Inference for Two Samples

Module 4C

Two Sample designs

- When comparing two groups, the fundamental ways data may be collected are:

  - Independent samples (4.5) (complete randomization): Subjects are randomly divided between 2 treatment groups.

  - Matched pair samples (4.6) (randomization within each matched pair): The two subjects of the pair are matched as closely as possible – same person, identical twins, same relevant characteristics.

Matched Pairs Comparisons

- Experimental units or subjects in the pair must be as alike as possible. Pairs are also called "blocks".
- The members of the pairs are assigned to treatments randomly.

Matched Pairs Design

Matched pair	Experimental units	
1	2	1
2	1	2
3	1	2
⋮	⋮	⋮
n	2	1

Units in each pair are alike, whereas units in different pairs may be dissimilar. In each pair, a unit is chosen at random to receive treatment 1, the other unit receives treatment 2.

Matched Pairs Design

- To compare two treatments, subjects are matched in pairs and each treatment is given to one subject in each pair (comparative inference)
- Subjects in the pair must be as alike as possible.
- The members of the pairs are assigned to treatments randomly.
- **Dependent samples**: individuals selected for sample 1 have a relationship to those selected for sample 2.

Matched Pairs Examples

- Pairing – pairing similar subjects according to some identifiable characteristics(e.g. age, gender) serves to remove this source of variation from the experiment (i.e. control potential lurking variables).

  - Example: Twin studies often try to sort out the influence of genetic factors (and age) by comparing a variable between sets of twins.
  - Example: Pre-test and post-test studies look at data collected on the same subjects before and after some experiment is performed (each subject receives both treatments).

Data Structure

Structure of Data for a Matched Pair Comparison

Pair	Treatment 1	Treatment 2	Difference
1	X_1	Y_1	$D_1 = X_1 - Y_1$
2	X_2	Y_2	$D_2 = X_2 - Y_2$
.	.	.	.
.	.	.	.
.	.	.	.
n	X_n	Y_n	$D_n = X_n - Y_n$

The differences $D_1, D_2, \ldots, D_n$ are a random sample.
Summary statistics:

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i \qquad S_D^2 = \frac{\sum_{i=1}^{n}(D_i - \overline{D})^2}{n - 1}$$

Be aware of the order of subtraction in interpreting *D-bar*.
The random variable D_i is used with a t statistic if distributed normally.

Matched Pairs t-procedures

In these cases, we use the paired data to test the difference in the two population means.

H_0: $\mu_{\text{Difference}} = 0$; H_a: $\mu_{\text{Difference}} > 0$ (or < 0, or $\neq 0$)

Test statistic:

$$t = \frac{\bar{x}_D - \mu_D}{s_D / \sqrt{n_D}}$$

Conceptually, this is not different from tests on one population.

Use of t-distribution

- The distribution of the differences needs to be normal (or not extremely non-normal).
- If so, the test statistic $\frac{\bar{x}_D - \mu_D}{s_D / \sqrt{n_D}}$ has a t-distribution with $n_D - 1$ degrees of freedom.
- Confidence interval: $\bar{x}_D \pm t^* \frac{s_D}{\sqrt{n_D}}$

Idea

- We obtain data from two dependent samples.
- Calculate the difference between each pair.
- Use the column of all the differences to obtain the mean and standard deviation of the differences
  - i.e. treat the column of differences as if it were data from one sample.

Example: Less Body Fat

In a physical conditioning class of 81 students, the % body fat is measured at the beginning & end of the semester. d_i = before – after %.

1. Obtain a CI(98%) of the decrease in % body fat.
2. Test the claim that the class reduces mean % body fat at the 1% level of significance.

Basic data		Confidence Interval		Hypothesis Test	
n =	81	α =	0.02	H_0: $\mu_D = 0$	0
d-bar =	3.322	$z_{\alpha/2}$ =	2.3263	H_1: $\mu_D > 0$	0
S_d =	2.728	ME =	0.7051	α =	0.01
Std err =	0.3031	Lower limit =	2.62	obs z =	10.96
		Upper limit =	4.03	crit z =	2.326
				p-value =	0.000
				Reject H_0	

Paired T example

Does the Pill lower blood pressure over 6-months?

