

# AP Statistics

## Review Week 5

### Inference Proportions

Advanced Placement AAP Review will be held in **room 315** and **312** on Tuesdays and Thursdays.

The week of April 20<sup>th</sup> we will be reviewing **Inference Proportions**.

The session will begin in room 315 with a brief review of the weekly topic.

Instruction will be from 3:15 pm to 3:30 pm

Once we have reviewed the topic you may begin practicing the questions in your review packet.

Answers will be posted in room 315 and 312 all week and will be posted on line after 3:00 pm on Friday the week of review.

If you have difficulty with a question look at the detailed answer postings BEFORE you ask your teacher for help.

Get a hint....**DON'T COPY THE ANSWER!!! THAT IS NOT HELPFUL!!**

When you have completed a question...**REFLECT!!!!** Ask yourself what skill you used to solve that problem and write that down!!

Once we have completed the weekly review, keep it to study from as we get closer to the exam.

## Inference: Proportions

### Brief Review

There are two things you need to be able to do with proportions...

- 1.) Make a confidence interval ... which is a range estimate of the TRUE proportion from your sample.
- 2.) Perform a test of significance ... which is a decision making process.

Before you can do a test or create a confidence interval you have to check your CONDITIONS and ASSUMPTIONS...

- 1.) Random sampling of independent subjects or groups.
- 2.) NP and NQ have to be greater than or equal to 10.

Type I Error: Rejecting a true  $H_0$ ...You think the  $H_0$  is not true but it is...

Type II Error: Not rejecting a false  $H_0$ ...You think the  $H_0$  is false but it's not...

Both of these lead to CONSEQUENCES...Usually MONEY spent on a bad program...or the chance to correct a problem is ignored because you throw away a good program.

The  $H_0$  is always =

The  $H_A$  is <, > or  $\neq$

You need to define P as the TRUE proportion of...CONTEXT.

1 Proportion Z interval ...confidence interval for one proportion

2 Proportion Z interval ...confidence interval for two proportions.

1 Proportion Z Test ...significance test for one proportion

2 Proportion Z Test ...significance test for two proportions.

Alpha is the significance level and the probability of a type I error.

Beta is the probability of a type II error.

## 2007 Exam

4. Independent random samples of 100 luxury cars and 250 non-luxury cars in a certain city are examined to see if they have bumper stickers. Of the 250 non-luxury cars, 125 have bumper stickers and of the 100 luxury cars, 30 have bumper stickers. Which of the following is a 90 percent confidence interval for the difference in the proportion of non-luxury cars with bumper stickers and the proportion of luxury cars with bumper stickers from the population of cars represented by these samples?

(A)  $(0.5 - 0.3) \pm 1.645 \sqrt{\frac{(0.5)(0.5)}{250} + \frac{(0.3)(0.7)}{100}}$

(B)  $(0.5 - 0.3) \pm 1.96 \sqrt{\frac{(0.5)(0.5)}{250} + \frac{(0.3)(0.7)}{100}}$

(C)  $(0.5 - 0.3) \pm 1.645 \sqrt{\left(\frac{155}{350}\right)\left(\frac{195}{350}\right)\left(\frac{1}{250} + \frac{1}{100}\right)}$

(D)  $(0.5 - 0.3) \pm 1.96 \sqrt{\left(\frac{155}{350}\right)\left(\frac{195}{350}\right)\left(\frac{1}{250} + \frac{1}{100}\right)}$

(E)  $(0.5 - 0.3) \pm 1.645 \sqrt{(0.4)(0.6)\left(\frac{1}{250} + \frac{1}{100}\right)}$

21. A city is interested in building a waste management facility in a certain area. One hundred randomly selected residents from this area were asked, "Do you support the city's decision to build a waste management facility in your area?" Of the 100 residents interviewed, 54 said no, 4 said yes, and

42 had no opinion. A large sample  $z$ -confidence interval,  $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ , was constructed from

these data to estimate the proportion of this area's residents who support building a waste management facility in their area. Which of the following statements is correct for this confidence interval?

- (A) This confidence interval is valid because a sample size of more than 30 was used.
- (B) This confidence interval is valid because each area resident was asked the same question.
- (C) The confidence interval is valid because no conditions are required for constructing a large sample confidence interval for a proportion.
- (D) This confidence interval is not valid because the quantity  $n\hat{p}$  is too small.
- (E) This confidence interval is not valid because "no opinion" was included as a response category for the question.

