Probability for proportions in a simple random sample

- categorical variable with values "success" and "failure"
- population with proportion p of successes and proportion $1 - p$ of failures
- draw SRS of size n and measure success or failure on each subject
- if population is large compared to n, then a good probability model for the count X of successes is the binomial distribution $B(n, p)$
- for a random variable X with binomial distribution $B(n, p)$, have

$$\mu_X = np \qquad \sigma_X^2 = np(1-p) \qquad \sigma_X = \sqrt{np(1-p)}$$

- another random variable is the *sample proportion* $\hat{p}$ for each value of X
- $\hat{p}$ and X are related by $\hat{p} = \dfrac{X}{n} = \dfrac{1}{n}X$
- can get mean and standard deviation for $\hat{p}$ from mean and standard deviation for X:

$$\mu_{\hat{p}} = \frac{1}{n}\mu_X = \frac{1}{n}np = p$$

$$\sigma_{\hat{p}}^2 = \frac{1}{n^2}\sigma_X^2 = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

- $\mu_{\hat{p}} = p$ tells us $\hat{p}$ is an *unbiased* estimator of p
- $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$ tells us we can make variability in $\hat{p}$ small by making n big