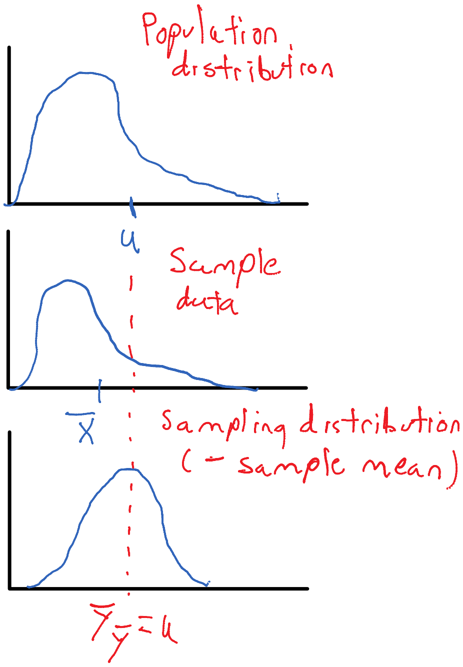
Lecture 7 chicken scratch

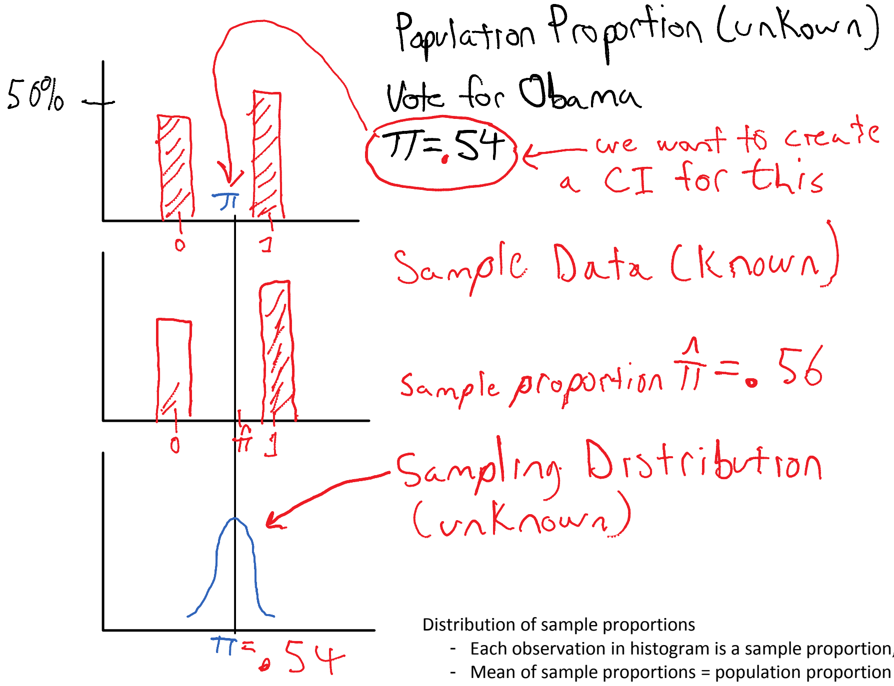
Saturday, September 29, 2012
9:19 AM

Variable: total enrollments at a college



For each p

Confidence Intervals for Proportions



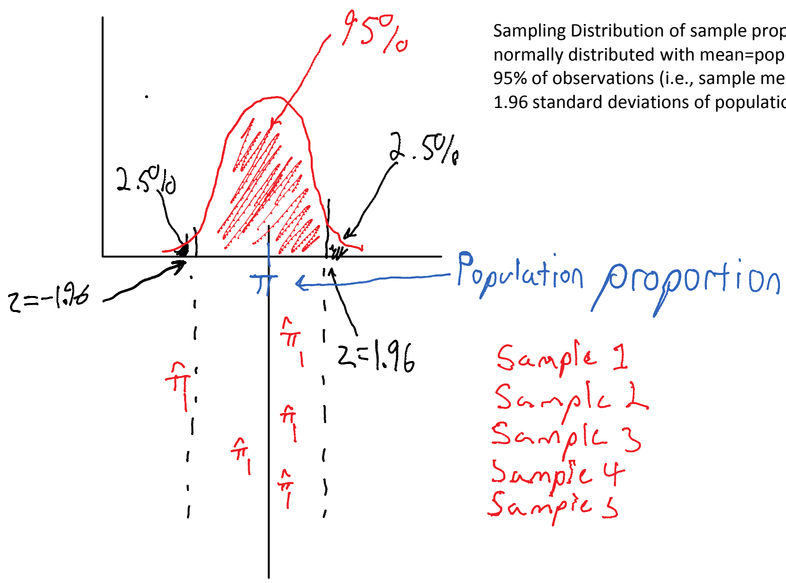
Distribution of sample proportions

- Each observation in histogram is a sample proportion, π
- Mean of sample proportions = population proportion
- Goal: we want an estimate of the population proportion π

- 95% of sample proportions (from the sampling distribution) will be within 1.96 standard deviations of the population proportion

Sampling Distribution of sample proportion: If we knew the sampling distribution

