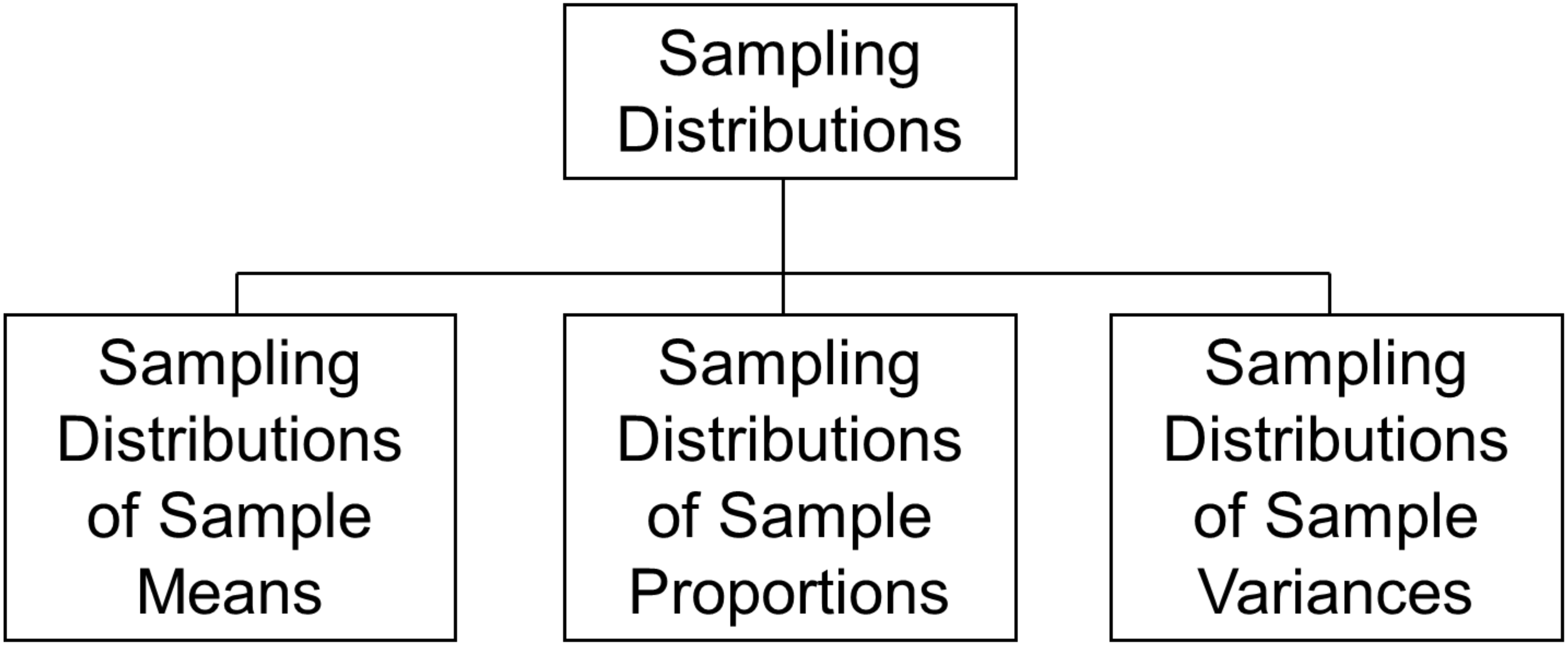
Section 6.3 Sampling Distributions of Sample Proportions



