samples, σ unknown	$t = \dfrac{\bar{D} - m_{\bar{D}}}{s_{\bar{D}}}$	$s_{\bar{D}} = \dfrac{s_D}{\sqrt{n_D}}$	$n_D - 1$
Two independent samples, σ unknown	$t = \dfrac{(\bar{X}_A - \bar{X}_B) - (m_A - m_B)}{s_{\bar{X}_A - \bar{X}_B}}$	$s_{\bar{X}_A - \bar{X}_B} = \sqrt{\dfrac{s_P^2}{n_A} + \dfrac{s_P^2}{n_B}}$	$n_A + n_B - 2$



Hypothesis testing formulas summary

Choice of test depends on test statistic and data availability

Means

Compare the sample mean to the population mean when std dev is known
1-sample z-test

Compare the sample mean to the population mean when std dev is unknown
1-sample t-test

Compare the sample means from 2 independent populations when std devs are known
2-sample ind. z-test

Compare the sample means from 2 independent populations when std devs are unknown
2-sample ind. t-test

Compare the sample means from 2 related populations when std devs are unknown
Paired t-test

Compare the sample means from 2 or more independent populations
ANOVA Test

Proportions

Compare the sample proportion to the population proportion
1-sample z-test

Compare the sample proportions from two populations
2-sample z-test

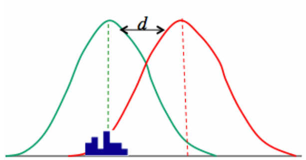
Variances

Compare the sample variance to the population variance
Chi-Square test

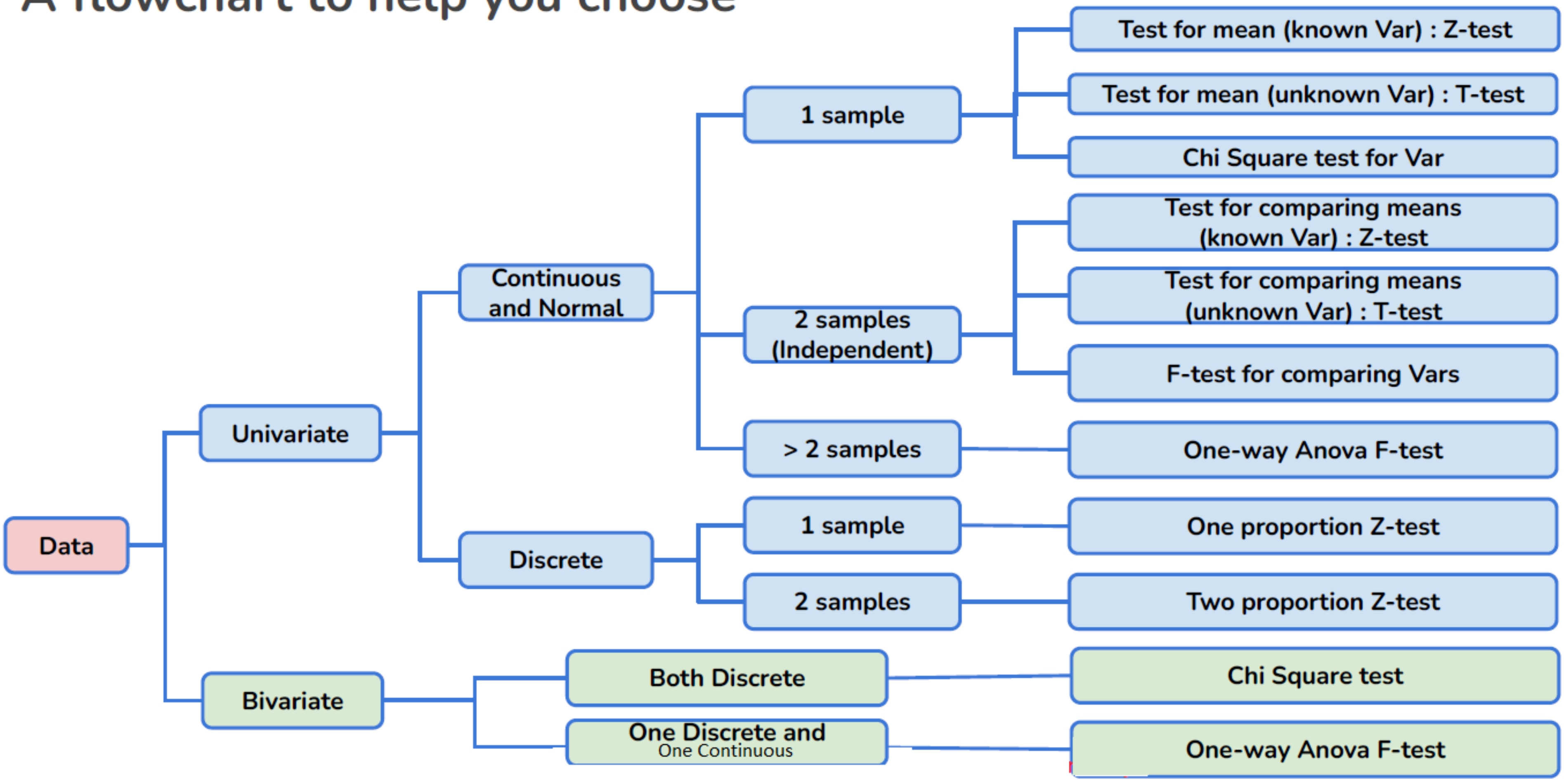
Compare the sample variances from two populations
F-test

