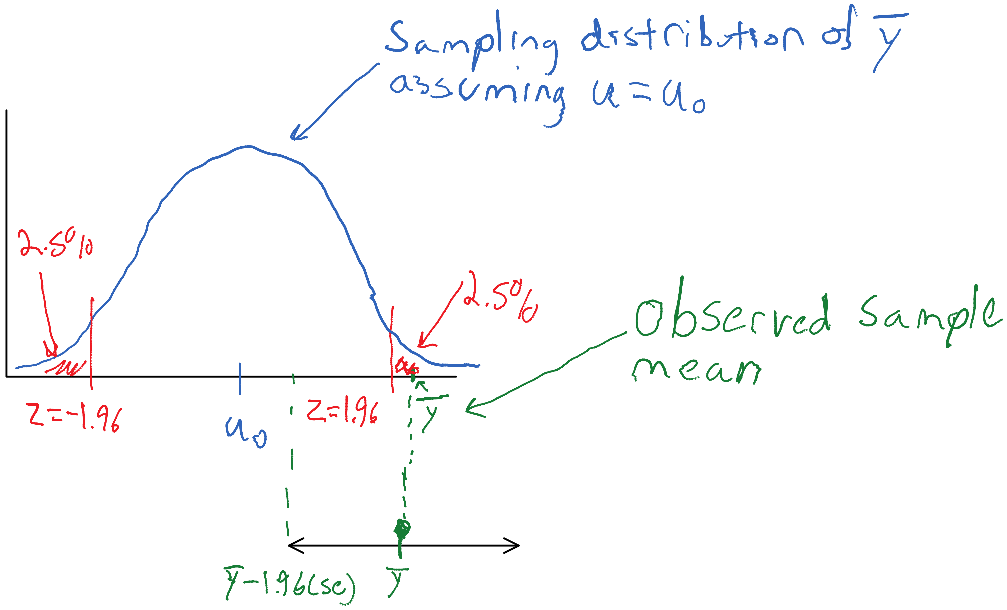
Lecture9-comparing groups

Saturday, October 13, 2012
10:27 AM

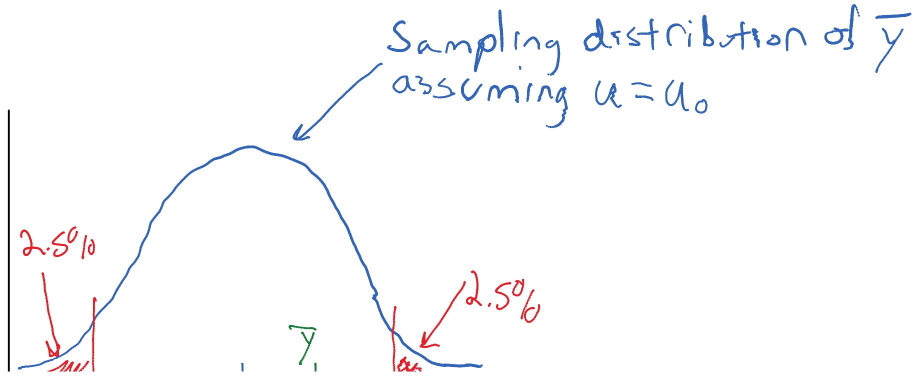
Equivalence between 95% CI and two-sided significance test

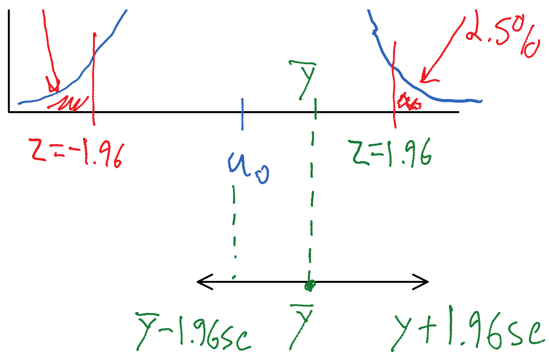
Imagine we have two-sided alternative hypothesis,
using an alpha level of .05
So, we reject H_0 if t>1.96 or t<-1.96