Data				Confidence Interval		Hypothesis Test		
Subject	Before (x)	After (y)	d = x-y	mean =	8.80	Step 1:	H_0: $\mu = 0$	
1	70	68	2	std dev =	10.98		H_1: $\mu > 0$	
2	80	72	8	std err =	2.83	Step 2:	α =	0.01
3	72	62	10	conf level =	0.95	Stap 3:	obs t =	3.105
4	76	70	6	df =	14	Step 4:	crit t =	2.624
5	76	58	18	$t_{\alpha/2}$ =	2.145		p-value =	0.004
6	76	66	10	ME =	6.08	Step 5:	Reject H_0	
7	72	68	4	Lower limit =	2.72			
8	78	52	26	Upper limit =	14.88			
9	82	64	18					
10	64	72	-8					
11	74	74	0					
12	92	60	32					
13	74	74	0					
14	68	72	-4					
15	84	74	10					

There is sufficient evidence to indicate that the Pill does indeed lower blood pressiure over a 6-month time period at a 1% level of significance. We are 98% confident that this reduction of blood pressure is between about 3 to 15 points.

Independent Samples

- Two samples are <u>independent</u> when the method of sample selection is such that those individuals selected for sample 1 do not have any relationship to those individuals selected for sample 2.
- The samples are unrelated, uncorrelated.

Limitations

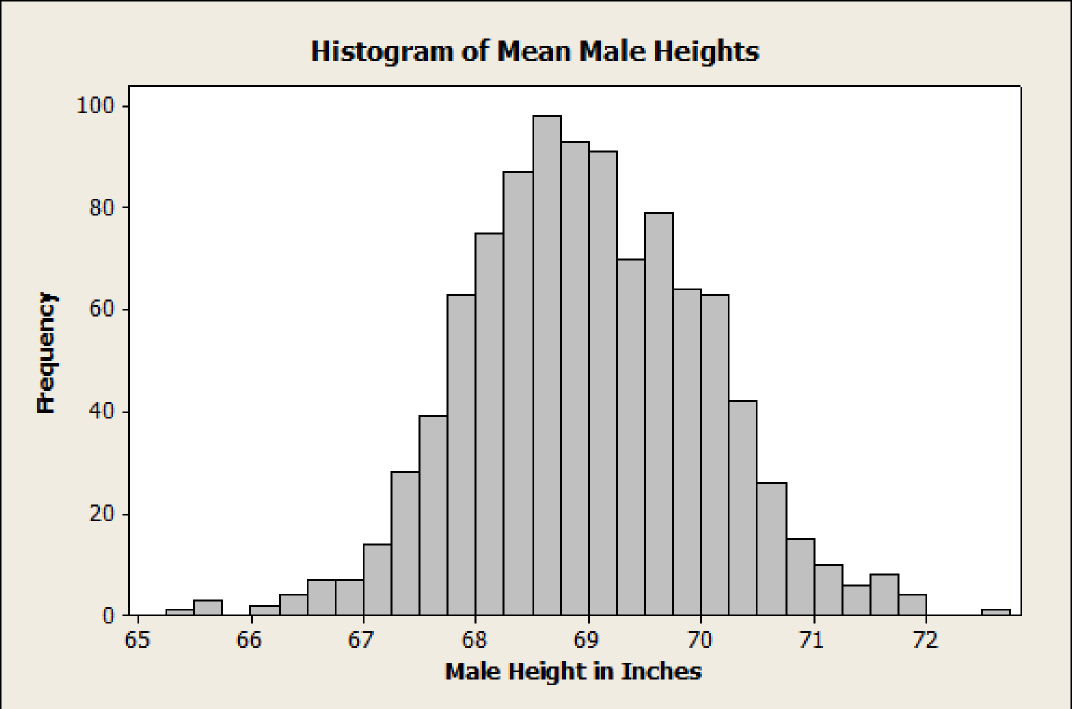
- With one sample tests, we compare a single sample mean to a known population mean (a value we believe to be true in the null hypothesis).
- However, this is unrealistic; we generally have no knowledge about the true population parameters.

Conditions for Inference

- When comparing two means…
- We have two independent SRSs from two distinct populations.
- Both populations are Normally distributed. In practice, it is enough that the distributions have similar shapes and no strong outliers.