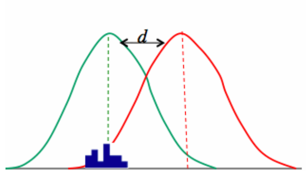
Frequencies

Check whether the categorical variables from a population are independent
Chi-Square Test of Independence



A flowchart to help you choose





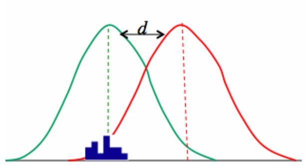
Case Study 1: **Coffee Sale**

A coffee vendor nearby Nalanda Academic complex has been having average sales of 500 cups per day. Because of the development of another wing in the complex, it expects to increase its sales. During the first 12 days, after the inauguration of the new wing, the daily sales were as under:

550 570 490 615 505 580 570 460 600 580 530 526

On the basis of this sample information, can we conclude that the sales of coffee have increased?

Consider 5% as the significance level of testing.



Case Study 1: Step 1

Step 1: Specification of hypotheses and level of significance α

Let us consider the hypotheses for the given problem as follows.

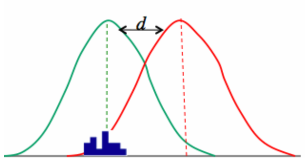
$H_0: \mu = 500$ cups per day

The null hypothesis that sales average 500 cups per day and they have not increased.

$H_1: \mu > 500$

The alternative hypothesis is that the sales have increased.

Given the value of $\alpha = 0.05$ $(i.e., 5\%\ significance\ level\ of\ the\ test)$



Case Study 1: Step 2

Step 2: Sample-based test statistics and the rejection region for specified H_0

Given the sample as

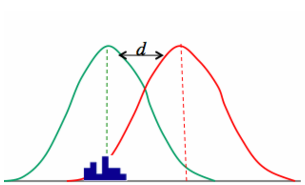
550 570 490 615 505 580 570 460 580 530 526

Since the sample size is small and the population standard deviation is not known, we shall use $t-test$ assuming normal population. The test statistics t is

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

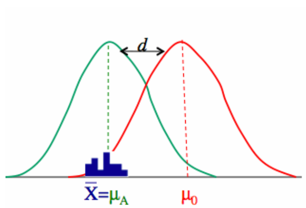
To find $\bar{X}$ and S, we make the following computations.

$$\bar{X} = \frac{\sum x_i}{n} = \frac{6576}{12} = 548$$



Case Study 1: Step 2

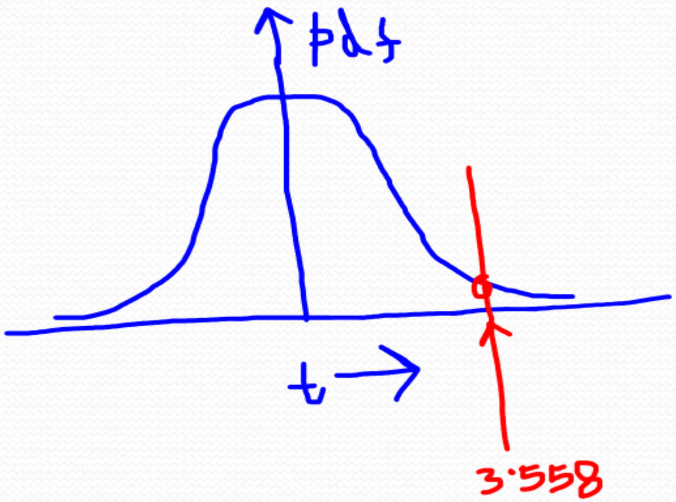
Sample #	X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
1	550	2	4
2	570	22	484
3	490	−58	3364
4	615	67	4489
5	505	−43	1849
6	580	32	1024
7	570	22	484
8	460	−88	7744
9	600	52	2704
10	580	32	1024
11	530	−18	324
12	526	−22	484
$n = 12$	$\sum X_i = 6576$		$\sum (X_i - \bar{X})^2 = 23978$



