STA 570 Spring 2011

Lecture 20 *Tuesday, April 5*

- **6.2 Small Sample Inference for Means**
- **7.3 Comparing Means of Two Independent Samples**
- **7.4 Comparing Means of Two Dependent Samples**

Correspondence Between Tests and Confidence Intervals

- Results of confidence intervals and of two-sided significance tests are consistent:
  - Whenever the hypothesized mean μ_0 is not in the confidence interval around $\bar{Y}$, then the *p*-value for testing $H_0 : \mu = \mu_0$ is smaller than 5% (significance at the 5%-level)
  - In general, a (1-alpha)-confidence interval corresponds to a test at significance level alpha
  - This is true for means as well as proportions

Multiple Choice Question I

- A 95% confidence interval for mu is (96,110). Which of the following statements about significance tests for the same data are correct?
  a) In testing the null hypothesis mu=100 (two-sided), P>0.05
  b) In testing the null hypothesis mu=100 (two-sided), P<0.05
  c) In testing the null hypothesis mu=x (two-sided), P>0.05 if x is any of the numbers inside the confidence interval
  d) In testing the null hypothesis mu=x (two-sided), P<0.05 if x is any of the numbers outside the confidence interval

Multiple Choice Question II

- The P-value for testing the null hypothesis mu=100 (two-sided) is P=.001. This indicates
  a) There is strong evidence that mu = 100
  b) There is strong evidence that mu does not equal 100
  c) There is strong evidence that mu > 100
  d) There is strong evidence that mu < 100
  e) If mu were equal to 100, it would be unusual to obtain data such as those observed

Multiple Choice Question II

- The P-value for testing the null hypothesis mu=100 (two-sided) is P=.001. Suppose that in addition you know that the z score of the test statistic was z=3.29. Then

a) There is strong evidence that mu = 100

b) There is strong evidence that mu > 100

c) There is strong evidence that mu < 100

Small Sample Confidence Interval for a Mean

- What if we want to make inference about the population mean, but our sample size is not big enough to meet the minimal sample size requirement n>25 to apply the central limit theorem?
- Confidence intervals are constructed in the same way as before, but now we are using ***t*-values instead of *z*-values**
- For a random sample ***from a normal distribution***, a 95% confidence interval for mu is

$$\bar{Y} \pm t_{0.025} \frac{s}{\sqrt{n}}$$

- where $t_{0.025}$ is a t-score (instead of z-score) from a site like *http://stattrek.com/Tables/T.aspx*
- degrees of freedom are *df=n-1*

Small Sample Hypothesis Test for a Mean

- Assumptions
  - Quantitative variable, random sampling, **population distribution is normal**, ***any* sample size**
- Hypotheses
  - Same as in the large sample test for the mean

$$H_0 : \mu = \mu_0 \text{ vs. } H_1 : \mu \neq \mu_0$$

$$\text{or} \quad H_0 : \mu = \mu_0 \text{ vs. } H_1 : \mu > \mu_0$$

$$\text{or} \quad H_0 : \mu = \mu_0 \text{ vs. } H_1 : \mu < \mu_0$$

Small Sample Hypothesis Test for a Mean

- Test statistic
  - Exactly the same as for the large sample test

$$t_{obs} = \frac{\bar{Y} - \mu_0}{s / \sqrt{n}}$$

- *p*-Value
  - Same as for the large sample test (one-or two-sided), but using an online tool for the *t* distribution
- Conclusion
  - Report p-value and make formal decision

Normality Assumption

- An assumption for the *t*-test is that the population distribution is normal
- In practice, it is impossible to be 100% sure if the population distribution is normal
- It is useful to look at histogram or stem-and-leaf plot (or normal probability plot) to check whether the normality assumption is reasonable

Normality Assumption

- Good news: The *t*-test is relatively ***robust*** against violations of the assumption that the population distribution is normal
- Unless the population distribution is highly skewed, the *p*-values und confidence intervals are fairly accurate

- However: The random sampling assumption must never be violated, otherwise the test results are completely invalid

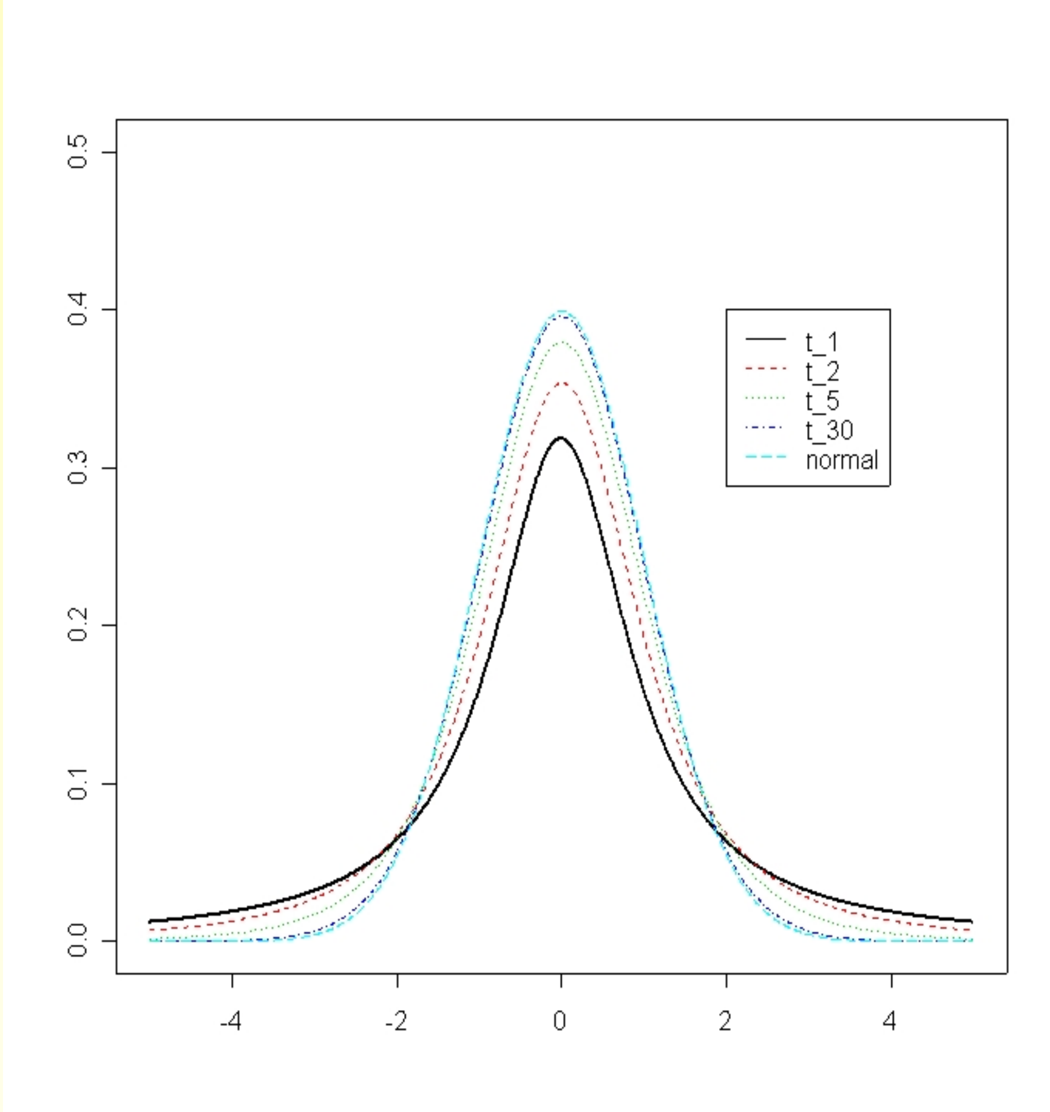
Summary: Small Sample… Significance Test for a Mean

(Assumption: Population distribution is normal)

	One-Sided Tests		Two-Sided Test
Null Hypothesis	$H_0 : \mu = \mu_0$		
Research Hypothesis	$H_1 : \mu < \mu_0$	$H_1 : \mu > \mu_0$	$H_1 : \mu \neq \mu_0$
Test Statistic	$t_{obs} = \dfrac{\overline{Y} - \mu_0}{s / \sqrt{n}}$, degrees of freedom $= n - 1$		
p-value	$P(T_{n-1} < t_{obs})$	$P(T_{n-1} > t_{obs})$	$2 \cdot P(T_{n-1} >\| t_{obs} \|)$

t-Distributions

- The t-distributions are bell-shaped and symmetric around 0
- The smaller the degrees of freedom, the more spread out is the distribution
- *t*-distributions look almost like a normal distribution
- In fact, the limit of the t-distributions is a normal distribution when *n* gets larger