Sampling Distributions of Sample Proportions

P = the proportion of the population having some characteristic

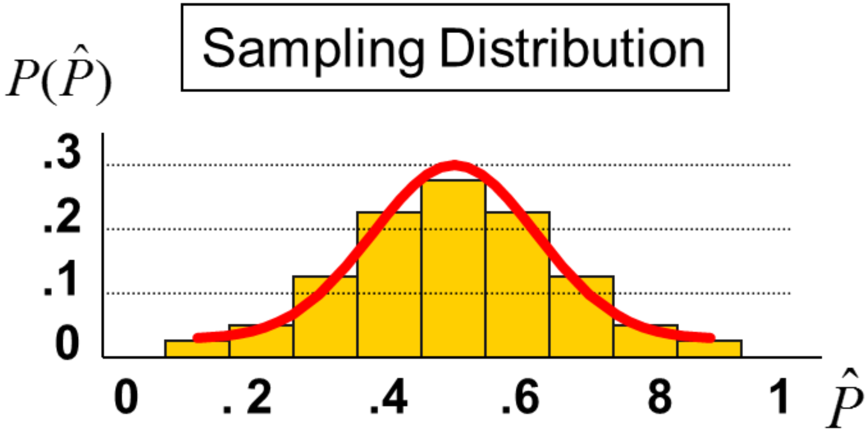
- Sample proportion $(\hat{p})$ provides an estimate of *P*:

$$\hat{p} = \frac{X}{n} = \frac{\text{number of items in the sample having the characteristic of interest}}{\text{sample size}}$$

- $0 \leq ?\hat{p} \leq$
- $\hat{p}$ has a binomial distribution, but can be approximated by a normal distribution when $nP(1-P) > 5$

Sampling Distribution of *p* Hat

- Normal approximation:



Properties: $E(\hat{p}) = P$ and $\sigma_{\hat{p}} = \sqrt{\frac{P(1-P)}{n}}$

(where *P* = population proportion)

Z-Value for Proportions

Standardize $\hat{p}$ to a Z value with the formula:

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

Where the distribution of Z is a good approximation to the standard normal distribution if $nP(1-P) > 5$

Example 2 (1 of 3)

- If the true proportion of voters who support Proposition *A* is *P* = 0.4, what is the probability that a sample of size 200 yields a sample proportion between 0.40 and 0.45?

- i.e.: **if *P* = 0.4 and *n* = 200, what is**

$$\boldsymbol{P\left(0.40 \le \hat{p} \le 0.45\right)?}$$

Example 2 (2 of 3)

- **if P = 0.4 and n = 200, what is**

$$\boldsymbol{P\left(0.40 \le \hat{p} \le 0.45\right)?}$$

Find $\sigma_{\hat{p}}$: $\sigma_{\hat{p}} = \sqrt{\dfrac{P(1-P)}{n}} = \sqrt{\dfrac{.4(1-.4)}{200}} = .03464$

Convert to standard normal:

$$P(.40 \le \hat{p} \le .45) = P\left(\frac{.40-.40}{.03464} \le Z \le \frac{.45-.40}{.03464}\right)$$

