MATH

AP* Statistics

Keeping Pace!

Teacher Packet

STUDENT PREP SESSIONS

Nothing takes the place of the qualified and creative teacher in the classroom every day. However, formally scheduled prep sessions provide students opportunities to spend additional time on the all-important task of preparing for upcoming examinations, measuring their current skills and knowledge against those of other students and against the high standards set by the instructor and the materials. Students are treated as mature individuals, capable of handling a rigorous, fast-paced review in six hours of content instruction. Often, they realize how well-prepared they actually are and gain confidence in their abilities. Sometimes their eyes are opened to the work that must still be done to successfully manage the challenges of upcoming exams. Prep sessions can provide the instruction, materials, and encouragement students need to reach their goals. Specifically, students who attend will have the opportunity to

- Review topics taught in the classroom
- Gain tips and strategies for problem-solving
- Learn information from teachers other than their own
- Interact with other students who share an all-important goal
- Understand relationships and connections among many concepts

Teachers are encouraged to attend prep sessions with their students. They will have an opportunity to

- Share with other successful instructors and gain new strategies for test review
- Receive valuable materials created by master teachers
- Observe students other than their own for purposes of gauging classroom progress

TABLE OF CONTENTS

Published by:
Laying the Foundation, Inc.
8350 North Central Expressway, Suite 300
Dallas, TX 75206

Grateful acknowledgement is given authors, publishers, and agents for permission to reprint copyrighted material. Every effort has been made to determine copyright owners. In case of any omission, the publisher will be pleased to make suitable acknowledgments in future editions.

AP* Statistics Review

Normal Models and Sampling Distributions

Teacher Packet

Characteristics of a normal model

- The shape is unimodal, symmetric, and mound-shaped.
- The mean is equal to the median.
- A normal model is continuous, although it is often used to approximate a discrete distribution like a histogram.
- The shape, center, and spread of a normal model can be quickly given by writing $N(\mu, \sigma)$.

The 68%-95%-99.7% (Empirical) Rule

- About 68% of the area is in the interval $(\mu - \sigma, \mu + \sigma)$, or within 1 standard deviation of the mean
- About 95% of the area is in the interval $(\mu - 2\sigma, \mu + 2\sigma)$, or within 2 standard deviations of the mean
- About 99.7% of the area is in the interval $(\mu - 3\sigma, \mu + 3\sigma)$, or within 3 standard deviations of the mean

What's a z score?

- A z score tells how many standard deviations above or below the mean a data point is. A z score of +1 means the point is one standard deviation above the mean. A z score of -2 means the point is two standard deviations below the mean.
- Calculate a z score using this formula: $z = \frac{x - \mu}{\sigma}$, where x is the specific value of interest in the distribution.
- Z scores make it possible to compare quantities measured on different scales, such as SAT scores and ACT scores.

Finding probabilities using a normal model (forwards!)

- Strategy: $x \rightarrow z \rightarrow P$.
- Sketch a normal model, draw a vertical line at the x value, and shade the area of interest.
- Take the x value and find the z score using the formula.
- Then look up the z score on the normal table to find the probability below (to the left of) the line. (See the last page for calculator usage)
- If you want to find the area to the right of the line, subtract the P value off the table from 1.
- If you want to find the area between two lines, find the z scores for each x value, look up P values on the table for each z score, and subtract the P values.

Finding an x value given a normal probability (backwards!)

- Strategy is $P \rightarrow z \rightarrow x$.
- Sketch a normal model, draw a vertical line where you think the x value will be, and label the area/probability you have been given.
- Remember, the table lists the area to the left of the line. If you have been given an area to the right, subtract from 1 to get the area on the left.
- Find the P value inside the chart and move outwards to read the z score.
- Plug the *z*-score as well as the mean and standard deviation into the formula and solve for x.

Sampling Distributions and The Central Limit Theorem

Consider taking many (theoretically, all possible) samples of size n from a population. Take the average $\bar{x}$ of each sample. All of these sample means make up the sampling distribution, which can be graphed as a histogram.

- The mean of the sampling distribution is the same as the mean of the population: $\mu_{\bar{x}} = \mu$
- The standard deviation of the sampling distribution gets smaller according to this equation: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
- The Central Limit Theorem states that as the sample size n increases, the sampling distribution becomes more normal (regardless of the shape of the population). In practice, if $n \geq 30$, we assume the distribution is approximately normal.

When do you use the Central Limit Theorem?

Use the Central Limit Theorem when a question asks you to calculate a probability about an average or mean.

Example:
The amount of dirt loaded into a dump truck varies normally with a mean of 750 pounds and a standard deviation of 40 pounds. A) Find the probability that a randomly chosen truck would hold over 800 pounds of dirt. B) Find the probability that the average weight in a random sample of 5 trucks would be over 800 pounds.

Solution:

A) x = weight of dirt in a dump truck

x follows a $N(750,40)$ model

$z = \frac{800-750}{40} = 1.25$, $P(x > 800)) = P(z > 1.25) = 1 - .8944 = .1056$

Nearly 11% of the trucks (10.56%) would hold over 800 pounds.

B) $\bar{x}$ = average weight of dirt in a sample of 5 dump trucks

$\bar{x}$ follows a $N\left(750, \frac{40}{\sqrt{5}} = 17.89\right)$ model

$z = \frac{800-750}{17.89} = 2.80$

$P(\bar{x} > 800)) = P(z > 2.80) = 1 - .9974 = .0026$

Less than one percent (0.26%) of groups of 5 dump trucks would have an average weight over 800 pounds.

Multiple Choice Questions on Normal Models and Sampling Distributions

1. If heights of 3rd graders follow a normal distribution with a mean of 52 inches and a standard deviation of 2.5 inches, what is the z score of a 3rd grader who is 47 inches tall?

(A) -5
(B) -2
(C) 2
(D) 5
(E) 26.2

2. Suppose that a normal model describes the acidity (pH) of rainwater, and that water tested after last week's storm had a *z*-score of 1.8. This means that the acidity of the rain

(A) had a pH 1.8 higher than average rainfall.
(B) had a pH of 1.8.
(C) varied with standard deviation 1.8
(D) had a pH 1.8 standard deviations higher than that of average rainwater.
(E) had a pH 1.8 times that of average rainwater.

3. In a factory, the weight of the concrete poured into a mold by a machine follows a normal distribution with a mean of 1150 pounds and a standard deviation of 22 pounds. Approximately 95% of molds filled by this machine will hold weights in what interval?

(A) 1084 to 1216 pounds
(B) 1106 to 1150 pounds
(C) 1106 to 1194 pounds
(D) 1128 to 1172 pounds
(E) 1150 to 1194 pounds

4. Which of the following are true?

I. In a normal distribution, the mean is always equal to the median.
II. All unimodal and symmetric distributions are normal for some value of μ and σ.
III. In a normal distribution, nearly all of the data is within 3 standard deviations of the mean, no matter the mean and standard deviation.

(A) I only
(B) II only
(C) III only
(D) I and III only
(E) I, II, and III

5. The height of male Labrador Retrievers is normally distributed with a mean of 23.5 inches and a standard deviation of 0.8 inches. (The height of a dog is measured from his shoulder.) Labradors must fall under a height limit in order to participate in certain dog shows. If the maximum height is 24.5 inches for male labs, what percentage of male labs are <u>not</u> eligible?

(A) 0.1056
(B) 0.1250
(C) 0.8750
(D) 0.8944
(E) 0.9750

6. The heights of mature pecan trees are approximately normally distributed with a mean of 42 feet and a standard deviation of 7.5 feet. What proportion of pecan trees are between 43 and 46 feet tall?

(A) 0.1501
(B) 0.2969
(C) 0.4470
(D) 0.5530
(E) 0.7031

7. Heights of fourth graders are normally distributed with a mean of 52 inches and a standard deviation of 3.5 inches. Ten percent of fourth graders should have a height below what number?

(A) -1.28 inches
(B) 45.0 inches
(C) 47.5 inches
(D) 48.9 inches
(E) 56.5 inches

8. A large college class is graded on a total points system. The total points earned in a semester by the students in the class vary normally with a mean of 675 and a standard deviation of 50. Another large class in a different department is graded on a 0 to 100 scale. The final grades in that class follow a normal model with a mean of 82 and a standard deviation of 6. Jessica earns 729 points in the first class, while Ana scores 90 in the second class. Which student did better and why?

(A) Jessica did better because her score is 54 points above the mean while Ana's is only 8 points above the mean
(B) The students did equally well because both scored above the mean.
(C) Ana did better because her score is 1.33 standard deviations above the mean while Jessica's is only 1.08 standard deviations above the mean.
(D) Neither student did better; they cannot be compared because their classes have different scoring systems.

9. The distance Jonathan can throw a shot put is skewed to the right with a mean of 14.2 meters and a standard deviation of 3.5 meters. Over the course of a month, Jonathan makes 75 throws during practice. Assume these throws can be considered a random sample of Jonathan's shot put throws. What is the probability that Jonathan's average shot put distance for the month will be over 15.0 meters?

(A) 0.0239
(B) 0.4096
(C) 0.5224
(D) 0.5904
(E) 0.9761

10. Heights of fourth graders are normally distributed with a mean of 52 inches and a standard deviation of 3.5 inches. For a research project, you plan to measure a simple random sample of 30 fourth graders. For samples such as yours, 10% of the samples should have an average height below what number?

(A) 47.52 inches
(B) 51.18 inches
(C) 51.85 inches
(D) 52.82 inches
(E) 56.48 inches

Free Response Questions on Normal Models and Sampling Distributions

Free Response 1.
A machine is used to fill soda bottles in a factory. The bottles are labeled as containing 2.0 liters, but extra room at the top of the bottle allows for a maximum of 2.25 liters of soda before the bottle overflows. The standard deviation of the amount of soda put into the bottles by the machine is known to be 0.15 liter.

(a) Overfilling the bottles causes a mess on the assembly line, but consumers will complain if bottles contain less than 2 liters. If the machine is set to fill the bottles with an average of 2.08 liters, what proportion of bottles will be overfilled?

(b) If management requires that no more than 3% of bottles should be overfilled, the machine should be set to fill the bottles with what mean amount?

(c) Complaints from consumers about underfilled bottles leads the company to set the mean amount to 2.15 liters. In this situation, what standard deviation would allow for no more than 3% of the bottles to be overfilled?

Free Response 2.
The distribution of scores for persons over 16 years of age on a common IQ test is approximately normal with mean 100 and standard deviation 15.

(a) What is the probability that a randomly chosen adult has an IQ score on this test over 105?

(b) What are the mean and standard deviation of the average IQ score on this test for an SRS of 60 people?

(c) What is the probability that the average IQ score on this test of an SRS of 60 people is 105 or higher?

(d) Would your *method* of answering (a) or (c) be affected if the distribution of IQ scores on this test in the adult population were distinctly nonnormal (for example, if they were skewed)? Explain which parts could be answered the same way, which could not, and how you know.

Key to Normal Models and Sampling Distributions Multiple Choice

1. B — Distractors include switching the order of x and μ, and forgetting to divide by the standard deviation.
2. D — z score is number of standard deviations from mean
3. C — Distractors include adding/subtracting 1 or 3 standard deviations instead of 2.
4. D — Unimodal and symmetric doesn't necessarily mean normal.
5. A — Distractor is finding eligible labs *under* the height limit instead of those who are not eligible.
6. A — Distractors are finding proportion of trees under or over 43 or 46 feet instead of those between 43 and 46 feet.
7. C — z = -1.28; distractor is solving z score formula incorrectly
8. C — z score (or number of standard deviations above the mean) is relevant; number of points above the mean is irrelevant
9. A — $\bar{x}$ follows $N\left(14.2, \frac{3.5}{\sqrt{75}} = .404\right)$ model with z score of 1.98; distractors include not modifying the standard deviation, and finding the probability below 15 m instead of above 15 m
10. B — z = -1.28; $\bar{x}$ follows $N\left(52, \frac{3.5}{\sqrt{30}} = .639\right)$ model; distractors include not modifying standard deviation, incorrectly solving z score formula for x

Rubrics for Normal Models and Sampling Distributions Free Response

1. **Solution**

Part (a):

Let x = amount of soda put into the bottle

x follows $N(2.08, 0.15)$ model

$$z = \frac{2.25 - 2.08}{0.15} = 1.13$$

$P(x > 2.25) = P(z > 1.13) = 1 - 0.871 = 0.129$

12.9% of bottles would be overfilled.