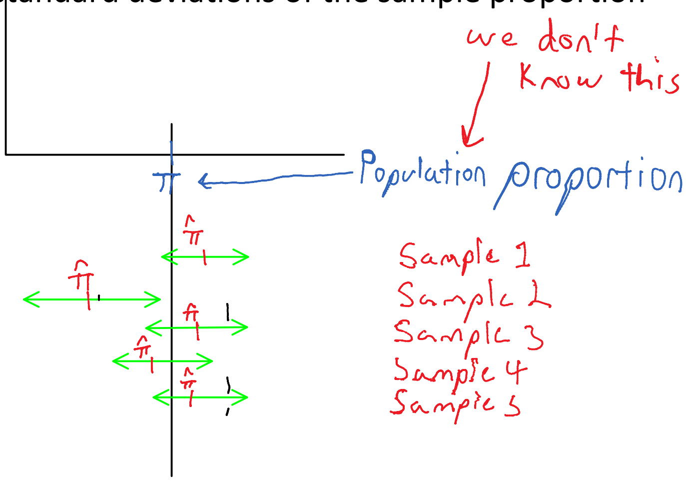
Sampling Distribution of sample proportion: normally distributed with mean=population mean. 95% of observations (i.e., sample means) fall within 1.96 standard deviations of population mean



- Equivalently, if we select a random sample and calculate the sample proportion, there is a 95% chance that the population proportion will be within 1.96 standard deviations of the sample proportion

Confidence intervals in green

We are 95% confident that The true population proportion lies somewhere in this confidence interval.



CHAPTER 6

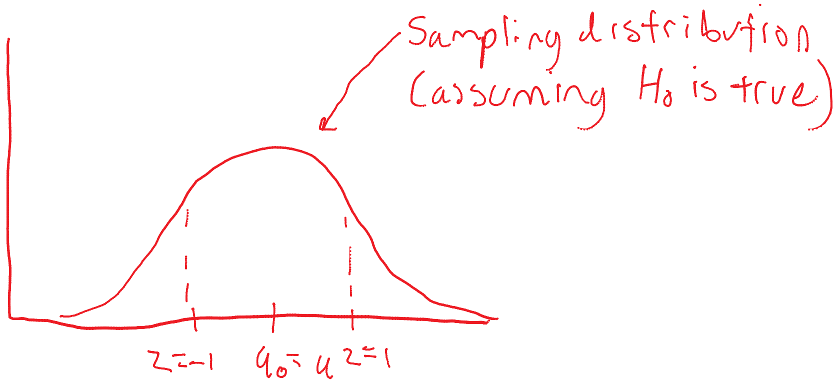
CHAPTER 6

Research question: Is the population mean number of hours worked equal to 40?