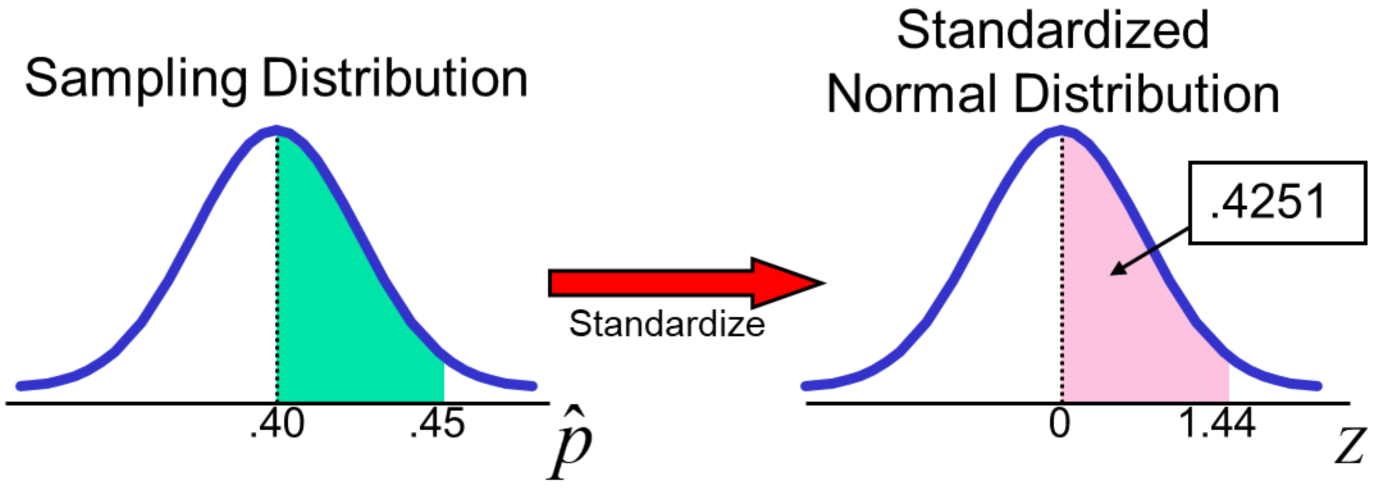
$$= P(0 \le Z \le 1.44)$$

Example 2 (3 of 3)

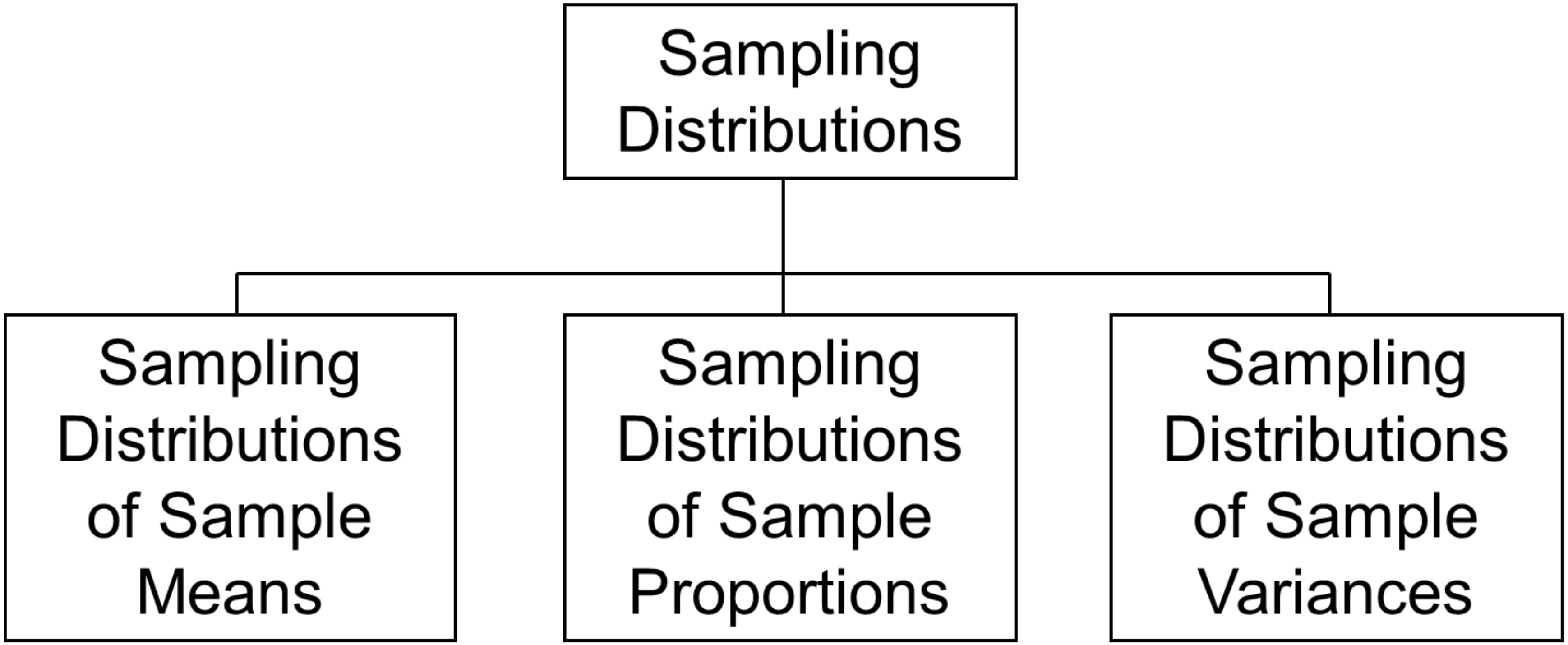
- **if *P* = 0.4 and *n* = 200, what is**