Part (b):

x follows $N(\mu, 0.15)$.

A z score of 1.88 separates the normal model into 97% below the line and 3% above the line: $P(z > 1.88) = 0.0300$

We require that only 3% have over 2.25 liters of soda.

$$z = \frac{x - \mu}{\sigma} \qquad 1.88 = \frac{2.25 - \mu}{0.15} \qquad \mu = 2.25 - (1.88)(0.15) = 1.97 \text{ liters}$$

Part (c):

x follows $N(2.15, \sigma)$.

A z score of 1.88 separates the normal model into 97% below the line and 3% above the line: $P(z > 1.88) = 0.0300$

We require that only 3% have over 2.25 liters of soda.

$$z = \frac{x - \mu}{\sigma} \qquad 1.88 = \frac{2.25 - 2.15}{\sigma} \qquad \sigma = \frac{2.25 - 2.15}{1.88} = 0.053 \text{ liters}$$

Scoring

Part (a) can be essentially correct (E) or incorrect (I). Parts (b) and (c) can be essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is essentially correct if the proportion is correct (except for minor arithmetic errors) and supporting work (not including calculator input) is shown.

Part (a) is incorrect if the proportion is calculated incorrectly or the correct proportion is given with no supporting work.

Part (b) is essentially correct if the mean is correct (except for minor arithmetic errors) and supporting work (not including calculator input) is shown.

Part (b) is partially correct if the z score is incorrect but the method for calculating the mean was correct based on that z score and work is shown
OR
the z score is correct but the mean is incorrectly calculated.
Part (b) is incorrect if the z score is wrong and the mean is wrong
OR
The z score is wrong and the mean is correct but no supporting work is given.

Part (c) is essentially correct if the standard deviation is correct (except for minor arithmetic errors) and supporting work (not including calculator input) is shown.
OR
if the same incorrect z score is used from part (b) and the standard deviation is calculated correctly based on that z score with work shown (a student will not be penalized twice for having the wrong z score).
Part (c) is partially correct if the z score is correct but the standard deviation is incorrectly calculated.
Part (c) is incorrect if the standard deviation is correct but no supporting work is given.

4 Complete Response

All parts essentially correct.

3 Substantial Response

Two parts essentially correct and one part partially correct

2 Developing Response

Two parts essentially correct and no parts partially correct
OR
One part essentially correct and two parts partially correct

1 Minimal Response

One part essentially correct and either zero or one part partially correct
OR
No parts essentially correct and two parts partially correct

2. **Solution** **Part (a):**

Let x = score on this IQ test

x follows $N(100, 15)$ model

$$z = \frac{105-100}{15} = 0.33$$

$$P(x > 105) = P(z > 0.33) = 1 - .6293 = 0.3707$$

Part (b):

The average of the sample means is equal to the average of the population, so $\mu_{\bar{x}} = 100$.

The standard deviation of the groups of size 60 is $\sigma_{\bar{x}} = \frac{15}{\sqrt{60}} = 1.94$.

Part (c):

$\bar{x}$ follows $N(100, 1.94)$ model

$$z = \frac{105-100}{1.94} = 2.58$$

$$P(\bar{x} > 105) = P(z > 2.58) = 1 - .9951 = 0.0049$$

Part (d):

If the population was not normal, then we could not use z scores and a normal model to calculate the probability in part (a). However, since the sample size is large ($n = 60$), the Central Limit Theorem states that the sampling distribution will be approximately normal, so we could do part (c) the same way.

Note: To receive complete credit for part (a) or part (c), a student must show how the probability is computed. Since part (a) and part (c) involve different normal distributions, it is important to identify which normal distribution is used in which part. This can be done by showing the z score calculation, displaying the mean and standard deviation within the probability statements, or by listing the mean and standard deviation and displaying an appropriate graph.

Scoring

Parts (a) and (c) can be either essentially correct (E) or incorrect (I). Parts (b) and (d) can be essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is essentially correct if the probability is calculated correctly (except for minor arithmetic errors) and work is shown. Calculator input does not count as work.

Part (a) is incorrect if no work is shown.

Part (b) is correct if both the mean and standard deviation are computed correctly and work is shown for the standard deviation calculation.

Part (b) is partially correct if only one of the mean or standard deviation is computed correctly (with work shown for the standard deviation).

Part (c) is essentially correct if the probability is calculated correctly using the mean and standard deviation from part (b) with work shown.

Part (c) is incorrect if no work is shown.

Part (d) is essentially correct if the student states that part (a) could NOT be done but part (c) could be done AND justification is given. Justification can be either stating that the sample size is large enough or referring to the Central Limit Theorem.

Part (d) is partially correct if the student correctly answers for either part (a) or part (c) with justification

OR

the student answers both part (a) and part (c) correctly with weak justification.

Part (d) is incorrect if the answers are incorrect

OR

the answers are correct but no justification is given.

Each essentially correct response is worth 1 point; each partially correct response is worth half a point.

4 Complete Response

3 Substantial Response

2 Developing Response

1 Minimal Response

If a response is between two scores (for example, 2.5 points), use a holistic approach to determine whether to score up or down depending on the strength of the response and communication.

(Optional) Using a TI-84 to find normal probabilities

Note: cdf stands for cumulative distribution function. This adds up probabilities for a range of outcomes. pdf stands for probability distribution function. Students do not generally use Normalpdf.

Finding the area under a normal curve

- First, draw and label a graph on paper, shading the area you are looking for and labeling the graph with the information you know.
- Go to DISTR/normalcdf.
- If you know the z-scores for the area you want, in the parentheses put the lower z limit and then the upper z limit separated by commas. Use –1E99 for negative infinity and 1E99 for positive infinity. After you hit ENTER, the calculator will tell you the area under the curve (probability) between your 2 z values.

$$P\left(z_{low} < z < z_{high}\right) = \text{normalcdf}\left(z_{low}, z_{high}\right)$$

- If you do not know the z scores, you may let the calculator find them for you. In the parentheses, put the lower x limit, the upper x limit, the mean of the normal distribution, and the standard deviation of the normal distribution. The latter two values are used to convert your x values into z values.

$$P\left(x_{low} < x < x_{high}\right) = \text{normalcdf}\left(x_{low}, x_{high}, \mu, \sigma\right)$$

Finding a z-score corresponding to a given area under the normal curve

- First, draw and label a graph on paper, shading the area you are looking for and labeling the graph with the information you know.
- Go to DISTR/invNorm.
- You must input the proportion (area) to the left of the z-score you are looking for. Let's call this A. If you are given the area to the right, subtract that from 1 to get A. Put this number in the parentheses, and the calculator will give you the z-score that separates the distribution into two parts: one with area A to the left of that z-score and one with area 1-A to the right of that z-score.

$$z = \text{InvNorm}(A)$$

- If you want to find the x value instead of the z score, you need to tell the calculator the mean and standard deviation, too.

$$x = \text{InvNorm}(A, \mu, \sigma)$$

AP Statistics Exam Connections

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

http://apcentral.collegeboard.com.

Free Response Questions	
2003	Question 3
2004 Form B	Question 3
2006 Form B	Question 3

AP* Statistics Review

Inference for Proportions

Teacher Packet

Inference for Proportions is used to determine the validity of a claim about a population proportion.

Suppose you are playing a dice game with a friend. You notice that the number 3 has been rolled many times and you think that your friend is using a weighted die. You decide to test the die by rolling it 100 times and recording the number of threes that are rolled. You expect that one sixth of the rolls should result in a three if the die is fair. If your test produces a count of threes that is far enough from your expected value of about 17 rolls, you will suspect that the die is weighted.

Procedures for Inference about a population proportion: (1-proportion z-test)

1. Describe the population parameter used in the hypothesis test.
2. State the null and alternate hypotheses using symbols or words.
3. State the alpha level that will be used.
4. Identify the inference procedure by name or by formula.
5. Verify any conditions that need to be met for that inference procedure.
   a. The sample is a simple random sample from the population of interest.
   b. The population is large relative to the sample size.
      $10n < N$ (N = the size of the population)
   c. The sample size is large, so use of a normal model is reasonable.
      $np_0 \geq 10$
      $n(1-p_0) \geq 10$
6. Carry out the procedure showing the test statistic and p-value.
   $z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$ for a one proportion test
7. Interpret the results in the context of the situation.

Example:

You suspect a number cube (die) used in a game of chance is weighted to favor a roll of three. You roll the die 100 times and count the number of threes rolled to be 22. Is this good evidence that the number three is more likely to be rolled than the other numbers?

The population of interest is the proportion of rolls that result in a three.
p = probability of rolling a head for the population of all rolls.
$\hat{p} = \dfrac{22}{100} = 0.22$ is the sample proportion of rolls of three.

Is this proportion (0.22) significantly different from the population proportion (one sixth) or could we get this merely by chance alone?

H_O: p = 1/6 The population proportion of each number on the die is one sixth

H_A: p > 1/6 The die is weighted to land on a three more often than other numbers

The alpha level will be set at 5% or $\alpha = 0.05$

We will use a one proportion z-test.

We can assume the 100 rolls are a simple random sample from the population of all rolls.
$10n = 1000$ The population of all rolls of the die would be infinite, so our population is large relative to our sample.

$np_0 = (100)(\tfrac{1}{6}) = 16.7$

$16.7 \geq 10$

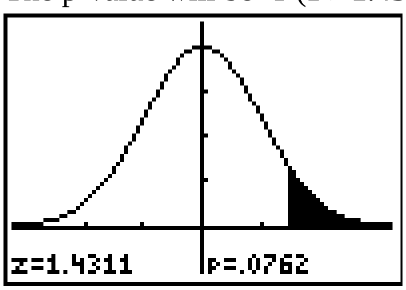
$n(1-p_0) = (100)(\tfrac{5}{6}) = 83.3$

$83.3 \geq 10$

We will now standardize our $\hat{p}$ value by finding the z-statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.22 - 0.1667}{\sqrt{\frac{0.1667(1-0.1667)}{100}}} = 1.431$$

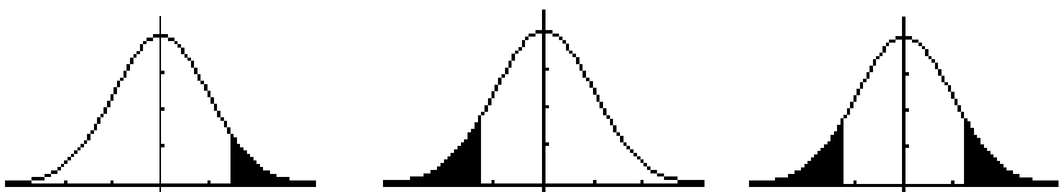
The p-value will be $P(z > 1.431) = 0.764$



Because the p-value (0.764) is greater than the chosen alpha level (0.05) we fail to reject the null hypothesis. There is insufficient evidence to support the claim that the true proportion of threes that are rolled on this die is greater than one sixth.

The different cases of a One Proportion z-test

In the example we just looked at we used a right-tailed test because we thought the number of threes rolled was too high ($H_A : p > 1/6$). We can also run tests that question if the proportion is too low (<, a left-tailed test) or not equal to (≠, a two-tailed test)



Right-tailed test (>) Left-tailed test (<) Two-tailed test (≠)

In each of these cases, the area under the curve (corresponding to your z-score) will be the p-value that is compared to the alpha value. The two-sided test requires that we double the area on one side.

