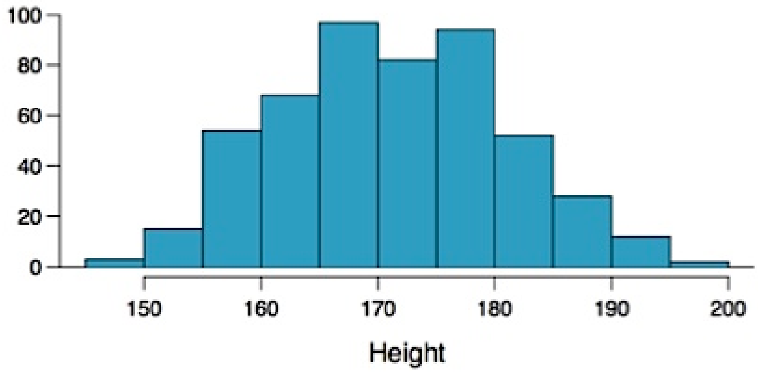
Feedback — Unit 3 Quiz - Foundations for inference

Help

You submitted this quiz on **Mon 17 Mar 2014 6:43 AM PDT**. You got a score of **11.00** 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

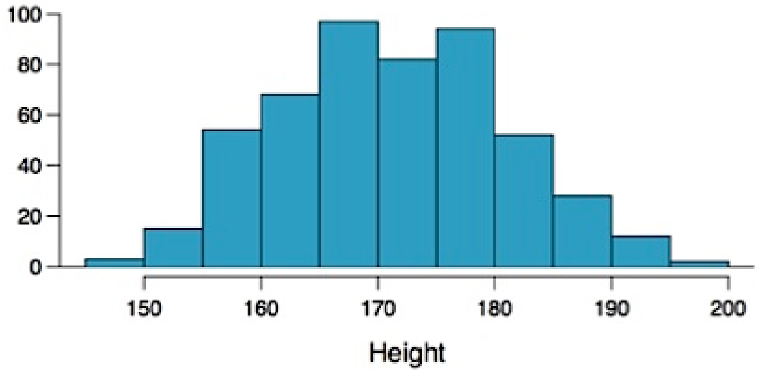
Your Answer		Score	Explanation
The sample median is 170.3 cm.			
The point estimate for the population mean is 171.1 cm.			
The sample mean is 171.1 cm.			
The population mean is 171.1 cm.	✔	1.00	While the sample statistics provided in the table are point estimates for the unknown parameters, it's necessarily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

Researchers studying anthropometry collected various body and skeletal measurements for 507 physically active individuals. The histogram below shows the sample distribution of heights in centimeters. If the 507 individuals are a simple random sample - and let's assume they are - then the sample mean is a point estimate for the mean height of all active individuals. What measure do we use to quantify the variability of such an estimate? Compute this quantity using the data from this sample and choose the **best** answer below.



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
standard deviation = 0.417			
standard deviation = 0.019			
standard error = 0.417			
mean squared error = 0.105			
standard error = 0.019	✖	0.00	We quantify variability in the sample mean by calculating the **standard error** (of the mean) $SE = \sigma/\sqrt{n}$.
Total		0.00 / 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

The ages of pennies at a particular bank follow a nearly normal distribution with mean 10.44 years with standard deviation 9.2 years. Say you take random samples of 30 pennies, find the mean age in each sample, and plot the distribution of these means. Which of the following are the **best** estimates for the center and spread of this distribution?

Your Answer		Score	Explanation
mean = 10.44, standard error = 9.2/30 = 0.31			
mean = 10.44/30 = 0.348, standard error = $(9.2/30)^2 = 0.094$			
mean = 10.44, standard error = 9.2			
mean = 10.44, standard error = $9.2/\sqrt{30} = 1.68$	✔	1.00	According to the CLT: $\bar{x} \sim N\left(mean = \mu, SE = \frac{\sigma}{\sqrt{n}}\right)$
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.			
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.	✔	1.00	Taking more samples from the original population (with each sample of the same fixed size n), does not ensure that the sampling distribution will be more likely to be nearly normal. However, if we let the **size** n of each sample increase, then the sampling distribution is more likely to be nearly normal. So the n of interest is not the number of samples but instead the sample size.
Total		1.00 / 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

A random sample of 100 runners who completed the 2012 Cherry Blossom 10 mile run yielded an average completion time of 95 minutes. A 95% confidence interval calculated based on this sample is 92 minutes to 98 minutes. Which of the following is false based on this confidence interval?

Your Answer		Score	Explanation
◉ 95% of the time the true average finishing time of all runners who completed the 2012 Cherry Blossom 10 mile run is between 92 minutes and 98 minutes.	✔	1.00	The ``true average finishing time" is a fixed number whose exact value we do not know. So it does not make sense to talk about the true average finishing time being between two numbers ``95\% of the time".
○ We are 95% confident that the true average finishing time of all runners who completed the 2012 Cherry Blossom 10 mile run is between 92 minutes and 98 minutes.			
○ Based on this 95% confidence interval, we would reject a null			

hypothesis stating that the true average finishing time of all runners who completed the 2012 Cherry Blossom 10 mile run is 90 minutes.

The margin of error of this confidence interval is 3 minutes.

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

Your friend likes to show off to his coworkers using statistical terminology, but he makes errors so much that you often have to correct him. He just completed the following hypothesis test:

$H_0 : \mu = 100$; $H_A : \mu \neq 100$

$\bar{x} = 105$, $s = 10$, $n = 40$

p-value = 0.0016

He claims the definition of this p-value is “the probability of obtaining a sample mean of 105 from a random sample of n = 40 when the true population mean is assumed to be 100.” Which of the following is true? (You may assume his calculations are correct, only focus on his interpretation.)

