$X_1, \dots, X_N \overset{iid}{\sim} \text{Bern}(p)$

Prob (mushrom lovers)

Population

$N \approx |\mathbb{N}|$

$p := \frac{\sum X_i}{N}$

pop prop of mushrooms lovers

Knowing p is impossible

Sample (survey)

(Simple random (SRS)

sample size $n << N$

$\underbrace{X_1, \dots, X_n}_{SRS} \overset{iid}{\sim} \text{Bern}(p)$

p unknown

Either likes or doesn't.

Goals

1. Estimate p (CI)
2. Test theories about p (HT's)
3. Make decisions using p

How to take a sample

$X_1 \mid C_1 = 1, \dots, X_n \mid C_n = 1 \overset{iid}{\sim} \text{Bern}(p_c)$

Chinise Yes/NO

s.t. $p_c > p$

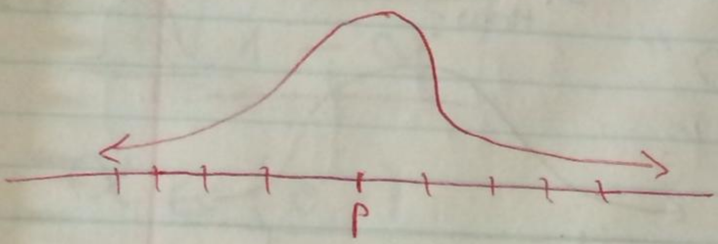
Biased sample

Non-representative Sample

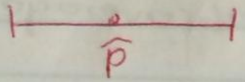
Estimates

① Point esti $\qquad \hat{p} = \frac{\sum x_i}{n}$

② Integral Estimates
(range of values)



$$\hat{p} \sim N\left(p, \left(\sqrt{\frac{p(1-p)}{n}}\right)^2\right)$$



$$\Longrightarrow \left[\hat{p} - \sqrt{\frac{p(1-p)}{n}}, \ \hat{p} + \sqrt{\frac{p(1-p)}{n}}\right]$$

$$(?) \quad \left[\hat{p} \pm \sqrt{\frac{p(1-p)}{n}}\right]$$

Converge or Confidence $:= P(p \in \quad)$

$$P\left(p \in \left[\hat{p} - \sqrt{\frac{p(1-p)}{n}}, \ \hat{p} + \sqrt{\frac{p(1-p)}{n}}\right]\right) =$$

$$= P\left(\hat{p} - \sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{p} + \sqrt{\frac{p(1-p)}{n}}\right)$$

$$= P\left(-\sqrt{\frac{p(1-p)}{n}} \leq p - \hat{p} \leq \sqrt{\frac{p(1-p)}{n}}\right)$$

$$= P\left(-1 \leq \frac{p - \hat{p}}{\sqrt{\frac{p(1-p)}{n}}} \leq 1\right)$$

$$= P\left(1 \geq \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \geq -1\right) = P(1 \geq Z \geq -1)$$

$$= P(Z \in [-1, 1]) = \underline{.68}$$

$z_{\frac{\alpha}{2}} := z$ st

$F_Z(z) = 1 - \frac{\alpha}{2}$

$\alpha = 5\%$

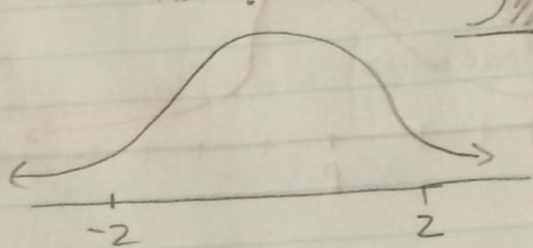
$F_Z(z) = .975$

$Z_{5\%} = 2$

How?

prob & z values

$\left[\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}\right]$



P of coverage is 95%.

