# Some finer points concerning inference on one proportion

**Confidence intervals** for $p$:

- Centered interval approach (Check the rules of thumb before applying)

critical value chosen for level of confidence

$$\widehat{p} \pm (z^*)(\mathrm{SE}_{\widehat{p}})$$

pt. est. is sample proportion

- Formula $\mathrm{SE}_{\widehat{p}} = \frac{\sqrt{p(1-p)}}{\sqrt{n}}$

$\mathrm{SE}_{\tilde{p}} = \sqrt{\frac{p(1-p)}{n}}$

Insert as $p$:
CI: $\hat{p}$
hyp. test: $p_0$

  Note that $\mathrm{SE}_{\widehat{p}}$ decreases as $n$ grows. Specifically,

  - can cut $\mathrm{SE}_{\widehat{p}}$ in half by quadrupling (4×) sample size
  - can cut $\mathrm{SE}_{\widehat{p}}$ to one third its size at $n$ by increasing sample size by a factor of 9
  - other similar statements?

- Can "tailor" a margin of error

Typical issue at election time: Want to estimate $p$ = population proportion who will vote for X using a CI.

Useless-sounding: We estimate $p$ to be $0.44 \pm 0.2$ w/ 95% Confidence

Better: We estimate $p$ to be ______ w/ M. of E. 0.03

$0.44 \pm 0.03$ or $0.75 \pm \underbrace{0.03}_{\text{better m. of E}}$

Q: Can I decide ahead of time what sample size $n$ is sufficient to give me a m. of error no bigger than 0.03 (or some specified amount)?

M. of error $= (z^*)\, SE$

$$0.03 \geq ME = (z^*)\frac{\sqrt{p(1-p)}}{\sqrt{n}}$$

Solve the inequality

$$0.03 \geq (z^*)\frac{\sqrt{p(1-p)}}{\sqrt{n}}$$

Multiply by $\sqrt{n}$, divide by $0.03$:

$$\sqrt{n} \geq (z^*)\frac{\sqrt{p(1-p)}}{0.03}$$

From calculus, a fact about numerator is that it can't <u>ever</u> get larger than 0.5.

So, we can replace numerator w/ $\frac{1}{2}$:

$$\sqrt{n} \geq (z^*)\frac{0.5}{0.03}$$

Now pick $z^*$ for desired level of confidence

95% conf. : $z^* = 1.96$
90% conf. : $z^* = 1.645$

$$\sqrt{n} \geq (1.96)\frac{.5}{0.03} \implies n \geq \left[1.96 \cdot \frac{0.5}{0.03}\right]^2$$

$= 1067.111$

$$n \geq \frac{(z^*)^2 \cdot (0.5)^2}{M^2} = \boxed{\frac{(z^*)^2}{4M^2}}.$$

In this example I took $M = 0.03$.

and I also replaced $\sqrt{p(1-p)}$ by $1/2$

but in some situations you may have reason to plug in a value for $p$.

**Hypothesis tests** involving null hypothesis $\mathbf{H}_0: p = p_0$:

- several **test statistics** one might use from the sample
  1. the sample proportion $\widehat{p}$
     - natural to use, as an unbiased estimator of $p$
     - approximating normal distribution: $\mathsf{Norm}(p_0, \sqrt{p_0(1-p_0)/n})$ for $\hat{p}$
  2. the count of successes $X$
     - natural to use, as an unbiased estimator of $p$
     - approximating normal distribution: $\mathsf{Norm}(np_0, \sqrt{np_0(1-p_0)})$ for $X$
  3. the $Z$-score of either of the previous
     - called the **standardized test statistic**
     - approximating normal distribution: $\mathsf{Norm}(0, 1)$
     - one advantage: immediate comparison with critical $z^*$ value

$H_0: p = 1/2$

On Monday: hyp. test for home-won soccer games

$X = 70$ times in $n = 120$ games the home team won.

