



# AP\* Statistics Review

## Inference

### Teacher Packet



# Review of Inference

Page 1 of 21

## *Deciding whether to use a test or a confidence interval*

Generally, use a significance test to test a claim. Use a confidence interval if you wish to estimate a population parameter ( $\mu$  or  $p$ ) based on statistics from your sample ( $\bar{x}$  or  $\hat{p}$ ). Once a confidence interval is constructed, you may use it to test claims (where you fail to reject a claim that falls within the confidence interval, and you reject a claim that falls outside of a confidence interval). The  $\alpha$ -level of a **two**-sided hypothesis test is related to the confidence level of an interval by  $C = 1 - \alpha$ .

## *Identifying the type of test or interval to use*

- Is there a single numerical variable being measured for each subject? Then we will perform a test for means.
- Is there a categorical variable being measured, and we are only concerned with how often a single response (a “success”) occurs? Then we will perform a test for a proportion. *For example, if we only care about the proportion of brown-eyed people, then we can do a  $z$ -test for the proportion of brown-eyed people.*
- Is there a categorical variable being measured, and we are concerned with how many responses fall into **each** category? Then we will perform a  $\chi^2$ -test. *For example, if we want to compare the occurrence of brown, blue, green, and grey eyes in two different groups, we will do a  $\chi^2$ -test because we are looking at multiple categories for the categorical variable “eye color.”*
- Are we looking at the relationship between two numerical variables? Then we will perform a  $t$ -test for the slope of a regression line.

## *How many samples?*

- Be careful to identify the source of each mean or proportion mentioned in a problem. If a mean or proportion does not clearly come from a sample (with an identifiable sample size  $n$ ), then it is probably a claim or a population proportion which should be used in the null hypothesis. A two-sample test should have two clearly identified sample sizes, and each sample should result in an  $\bar{x}$  or a  $\hat{p}$ .
- Some problems have two lists of numerical data that are linked in some way. For example, they could be pre-test and post-test scores for a list of students or temperatures in the sun and temperatures in the shade for a list of days. In these cases, the improvement (post-test score minus pre-test score) or the difference (temperature in sun minus temperature in shade) is the important variable. These are called matched-pair  $t$  tests. Begin by subtracting the two lists of data to obtain one list of improvements or differences. Then do a one sample  $t$ -test for a mean. (You will ignore the original two lists after you subtract.) Your null hypothesis will often be that the mean improvement was zero. For example in the pre- and

- post-test problem, you might use  $H_0 : \mu_{\text{improvement}} = 0$  and  $H_a : \mu_{\text{improvement}} > 0$ , where improvement is defined as post-test score minus pre-test score.
- In some  $\chi^2$ -tests, there is a clear claim. *For example, a company claims that 50% of the prizes in the popcorn boxes are stickers, 20% are rings, and 30% are temporary tattoos.* In this case, we are comparing the data from one sample to a claim, using a  $\chi^2$ -test for goodness of fit. The data table for this test consists of a single row of data.
  - In other  $\chi^2$ -tests, we are comparing two groups to see if they have the same percentages in each category. *For example, we could compare eye colors of a group of men and a group of women.* In this case, do a  $\chi^2$ -test for homogeneity. The data table would have two rows (one for male and one for female) and multiple columns (for brown eyes, green eyes, etc.). Rows and columns may be switched.
  - The data for some  $\chi^2$ -tests consists of two categorical questions asked to a single sample of people. *For example, we could ask a group of teachers whether they exercise frequently, often, or never and whether or not they missed any days of school last year due to illness.* We would like to see if their answers to the two questions are independent of each other. If they are independent, then the proportion who missed school due to illness should be the same for all three exercise categories. In this case, do a  $\chi^2$ -test for independence. The data table would have two rows (one for people who missed school due to illness and one for those who did not) and multiple columns (for the different exercise categories). Rows and columns could be switched. **The mechanics of the  $\chi^2$ -test for independence and the  $\chi^2$ -test for homogeneity are exactly the same.**