Case Study 1: Step 2

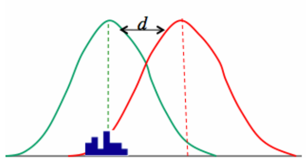
$$S = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{23978}{12-1}} = 46.68$$

Hence, $t = \frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{48}{46.68/\sqrt{12}} = \frac{48}{13.49} = 3.558$



Note:

This gives a t-value given a sample of size n.



Case Study 1: Step 3

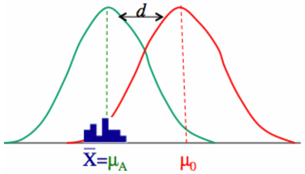
Step 3: Decide the critical value for the hypothesis test

$$Degree\ of\ freedom = n - 1 = 12 - 1 = 11$$

As H_1 is one-tailed, we shall determine the rejection region applying one-tailed in the right tail because H_1 is more than type) at 5% level of significance.

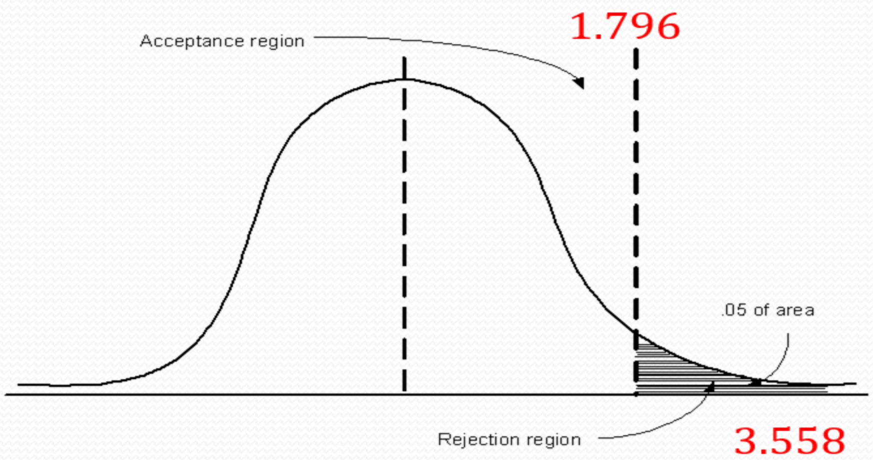
Using table of $t - distribution$ for 11 degrees of freedom and with 5% level of significance,

Rejection region: $\boldsymbol{t > 1.796}$

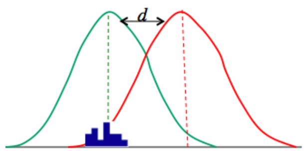


Case Study 1: Step 4

Step 4: Make a decision to either reject or fail to reject H_0



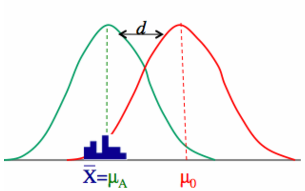
The observed value of $t = 3.558$ which is in the rejection region and thus H_0 is rejected at 5% level of significance.



Case Study 1: Step 5

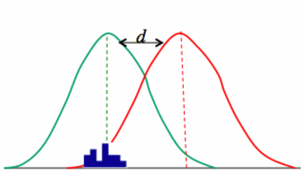
Step 5: Final comment and interpret the result

We can conclude that the sample data indicate that coffee sales have increased.



Case Study 2: **Machine Testing**

A medicine production company packages medicine in a tube of 8 ml. In maintaining the control of the amount of medicine in tubes, they use a machine. To monitor this control a sample of 16 tubes is taken from the production line at random time interval and their contents are measured precisely. The mean amount of medicine in these 16 tubes will be used to test the hypothesis that the machine is indeed working properly. Maximum variance that can be allowed is 0.2.