Your Answer		Score	Explanation
Your friend is wrong, the statement should be revised as “the probability of obtaining a sample mean of 105 from a random sample of n = 40 when the true population mean is assumed to be different than 105.”			
Your friend is wrong, the sample size is irrelevant.	✖	0.00	Your friend is wrong, but this is not the reason why. Review the associated learning objective.
Your friend is wrong, the statement should be revised as “the probability of obtaining a sample mean of 105 or more extreme from a random sample of n = 40 when the true population mean is assumed to be 100."			
Your friend is right.			

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

Which of the following is **false** about confidence intervals? All else held constant,

Your Answer		Score	Explanation
as the standard deviation of the sample increases, the width increases.	✖	0.00	As the standard deviation of the sample increases (all else held constant), the width of the confidence interval will increase because the data is more variable, the standard error increases, which in turn increases the margin of error and the width of the confidence interval.
as the sample size increases, the margin of error decreases.			
as the confidence level increases, the width decreases.			
as the sample mean increases, the margin of error stays constant.			
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

A researcher found a 2006 - 2010 survey showing that the average age of women at first marriage is 23.44. Suppose a researcher believes that this value may have increased more recently, but as a good scientist he also wants to consider the possibility that the average age may have decreased. The researcher has set up his hypothesis test; which of the following states the appropriate H_A correctly?

Your Answer		Score	Explanation
$H_A : \mu = 23.44$ years old.			
$H_A : \mu \neq 23.44$ years old.	✔	1.00	Because the researcher is interested in both an increase or a decrease, H_A should be two-sided.
$H_A : \mu < 23.44$ years old.			
$H_A : \mu > 23.44$ years old.			
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 1 error occurs when the null hypothesis is

Your Answer		Score	Explanation
rejected when it is false			
rejected when it is true	✔	1.00	
not rejected when it is false			
not 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
(III), because the sample size is too small compared to all Italians in the age range 75-80.			
(II), because she only has data from a small proportion of the whole population.			
(I), because the sample may not be random and hence observations may not be independent.	✔	1.00	The correct answer is that the data arose as a result of an email survey. This data collection would likely result in a sample which is not a simple random sample of Italians aged 75-80, which would violate the independence of observations condition necessary for the CLT.
(IV), because there is some skew in the sample distribution.			

Total	1.00 / 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 ($n \geq 30$ 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			
553			
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			
393			

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

True / False: Decreasing the significance level (α) will increase the probability of making a Type 1 error.

Your Answer		Score	Explanation
False			
True	✖	0.00	Note that the probability of making a Type 1 error is equivalent to the significance level when the null hypothesis is true.
Total		0.00 / 1.00	

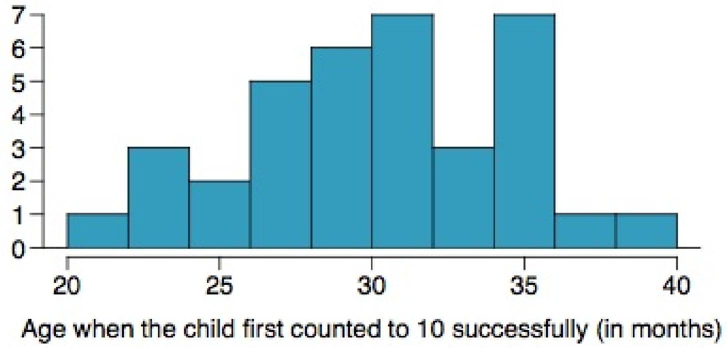
Question Explanation

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

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. Suppose you read online that children first count to 10 successfully when they are 32 months old, on average. You perform a hypothesis test evaluating whether the average age at which gifted children first count to 10 is different than the general average of 32 months. What is the p-value of the hypothesis test? Choose the closest answer.



n	36
min	21
mean	30.69
sd	4.31
max	39

Your Answer		Score	Explanation
0.0688	✔	1.00	$$H_0 : \mu = 32, H_A : \mu \neq 32$$ $$Z = \frac{30.69 - 32}{\frac{4.31}{\sqrt{36}}} = -1.82$$ $$p - value = P(\bar{x}1.82)$$ $$= 2 \times 0.0344 = 0.0688$$
0.0344			
0.7183			
0.9656			
0.9312			
Total		1.00 / 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

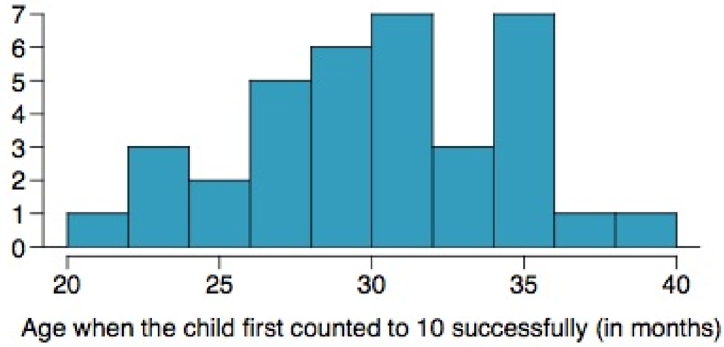
$$Z = \frac{sample\ statistic - null\ value}{SE} = \frac{\bar{x} - \mu_0}{SE}$$

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
(30.12, 31.26)			
(29.28, 32.10)			
(30.49, 30.89)			
(29.50, 31.88)	✔	1.00	The 90% confidence interval can be calculated as follows: $\bar{x} \pm z^\star se(\bar{x}) = 30.69 \pm 1.65 \times \frac{4.31}{\sqrt{36}} = 30.69 \pm 1.19 = (29.50, 31.88)$
Total		1.00 / 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.