### What to put in your significance test

- A null and an alternative hypothesis (define the parameter of interest in words)  
*Note: Always hypothesize about the unknown population parameters ( $\mu$  and  $p$ ), not the sample statistics ( $\bar{x}$  and  $\hat{p}$ ), which are known from the data.*
- Identify the test you are using and check the conditions necessary for doing that test.
- Formula for the test statistic ( $z$  or  $t$  or  $\chi^2$ )
- Value for the statistic (can be from calculator if you have written the formula) and a shaded picture of the distribution if you have time to draw it
- The P value (from calculator or table) **related to** the  $\alpha$ -level, plus  $df$  for  $t$ -tests or the expected values for  $\chi^2$  tests
- Two conclusions: either reject  $H_0$  or don't reject it based on the relationship of the P-value and the  $\alpha$ -level AND write a conclusion about the alternative hypothesis in the context of the problem

## Phrasing to put in your conclusions

*Example: A seed manufacturer claims that at least 97% of their seeds will germinate. You suspect that the germination rate is less, so you buy a random selection of these seeds to test this claim. You calculate a P-value based on the hypotheses  $H_0 : p = 0.97$  and  $H_a : p < 0.97$ , where  $p$  is the germination rate of all seeds sold by this company.*

- **When  $P < \alpha$ , reject  $H_0$ .**

*First, draw a mathematical conclusion about  $H_0$ : “Since the P-value of 0.017 is less than the  $\alpha$ -level of 0.05, reject  $H_0$ . A value as extreme as my sample’s germination rate should only occur 1.7% of the time by random chance if the company’s claim is true.” Then, write a conclusion about  $H_a$  in the context of the problem: “We can conclude that the germination rate of the seeds is significantly lower than the 97% claimed by the company.”*

- **When  $P > \alpha$ , do not reject  $H_0$ .**

*First, draw a mathematical conclusion about  $H_0$ : “Since the P-value of 0.209 is greater than the  $\alpha$ -level of 0.05, there is not sufficient evidence to reject  $H_0$ . A value as extreme as my sample’s germination rate would occur 20.9% of the time by random chance if the company’s claim is true.” Then, write a conclusion about  $H_a$  in the context of the problem: “We cannot conclude that the germination rate of the seeds is significantly lower than the 97% claimed by the company.”*

- Note that we never accept  $H_0$ .

## How to check the conditions necessary to do the tests for means and proportions

Always check to see if the sample is independent: randomly taken **from the population of interest** and that the population is at least 10 times the sample size.

### One Sample

What you are studying	What you know	Use this	Why?
Mean	$s$ (standard deviation of sample) <b>Graph the sample if <math>n &lt; 40</math> to see if it is approx. normal</b>	$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$	properties of $t$ distribution



# Review of Inference

Proportion	$np_o > 10$ $n(1 - p_o) > 10$ <i>Use the claimed proportion here.</i>	$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$	you can approximate the binomial distribution by the normal distribution
Mean (Note: It is unusual to know $\sigma$ of the population.)	$\sigma$ population is normal	$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$	because you can always use $z$ scores for normal distributions
Mean (Note: It is unusual to know $\sigma$ of the population.)	$\sigma$ population may not be normal, but $n > 30$	$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$	Central Limit Theorem says that distributions get more normal as $n$ increases

## Two Samples

What you are studying	What you know	Use this	Why?
Difference of means (two independent samples)	$s_1$ and $s_2$ (standard deviation of samples), graph each one to see if it is approx. normal if $n < 40$	$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	properties of $t$ distribution
Difference of two proportions	$n_1\hat{p} > 5$ $n_1(1 - \hat{p}) > 5$ $n_2\hat{p} > 5$ $n_2(1 - \hat{p}) > 5$ <i>Use the pooled <math>\hat{p}</math> here.</i>	$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ where $\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$	you can approximate the binomial distribution by the normal distribution.
Difference of two dependent means	<b>THIS IS MATCHED PAIRS!!!</b>	Find the difference between each pair and do a one sample $t$ test.	We are only interested in one piece of data for each subject; usually the improvement or difference (after-before).
Difference of means (two ind. samples) (Note: It is unusual to know $\sigma$ .)	$\sigma_1$ and $\sigma_2$ population is normal	$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	because you can always use $z$ scores for normal distributions