Case Study 2: Step 1

Step 1: Specification of hypothesis and level of significance α

The hypotheses are given in terms of the population mean of medicine per tube.

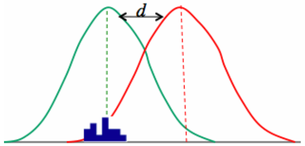
The null hypothesis is

$$H_0 : \mu = 8$$

The alternative hypothesis is

$$H_1 : \mu \neq 8$$

We assume α, the significance level in our hypothesis testing ≈ 0.05.
(This signifies the probability that the machine needs to be adjusted less than 5%).



Case Study 2: Step 2

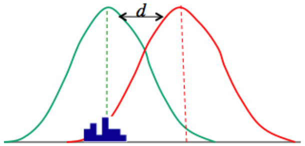
Step 2: Collect the sample data and calculate the test statistics

Sample results: $n = 16$, $\bar{x} = 7.89$, $\sigma = 0.2$

With the sample, the test statistics is

$$z = \frac{7.89-8}{0.2/\sqrt{16}} = -2.20$$

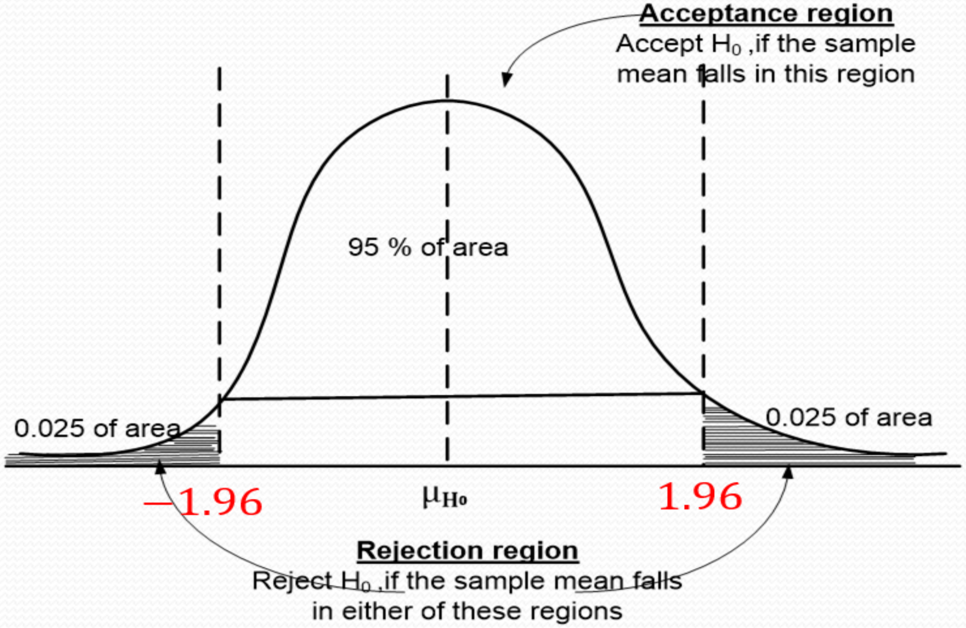
Hence, $|Z| = 2.20$

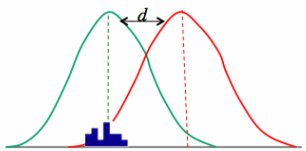


Case Study 2: Step 3

Step 3: To decide the critical region for specified H_0

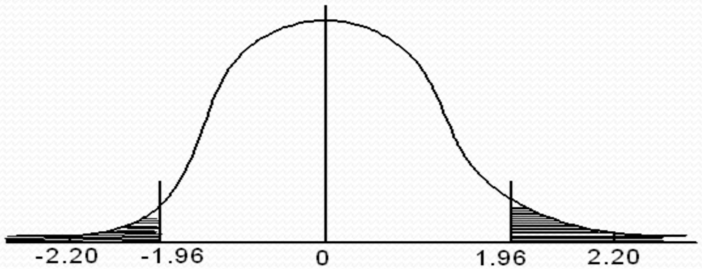
Rejection region: Given $\alpha = 0.05$, which gives $(P|Z| > 1.96)$ (obtained from standard normal calculation for $n(Z: 0{,}1) = 0.025$ for a rejection region with two-tailed test).