Small Sample Inference for a Mean: Example

- Several food products are fortified by adding nutrients, especially vitamins. A study of production quality takes a random sample from a production run and determines the actual amount of vitamin C.
- The vitamin C content is supposed to be 40mg/100g, but it is suspected that it is actually lower. We would like to test whether vitamin C content conforms to the specifications, or whether it is lower.
- A random sample of size 8 has sample mean 22.5 and sample standard deviation 7.19.
- Can we conclude that the vitamin C content for this run is too low?

Statistical Methods for One Sample Summary I

- Testing the Mean
  - Large sample size (*30 or more)*:

    Use the large sample test for the mean

    (*z*-Scores, normal distribution)
  - Small sample size:

    Check whether the data is not very skewed

    Use the *t* test for the mean

    (*t*-Scores, *t* distribution)

Statistical Methods for One Sample Summary II

- Testing the Proportion
  - Large sample size (*np>5, n(1-p)>5)*:

    Use the large sample test for the proportion

    (*z*-Scores, normal distribution)
  - Small sample size:

    Binomial distribution

    (not covered in class)

Statistical Methods for Two Samples
Independent vs. Dependent Samples

- Two ***Independent*** Samples
  - Different subjects in the different samples
  - The two samples constitute independent samples from two subpopulations
- Two ***Dependent*** Samples
  - Natural matching between an observation in one sample and an observation in the other sample
  - For example, two measurements on the same subject
- Data sets with dependent samples require different statistical methods than data sets with independent samples

Independent Samples

Confidence Interval for Difference of Two Means

- The large sample (both samples sizes at least 20) confidence interval for $\mu_2 - \mu_1$ is

$$(\bar{Y}_2 - \bar{Y}_1) \pm z \cdot \sqrt{\frac{s_1^{\ 2}}{n_1} + \frac{s_2^{\ 2}}{n_2}}$$

Significance Test for the Difference of Two Means

- For large samples (both sample sizes at least 20),
- the significance test for the null hypothesis that both population means are equal,

$H_0 : \mu_1 = \mu_2$ which is equivalent to $H_0 : \mu_2 - \mu_1 = 0$, is

$$z_{obs} = \frac{\text{parameter estimate} - \text{parameter under null hypothesis}}{\text{standard error of estimator}}$$

$$= \frac{\bar{Y}_2 - \bar{Y}_1}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Significance Test for the Difference of Two Means

- Most commonly, the alternative hypothesis is two-sided
- Then, the P-value is the two-tail probability of “anything at least as extreme as observed”
- The probability is taken from the z-score online tool (normal distribution)

Significance Test for the Difference of Two Means: Example (contd.)

- Do Americans consume more caramel apples in the Fall than five years ago?
- 2002 survey: 350 subjects had sample mean 2.3 caramel apples, sample standard deviation 1.2 caramel apples
- 2007 survey: 485 subjects, sample mean 2.8 caramel apples, sample standard deviation 1.4 caramel apples
- *Set up the hypotheses of a significance test to analyze whether the population means differ in 2002 and 2007*
- *Construct the test statistic*
- *Report and interpret the P-value*
- *Is the result consistent with the confidence interval?*

Correspondence Between Confidence Intervals and Tests

- As before, confidence intervals and tests are equivalent in the sense that
  - If the two-sided test has a P-value less than 0.01 (significant at level alpha=0.01),
  - then the 99% confidence interval does not contain the null hypothesis value
  - The same is true for alpha=0.05 and 95% confidence intervals,
  - or any alpha-level test and the corresponding 1-alpha confidence interval

Correspondence Between Confidence Intervals and Tests

- As before, confidence intervals are preferred over tests because they provide a range of plausible values
- In the two-sample case, it is a range of plausible values for the difference between the population means

Small Sample Inference

- How to make inference about the difference between two population **means** when the sample sizes are small?
- $n_1 < 20$ or $n_2 < 20$
- There is a method that works for every sample size (if $n_1 > 1$ and $n_2 > 1$): ***two sample t test for independent data / unpaired samples***
- Assumption: Both samples come from normal population distributions
- P-value calculation requires the t distribution
- Formulas too complicated to be practical "by hand"
- http://www.fon.hum.uva.nl/Service/Statistics/2Sample_Student_t_Test.html
- http://graphpad.com/quickcalcs/ttest1.cfm

What if all the assumptions fail?

