MATH 1700

Instructor: Mehdi Maadooliat

Chapter 9A
Inferences Involving One Population

Department of Mathematical and Statistical Sciences

INFERENCES ABOUT THE BINOMIAL PROBABILITY OF SUCCESS, p

- Perhaps the most common inference involves the **binomial parameter** p, the "probability of success." Yes, every one of us uses this inference, even if only casually.

- In thousands of situations, we are concerned about something either "happening" or "not happening." There are only two possible outcomes of concern, and that is the fundamental property of a **binomial experiment**.

- The other necessary ingredient is multiple independent trials.

INFERENCES ABOUT THE BINOMIAL PROBABILITY OF SUCCESS, p

- **These are the makings of a binomial inference:**
  1. **Asking five people whether they are "for" or "against" some issue can create five independent trials;**
  2. **If 200 people are asked the same question, 200 independent trials may be involved;**
  3. **If 30 items are inspected to see if each "exhibits a particular property" or "not," there will be 30 repeated trials;**

- **The binomial parameter p is defined to be the probability of success on a single trial in a binomial experiment.**

INFERENCES ABOUT THE BINOMIAL PROBABILITY OF SUCCESS, p

Sample Binomial Probability

$$p' = \frac{x}{n} \quad (9.3)$$

where the **random variable** x represents the number of successes that occur in a sample consisting of n trials

- We know that the mean and standard deviation of the binomial variable x are found by the formulas
  - $\mu = np$, and
  - $\sigma = \sqrt{npq}$

    where $q = 1 - p$.

INFERENCES ABOUT THE BINOMIAL PROBABILITY OF SUCCESS, p

MARQUETTE UNIVERSITY
Be The Difference.

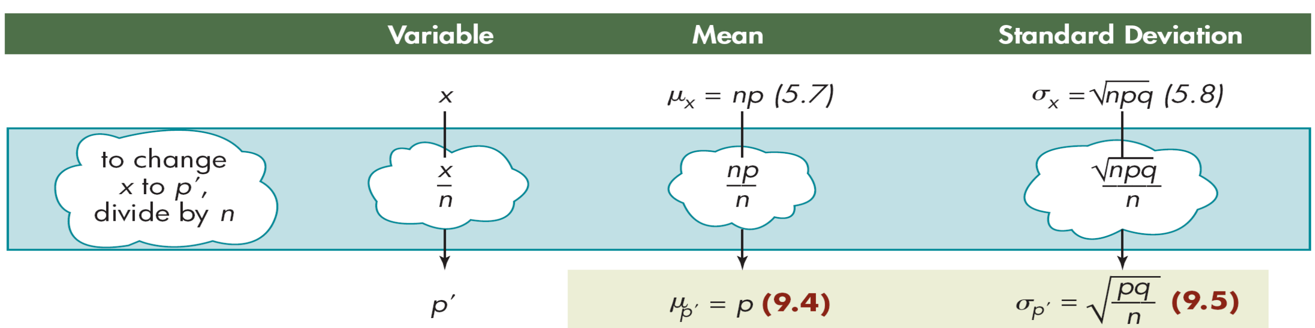
- **The distribution of x is considered to be approximately normal if n is greater than 20 and if np and nq are both greater than 5.**

- **This commonly accepted rule of thumb allows us to use the standard normal distribution to estimate probabilities for the binomial random variable x, the number of successes in n trials, and to make inferences concerning the binomial parameter p, the probability of success on an individual trial.**

INFERENCES ABOUT THE BINOMIAL PROBABILITY OF SUCCESS, p

- **Generally, working with the distribution of p' (the observed probability of occurrence) is easier and more meaningful with x (the number of occurrences).**

- **Consequently, we will convert formulas $\sigma = \sqrt{npq}$ and $\mu = np$ from units of x (integers) to units of proportions (percentages expressed as decimals) by dividing each formula by n, as shown in Table 9.1.**



Formulas (9.4) and (9.5)