  $$P\left(0.40 \le \hat{p} \le 0.45\right)?$$

Use standard normal table: $P\left(0 \le Z \le 1.44\right) = \boxed{.4251}$



Section 6.4 Sampling Distributions of Sample Variances



Sample Variance

- Let $x_1, x_2, \ldots, x_n$ be a random sample from a population. The sample variance is

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

- the square root of the sample variance is called the sample standard deviation

- the sample variance is different for different random samples from the same population

Sampling Distribution of Sample Variances

- The sampling distribution of s^2 has mean σ^2

$$E(s^2) = \sigma^2$$

- If the population distribution is normal, then

$$Var(s^2) = \frac{2\sigma^4}{n-1}$$

Chi-Square Distribution of Sample and Population Variances

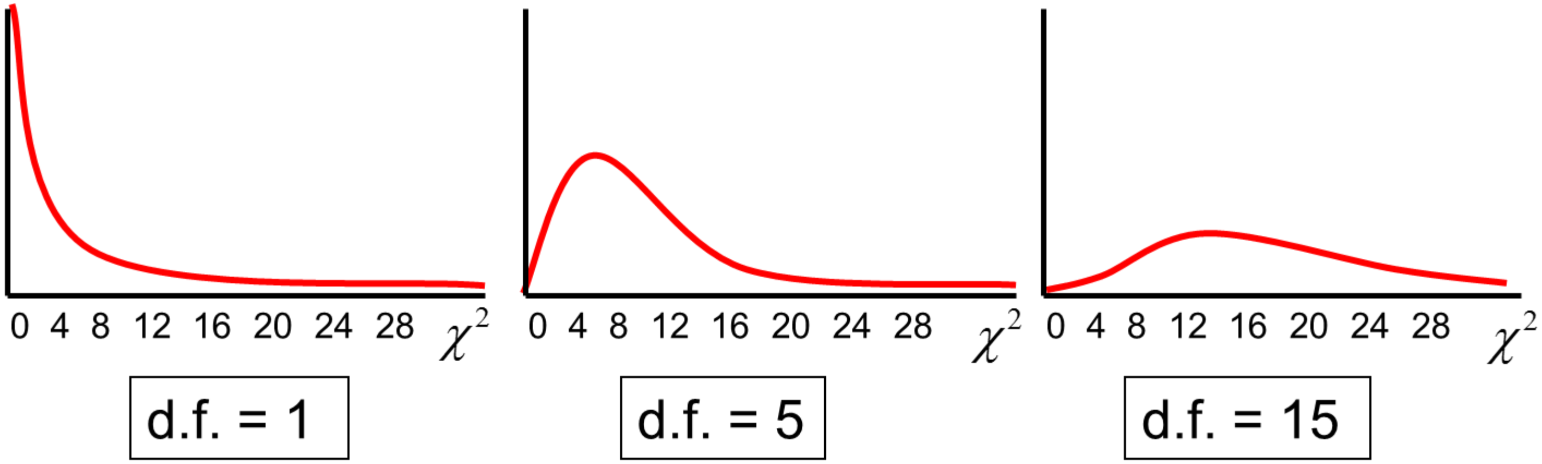
- If the population distribution is normal then

$$\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma^2}$$

has a chi-square (χ^2) distribution
with $n - 1$ degrees of freedom

The Chi-Square Distribution

- The chi-square distribution is a family of distributions, depending on degrees of freedom:
- d.f. = $n - 1$



- Text Appendix Table 7 contains chi-square probabilities

Degrees of Freedom (df)

Idea: Number of observations that are free to vary after sample mean has been calculated

Example: Suppose the mean of 3 numbers is 8.0

Let $X_1 = 7$
Let $X_2 = 8$
What is $\boldsymbol{X_3}$?

