Binomial Test:

- Uses data to test whether a population proportion (p) matches a null expectation (p_0) for the proportion

- Binomial sampling distribution provides the hypotheses:

H_0: The relative frequency of successes in the population is p_0
H_A: The relative frequency of successes in the population is **NOT** p_0

Proportions

Binomial Test:

- Uses data to test whether a population proportion (p) matches a null expectation (p_0) for the proportion
- Binomial sampling distribution provides the hypotheses:

H_0: The relative frequency of successes in the population is p_0

H_A: The relative frequency of successes in the population is **NOT** p_0

The core of using the binomial to test proportions is:

- *The sampling distribution is the Binomial distribution (assuming that H_0 is true.)*
- *Calculate exact p-value from null distribution*

Proportions

Example: Imagine a student taking a multiple choice test before starting a statistics class. Each of the 10 questions on the test have only 5 possible answers, only one of which is correct. This student gets 4 answers right. Can we deduce from this that this student knows anything at all about statistics?

Proportions

Example: Imagine a student taking a multiple choice test before starting a statistics class. Each of the 10 questions on the test have only 5 possible answers, only one of which is correct. This student gets 4 answers right. Can we deduce from this that this student knows anything at all about statistics?

1. Formulate hypothesis

H_0:

H_A:

Proportions

Example: Imagine a student taking a multiple choice test before starting a statistics class. Each of the 10 questions on the test have only 5 possible answers, only one of which is correct. This student gets 4 answers right. Can we deduce from this that this student knows anything at all about statistics?

1. Formulate hypothesis

H_0: Student answered questions randomly

H_A: Student got more answers correct than random

Proportions

Example: Imagine a student taking a multiple choice test before starting a statistics class. Each of the 10 questions on the test have only 5 possible answers, only one of which is correct. This student gets 4 answers right. Can we deduce from this that this student knows anything at all about statistics?

1. Formulate hypothesis

   H_0: Student answered questions randomly

   H_A: Student got more answers correct than random

2. Identify the test statistic

Proportions

Example: Imagine a student taking a multiple choice test before starting a statistics class. Each of the 10 questions on the test have only 5 possible answers, only one of which is correct. This student gets 4 answers right. Can we deduce from this that this student knows anything at all about statistics?

1. Formulate hypothesis

   H_0: Student answered questions randomly

   H_A: Student got more answers correct than random

2. Identify the test statistic

   *Assume null hypothesis is true so $p_0 = 0.20$ (since there are 5 possible answers for each question)

   * binomial test describes the situation

$$H_0: p_0 = 0.2$$
$$H_A: p_0 > 0.2$$

Proportions

<u>Example:</u> Imagine a student taking a multiple choice test before starting a statistics class. Each of the 10 questions on the test have only 5 possible answers, only one of which is correct. This student gets 4 answers right. Can we deduce from this that this student knows anything at all about statistics?

1. <u>Formulate hypothesis</u>

   H_0: Student answered questions randomly

   H_A: Student got more answers correct than random

2. Identify the test statistic

   * Binomial Test

   H_0: $p_0 = 0.2$
   H_A: $p_0 > 0.2$

3. Calculate P-value

Pval=P[4]+P[5]+P[6]+P[7]+P[8]+P[9]+P[10]

$$P(X \geq 4) = \binom{10}{4}(0.2)^4(0.8)^6 + \binom{10}{5}(0.2)^5(0.8)^5 + \ldots + \binom{10}{10}(0.2)^{10}(0.8)^0$$

$= 0.12$

Proportions

Example: Imagine a student taking a multiple choice test before starting a statistics class. Each of the 10 questions on the test have only 5 possible answers, only one of which is correct. This student gets 4 answers right. Can we deduce from this that this student knows anything at all about statistics?