- If the samples are *small*

  and

- the assumption of *normal population is not justifiable*,
- one should use the ***Wilcoxon-Mann-Whitney test for independent samples***
- http://www.fon.hum.uva.nl/Service/Statistics/Wilcoxon_Test.html

- However:
  - The random sampling assumption must never be violated
  - The two samples must be independent for this test

Summary
Selecting the Correct Method for Comparing Means in of Two Populations

- Independent Samples
  - Large Samples
  - Small Samples
    - Normal populations
      (sample histograms appear symmetric)
    - Not normal populations
- Dependent Samples (Matched Pairs)
  - Large Samples
  - Small Samples

Comparing Dependent Samples

- Comparing Dependent Means
  - Example: Special exam preparation training for STA 570 students
    - Choose 10 students and compare before vs. after special training.
    - Better design: Choose *n=10* pairs of students such that the students matched in any given pair are very similar with regard to previous exam/quiz results

      For each pair, one student is randomly selected for the special training (group 1)

      The other student receives normal instruction (group 2)

Comparing Dependent Samples

- Example, contd.
  - Consider the first (simple) design
  - For the *i*th student, define

    D_i = *Score after special training – Score before*

Comparing Dependent Samples

- The sample mean of the difference scores,

$$\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i = (D_1 + D_2 + \cdots + D_n)/n$$

- is an estimator for the difference between the population means
- We can now use exactly the same methods as for one sample
- Just replace the Y_i by D_i
- *When the data set is small, we need the assumption that the population distribution of the score differences is normal*

Comparing Dependent Samples

- The *small sample* confidence interval is

$$\bar{D} \pm t_{n-1} \frac{s_D}{\sqrt{n}} = \bar{D} \pm t_{n-1} \frac{\sqrt{\frac{\sum_{i=1}^{n} (D_i - \bar{D})^2}{n-1}}}{\sqrt{n}}$$

- When n exceeds 30, we *may* use the z-scores instead of the t-scores

Comparing Dependent Samples

- The small sample test statistic for testing difference in the population means is