Two-Proportion z-test: (2-proportion z-test)

1. Describe both population parameters used in the hypothesis test.
2. State the null and alternate hypotheses using symbols or words.

   $H_O: \; p_1 = p_2 \text{ or } p_1 - p_2 = 0$

   $H_A: \begin{cases} p_1 > p_2 \\ p_1 < p_2 \\ p_1 \neq p_2 \end{cases}$

   Because the null assumption is that the population proportions are equal, we will pool the sample proportion

   $\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$
3. State the alpha level that will be used.
4. Identify the inference procedure by name or by formula.

   2-proportion z-test
5. Verify any conditions that need to be met for that inference procedure.
   a. The samples are independent random samples from the population of interest.
   b. The population sizes are large relative to the respective sample sizes.

      $10n_1 < N_1$

      $10n_2 < N_2$
   c. The sample sizes are large, so the use of normal models is reasonable.

      $n_1\hat{p} \geq 10 \quad n_1(1-\hat{p}) \geq 10$

      $n_2\hat{p} \geq 10 \quad n_2(1-\hat{p}) \geq 10$

6. Carry out the procedure showing the test statistic and the p-value.

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\left(1-\hat{p}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$ for a two-proportion test

7. Interpret the results in the context of the situation.

Example:
A major restaurant believes that male customers prefer their new spicy hot wings over their regular hot wings more than female customers. To test this claim, a random sample of 45 men and 50 women were given a taste test of both types of wings and asked which they preferred. Of the 45 men, 34 preferred the new spicy hot wings and of the 50 women, 32 preferred the new spicy hot wings. Is this significant evidence that men prefer the new spicy wings more than women do?

p_m = population proportion of men that prefer the new spicy hot wings over the regular hot wings.
p_w = population proportion of women that prefer the new spicy hot wings over the regular hot wings.
$\hat{p}_m = \frac{34}{45} = 0.7556 =$ sample proportion of men that prefer the new spicy hot wings over the regular hot wings.
$\hat{p}_w = \frac{32}{50} = 0.64 =$ sample proportion of women that prefer the new spicy hot wings over the regular hot wings.
$\hat{p} = \frac{34+32}{45+50} = 0.6947 =$ pooled sample proportion of people that prefer the new spicy hot wings over the regular hot wings.

$H_O:\ p_m = p_w$ The population proportion of men is equal the population proportion of women that prefer the new spicy hot wing over the regular.
$H_A:\ p_m > p_w$ The population proportion of men is greater than the population proportion of women that prefer the new spicy hot wing over the regular.

$\alpha = 0.05$

2-proportion z-test

Both samples are a simple random sample from the population of interest as stated by the problem.

The population is large relative to the sample size
$10n_1 = (10)(45) = 450 < N_1$
$10n_2 = (10)(50) = 500 < N_2$
It is reasonable to assume that there are more than 450 men or 500 women.
The sample size is large
$n_1\hat{p} = (45)(0.6947) = 31.26 \geq 10 \quad n_1(1-\hat{p}) = (45)(1-0.6947) = 13.74 \geq 10$
$n_2\hat{p} = (50)(0.6947) = 34.74 \geq 10 \quad n_2(1-\hat{p}) = (50)(1-0.6947) = 15.27 \geq 10$

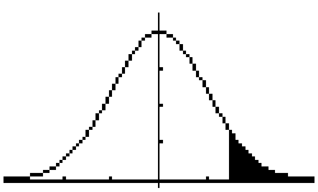
$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.756-0.64}{\sqrt{0.6947(1-0.6947)\left(\frac{1}{45}+\frac{1}{50}\right)}} = 1.226$$

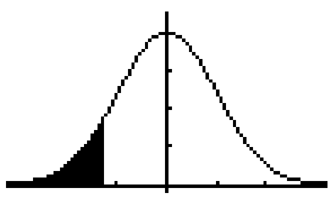
$P(z > 1.226) = 0.1093$ or 0.111 from the calculator

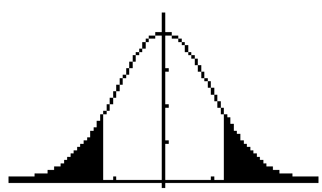
Because the p-value (0.1093) is greater than the alpha level (0.05) we fail to reject the null hypothesis. We have insufficient evidence to support the claim that men prefer the new spicy hot wings more than women do.

The different cases of a Two Proportion z-test

In the example we just looked at we used a right-tailed test because we thought that men preferred the new hot wings more than women ($H_A:\ p_m > p_w$). We can also run tests that question if men prefer it less than women (<, a left-tailed test) or not equal to (≠, a two-tailed test)


Right-tailed test (>)


Left-tailed test (<)


Two-tailed test (≠)

Multiple Choice Questions on Inference for Proportions:

1. While visiting a major city, a travel agent reads a pamphlet about the hotels in the city. The pamphlet states that there are the same proportion of non-smoking hotel rooms as smoking hotel rooms in the city. The travel agent does not believe this claim and decides to test it when she returns to her office. Which of the following would be an appropriate set of hypothesis for this test?

(A) $H_O : p_{non-smoke} = p_{smoke} \quad H_A : p_{non-smoke} \neq p_{smoke}$
(B) $H_O : \hat{p}_{non-smoke} = \hat{p}_{smoke} \quad H_A : \hat{p}_{non-smoke} \neq \hat{p}_{smoke}$
(C) $H_O : p_{non-smoke} = p_{smoke} \quad H_A : p_{non-smoke} > p_{smoke}$
(D) $H_O : \hat{p}_{non-smoke} = \hat{p}_{smoke} \quad H_A : \hat{p}_{non-smoke} > \hat{p}_{smoke}$
(E) $H_O : p_{non-smoke} \neq p_{smoke} \quad H_A : p_{non-smoke} = p_{smoke}$

2. Greyhound dogs are used at dog race tracks for people to make wagers on which dog will win the race. After the dogs get old or are injured, they are retired from racing. Rescue organizations find homes for these greyhounds. One rescue organization claims that only 55% of all retired greyhounds find a home. A pet store owner and pet advocate believes the rescue organization is significantly overestimating the percentage of retired greyhounds who find a new home. The pet store owner checks the status (found home or not found home) of a random sample of 150 registered greyhounds and finds that 71 have been placed in a new home. At an alpha level of 0.05, what should the pet store owner conclude?

(A) Because the p-value is greater than the alpha level, the pet store owner has evidence that 55% of retired greyhounds have been placed in homes.

(B) Because the p-value is greater than the alpha level, the pet store owner has evidence that less than 55% of retired greyhounds have been placed in homes.

(C) Because the p-value is greater than the alpha level, the pet store owner has insufficient evidence that the proportion of retired greyhounds is more than 55%.

(D) Because the p-value is less than the alpha level, the pet store owner has evidence that less than 55% of retired greyhounds have been placed in homes.

(E) Because the p-value is less than the alpha level, the pet store owner has insufficient evidence that less than 55% of retired greyhounds have been placed in homes.

3. A popular lakeside resort on the Great Lakes promotes in their brochure that you have an 85% chance of seeing a bald eagle while you are staying at their resort. A local businessman believes that this claim is outrageous. He has lived in the area for years and seen very few bald eagles. He decides to conduct a test to disprove the resorts claim of 85%. Over the following month, he asks every visitor to his store if they are staying at the resort and if they have seen a bald eagle while staying there. He finds that 35 people have stayed at the resort and only 15 of them have seen a bald eagle. Which of the following is true about this data?
(A) The p-value is essentially zero, so at any reasonable alpha level he has disproven the resorts claim.
(B) The p-value is essentially zero, so there is no proof that the resorts claim is false.
(C) The p-value is essentially zero, which means that no inference can be made.
(D) The sample size is too small, so not inference can be made.
(E) The sample data does not come from a simple random sample and therefore cannot be used to conduct this hypothesis test.

4. A random sample of 100 individuals are polled to determine the number that own a pet. The poll returns a result of 62 individuals owning a pet. The researcher is interested in determining if the number of individuals owning a pet has decreased from the previous believed proportion of 70%. Which of the following is true at the 5% level of significance?
(A) The p-value is 0.04 which indicates that the proportion of individuals owning a pet may have decreased.
(B) The p-value is 0.04 which indicates that the proportion of individuals owning a pet has not changed.
(C) The p-value is 0.08 which indicates that the proportion of individuals owning a pet may have changed.
(D) The p-value is 0.08 which indicates that there is no evidence that the proportion of individuals owning a pet has decreased.
(E) The sample size is too small to make an inference on this proportion.

5. A production line that makes digital cameras has a very high proportion of defective cameras that are being produced, 34%. After redesigning the production process and improving employee training, the production line is restarted. A random sample of 30 cameras is tested to see if they are defective. After testing these 30 cameras, only 8 were found to be defective. Is this good evidence that the improvements to the production line have decreased the proportion of defective cameras produced?
(A) yes, the new proportion, 27%, is much lower than the previous proportion.
(B) yes, the redesign and training have improved camera production.
(C) yes, there is significant evidence that the proportion of cameras that are defective has reduced.
(D) no, this data does show a reduction in the proportion of defective cameras, but a proportion this extreme could occur simply by chance 19.8% of the time.
(E) no, the redesign and training produced a p-value that was lower than a reasonable alpha level. Therefore we believe that the proportion of cameras that are defective is still 34%.

6. Are women better at face recognition than men? A psychology student wanted to test the idea that women can recognize a face that they have seen before better than men. A random sample of 50 women and 60 men are shown 20 black-and-white pictures of faces. After exactly one minute, they are shown a grid of 50 faces and asked to mark each of the 20 faces they were shown before. The women scored 78% recognition while the men scored only 70% recognition. Which of the following is true?
(A) p-value = 0.0858
(B) p-value = 0.1715
(C) p-value = 0.3430
(D) p-value = 0.7364
(E) p-value = 0.9861

7. A one proportion z-test is performed using a left tailed test. After further researching the population in question, it is determined that a two tailed test would be more appropriate. Which of the following would be true?
(A) The p-value from the original test should be halved for the two-tailed test.
(B) The alpha level from the original test should be doubled for the two-tailed test.
(C) The p-value from the original test should be doubled for the two-tailed test.
(D) The z-statistic from the original test should be doubled for the two-tailed test.
(E) A new random sample must be taken to run the two tailed test.

8. A lumber company requires that 68% of the trees in an area to be logged have usable wood for them to make a profit. They are offer a plot of land to log. The lumber company takes a random sample of 40 trees and finds that 25 of the trees have usable wood. Should the timber company agree to log this plot of land?
(A) yes, because the sample proportion produces a p-value that is very large
(B) yes, because the sample proportion is very close to the required 68%
(C) no, because the sample proportion is lower than the required 68%
(D) no, because the sample proportion produces a p-value that is very small
(E) no, there is too much error associated with this sample proportion.

9. Which of the following true about a one-proportion z-test?
(A) There is no restriction on the sample size.
(B) This test works even if the sampling process is not random.
(C) This test works when using a census.
(D) The population should be large compared to the sample size.
(E) For this test to be affective, the alpha level must be set at 5%

10. The owner of a large, vacant downtown building has agreed to tear down the building to make way for a new parking garage if at least 51% of the city's population agrees with the building's destruction. Because a poll of every citizen of the city would be too expensive, a random sample of 135 city residents is polled. The building owner has agreed to use a left tailed one proportion z test at an alpha level of 5% to determine if the actual proportion of residents is lower than 51%. Which of the following would be the minimum number of agreeing city residents required to make the owner decide to tear down the building?
(A) 69
(B) 67
(C) 60
(D) 55
(E) 51

Free Response Questions on Inference for Proportions

Free Response 1.

