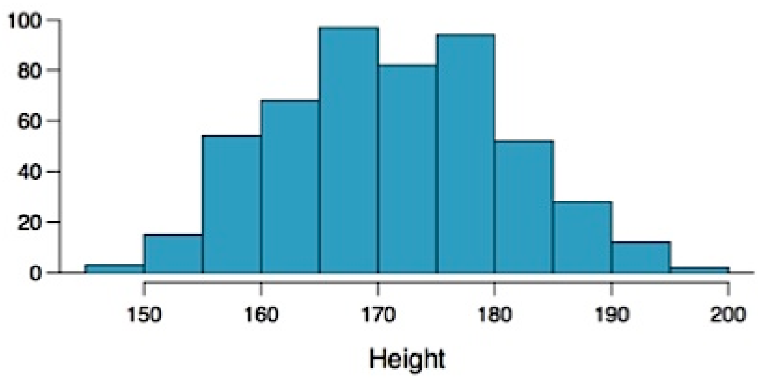
Feedback — Unit 3 Quiz - Foundations for inference

Help

You submitted this quiz on **Sat 8 Mar 2014 3:40 AM PST**. You got a score of **8.00** out of **15.00**. You can attempt again, if you'd like.

Question 1

Researchers studying anthropometry collected body girth measurements and skeletal diameter measurements, as well as age, weight, height and gender, for 507 physically active individuals. The histogram below shows the sample distribution of heights in centimeters, and the table shows sample statistics calculated based on this sample. Which of the following is **not necessarily true**?



Min	147.2
Q1	163.8
Median	170.3
Mean	171.1
SD	9.4
Q3	177.8
Max	198.1

Your Answer		Score	Explanation
○ The sample mean is 171.1 cm.			
○ The point estimate for the population mean is 171.1 cm.			
◉ The population mean is 171.1 cm.	✔	1.00	While the sample statistics provided in the table are point estimates for the unknown parameters, it's necessarily not true that the true population parameters will be exactly equal to these values.
○ The sample median is 170.3 cm.			
Total		1.00 / 1.00	

Question Explanation

This question refers to the following learning objective(s): Define sample statistic as a point estimate for a population parameter, for example, the sample mean is used to estimate the population mean, and note that point estimate and sample statistic are synonymous.

Question 2

Which of the following is **false**?

Your Answer		Score	Explanation
Standard error computed based on a sample standard deviation will always be lower than the standard deviation of that sample.			
In order to reduce the standard error by half, sample size should be doubled.	✔	1.00	Since n is underneath a square root in the denominator of the formula for standard error. Because of the square root, to reduce the standard error by half you would actually need a sample size four times (2^2) as high.
As the sample size increases, the variability of the sampling distribution decreases.			
Standard error measures the variability in means of samples of the same size taken from the same population.			
Total		1.00 / 1.00	

Question Explanation

This question refers to the following learning objective(s): Calculate the sampling variability of the mean, the standard error, as $SE = \sigma/\sqrt{n}$.

Question 3

Students are asked to count the number of chocolate chips in 22 cookies for a class activity. They found that the cookies on average had 14.77 chocolate chips with a standard deviation of 4.37 chocolate chips. After collecting the data, a student reports the standard error of the mean to be 0.93 chocolate chips. What is the **best** way to interpret the student's result?

Your Answer		Score	Explanation
0.93 chocolate chips is a measure of the variability in the mean number of chocolate chips across all chocolate chip cookies.			
0.93 is the standard deviation of the number of chocolate chips in a chocolate chip cookie.			
0.93 chocolate chips is a measure of the variability we'd expect in calculations of the mean number of chocolate chips if we took repeated random samples of 22 cookies.	✔	1.00	
The student either made a calculation error or his result is meaningless, because it does not make sense to talk about 0.93 chocolate chips.			
Total		1.00 / 1.00	

Question Explanation

This question refers to the following learning objective(s): Distinguish standard deviation (σ or s) and standard error (SE): standard deviation measures the variability in the data, while standard error measures the variability in point estimates from different samples of the same size and from the same population, i.e. measures the sampling variability.

Question 4

Which of the following is false about the central limit theorem (CLT)?

Your Answer		Score	Explanation
As the sample size increases, the sampling distribution of the mean is more likely to be nearly normal, regardless of the shape of the original population distribution.			
If the population distribution is normal, the sampling distribution of the mean will also be nearly normal, regardless of the sample size.	✖	0.00	Review the associated learning objective.
The CLT states that the sampling distribution will be centered at the true population parameter.			
If we take more samples from the original population, the sampling distribution is more likely to be nearly normal.			