$$t_{obs} = \frac{\bar{D}}{s_D / \sqrt{n}}$$

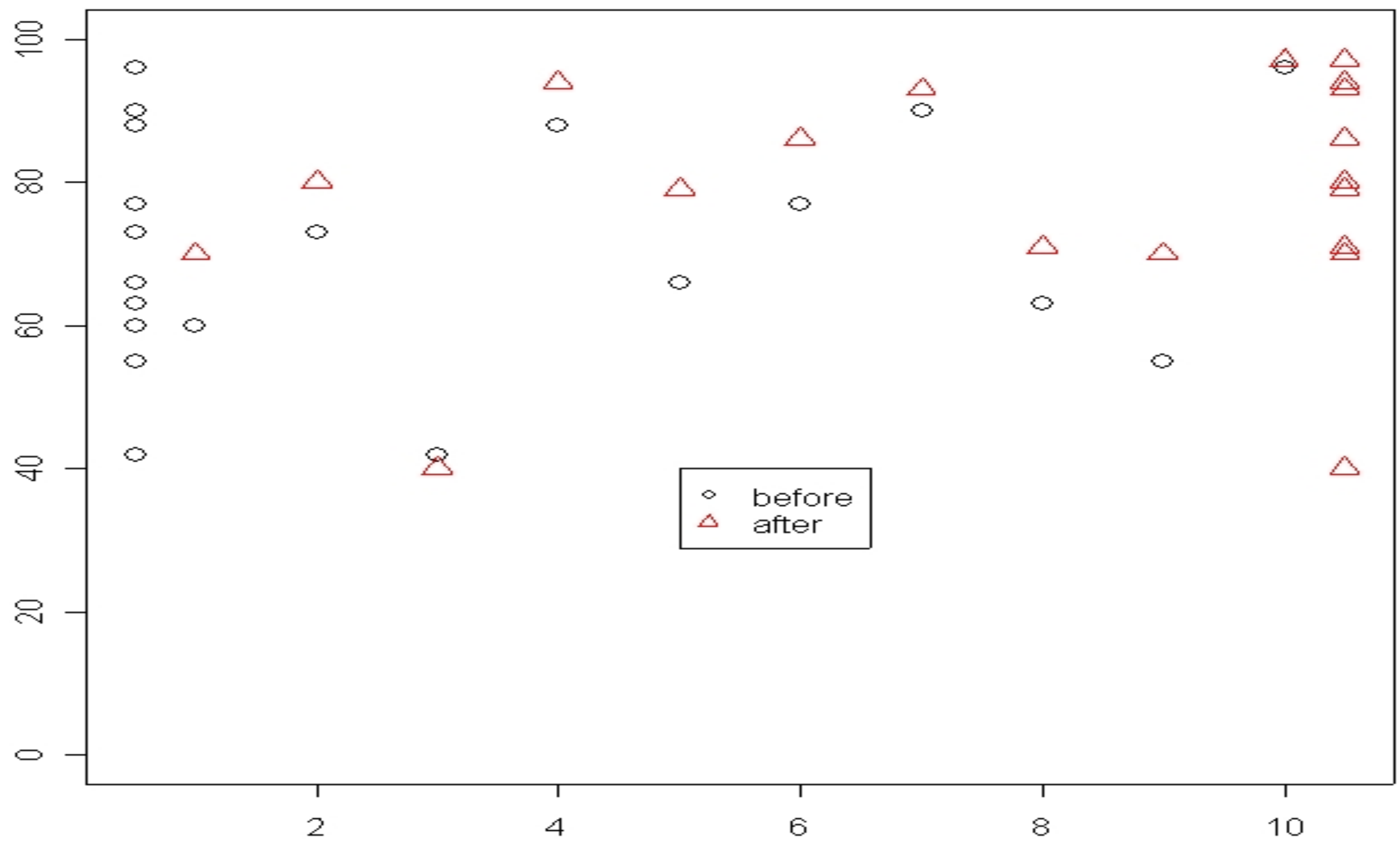
- For small n, use the t-distribution with $df=n-1$
- When n exceeds 30, we *may* use the z-scores (normal distribution) instead of the t-distribution

Comparing Dependent Samples: Example

- Ten STA 570 students take a statistics test both before and after undergoing intensive training using the online study tools
- Then, the scores for each student are paired, as in the following table

Student	1	2	3	4	5	6	7	8	9	10
Before	60	73	42	88	66	77	90	63	55	96
After	70	80	40	94	79	86	93	71	70	97

Comparing Dependent Samples: Example



Comparing Dependent Samples: Example (contd.)

Student	1	2	3	4	5	6	7	8	9	10
Before	60	73	42	88	66	77	90	63	55	96
After	70	80	40	94	79	86	93	71	70	97

a) Find a point estimator for the difference of population means.

b) Calculate and interpret the P-value for testing whether the mean change equals 0

c) Compare the mean scores after and before the training course by constructing and interpreting a 90% confidence interval for the population mean difference

Comparing Dependent Samples: Example (contd.)

Output from Statistical Software Package SAS

N	10
Mean	7
Std Deviation	5.24933858

Tests for Location: Mu0=0

Test	-Statistic-		-----p Value------	
Student's t	**t**	**4.216901**	**Pr > |t|**	**0.0022**
Sign	M	4	Pr >= |M|	0.0215
Signed Rank	S	25.5	Pr >= |S|	0.0059

Comparing Dependent Samples: Example (contd.)
Using the Wrong Method (for independent samples)

Output from SAS

The TTEST Procedure

Variable	Sample	N	Mean	Std Dev	Std Err
score	1	10	71	17.068	5.3975
score	2	10	78	16.773	5.3041
score	Diff (1-2)		-7	16.921	7.5675