# Review of Inference

Difference of means (two ind. samples) <i>(Note: It is unusual to know <math>\sigma</math>.)</i>	$\sigma_1$ and $\sigma_2$ population may not be normal, but $n_1 > 30$ and $n_2 > 30$	$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	Central Limit Theorem says that distributions get normal as $n$ increases
---	---	--	---

### How to check the conditions for the chi-squared tests

- Data must be counts (not averages or proportions).
- Data in sample are independent (chosen randomly and  $n < 10\%$  of the population)
- Groups are large enough that all expected values  $\geq 5$ .

### How to check the conditions for the t- test for the slope of a regression line

- The scatterplot must look linear.
- There must be no pattern in the residual plot (errors are independent).
- The residual plot has a constant spread (errors have constant variability).
- Histogram of residuals is approximately normal.

### Inference using confidence intervals

- In general, you must put the following three things in a confidence interval problem:
  1. Identify the interval you will use and check the conditions necessary to use the interval.
  2. Calculate the interval.
  3. Interpret the interval in the context of the problem.
- The conditions we must check are the same as for the associated significance tests (as shown in the table above), with one exception. When performing significance tests for one or two proportions, you check to see that  $n$  is large enough by examining  $np_o$  and  $n(1 - p_o)$  where  $p_o$  is the claimed proportion which appears in the null hypothesis. Since there is no claim in a confidence interval problem, use the sample proportion  $\hat{p}$  in these checks. For one-sample intervals, we require that  $n\hat{p}$  and  $n(1 - \hat{p})$  be over 10. In two-sample intervals, we require that  $n_1\hat{p}_1$ ,  $n_1(1 - \hat{p}_1)$ ,  $n_2\hat{p}_2$ , and  $n_2(1 - \hat{p}_2)$  be over 5.
- Also, in proportion confidence intervals, we must use the sample proportion  $\hat{p}$  to calculate the standard error.

- Formulas:

One mean ( $\sigma$ unknown)	$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$
Difference in two means ( $\sigma$ unknown)	$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
One mean ( $\sigma$ known; unusual case)	$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$
Difference in two means ( $\sigma$ known; unusual case)	$(\bar{x}_1 - \bar{x}_2) \pm z^* \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
One proportion	$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
Difference in two proportions	$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

### Good confidence interval conclusions

Make sure to state these in the context of the question.

- I am C% confident that my interval captures the population value  $\mu$  or p.
- C out of 100 intervals constructed using this method would capture the population value  $\mu$  or p.

### Bad confidence interval conclusions

Avoid making these statements:

- C% of the  $\bar{x}$  values or  $\hat{p}$  values would fall in my interval.
- C% of the data is in my interval.
- There is a C% chance that the population value  $\mu$  or p is in my interval.



## Review of Inference

Page 7 of 21

### Multiple Choice Questions on Inference

1. A study found that 63 of 211 randomly selected men and 130 out of 651 randomly selected women prefer cats to dogs. You want to test the hypothesis that women like cats more. Choose the correct hypotheses and pooled  $\hat{p}$ .

- (A)  $H_0: p_M = p_F; \quad H_A: p_M < p_F; \quad \hat{p} = .224$
- (B)  $H_0: p_M = p_F; \quad H_A: p_M > p_F; \quad \hat{p} = .249$
- (C)  $H_0: p_M = p_F; \quad H_A: p_M < p_F; \quad \hat{p} = .249$
- (D)  $H_0: p_M = p_F; \quad H_A: p_M > p_F; \quad \hat{p} = .224$
- (E)  $H_0: p_M < p_F; \quad H_A: p_M > p_F; \quad \hat{p} = .224$

2. An independent testing lab obtained random samples of new halogen bulbs and standard incandescent bulbs made by the same company to establish the company's claim that, on average, the halogen bulb lasts longer than the incandescent one. Which test would you use?