INFERENCES ABOUT THE BINOMIAL PROBABILITY OF SUCCESS

- If a random sample of size n is selected from a large population with $p = P(\text{success})$, then the sampling distribution of p' has
  - A mean, $\mu_{p'}$ equal to p
  - A standard error, $\sigma_{p'}$ equal to $\sqrt{\frac{pq}{n}}$
  - an approximately normal distribution if n is sufficiently large
- In practice, using these guidelines will ensure normality:
  - The sample size is greater than 20.
  - The products np and nq are both greater than 5.
  - The sample consists of less than 10% of the population.
- We are now ready to make inferences about the population parameter p. Use of the z distribution.

Confidence Interval Procedure

CONFIDENCE INTERVAL PROCEDURE

- **Inferences concerning the population binomial parameter $p, P\ (\text{success})$, are made using procedures that closely parallel the inference procedures used for the population mean μ.**

- **When we estimate the population proportion p, we will base our estimations on the unbiased estimator p'.**

- **The point estimate, the sample statistic p', becomes the center of the confidence interval, and the maximum error of the estimate is a multiple of the standard error.**

CONFIDENCE INTERVAL PROCEDURE

- **The level of confidence determines the confidence coefficient, the number of multiples of the standard error.**

Confidence Interval for a Proportion

$$p' - z(\alpha/2)\left(\sqrt{\frac{p'q'}{n}}\right) \quad \text{to} \quad p' + z(\alpha/2)\left(\sqrt{\frac{p'q'}{n}}\right) \qquad (9.6)$$

where $p' = \frac{x}{n}$ and $q' = 1 - p'$

- **Notice that the standard error, $\sqrt{\frac{pq}{n}}$, has been replaced by**

$$\sqrt{\frac{p'q'}{n}}$$

EXAMPLE 8 – CONFIDENCE INTERVAL FOR p

- **In a discussion about the cars that fellow students drive, several statements were made about types, ages, makes, colors, and so on. Dana decided he wanted to estimate the proportion of cars students drive that are convertibles, so he randomly identified 200 cars in the student parking lot and found 17 to be convertibles.**

- **Find the 90% confidence interval for the proportion of cars driven by students that are convertibles.**

EXAMPLE 8 – SOLUTION

Step 1 The Set-Up:

- **Describe the population parameter of interest.**
- p, **the proportion (percentage) of students' cars that are convertibles**

Step 2 The Confidence Interval Criteria:

a. **Check the assumptions.**
   - **The sample was randomly selected, and each student's response is independent of those of the others surveyed.**

EXAMPLE 8 – SOLUTION

b. Identify the probability distribution and the formula to be used.

- The standard normal distribution will be used with formula (9.6) as the test statistic. p' is expected to be approximately normal because:

1. $n = 200$ is greater than 20, and
2. Both np [approximated by $np' = 200\left(\frac{17}{200}\right) = 17$]

   and nq [approximated by $nq' = 200\left(\frac{183}{200}\right) = 183$]

   are greater than 5.

c. State the level of confidence:

$$1 - \alpha = 0.90$$

EXAMPLE 8 – SOLUTION CONT.

Step 3 The Sample Evidence:

- **Collect the sample information.**
- $n = 200$ **cars, and** $x = 17$ **were convertibles:**

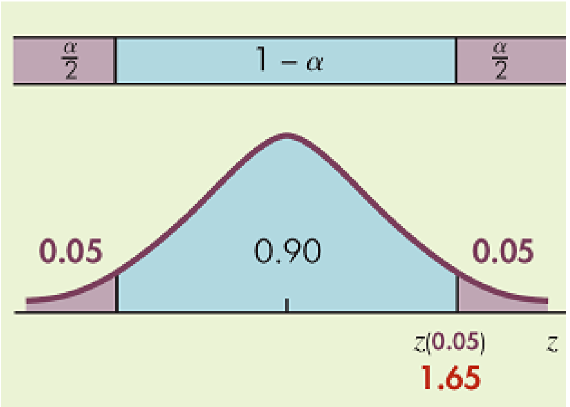
$$p' = \frac{x}{n} = \frac{17}{200}$$

$$= 0.085$$

Step 4 The Confidence Interval:

a. Determine the confidence coefficient.