Total	0.00 / 1.00

Question Explanation

This question refers to the following learning objective(s):

Recognize that the Central Limit Theorem (CLT) is about the distribution of point estimates, and that given certain conditions, this distribution will be nearly normal.

- In the case of the mean the CLT tells us that if
  (1a) the sample size is sufficiently large $(n \geq 30)$ and the data are not extremely skewed or
  (1b) the population is known to have a normal distribution, and
  (2) the observations in the sample are independent,
  then the distribution of the sample mean will be nearly normal, centered at the true population mean and with a standard error of $\frac{\sigma}{\sqrt{n}}$.

$$\bar{x} \sim N\left(mean = \mu, SE = \frac{\sigma}{\sqrt{n}}\right)$$

- When the population distribution is unknown, condition (1a) can be checked using a histogram or some other visualization of the distribution of the observed data in the sample.
- The larger the sample size (n), the less important the shape of the distribution becomes, i.e. when n is very large the sampling distribution will be nearly normal regardless of the shape of the population distribution.

Question 5

To get an estimate of consumer spending in the U.S. following the Thanksgiving holiday, 436 randomly sampled American adults were surveyed. Their daily spending for the six-day period following Thanksgiving averaged $84.71. A 95% confidence interval based on this sample is ($80.31, $89.11). Which of the following are true?

I. We are 95% confident that the average spending of the 436 American adults in this sample is between $80.31 and $89.11.
II. If we collected many random samples of the same size and calculated a confidence interval for daily spending for each sample, then we would expect 95% of the intervals to contain the true population parameter.
III. We are 95% confident that the average spending of all American adults is between $80.31 and $89.11.

Your Answer		Score	Explanation
◉ I and II	✖	0.00	I is wrong because the confidence interval is not about the sample mean, indeed we're 100% confident that the sample mean is in the interval since the interval is built around the sample mean.
○ I and			

III

II and III

I, II, and III

None

Total 0.00 / 1.00

Question Explanation

This question refers to the following learning objective(s):

- Interpret a confidence interval as "We are XX% confident that the true population parameter is in this interval", where XX% is the desired confidence level.
- Define margin of error as the distance required to travel in either direction away from the point estimate when constructing a confidence interval.

Question 6

A study suggests that the average college student spends 2 hours per week communicating with others online. You believe that this is an underestimate and decide to collect your own sample for a hypothesis test. You randomly sample 60 students from your dorm and find that on average they spent 3.5 hours a week communicating with others online. Which of the following is the correct set of hypotheses for this scenario?

Your Answer		Score	Explanation
$H_0 : \mu = 2$ $H_A : \mu > 2$	✔	1.00	
$H_0 : \mu = 2$ $H_A : \mu < 2$			
$H_0 : \bar{x} = 2$ $H_A : \bar{x} < 2$			
$H_0 : \mu = 3.5$ $H_A : \mu < 3.5$			
$H_0 : \bar{x} = 2$ $H_A : \bar{x} > 2$			
Total		1.00 / 1.00	

Question Explanation

This question refers to the following learning objective(s):

- Always construct hypotheses about population parameters (e.g. population mean, μ) and not the sample statistics (e.g. sample mean, $\bar{x}$). Note that the population parameter is unknown while the sample statistic is measured using the observed data and hence there is no point in hypothesizing about it.
- Define the null value as the value the parameter is set to equal in the null hypothesis.
- Note that the alternative hypothesis might be one-sided (μ the null value) or two-sided ($\mu \neq$ the null value), and the choice depends on the research question.

Question 7

Which of the following is the correct definition of the p-value?

Your Answer		Score	Explanation
P(H0 true)			
P(observed or more extreme sample statistic | H_0 true)			
P(H0 true | observed data)	✖	0.00	Review the associated learning objective. This sounds more like the posterior probability: P(hypothesis | data).
P(H0 true | HA false)			
Total		0.00 / 1.00	

Question Explanation

This question refers to the following learning objective(s): Define a p-value as the conditional probability of obtaining a sample statistic at least as extreme as the one observed given that the null hypothesis is true.

p-value = P(observed or more extreme sample statistic | H_0 true)

Question 8

Suppose we collected a sample of size n = 100 from some population and used the data to calculate a 95% confidence interval for the population mean. Now suppose we are going to increase the sample size to n = 300. Keeping all else constant, which of the following would we expect to occur as a result of increasing the sample size?

I. The standard error would decrease.
II. Width of the 95% confidence interval would increase.

III. The margin of error would decrease.

Your Answer		Score	Explanation
◯ II and III			
◯ I and III			
◉ I and II	✖	0.00	Increasing the sample size (while keeping the same confidence level) will decrease the standard error, which decreases the margin of error and hence the width of intervals.
◯ I, II, and III			
◯ None			
Total		0.00 / 1.00	

Question Explanation

This question refers to the following learning objective(s):

- Recognize that when the sample size increases we would expect the sampling variability to decrease.
- Define margin of error as the distance required to travel in either direction away from the point estimate when constructing a confidence interval, i.e. $z^{\star} \times SE$.

