

## Homework 3

## Part 1 of 1 -

## Question 1 of 23

2.0 Points

The deciles of any distribution are the nine values that divide the sorted data into ten equal parts, so that each part represents 10% of the sample or population. How many standard deviations from the mean are the deciles that mark off the highest and lowest 10% of a normal distribution?



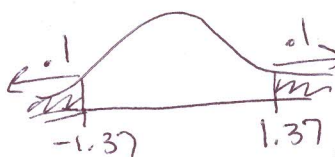
STATA:

$$d: \text{invnormal}(0.1) = 1.28$$

## Question 2 of 23

2.0 Points

The deciles of any distribution are the nine values that divide the sorted data into ten equal parts, so that each part represents 10% of the sample or population. How many standard deviations from the mean are the deciles that mark off the highest and lowest 10% of a t distribution with 10 degrees of freedom?



$$d: \text{invttail}(10, 0.1) = 1.37$$

## Question 3 of 23

2.0 Points

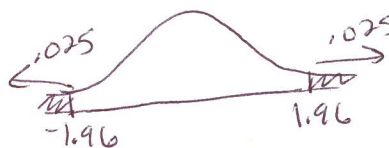
The deciles of any distribution are the nine values that divide the sorted data into ten equal parts, so that each part represents 10% of the sample or population. How many standard deviations from the mean are the deciles that mark off the highest and lowest 10% of a t distribution with 100 degrees of freedom?



## Question 4 of 23

2.0 Points

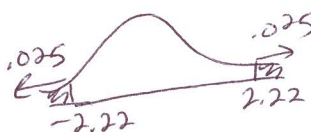
The deciles of any distribution are the nine values that divide the sorted data into ten equal parts, so that each part represents 10% of the sample or population. How many standard deviations from the mean are the deciles that mark off the highest and lowest 2.5% of a normal distribution?



## Question 5 of 23

2.0 Points

The deciles of any distribution are the nine values that divide the sorted data into ten equal parts, so that each part represents 10% of the sample or population. How many standard deviations from the mean are the deciles that mark off the highest and lowest 2.5% of a t distribution with 10 degrees of freedom?

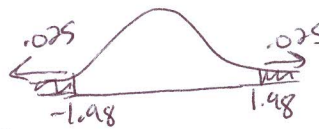


## Question 6 of 23

2.0 Points

The deciles of any distribution are the nine values that divide the sorted data into ten equal parts, so that each part represents 10% of the sample or population. How many standard deviations from

the mean are the deciles that mark off the highest and lowest 2.5% of a t distribution with 100 degrees of freedom?



Question 7 of 23

5.0 Points

Using an online calculator, you find that your 3 year old daughter has a BMI of 13.9 and are concerned that she may be underweight. Assuming the distribution of BMI of 3 year old girls is normal with mean 15.4 and standard deviation 1.2, the probability that a three year old girl has BMI less than or equal to 13.9 is .

$$Z = \frac{x - \mu}{\sigma} = \frac{13.9 - 15.4}{1.2} = -1.25$$



$$\text{di normal}(-1.25) \approx \boxed{.11}$$

Question 8 of 23

10.0 Points

At a routine physical examination, you learn that your 3 year old son has a BMI of 18.5 and are shocked to hear the doctor warn you that he is overweight. Assuming the BMI of 3 year old boys is normal with mean 15.6 and standard deviation 1.3, explain why the doctor considers your son to be overweight.

$$Z = \frac{x - \mu}{\sigma} = \frac{18.5 - 15.6}{1.3} \approx 2.23$$

$$\text{Find } P(Z \geq 2.23)$$

$$\text{di } 1 - \text{normal}(2.23) \approx .0129$$

(Maximum number of characters: 60000)

Show/Hide Rich-Text Editor

One can see that only a ~~small~~ small proportion of boys have BMI 18.5 or greater. [Also, this value (2.23) is greater than 1.96 and so the child's BMI is more than 2 standard deviations above the mean.]

