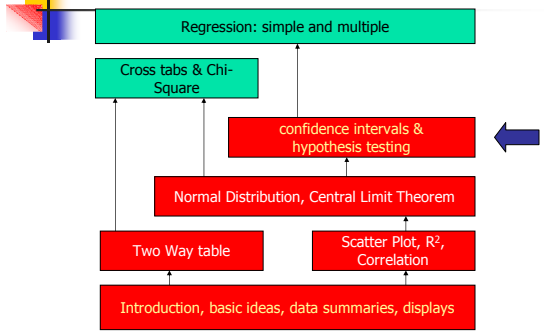
PUB – POS 316
Week 10

Inference for the mean and the proportion

Navid Ghaffarzadegan
navidg@gmail.com
Last updated – March 27, 10

Course Road Map



Introduction

- Inference for the mean
  - Inference for the mean of a population
  - t-distribution
  - Comparing of two means
- Inference for the proportion
  - Inference for single proportion
  - Comparing two proportions

Introduction

- We talked about margin of error and confidence intervals:
  - Example: Average price of coffee in Albany:
  - $ 1.85 ±$0.35 (with 95% confidence)

$$x = \mu \pm z.s \Rightarrow x = \mu \pm z.\frac{\sigma}{\sqrt{n}}$$

- We also talked about hypothesis testing.

Introduction

- A typical question for hypothesis testing:
  - You're an analyst for Ford. You want to find out if the average miles per gallon of Escorts is at least 32 mpg. Similar models have a standard deviation of 3.8 mpg. You take a sample of 60 Escorts & compute a sample mean of 30.7 mpg.
  - At the 0.05 level, is there evidence that the miles per gallon is less than 32?

(source: Carnegie Mellon University, 90-711, Empirical Methods)

Introduction

- A typical question for hypothesis testing:
  - You're an analyst for Ford. You want to find out if the average miles per gallon of Escorts is at least 32 mpg. Similar models have a standard deviation of 3.8 mpg. You take a sample of 60 Escorts & compute a sample mean of 30.7 mpg.
  - At the 0.05 level, is there evidence that the miles per gallon is at least 32?

(source: Carnegie Mellon University, 90-711, Empirical Methods)

Introduction

- In many situations:
  - We do not know the standard deviation in our population
    - How are we going to find z?

$$z = \frac{x-\mu}{\sigma_{\bar{x}}} = \frac{x-\mu}{\frac{\sigma}{\sqrt{n}}}$$

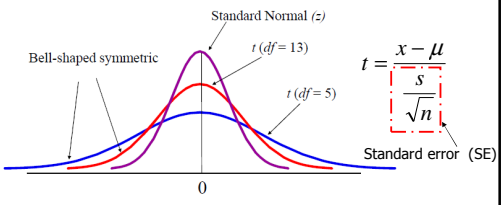
  - In many situations our sample size is small (some times even less than 30).
- What if we look at the standard deviation in our sample?

$$\frac{x-\mu}{\frac{s}{\sqrt{n}}}$$

It turned out that this does not follow z-distribution

But follows another distribution, called **Student's t distribution** AKA **t-distribution**

t-distribution



- Degree of freedom = n-1 = (sample size-1)
- Table

t-distribution

- What are we going to do with the t-distribution?
- When we don't have the standard deviation in population (and the sample size is small), replace all z's with t's! and use the t-table rather than z-table.
- Let's work with the table:
- What is the t-statistics for under tail probability of 0.025? When the sample size is 10? What if our sample size is 60? 100?
- t-distribution converges to z-distribution when the sample size is large.

t-distribution

- Example: We take a sample of 10 students in the class and ask how many hours a day they study. The results are:
- 1.5,2,0.5, 3, 4, 3, 3.5, 2.5, 2.5, 2
- What is the mean?
- What can we say about the population (all students in UAlbany)? We should find the margin of error.

t-distribution

$$x = \mu \pm z.\frac{\sigma}{\sqrt{n}} \qquad z = \frac{x-\mu}{\sigma_{\bar{x}}} = \frac{x-\mu}{\frac{\sigma}{\sqrt{n}}}$$

$$x = \mu \pm t.\frac{s}{\sqrt{n}} \qquad t = \frac{x-\mu}{\frac{s}{\sqrt{n}}}$$

t-distribution

- Example: In the UK people listen 8.3 hours per month to music. In the US, we take a sample of 10 and find that in this sample in average people listen to music for 5 hours per month. The standard deviation in this sample was 3.63.
- What is the margin of error of our study.
- Compare our results for the US with the UK.