1. $\hat{p} = 70/120$ has roughly a normal dist. like $\mathsf{Norm}\left(\frac{1}{2}, \sqrt{\frac{(1/2)(1/2)}{120}}\right)$, $\doteq 0.0456$

2. $X = 70$ has roughly a normal dist $\mathsf{Norm}\left(120 \cdot \frac{1}{2}, \sqrt{120(\frac{1}{2})(1-\frac{1}{2})}\right)$, $= 60$, $= 5.477$

3. $Z$-score from $\hat{p}$:

$$= \frac{70/120 - 0.5}{0.0456} = 1.83$$

from $X$:

$$Z = \frac{70 - 60}{5.477} = 1.83$$

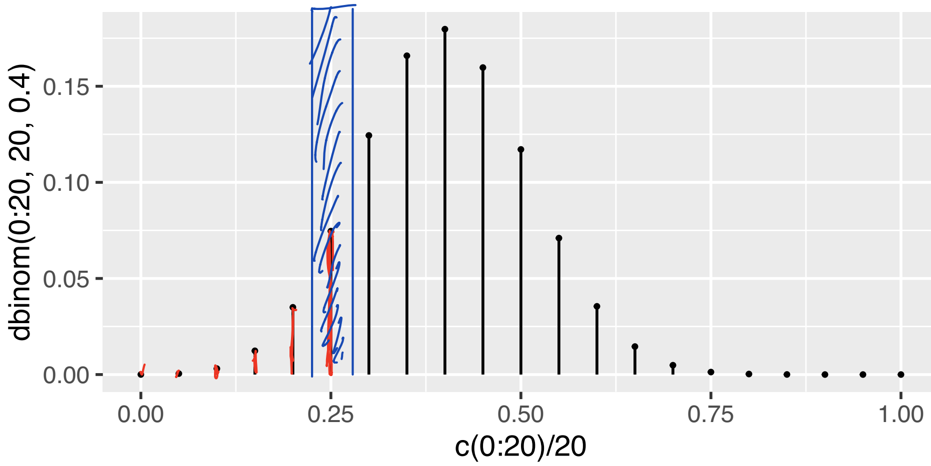
has $\mathsf{Norm}(0, 1)$

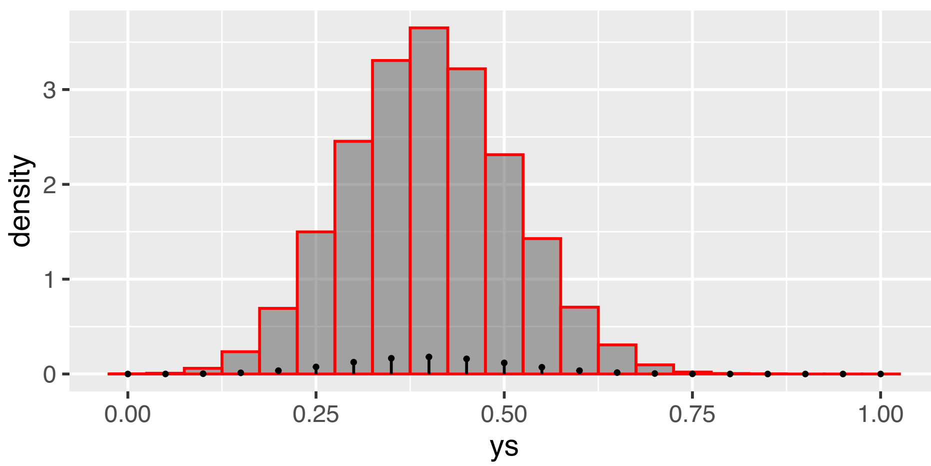
A normal model for single proportion replaces heights w/ areas.