Question 9 of 23

5.0 Points

The 2 hour oral glucose tolerance test is often used to classify individuals into three groups: "healthy" (glucose level < 140 mg/dL), diabetic (glucose level 200 mg/dL or higher), or impaired glucose tolerance without diabetes (glucose levels in between). Suppose you are studying a population in which the distribution of glucose levels is normal with mean 100 mg/dL and standard deviation 30 mg/dL. The probability that a randomly sampled person from your population has a healthy glucose level is .

$$\begin{aligned} X &= \text{glucose level} \\ \text{want } P(X < 140) \\ &= P\left(Z < \frac{140 - 100}{30}\right) \approx .91 \end{aligned}$$

Question 10 of 23

5.0 Points

The 2 hour oral glucose tolerance test is often used to classify individuals into three groups: "healthy" (glucose level < 140 mg/dL), diabetic (glucose level 200 mg/dL or higher), or impaired glucose tolerance without diabetes (glucose levels in between). Suppose you are studying a population in which the distribution of glucose levels is normal with mean 100 mg/dL and standard deviation 30 mg/dL. The probability that a randomly sampled person from this population is diabetic is .

$$\begin{aligned} P(X > 200) \\ &= P\left(Z > \frac{200 - 100}{30}\right) \approx .0004 \end{aligned}$$

## Question 11 of 23

5.0 Points

The 2 hour oral glucose tolerance test is often used to classify individuals into three groups: "healthy" (glucose level < 140 mg/dL), diabetic (glucose level 200 mg/dL or higher), or impaired glucose tolerance without diabetes (glucose levels in between). Suppose you are studying a population in which the distribution of glucose levels is normal with mean 100 mg/dL and standard deviation 30 mg/dL. The probability that a randomly sampled person from this population has impaired glucose tolerance but is not diabetic is .

$$1 - .91 - .0004 \approx .09$$

## Question 12 of 23

5.0 Points

An audit of over 5000 (flagged as suspicious) charges to Medicare estimated the average overcharge as \$2150 with 95% CI=(\$2013, \$2287). Mark ALL of the statements below that are correct.

☒ A. In the long run, if we repeated the random sampling and calculated confidence intervals a large number of times, we would know that 95% of those intervals would contain the true population mean. However, for any one sample, we do not know whether the true population mean is included or not.

Correct

☐ B. If we took a very large number of samples of 5000 suspicious charges and calculated 95% confidence intervals for the mean each time, 95% of those intervals would include the true population mean, and 5% would not.

Correct

☐ C. The probability is 0.95 that the true population mean lies between \$2013 and \$2287.

Incorrect

☒ D. There is a 95% chance that this interval contains the true population mean.

Incorrect

## Question 13 of 23

2.0 Points

Using the dataset chinayoung.dta, which contains information on BMI of China Health and Nutrition Survey participants in their 20's, calculate a 95% confidence interval for the BMI of women in this age group using the t distribution. The lower limit of this 95% confidence interval is .

For 13 & 14: in Stata:

sort gender

by gender: ci bmi

20.8

## Question 14 of 23

2.0 Points

Using the dataset chinayoung.dta, which contains information on BMI of China Health and Nutrition Survey participants in their 20's, calculate a 95% confidence interval for the BMI of women in this age group using the t distribution. The upper limit of this 95% confidence interval is .

21.42



## Question 15 of 23

2.0 Points

Using the dataset chinayoung.dta, which contains information on BMI of China Health and Nutrition Survey participants in their 20's, calculate a 99% confidence interval for the BMI of men in this age group using the t distribution. The lower limit of this ~~98~~ 99% confidence interval is  . 21.7

sort gender  
by gender : ci bmi, level(99)

## Question 16 of 23

2.0 Points

Using the dataset chinayoung.dta, which contains information on BMI of China Health and Nutrition Survey participants in their 20's, calculate a 99% confidence interval for the BMI of men in this age group using the t distribution. The upper limit of this ~~98~~ 99% confidence interval is  . 22.7

## Question 17 of 23

10.0 Points

Using the dataset chinayoung.dta, which contains information on BMI of China Health and Nutrition Survey participants in their 20's, calculate and interpret the 90% confidence interval for the BMI of smokers (smoker=1 indicates a smoker, and smoker=0 indicates a nonsmoker) in this age group using the t distribution.

sort smoker  
by smoker : ci bmi, level(90)  
↑  
make sure  
you did a  
90% CI

(Maximum number of characters: 60000)

Show/Hide Rich-Text Editor

(21.87, 22.76)

If we took a large number of random samples of smokers, then approximately 90% of the intervals created from these samples would cover the true population mean of BMI for smokers.