- 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**

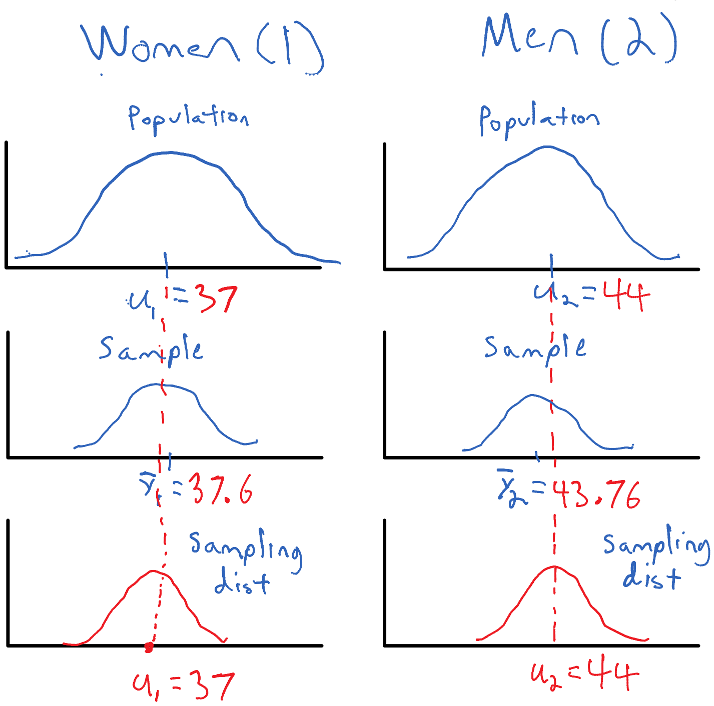




Chapter 7

- Hours worked data
  - Women (group 1): $y_1 = 37.60; n_1 = 1501; s_1 = 13.94$
  - Mean (group 2): $y_2 = 43.76; n_2 = 1319; s_2 = 15.18$

Question: Do men and women work the same number of hours per week?



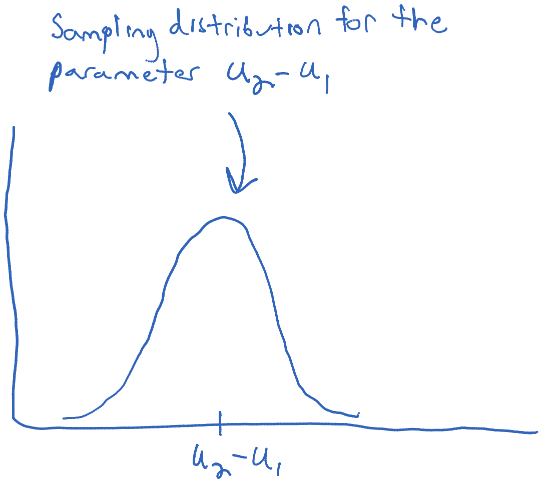
What is our strategy for figuring out whether population mean hours worked for women, μ_1, is different than population mean hours worked for men, μ_2?

Instead of thinking of two separate parameters, μ_1 *and* μ_2, we think of $(\mu_2 - \mu_1)$ as a single parameter.

If the parameter $(\mu_2 - \mu_1) \neq 0$, then we know that the parameter μ_2 is not equal to the paramter μ_1

H_0: $\mu_2 = \mu_1$
Same as this:
H_0: $\mu_2 - \mu_1 = 0$

H_a: $\mu_2 \neq \mu_1$
Same as this:
H_a: $\mu_2 - \mu_1 \neq 0$



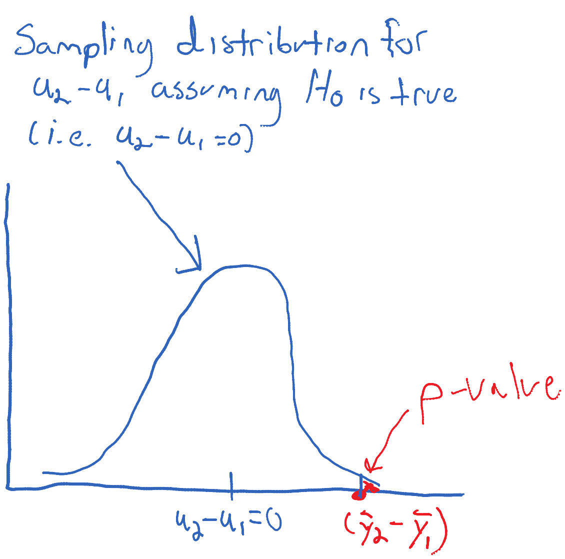
Question: What does each observation in the sampling distribution $(\mu_2 - \mu_1)$ represent?

So we want to test this:
H_0: $\mu_2 - \mu_1 = 0$
H_a: $\mu_2 - \mu_1 \neq 0$

We use sample data to make inferences about the population parameter $(\mu_2 - \mu_1)$.

Question: What is our best estimate of the population parameter $(\mu_2 - \mu_1)$?