Example

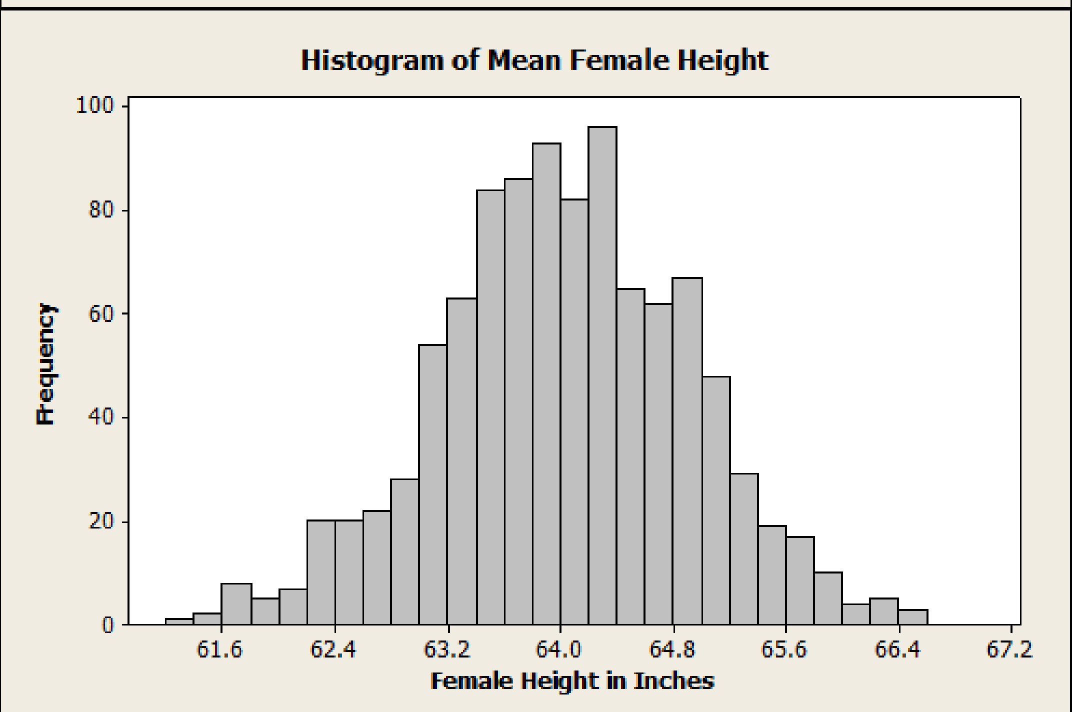
- Heights of males in the US:
  - $X_1 \sim N(69, 3.2)$
- Heights of females in the US:
  - $X_2 \sim N(64, 2.8)$
- I built two sampling distributions in Minitab. Describe them…