- (A) a matched pair  $t$  test
- (B) a  $t$ -test for the difference in two means
- (C) a  $z$ -test for the difference in two proportions
- (D) a  $t$ -test for the slope of the regression line
- (E) a  $\chi^2$ -test for homogeneity

3. A certain population follows a normal distribution with mean  $\mu$  and standard deviation  $\sigma$ . You construct a 95% confidence interval for  $\mu$  and find it to be  $1.1 \pm 0.9$ . Which of the following is true?

- (A) In a test of the hypotheses  $H_0: \mu=1.2, H_A: \mu \neq 1.2, H_0$  would be rejected at the .05 level.
- (B) In a test of the hypotheses  $H_0: \mu=1.9, H_A: \mu \neq 1.9, H_0$  would be rejected at the .05 level.
- (C) In a test of the hypotheses  $H_0: \mu=1.9, H_A: \mu \neq 1.9, H_0$  would be rejected at the .025 level.
- (D) In a test of the hypotheses  $H_0: \mu=0, H_A: \mu \neq 0, H_0$  would be rejected at the .05 level.
- (E) A conclusion about hypotheses cannot be made from a confidence interval.





## Review of Inference

Page 8 of 21

4. Which of the following is a condition for choosing a  $t$ -interval rather than a  $z$ -interval when constructing a confidence interval for the mean of a population?

- (A) The standard deviation of the population is unknown.
- (B) There is an outlier in the sample data.
- (C) The sample may not have been a simple random sample.
- (D) The population is not normally distributed.
- (E) The data are linked so a matched-pairs test is necessary.

5. You want to see whether or not high school changes children's educational plans. You take a random sample of 6<sup>th</sup> graders and of 12<sup>th</sup> graders and ask them whether they plan to get a job right after high school, go to college, or get an advanced degree. Which test do you perform?

- (A) a  $\chi^2$  test for homogeneity
- (B) a two-sample  $z$ -test for proportions
- (C) a matched pair  $t$ -test
- (D) a  $\chi^2$  test for goodness of fit
- (E) a  $t$ -test for the slope of the regression line

6. The Centers for Disease Control report a survey of randomly chosen Americans age 45 and older, which found that 51 of 100 men and 80 of 782 women suffered from some form of arthritis. You want to estimate the difference in the proportions of men and women over 45 who have arthritis with a 95% confidence interval. What standard error will you use?

- (A) 0.0192
- (B) 0.0378
- (C) 0.0511
- (D) 0.1485
- (E) 1.96

7. A two-sided hypothesis test for a population mean is significant at the 1% level of significance. Which of the following is necessarily true?

- (A) The 99% confidence interval of the mean contains 0.
- (B) The 99% confidence interval of the mean does not contain 0.
- (C) The 99% confidence interval of the mean contains the hypothesized mean.
- (D) The 99% confidence interval of the mean does not contain the hypothesized mean.
- (E) The 99% confidence interval is not useful here.



## Review of Inference

Page 9 of 21

8. Which of the following is **not** a characteristic of the  $\chi^2$  distribution?
- (A) Its shape is based on the sample size.
  - (B) It is skewed to the right.
  - (C) It approaches a normal distribution as the degrees of freedom increase.
  - (D) It can never take on negative values.
  - (E) It is always used for one-sided significance tests.
9. Which of the following would be the most appropriate for measuring the association between gender and favorite color based on a random sample of subjects?
- (A) a two-sample  $t$ -test
  - (B) a correlation coefficient
  - (C) a  $\chi^2$ -test for independence
  - (D) a one-sample  $z$ -test for a proportion
  - (E) a  $t$ -test for the slope of the regression line
10. Sixty senior account executives were classified into three groups, labeled A, B, and C. There were 26 in group A, 19 in group B and 15 in group C. At the .05 significance level, we would like to test if it is reasonable to conclude that the proportion of the population that falls into each group is the same. Which of the following is a correct conclusion?
- (A) Reject  $H_0$ . The proportion in the three groups is not significantly different.
  - (B) Reject  $H_0$ . The proportion in the three groups is not the same.
  - (C) Do not reject  $H_0$ . The proportion in the three groups is not significantly different.
  - (D) Do not reject  $H_0$ . The proportion in the three groups is not the same.
  - (E) We cannot perform a significance test because there are three groups.