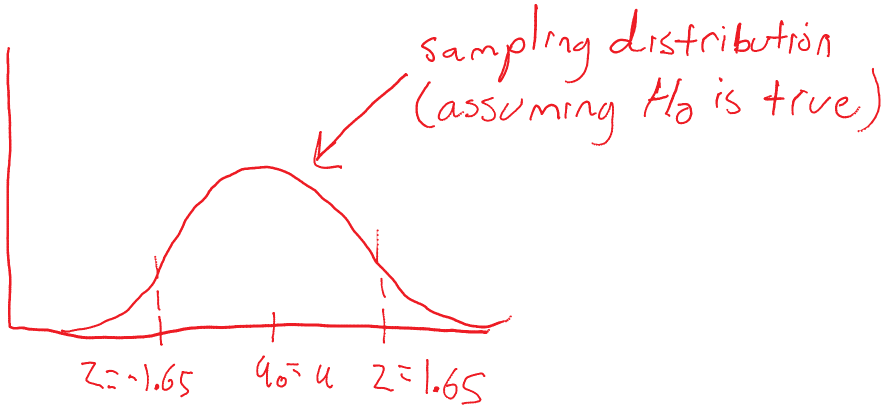
- Null hypothesis (H_o)
  - H_o: $\mu = \mu_o = 40$
- Two sided alternative hypothesis (H_a)
  - Two-sided: H_a: $\mu \neq 40$

Visualizing the test statistic
Draw sampling distribution under the assumption that the null hypothesis is true

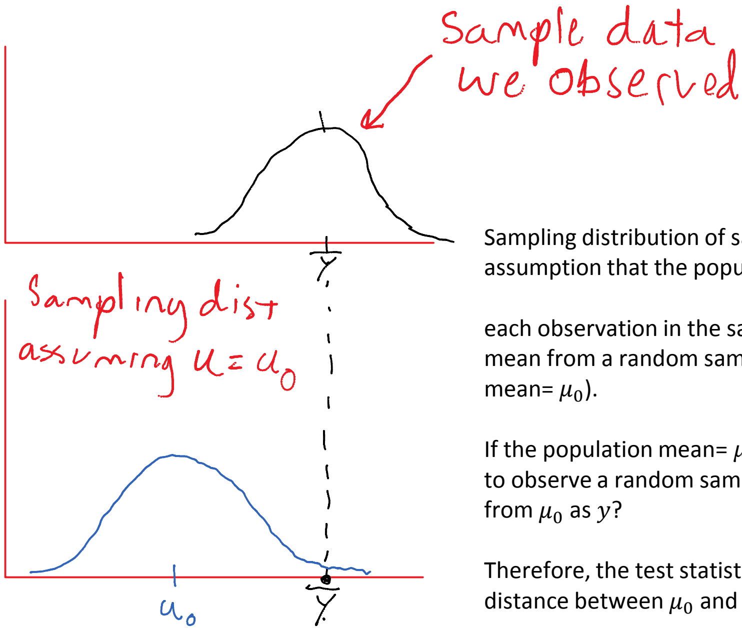
If the null hypothesis is true, what would be the probability of finding a sample mean more than one 1 standard error away from the population mean?



If the null hypothesis is true, what would be the probability of finding a sample mean more than one 1.65 standard errors away from the population mean?



If, under the assumption that the null hypothesis is true, the probability of picking the sample mean we observed is very low, then the null hypothesis is probably not true.