A major toy company has started production of a new building block toy system that uses blocks that snap together. Each day the production line uses one batch of melted plastic to produce millions of blocks. As production progresses through the day, there is less and less liquid plastic left in the batch. The quality control engineer suspects that the amount of liquid plastic left in the batch effects the quality of the fit of the blocks. To test this, a random sample of 100 blocks is taken at the beginning of the batch and another random sample of 100 blocks at the end of the batch. The quality control engineer plans to run a test that will determine if there is a difference in the proportion of blocks that fit together at the beginning of the batch as compared to the end of the batch. The proportion of blocks that fit together at the beginning of the batch is 88% and the proportion that fit together at the end of the batch are 79%. Is there a significant difference in the proportion of blocks that fit together from the beginning of the batch to the end of the batch?

Free Response 2.

A cookie making company ships hundreds of thousands of their cookies to costumers all over the United States. One of the largest complaints the cookie company has had is the number of individually -wrapped cookies that arrive broken after shipping. In an effort to reduce the proportion of broken cookies, the cookie company has implemented a new packing procedure. Before the new procedure was implemented, the cookie company had 13% of their individually-wrapped cookies broken in shipping. The cookie company wanted to test the results of their new packing procedure. To do this, they placed a postage-paid survey in with the new packing materials for the costumers to return. The customer simply wrote down the number of cookies that were broken and dropped the survey in the mail. Of the 450 individually-wrapped cookies represented by the surveys returned, there were 68 that were reported broken.

(a) Using a test of significance, is there evidence that the new packing procedure reduced the proportion of cookies broken?

(b) Does the method of collecting this data introduce bias into the sample? If so, discuss this bias and how it would affect the results of the test.

Key to Inference for Proportions Multiple Choice

1. A appropriate hypotheses for a 2-prop-z-test
2. D p-value = 0.0296
3. E conditions of 1-prop-z-test
4. A 1-prop-z-test p-value and conclusion
5. D definition of p-value
6. B 2-prop-z-test
7. C one-tailed test vs two-tailed test
8. A conclusion based on 1-prop-z-test
9. D conditions of a 1-prop-z-test
10. C find x which results in the rejection of the null hypothesis

<u>Rubrics for Inference for Proportions Free Response</u>

Free Response 1. ***Solution***

1. **Describe both population parameters used in the hypothesis test.**

p_B = population proportion of blocks that fit at the beginning of the batch.

p_E = population proportion of blocks that fit at the end of the batch.

$\hat{p}_B = 0.88$ = sample proportion of blocks that fit at the beginning of the batch.

$\hat{p}_E = 0.79$ = sample proportion of blocks that fit at the end of the batch.

$\hat{p} = \dfrac{88+79}{100+100} = 0.835$ = pooled sample proportion of blocks that fit together.

2. **State the null and alternate hypotheses using symbols or words.**

$H_O:\ p_B = p_E$ The population proportion of beginning blocks that fit is equal to the population proportion of ending blocks that fit.

$H_A:\ p_B \neq p_E$ The population proportion of beginning blocks that fit is equal to the population proportion of ending blocks that fit. This signifies a two tailed test.

3. **State the alpha level that will be used.**

$\alpha = 0.05$

4. **Identify the inference procedure by name or by formula.**

2-proportion z-test

5. **Verify any conditions that need to be met for that inference procedure.**

Both samples are independent random samples from the population of interest as stated by the problem.

The populations are both large relative to the sample sizes

$10n_1 = (10)(100) = 1000 < N_1$

$10n_2 = (10)(100) = 1000 < N_2$

The problem states that millions of blocks are made from one batch.

The sample sizes are large

$n_1\hat{p} = (100)(0.835) = 83.5 \geq 10 \quad n_1(1-\hat{p}) = (100)(1-0.835) = 16.5 \geq 10$

$n_2\hat{p} = (100)(0.835) = 83.5 \geq 10 \quad n_2(1-\hat{p}) = (100)(1-0.835) = 16.5 \geq 10$

6. **Carry out the procedure showing the test statistic and the p-value.**

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.88-0.79}{\sqrt{0.835(1-0.835)\left(\frac{1}{100}+\frac{1}{100}\right)}} = 1.7145$$

$\mathbf{2}\,P(z > 1.7145) = 0.0872$ or 0.0864 from the calculator

7. Interpret the results in the context of the situation.
Because the p-value (0.1744) is greater than the alpha level (0.05) we fail to reject the null hypothesis. We have insufficient evidence to support the claim that the proportion of blocks that fit together is different at the end of the batch from the beginning of the batch.

Scoring

The problem is divided into four parts. Each part is essentially correct (E), partially correct (P), or incorrect (I).

Part (a) (1 and 2) identify the parameters and set up the hypotheses

Part (a) is essentially correct if the parameters are identified and the correct null and alternate hypotheses are stated.
Part (a) is partially correct if only one of the above are provided in the answer
Part (a) is incorrect if neither component is provided in the answer

Part (b) (3 and 4) Identify the alpha level and the test to be used by name or equation.

Part (b) is essentially correct if the student provides an acceptable alpha level and clearly labels the test to be used. These responses can be written in any part of the question.
Part (b) is partially correct if the student only provides one of the two components.
Part (b) is incorrect if fails to provide either of these components.

Part (c) (5) Conditions

Part (c) is essentially correct if all three conditions have been checked for both samples.
Independent random samples
Sample sizes small relative to the populations
Large enough sample sizes
Part (c) is partially correct if the student checks only two of the conditions
or
The student checks all three of the conditions for only one of the samples.
Part (c) is incorrect if only one condition is checked, or two conditions checked on one sample.

Part (d) (6 and 7) Mechanics and conclusion

Part (d) is essentially correct if the student calculates an appropriate p-value and interprets it in the context of the problem correctly.
Part (d) is partially correct if the student calculates the correct p-value and has a weak conclusion with no direct connection to the context of the problem.
Part (d) is incorrect if no conclusion is given.

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

Free Response 2. **Solution**

Part (a)

1. **Describe the population parameter used in the hypothesis test.**

p = the population proportion of shipped cookies that are broken

$\hat{p} = \frac{68}{450} = 0.1511 =$ the sample proportion of cookies broken

n = 450

2. **State the null and alternate hypothesis using symbols or words.**

$H_O:\ p = 0.13$ The population proportion of cookies broken is 0.13

$H_A:\ p < 0.13$ The population proportion of cookies broken is less than 0.13

3. **State the alpha level that will be used.**

$\alpha = 0.05$

4. **Identify the inference procedure by name or by formula.**

We will use the one proportion z-test

5. **Verify any conditions that need to be met for that inference procedure.**

The sample is a simple random sample from the population of interest
The sample is not stated as being a simple random sample (in fact it is not a random sample at all). This condition has not been met, but we will continue with caution.

The population is large relative to the sample size
$10n = (10)(450) = 4500 < N$ the problem states that hundreds of thousands of cookies are shipped. This condition has been met.

The sample size is large
$np_0 = (450)(0.13) = 58.5 \geq 10$
$n(1-p_0) = (450)(1-0.13) = 391.5 \geq 10$

6. **Carry out the procedure showing the test statistic and the p-value.**

$$z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.1511-0.13}{\sqrt{\frac{0.13(1-0.13)}{450}}} = 1.33$$

$P(z < 1.33) = 0.9082$

7. **Interpret the results in the context of the situation.**

Because the p-value (0.9082) is very large compared to the alpha value (0.05) we fail to reject the null hypothesis. We have no evidence that the population proportion is less than 0.13. We can not be too confident in this conclusion because a very important condition was not met, random sampling.

Part (b)

The method of collecting this data is subject to nonresponse bias (voluntary response bias). Because the respondents were not randomly selected, there may be bias in the data. It might be that only costumers that received a large number of broken cookies returned their surveys. This would influence the sample proportion to be higher than the true proportion of broken cookies.

Scoring
Part (a) and (b) will be graded as essential correct, partially correct or incorrect.

Part (a) is graded essentially correct if each of the following is included in the response:
- Appropriate H_O and H_A
- Conditions checked with comment on the random sample condition not being met
- A conclusion in the context of the problem

Part (a) is graded partially correct if two of the three are included in the response
Part (a) is graded incorrect if less than two of the three are included in the response

Part (b) is graded essential correct if the student:
- Indicates that nonresponse bias is present
- Communicates how the bias could affect the data by example.

Part (b) is graded partially correct if the student gives an example of how the bias could affect the data but does not communicate it clearly.
Part (c) is graded incorrect if they state that there is bias with no explanation or if they say there is no bias.

4 Complete Response

Both parts essentially correct.

3 Substantial Response

One part essentially correct and one part partially correct

2 Developing Response

One part essentially correct and one part incorrect
OR
Partially correct on both parts

1 Minimal Response
Partially correct on one part.

AP Statistics Exam Connections

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

http://apcentral.collegeboard.com.

Free Response Questions	
2003	Question 3
2004	Question 5
2004 B	Question 6
2006 B	Question 6

AP* Statistics Review

Confidence Intervals

Teacher Packet

Confidence Intervals

A **Confidence Interval** is an interval that is computed from sample data and provides a range of plausible values for a population parameter.
A **Confidence Level** is a number that provides information on how much "confidence" we have in the method used to construct a confidence interval estimate. This level specifies the percentage of all possible samples that produce an interval containing the true value of the population parameter.

Constructing a Confidence Interval

The steps listed below should be followed when asked to calculate a confidence interval.

1. Identify the population of interest and define the parameter of interest being estimated.
2. Identify the appropriate confidence interval by name or formula.
3. Verify any conditions (assumptions) that need to be met for that confidence interval.
4. Calculate the confidence interval.
5. Interpret the interval in the context of the situation.

The general formula for a confidence interval calculation is:

Statistic ± (critical value) · (standard deviation of statistic)

The critical value is determined by the confidence level.
The type of statistic is determined by the problem situation. Here we will discuss two types of statistics: mean and proportion.

Confidence Interval for Proportions: (1-proportion z-interval)
We are finding an interval that describes the population proportion (p *or* π).
The general formula uses the sample proportion ($\hat{p}$) and the sample size (n).

Statistic ± (critical value) · (standard deviation of statistic)

$$\hat{p} \pm \left(z^*\right)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

The conditions that need to be met for this procedure are:

1. The sample is a simple random sample.
2. The population is large relative to the sample.
   $10n < N$ (N = the size of the population)
3. The sampling distribution of the sample proportion is approximately normal.
   $np \geq 10$
   $n(1-p) \geq 10$

Complete the calculations after showing that the conditions are met.
Write the answer in the context of the original problem.

Sample question:

The owner of a popular chain of restaurants wishes to know if completed dishes are being delivered to the customer's table within one minute of being completed by the chef. A random sample of 75 completed dishes found that 60 were delivered within one minute of completion. Find the 95% confidence interval for the true population proportion.

1. Identify the population of interest and define the parameter of interest being estimated.

The population of interest is the dishes that are being served at this chain of restaurants.

p = the population proportion of dishes that are served within one minute of completion.

$\hat{p} = \frac{60}{75} = 0.8$ = the sample proportion of dishes that are served within one minute of completion.

$n = 75$ is the sample size.

2. Identify the appropriate confidence interval by name or formula.

We will use a 95% confidence z-interval for proportions.

3. Verify any conditions (assumptions) that need to be met for that confidence interval.

The problem states that this is a simple random sample.

$$10n < N$$
$$10(75) < N$$
$$750 < N$$

It is reasonable to assume that a popular chain of restaurants will serve more than 750 dishes.

$$n\hat{p} \geq 10 \qquad n(1-\hat{p}) \geq 10$$
$$75(0.8) = 60 \qquad 75(1-.8) = 15$$
$$60 \geq 10 \qquad 15 \geq 10$$

It is reasonable to use a normal model.