Question 9

One-sided alternative hypotheses are phrased in terms of:

Your Answer		Score	Explanation
◯ ≠			
◯ ≈ or =			
◉ < or >	✔	1.00	
◯ ≤ or ≥			
Total		1.00 / 1.00	

Question Explanation

This question refers to the following learning objective(s): Note that the alternative hypothesis might be one-sided (μ the null value) or two-sided ($\mu \neq$ the null value), and the choice depends on the research question.

Question 10

A Type 2 error occurs when the null hypothesis is

Your Answer		Score	Explanation
not rejected when it is false	✔	1.00	
rejected when it is false			
not rejected when it is true			
rejected when it is true			
Total		1.00 / 1.00	

Question Explanation

This question refers to the following learning objective(s): Note that the conclusion of a hypothesis test might be erroneous regardless of the decision we make.

- Define a Type 1 error as rejecting the null hypothesis when the null hypothesis is actually true.
- Define a Type 2 error as failing to reject the null hypothesis when the alternative hypothesis is actually true.

Question 11

A statistician is studying blood pressure levels of Italians in the age range 75-80. The following is some information about her study:

I. The data were collected by responses to a survey conducted by email, and no measures were taken to get information from those who did not respond to the initial survey email.
II. The sample observations only make up about 4% of the population.
III. The sample size is 2,047.
IV. The distribution of sample observations is skewed - the skew is easy to see, although not very extreme.

The researcher is ready to use the Central Limit Theorem (CLT) in the main part of her analysis. Which aspect of the her study is most likely to prevent her from using the CLT?

Your Answer		Score	Explanation
(II), because she only has data from a small proportion of the whole population.			
(III), because the sample size is	✖	0.00	Recall that having a sample of less than 10%

too small compared to all Italians in the age range 75-80.		of the population is a necessary condition for use of the CLT.
(IV), because there is some skew in the sample distribution.		
(I), because the sample may not be random and hence observations may not be independent.		
Total	0.00 / 1.00	

Question Explanation

This question refers to the following learning objective(s):

- Recognize that the Central Limit Theorem (CLT) is about the distribution of point estimates, and that given certain conditions, this distribution will be nearly normal. In the case of the mean the CLT tells us that if (1a) the sample size is sufficiently large (n ≥ 30 or larger if the data are considerably skewed), or (1b) the population is known to have a normal distribution, and (2) the observations in the sample are independent, then the distribution of the sample mean will be nearly normal, centered at the true population mean and with a standard error of $\frac{\sigma}{\sqrt{n}}$:

$$\bar{x} \sim N\left(mean = \mu, SE = \frac{\sigma}{\sqrt{n}}\right)$$

  When the population distribution is unknown, condition (1a) can be checked using a histogram or some other visualization of the distribution of the observed data in the sample. The larger the sample size (n), the less important the shape of the distribution becomes, i.e. when n is very large the sampling distribution will be nearly normal regardless of the shape of the population distribution.
- If the conditions necessary for the CLT to hold are not met, note this and do not go forward with the analysis. (We will later learn about methods to use in these situations.)

Question 12

SAT scores are distributed with a mean of 1,500 and a standard deviation of 300. You are interested in estimating the average SAT score of first year students at your college. If you would like to limit the margin of error of your 95% confidence interval to 25 points, at least how many students should you sample?

Your Answer	Score	Explanation
13,830		
393		

554 ✔	1.00	$ME = z^\star \frac{s}{\sqrt{n}} \rightarrow 25 = 1.96\,\frac{300}{\sqrt{n}} \rightarrow n = \frac{1.96^2 \times 300^2}{25^2} \rightarrow n = 553.1904 \rightarrow$ n should be at least 554, since rounding down would result in a slightly larger margin of error than we desire.
392		
553		
Total	1.00 / 1.00	

Question Explanation

This question refers to the following learning objective(s): Calculate the required sample size to obtain a given margin of error at a given confidence level by working backwards from the given margin of error.

Question 13

If it's relatively riskier to reject the null hypothesis when it might be true, should a smaller or a larger significance level be used?

Your Answer	Score	Explanation
smaller ✔	1.00	If it's relatively riskier to reject the null hypothesis when it might be true, that means it's relatively riskier to make a Type 1 error, therefore we should decrease the probability of making a Type 1 error, which means decreasing the significance level.
larger		
Total	1.00 / 1.00	

Question Explanation

This question refers to the following learning objective(s): Note that the probability of making a Type 1 error is equivalent to the significance level when the null hypothesis is true, and choose a significance level depending on the risks associated with Type 1 and Type 2 errors.

