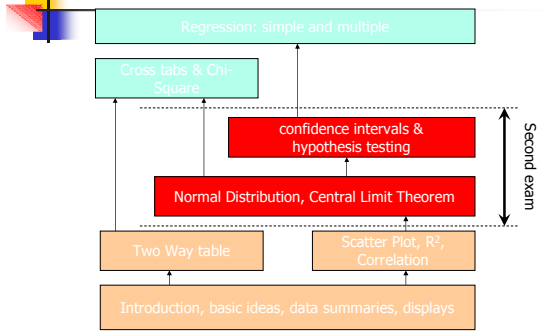
PUB – POS 316
Week 12

Review

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

Course Road Map

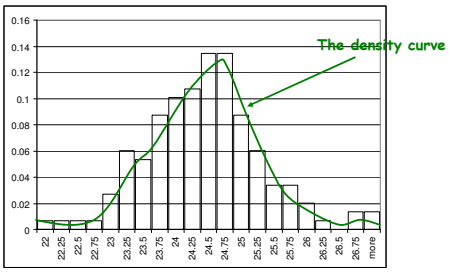


Introduction

- 1. Normal Distribution
- 2. Samples
  - 2.2. Sample proportion
  - 2.3. Sample means
- 3. Confidence interval
- 4. Hypothesis testing
  - 5. t-test

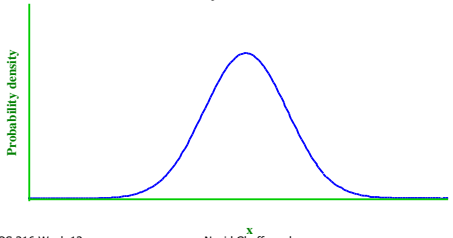
1. Normal Distribution

- Example: Biological clocks



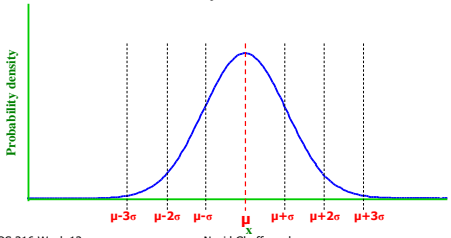
1. Normal Distribution

- Normal distribution is a bell-shaped, symmetric density curve. Average is shown by μ and variance is shown by σ.



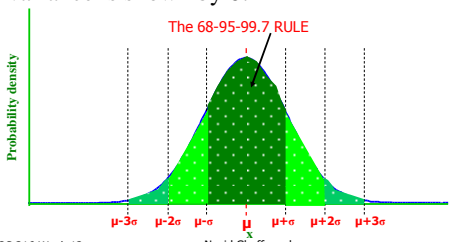
1. Normal Distribution

- Normal distribution is a bell-shaped, symmetric density curve. Average is shown by μ and variance is shown by σ.

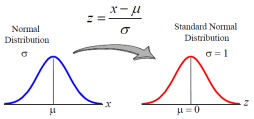


1. Normal Distribution

- Normal distribution is a bell-shaped, symmetric density curve. Average is shown by μ and variance is shown by σ.



1. Normal Distribution



One table!

Clarify the **normal distribution** characteristics → Find the equivalent **z-distribution** (standard normal dis)

Make inferences for the **normal distribution** ← Analyze **z-distribution**

1. Normal Distribution

- **(1)** Assume that the level of income per family in Albany is normally distributed with the average of 29K.
- **1-a:** If 68% of the families earn between $20,000 and $38,000, what is the standard deviation of this distribution?
- **1-b:** Assume the poverty line is on 15K. What percentage of families is below the poverty line?

2. Sample

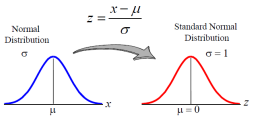
- Almost (always) we study samples. So, we don't have the distribution of population.
- Two kind of questions:
  - Sample mean (e.g., average income of UAlbany students)
  - Sample proportion (e.g., percentage of people who agree with Obama's health care plan)

2-1- Sample mean

- An example: We would like to know the average income of UAlbany students. We take a sample of 20 students, …..
  - **The central limit theorem**: when n is large, the sampling distribution of the sample mean is approximately normal with mean of μ and

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

2-1- Sample mean



One table!

Clarify the **normal distribution** characteristics → $\mu_{\bar{x}} = \mu$, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ → Find the equivalent **z-distribution** (standard normal dis)

Make inferences for the **normal distribution** ← Analyze **z-distribution**

2-1- Sample mean

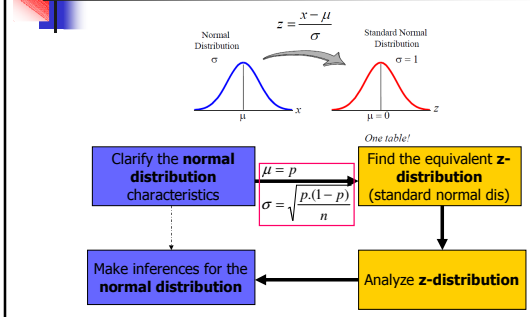
- **(2) Assume the average of years of education in Albany is 13 years with the standard deviation of 5 years. We take a sample of 100 people.**
- **2-a: What is the probability that the average of years of education in our sample becomes less than 12 years?**
- **2-b: What is the probability that the average of years of education in our sample becomes between 12 years and 13.5 years?**