⇨

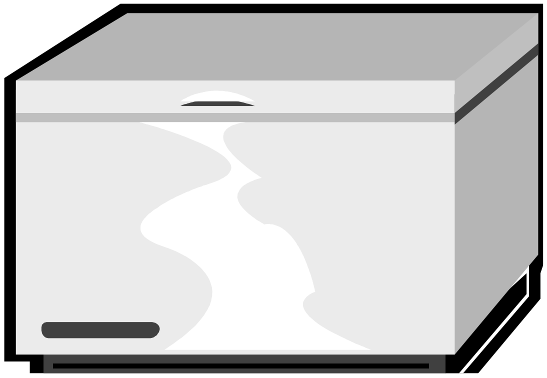
If the mean of these three values is 8.0,
then X_3 must be 9
(i.e., X_3 is not free to vary)

Here, n = 3, so degrees of freedom = n − 1 = 3 − 1 = 2

(2 values can be any numbers, but the third is not free to vary for a given mean)

Chi-Square Example (1 of 2)

- A commercial freezer must hold a selected temperature with little variation. Specifications call for a standard deviation of no more than 4 degrees (a variance of 16 degrees^2).
- A sample of 14 freezers is to be tested
- What is the upper limit (*K*) for the sample variance such that the probability of exceeding this limit, given that the population standard deviation is 4, is less than 0.05?

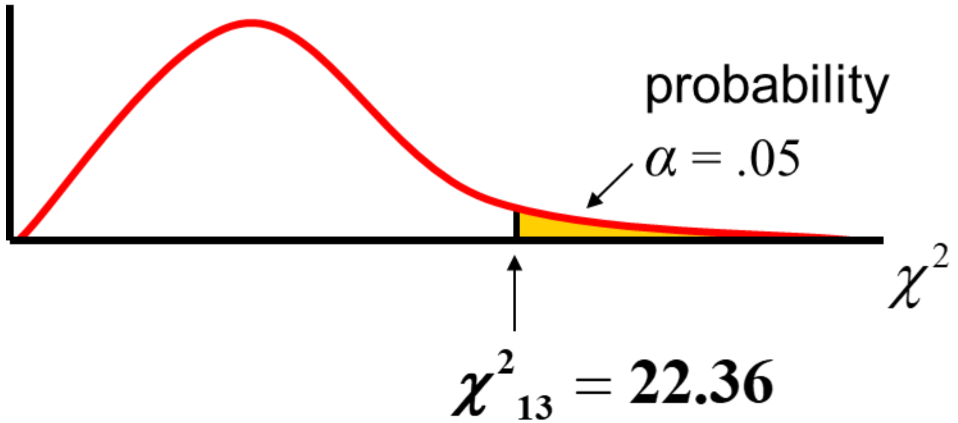


Finding the Chi-Square Value

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$ Is chi-square distributed with $(n-1) = 13$ degrees of freedom

- Use the the chi-square distribution with area 0.05 in the upper tail:

$$\chi^2_{13} = 22.36 \ (\alpha = .05 \text{ and } 14 - 1 = 13 \text{ d.f.})$$



Chi-Square Example (2 of 2)

$$\chi^2_{13} = 22.36 \ (\alpha = .05 \text{ and } 14-1=13 \text{ d.f.})$$

So: $$P(s^2 > K) = P\left(\frac{(n-1)s^2}{16} > \chi^2_{13}\right) = 0.05$$

or $$\frac{(n-1)K}{16} = 22.36$$ (where n = 14)

so $$K = \frac{(22.36)(16)}{(14-1)} = 27.52$$

If s^2 from the sample of size n = 14 is greater than 27.52, there is strong evidence to suggest the population variance exceeds 16.