December 1, Lecture 21

~~• Recall if $X_1, \ldots, X_n$ are iid with~~

• Recall if $X_1, \ldots, X_n \overset{iid}{\sim}$ with mean μ & standard error σ, and n large...

(II) $\dfrac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$

(III) $\bar{X} \overset{d}{\approx} N(\mu, (\frac{\sigma}{\sqrt{n}})^2)$

(IV) $T \overset{d}{\approx} N(n\mu, (\sqrt{n}\,\sigma)^2)$

• Shipments are late 2% of the time. In 10,000 shipments, what is the probability more than 3% are late?

* $X_1, \ldots, X_{10000} \overset{iid}{\sim} \text{Bern}(p = 2\%)$
  - → Assumption: assuming each shipment iid Bern(p=2%).
  - → Assume its right.

* $P(\bar{X} \geq 3\%) = P(\hat{p} \geq 0.3)$ → next page...

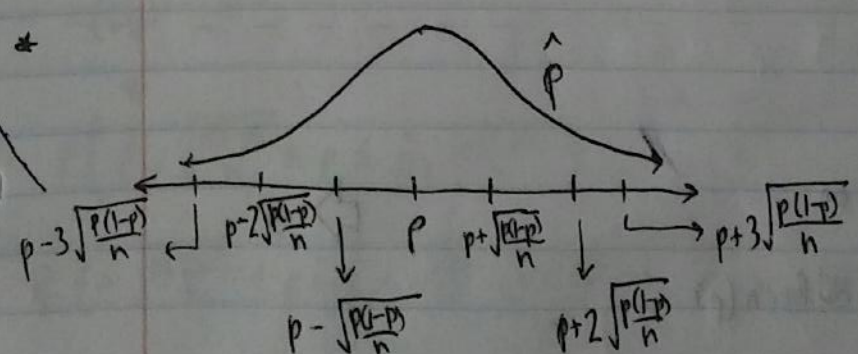
* $\bar{X} \overset{\text{by III}}{\approx} N(\mu, (\frac{\sigma}{\sqrt{n}})^2)$ → If $X_1, \ldots, X_n$, what is μ & σ?
  - $\mu = p$
  - $\sigma = \sqrt{p(1-p)}$

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

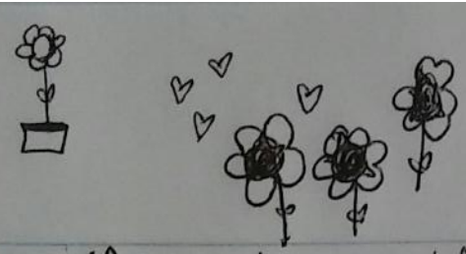
  $\hat{p} = \bar{X}$ in Bernoulli case

* $\hat{p} := \bar{X} = \dfrac{\Sigma \#1\text{'s}}{n}$ "sample proportion"

*



Next



$$* P(\bar{X} \geq 3\%) = P(\hat{P} \geq 0.03) = P\left(\frac{\hat{P} - .02}{\sqrt{\frac{.02(1-.02)}{10000}}} \geq \frac{.03 - .02}{\sqrt{\frac{.02(1-.02)}{10000}}}\right)$$

$$P(Z \geq 7.14) \approx 0$$

Implication - Never going to have more than 3% late

- Mushrooms → 11 yes, 23 total subjects

$n = 23$

$\hat{p} = 11/23 = 0.48$ ("data")

[$\hat{p}$ came from $\hat{P}$] - little $\hat{p}$ with no hat is ⇨ true expectation of someone like mushrooms or not.

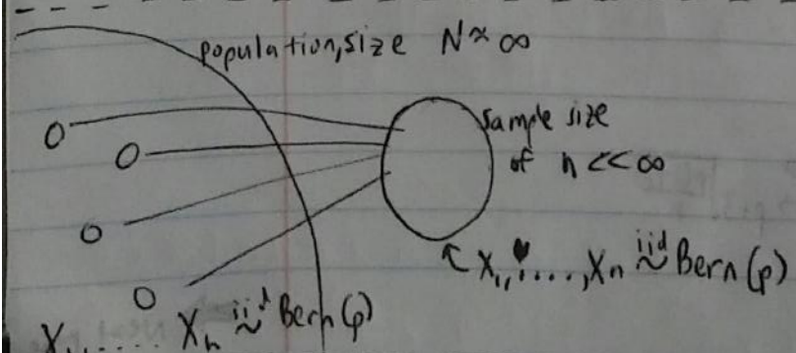
* "p" is the "true" population parameter. *
* we don't know p, too many people to ask. Need to ask everyone, however with a population of 7.5 billion, too difficult. Unknowable *

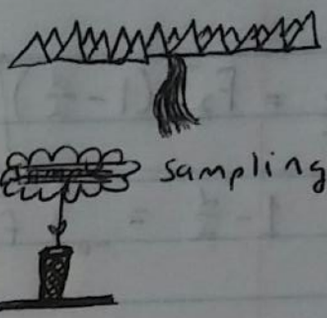
[Our goal is to know something about p.
Can we say anything we know about p.]

Statistics has started, probability ended.

we have data, but don't know parameter.
Trying to infer something from data about parameter.