Actual $\bar{x}_1 = 68.988$
Actual SD of means = 1.049

This approximates

$\bar{X}_1 \sim N(69, \frac{3.2}{\sqrt{10}} = 1.012)$



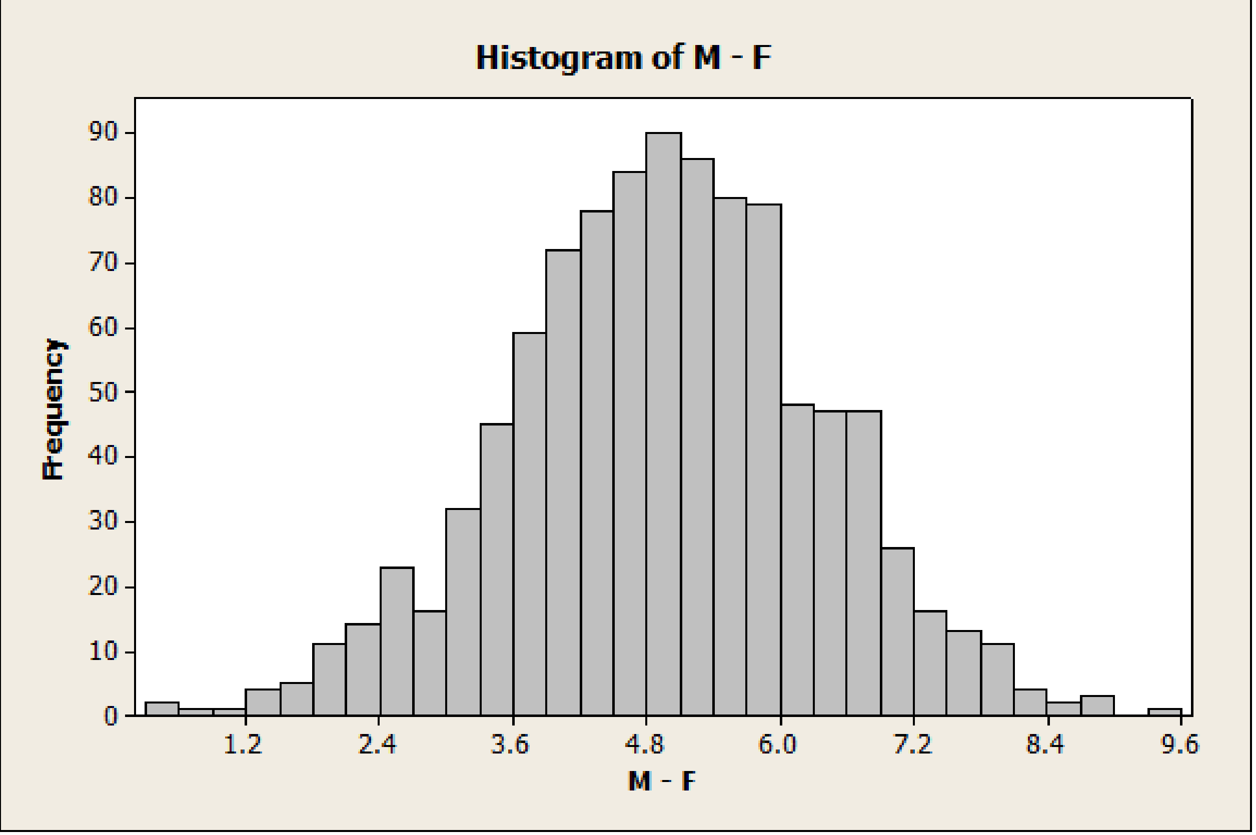
Actual $\bar{x}_2 = 64.023$
Actual SD of means = 0.889

This approximates

$\bar{X}_2 \sim N(64, \frac{2.8}{\sqrt{10}} = 0.885)$

Difference Distribution

- What if we take every male height mean and subtract it from every female height mean, what would the distribution of these differences look like?



This histogram represents the 1000 male mean heights each subtracted from the corresponding female height.

The shape of this distribution is normal with a mean 4.9645 and SD 1.3888.

This approximates the distribution of the difference $\overline{X}_1 - \overline{X}_2$

The mean of $\overline{X}_1 - \overline{X}_2$ is defined to be $\mu_1 - \mu_2 = 5$

The standard deviation of $\overline{X}_1 - \overline{X}_2$ is $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 1.3446$

$\mu_1 - \mu_2$: Large Samples

- Appropriate normality when $n_1 > 30$ and $n_2 > 30$.
- Therefore the random variable (*Xbar – Ybar*) is normally distributed by CLT.
- Use S^2 as an estimator of σ^2 if we are not given but our sample size is big enough.

$$E\left(\bar{X}-\bar{Y}\right)=\mu_1-\mu_2$$

$$Var\left(\bar{X}-\bar{Y}\right)=\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$$

$$SE\left(\bar{X}-\bar{Y}\right)=\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

$$Z=\frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$$

$$CI\left(1-\alpha\right)=\left(\bar{X}-\bar{Y}\right)\pm z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

$\mu_1 - \mu_2$ HT: "Big" Samples

Is there a difference in the mean job satisfaction score between firefighters and office supervisors? (use alpha=0.02)

$$\text{obs } Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

Step 1:	H_0: $\mu_1 - \mu_2 = 0$	
	H_1: $\mu_1 - \mu_2 \neq 0$	
Step 2:	0.02	= α
Step 3:	Fire	Office
n =	226	247
mean =	3.673	3.547
stddev =	0.7325	0.6089
obs z =	2.0241	
Step 4:		
crit z =	±2.3263	-2.3263
p-value =	0.0430	
Step 5:	Do not reject H_0	
since obs z not in rej region		
	since p-value > α	

$\mu_1 - \mu_2$ Confidence Interval: Lg Samples

$$CI(1-\alpha) = \left(\bar{X} - \bar{Y}\right) \pm z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Job satisfaction can be measured on a 4-point scale. For the data given, construct a 95% confidence interval of the difference in job satisfactions scores.