4. Calculate the confidence interval.

At a 95% CI, the critical value is $z^* = 1.96$.

This value is found on the last row of the t-distribution table.

$$0.8 \pm 1.96\sqrt{\frac{0.8(1-0.8)}{75}}$$
$$0.8 \pm 0.091$$
$$(0.709, 0.891)$$

5. Interpret the interval in the context of the situation.

We are 95% confident that the true proportion of dishes that are served within one minute of completion for this chain of restaurants is between 0.709 and 0.891.

<u>Confidence Interval for Means with σ known:</u> (z-interval)
We are finding an interval that describes the population mean (μ).
The general formula uses the sample mean ($\overline{x}$), the population standard deviation (σ) and the sample size (n).

Statistic ± (critical value) · (standard deviation of statistic)

$$\overline{x} \pm (z^*)\left(\frac{\sigma}{\sqrt{n}}\right)$$

The conditions that need to be met for this procedure are:
1. The sample is a simple random sample.
2. The population is normal or $n \geq 30$
3. The population standard deviation (σ) is known.

Complete the calculations after showing that the conditions are met.
Write your answer in the context of the original problem.

<u>Sample question:</u>

An asbestos removal company places great importance on the safety of their employees. The protective suits that the employees wear are designed to keep asbestos particles off the employee's body. The owner is interested in knowing the average amount of asbestos particles left on the employee's skin after a days work. A random sample of 100 employees had skin tests after removing their protective suit. The average number of particles found was .481 particles per square centimeter. Assuming that the population standard deviation is 0.35 particles per square centimeter, calculate a 95% confidence interval for the number of particles left on the employee's skin.

<u>Solution:</u>

The population of interest is the employees of the asbestos removal company.
μ = the population mean of particles per square centimeter after a days work wearing the protective suit.
$\overline{x} = 0.481$ = the sample mean of the number of particles per square centimeter.
$n = 100$ is the sample size.
$\sigma = 0.35$ =population standard deviation.

We will use a 95% confidence z-interval for means (z-interval).
The problem states that this is a simple random sample.
n = 100 is greater than 30 so we can assume that use of a normal model is reasonable.

At a 95% CI, the critical value is $z^* = 1.96$.
This value is found on the last row of the t-distribution table.

$$0.481 \pm 1.96\left(\frac{0.35}{\sqrt{100}}\right)$$

$$0.481 \pm 0.069$$

$$(0.412, 0.550)$$

We are 95% confident that the true mean number of particles of asbestos found on the skin of an employee after a days work is between 0.412 and 0.550 particles per square centimeter.

<u>Confidence Interval for Means with σ unknown:</u> (t-interval)
You will be finding an interval that will describe the population mean (μ).
The general formula will use the sample mean ($\bar{x}$), the sample standard deviation (s), the sample size (n), and the degrees of freedom ($n-1$).

Statistic ± (critical value) · (standard deviation of statistic)

$$\bar{x} \pm (t^*)\left(\frac{s}{\sqrt{n}}\right)$$

The conditions that need to be met for this procedure are:
1. The sample is a simple random sample.
2. The population is approximately normal (graphical support required) or $n \geq 40$
3. The population standard deviation (σ) is unknown.

Complete the calculations after showing that the conditions are met.
Write your answer in the context of the original problem.

This situation (where the population standard deviation is unknown) is much more realistic that the previous case.

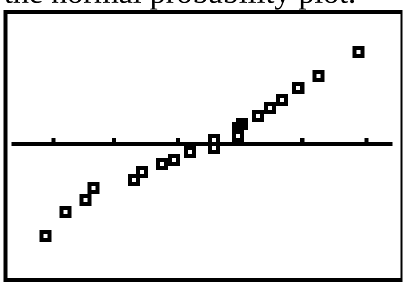
<u>Sample question:</u>
A biology student at a major university is writing a report about bird watchers. She has developed a test that will score the abilities of a bird watcher to identify common birds. She collects data from a random sample of people that classify themselves as bird watchers (data shown below). Find a 90% confidence interval for the mean score of the population of bird watchers.

4.5	*9.1*	*8*	*5.9*	*7.0*	*5.2*	*7.3*	*7.0*	*6.6*	*5.1*
7.6	*8.2*	*6.4*	*4.8*	*5.8*	*6.2*	*8.5*	*7.3*	*7.8*	*7.4*

Solution:

The population of interest is people that classify themselves as bird watchers.
μ = the population mean score on the bird identification ability test.
$\bar{x} = 6.785$ = the sample mean of the scores on the ability test.
$s = 1.2828$ = sample standard deviation of test scores.
$n = 20$ is the sample size.
df = 20 – 1 = 19 degrees of freedom.

We will use a 90% confidence t-interval for means (t-interval).
The problem states that this is a simple random sample.
The sample size, 20, is smaller than 40 so we will assess normality by looking at the normal probability plot.



The normal probability plot appears to be linear so we will assume an approximately normal distribution of scores.
The population standard deviation is unknown.

At a 90% CI, the critical value is $t^* = 1.729$.
This value is found on the t-distribution table using 19 degrees of freedom.

$$6.785 \pm 1.729\left(\frac{1.2828}{\sqrt{20}}\right)$$

$$6.785 \pm 0.49595$$

$$(6.289, 7.281)$$

We are 90% confident that the true mean score on the bird identification ability test of the population of persons that classify themselves as bird watchers is between 6.289 and 7.281.

Summary of Confidence Intervals with One Sample

Confidence Interval Type	**Formula**	**Conditions**	**Calculator Test**
Proportions	$\hat{p} \pm (z^*)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	1. The sample is a simple random sample. 2. The population is large relative to the sample $10n < N$ 3. $np \geq 10$ $n(1-p) \geq 10$	1-PropZInterval
Means (σ known)	$\bar{x} \pm (z^*)\left(\frac{\sigma}{\sqrt{n}}\right)$	1. The sample is a simple random sample. 2. The population is normal or $n \geq 30$ 3. The population standard deviation (σ) is known.	ZInterval
Means (σ unknown)	$\bar{x} \pm (t^*)\left(\frac{s}{\sqrt{n}}\right)$	1. The sample is a simple random sample. 2. The population is approximately normal (graphical support required) or $n \geq 40$ 3. The population standard deviation (σ) is unknown.	TInterval

Confidence Intervals with TWO Samples

We may be asked to work a confidence interval problem to estimate the difference of two population parameters. The process is very similar to that of a one sample confidence interval. We must make sure that the conditions are met for both samples and that the samples are independent of each other.

The formulas for the two sample confidence intervals are as follows:

Confidence Interval Type	Formula	Conditions	Calculator Test
Difference of Proportions $(p_1 - p_2)$	$\hat{p}_1 - \hat{p}_2 \pm (z^*)\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	1. The samples are independently selected simple random samples. 2. Each population is large relative to the sample $10n_1 < N_1$ $10n_2 < N_2$ 3. $n_1 p_1 \geq 10$ $n_1(1-p_1) \geq 10$ $n_2 p_2 \geq 10$ $n_2(1-p_2) \geq 10$	2-PropZInterval
Difference of Means (σ known) $(\mu_1 - \mu_2)$	$\bar{x}_1 - \bar{x}_2 \pm (z^*)\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	1. The samples are independently selected simple random samples. 2. Both populations are normal or $n_1 \geq 30$ $n_2 \geq 30$ 3. Both population standard deviations (σ_1 *and* σ_2) are known.	2-samp-ZInterval
Difference of Means (σ unknown) $(\mu_1 - \mu_2)$	$\bar{x}_1 - \bar{x}_2 \pm (t^*)\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ Note: The formula for degrees of freedom is $df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$ This formula is not provided on the formula chart. A more conservative quantity used by some statisticians is the smaller of the two individual degrees of freedom; either $n_1 - 1$ or $n_2 - 1$.	1. The samples are independently selected simple random samples. 2. Both populations are approximately normal (graphical support is required) or $n_1 \geq 40$ $n_2 \geq 40$ 3. Both population standard deviations (σ_1 *and* σ_2) are unknown.	2-samp-TInterval

Sample question:

Two popular strategy video games, AE and C, are known for their long play times. A popular game review website is interested in finding the mean difference in play time between these games. The website selects a random sample of 43 gamers to play AE and finds their sample mean play time to be 3.6 hours with a standard deviation of 0.9 hours. The website also selected a random sample of gamers to test the game C. There test included 40 gamers with a sample mean of 3.1 hours and a standard deviation of 0.4 hours. Find the 98% confidence interval for the difference $\mu_{AE} - \mu_C$.

Solution:

The population of interest is play time of games AE and C.

μ_{AE} = the population mean play time for game AE.

μ_C = the population mean play time for game C.

$\bar{x}_{AE} = 3.6$ = the sample mean of play time for game AE.

$\bar{x}_C = 3.1$ = the sample mean of play time for game C.

$s_{AE} = 0.9$ = sample standard deviation play time for game AE.

$s_C = 0.4$ = sample standard deviation play time for game C.

$n_{AE} = 43$ is the sample size of the players of game AE.

$n_C = 40$ is the sample size of the players of game C.

df = 43 – 1 = 42 degrees of freedom (using the simplified approximation for df).

We will use a 98% confidence t-interval for the difference of means (2-sample-t-interval).
The problem states that each is a simple random sample.
The sample sizes are 43 and 40 which are both at least 40.
The population standard deviation is unknown.

At a 98% CI, the critical value is $t^* = 2.457$.
This value is found on the t-distribution table using 40 degrees of freedom (because the table does not have a value for 42).

$$(3.6-3.1) \pm 2.423\sqrt{\frac{0.9^2}{43}+\frac{0.4^2}{40}}$$

$$0.5 \pm 0.3662$$

$$(0.1338, 0.8662)$$

We are 98% confident that the true difference in mean play time between games AE and C falls between 0.1338 and 0.8662 hours.

Note: The calculator gives a result of (0.1386, 0.8614) for this confidence interval. The difference in these answers is due to its use of 58.872 degrees of freedom. Our answer has a slightly wider spread, and thus is more conservative.

Other Confidence Interval Topics:

Statistic ± (critical value) · (standard deviation of statistic)

Margin of Error (ME) – the margin of error is the value of the critical value times the standard deviation of the statistic. It is the plus or minus part of the confidence interval.

Some problems might ask you to determine the sample size required given a margin of error. This requires a little algebra to work backward through the equations. The equations are listed below.

Confidence Interval Type	Formula for finding the sample size within a margin of error ME
1-sample proportion	$n = \hat{p}(1-\hat{p})\left(\frac{z*}{ME}\right)^2$ Note: if $\hat{p}$ is not given in this type of problem, a conservative value to use if 0.5.
1-sample mean (z-interval)	$n = \left(\frac{z*\sigma}{ME}\right)^2$

Keep in mind the effects of changing the confidence level. A large confidence level (say 99% as compared to 90%) produces a larger margin of error. To be more confident we must include more values in our range.

Do not confuse the meaning of a confidence level: A 95% confidence level means that if we repeated the sampling process many times, the resulting confidence interval would capture the true population parameter 95% of the time.

Multiple Choice Questions on Confidence Intervals

1. A random sample of 100 visitors to a popular theme park spent an average of $142 on the trip with a standard deviation of $47.5. Which of the following would the 98% confidence interval for the mean money spent by all visitors to this theme park?

   (A) ($130.77, $153.23)
   (B) ($132.57, $151.43)
   (C) ($132.69, $151.31)
   (D) ($140.88, $143.12)
   (E) ($95.45, $188.55)

2. How large of a random sample is required to insure that the margin of error is 0.08 when estimating the proportion of college professors that read science fiction novels with 95% confidence?