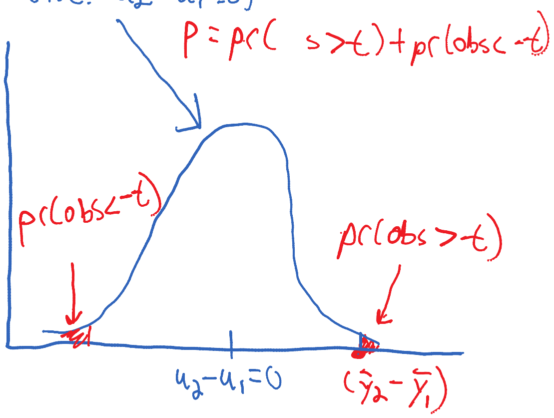
Answer: The difference between the two sample means that we actually observed, $(y_2 - y_1)$



Test statistic: if it would be very unlikely to find $(y_2 - y_1)$ under the presumption that $\mu_2 - \mu_1 = 0$ then the null hypothesis is probably not true.

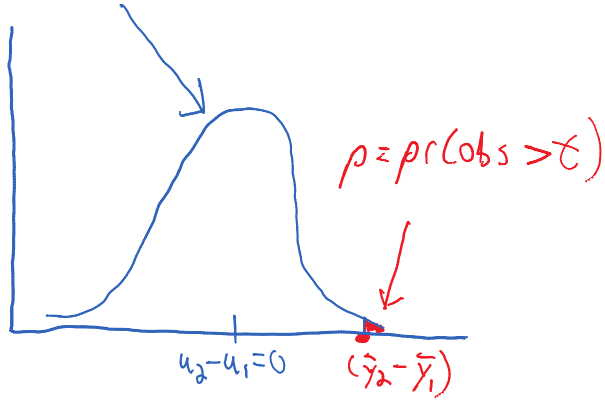
$H_a: u_2 - u_1 \neq 0$

Sampling distribution for $u_2 - u_1$ assuming H_0 is true (i.e. $u_2 - u_1 = 0$)



$H_a: u_2 - u_1 > 0$

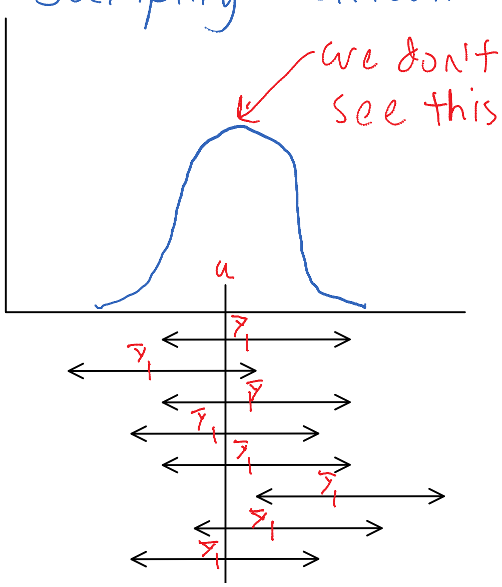
Sampling distribution for $u_2 - u_1$ assuming H_0 is true (i.e. $u_2 - u_1 = 0$)



Confidence Intervals

Confidence Interval for a single population mean

Sampling distribution of $\bar{y}$



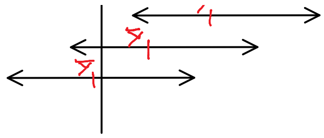
Note: for Confidence intervals we do not assume that the sampling distribution is based on some null hypothesis

Center of sampling distribution is population parameter

Eight different 95% confidence intervals drawn from eight random samples from population.

For each CI: we are 95% confident that the true population mean lies somewhere within the confidence interval

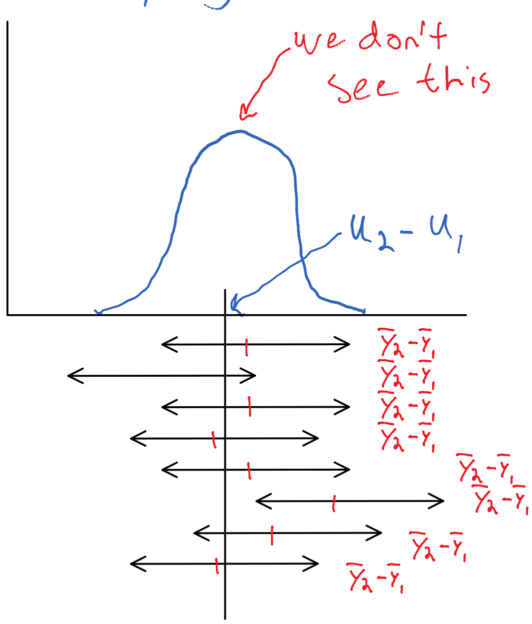
In this example, only one of the eight CIs does not contain the population mean



In this example, only one of the eight CIs does not contain the population mean

Confidence interval for a difference in Means

Sampling distribution of $\bar{y}_2 - \bar{y}_1$



If point estimate is unbiased then center of the sampling distribution equals the population parameter

Eight different 95% confidence intervals drawn from eight random samples from population.

