

# The Sampling Distribution of the Sample Proportion

David Armstrong

UCI

# Binomial Distribution

Suppose we are interested in  $X$ , where  $X$  = the number of successes in  $n$  independent trials.

- Fixed number of trials
- Fixed probability of success
- Trials are independent
- $X \sim \text{Binomial}(n, p)$

$$E(X) = np.$$

$$\text{VAR}(X) = np(1 - p) = npq$$

# Normal Approximation to the Binomial

When you have a large SRS from a large population:

- Same assumptions as the Binomial Distribution
- Rule of Thumb
  - Expected number of successes =  $np \geq 10$
  - Expected number of failures =  $nq \geq 10$
- We would have to use MANY equations to find cumulative probabilities if we used the Binomial Distribution.
- We would only have to use 1 equation to find cumulative probabilities if we approximated the Binomial distribution with the Normal distribution.

Suppose we are interested in  $X$ :

- $X$  = the number of successes in  $n$  independent trials.
- We say: " $X$  follows an Approximately Normal distribution with mean of the number of successes equal to  $np$  and standard deviation of the number of successes equal to the square root of  $npq$ "
- $X \sim \text{Approximately Normal}(\mu_X = np, \sigma_X = \sqrt{npq})$

$$E(X) = np.$$

$$\text{VAR}(X) = np(1 - p) = npq$$

- To calculate probabilities we

**STANDARDIZE**

$$Z = \frac{X - \mu_X}{\sigma_X} = \frac{X - np}{\sqrt{npq}}$$

Example: Eighty percent of all patrons at a local restaurant request the non-smoking section. Suppose we randomly select 7 customers.

Which statement below describe the correct distribution of  $X$ = number of patrons that request the non-smoking section?

- A.  $X \sim B(7, 0.80)$
- B.  $X \sim AN(5.6, 1.058)$
- C.  $X \sim N(5.6, 1.058)$
- D.  $X \sim B(5.6, 1.058)$
- E.  $X \sim AN(7, 0.80)$

What is the probability that at least 6 of the 7 customers selected will request the non-smoking section?

Suppose we now take a larger random sample of 119 customers.

What are the mean and standard deviation of the number that request the non-smoking section?

Using the mean and standard deviation from above, what is the probability that at most 100 of the 119 customers selected will request the non-smoking section?

## Linear Transformation of the Normal Approximation

Let  $X \sim \text{Approximately Normal}(\mu_X = np, \sigma_X = \sqrt{npq})$

Let  $\hat{p} = \frac{X}{n}$ .

- Calculate  $E(\hat{p})$
- Calculate  $\text{VAR}(\hat{p})$
- Calculate the standard deviation of  $\hat{p}$

# The Distribution of a Sample Proportion

Suppose we are interested in  $\hat{p} = \frac{X}{n}$  and you have a large SRS from a large population:

- Fixed number of trials
- Fixed probability of success
- Fixed probability of failure
- Trials are independent
- $\hat{p}$  = the sample proportion of successes in  $n$  independent trials.
- We say: “p-hat follows an Approximately Normal distribution with mean of the sample proportion of successes equal to  $p$  and standard deviation of the sample proportion of successes equal to the square root of  $pq$  over  $n$ ”

Then:  $\hat{p} \sim \text{Approximately Normal}\left(\mu_{\hat{p}} = p, \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}\right)$

- To calculate probabilities we **STANDARDIZE**

$$Z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

- Conservative Rule of Thumb

$$\text{Observed number of successes} = X \geq 10$$

$$\text{Observed number of failures} = n - X \geq 10$$

Ex: The Associated Press reported that 71% of Americans ages 25 and older are overweight. A researcher wants to know whether the proportion of such individuals in his state that are overweight differs from the national proportion. A random sample of 600 adults in his state results in 405 who are classified as overweight.

- a. What is the sample proportion of overweight Americans?
- b. Check and verify all of the assumptions and conditions.
- c. Describe the sampling distribution of the sample proportion for size 600 using the appropriate notation.
- d. Find the probability that at most 405 of the 600 sampled adults are classified overweight.

Ex: According to the 2001 Youth Risk Behavior Surveillance by the Center for Disease Control and Prevention, 39% of the 10th-graders surveyed said that they watch three or more hours of television on a typical school day. Assume that this percentage is true for the current population of all 10th graders. Suppose in a random sample of 200 10th-graders, 86 watched three or more hours of television on a typical school day.

- Find the probability that 86 or more out of the 200 students watched three or more hours of television on a typical day.

Ex: A nationwide survey by the University of Connecticut Center for Survey Research and Analysis found that 30% of men aged 18 to 29 had tattoos in 2002. Suppose this result holds true for the current population of all men in this age group.

- Find the probability that in a random sample of 500 men aged 18 to 29, between 28.4% and 32.6% have tattoos.