$$P\left(p \in \left[\hat{p} - z\sqrt{\frac{p(1-p)}{n}},\ \hat{p} + z\sqrt{\frac{p(1-p)}{n}}\right]\right)$$

$$= P\left(\hat{p} - z\sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{p} + z\sqrt{\frac{p(1-p)}{n}}\right)$$

$$P\left(-z\sqrt{\frac{p(1-p)}{n}} \leq p - \hat{p} \leq z\sqrt{\frac{p(1-p)}{n}}\right)$$

$$P\left(-z \leq \frac{p - \hat{p}}{\sqrt{\frac{p(1-p)}{n}}} \leq z\right)$$

$$= P\left(z \geq \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \geq -z\right)$$

$$P(z \geq Z \geq -z)$$

$$P\left(Z \in \left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)\right) = F\left(z_{\frac{\alpha}{2}}\right) - F\left(-z_{\frac{\alpha}{2}}\right)$$

$$= F(z_{\frac{\alpha}{2}}) - (1 - F(z_{\frac{\alpha}{2}}))$$

$$= 2F(z_{\frac{\alpha}{2}}) - 1$$

$$= 2(1 - \frac{\alpha}{2}) - 1$$

$$= 2 - \alpha - 1$$

$$= \boxed{1 - \alpha}$$ "Coverage"

"Confidence"

$\alpha \in (0,1)$

2.5% do not cover

2.5% probability of coverage $= \underline{97.5\%}$ (question)

inside

point estim

$$\Big[\hat{p} \pm \underbrace{z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}}_{\text{margin of error}}\Big] \approx \Big[\hat{p} \pm \underbrace{z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}_{\text{margin of error}}\Big]$$

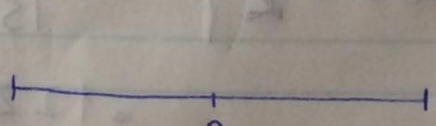
This works if n is large and p is not small, not large

Confident interval

$$CI_{p,\, 1-\alpha} := \Big[\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\Big]$$

Parameter

conf



means

$1-\alpha$ confidence $\neq$ $1-\alpha$ probability

Objectivist / Frequentist Interpretation

1. Before you make a CI,

$P(p \in CI) = 1-\alpha$ I already made the CI

2. If I repeat the probability many time

$1-\alpha \approx \frac{\sum \mathbb{1}_{p \in CI_i}}{n}$ l.r.f (long run frequency)

CI only once.

3. After CI is computed GARBAGE

$P(p \in CI) \in \{0,1\} \neq 1-\alpha$

Subjectivis / Bergesians

1. Under certain prior belief about p

$P(p \in CI) = 1-\alpha$ after construction.

$n = 15$

$\hat{p} = \frac{10}{15} = .667$

$CI_{p,95\%} = \left[.667 \pm 2\sqrt{\frac{.667 \times .333}{15}}\right]$

$.122 \times 2 = .243$

Ans

$= [.424, .910]$

Diverse college prior beliefs

* $\alpha = 5\% \Rightarrow z = 2$

$\alpha = 0.3\% \Rightarrow z = 3$

$$CI_{p,95\%} = \left[.667 \pm 3\sqrt{\frac{.667 \times .333}{15}}\right]$$

$$\overset{\text{Ans}}{=} [.300, 1]$$ uselessness

α goes down it's useless

Making a better integral n bigger

* $\hat{p} = \frac{100}{150} = .667$ $\alpha = 5\%$ then 2

$$CI_{p,95\%} = \left[.667 \pm 2\sqrt{\frac{.667 \times .333}{150}}\right]$$

$$= [.59, .74]$$ nice range useful

Play with n and α

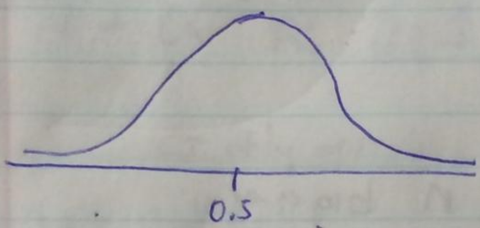
Confidence integral done (CI)
All about estimating p

② . Is my theory about p true?

Theory: Male / female births are same proportion

$p := P(\text{male}) = 0.5$

$X_1, \ldots, X_n \overset{iid}{\sim} \text{Bern}(p = 0.5)$



$\hat{p} \sim N\left(0.5, \left(\sqrt{\frac{0.25}{n}}\right)^2\right)$

Theory is OK

$\hat{p}$ here then Theory is wrong