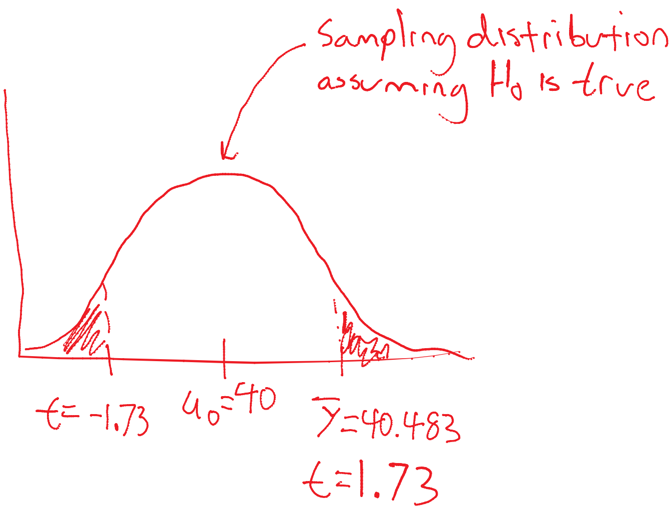


Sampling distribution of sample mean, y, under the assumption that the population mean, $\mu = \mu_0$

each observation in the sampling distribution is a sample mean from a random sample of data (assuming population mean= μ_0).

If the population mean= μ_0 then how likely would be it be to observe a random sample mean at least as far away from μ_0 as y?

Therefore, the test statistic is based on measuring the distance between μ_0 and y



Example 2

Research question: Is the population mean number of hours worked equal to 40?

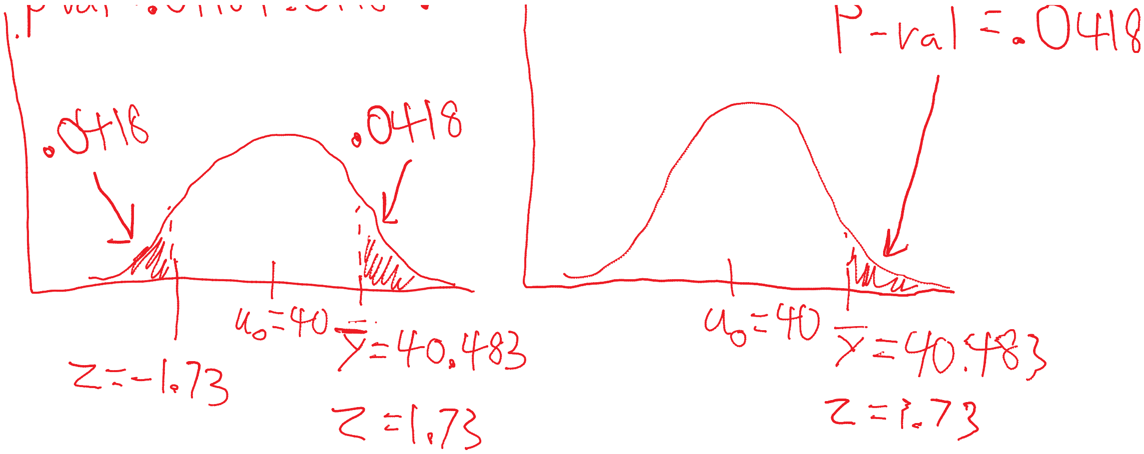
- Null hypothesis (H_o)
  - H_o: $\mu = \mu_o = 40$
- One sided alternative hypothesis (H_a)
  - One-sided: H_a: $\mu > 40$

p-value two sided H_a)

p-val = .0418 + .0418 = .0836

P-value (one sided H_a)