Use the following information to answer questions 11 and 12.

A one sample  $t$  test yields a  $t$  statistic of 2.089. The sample size was 16.

11. The alternative hypothesis was in the form  $H_a : \mu > 37.5$ . Is there significant evidence at the  $\alpha = .05$  level to reject the null hypothesis?
- (A) No, because the P-value is between 0.05 and 0.10.
  - (B) No, because the P-value is between 0.025 and 0.05.
  - (C) No, because the sample mean was significantly above 37.5.
  - (D) Yes, because the P-value is between 0.05 and 0.10.
  - (E) Yes, because the P-value is between 0.025 and 0.05.



## Review of Inference

Page 10 of 21

12. If the alternative hypothesis was  $H_a : \mu \neq 37.5$  instead, would you reject the null hypothesis at the  $\alpha = 0.05$  level?

- (A) No, because the P-value is between 0.05 and 0.10.
- (B) No, because the P-value is between 0.025 and 0.05.
- (C) No, because the sample mean was significantly above 37.5.
- (D) Yes, because the P-value is between 0.05 and 0.10.
- (E) Yes, because the P-value is between 0.025 and 0.05.



## Review of Inference

Page 11 of 21

### Free Response Questions on Inference

1. A fitness trainer wants to know if her weight-lifting program can quickly improve upper body strength in older people. To find out, she has a group of randomly selected people over 55 years old do push-ups for 90 seconds and counts the number each can do. After these people participate in her weightlifting program for 3 weeks, she tests them again in the same way. Here are the results:

Person	1	2	3	4	5	6	7	8	9	10	11	12
Before	15	12	21	22	17	19	10	25	12	17	8	19
After	17	15	22	22	21	24	11	28	14	16	12	21

Does the program help?



## Review of Inference

Page 12 of 21

2. There are two main dog parks in Dallas, one near White Rock Lake and one near downtown. The downtown dog park is smaller and is located underneath several highway overpasses. There are many apartments, townhomes, and lofts nearby. The White Rock Lake dog park is larger and provides a place for dogs to swim in the lake. The neighborhoods nearby are a mix of single family homes with some apartments. Jessica believes that since the downtown dog park is near many apartments, many of the dogs there will be smaller breeds, while the White Rock Lake park will attract larger, more active breeds. In order to test this assertion, she chooses random times during a month to visit each park. She categorizes the dogs there by size.

	Toy ( $< 10$ lbs)	Small (11 – 20 lbs)	Medium (21-50 lbs)	Large (51-100 lbs)	Giant (over 100 lbs)
Downtown	39	72	101	89	12
White Rock Lake	77	158	188	275	51

Does the breed distribution for the downtown dog park differ significantly from the White Rock Lake dog park at the  $\alpha = 0.05$  level?



## Review of Inference

Page 13 of 21

3. Are female or male students more likely to attend college outside their home state? In order to find out, random samples of male and female college-bound high school seniors were taken in the Dallas/Fort Worth metropolitan area. In September following their high school graduations, the students in the samples were contacted to see if they were attending college in Texas or outside of it. (Students who were not attending college were eliminated from the study.) The results are summarized in the following table.

	Attending college outside of Texas	Attending college in Texas
female	121	329
male	94	212

a) Write and interpret a 95% confidence interval for the difference in proportion of male and female students attending college outside of Texas.

b) Based only on your confidence interval, does the data from the random samples indicate that there is a difference in proportions of male and female students attending college outside of Texas? Justify your answer.