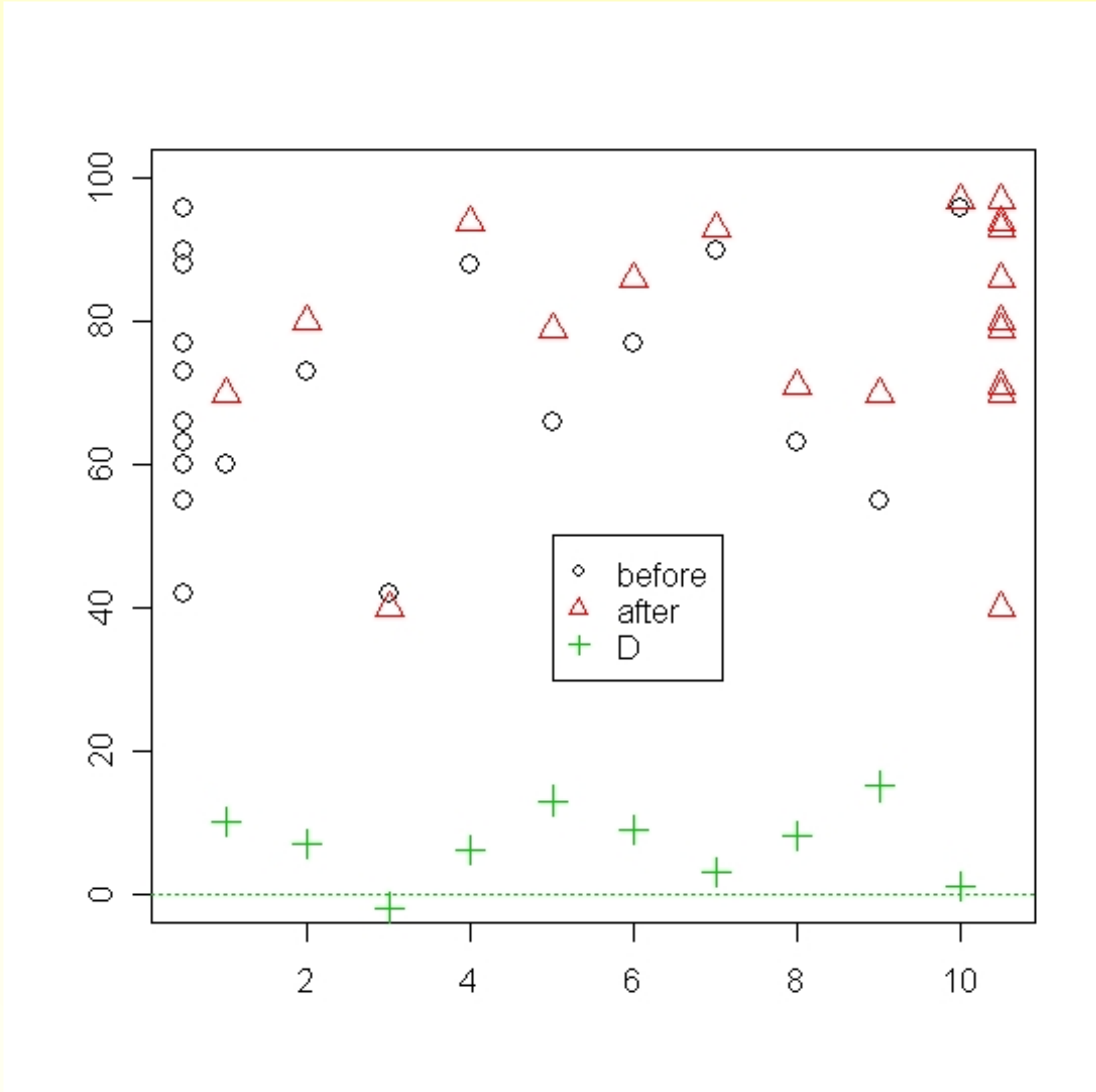
T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
score	**Pooled**	**Equal**	**18**	**-0.93**	**0.3672**
score	**Satterthwaite**	**Unequal**	**18**	**-0.93**	**0.3672**

Comparing Dependent Samples: Reducing variability

- Variability in the difference scores may be less than the variability in the original scores
- This happens when the scores in the two samples are strongly associated
- Subjects who score high before the intensive training also tend to score high after the intensive training

Comparing Dependent Samples: Example



Comparing Proportions in Two (Large) Independent Samples

- Response variable: Qualitative
- Inference about the population proportions that are classified in a particular category of the response variable
- Explanatory variable: The variable that defines the group membership.
- Are the proportions different for the two groups?

$$p_2 - p_1 = 0\,?$$

- Confidence interval for the difference
- Significance test about whether the difference equals zero

Comparing Two Proportions: Examples

1. Gender Gap in Party Identification

   Explanatory variable: Male/Female

   Response variable: Party Identification

   Is the proportion of Republicans different between male and female voters?