- Use a smaller α if Type 1 error is relatively riskier.
- Use a larger α if Type 2 error is relatively riskier.

Question 14

The nutrition label on a bag of potato chips says that a one ounce (28 gram) serving of potato chips has 130 calories and contains ten grams of fat, with three grams of saturated fat. A random sample of 35 bags yielded a sample mean of 134 calories with a standard deviation of 17 calories. We are evaluating whether these data provide convincing evidence that the nutrition label does not provide an accurate measure of calories in the bags of potato chips at the 10% significance level. Which of the following is correct?

Your Answer		Score	Explanation
◯ The p-value is approximately 16%, which means we should reject the null hypothesis and determine that these data provide convincing evidence the nutrition label does not provide an accurate measure of calories in the bags of potato chips.			
◯ The p-value is approximately 16%, which means we should fail to reject the null hypothesis and determine that these data **do not** provide convincing evidence the nutrition label does not provide an accurate measure of calories in the bags of potato chips.			
◉ The p-value is approximately 8%, which means we should reject the null hypothesis and determine that these data provide convincing evidence the nutrition label does not provide an accurate measure of calories in the bags of potato chips.	✖	0.00	Note that the alternative hypothesis is two-sided: $$H_0 : \mu = 130; H_A : \mu \neq 130$$ $$Z = \frac{134 - 130}{\frac{17}{\sqrt{35}}} = 1.39$$ $$p - value = P(\bar{x}134 \mid \mu = 130) = P(z1.39) = 2 \times 0.0823 = 0.1646$$ Since p-value > 10%, fail to reject H_0.
◯ The p-value is approximately 8%, which means we should fail to reject the null hypothesis and determine that these data **do not** provide convincing evidence the nutrition label does not provide an accurate measure of calories in the bags of potato chips.			
Total		0.00 / 1.00	

Question Explanation

This question refers to the following learning objective(s): Calculate a p-value as the area under the normal curve beyond the observed sample mean (either in one tail or both, depending on the alternative hypothesis). Note that in doing so you can use a Z score, where

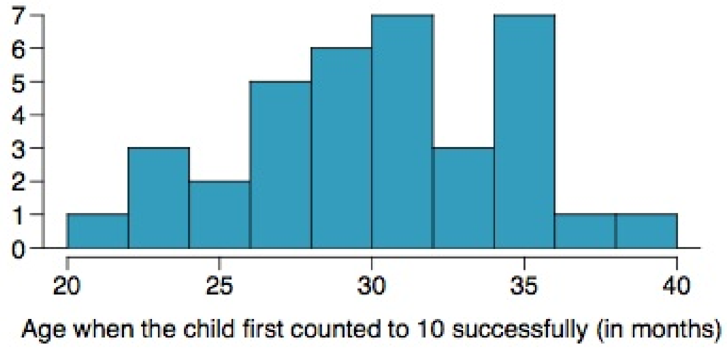
$$Z = \frac{sample\ statistic - null\ value}{SE} = \frac{\bar{x} - \mu_0}{SE}$$

Always sketch the normal curve when calculating the p-value, and shade the appropriate area(s) depending on whether the alternative hypothesis is one- or two-sided.

Question 15

Researchers investigating characteristics of gifted children collected data from schools in a large city on a random sample of thirty-six children who were identified as gifted children soon after they reached the age of four. The following histogram shows the distribution of the ages (in months) at which these children first counted to 10 successfully. Also provided are some sample statistics.

Calculate a 90% confidence interval for the average age at which gifted children first count to 10 successfully. Choose the closest answer.



n	36
min	21
mean	30.69
sd	4.31
max	39

Your Answer		Score	Explanation
(30.49, 30.89)			
(30.12, 31.26)			
(29.50, 31.88)			
(29.28, 32.10)	✖	0.00	Note that the $z^\star$ for the 90% confidence level is 1.65 (not 1.96).
Total		0.00 / 1.00	

Question Explanation

This question refers to the following learning objective(s): Recognize that the nearly normal distribution of the point estimate (as suggested by the CLT) implies that a confidence interval can be calculated as

$$point\ estimate \pm z^{\star} \times SE,$$

where $z^{\star}$ corresponds to the cutoff points in the standard normal distribution to capture the middle XX\% of the data, where XX\% is the desired confidence level.

- For means this is: $\bar{x} \pm z^{\star} \frac{s}{\sqrt{n}}$
- Note that $z^{\star}$ is always positive.