30. A researcher has conducted a survey using a simple random sample of 50 registered voters to create a confidence interval to estimate the proportion of registered voters favoring the election of a certain candidate for mayor. Assume that a sample proportion does not change. Which of the following best describes the anticipated effect on the width of the confidence interval if the researcher were to survey a random sample of 200, rather than 50, registered voters?

- (A) The width of the new interval would be about one-fourth the width of the original interval.
- (B) The width of the new interval would be about one-half the width of the original interval.
- (C) The width of the new interval would be about the same width as the original interval.
- (D) The width of the new interval would be about twice the width of the original interval.
- (E) The width of the new interval would be about four times the width of the original interval.

32. An independent research firm conducted a study of 100 randomly selected children who were participating in a program advertised to improve mathematics skills. The results showed no statistically significant improvement in mathematics skills, using  $\alpha = 0.05$ . The program sponsors complained that the study had insufficient statistical power. Assuming that the program is effective, which of the following would be an appropriate method for increasing power in this context.

- (A) Use a two-sided test instead of a one-sided test.
- (B) Use a one-sided test instead of a two-sided test.
- (C) Use  $\alpha = 0.01$  instead of  $\alpha = 0.05$ .
- (D) Decrease the sample size to 50 children.
- (E) Decrease the sample size to 200 children.

34. A planning board in Elm County is interested in estimating the proportion of its residents that are in favor of offering incentives to high-tech industries to build plants in that county. A random sample of Elm County residents was selected. All of the selected residents were asked, "Are you in favor of offering incentives to high-tech industries to build plants in your county?" A 95 percent confidence interval for the proportion of residents in favor of offering incentives was calculated to be  $0.54 \pm 0.05$ . Which of the following statements is correct?

- (A) At the 95% confidence level, the estimate of 0.54 is within 0.05 of the true proportion of county residents in favor of offering incentives to high-tech industries to build plants in the county.
- (B) At the 95% confidence level, the majority of area residents are in favor of offering incentives to high-tech industries to build plants in the county.
- (C) In repeated sampling, 95% of sample proportions will fall in the interval (0.49, 0.59)
- (D) In repeated sampling, the true proportion of county residents in favor of offering incentives to high-tech industries to build plants in the county will fall in the interval (0.49, 0.59).
- (E) In repeated sampling, 95% of the time the true proportion of county residents in favor of offering incentives to high-tech industries to build plants in the county will be equal to 0.54.

39. A polling organization asks a random sample of 1,000 registered voters which of two candidates they plan to vote for in an upcoming election. Candidate A is preferred by 400 respondents, Candidate B is preferred by 500 respondents, and 100 respondents are undecided. George uses a large sample confidence interval for two proportions to estimate the difference in the population proportions favoring the two candidates. This procedure is not appropriate because

- (A) the two sample proportions were not computed from independent samples
- (B) the sample size was too small
- (C) the third category, undecided, makes the procedure invalid
- (D) the sample proportions are different; therefore the variances are probably different as well
- (E) George should have taken the difference  $\frac{500 - 400}{1,000}$  and then used a large sample confidence interval for a single proportion instead

## 2002 Exam

2. A manufacturer of balloons claims that  $p$ , the proportion of its balloons that burst when inflated to a diameter of up to 12 inches, is no more than 0.05. Some customers have complained that the balloons are bursting more frequently. If the customers want to conduct an experiment to test the manufacturer's claim, which of the following hypotheses would be appropriate?

- (A)  $H_0: p \neq 0.05, H_a: p = 0.05$
- (B)  $H_0: p = 0.05, H_a: p > 0.05$
- (C)  $H_0: p = 0.05, H_a: p \neq 0.05$
- (D)  $H_0: p = 0.05, H_a: p < 0.05$
- (E)  $H_0: p < 0.05, H_a: p = 0.05$

24. A consulting statistician reported the results from a learning experiment to a psychologist. The report stated that on one particular phase of the experiment a statistical test result yielded a  $p$ -value of 0.24. Based on this  $p$ -value, which of the following conclusions should the psychologist make?