### Key to Inference Multiple Choice

1. A We are trying to prove that  $p_F > p_M$ , and the pooled  $\hat{p} = \frac{63+130}{211+651} = 0.224$ .
2. B We are comparing two means, that of halogen bulbs and that of incandescent bulbs.
3. D The confidence interval is (0.2, 2.0). Zero is not in this interval, so we reject that claim. A 95% confidence level corresponds to an  $\alpha$ -level of 0.05 for a two-sided test.
4. A Use a  $z$  test if  $\sigma$  is known; use a  $t$  test if we must approximate  $\sigma$  using  $s$  from the sample.
5. A We are comparing two groups on their answer to a categorical question.
6. C Standard error =  $\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} = 0.0511$  (Answer choice B incorrectly uses the pooled  $\hat{p}$ , which would be correct in a significance test, but not a confidence interval.)
7. D When a test is significant,  $H_o$  was rejected. The claim (or hypothesized mean) was NOT in the interval.
8. C The  $t$  distribution approaches the normal distribution as  $n$  increases, but the  $\chi^2$  distribution is always skewed.
9. C We are looking at one group's answers to two questions with categorical answers.
10. C The expected value for each group is 20. The value of the  $\chi^2$  statistics is 3.1, and the P-value of the test is 0.212. This is higher than any  $\alpha$  level, so we do not reject  $H_o$ , which says that the groups are the same. We can't say that the groups differ significantly.
11. E The  $df = 15$ , so the P value is 0.0271, which is less than 0.05, so we reject  $H_o$ .
12. A For a two-sided test, double the P value to 0.054, which is greater than 0.05, so we do not reject  $H_o$ .

## Rubric for Inference for Free Response

### 1. Solution

Part 1: Identify the correct test by name or formula, subtract to get the improvement, check conditions.

Since the two lists are “before” and “after” results for the same 12 people, this is a

matched pair  $t$  test. OR  $t = \frac{\bar{x}_{diff} - 0}{s_{diff} / \sqrt{n}}$

Subtract “after” – “before” to get each participant’s improvement:

Person	1	2	3	4	5	6	7	8	9	10	11	12
Before	15	12	21	22	17	19	10	25	12	17	8	19
After	17	15	22	22	21	24	11	28	14	16	12	21
Improvement	2	3	1	0	4	5	1	3	2	-1	4	2

Check conditions for one-sample  $t$  test:

The data are independent because the participants were randomly chosen and  $n$  is less than 10% of the population of adults over 55.

AND

The data is nearly normal because an examination of the dotplot of the differences shows a unimodal graph with no outliers:

```

-1 | X
0  | X
1  | XX
2  | XXX
3  | XX
4  | XX
5  | X
  
```

Part 2: Write hypotheses, identifying the parameter of interest.

$$H_o : \mu_{\text{improvement}} = 0$$

$$H_a : \mu_{\text{improvement}} > 0$$

where  $\mu_{\text{improvement}}$  = the true mean improvement in number of push-ups done

Part 3: Perform the test, using correct mechanics, including value of the  $t$  statistic, the degrees of freedom, and the P value.

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{2.167 - 0}{1.749 / \sqrt{12}} = 4.29$$

With  $df = 11$ ,  $P(\bar{x}_{\text{improvement}} > 2.167) = P(t > 4.29) = 0.00064$





## Review of Inference

Page 16 of 21

Part 4: Using the calculations, write a conclusion in the context of the problem.

Since the P value of 0.00064 is less than any reasonable  $\alpha$ , we have evidence to reject  $H_0$ . We can conclude that the improvement is significantly above zero; the participants did improve the number of push-ups they could do.

### Scoring

Each part is *essentially correct* (E), *partially correct* (P), or *incorrect* (I).