   (A) 600
   (B) 300
   (C) 150
   (D) 75
   (E) 25

3. A quality control specialist at a plate glass factory must estimate the mean clarity rating of a new batch of glass sheets being produced using a sample of 18 sheets of glass. The actual distribution of this batch is unknown, but preliminary investigations show that a normal approximation is reasonable. The specialist decides to use a t-distribution rather than a z-distribution because

   (A) The z-distribution is not appropriate because the sample size is too small.
   (B) The sample size is large compared to the population size.
   (C) The data comes from only one batch.
   (D) The variability of the batch is unknown.
   (E) The t-distribution results in a narrower confidence interval.

4. An independent random sample of 200 college football players and 150 college basketball players in a certain state showed that 65% of football players received academic tutors while 58% of basketball players received academic tutors. Which of the following is a 90 percent confidence interval for the difference in the proportion of football players that received tutors and the proportion of basketball players that received tutors for the population of this state?

(A) $(.65-.58)\pm 1.96\sqrt{(0.65)(0.58)\left(\frac{1}{200}+\frac{1}{150}\right)}$

(B) $(.65-.58)\pm 1.645\sqrt{(0.65)(0.58)\left(\frac{1}{200}+\frac{1}{150}\right)}$

(C) $(.65-.58)\pm 1.96\sqrt{\frac{(0.65)(0.35)}{200}+\frac{(0.58)(0.42)}{150}}$

(D) $(.65-.58)\pm 1.645\sqrt{\frac{(0.65)(0.35)}{200}+\frac{(0.58)(0.42)}{150}}$

(E) $(.65-.58)\pm 1.645\sqrt{(0.435)(0.867)\left(\frac{1}{200}+\frac{1}{150}\right)}$

5. The board of directors at a city zoo is considering using commercial fast food restaurants in their zoo rather than the current eateries. They are concerned that major donors to the zoo will not approve of the proposed change. Of the 280 major donors to the zoo, a random sample of 90 is asked "Do you support the zoo's decision to use commercial fast food restaurants in the zoo?" 50 of the donors said no, 38 said yes, and 2 had no opinion on the matter. A large sample z-interval, $\hat{p}\pm(z^*)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, was constructed from these data to estimate the proportion of the major donors who support using commercial fast food restaurants in the zoo. 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 no conditions are required for constructing a large sample confidence interval for proportions.
(C) This confidence interval is not valid because the sample size is too large compared to the population size.
(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 allowed as a response.

6. A research and development engineer is preparing a report for the board of directors on the battery life of a new cell phone they have produced. At a 95% confidence level, he has found that the battery life is 3.2 ± 1.0 days. He wants to adjust his findings so the margin of error is as small as possible. Which of the following will produce the smallest margin of error?

(A) Increase the confidence level to 100%. This will assure that there is no margin of error.
(B) Increase the confidence level to 99%.
(C) Decrease the confidence level to 90%.
(D) Take a new sample from the population using the exact same sample size.
(E) Take a new sample from the population using a smaller sample size.

7. A biologist has taken a random sample of a specific type of fish from a large lake. A 95 percent confidence interval was calculated to be 6.8 ± 1.2 pounds. Which of the following is true?

(A) 95 percent of all the fish in the lake weigh between 5.6 and 8 pounds.
(B) In repeated sampling, 95 percent of the sample proportions will fall within 5.6 and 8 pounds.
(C) In repeated sampling, 95% of the time the true population mean of fish weights will be equal to 6.8 pounds.
(D) In repeated sampling, 95% of the time the true population mean of fish weight will be captured in the constructed interval.
(E) We are 95 percent confident that all the fish weigh less than 8 pounds in this lake.

8. A polling company is trying to estimate the percentage of adults that consider themselves happy. A confidence interval based on a sample size of 360 has a larger than desired margin of error. The company wants to conduct another poll and obtain another confidence interval of the same level but reduce the error to one-third the size of the original sample. How many adults should they now interview?

(A) 40
(B) 180
(C) 720
(D) 1080
(E) 3240

9. A researcher is interested in determining the mean energy consumption of a new compact florescent light bulb. She takes a random sample of 41 bulbs and determines that the mean consumption is 1.3 watts per hour with a standard deviation of 0.7. When constructing a 97% confidence interval, which would be the most appropriate value of the critical value?

(A) 1.936
(B) 2.072
(C) 2.250
(D) 2.704
(E) 2.807

10. A 98 percent confidence interval for the mean of a large population is found to be 978 ± 25. Which of the following is true?

(A) 98 percent of all observations in the population fall between 953 and 1003
(B) The probability of randomly selecting an observation between 953 and 1003 from the population is 0.98
(C) If the true population mean is 950, then this sample mean of 978 would be unlikely to occur.
(D) If the true population mean is 990, then this sample mean of 978 would be unexpected.
(E) If the true population mean is 1006, then this confidence interval must have been calculated incorrectly.

Free Response Questions on Confidence Intervals

Free Response 1.

A random sample of 9^{th} grade math students was asked if they prefer working their math problems using a pencil or a pen. Of the 250 students surveyed, 100 preferred pencil and 150 preferred pen.

(a) Using the results of this survey, construct a 95 percent confidence interval for the proportion of 9^{th} grade students that prefer to work their math problems in pen.

(b) A school newspaper reported on the results of this survey by saying, "Over half of ninth-grade math students prefer to use pen on their math assignments." Is this statement supported by your confidence interval? Explain.

Free Response 2.

A major city in the United States has a large number of hotels. During peak travel times throughout the year, these hotels use a higher price for their rooms. A travel agent is interested in finding the difference of the average cost of a hotel rooms from the peak season to the off season. He takes a random sample of hotel room costs during each of these seasons. Plots of both samples of data indicate that the assumption of normality is not unreasonable.

Season	Cost	Standard Deviation	Sample Size
Peak Season	$245	$45	33
Off Season	$135	$65	38

(a) Construct a 95 percent confidence interval for the difference of the mean cost of hotel rooms from peak season to the off season.

(b) One particular hotel has an off season rate of $88 and a peak season rate of $218. Based on your confidence interval, comment on the price difference of this hotel.

Key to Confidence Interval Multiple Choice

1. A	Use the t-interval
2. C	Make a conservative assumption that the proportion is 0.5
3. D	The standard deviation of the population (batch) is not known.
4. D	2-prop-z-interval
5. C	Conditions of 1-proportion-z-interval
6. C	Decreasing the confidence level decreases the margin of error
7. D	Correct interpretation of the confidence level
8. E	Increasing the sample size by 9 decreases the error by one-third
9. C	Estimates from the t-table or invnorm(.985) produce this result.
10. C	Reasonable alternative interpretation of a confidence interval

Rubric for Confidence Interval Free Response

Free Response 1. **Solution**
Part (a)

1. Identify the population of interest and define the parameter of interest being estimated.

The population of interest is all ninth grade math students.
p = the population proportion of ninth grade math students that prefer to use pen on their math work.
$\hat{p} = \frac{150}{250} = 0.6$ = the sample proportion of ninth grade math students that prefer to use pen.
$n = 250$ is the sample size.

2. Identify the appropriate confidence interval by name or formula.

We will use a 95% confidence z-interval for proportions.

3. Verify any conditions (assumptions) that need to be met for that inference procedure.

The problem states that this is a simple random sample.

$10n < N$

$10(250) < N$

$2500 < N$

It is reasonable to assume that there are more than 2500 ninth grade math students.

$n\hat{p} \ge 10$	$n(1-\hat{p}) \ge 10$
$250(0.6) = 150$	$250(1-.6) = 100$
$150 \ge 10$	$100 \ge 10$

It is reasonable to use the normal approximation.

4. Calculate the confidence interval.

At a 95% CI, the critical value is $z^* = 1.96$.

$$0.6 \pm 1.96\sqrt{\frac{0.6(1-0.6)}{250}}$$

$$0.6 \pm 0.0607$$

$$(0.5393, 0.6607)$$

5. **Interpret your results in the context of the situation.**
We are 95% confident that the true proportion of ninth grade math students that prefer to do their math work with pen is between 0.5393 and 0.6607.

Part (b)
The school paper states that more than half of ninth grade math students prefer to do their math in pen. 50% or 0.5 is below our confidence interval. Because our confidence interval contains values that are above 0.5, we believe that the newspaper has made an accurate statement.

Scoring

Parts (a) and (b) are essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is correct if each of the following is included:
Identify the appropriate confidence interval by name or formula
Check appropriate conditions
Correct mechanics
Interpret the confidence interval in context

Part (a) is partially correct if only three of the above are provided in the answer.

Part (a) is incorrect if less than three of components are provided in the answer.

Part (b) is essentially correct if the student connects the paper's claim of 0.5 to the confidence interval and says that the interval supports the paper's claim.

Part (b) is partially correct if the student says that the confidence interval does support the paper's claim, but have weak or incomplete explanation.

Part (b) is incorrect if the student fails to provide an explanation.

4 Complete Response

Both parts essentially correct.

3 Substantial Response

One part essentially correct and one part partially correct

2 Developing Response

One part essentially correct and one part incorrect

OR

Partially correct on both parts

1 Minimal Response

Partially correct on one part.

Free Response 2. **Solution**

Part (a)

1. Identify the population of interest and define the parameters of interest being estimated.

The population of interest is price of hotel rooms in a major city.

μ_{peak} = the population mean cost of a hotel room in peak season.

μ_{off} = the population mean cost of a hotel room in off season.

$\bar{x}_{peak} = 245$ = the sample mean cost of a hotel room in peak season.

$\bar{x}_{off} = 135$ = the sample mean cost of a hotel room in off season.

$s_{peak} = 45$ = sample standard deviation cost of a hotel room in peak season.

$s_{off} = 65$ = sample standard deviation cost of a hotel room in off season.

$n_{peak} = 33$ is the sample size of the cost of a hotel room in peak season.

$n_{off} = 38$ is the sample size of the cost of a hotel room in off season.

df = 33 – 1 = 32 degrees of freedom (using the simplified approximation for df).

2. Identify the appropriate confidence interval by name or formula.

We will use a 95% confidence t-interval for the difference of means (2-sample-t-interval).

3. Verify any conditions (assumptions) that need to be met for that inference procedure.

The problem states that each is a simple random sample.
The problem states that the assumption of normality is reasonable.
The population standard deviation is unknown.

4. Calculate the confidence interval.

At a 95% CI, the critical value is $t^* = 2.042$ from the table using df = 30.
This value is found on the t-distribution table using 30 degrees of freedom (because the table does not have a value for 32).

$$(245-135) \pm 2.042\sqrt{\frac{45^2}{33}+\frac{65^2}{38}}$$

$$110 \pm 26.82$$

$$(83.18, 136.82)$$

5. Interpret your results in the context of the situation.

We are 95% confident that the true difference in mean cost of a hotel room in this city between peak season and off season falls between \$83.18 and \$136.82.

Note: The calculator gives a result of (\$83.77, \$136.23) for this confidence interval. df = 65.9

Part (b)

The difference in cost from peak season to off season is 218-88 = \$130

Because \$130 falls inside our confidence interval, we would not think that this is an extreme change in price from peak to off season.

Scoring

Parts (a) and (b) are essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is correct if each of the following is included:

Identify the appropriate confidence interval by name or formula
Check appropriate conditions
Correct mechanics
Interpret the confidence interval in context

Part (a) is partially correct if only three of the above are provided in the answer.

Part (a) is incorrect if less than three of components are provided in the answer.

Part (b) is essentially correct if the student compares the difference in prices to the confidence interval and concludes that this is not an unusual price change for this city.

Part (b) is partially correct if the student says that this in not an unusual price change but does not directly tie their conclusion to the confidence interval.

Part (b) is incorrect if fails to provide an explanation.

4 Complete Response

Both parts essentially correct.

3 Substantial Response

One part essentially correct and one part partially correct

2 Developing Response

One part essentially correct and one part incorrect

OR

Partially correct on both parts

1 Minimal Response
Partially correct on one part.

AP Statistics Exam Connections

The list below identifies free response questions that have been previously asked on the topic of Confidence Intervals. These questions are available from the CollegeBoard and can be downloaded free of charge from AP Central http://apcentral.collegeboard.com.