1. Formulate hypothesis

   H_0: Student answered questions randomly

   H_A: Student got more answers correct than random

2. Identify the test statistic

   * Binomial Test

   H_0: $p_0 = 0.2$

   H_A: $p_0 > 0.2$

3. Calculate P-value

Pval=P[4]+P[5]+P[6]+P[7]+P[8]+P[9]+P[10]

$$P(X \geq 4) = \binom{10}{4}(0.2)^4(0.8)^6 + \binom{10}{5}(0.2)^5(0.8)^5 + \ldots + \binom{10}{10}(0.2)^{10}(0.8)^0$$

$$= 0.12$$

4. Compare to fixed significance

   If α **= 0.05, we FTR null.**

It is plausible that the student got four answers correct by randomly guessing

Confidence Interval for a proportion:

(Agresti-Coull confidence interval)

$$p' = \frac{X+2}{n+4}$$

$$\left(p' - Z\sqrt{\frac{p'(1-p')}{n+4}} \right) \le p \le \left(p' + Z\sqrt{\frac{p'(1-p')}{n+4}} \right)$$

Where Z = 1.96 for a 95% confidence interval <- we are done with just using 2*SE

Example 7.3: Male radiologists have long suspected that they tend to have fewer sons than daughters. What is the proportion of males among the offspring of radiologists? In a sample of 87 offspring of "highly irradiated" radiologists, 30 were male (Hama et al, 2001). Assume that this is random sample.

$$\hat{p} = \frac{X}{n} = \frac{30}{87}$$

Proportions

Example 7.3: Male radiologists have long suspected that they tend to have fewer sons than daughters. What is the proportion of males among the offspring of radiologists? In a sample of 87 offspring of "highly irradiated" radiologists, 30 were male (Hama et al, 2001). Assume that this is random sample. What is the 95% confidence interval for this estimate?

$$\hat{p} = \frac{X}{n} = \frac{30}{87} = 0.345$$

$$p' = \frac{X+2}{n+4} = \frac{32}{91} = 0.352$$

Proportions

Example 7.3: Male radiologists have long suspected that they tend to have fewer sons than daughters. What is the proportion of males among the offspring of radiologists? In a sample of 87 offspring of "highly irradiated" radiologists, 30 were male (Hama et al, 2001). Assume that this is random sample. What is the 95% confidence interval for this estimate?

$$\hat{p} = \frac{X}{n} = \frac{30}{87} = 0.345$$

$$p' = \frac{X+2}{n+4} = \frac{32}{91} = 0.352$$

$$p' \pm Z\sqrt{\frac{p'(1-p')}{n+4}} = 0.352 \pm 1.96\sqrt{\frac{0.352(0.648)}{91}}$$

Proportions

Example 7.3: Male radiologists have long suspected that they tend to have fewer sons than daughters. What is the proportion of males among the offspring of radiologists? In a sample of 87 offspring of "highly irradiated" radiologists, 30 were male (Hama et al, 2001). Assume that this is random sample. What is the 95% confidence interval for this estimate?

$$\hat{p} = \frac{X}{n} = \frac{30}{87} = 0.345$$

$$p' = \frac{X+2}{n+4} = \frac{32}{91} = 0.352$$

$$p' \pm Z\sqrt{\frac{p'(1-p')}{n+4}} = 0.352 \pm 1.96\sqrt{\frac{0.352(0.648)}{91}}$$

$$= 0.352 \pm 0.098$$

$$\boxed{0.254 < p < 0.450}$$

A newborn baby whose Apgar score is over 6 is classified as normal and this happens in 80% of births. As a quality control check, an auditor examined the records of 100 births. He would be suspicious if the number of normal births in the sample of 100 births fell above the upper limit of a "95%-normal-range". What is this upper limit (approx)?

a. 112

b. 72

c. 87

d. 8

e. none of these

Babies that have Apgar scores of 6 or lower require more expensive medical care. What is the probability that in the next 10 births, 2 or more babies will have Apgar scores of 6 or lower?

a. 0.2013
b. 0.3758
c. 0.6242
d. 0.0001
e. 0.1536