* Statistical Inference: Infer the population parameter using a statistic of the data. Here, $\hat{p}$.





*What constitutes a "good" sample? ★

A representative sample. ie completely random sampling.

*Goals of Inference

① Give me the best guess of p.

② Give me a ~~reasonable~~ reasonable interval of values for p.

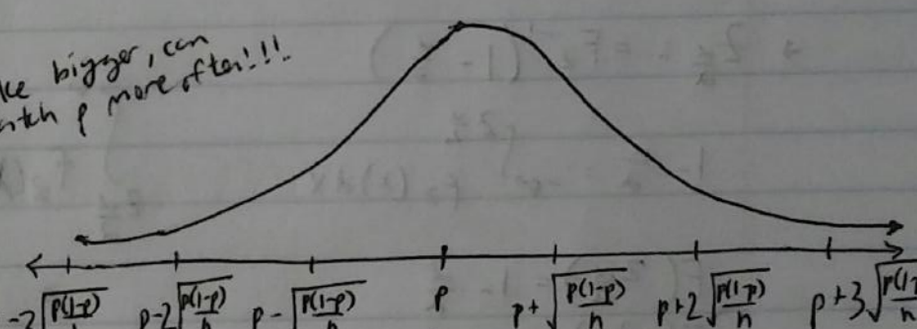
③ Test theories about p.

① point estimation

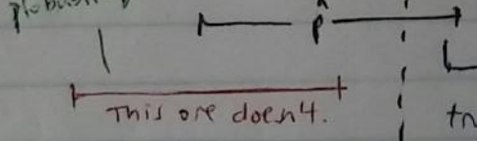
$p \approx \hat{p}$

② Confidence Intervals — If make bigger, can catch p more often!!!

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



probability here is 0 — probability caught it here 1



→ Did this interval catch true p? Is $p \in [\hat{p} \pm \sqrt{\frac{p(1-p)}{n}}]$?

→ yes, because drew it that way.

If do this many many times, how often will catch the truth? - What's probability that catch it?

have to amplify, make bigger, to catch p more.

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

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

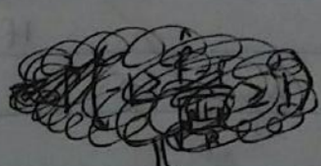
(big)

$\hat{p}, \hat{P}, p, P()$

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

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

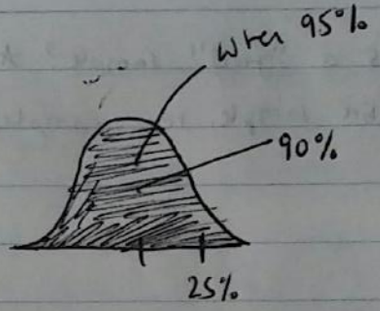
$$= P\left(-1 \le \frac{p-\hat{p}}{\sqrt{\frac{p(1-p)}{n}}} \le 1\right)$$



$$\overset{CLT}{\approx} P(-1 \le -Z \le 1) = P(Z \in [-1,1]) = 68\%$$

* $\boxed{z_{\frac{\alpha}{2}} = F_z^{-1}(1-\frac{\alpha}{2})}$

$1-\frac{\alpha}{2} = \int_{-\infty}^{z_{\frac{\alpha}{2}}} f_z(x)\,dx$



$\alpha = 10\% \Rightarrow \frac{\alpha}{2} = 5\%$

$1-\frac{\alpha}{2} = 95\%$

If do calculation again...

* $\left[\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\right]$ ... $\rightarrow P(z \in [-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}])$ ... $F(z_{\frac{\alpha}{2}}) - F(z_{\frac{\alpha}{2}})$ ... $(1-\frac{\alpha}{2}) - \frac{\alpha}{2} = \boxed{1-\alpha}$

$\rightarrow z_{\frac{\alpha}{2}} := F_z^{-1}(1-\frac{\alpha}{2})$

$1-\frac{\alpha}{2} = \int_{-\infty}^{z_{\frac{\alpha}{2}}} f_z(x)\,dx$

$\int_{z_{\frac{\alpha}{2}}}^{\infty} f_z(x)\,dx = 1-(1-\frac{\alpha}{2}) = \frac{\alpha}{2} = \int_{-\infty}^{-z_{\frac{\alpha}{2}}} f_z(x)\,dx$

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

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

* Mushroom example

$\hat{p} = 0.48$

~~0.48 ± 2 for 95% coverage.~~

$\left[0.48 \pm 2\sqrt{\frac{p(1-p)}{28}}\right]$ for 95% coverage. — absurd, don't know p, what we are trying to look for anyway.

change to →

* $\left[\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\right] \approx \left[\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$ * $\Rightarrow$ $\left[0.48 \pm 2\sqrt{\frac{0.48(1-.48}{23}}\right]$

~~If the p is not~~ If the p is not $p \approx 0$ or $p \approx 1$.

debated for 100 years.

* Definition: A $1-\alpha$ size C confidence interval for population proportion p

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

* $[0.48 \pm 2\sqrt{\frac{0.48(1-.48)}{23}}] \Rightarrow [.272, .688]$

for 95% coverage

* Can you say this? $P(p \in [.272, .688]) \neq 95\%$

No, the real p is a single #. This is range of #s & is not random.

So no, can't say this.