Part 1 is *essentially correct* if the student correctly identifies the test, subtracts to find the improvement, checks for independence, and graphs the improvements to show that they are unimodal and symmetric.

Part 1 is *partially correct* if the student correctly does two or three of those.

Part 1 is *incorrect* if the student does only zero or one of those.

Part 2 is *essentially correct* if the student correctly gives both hypotheses and identifies the parameter.

Part 2 is *partially correct* if the student does only one of those.

Part 3 is *essentially correct* if the student correctly gives the value of the t statistic, the degrees of freedom, and the P-value.

Part 3 is *partially correct* if the student gives only one or two of these.

Part 4 is *essentially correct* if the student correctly links the P-value to the alpha-level in order to reject  $H_0$  AND gives the conclusion (that the program does help) in context.

Part 4 is *partially correct* if the student gives only one of these conclusions.

**To assign a score to this question let an E = 1point, a P = 0.5 points, and an I = 0 points. Sum the total points for the student's score. If a student has a half point, look at the question holistically to determine if the score should be rounded up or truncated.**

**4 Complete Response**

**3 Substantial Response**

**2 Developing Response**

**1 Minimal Response**



# Review of Inference

## 2. Solution

Part 1: Identify the correct test by name or formula, check conditions necessary to do test.

Since we are comparing two different random samples on multiple categories, we will do a  $\chi^2$  test for homogeneity OR

Check the conditions for a  $\chi^2$  test:

The data are counts.

The two samples are independent.

Each expected value is at least 5.

Part 2: Write hypotheses.

$H_0$ : The White Rock Lake dog park and the downtown dog park have the same distribution of dogs by size (are homogeneous).

$H_A$ : The White Rock Lake dog park and the downtown dog park do not have the same distribution of dogs by size.

Part 3: Perform the test, using correct mechanics, including value of the  $\chi^2$  statistic, the degrees of freedom, the expected values, and the P value.

Expected values are in parentheses:

	Toy (< 10 lbs)	Small (11 – 20 lbs)	Medium (21-50 lbs)	Large (51-100 lbs)	Giant (over 100 lbs)
Downtown	39 (34)	72 (68)	101 (85)	89 (107)	12 (19)
White Rock Lake	77 (82)	158 (162)	188 (204)	275 (257)	51 (44)

$$df = (rows - 1)(columns - 1) = (2 - 1)(5 - 1) = 4$$

$$\chi^2 = \sum \frac{(O - E)^2}{E} = \frac{(39 - 34)^2}{34} + \frac{(72 - 68)^2}{68} + \frac{(101 - 85)^2}{85} + \dots + \frac{(51 - 44)^2}{44} = 13.21$$

$$P(\chi^2 > 13.21) = 0.0103$$

Part 4: Using the calculations, write a conclusion in the context of the problem.

Since the P-value of 0.0103 is less than the  $\alpha$ -level of 0.05, we can reject  $H_0$ . Based on these samples, the White Rock Lake dog park and the downtown dog park have a significantly different distribution of dogs by size.

## Scoring

Parts 1, 3, and 4 can be *essentially correct* (E), *partially correct* (P), or *incorrect* (I). Part 2 can be *essentially correct* (E) or *incorrect* (I).



## Review of Inference

Page 18 of 21

Part 1 is *essentially correct* if the student correctly identifies the test, mentions independence, and states that the expected values are over 5.

Part 1 is *partially correct* if the student correctly does one or two of those.

Part 2 is *essentially correct* if the student correctly gives both hypotheses in words.

Part 3 is *essentially correct* if the student correctly gives the value of the  $\chi^2$  statistic, the degrees of freedom, the expected values, and the P-value.

Part 3 is *partially correct* if the student gives only two or three of these.

Part 3 is *incorrect* if the student gives only zero or one of these.

Part 4 is *essentially correct* if the student correctly links the P-value to the alpha-level in order to reject  $H_0$  AND gives the conclusion (that the sizes of the dogs at each park differ) in context.

Part 4 is *partially correct* if the student gives only one of these conclusions.