This is the z-score
[$z(\alpha/2)$, "z of one-half of alpha"]



EXAMPLE 8 – SOLUTION

b. Find the maximum error of estimate.

Use the maximum error part of formula (9.6):

$$E = z(\alpha/2)\left(\sqrt{\frac{p'q'}{n}}\right) = 1.65\left(\sqrt{\frac{(0.085)(0.915)}{200}}\right)$$

$$= (1.65)\sqrt{0.000389}$$

$$= (1.65)(0.020)$$

$$= \mathbf{0.033}$$

EXAMPLE 8 – SOLUTION

- **c. Find the lower and upper confidence limits.**

$p' - E$	**to**	$p' + E$
$0.085 - 0.033$	**to**	$0.085 + 0.033$
0.052	**to**	0.118

Step 5 The Results:

- **State the confidence interval.**
  0.052 **to** 0.118 **is the** 90% **confidence interval for** $p = P\,(\text{drives convertible})$**. That is, the true proportion of students who drive convertibles is between** 0.052 **and** 0.118**, with** 90% **confidence.**

Determining the Sample Size

DETERMINING THE SAMPLE SIZE

- By using the maximum error part of the confidence interval formula, it is possible to determine the size of the sample that must be taken in order to estimate p with a desired accuracy.

- Here is the formula for the maximum error of estimate for a proportion:

$$E = z(\alpha/2)\left(\sqrt{\frac{pq}{n}}\right)$$

DETERMINING THE SAMPLE SIZE

- The desired preciseness will determine the maximum error of estimate, E. (Remember that we are estimating p, the binomial probability; therefore, E will typically be expressed in hundredths.)

Sample Size for $1 - \alpha$ Confidence Interval of p

$$n = \frac{[z(\alpha/2)]^2 \cdot p^* \cdot q^*}{E^2} \quad (9.8)$$

where p^* and q^* are provisional values of p and q used for planning

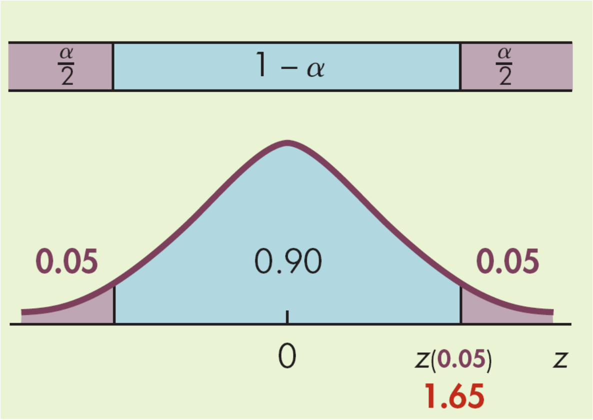
EXAMPLE 10 – SAMPLE SIZE FOR ESTIMATING p (NO PRIOR INFORMATION)

Determine the sample size that is required to estimate the true proportion of community college students who are blue-eyed if you want your estimate to be within 0.02 with 90% confidence.

- **If no provisional values for p and q are available, then use $p^* = 0.5$ and $q^* = 0.5$. Using $p^* = 0.5$ is safe because it gives the largest sample size of any possible value of p.**

Solution:

- **Step 1 : The level of confidence is $1 - \alpha = 0.90$; therefore, the confidence coefficient is $z(\alpha/2) = z(0.05) = 1.65$**



EXAMPLE 10 – SOLUTION

Step 2: The desired maximum error is $E = 0.02$

Step 3: No estimate was given for p, so use $p^* = 0.5$ and $q^* = 1 - p^* = 0.5$.

Step 4: Use formula (9.8) to find n:

$$n = \frac{[z(\alpha/2)]^2 \cdot p^* \cdot q^*}{E^2} \; : \; n = \frac{(1.65)^2 \cdot 0.5 \cdot 0.5}{(0.02)^2}$$

$$= \frac{0.680625}{0.0004}$$

$$= 1701.56 = \mathbf{1702}$$

Hypothesis-Testing Procedure

HYPOTHESIS-TESTING PROCEDURE

- When the binomial parameter p is to be tested using a hypothesis-testing procedure, we will use a test statistic that represents the difference between the observed proportion and the hypothesized proportion, divided by the standard error.
- This test statistic is assumed to be normally distributed when the null hypothesis is true, when the assumptions for the test have been satisfied, and when n is sufficiently large ($n > 20$, $np > 5$, and $nq > 5$).

Test Statistic for a Proportion

$$z\bigstar = \frac{p' - p}{\sqrt{\frac{pq}{n}}} \quad \text{with } p' = \frac{x}{n} \tag{9.9}$$

EXAMPLE 12 – ONE-TAILED HYPOTHESIS TEST FOR PROPORTION p

- **Many people sleep late on the weekends to make up for "short nights" during the workweek. The Better Sleep Council reports that 61% of us get more than 7 hours of sleep per night on the weekend.**

- **A random sample of 350 adults found that 235 had more than 7 hours of sleep each night last weekend. At the 0.05 level of significance, does this evidence show that more than 61% sleep 7 hours or more per night on the weekend?**

EXAMPLE 12 – SOLUTION

Step 1 The Set-Up:

a. Describe the population parameter of interest.

p, **the proportion of adults who get more than** 7 **hours of sleep per night on weekends**

b. State the null hypothesis (H_o**) and the alternative hypothesis (**H_a**).**

- H_o: $p = 0.61 (\leq)$ (no more than 61%)
- H_a: $p > 0.61$ (more than 61%)

EXAMPLE 12 – SOLUTION

Step 2 The Hypothesis Test Criteria:

a. Check the assumptions.

- **The random sample of 350 adults was independently surveyed.**

b. Identify the probability distribution and the test statistic to be used.

- **The standard normal z will be used with formula (9.9). Since $n = 350$ is greater than 20 and both $np = (350)(0.61) = 213.5$ and $nq = (350)(0.39) = 136.5$ are greater than 5, p' is expected to be approximately normally distributed.**

c. Determine the level of significance:

- $\alpha = 0.05$.

EXAMPLE 12 – SOLUTION

Step 3 The Sample Evidence:

a. Collect the sample information:

- $n = 350$ **and** $x = 235$**:**

$$p' = \frac{x}{n} = \frac{235}{350} = 0.671$$

b. Calculate the value of the test statistic.

- **Use formula (9.9):**

$$z\bigstar = \frac{p' - p}{\sqrt{\frac{pq}{n}}} : z\bigstar = \frac{0.671 - 0.61}{\sqrt{\frac{(0.61)(0.39)}{350}}}$$

$$= \frac{0.061}{\sqrt{0.0006797}}$$

$$= \frac{0.061}{0.0261} = \mathbf{2.34}$$

EXAMPLE 12 – SOLUTION

Step 4 The Probability Distribution:

Using the classical procedure:

a. Determine the critical region and critical value (s).

- **The critical region is the right-hand tail because H_a expresses concern for values related to "more than." The critical value is obtained from Table 4A: $z(0.05) = 1.65$**

EXAMPLE 12 – SOLUTION

b. Determine whether or not the calculated test statistic is in the critical region.

- $z\bigstar$ **is in the critical region, as shown in red in the accompanying figure.**

Step 5 The Results:

a. State the decision about H_o: Reject H_o.

b. State the conclusion about H_a.

