STA 291
Summer 2008

Lecture 12

Multiple Choice Question I

- The P-value for testing the null hypothesis μ=100 (two-sided) is P=.001. This indicates
  a) There is strong evidence that μ = 100
  b) There is strong evidence that μ does not equal 100
  c) There is strong evidence that μ > 100
  d) There is strong evidence that μ < 100
  e) If μ 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 μ=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 μ = 100
  b) There is strong evidence that μ > 100
  c) There is strong evidence that μ < 100

Summary: Large Sample (n > 30)
Significance Test for a Population Mean

	One-Sided Tests		Two-Sided Test
Null Hypothesis	$H_0 : m = m_0$		
Research Hypothesis	$H_1 : m < m_0$	$H_1 : m > m_0$	$H_1 : m \neq m_0$
Test Statistic	$z_{obs} = \frac{\bar{X} - m_0}{s / \sqrt{n}}$		
p-value	$P(Z < z_{obs})$	$P(Z > z_{obs})$	$2 \cdot P(Z > \vert z_{obs} \vert)$

Summary: Large Sample $n > 5/p_0$ and $n > 5/(1-p_0)$
Significance Test for a Population Proportion

	One-Sided Tests		Two-Sided Test
Null Hypothesis	$H_0 : p = p_0$		
Research Hypothesis	$H_1 : p < p_0$	$H_1 : p > p_0$	$H_1 : p \neq p_0$
Test Statistic	$z_{obs} = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$		
p-value	$P(Z < z_{obs})$	$P(Z > z_{obs})$	$2 \cdot P(Z > \vert z_{obs} \vert)$

12.1 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>30 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 μ is

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

- where $t_{0.025}$ is a t-score (instead of z-score) from Table B4 (p.B-9) or better, from a site like *surfstat*:
- http://www.anu.edu.au/nceph/surfstat/surfstat-home/tables/t.php
- 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

$H_0 : m = m_0$ vs. $H_1 : m \neq m_0$

or $H_0 : m = m_0$ vs. $H_1 : m > m_0$

or $H_0 : m = m_0$ vs. $H_1 : m < m_0$

Small Sample Hypothesis Test for a Mean

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

$$t_{obs} = \frac{\bar{X} - m_0}{s / \sqrt{n}}$$

- *p*-Value
  - Same as for the large sample test (one-or two-sided), but using the table/online tool for the *t* distribution
  - Table B4 only provides very few values
- Conclusion
  - Report p-value and make formal decision

Small Sample Hypothesis Test for a Mean: Example

- A study was conducted of the effects of a special class designed to improve children/s verbal skills
- Each child took a verbal skills test twice, both before and after a three-week period in the class
- X=2nd exam score – 1st exam score
- If the population mean for X, E(X)=µ equals 0, the class has no effect
- Test the null hypothesis of no effect against the alternative hypothesis that the effect is positive
- Sample (*n=4*): 3,7,3,3

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

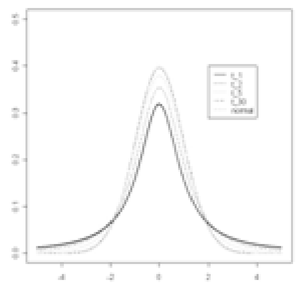
- Good news: The *t*-test is relatively ***robust*** against violations of the assumption that the population distribution is normal
- Unless the population distribution is highly skewed, the *p*-values und confidence intervals are fairly accurate
- However: The random sampling assumption must never 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	$H_0 : m = m_0$		
Research Hypothesis	$H_1 : m < m_0$	$H_1 : m > m_0$	$H_1 : m \neq m_0$
Test Statistic	$t_{obs} = \frac{\bar{X} - m_0}{s / \sqrt{n}}$, degrees of freedom = $n-1$		
p-value	$P(T_{n-1} < t_{obs})$	$P(T_{n-1} > t_{obs})$	$2 \cdot P(T_{n-1} > \lvert t_{obs} \rvert)$

t-Distributions (Section 8.4)

- 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



Statistical Methods for One Sample Summary I

- Testing the Mean
  - Large sample size (*30 or more)*:
    Use the large sample test for the mean
    (Table B3, normal distribution)
  - Small sample size:
    Check whether the data is not very skewed
    Use the *t* test for the mean
    (Table B4, *t* distribution)

Statistical Methods for One Sample Summary II

- Testing the Proportion
  - Large sample size ($np_0>5$, $n(1-p_0)>5$):
    Use the large sample test for the proportion
    (Table B3, normal distribution)
  - Small sample size:
    Binomial distribution

13.1 Comparison of **Two** Groups Independent Samples

- Two ***Independent*** Samples
  - Different subjects in the different samples
  - Two subpopulations (e.g., male/female)
  - The two samples constitute independent samples from two subpopulations
  - For example, stratified samples