t-distribution

- Example: Suppose in the same question, that two years ago Americans listened 4.7 hours per month to music. We want to test to see if listening to music is significantly increased or not.
- State and test the hypotheses.

t-distribution

- Example: Based on a sample of 16 people at UAlbany, we know that the average income per month is $620. How should we find the margin of error in our study?
- Assume the standard deviation in our sample is $50.
- In Albany, a student in order to survive (!), needs at least $585. Can we state at the 5% significance level that the average income of UAlbany students is higher than $585?

t-distribution

- t-test in an Excel sheet

t-distribution

- *Many times we want to compare two samples.*

t-distribution

- We conduct a study in a university to see the average level of satisfaction of students (scale 1-10). 10 male and 12 female students participate. The average result is 4.5 and 5.6 respectively. Assume the standard deviation in the samples are 0.5 and 2 respectively. We would like to see if female students are happier with their school in comparison to male students.
- Two kinds of questions you may see:
- Q1: Find the margin of error for each group; And compare the results.
- Q2: State Hypotheses, and test them..

t-distribution

- We state the hypotheses as following and look at the difference between the means.
- H_0: $\mu_1=\mu_2$
- H_a: $\mu_1\neq\mu_2$

$$\frac{s}{\sqrt{n}} \Rightarrow \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

$$t=\frac{x-\mu}{\frac{s}{\sqrt{n}}} \Rightarrow t=\frac{(\overline{x_1}-\overline{x_2})-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

df=minimum(df1, df2)

You don't need to memorize it!

Introduction

- Inference for the mean
  - Inference for the mean of a population
  - t-distribution
  - Comparing of two means
- Inference for the proportion
  - Inference for single proportion
  - Comparing two proportions

Inference for the proportion

$$x = \mu \pm z.\frac{\sigma}{\sqrt{n}}$$

$$x = \mu \pm z.\sqrt{\frac{p(1-p)}{n}}$$

Or you can write the second one as: $x = \hat{p} \pm \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

The same! As long as you know what is the equation above.

Inference for the proportion

- If you don't have enough data:

$$x = \mu \pm z.\frac{\sigma}{\sqrt{n}}$$ Use t-distribution instead of z-distribution

$$x = \mu \pm z.\sqrt{\frac{p(1-p)}{n}}$$ NO way. Gather more data!!

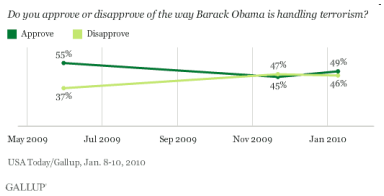
Confidence intervals for proportion

From the survey that we conducted in the class, n=36 (we added data from another class), we know that 60% of the students want to work for the government.

Report your estimation of the true parameter in the public administration department students based on our sample with 95% confidence.

An Example from the first class

- Obama Approval on Terrorism Up to 49%
- http://www.gallup.com/poll/125033/Obama-Approval-Terrorism-Up-49.aspx



Comparison of two proportions

How to compare two proportions?

- Again: 1) compare with looking at confidence intervals and if they don't overlap you can say one is bigger than the other.
- 2) state your hypotheses, and test it.

Comparison of two proportions

- H0: p1=p2
- Ha: p1≠p2

Or

Ha: p1>p2 or Ha: p2>p1

- Variance of A-B = variance of A + variance of B

$$\text{var} = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

- Variance of (p1-p2) = Variance of (p1)+ Variance of (p2)

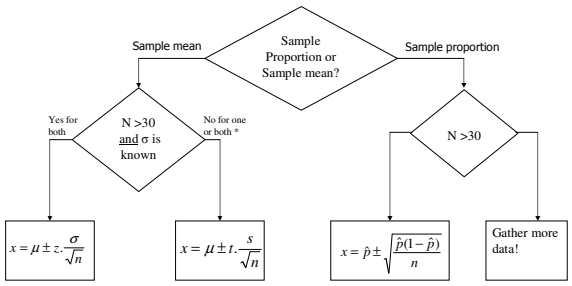
$$SE_D = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Comparison of two proportions

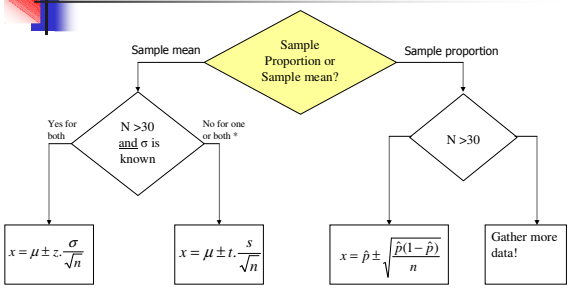
Binge drinkers: Based on a study which involved 5,348 male, and 8,471 female, we got that 0.260 of male and 0.206 of female are frequent binge drinkers. We would like to know if the proportion of men who are frequent binge drinkers is higher than women.