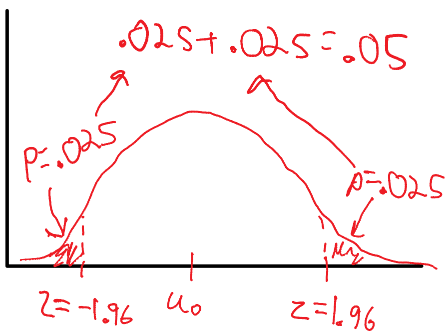
P-val = .0418

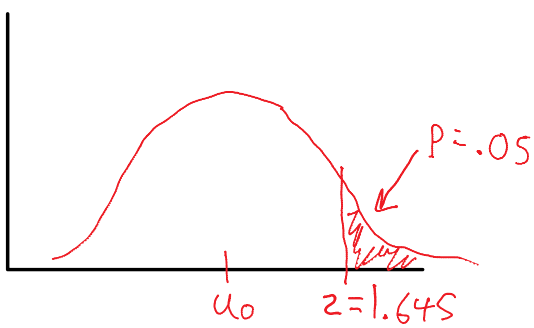


Rejection region:
This is something we decide *before* we test hypotheses

Two sided hypothesis
α = rejection region = .05

One sided hypothesis
$H_a: u > u_0$



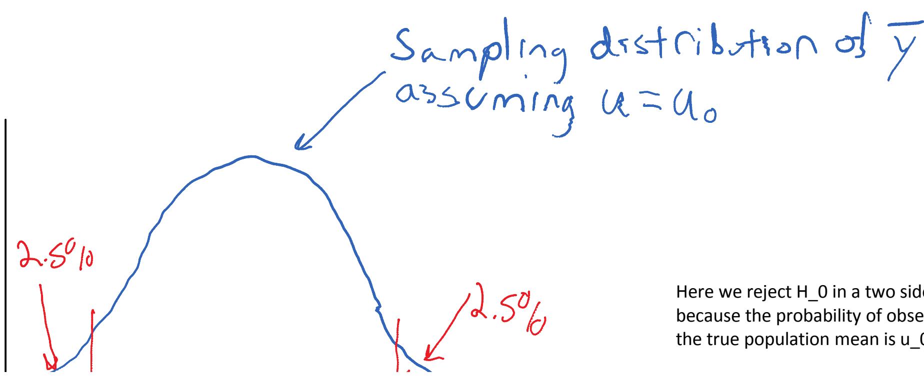


reject H_0 if t>1.96 or t<-1.96

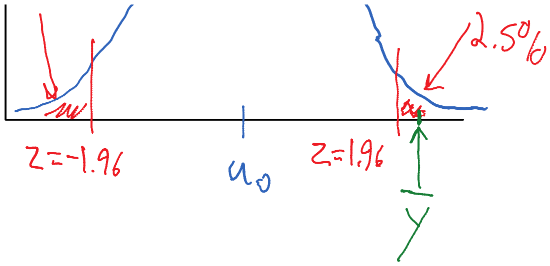
reject H_0 if t>1.645

Equivalence between 95% CI and two-sided significance test

Imagine we have two-sided alternative hypothesis,
using an alpha level of .05
So, we reject H_0 if t>1.96 or t<-1.96