Comparison of Two Groups Dependent Samples

- Two ***Dependent*** Samples
  - Natural matching between an observation in one sample and an observation in the other sample
  - For example, two measurements at the same subject (left and right hand, performance before and after training)
- Data sets with dependent samples require different statistical methods than data sets with independent samples

Comparing Two Means (Large Samples)

- Response variable: Quantitative
- Inference about the population means for the two groups, and their difference

$$m_2 - m_1$$

- Confidence interval for the difference
- Significance test about the difference

Confidence Interval for the Difference of Two Means

- The large sample (both samples sizes at least 30) confidence interval for $\mu_2 - \mu_1$ is

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

Confidence Interval for the Difference of Two Means: Example

- In a 1994 survey, 350 subjects reported the amount of turkey consumed on Thanksgiving day. The sample mean was 3.1 pounds, with standard deviation 2.3 pounds
- In a 2006 survey, 1965 subjects reported an average amount of consumed Thanksgiving turkey of 2.8 pounds, with standard deviation 2.0 pounds
- *Construct a 95% confidence interval for the difference between the means in 1994 and 2006.*
- *Is it plausible that the population mean was the same in both years?*

Significance Test for the Difference of Two Means

- For large samples (both samples sizes at least 30),
- 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

$$z_{obs} = \frac{\text{parameter estimate} - \text{parameter under null hypothesis}}{\text{standard error of estimator}} = \frac{\bar{X}_2 - \bar{X}_1}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

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 Table B3 (normal distribution)

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

- 1994 survey, 350 subjects, sample mean 3.1 pounds, sample standard deviation 2.3 pounds
- 2006 survey, 1965 subjects, sample mean 2.8 pounds, sample standard deviation 2.0 pounds
- *Set up the hypotheses of a significance test to analyze whether the population means differ in 1994 and 2006*
- *Construct the test statistic*
- *Report and interpret the P-value*

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 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?
- $n_1 < 30$ or $n_2 < 30$
- There is a method that works for every sample size (well: $n_1 > 1$ 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, with degrees of freedom n_1+n_2-2
- Formulas too complicated to be used "by hand"
- http://www.fon.hum.uva.nl/Service/Statistics/2Sample_Student_t_Test.html
- http://graphpad.com/quickcalcs/ttest1.cfm
- However, a (1-a) confidence interval can be found to test a hypothesis at the a significance level.

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

Paired Experiment: focus on the differences

- One subject contributes two results, we often can focus on the difference of the two from the same subject.
- Sometimes there is no alternative…… how long mice live before cancer kill. Same mouse cannot be used twice. Strength of the shipping packaging …… test of strength would destroy the package

13.3 Comparing Dependent Samples

- Comparing Dependent Means
  - Example: Special exam preparation training for STA 291 students
  - 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.
  - "Matched Pairs" plan
  - Each sample (Group 1, Group 2) has the same number of observations
  - Each observation in one sample pairs with an observation in the other sample
  - For the *i*th pair, let

  D_i = Score for student who receives special training – Score for student who receives normal instruction

Comparing Dependent Samples

- The sample mean of the difference scores,

$$\bar{X}_D = \frac{\sum_{i=1}^{n} D_i}{n} = \frac{D_1 + D_2 + \cdots + D_n}{n}$$

- 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 X_i by D_i
- *When the data set is small, we need the assumption that the population distribution of the difference scores is normal*

Comparing Dependent Samples

- The *small sample* confidence interval is

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

- When *n* exceeds 30, we can 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

$$t = \frac{\bar{X}_D}{s_D / \sqrt{n}}$$

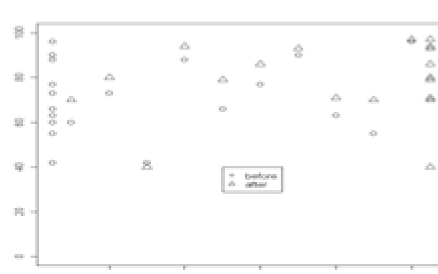
- For small *n*, use the *t*-distribution with df=n-1
- When *n* exceeds 30, we can use Table B3 (normal distribution) instead of the *t*-distribution

Comparing Dependent Samples: Example

- Ten STA 291 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

13.5 Comparing Two Proportions (Large Samples)

- Response variable: Qualitative
- Inference about the population proportions that are classified in a particular category of the response variable
- 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,

$H_0: p_1 = p_2$ which is equivalent to $H_0: p_2 - p_1 = 0$, is

$$z = \frac{\text{estimate} - \text{null hypothesis value}}{\text{standard error of estimator}} = \frac{\hat{p}_2 - \hat{p}_1}{\sqrt{\hat{p}\cdot(1-\hat{p})\cdot\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$\hat{p}$ 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 Table B3 (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.*

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