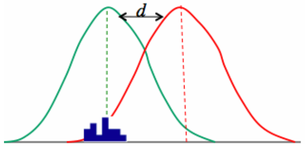


Case Study 2: Step 4

Step 4: Make a decision to either reject or fail to reject H_0



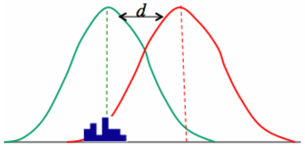
Since $Z > 1.96$, we reject H_0



Case Study 2: Step 5

Step 5: Final comment and interpret the result

We conclude $\mu \neq 8$ and recommend that the machine be adjusted.



Case Study 2: Comment 1 ($\alpha = 1\%$)

Suppose that in our initial setup of hypothesis test, if we choose $\alpha = 0.01$ instead of 0.05, then the test can be summarized as:

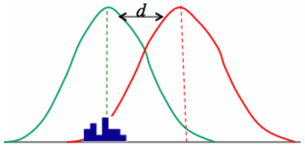
1. $H_0: \mu = 8\,,\quad H_1: \mu \neq 8 \qquad \alpha = 0.01$

2. Reject H_0 if $Z > 2.576$

3. Sample result n =16, $\sigma = 0.2$, $\overline{X}$=7.89, Z $= \frac{7.89-8}{0.2/\sqrt{16}} = -2.20$, $|Z| = 2.20$

4. $|Z| < 2.20$, we fail to reject H_0= 8

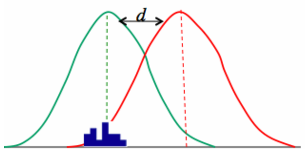
5. We do not recommend that the machine be adjusted.



Case Study 2: Comment 1 ($\bar{X}$= 7.91)

Suppose that in our initial setup of hypothesis test, we choose $\alpha = 0.05$, and the collected sample is $\overline{X}$=7.91 of size 16 with $\sigma = 0.2$. In this case, the test can be summarized as:

1. $H_0: \mu = 8$, $H_1: \mu \neq 8$ $\alpha = 0.05$
2. Reject H_0 if $Z > 1.96$
3. Sample result n =16, $\sigma = 0.2$, $\overline{X}$=7.91, Z $= \frac{7.91-8}{0.2/\sqrt{16}} = -1.80$, $|Z| = 1.80$
4. $|Z| < 1.96$, we fail to reject H_0= 8
5. We do not recommend that the machine be adjusted.



- Dr. Ruthie asked each member of 10 married couples with children to rate the importance of "nights out" (1=not important, 5= very important). Dr. Ruthie hypothesized that wives would consider the nights out more important than the husbands. Test her hypothesis with a probability of making a type I error = 0.05.

Husbands: 5, 3, 4, 2, 3, 4, 5, 1, 1, 4

Wives: 5, 4, 5, 3, 2, 4, 5, 2, 3, 4

Related samples t-test

One tailed *"more important"*

$$t = \frac{\bar{D} - \mu_{\bar{D}}}{s_{\bar{D}}} \qquad s_{\bar{D}} = \frac{s_D}{\sqrt{n_D}}$$

H_0: $\mu_D > 0$

H_A: $\mu_D \leq 0$

α-level =0.05

$$df_D = n_D - 1 = 9$$

$$t = \frac{0.5 - 0}{.26} = 1.9$$

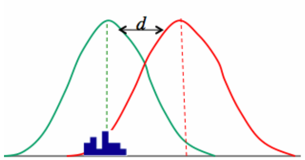
$$D = (X_A - X_B) \qquad SS_D = \sum (D - \bar{D})^2$$

$$t_{crit} = 1.83$$

$$\bar{D} = \sum \frac{D}{n_D} \qquad s_D = \sqrt{\frac{SS_D}{n_D - 1}}$$

Reject H_0

Evidence suggests that wives do care more



Make your flash cards

- Dr. Psychic examined the performance of 28 students who answered multiple-choice items on the SAT test without having read the passages to which the items referred. The mean score was 46.6 (out of 100), with a standard deviation of 6.8. Test whether these students performed different than chance (chance performance would result in 20 correct scores) with an α-level = 0.01.