Free Response Questions	
2002 B	Question 4
2004	Question 6
2005	Question 5
2006	Question 4

AP* Statistics Review

Inference for Means and Errors

Teacher Packet

The t-distribution

- In order to conduct a hypothesis test or write a confidence interval for a population mean, a standard deviation must be known. Since it is not reasonable to know the standard deviation of the population, the standard deviation of the sample is used in its place. When William Gossett investigated the shape of the sampling distribution of the sample means using the sample standard deviation, he found more variability in the tails of the distribution than in a normal model. He also found that increasing the sample size decreased the variability of the distribution. From his work we have the student t-distribution to use as the model when the standard deviation of the population is not known.
- Properties of the t-distribution
  a. Has a mean of zero
  b. Is symmetric and bell-shaped
  c. Has more area in the tails than the standard normal distribution
  d. Has a different variance for each different degrees of freedom, where degrees of freedom = sample size minus 1. (df = n-1)
  e. The larger the degree of freedom, the more the graph resembles the standard normal distribution.

Hypothesis Test for a Population Mean

- Formulate the hypotheses to be tested.
  a. The Null Hypothesis (H_O) – a statement describing the believed value of a population mean.
  b. The Alternative Hypothesis (H_A) – an inequality describing the value of that same population mean. Usually this is the statement that the test is trying to provide evidence to support.

  Example: Suppose we wish to test whether a new soft drink "Refresh" contains the 355 ml as shown on the label. Some recently purchased cans seemed under filled. (And we wouldn't care if they were over filled.) The hypotheses for this test are: H_O: $\mu = 355\text{ ml}$ and H_A: $\mu < 355\text{ ml}$, where μ is the true mean content of the "Refresh" cans.

- Identify the appropriate test and verify that the conditions for that test are reasonably satisfied. The name of the test here is the <u>one-sample t-test for a mean</u>. The conditions that must be satisfied to use the test are that the data was obtained <u>randomly</u> from a <u>normal</u> population. The randomization condition is satisfied if a random sample was used to obtain the data or if random assignment was used in an experiment. The normal population condition may be difficult to satisfy. In practice, we rely on the sample size and the graph of the data. If the sample size is 15 or less, then the graph of the data can show no outliers or skewness. If the sample size is between 15 and 40, then the graph can show no outliers or extreme skewness. If the sample size is 40 or more, there are no restrictions on the data.

Example continued: Ideally, cans of "Refresh" should be selected at random from the population of all such cans. Due to the cost involved, a reasonable sample size should be selected and a graph of the data, like a boxplot, must be examined to check for skewness and outliers.

- Calculate the value of the test statistic and find the p-value. The test statistic for the t-test is $t_{\text{df}} = \dfrac{\bar{x} - \mu}{\dfrac{s}{\sqrt{n}}}$, where $\bar{x}$ is the sample mean and s is the sample standard deviation. The p-value is the probability of obtaining the sample statistic or something more extreme given that the null hypothesis is true. In other words it is a conditional probability.

Example continued: Suppose the mean of 15 randomly chosen "Refresh" cans is 352 ml with a standard deviation of 6 ml. Then the test statistic is $t_{14} = \dfrac{352 - 355}{\dfrac{6}{\sqrt{15}}} = -1.94$.
The p-value = $P(\bar{x} \le 352 \mid \mu = 355) = P(t \le -1.94) = 0.0366$. *There is a 3.66% chance of obtaining a mean of 352 ml or less if the true mean is 355.*

- Write a conclusion with two parts. The first part rejects or fails to reject the null hypothesis. The decision is made based on the p-value and the desired significance level (alpha level) of the test. If the p-value is less than the alpha level (α), the null hypothesis is rejected. If the p-value exceeds α, the null hypothesis is not rejected. If the values are equal, no decision is made. The second conclusion states the results in the context of the question. If the null hypothesis is rejected, there is evidence to support the alternative hypothesis. If the null hypothesis is not being rejected, there is no evidence to support the alternative hypothesis.

Example continued: If α =0 .05, then a p-value of 0.0366 is smaller than α, so H_O will be rejected. There is evidence that the mean amount of "Refresh" in the cans is less than the advertised 355 ml. If α = 0.01, then we fail to reject H_O. There is no evidence that the mean amount of "Refresh" in the cans is less than 355 ml.

Errors in Hypothesis Testing and Power

- Type I error – incorrectly rejecting a true null hypothesis. Making this type of error is similar to finding an innocent person guilty in a jury trial.
  $P(\text{Type I error}) = \alpha$, the significance level of the test.
- Type II error – failing to reject a false null hypothesis (incorrectly rejecting a true alternative hypothesis). Making this error is similar to finding a guilty person innocent in a jury trial.
- Power – the probability of correctly rejecting a false null hypothesis in favor of a particular true alternative.
  $P(\text{Type II error}) = \beta$

  Power = $1-\beta$

  The calculation of β and the power depend on the value of the true alternative.

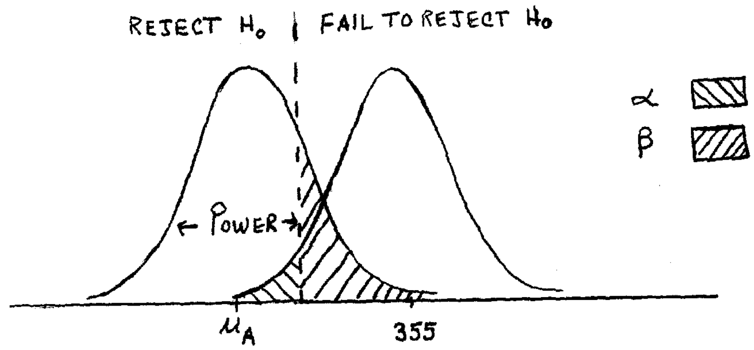
  The course description does not include these calculations but an investigative task question might incorporate this concept.

> *Example continued: A Type I error occurs if the results of the hypothesis test suggest that the "Refresh" cans are being under filled, but they really aren't. A possible consequence is that the company can be incorrectly accused of false advertising. A Type II error occurs if the results of the test suggest that the "Refresh" cans are not being under filled but they really are. A possible consequence is that the company continues to sell a product that is falsely advertised. Consumers get less than they are paying for. The power is the probability that the results of the test show that the true mean content of the cans is not 355 ml but is some smaller value.*

Increasing the Power of a Test

- Increase the sample size n
- Increase the probability of making a Type I error α
- Increase the distance between the hypothesized parameter and the true alternative value.

> *Example continued: The diagram below illustrates the relationship between the chances of making a Type I error, making a Type II error, and the power of the "Refresh" hypothesis test. If the sample size is increased the two curves will overlap less because the standard error will be smaller, thus increasing the power. If α is increased, then the decision line for the test will move right in this case and increase the power. Increasing the distance between 355 and the true alternative mean moves the left curve farther left, increasing the power.*



REJECT H_0
FAIL TO REJECT H_0
α
β
← POWER →
μ_A
355

Multiple Choice Questions on Inference for Means and Errors

Use the following to answer questions 1 and 2.
A city representative claims that the policemen in the city earn an average of $52,000 per year. The local paper believes that the mean salary is less for the beat cops. A survey conducted by the paper selected a random sample of 20 beat cops and found a mean salary of $51,300 with a standard deviation of $1900. Assume the normality for the population of the salaries of beat cops is reasonable so a t-test can be conducted using the data.

1. What is the correct null and alternative hypothesis?
   (A) $H_O: \mu = 51,300 \quad H_A: \mu > 51,300$
   (B) $H_O: \mu > 51,300 \quad H_A: \mu \geq 51,300$
   (C) $H_O: \mu = 52,000 \quad H_A: \mu = 51,300$
   (D) $H_O: \mu = 52,000 \quad H_A: \mu < 52,000$
   (E) $H_O: \mu < 52,000 \quad H_A: \mu \geq 52,000$

2. What are the value of the test statistic and the p-value of the test?
   (A) $t = -0.368, \quad p - value > .25$
   (B) $t = -1.65, \quad 0.01 < p - value < 0.05$
   (C) $t = -1.65, \quad 0.05 < p - value < 0.10$
   (D) $t = 1.65, \quad 0.05 < p - value < 0.10$
   (E) $t = .368, \quad p - value > .25$

3. Compare the standard normal distribution to the t-distribution. Which of the following statements are true?
   I. Both distributions have a mean of zero.
   II. Both distributions are symmetric and bell-shaped.
   III Both distributions have approximately 68% of the data within one standard deviation of the mean.

   (A) I only
   (B) I and II only
   (C) II only
   (D) III only
   (E) I, II and III

Use the following to answer questions 4 and 5.
A particular tire manufacturer recommends 32 pounds per square inch (psi) of pressure for its passenger car tires. Two independent car driving associations A and B wanted to conduct a two-tailed t-test of whether tire owners are really keeping the tires at the recommended pressure. Association A chooses a random sample of 18 owners. Association B chose a random sample of 30 owners. Plots of both samples showed no skewness or outliers. Surprisingly, both sets of data yielded a mean of 33 psi and standard deviation of 3 psi.

4. Which of the following hypotheses are used in the t-test by each association?
   (A) The null hypothesis is that the mean pressure on the tires is less than 32 psi.
   (B) The null hypothesis is that the mean pressure on the tires in not 32 psi.
   (C) The alternative hypothesis is that the mean tire pressure is greater than 32 psi.
   (D) The alternative hypothesis is that the mean tire pressure is less than 32 psi.
   (E) The alternative hypothesis is that the mean tire pressure is not 32 psi.

5. Which of the following is a true statement about the results of the tests at the 5% level of significance?
   (A) Neither test led to a rejection of the null hypothesis.
   (B) Both tests led to a rejection of the null hypothesis.
   (C) Only Association B's test led to a rejection of the null hypothesis.
   (D) Only Association A's test led to rejection of the null hypothesis.
   (E) Both tests had the same p-value.

6. Suppose a machine, that makes pegs to be used in holes to hold furniture parts together, is malfunctioning, but the manufacturer doesn't know it. A quality control test is conducted bimonthly with the null hypothesis stating that the machine works properly. The p-value of the most recent test was 0.185. What probably happens as a result of this test?
   (A) The test correctly fails to reject H_O.
   (B) The test correctly rejects H_O.
   (C) H_O is rejected, resulting in a Type I error.
   (D) H_O is not rejected, resulting in a Type I error.
   (E) H_O is not rejected, resulting in a Type II error.

7. Which sample size and significance level will give a test of highest power?
   (A) $n = 25, \quad \alpha = 0.01$
   (B) $n = 25, \quad \alpha = 0.05$
   (C) $n = 50, \quad \alpha = 0.10$
   (D) $n = 50, \quad \alpha = 0.05$
   (E) $n = 50, \quad \alpha = 0.01$

8. Which of the following statements are false concerning Type I and Type II errors?
   I. A Type I error is always worse than a Type II error
   II. The higher the probability of a Type I error the lower the probability of a Type II error.
   III. A Type I error incorrectly rejects a true alternative hypothesis.
   (A) I only
   (B) II only
   (C) III only
   (D) I and III only
   (E) II and III only

9. A t-test should not be used to conduct a hypothesis test for a mean if which of the following is true?
   (A) The sample size was only 20.
   (B) The standard deviation of the population was unknown.
   (C) The data was obtained by random sampling.
   (D) A histogram of the data was strongly skewed left.
   (E) A boxplot of the data showed no outliers.

10. Which of the following is the correct interpretation of a p-value of 0.003?
   (A) The probability of seeing the observed statistic or something more extreme, if the alternative hypothesis is true, is 0.003.
   (B) The probability of seeing the observed statistic or something more extreme, if the null hypothesis is true, is 0.003.
   (C) The probability of failing to reject the null hypothesis is 0.003.
   (D) The probability that the null hypothesis is true is 0.003.
   (E) The probability that the alternative hypothesis true is 0.003.