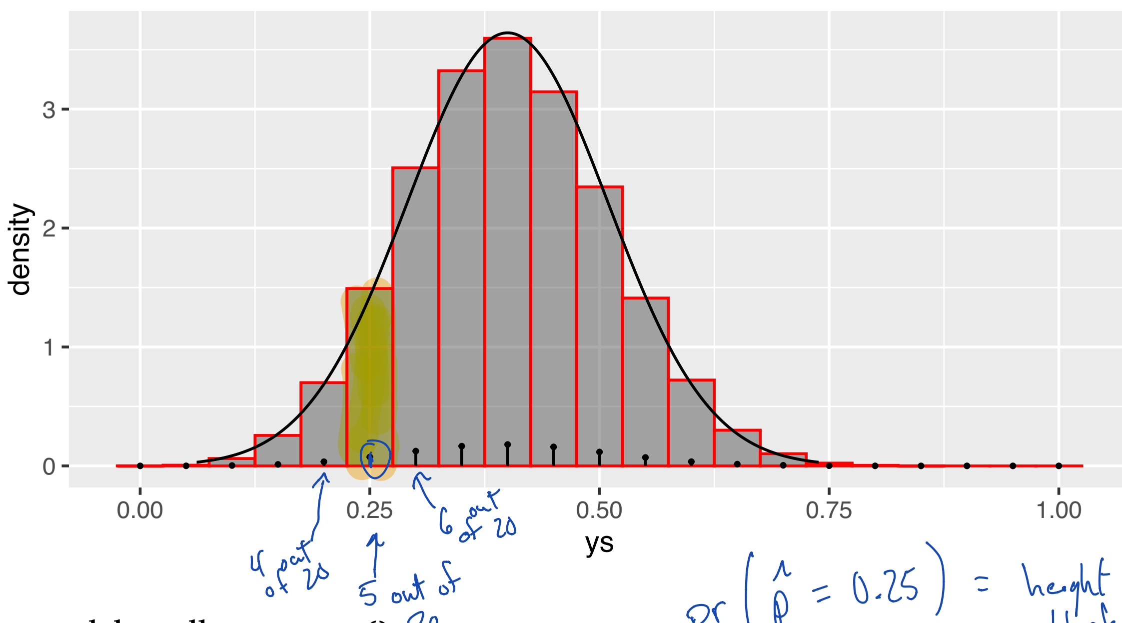
- continuity correction

  Here is the null distribution that goes with $\mathbf{H}_0: p = .4$ when $n = 20$:

Say $\hat{p} = 0.25$




$pr(\hat{p} = 0.25)$ = height of circled black segment = area of shaded rect.

**one-command does all**: `prop.test()`



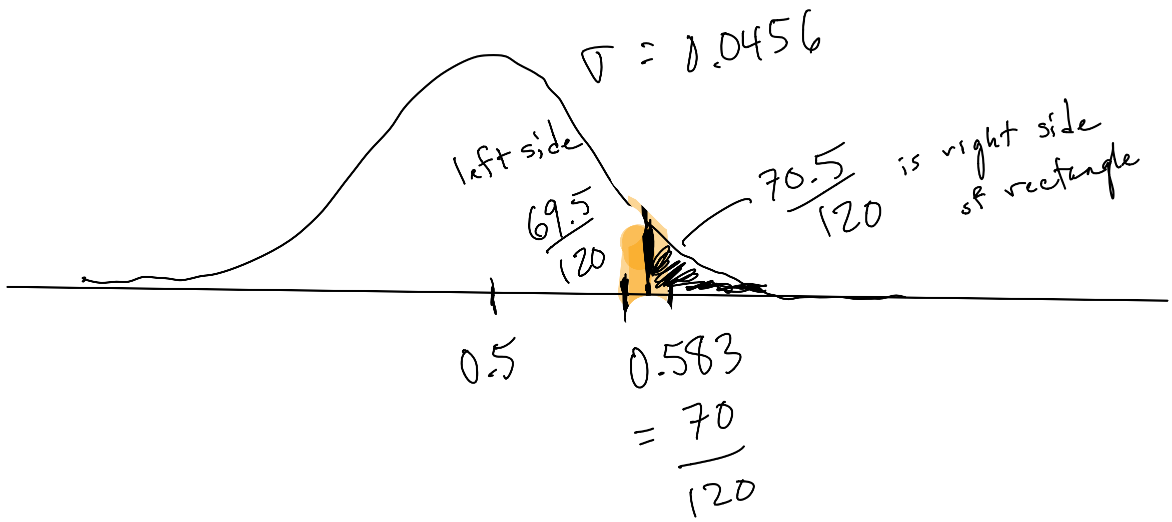
$Pr(\hat{p} = 0.25) =$ ?r of normal dist between 5.5 & 6.5

Monday:

$H_0: p = 0.5$ vs. $H_a: p > 0.5$

Data $\hat{p} = \frac{70}{120} \approx 0.583$

Used normal dist Norm$(0.5, 0.0456)$




area $\approx 1 - \text{pnorm}(0.583, 0.5, 0.0456)$

better approach using continuity correction

$1 - \text{pnorm}\left(\frac{69.5}{120}, 0.5, 0.0456\right)$