- **There is sufficient reason to conclude that the proportion of adults in the sampled population who are getting more than 7 hours of sleep nightly on weekends is significantly higher than 61% at the 0.05 level of significance.**

EXAMPLE 13 – TWO-TAILED HYPOTHESIS TEST FOR PROPORTION p

- **While talking about the cars that fellow students drive, Tom claimed that 15% of the students drive convertibles. Jody finds this hard to believe, and she wants to check the validity of Tom's claim using Dana's random sample.**

- **At a level of significance of 0.10, is there sufficient evidence to reject Tom's claim if there are 17 convertibles in his sample of 200 cars?**

EXAMPLE 13 – SOLUTION

Step 1 The Set-Up:

a. Describe the population parameter of interest.

- $p = P(\text{student drives convertible})$

b. State the null hypothesis (H_o) and the alternative hypothesis (H_a).

- $H_o: p = 0.15$ (15% **do drive convertibles)**
- $H_a: p \neq 0.15$ **(the percentage is different from** 15%**)**

EXAMPLE 13 – SOLUTION

Step 2 The Hypothesis Test Criteria:

a. Check the assumptions.

- **The sample was randomly selected, and each subject's response was independent of other responses.**

b. Identify the probability distribution and the test statistic to be used.

- **The standard normal z and formula (9.9) will be used. Since $n = 200$ is greater than 20 and both np and nq are greater than 5, p is expected to be approximately normally distributed.**

c. Determine the level of significance: $\alpha = 0.10$

EXAMPLE 13 – SOLUTION

MARQUETTE UNIVERSITY
Be The Difference.

Step 3 The Sample Evidence:

a. Collect the sample information:

- $n = 200$ and $x = 17$

$$p' = \frac{x}{n} = \frac{17}{200} = 0.085$$

b. Calculate the value of the test statistic.

- Use formula (9.9):

$$z\star = \frac{p' - p}{\sqrt{\frac{pq}{n}}} \; : \; z\star = \frac{0.085 - 0.150}{\sqrt{\frac{(0.15)(0.85)}{200}}}$$

$$= \frac{-0.065}{\sqrt{0.00064}} = \frac{-0.065}{0.02525} = -\mathbf{2.57}$$

EXAMPLE 13 – SOLUTION

Step 4 The Probability Distribution:

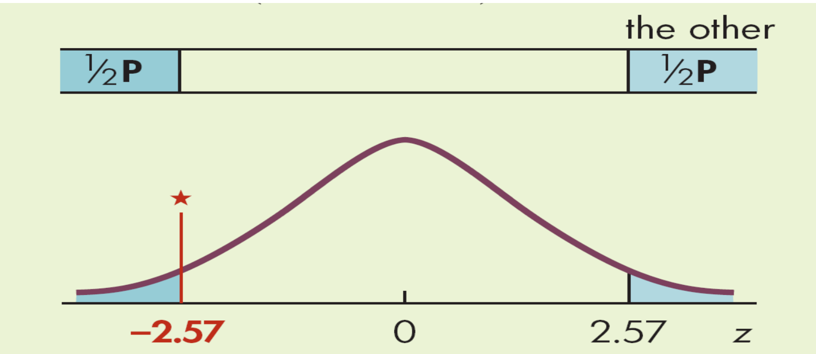
Using the p-value procedure:

a. Calculate the $p - value$ for the test statistic.

- **Use both tails because H_a expresses concern for values related to "different from."**
- $p - value = P(z < -2.57) + P(z > 2.57)$

$$= 2 \times P(z < -2.57)$$

as shown in the figure $= 2 \times 0.0051 = 0.0102$



- **The $p - value$ is smaller than α.**

EXAMPLE 13 – SOLUTION

Step 5 The Results:

a. State the decision about H_o: **Reject H_o.**

b. State the conclusion about H_a.

- **There is sufficient evidence to reject Tom's claim and conclude that the percentage of students who drive convertibles is different from 15% at the 0.10 level of significance.**

QUESTIONS?

- ANY QUESTION?