Degree	Scale	Fire-fighter	Office Supervisor	
Very dissatisfied	1	226	247	= *n*
Little dissatisfied	2	3.673	3.547	= *Xbar*
Slightly satisfied	3	0.7235	0.6089	= *S*
Very satisfied	4			
		Std error =	0.0618	
		z =	1.96	
		Margin of error =	0.1211	
		$x_1$bar - $x_2$bar =	0.126	
		Lower limit =	0.0049	
		Upper limit =	0.2471	

Two-Sample *t* Procedures

- *Problem:* We don't know the population standard deviations σ_1 and σ_2 and our sample sizes isn't big enough to assume normality

- *Solution:* Estimate them with s_1 and s_2. The result is called the standard error, or estimated standard deviation, of the difference in the sample means.

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- If we are using t procedures, we need to ask are the variances of the two groups similar? If so we will treat them differently

Looking at differences in means

Ways to analyze the mean difference

- Create a comparative boxplot.
- Run a formal 2 sample HT for $\mu_1 = \mu_2$ (pooling if appropriate), to see if there is a difference.
- If you find a difference, calculate a CI (pooling if appropriate) to estimate it

Equal Variance 2 Independent Sample T Test ($\sigma_1^2 = \sigma_2^2$ or "Pooled")

- Pooled variance:

$$s_{pooled}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- Pooled TS:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$t^* = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- w/ DF:

$$n_1 + n_2 - 2$$

$\mu_1 - \mu_2$: Small Samples ($\sigma_1^2 = \sigma_2^2$) "Pooled"

Assumptions:

1. Both populations are normal
2. Population standard deviations are equal ($\sigma_1 = \sigma_2$), therefore sample variances are pooled.

$$S^2_{pooled} = \frac{\sum_{i=1}^{n_1}\left(X_i - \bar{X}\right)^2 + \sum_{i=1}^{n_2}(Y_i - \bar{Y})^2}{n_1 + n_2 - 2}$$

$$= \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$t^* = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Example 8		
	x	y
	8	2
	5	6
	7	4
	6	7
	9	6
	7	
n =	6	5
mean =	7.0	5.0
variance =	2.00	4.00
pool var =	2.889	

Assuming ($\sigma_1 = \sigma_2$) is appropriate if:

$$\frac{1}{2} \leq \frac{S_1^2}{S_2^2} \leq 2$$

Note: (We could also do a formal HT for equal variances)

Unequal Variance 2 Independent Sample T Test ("Un-pooled")

- Test Statistic:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

- w/ DF:

min(n_1-1, n_2-1)
OR
Satterthwaite

$\mu_1 - \mu_2$: Small Samples ($\sigma_1^2 \neq \sigma_2^2$) "un-pooled"

For an un-pooled 2 sample t test, the formal formula (called Satterthwaite's df correction) is below:

$$df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{S_2^2}{n_2}\right)^2}$$

In practice we might prefer to use:

min{n_1 - 1, n_2 - 1}.

This is called the conservative or Welch's approximation.

Example 8 Data					
	x	y			
	8	2			
	5	6			
	7	4			
	6	7			
	9	6			
	7				
n =	6	5			
mean =	7.0	5.0			
variance =	2.00	4.00			
pool var =	2.889				
	df	t	SE	ME	
pooled	9	2.69	1.03	2.76	% inc
unpooled	4	3.50	1.06	3.72	35%
Satterthwaite	6	2.97	1.06	3.16	14%
num	1.13				
denom	0.18				

CI Formulas

Comparative formulas

$$CI\left(95\%\right)=\left(\bar{X}-\bar{Y}\right)\pm t_{\alpha/2}S_{pooled}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$=\left(\bar{X}-\bar{Y}\right)\pm t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

Independent Samples – 3C

Sample size(s)	If	Assumption	SE^2	df
Large	$n_1, n_2 \geq 30$	CLT (Z)	$s_1^2 / n_1 + s_2^2 / n_2$	does not apply
Small	$n_1, n_2 \geq 2$	$\sigma_1^2 \neq \sigma_2^2$ ("un-pooled")	$s_1^2 / n_1 + s_2^2 / n_2$	$\min(n_1 - 1, n_2 - 1)$ OR Satterthwaite
Small	$n_1, n_2 \geq 2$	$\sigma_1^2 = \sigma_2^2$ ("pooled")	$s_{pooled}^2 \left(1/n_1 + 1/n_2\right)$	$n_1 + n_2 - 2$

$$If\ 0.5 < \frac{S_1^2}{S_2^2} < 2.0,\ \text{then}\ \sigma_1^1 = \sigma_2^2$$

Summary

- Distinguish between independent & matched pair samples.
- Difference of population means under complete randomization.
- Large sample inference with z statistic.
- Small sample inference under 3 conditions with t statistic.