2. Explanatory variable: Treatment (Drug / Placebo)

   Response variable: Pain (Yes/No)

   Is the proportion who suffers from pain different for the two treatment groups?

Confidence Interval for the Difference of Two Proportions

- Here, large sample means at least five observations in each category of interest in each of the samples
- The large sample confidence interval for $p_2 - p_1$ is

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

Confidence Interval for the Difference of Two Proportions: Example

- Famous five-year study on the effect of Aspirin to reduce heart disease
- Study subjects: 22,071 male physicians
- Every other day, participants took either an aspirin tablet or a placebo
- 11,034 who took placebo: 189 had a heart attack
- 11,037 who took aspirin: 104 had heart attacks
- *Estimate the heart attack rates for the two groups.*
- *Construct a 95% confidence interval to compare them.*
- *Interpret.*

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

Significance Test for the Difference of Two Proportions

- The large sample (see above) significance test for the null hypothesis that both population proportions are equal,

$$H_0: p_1 = p_2 \quad \text{which is equivalent to } H_0: p_2 - p_1 = 0, \text{is}$$

$$z_{obs} = \frac{\text{estimate } - \text{null hypothesis value}}{\text{standard error of estimator}}$$

$$= \frac{\widehat{p}_2 - \widehat{p}_1}{\sqrt{\widehat{p} \cdot (1 - \widehat{p}) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$\widehat{p}$ is the "pooled" proportion of the total sample (both samples together) in the category of interest

Significance Test for the Difference of Two Proportions

- As above, most commonly, the alternative hypothesis is two-sided
- Then, the P-value is the two-tail probability of “anything at least as extreme as observed”
- The probability is taken from a z-score applet (normal distribution)

Significance Test for the Difference of Two Proportions: Example

- Effect of Aspirin to reduce heart disease
- Study subjects: 22,071 male physicians
- Every other day, participants took either an aspirin tablet or a placebo
- 11,034 who took placebo: 189 had a heart attack
- 11,037 who took aspirin: 104 had heart attacks
- *Test whether the rates are significantly different. Report the P-value and interpret.*

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

Confidence Interval and Hypothesis Test for the Difference of Two Proportions: Example

- A serious side effect of many allergy medicines is that they cause drowsiness, which makes them dangerous for industrial workers. In recent years, different nondrowsy allergy medicines have been developed. One of them was Hismanal. The manufacturer claimed that this was the first once-a-day nondrowsy allergy medicine.
- The nondrowsiness claim is based on a clinical experiment in which 1,604 patients were given Hismanal, and 1,109 patients were given a placebo.
- Of the first group, 7.1% reported drowsiness; of the second group, 6.4% reported drowsiness.
- Do these results allow us to infer at the 5% significance level that Hismanal's claim is false?
- Report a 95% confidence interval for the difference of the proportions.

Contingency Table

- The proportions are usually listed in a table called ***contingency table***
- How are the outcomes of the response variable *contingent* on the category of the explanatory variable
- Each row represents a category of one variable, and each column represents a category of the other variable
- The cells of the table contain frequency counts for the four possible combinations of outcomes

	Drowsy	Not Drowsy	
Hismanal			
Placebo			

Summary

Large Sample Significance Test for the Difference of Two Proportions

	One-Sided Tests		Two-Sided Test
Null Hypothesis	$H_0 : p_1 = p_2$		
Research Hypothesis	$H_1 : p_2 < p_1$	$H_1 : p_2 > p_1$	$H_1 : p_1 \neq p_2$
Test Statistic	$z_{obs} = \dfrac{\hat{p}_2 - \hat{p}_1}{\sqrt{\hat{p} \cdot (1 - \hat{p}) \cdot \left(\dfrac{1}{n_1} + \dfrac{1}{n_2} \right)}}$		
p-value	$P(Z < z_{obs})$	$P(Z > z_{obs})$	$2 \cdot P(Z > \lvert z_{obs} \rvert)$