- (A) The test was statistically significant because a  $p$ -value of 0.24 is greater than a significance level of 0.05.
- (B) The test was statistically significant because  $p = 1 - 0.24 = 0.76$  and this is greater than a significance level of 0.05.
- (C) The test was not statistically significant because  $2 \text{ times } 0.24 = 0.48$  and that is less than 0.5.
- (D) The test was not statistically significant because, if the null hypothesis is true, one could expect to get a test statistic at least as extreme as that observed 24% of the time.
- (E) The test was not statistically significant because, if the null hypothesis is true, one could expect to get a test statistic at least as extreme as that observed 76% of the time.

38. Suppose that public opinion in a large city is 65 percent in favor of increasing taxes to support the public school system and 35 percent against such an increase. If a random sample of 500 people from this city are interviewed, what is the approximate probability that more than 200 of these people will be against increasing taxes?

(A)  $\binom{500}{200} (0.65)^{200} (0.35)^{300}$

(B)  $\binom{500}{200} (0.35)^{200} (0.65)^{300}$

(C)  $P\left(z > \frac{0.40 - 0.65}{\sqrt{\frac{(0.65)(0.35)}{500}}}\right)$

(D)  $P\left(z > \frac{0.40 - 0.35}{\sqrt{\frac{(0.4)(0.6)}{500}}}\right)$

(E)  $P\left(z > \frac{0.40 - 0.35}{\sqrt{\frac{(0.35)(0.65)}{500}}}\right)$

39. As lab partners, Sally and Betty collected data for a significance test. Both calculated the same z-test statistic, but Sally found the results were significant at the  $\alpha = 0.05$  level while Betty found that the results were not. When checking their results, the women found that the only difference in their work was that Sally had used a two-sided test, while Betty used a one-sided test. Which of the following could have been their test statistic?

(A) -1.980

(B) -1.690

(C) 1.340

(D) 1.690

(E) 1.780

## 1997 Exam

1. *USA Today* reported that speed skater Bonnie Blair had “won the USA’s heart,” according to a *USA Today*/CNN/Gallup poll conducted on the final Thursday of the 1994 Winter Olympics. When asked who was the hero of the Olympics, 65 percent of the respondents chose Blair, who won five gold medals. The poll of 615 adults, done by telephone, had a margin of error of 4 percent. Which of the following statements best describes what is meant by the 4 percent margin of error?
  - (A) About 4 percent of adults were expected to change their minds between the time of the poll and its publication in *USA Today*.
  - (B) About 4 percent of adults did not have telephones.
  - (C) About 4 percent of the 615 adults polled refused to answer.
  - (D) Not all of the 615 adults knew anything about the Olympics.
  - (E) The difference between the sample percentage and the population percentage is likely to be less than 4 percent.
  
35. A survey was conducted to determine what percentage of college seniors would have chosen to attend a different college if they had known then what they know now. In a random sample of 100 seniors, 34 percent indicated that they would have attended a different college. A 90 percent confidence interval for the percentage of all seniors who would have attended a different college is
  - (A) 24.7% to 43.3%
  - (B) 25.8% to 42.2%
  - (C) 26.2% to 41.8%
  - (D) 30.6% to 37.4%
  - (E) 31.2% to 36.8%

## 2013 Question 5

5. Psychologists interested in the relationship between meditation and health conducted a study with a random sample of 28 men who live in a large retirement community. Of the men in the sample, 11 reported that they participate in daily meditation and 17 reported that they do not participate in daily meditation.

The researchers wanted to perform a hypothesis test of

$$H_0 : p_m - p_c = 0$$

$$H_a : p_m - p_c < 0,$$

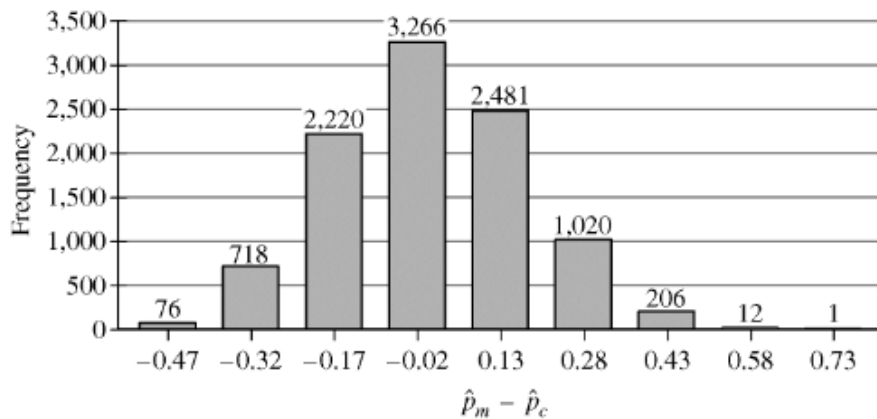
where  $p_m$  is the proportion of men with high blood pressure among all the men in the retirement community who participate in daily meditation and  $p_c$  is the proportion of men with high blood pressure among all the men in the retirement community who do not participate in daily meditation.

