Lecture 8

David Armstrong

STATS 67 - UCI

Recall: Binomial Distribution

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

- Fixed number of trials

  n trials

- Fixed probability of success

  p

- Fixed probability of failure

  $q = 1 - p$

- Trials are independent

- $X \sim Binomial(n, p)$

  $\mathrm{E}(X) = np.$

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

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.

  $\mathrm{E}(X) = np.$

  $\mathrm{VAR}(X) = np(1-p) = npq$
- $X \sim$ Approximately Normal$(\mu_X = np, \sigma_X = \sqrt{npq})$

  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"

  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?

The Distribution of a Sample Proportion

- Fixed number of trials
- Fixed probability of success
- Fixed probability of failure
- Trials are independent
- Recall that: $X \sim Binomial(n, p)$
- Recall that: $X \sim$ Approximately Normal$\left(\mu_X = np, \sigma_X = \sqrt{npq}\right)$
  if the expected number of successes $= np \geq 10$ AND
  the expected number of failures $= nq \geq 10$

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

- $\hat{p}$ = the sample proportion of successes in n independent trials.
- Then: $\hat{p} \sim$ Approximately Normal$\left(\mu_{\hat{p}} = p, \sigma_{\hat{p}} = \sqrt{\dfrac{pq}{n}}\right)$

  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"

  To calculate probabilities we **STANDARDIZE**

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

- Conservative Rule of Thumb

  Observed number of successes $= X \geq 10$

  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.

a. Check the general properties and describe the sampling distribution of the sample proportion of size 200 using the appropriate notation.

b. 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.

Supplimental Questions

Ex: Suppose a student takes a quiz that consists of five multiple choice questions. Each question has four options. Unfortunately the student has not reviewed the material and has to guess on each question. Let X = the number of questions the student guesses correctly.

Write the distribution for the number of questions that the student answers correctly:

Draw out the sampling distribution for the number of questions that the student answers correctly:

Draw out the corresponding histogram below:

What is the probability that the student answers none of the questions correctly?

What is the probability that the student answers one of the questions correctly?

What is the probability that the student answers two of the questions correctly?

What is the probability that the student answers three of the questions correctly?

What is the probability that the student answers four of the questions correctly?

What is the probability that the student answer five of the questions correctly?

What is the probability that the student answers at most 3 questions correctly?

What is the probability that the student answers at least 4 questions correctly?

Ex: The juror pool for an upcoming murder trial of a celebrity actor contains the names of 100,000 individuals in the population who may be called for jury duty. The percent of available jurors on the population list who are Hispanic is 40%. Suppose we take a random sample of 300 potential jurors. Let X=the number of Hispanics selected to be jurors for this jury.
What is the mean and standard deviation of the number of Hispanics selected?
a) 110, 8.49
b) 180, 13.42
c) 120, 10.95
d) 180, 8.66
e) 120, 8.49

Using the information you found in the problem above, find the probability that 110 or fewer Hispanics will be selected for jury duty.

Ex: The American Journal of Public Health published a study in the mid 90's showing a relationship between passive smoking and nasal allergies in American teenagers. The study revealed that 65% of teens from heavy-smoking families had indications of nasal allergies on physical examinations. In a random sample of 400 American teenagers living in heavy-smoking households in San Diego, 285 had indications of nasal allergies on their physical examinations.
a. Are we able to use the normal approximation? (Check the rule of thumb and use the appropriate notation)

b. Find the mean and standard deviation for the number of American teenagers having nasal allergies.

c. Is the San Diego County result of 285 unusually high? Why or why not? Check the probability of at least 285: