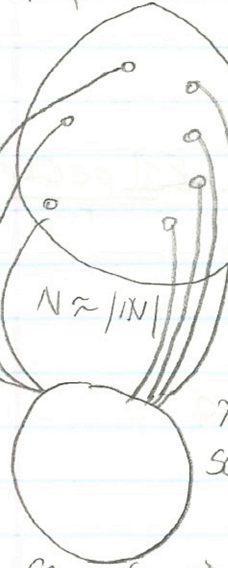
December 1, 2015 Lecture 21

population



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

^population proportion of mushroom lovers

prob of loving mushrooms

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

p is unknowable

Goals

① Estimate p * } Inference

② Test theories about p }

X③ Make decisions based on estimates of p.

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

Sample size

$n < N$

Sample (survey) ⇒ must be "representitive"

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

$p_c > p$

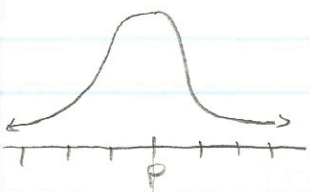
non-representitive or

biased sample

Simple random sample (SRS)

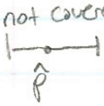
* Estimating p

Point estimate of p: $\hat{p} := \frac{\Sigma X_i}{n}$



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

not covered



Interval Estimates

(range of values for p)

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

^point estimate ^margin of error

$[3 \pm 1] = [2, 4]$

Coverage/confidence := $P(p \in \text{interval})$

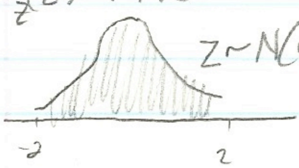
$= P\left(p \in \left[\hat{p} \pm \sqrt{\frac{p(1-p)}{n}}\right]\right) = P\left(\hat{p} - \sqrt{\frac{p(1-p)}{n}} \le p \le \hat{p} + \sqrt{\frac{p(1-p)}{n}}\right)$

$z_{\frac{\alpha}{2}}$ is the z s.t.

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

$\alpha = 5\%$ $\underset{z}{\overset{\|}{z_{2.5\%}}}$ is the z s.t.

$F_Z(z) = .975$



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

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

$= P\left(Z \in \left[-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right]\right) = F\left(z_{\frac{\alpha}{2}}\right) - F\left(-z_{\frac{\alpha}{2}}\right)$ $\alpha \in (0,1)$

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

$= 2\left(1 - \frac{\alpha}{2}\right) - 1 \Rightarrow 2 - \alpha - 1 = \boxed{1 - \alpha}$ "coverage confidence"

"Confidence level"

$\alpha = .2 \Rightarrow 1 - \alpha = .8$

$z = 1.64$

$\alpha \in \{1\%, 5\%, 10\%, 20\%\}$

$$\left[\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}\right] \approx \left[\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$ If n big, p not near 0, or 1, works well

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

CI: confidence interval; p: population proportion; $1-\alpha$: level of confidence

Frequentist / Objectivist Interpretation

1. Before I build the CI, $P(p \in CI) = 1-\alpha$
2. If I repeat this procedure many times $1-\alpha \approx \frac{\sum \mathbb{1}_{p \in CI_i}}{n}$ l.r.f.
3. After CI is built $\underbrace{P(p \in CI)}_{\text{makes no sense}} \in \{0,1\} \neq 1-\alpha$

$1-\alpha$ confidence

$\neq 1-\alpha$ probability

$$\hat{p} = \frac{10+9}{15+13} = \frac{19}{28} \approx .679$$

$$CI_{p,95\%} = \left[.679 \pm 2\sqrt{\frac{.679(1-.679)}{28}}\right]$$

$\alpha = 5\%$

$z_{\frac{\alpha}{2}} = 2$

$.679 - .176 = .503$

$.679 + .176 = .855$

$= [.503, .855]$

$$CI_{p,99.7\%} = \left[.679 \pm 3\sqrt{\frac{.679(1-.679)}{28}}\right] = [.415, .943]$$

$\alpha = .3\%$

$z_{\frac{\alpha}{2}} = 3$

$$CI_{p\,68\%} = \left[.679 \pm 1\sqrt{\frac{.679(1-.679)}{28}}\right] = [.59, .77]$$

$\alpha = 32\%$

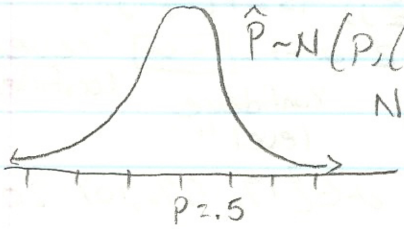
$z_{\frac{\alpha}{2}} = 1$

"Hypothesis testing" I have a theory about p

Does my data comport with my theory?

Theory: prop of male/female humans born is the same.

$p := P(\text{male}) = .5$ sample size n



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

$$N\left(.5, \left(\sqrt{\frac{0.5}{n}}\right)^2\right)$$