- State the hypotheses and test them.

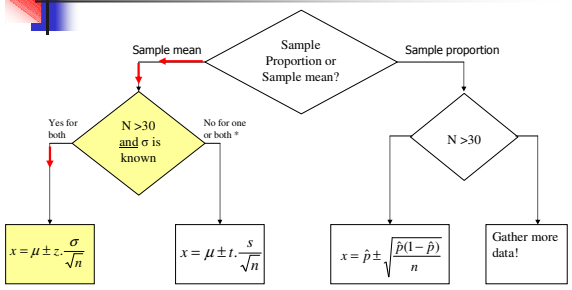
Review



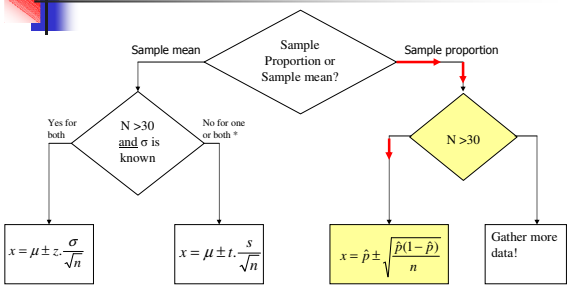
Review



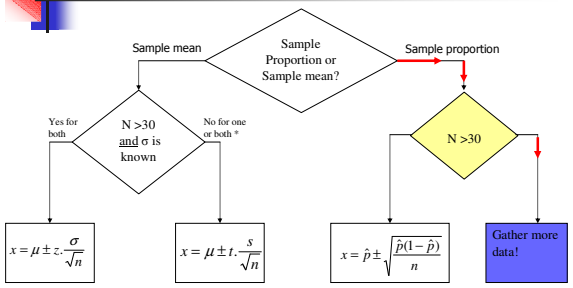
Review



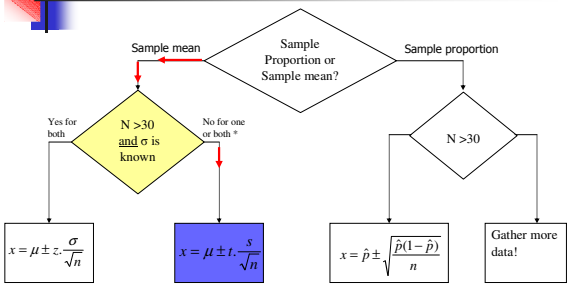
Review



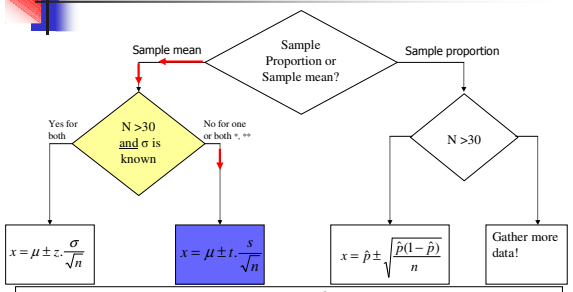
Review



Review



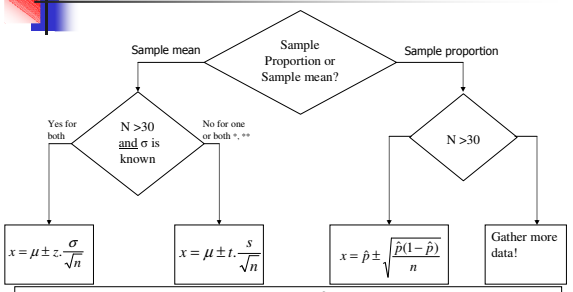
Review



*: in case n is large, but you don't have because σ, you use $x = \mu \pm t.\frac{s}{\sqrt{n}}$, however your t will be very close to z.

**: in case you have σ but your N is small, you need to use t, but you may follow: $x = \mu \pm t.\frac{\sigma}{\sqrt{n}}$

Review



*: in case n is large, but you don't have because σ, you use $x = \mu \pm t.\frac{s}{\sqrt{n}}$, however your t will be very close to z.

**: in case you have σ but your N is small, you need to use t, but you may follow: $x = \mu \pm t.\frac{\sigma}{\sqrt{n}}$

Review

- For comparison on mean samples or proportion samples:
- Two ways:
  - Comparing based on margin of error and confidence intervals
  - State Hypotheses (Ho: they are equal, Ha: they are either different or one is bigger than the other). And calculate the specific Standard deviation.