**To assign a score to this question let an E = 1 point, a P = 0.5 points, and an I = 0 points. Sum the total points for the student's score. If a student has a half point, look at the question holistically to determine if the score should be rounded up or truncated.**

**4 Complete Response**

**3 Substantial Response**

**2 Developing Response**

**1 Minimal Response**



## Review of Inference

Page 19 of 21

### 3. Solution

Part 1: Identify the correct interval by name or formula, check conditions.

This is a two-sample  $z$  interval for the difference in two proportions OR

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

Check conditions for two-proportion  $z$  interval:

The data are independent because the participants were randomly chosen and each  $n$  is less than 10% of the population of college-bound high school male or female seniors in the Dallas/Fort Worth area.

AND

The sample sizes are large enough because

$$n\hat{p} = (450)(0.269) = 121 > 10 \text{ or } 5$$

$$n(1 - \hat{p}) = (450)(0.731) = 329 > 10 \text{ or } 5$$

$$n\hat{p} = (306)(0.307) = 94 > 10 \text{ or } 5$$

$$n(1 - \hat{p}) = (306)(0.693) = 212 > 10 \text{ or } 5$$

Part 2: Calculate the interval. (This may be done in either order, male-female or female-male.)

$$\hat{p}_M = \frac{94}{306} = 0.307 \quad \hat{p}_F = \frac{121}{450} = 0.269$$

$$(\hat{p}_M - \hat{p}_F) \pm z^* \sqrt{\frac{\hat{p}_M(1 - \hat{p}_M)}{n_M} + \frac{\hat{p}_F(1 - \hat{p}_F)}{n_F}}$$

$$= (0.307 - 0.269) \pm 1.96 \sqrt{\frac{0.307(0.693)}{306} + \frac{0.269(0.731)}{450}}$$

$$= 0.038 \pm 0.066 = (-0.028, 0.104) = (-2.8\%, 10.4\%)$$

Part 3: Interpret the interval.

Based on these samples, I am 95% confident that the interval (-2.8%, 10.4%) captures the true difference between the population proportion of male DFW students who attend college outside Texas and the population proportion of female DFW students who attend college outside Texas.

OR

Based on these samples, I am 95% confident that the true difference between the population proportion of male DFW students who attend college outside Texas and the population proportion of female DFW students who attend college outside Texas is between -2.8% and 10.4%.



## Review of Inference

Page 20 of 21

Part 4:

Since 0 is in the 95% confidence interval, zero is a plausible value for the difference in proportions,  $p_M - p_F$ . The evidence shows no significant difference between the proportion of male students attending college outside of Texas and the proportion of female students attending college outside of Texas.

### **Scoring**

Each of the four parts can be essentially correct (E) or incorrect (I).

Part 1 is essentially correct if the interval is identified and the student comments on both independence and large sample size. The minimum amount necessary is an indication that the number of successes and failures for both samples is over 10 (or 5) AND a mention of independence (or independence with a check mark). The student does not have to repeat the fact that the samples are random.

Part 2 can be essentially correct even if there is an identifiable minor arithmetic error.

Part 4 is essentially correct if the student states that zero is not in the interval and links this to either a 95% confidence level or a 5% significance level. Part 4 is incorrect if the student says no without justification or if the student says no because zero is not in the interval.

- 4    **Complete Response (4E)**  
All four parts essentially correct
- 3    **Substantial Response (3E)**  
Three parts essentially correct
- 2    **Developing Response (2E)**  
Two parts essentially correct
- 1    **Minimal Response (1E)**  
One part essentially correct



# Review of Inference

Page 21 of 21

## AP Statistics Exam Connections

The list below identifies free response questions that have been previously asked on the topic of Inference on the AP Statistics Exam. These questions are available from the CollegeBoard and can be downloaded free of charge from AP Central.

<http://apcentral.collegeboard.com>.

Free Response Questions	
2001	Question 5
2002	Question 6
2004	Question 6