2-2- Sample proportion

- An example: Suppose 60% of people agree with Obama's health reform. We poll 36 people ….
  - **Every thing is the same BUT the way you calculate the standard deviation**

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ $\quad$ $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

2-2- Sample proportion



2-2- Sample proportion

- **(3) Assume 35% of adults smoke. If our class is an SRS of the population (assume we have 36 students), what is the probability that more than half of the students smoke?**

3. Confidence interval

- An interval containing the true value of the parameter with some probability.
  - Example: Price of a medium size cup of coffee in Albany is:
    - $ 1.85 ±$0.35 (with 95% confidence)
  - Margin of error
  - X= (the result from the sample) ± z. (proper standard deviation)
  - Sample mean: $x = \mu \pm z.\frac{\sigma}{\sqrt{n}}$
  - Sample proportion: $x = \mu \pm z.\sqrt{\frac{p(1-p)}{n}}$

4. Hypothesis testing

- Hypothesis testing is a way to make a systematic conclusion from data.
  - First you state your hypotheses.
  - Then assume the null hypothesis is correct, and draw a distribution curve for it.
  - Then, find p-value
    - P-Value: The probability that H_0 is true, based on our data.
  - Then compare p-value with a critical number, usually 0.05. If it is lower you can reject the null hypothesis under 0.05 level of significance.

3,4 - Confidence interval and Hypothesis testing

- **(4) Gasoline consumption (Fall09-test2):** A record was kept of the miles per gallon of gasoline used by a particular high-milage hybrid car. We have twenty observations, found by computing the miles per gallon each time the gas tank was filled. At the right are the 20 observations of the miles per gallon (mpg) used in successive tankfuls of gas for this particular car. The mean of this sample of 20 observations is = 43.2.
- **(4a)** Suppose that the standard deviation for miles per gallon for this type of car is known to be σ = 3.5 mpg. Find a 95% confidence interval for μ, the mean mpg for this type of vehicle.
- **(4b)** The manufacturer claims the average for this car is 45 miles per gallon. The data suggest it's less. Can you reject the manufacturer's statement at some level of significance (.05 or below)? State proper hypotheses and alternate hypothesis to test the manufacturer's claim.

3,4 - Confidence interval and Hypothesis testing

- **(5) (Fall09-test2) The online poll site www.twiigs.com** posted the following poll question September 5, 2009.
- "Some critics have argued that the Obama administration is taking on too much, while others believe the president is fulfilling campaign promises. What's your assessment?"

- The results it lists now on its web site are:
- Obama has taken on too much: 26%
- The White House is taking the right approach: 36%
- The president isn't taking on enough: 3%
- The administration isn't acting on the right issues: 12%
- Right issues, wrong approach: 21%
- Total Votes: 2,576

- (a) Assuming that this poll is representative of the U.S. adult population, find a 95% confidence interval for the proportion of U.S. adults who think President Obama has taken on too much.

5. t-test

- 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?

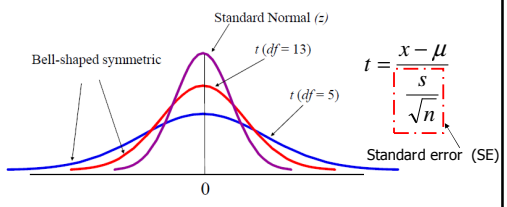
(source: Carnegie Mellon University, 90-711, Empirical Methods)

5. t-test

- 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)

5. t-test

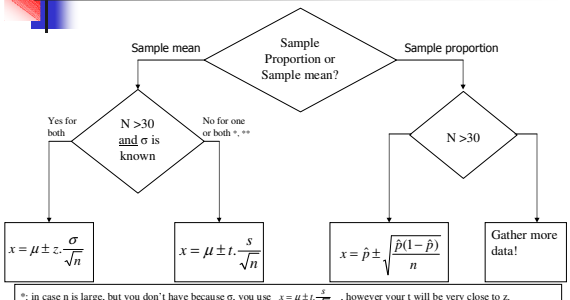


$$t = \frac{x - \mu}{\frac{s}{\sqrt{n}}}$$

Standard error (SE)

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

5. t-test (vs. z-test)



*: 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: $x = \mu \pm t.\frac{\sigma}{\sqrt{n}}$

5. t-test

- **(6) t-test:** We would like to know about the average expenditure of UAlbany students. We take an SRS, and the average result is $650 per month per person. We would like to state about the true value in the population with 95% confidence under each of these conditions:
- a – when the standard deviation in our population is $100 and n=40.
- b- when we don't know the standard deviation in our population, but the standard deviation in our sample is $100 and n=40.
- c- when the standard deviation in our population is $100 and n=10.

Short Questions

- **Short questions, short answers:**
- (a) What value of Z would be used to compute the margin of error for a 90% confidence interval?
- (b) Which is larger, a 95% confidence interval or a 99% confidence interval?
- (c) Which is larger, a 95% confidence interval based on a sample size of 400 or 800?
- (d) What is the intuitive meaning of "p-value" in a hypothesis test?
- (e) Under what conditions we should perform t-test instead of z-test?