Wrong to say that there's a 90% chance the true mean lies in the calculated interval

## Question 18 of 23

10.0 Points

Suppose that in Chapel Hill in October, the daily high temperature follows a normal distribution with mean 72 degrees F and standard deviation 3 degrees F. Suppose that in February, the daily low temperature follows a normal distribution with mean 31 degrees F and standard deviation 5 degrees F. What would be more unusual: having a high temperature of 85 degrees in Chapel Hill on a day in October, or having a low temperature of 10 degrees in Chapel Hill on a day in February? Explain how you obtained your solution based on the data provided.

(Maximum number of characters: 60000)

[Show/Hide Rich-Text Editor](#)


$$\text{For Feb: } \frac{x - \mu}{\sigma} = \frac{10 - 31}{5} = -4.2$$

$$\text{For Oct: } \frac{x - \mu}{\sigma} = \frac{85 - 72}{3} \approx 4.33$$

$$d1 \text{ normal}(-4.2) \approx .000013$$

$$d1 \text{ 1-normal}(4.33) \approx .000007$$

Since the probability of being 85 or higher in Oct is lower than the probability of being 10 or lower in Feb, 85 in Oct is more unusual

## Question 19 of 23

5.0 Points

Based on the previous data, the probability that the average of the daily high temperatures for Chapel Hill in October will be greater than 74 degrees F is .

$$z = \frac{74 - 72}{3/\sqrt{31}} \approx 3.712$$

$$d1 \text{ 1-normal}(3.712) \approx \boxed{.0001}$$

## Question 20 of 23

5.0 Points

Based on the previous data, the probability that the average of the daily high temperatures for Chapel Hill in October is between 71.5 and 72.5 degrees F is .

$$P(-.928 \leq z \leq .928) \approx \boxed{.65}$$

## Question 21 of 23

5.0 Points

Suppose you are designing a study of total testosterone levels in men. Suppose that in the population as a whole, total testosterone levels are normally distributed with mean 750 ng/dL and standard deviation 175 ng/dL. You want to estimate the mean total testosterone level in your study sample. How many subjects are needed in your sample if you want to have a sample size for which 95% of the sample averages in similar samples would be within +/- 25 ng/dL of the population mean?

$$\text{want } 1.96 \frac{\sigma}{\sqrt{n}} = 25$$

$$\text{so } n \approx \boxed{189}$$

## Question 22 of 23

5.0 Points

Suppose you are designing a study of total testosterone levels in men. Suppose that in the population as a whole, total testosterone levels are normally distributed with mean 750 ng/dL and standard deviation 175 ng/dL. You want to estimate the mean total testosterone level in your study sample. How many subjects are needed in your sample if you want to have a sample size for which 95% of the sample averages in similar samples would be within +/- 10 ng/dL of the population mean?

want  $1.96 \frac{\sigma}{\sqrt{n}} = 10$

so  $n \approx 1177$

## Question 23 of 23

5.0 Points

In this population, which event has the lowest probability?

- ☒ A. A group of 1000 men with average total testosterone level <725 ng/dL
- ☒ B. A group of 10 men with average total testosterone level <600 ng/dL
- ☒ C. An individual man having a total testosterone level <200 ng/dL
- ☐ D. A group of 500 men with average total testosterone level >775 ng/dL

[Reset Selection](#)

Correct answer but will be B on the HW you did (I believe)

Letter varies because response order is randomized.

Assessment Preview - This is an example student view of this assessment