Testing hypotheses

Ex.) Is a coin "fair"?

Let p = population proportion of "heads" in my coin

Two hypotheses competing

null hypoth. $\longrightarrow H_0: \; p = 0.5$ (meaning of "fair" coin)

alternative hyp. $\longrightarrow H_a: \; p \neq 0.5$

Last week
- Built CI for p
- Called $\hat{p}$ a "point estimate" of p

This week
- When testing hypotheses statements for p
- Call $\hat{p}$ a "test statistic"

Test using random sample

$n = 50$ (50 flips)

H H H T H T T H T T . . .

count heads: 21, say

sample proportion $\hat{p} = \dfrac{21}{50} = \underline{0.42}$

test statistic

Is this notable for a fair coin?

Answer comes via constructing a special sampling dist called a "null distribution"
- sampling dist for our test statistic
- takes our sample size into account
- takes our null hypothesis as true (hypothetically true)

> P-value: relative frq. of seeing a result at least as extreme as our test statistic in a world where H_0 holds true

In our case: $\hat{p} = 0.42$ corresponds to

P-value of ~ 0.33

Note: A P-value is a relative frequency, so lies between 0 and 1.

Note: The smaller the P-value, the stronger the evidence <u>against</u> H_0 (in favor of H_a).

Typically, one sets a threshold, called "significance level" $\underline{\alpha}$, and require P to be smaller than α before we are willing to reject H_0 in favor of H_a.

Typical values used

$$\left.\begin{array}{l} \alpha = 0.1 \\ \longrightarrow \alpha = 0.05 \\ \alpha = 0.01 \end{array}\right\}$$ Require $P < \alpha$ in order to conclude H_0 is false, H_a true.

In case of 50 flips of coin: P-value $= 0.328$

Not significant at any of these levels α

Say, when P not significant,

"Fail to reject H_0"

"My sample is consistent w/ H_0 being true"

Never say: H_0 is true (can't be proven)

<u>Ex.</u> Quant. var.: body temperature

H_0: $\mu = 98.6$ ← "null value" is where null distribution should be centered

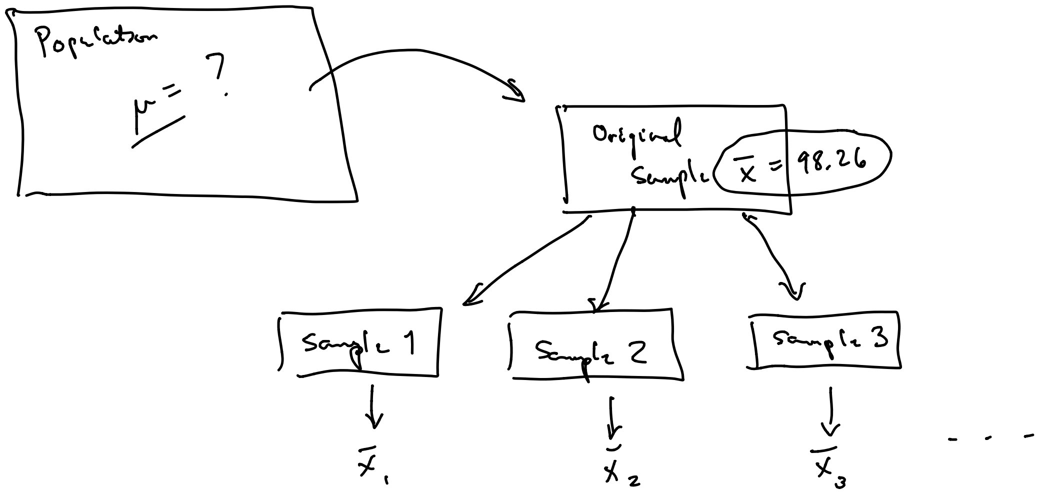
H_a: $\mu \neq 98.6$ (2-sided alt. hyp. — use both tails in null dist.)

Sample $n = 50$

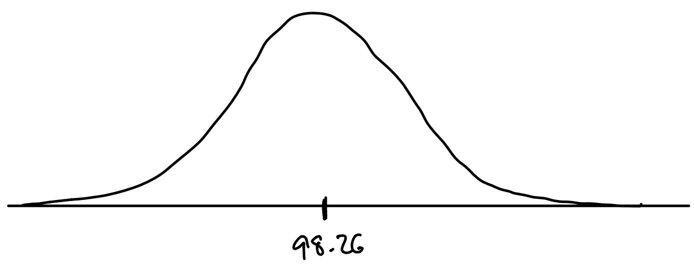
$\bar{x} = 98.26$ (test statistic)

Want null distribution – sampling dist for $\bar{x}$ where H_0 is true, and samples are of size 50.

Last week we bootstrapped

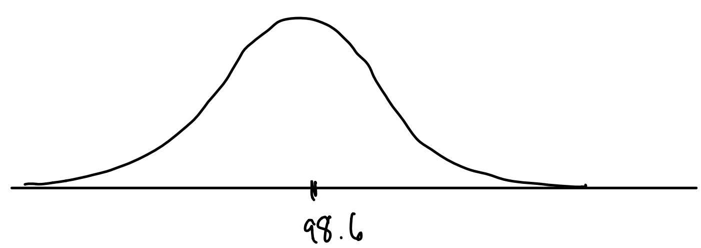


Bootstrap Distribution



Last week: great for CI const.

Get "randomization distribution" (an approximate null distribution) by sliding over so it is centered at null value



Examples above involve univariate data

1. Coin tosses, a categorical variable which is binary (2 values: H, T)
2. Body temperatures, a quantitative variable

Bivariate Scenarios (all have "apps" in the Randomization Hyp. Tests panel)

3. Explanatory variable: binary categorical (group-identifier var.)
   Response variable: quantitative

   H_0: $\mu_1 - \mu_2 = 0$ (or $\mu_1 = \mu_2$)

   test statistic: $\bar{x}_1 - \bar{x}_2$

4. Explanatory variable and response var. are both binary categorical

   H_0: $p_1 - p_2 = 0$ (or $p_1 = p_2$)

   test statistic: $\hat{p}_1 - \hat{p}_2$.

5. Explanatory and response are both quantitative

   H_0 can be either $\rho = 0$ (true correlation is zero)

   or $\beta_1 = 0$ (true regression line slope $= 0$)

   test statistic is r or b_1 depending on which H_0 used.

In all of cases 3-5,

- the null hypothesis expresses "no association between vars"
- the alternative hypothesis, whether 1- or 2-sided, says "there is an association".