- (a) If the study were to provide significant evidence against  $H_0$  in favor of  $H_a$ , would it be reasonable for the psychologists to conclude that daily meditation causes a reduction in blood pressure for men in the retirement community? Explain why or why not.

The psychologists found that of the 11 men in the study who participate in daily meditation, 0 had high blood pressure. Of the 17 men who do not participate in daily meditation, 8 had high blood pressure.

- (b) Let  $\hat{p}_m$  represent the proportion of men with high blood pressure among those in a random sample of 11 who meditate daily, and let  $\hat{p}_c$  represent the proportion of men with high blood pressure among those in a random sample of 17 who do not meditate daily. Why is it not reasonable to use a normal approximation for the sampling distribution of  $\hat{p}_m - \hat{p}_c$ ?

Although a normal approximation cannot be used, it is possible to simulate the distribution of  $\hat{p}_m - \hat{p}_c$ . Under the assumption that the null hypothesis is true, 10,000 values of  $\hat{p}_m - \hat{p}_c$  were simulated. The histogram below shows the results of the simulation.



- (c) Based on the results of the simulation, what can be concluded about the relationship between blood pressure and meditation among men in the retirement community?



## 2012 Question 4

4. A survey organization conducted telephone interviews in December 2008 in which 1,009 randomly selected adults in the United States responded to the following question.

At the present time, do you think television commercials are an effective way to promote a new product?

Of the 1,009 adults surveyed, 676 responded “yes.” In December 2007, 622 of 1,020 randomly selected adults in the United States had responded “yes” to the same question. Do the data provide convincing evidence that the proportion of adults in the United States who would respond “yes” to the question changed from December 2007 to December 2008 ?

## 2012 Question 5

5. A recent report stated that less than 35 percent of the adult residents in a certain city will be able to pass a physical fitness test. Consequently, the city's Recreation Department is trying to convince the City Council to fund more physical fitness programs. The council is facing budget constraints and is skeptical of the report. The council will fund more physical fitness programs only if the Recreation Department can provide convincing evidence that the report is true.

The Recreation Department plans to collect data from a sample of 185 adult residents in the city. A test of significance will be conducted at a significance level of  $\alpha = 0.05$  for the following hypotheses.

$$H_0 : p = 0.35$$

$$H_a : p < 0.35,$$

where  $p$  is the proportion of adult residents in the city who are able to pass the physical fitness test.

- Describe what a Type II error would be in the context of the study, and also describe a consequence of making this type of error.
- The Recreation Department recruits 185 adult residents who volunteer to take the physical fitness test. The test is passed by 77 of the 185 volunteers, resulting in a  $p$ -value of 0.97 for the hypotheses stated above. If it was reasonable to conduct a test of significance for the hypotheses stated above using the data collected from the 185 volunteers, what would the  $p$ -value of 0.97 lead you to conclude?
- Describe the primary flaw in the study described in part (b), and explain why it is a concern.

### 2010 Question 3

3. A humane society wanted to estimate with 95 percent confidence the proportion of households in its county that own at least one dog.

(a) Interpret the 95 percent confidence level in this context.

The humane society selected a random sample of households in its county and used the sample to estimate the proportion of all households that own at least one dog. The conditions for calculating a 95 percent confidence interval for the proportion of households in this county that own at least one dog were checked and verified, and the resulting confidence interval was  $0.417 \pm 0.119$ .

- (b) A national pet products association claimed that 39 percent of all American households owned at least one dog. Does the humane society's interval estimate provide evidence that the proportion of dog owners in its county is different from the claimed national proportion? Explain.
- (c) How many households were selected in the humane society's sample? Show how you obtained your answer.