For each CI: we are 95% confident that the true population parameter $(\mu_2 - \mu_1)$ lies somewhere within the confidence interval

In this example, only one of the eight CIs does not contain the population parameter $(\mu_2 - \mu_1)$

$\hat{\pi}_1 = \frac{778}{1{,}440} = .5403$

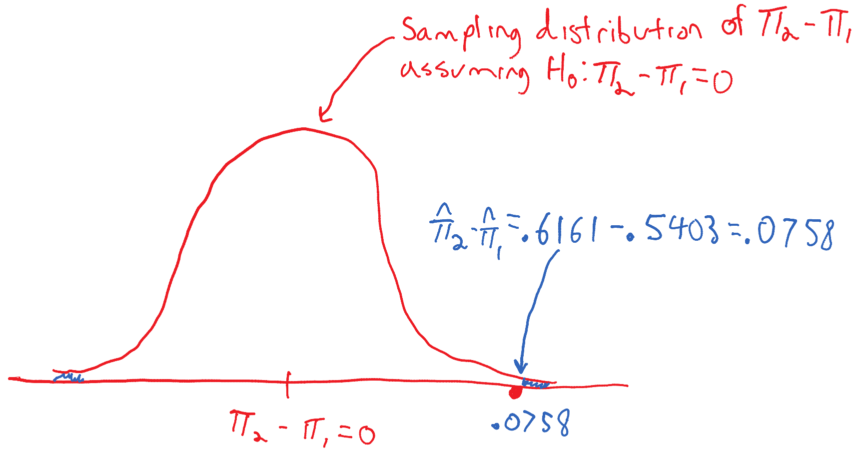
$\hat{\pi}_2 = \frac{1{,}194}{1{,}938} \approx .6161$

overall proportion or "pooled" proportion

$\hat{\pi} = \frac{1{,}972}{3{,}378} \approx .58.38$

0/1, responded voted for Obama in 2008 election	respondents sex 1. male	2. female	Total
0. no	662 45.97	744 38.39	1,406 41.62
1. yes	778 54.03	1,194 61.61	1,972 58.38
Total	1,440 100.00	1,938 100.00	3,378 100.00

$\hat{\pi}_2 - \hat{\pi}_1 = .6161 - .5403 = .0758$



Comparing means in Stata
(hours worked example)

$\bar{Y}_1$ $\bar{Y}_2$ $\bar{Y}_1 - \bar{Y}_2$ se of difference

95% CI for $\mu_1 - \mu_2$

```
. ttest hrs1, by(sex) unequal

Two-sample t test with unequal variances

   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
 1. male |    1319    43.76346    .4178559    15.17571    42.94372    44.58319
2. femal |    1501    37.60027    .3597864    13.93911    36.89453      38.306
---------+--------------------------------------------------------------------
combined |    2820    40.48298     .279642    14.85001    39.93466     41.0313
---------+--------------------------------------------------------------------
    diff |            6.163191    .5514071                5.081967    7.244414
------------------------------------------------------------------------------
    diff = mean(1. male) - mean(2. femal)                        t =  11.1772
Ho: diff = 0                     Satterthwaite's degrees of freedom =  2695.09

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

n_1 n_2

$t = \frac{6.16}{.55}$

$H_0: \mu_1 - \mu_2 = 0$

$H_a: \mu_1 - \mu_2 \neq 0$

$H_a: \mu_1 - \mu_2 > 0$

diff = male − female

p-value

p-value

Comparing proportions in Stata
(Voting for Obama in 2008 example)

$\hat{\pi}_{women}$ $\hat{\pi}_{men}$ se (of difference) se_0 (of difference)

95% CI for $\pi_{women} - \pi_{men}$

```
. prtest obama08, by(male)

Two-sample test of proportions                   0: Number of obs =      1938
```

se_0 (of difference)

```
. prtest obama08, by(male)

Two-sample test of proportions                  0: Number of obs =       1938
                                                1: Number of obs =       1440

    Variable |       Mean   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           0 |   .6160991   .0110473                       .5944467    .6377515
           1 |   .5402778   .0131333                       .5145369    .5660186
-------------+----------------------------------------------------------------
        diff |   .0758213   .0171618                       .0421847    .1094579
             |  under Ho:   .0171498     4.42   0.000
------------------------------------------------------------------------------
        diff = prop(0) - prop(1)                                  z =   4.4211
    Ho: diff = 0

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(Z < z) = 1.0000         Pr(|Z| < |z|) = 0.0000          Pr(Z > z) = 0.0000
```

women (0)

men (1)

= women - men = .0758

z-score

$H_0: \pi_{women} - \pi_{men} = 0$

$H_a: \pi_{women} - \pi_{men} \neq 0$

$H_a: \pi_{women} - \pi_{men} > 0$