Free Response Questions on Inference for Means and Errors

Free Response 1.
An English professor at a local community college is disappointed that the average score on the final exam in her previous Freshman Survey courses has been 78. A friend who is a computer professor suggests that she set up a website containing tutorials created by a local actor and require students to log on for a weekly presentation. There would be no charge for the first semester of the program, but thereafter students would have to pay an extra $50 to enroll in the course. The English professor decides to give the tutorials a try for one semester for a random sample of her students.
(a) Define the parameter of interest and state the null and alternative hypothesis that the English professor is testing.

(b) In the context of the problem, describe a Type I error and the possible consequences to the students.

(c) In the context of the problem, describe a Type II error and the possible consequences to the students.

Free Response 2.
An AP Statistics student enjoys eating ChocoChewys, a new chewy candy bar with peanuts, caramel, and chocolate that comes in 1.45 oz. sealed packaging. He begins questioning the quality control of the candy maker when he gets a bar that is much smaller than he expected. He decides to buy 10 ChocoChewys at 10 randomly chosen grocery stores in his city and conduct a hypothesis test that he learned about in class. He weighs each bar with a digital scale that is accurate to a thousandth of an ounce.

ChocoChew Bar	1	2	3	4	5	6	7	8	9	10
Weight (oz.)	1.432	1.471	1.406	1.443	1.469	1.398	1.439	1.458	1.448	1.429

(a) Based on the data that he obtains, is there evidence that the candy maker needs to check his quality control process?

(b) Interpret the p-value in the context of the problem.

Free Response 3.

Joey, a soccer player, wants to improve his kicking accuracy so he can be the primary team member to do corner kicks. After watching David Beckham videos and practicing all summer, he tells his select soccer coach he can now get 75% of his corner kicks into the goal. He had only been 50% accurate the year before. The coach decides to let Joey take 12 practice corner kicks. If he can get at least 10 of them in the goal, the coach will make him the primary corner kicker.

(a) Suppose Joey really didn't improve – he still gets 50% of his corner kicks in the goal. If we assume that each of his kicks is independent of each other, what is the probability that he can make at least 10 goals on the 12 practice kicks?

(b) Suppose Joey didn't improve but does make at least 10 goals on the 12 kicks. If a hypothesis test associated with this situation has a significance level of 0.05, what type of error does the coach make? Explain

(c) If Joey really can make 75% of the goals and the coach insists on his making at least 10 out of the 12 goals to prove it, define the power associated with this hypothesis test and find its value.

(d) Based on the results in part c, is this considered a powerful test? If yes, explain why? If no, explain how to improve the power of the test?

Key to Inference for Means and Errors Multiple Choice

1. D $52,000 is the hypothesized mean. The alternative hypothesis asserts this value should be smaller.

2. C The test statistic = $\dfrac{51,300-52,000}{\frac{1900}{\sqrt{20}}}$ = -1.65. Using the t-test in the calculator gives a p-value of .0579.

3. B III is false. Less than 68% of the data is within one standard deviation of the mean for a t-distribution.

4. E These are two-tailed tests so the alternative hypothesis must state not equals.

5. A The p-value for Association A's test is 0.17. The p-value for Association B's test is 0.078. Both are greater than the given alpha.

6. E A p-value greater than any alpha results in a failure to reject H_O. H_O is false.

7. C The largest sample size and the largest alpha give the largest power.

8. D The situation related to a specific test determines whether a Type I or Type II error is worse. Increasing alpha decreases beta.

9. D Extreme skewness is an indicator of nonnormality of the population.

10. B The p-value describes the chance that the observed data belongs to the given distribution.

Rubric for Inference for Means and Errors Free Response

Free Response 1. **Solution**

Part (a)

$H_O : \mu = 78$
$H_A : \mu > 78$
where μ is the true mean final exam score for internet students

Part (b)

A Type I error results if the teacher believes that the final exam score has increased when it really hasn't. The following semester students will have to pay extra and waste time on an ineffectual program.

Part (c)

A Type II error results if the teacher believes that the final exam score has not increased when it really has. The following semester students will miss out on the opportunity to achieve a higher grade in the professor's class.

Scoring

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

Part (a) is *essentially correct* if it correctly defines the parameter AND gives the correct hypotheses.

Part (a) is *partially correct* if:
It correctly defines the parameter of interest or gives the correct hypotheses.

Part (b) is *essentially correct* if:
The Type I error is correctly described in context for the stated hypotheses AND the consequence is consistent with the Type I error described.

Part (b) is *partially correct* if:
The Type I error is correct but not in context and the consequence is consistent with the error described
OR
The Type I error is correct and in context but the consequence is inconsistent with the error described.
OR
The Type I error is backwards for the stated hypotheses but in context and the consequence is consistent with the error as described.

Part (c) is *essentially correct* if:
The Type II error is correctly described in context for the stated hypotheses and the consequence is consistent with the Type II error described.

Part (c) is *partially correct* if:
The Type II error is correct but not in context and the consequence is consistent with the error described
OR
The Type II error is correct and in context but the consequence is inconsistent with the error described.

If both errors are completely reversed, only one deduction is made for that error.

4 **Complete Response** (3E)
All three parts essentially correct

3 **Substantial Response** (2E 1P)
Two parts essentially correct and one part partially correct

2 **Developing Response** (2E 0P or 1E 2P or 3P)
Two parts essentially correct and no parts partially correct
OR
One part essentially correct and two parts partially correct
OR
Three parts partially correct

1 **Minimal Response** (1E 1P or 1E 0P or 2P)
No parts essentially correct and either zero or one parts partially correct
OR
No parts essentially correct and two parts partially correct

Free Response 2. **Solution**

Part (a)
Step 1: State a correct pair of hypotheses.

$H_O: \mu = 1.45$
$H_A: \mu < 1.45$ where μ = the true mean weight of all ChocoChewy bars

Step 2: Identify the correct test by name or formula and check appropriate conditions.

A one-sample t-test for a mean

OR $t = \dfrac{\overline{x} - \mu}{\dfrac{s}{\sqrt{n}}}$

Conditions: A random sample is taken from a normal population. The random sample is given and doesn't have to be repeated. The data needs to be plotted to establish that there is no obvious skewness or outliers in the data set and students need to note that the normal assumption is not unreasonable.

Step 3: Use correct mechanics, including the value of the test statistic, degrees of freedom and the p-value.

$$t = \frac{1.4384 - 1.45}{\dfrac{.0247}{\sqrt{10}}} - 1.483$$

$p - value = 0.0861$

$\text{df} = 9$

Step 4: Using the results of the statistical test, state a conclusion in the context of the problem.

Since the p-value = 0.0861 is greater than α = 0.05, then we cannot reject the null hypothesis that the mean weight is 1.45 oz. There is no evidence that the candy maker needs to improve their quality control process.

OR

Since the p-value =0.0861 is less than α = 0.10, then we can reject the null hypothesis that the mean weight is 1.45 oz. There is some evidence that the candy maker needs to improve their quality control process

If both an α and a p-value are given, the linkage is implied. If no α is given, the solution must explain how the conclusion follows from the p-value.

If the p-value is incorrect in step 3, but the conclusion is consistent with the computed p-value, step 4 is considered as correct.

Part (b)

The p-value = 0.0861 describes the probability that the sample mean of 1.438 oz or some smaller value could have occurred by chance given that the true mean weight of the candy is 1.45 oz.

OR

If the mean weight of the candy bars had not decreased, we could expect to find a weight of at most 1.438 ounces 8.61% of the time.

Scoring

Part (a) is scored as two parts. Step 1 and Step 2 are scored together and Step 3 and Step 4 are scored together. Each pair can be scored *essentially correct* (E), *partially correct* (P), or *incorrect* (I).

Step 1 and Step 2 are together *essentially correct* if both steps are correct.

Step 1 and Step 2 are together *partially correct* if one of the two parts is incorrect

Step 3 and Step 4 are together *essentially correct* if both steps are correct.

Step 3 and Step 4 are together *partially correct* if one of the two steps is incorrect.

Part (b) is *essentially correct* (E), *partially correct* (P), or *incorrect* (I).

Part (b) is *essentially correct* if the student correctly defines the p-value and uses context of the situation.

Part (b) is *partially correct* if the student correctly defines the p-value but does not use context.

4 **Complete Response** (3E)
All three parts essentially correct

3 **Substantial Response** (2E 1P)
Two parts essentially correct and one part partially correct

2 **Developing Response** (2E 0P or 1E 2P or 3P)
Two parts essentially correct and no parts partially correct
OR
One part essentially correct and two parts partially correct
OR
Three parts partially correct

1 **Minimal Response** (1E 1P or 1E 0P or 2P)
No parts essentially correct and either zero or one parts partially correct
OR
No parts essentially correct and two parts partially correct

Free Response 3. **Solution**

Part (a)

Let X be the number of goals made out of the 12 corner kicks. Then X has a binomial distribution with parameters $n = 12$ and $p = 0.5$.

$$P(X \geq 10) = \sum_{k=10}^{12} \binom{12}{k} (0.5)^{10} (0.50)^2 = 1 - \text{binomcdf}(12,0.50,9) = .0193$$

Part (b)

The coach makes a Type I error. He incorrectly concludes that Joey has improved his scoring ability, when he really hasn't.

Part (c)

The power is the probability that the coach correctly concludes from the test that Joey has improved to 75% accuracy. This will happen if Joey makes at least 10 goals on the 12 kicks. Power = $P(X \geq 10 \mid p = 0.75) = \sum_{k=10}^{12} \binom{12}{k} (0.75)^{10} (0.25)^2 = 0.3907$

Part (d)

No, this is a low powered test. It will be difficult for Joey to prove that he has improved. To increase the power of the test he should make more practice corner kicks or convince the coach to lower the number of goals needed to show improvement.

Scoring

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

Part (a) is *essentially correct* if the students identifies that a binomial probability is needed, identifies its parameters, and correctly finds the probability (except for minor arithmetic errors)

Part (a) is *partially correct* if:

The student indicates that a binomial probability is needed, identifies the parameters, but finds the probability that X = 10.

OR

The student correctly uses the formula for the binomial probability by substituting in numbers but never identifies the problem as a binomial situation.

Part (a) is incorrect if a correct probability is given but no connection is made to a binomial distribution.

Part (b) is *essentially correct* if the type of error is correctly identified and explained in the context of the problem.

Part (b) is *partially correct* if the type of error is correctly identified but the student gives an explanation with no context.

Part (b) is *incorrect* if the type of error is correct but no explanation is given or if the type of error is incorrect.

Part (c) is *essentially correct* if the power is correctly defined in context and the correct value is found (except for minor arithmetic errors). The student does not have to describe it as a binomial probability.

Part (c) is *partially correct* if:
The student correctly defines the power in context but does not correctly calculate its value
OR
The student correctly defines the power without context and calculates its value correctly.

Part (d) is *essentially correct* if the student considers that the test is not powerful and explains at least one of the given ways to improve the power.

Part (d) is *partially correct* if the student considers the test is not powerful and has a weak explanation in context to support his answer.

Part (d) is *incorrect* if:
The student considers the test not to be powerful but offers no explanation
OR
The student considers the test to be powerful.

Each essentially correct response is worth 1 point; each partially correct response is worth half a point.

4 Complete Response

3 Substantial Response

2 Developing Response

1 Minimal Response

If a response is between two scores (for example, 2.5 points), use a holistic approach to determine whether to score up or down depending on the strength of the response and communication.

AP Statistics Exam Connections

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

http://apcentral.collegeboard.com.

Free Response Questions	
1999	Question 6 (a)
2003	Question 2
2005 B	Question 6