Choosing The 2-Sample Model

- Pairing reduces the df
  - Makes rejection more difficult.
- Pairing eliminates inter-sample variation
  - Gets at the heart of the experiment.
- Paired sampling is best when an appreciable reduction in variability can be anticipated by pairing – when it is appropriate.

Example

- A physical therapist wanted to know whether the mean step pulse of men was less than the mean step pulse of women. She randomly selected 16 men and 16 women to participate in the study. Each subject was required to step up and down onto a 6-inch platform for 3 minutes. The pulse of each subject (in beats per minute) was then recorded. After the data were entered into MINITAB, the following Summary statistics were obtained.

	n	Mean	S
Men	21	112.3	8.45
Women	16	118.3	14.2

Does Octane make a difference?

- Some people believe that higher-octane fuels result in better gas mileage for their car. To test this claim, a researcher randomly selected 11 individuals (and their cars) to participate in a study. Each participant received 10 gallons of gas and drove their car on a closed course that simulated both city and highway driving. The number of miles driven until the car ran out of gas was recorded. A coin flip was used to determine whether the car was filled with 87-octane or 92-octane fuel first, and the driver did not know which type of fuel was in his or her tank.

- Test the claim that the mileage from 87 octane is different than the mileage from 92 octane using $\alpha = 0.10$.

Two Proportions ($p_1 - p_2$)

- The goal here is to conduct a hypothesis test comparing two population proportions.

- Rather than making a claim about what the two population proportions are, we will begin with the belief that they are the same (i.e. there is no difference between them).

Basic Approach

Our comparison of two populations proportions will be based on the statistic

$$\hat{p}_1 - \hat{p}_2 .$$

This statistic is used to estimate the difference between two population proportions

$$p_1 - p_2 .$$

Sampling Distribution of $p_1 - p_2$

- Take a sample of size n_1 from population 1 and a sample of size n_2 from population 2.

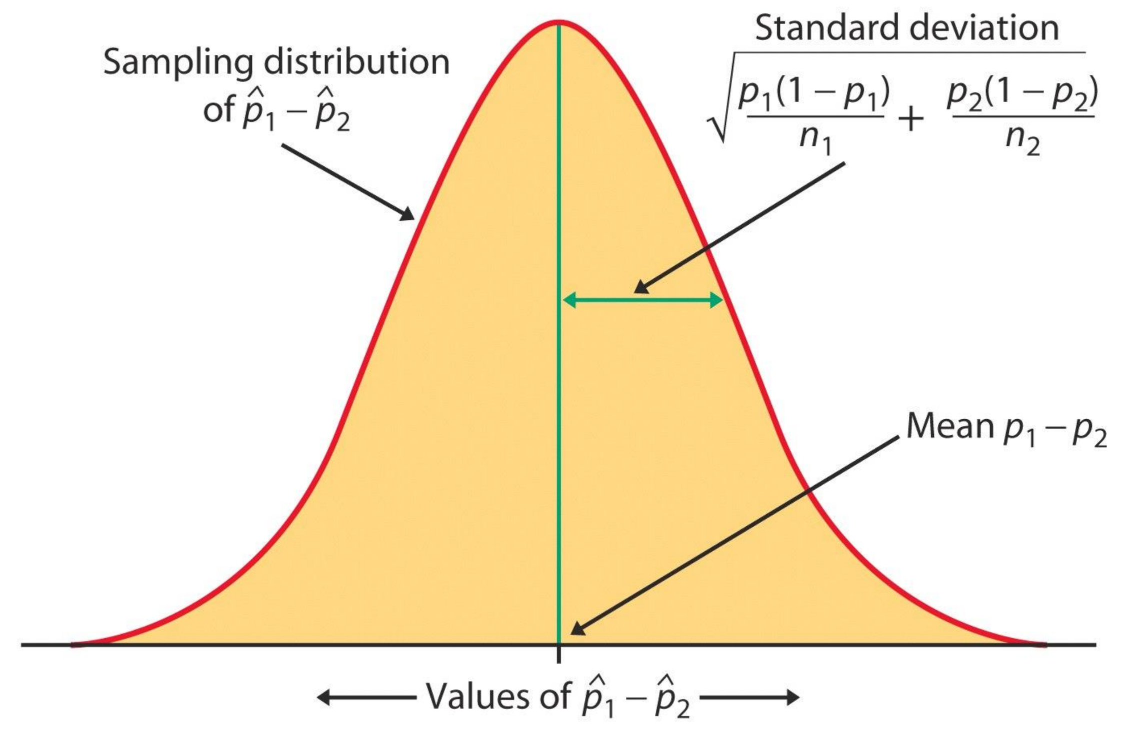
- Obtain the number of successes and divide by the number of trials then:

  $\hat{p}_1 = x_1/n_1 \qquad \hat{p}_2 = x_2/n_2$

- We know that $\hat{p}_1 - \hat{p}_2$ is an unbiased estimator of $p_1 - p_2$

Sampling Distribution

We often need to compare two treatments used on independent samples. We can compute the difference between the two sample proportions and compare it to the corresponding, approximately normal sampling distribution for $\hat{p}_1 - \hat{p}_2$:



CI for the difference between two proportions

- General Form of A CI:
  - Based on the sampling distribution of the Point estimate
  - P.E. $\pm$ ME
    - Where ME = CV * SE
- For two independent SRSs of sizes n_1 and n_2 with sample proportion of successes $\hat{p}_1$ and $\hat{p}_2$ respectively, an approximate confidence interval for $p_1 - p_2$ is

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

Example

- In 2002, a Pew Poll based on a random sample of 1500 people suggested that 43% of Americans approved of stem cell research. In 2009 a new poll of a different sample of 1500 people found that 58% approved.
- Did American opinion really change? Or do the sample proportions differ just by chance? Could the population proportions be the same even though the sample proportions are different?

Sample Proportions

- Even if two population proportions are equal, the sample proportions drawn from these populations are usually different.

- Confidence intervals are one method for determining whether different sample proportions indicate there are “real” differences in the population proportions.

Example

- Construct a 95% confidence interval for the difference between the proportion of people who support stem cell research.

- Based on your confidence interval, is there a difference in the population proportions? Explain.

Solution

- Work (conditions hold)

$$(\hat{p}_{09} - \hat{p}_{02}) \pm z^* \sqrt{\frac{\hat{p}_{09}(1-\hat{p}_{09})}{n_{09}} + \frac{\hat{p}_{02}(1-\hat{p}_{02})}{n_{02}}}$$

$$(0.58 - 0.43) \pm 1.96\sqrt{\frac{0.58(1-0.58)}{1500} + \frac{0.43(1-0.43)}{1500}}$$

$$0.15 \pm 0.035$$

(0.115, 0.185)

Interpretation

- We are 95% confident that the interval 0.115 to 0.185 *captures* the true difference between population proportions describing support of stem cell research.

- Also another thing we want to look for when making a confidence interval for the difference of two parameters

Basic Approach

1. Find a confidence interval for the difference in proportions $p_1 - p_2$.

2. Check to see if 0 is included in the interval.

   - If 0 is in the interval, this suggests the two population proportions might be the same because if $p_1 - p_2 = 0$, then $p_1 = p_2$ and the proportions are the same.

   - If 0 is not in the interval, the confidence interval tells us how much greater one of the proportions might be than the other.

Interpreting Confidence Interval for Two Proportions

When constructing a confidence interval for $p_1 - p_2$:

Interval	Interpretation
Contains 0	The population proportions may be equal
Both values are positive (+,+)	The population proportions are most likely different and $p_1 > p_2$
Both values are negative (–, –)	The population proportions are most likely different and $p_1 < p_2$

Problem

- However, the population proportions used in the formula for standard deviation of the sampling distribution are unknown.
- Methods of estimating the population proportions differ depending on whether you are constructing a confidence interval or conducting a hypothesis test.

Hypothesis Testing

- Assuming the null hypothesis is true, then we can rely on the properties of the sampling distribution to estimate the probability of drawing 2 samples with proportions $\hat{p}_1$ and $\hat{p}_2$ at random.