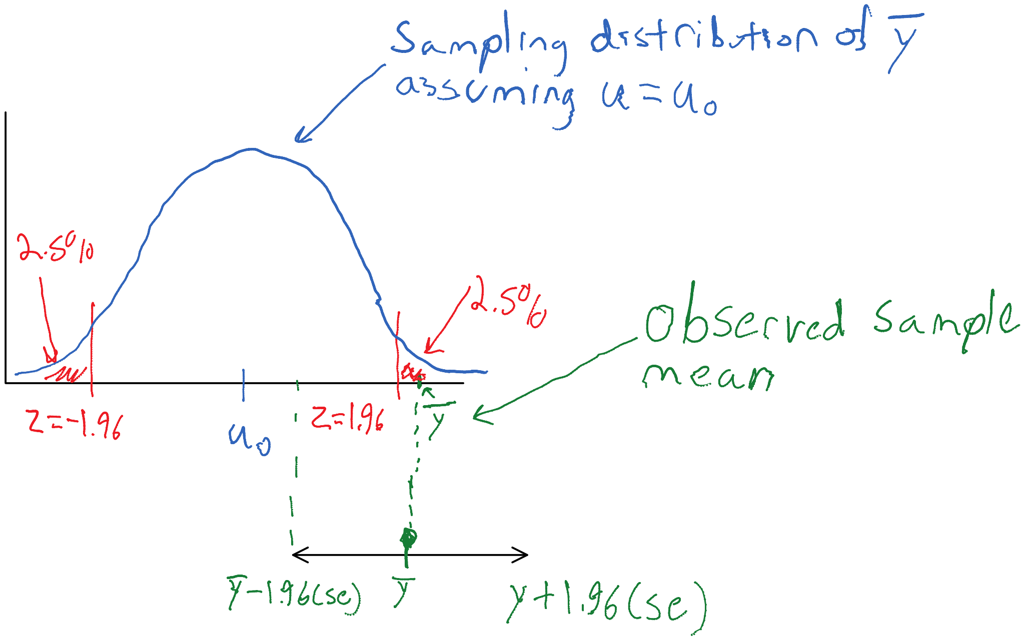


Here we reject H_0 in a two sided test where alpha=0
because the probability of observing a sample mean ybar if
the true population mean is u_0 is less than 5%

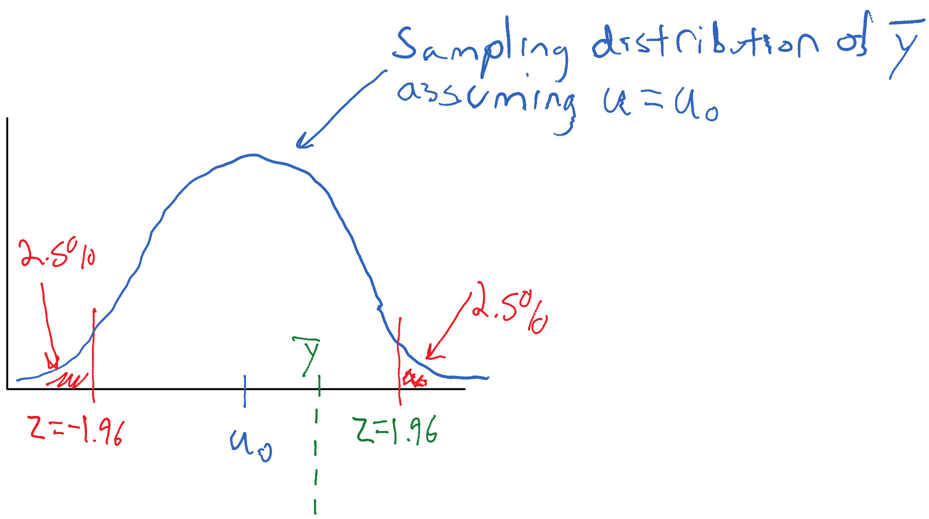


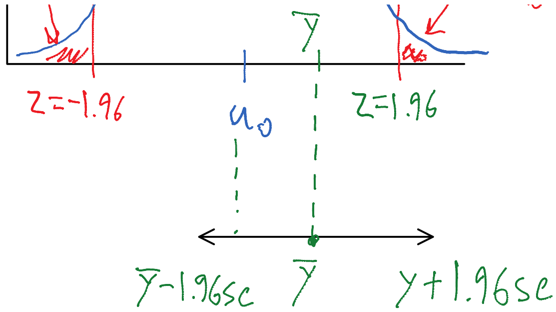
Here we reject H_0 in a two sided test where alpha=0 because the probability of observing a sample mean ybar if the true population mean is u_0 is less than 5%

- If p-value<=.05 (i.e., reject H_0)
  - If p-value<=.05 in a two-sided test, a 95% CI for μ does not contain μ_0
  - Equivalently, if 95% CI for μ does not contains μ_0 then we reject H_0



- If p-value>.05 (i.e., do not reject H_0)
  - When p-value>.05 in a two-sided test, the 95% CI for μ contains μ_0 (associated with null hypothesis, H_0)
  - Equivalently, If 95% CI for μ contains μ_0 then we do not reject H_0




$\bar{y}$
$z=-1.96$
u_0
$z=1.96$
$\bar{y}-1.96se$
$\bar{y}$
$y+1.96se$