One sample t-test

Two tailed *"different than"*

$$t = \frac{\bar{X} - \mu_{\bar{X}}}{s_{\bar{X}}} \qquad s_{\bar{X}} = \frac{s_X}{\sqrt{n_X}}$$

H_0: $\mu = 20$ α-level =0.01

H_A: $\mu \neq 20$

$$df = n - 1 = 27$$

$$t = \frac{46.6 - 20}{1.285} = 20.7$$

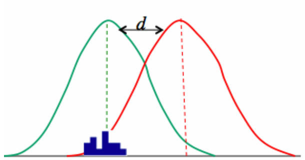
$$SS = \sum \left(X - \bar{X}\right)^2$$

$$t_{crit} = \pm 2.771$$

$$\bar{X} = \sum \frac{X}{n} \qquad s = \sqrt{\frac{SS}{n - 1}}$$

Reject H_0

Evidence suggests that performed differently from chance



Make your flash cards

- Dr. Mnemonic develops a new treatment for patients with a memory disorder. He isn't certain what impact, if any, it will have. To test it he randomly assigns 8 patients to one of two samples. He then gives one sample (A) the new treatment but not the other (B) and then tests both groups with a memory test. Use α = 0.05.

$\bar{X}_A = 44.5 \quad \bar{X}_B = 50$

$s_A = 7.19 \quad s_B = 9.13$

Side 1: Write out the problem

Independent samples t-test

Two tailed *"any impact"*

$H_0: \mu_A = \mu_B$ α-level =0.05

$H_A: \mu_A \neq \mu_B$

$$t = \frac{(\bar{X}_A - \bar{X}_B) - (m_A - m_B)}{s_{\bar{X}_A - \bar{X}_B}}$$

$$df = n_A + n_B - 1 = 6$$

$$s_{\bar{X}_A - \bar{X}_B} = \sqrt{\frac{s_P^2}{n_A} + \frac{s_P^2}{n_B}}$$

$$t = \frac{(50 - 44.5) - 0}{5.81} = 0.95 \quad t_{crit} = \pm 2.45$$

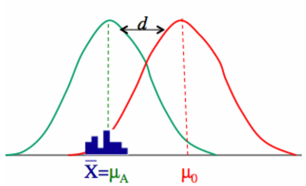
$$s_p^2 = \frac{(s_A^2 df_A) + (s_B^2 df_B)}{df_A + df_B}$$

Fail to Reject H_0

Evidence suggests that the treatment has no effect

Side 2: Write out the solution

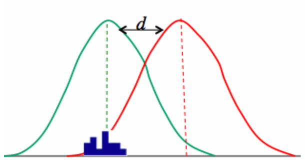
Make your flash cards



Reasoning in Psychology Using Statistics

- There are two types of estimations:
  - Single point estimate
    - For example, sample mean is a single point estimate.
    - This may vary from one sample to another.
    - This is called zero probability of being correct.
    - Not robust and reliable.

  - Interval estimated
    - Estimate with a range of values, for example, population mean is
    - Reliable and robust with essentially non-zero probability of being correct.
    - An alternative method to statistical learning.
    - Popularly known as Confident Interval measurement.



Procedure Confidence Interval Measurement

$$z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \implies \bar{X} - \mu = z \cdot \frac{\sigma}{\sqrt{n}}$$

$$\implies \bar{X} - z \cdot \frac{\sigma}{\sqrt{n}} = \mu$$

This implies that

$$P(Z > z_0) = P\left(\bar{X} - z \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Similarly,

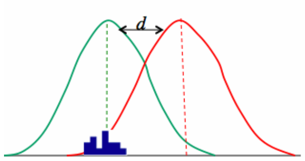
$$P(Z < z_0) = P\left(\bar{X} + z \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Thus,

$P\left(\bar{X} - z \cdot \frac{\sigma}{\sqrt{n}}\right) < \mu < P\left(\bar{X} + z \cdot \frac{\sigma}{\sqrt{n}}\right)$ with probability 1-α

Therefore, the interval estimate of μ is customarily written as

$$\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \text{ to } \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$



Example: Confidence Interval Measurement

Suppose, a hypothesis testing for a population mean $\mu = 8.0$ is as below.

$$\bar{X} = 7.89,\ n = 16, = 0.2,\ \sigma = 0.2 \text{ and } \alpha = 0.05$$

For this testing, we have

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \Longrightarrow 1.96$$

Thus,

$$z_{\alpha/2} = 1.96$$

Hence,

$$\text{Confidence interval is } 7.89 \pm 1.96(0.2)/\sqrt{16}$$

- This is the interval estimate with 95% confidence (i.e., accuracy)
  - We are 95% confident that the true mean is between 6.91 to 8.87
- Here, the term $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ is called maximum error (also called error margin)
- Alternatively,

$$\text{CI estimate is } \bar{X} \pm E$$