- When we state H_0: $p_1 = p_2$ (i.e. $p_1 - p_2 = 0$), we find that the best estimate of the standard error for the test statistic uses the pooled sample proportion estimate (instead of using the standard error seen in the confidence interval)

$$\hat{p}_p = \frac{x_1 + x_2}{n_1 + n_2}$$

Test Statistic

- Instead of using $\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ as the standard error in the denominator, replace the estimates of $\hat{p}_1$ and $\hat{p}_2$ with one estimate, called the pooled sample proportion.

$$\hat{p}_p = \frac{x_1 + x_2}{n_1 + n_2}$$

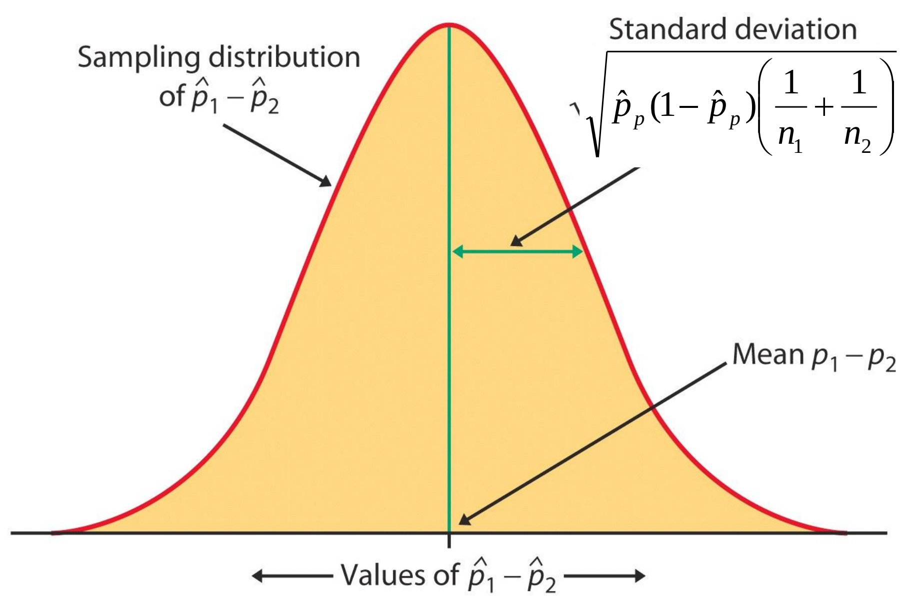
- Thus, the standard error of the test statistic is

$$\sqrt{\hat{p}_p(1-\hat{p}_p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

Hypothetical Sampling Distribution (Assuming Null is True)

- Test statistic:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}_p(1-\hat{p}_p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$



Hypotheses: One-Sided vs. Two-Sided Tests

Hypothesis	Symbols	The Alternative in Words
Two-sided	H_0: $p_1 = p_2$ H_a: $p_1 \neq p_2$	The proportions are different in the two populations.
One-sided (left)	H_0: $p_1 = p_2$ H_a: $p_1 < p_2$	The proportion in population 1 is less than the proportion in population 2.
One-sided (right)	H_0: $p_1 = p_2$ H_a: $p_1 > p_2$	The proportion in population 1 is greater than the proportion in population 2.

Example

- Data are from two independent random samples of male and female seniors on whether they expect to attend grad school. Test to see whether there is a significant difference between genders at $\alpha = 0.01$

	Male	Female	
Yes	18	33	
No	30	19	
Total	48	52	100

Solution

H_0: $p_m = p_f$
H_a: $p_m \neq p_f$

$\alpha = 0.01$

Conditions (first calculate the sample proportions)

$$\hat{p}_m = \frac{18}{48} = 0.375 \qquad \hat{p}_f = \frac{33}{52} = 0.6346$$

Solution

- Pooled proportion:

$$\hat{p}_p = \frac{18+33}{48+52} = \frac{51}{100} = 0.51$$

- Conditions: Random and Independent Samples.

$$48(0.51) = 24.48 \geq 10 \qquad 52(0.51) = 26.52 \geq 10$$

$$48(1-0.51) = 23.52 \geq 10 \qquad 52(1-0.51) = 25.48 \geq 10$$

Solution

- Test Statistic

$$z = \frac{(0.375 - 0.6346) - 0}{\sqrt{0.51(1 - 0.51)\left(\frac{1}{48} + \frac{1}{52}\right)}} = \frac{-0.2596}{0.1001} = -2.59$$

- P–value = 2(Z < -2.59) = 0.0096

Solution

- P–value = 2(Z < –2.59) = 0.0096

- 0.0096 < 0.01

- There is significant evidence of a difference between genders' interest in graduate school.