Primary Research Question

How many letters long is the typical UT student's name? How does our estimate change as we increase the size of our sample?

Breakdown Your Analysis

Let's break this analysis into its required steps:

Determine the population parameters:
1. Visualize the shape of the population data by making a histogram.
2. Calculate the "true" mean and standard deviation of the population.

Compare the sample statistics:
3. Draw 1,000 samples of size n=5 from the population data. Calculate the mean of each sample.
4. Graph these 1,000 sample means in a histogram and examine the shape.
5. Calculate the mean and standard deviation of the sampling distribution.
6. Repeat this process for samples of size n=15 and n=25.
7. Compare the results you get to the predictions of the Central Limit Theorem.

Here is the code you will use:

```
# Calculate the population parameters
hist(survey$name_letters)
fivenum(survey$name_letters)
mean(survey$name_letters)
sd(survey$name_letters)


# Draw 1,000 samples of n=5 and find the mean of each sample.
xbar5 <-rep(NA, 1000)
```

```
for (i in 1:1000)
{x <-sample(survey$name_letters, size =5)
xbar5[i] <-  mean(x)}


# Graph the histogram of 1,000 sample means.
hist(xbar5,xlim=c(2,10))


# Calculate the mean and sd of the sampling distribution.
mean(xbar5)
sd(xbar5)

# Compare to the std dev predicted by the CTL.
sd(survey$name_letters)/sqrt(5)


#Repeat for samples of size n=15
xbar15 <-rep(NA, 1000)
for (i in 1:1000)
{x <-sample(survey$name_letters, size =15)
xbar15[i] <- mean(x)}
hist(xbar15,xlim=c(2,10))
mean(xbar15)
sd(xbar15)
sd(survey$name_letters)/sqrt(15)


#Repeat for samples of size n=25
xbar25 <-rep(NA, 1000)
for (i in 1:1000)
{x <-sample(survey$name_letters, size =25)
xbar25[i] <- mean(x)}
hist(xbar25,xlim=c(2,10))
mean(xbar25)
sd(xbar25)
sd(survey$name_letters)/sqrt(25)
```

(3/3 points)
Focus on this portion of the code in order to answer the following questions:

```
xbar5<-rep(NA, 1000)
for (i in 1:1000)
{x<-sample(survey$name_letters, size =5)
```

```
xbar5[i] <- mean(x)}
```

1a) What is *x*?

- *x* is the number 5
- *x* is the number of letters in the name of one individual drawn from the population
- *x* is a sample of 5 data values drawn from the population ✔

1b) What is mean(*x*)?

- It is the mean of all the values in the population.
- It is the mean of the 5 data points drawn in each sample. ✔
- It is the mean of 5 sample means.

1c) When the loop is in the 200th iteration (i=200), what will the following code be doing:

```
xbar5[i] <- mean(x)
```

- Calculating the mean of 200 samples.
- Calculating the mean of the 200th sample, and placing it in the 200th position of xbar5 vector. ✔
- Taking 200 observations from the population and then calculating the mean.

Click here for a video explanation of how to answer this question.

You have used 1 of 1 submissions

(1/1 point)

2) The standard deviation of a sampling distribution is called a "standard error." What goes in the denominator of this equation to solve for standard error (SE)?

$SE = \sigma / ?$

- 15
- μ
- $\sqrt{n}$ ✔

Click here for a video explanation of how to answer this question.

You have used 1 of 1 submissions

(1/1 point)

3) We used the following code to try to show the sampling distribution of ages:

```
xbar5 <-rep(NA, 1000)
for (i in 1:1000)
{x <-sample(survey$age, size =5)
xbar5[i] <- mean(x)}
hist(xbar5,xlim=c(2,10))
```

Why was the histogram that R produced blank?

- The xbar5 vector (of sample means) is empty.
- Since we are taking random samples, it is not unusual to have this histogram.
- The scale of the x-axis is set from 2 to 10, but the ages are not in this range. ✔
- The sample size in line 4 needs to be 1000 to match the previous lines.

Click here for a video explanation of how to answer this question.

You have used 1 of 1 submissions