Problem Set "Assessing Normality - Normal Distn Questions" id:[14283]

1) Assistment #113499 "113499 - Assessing Normality - Normal Distn Questions"

A) One of the first things one should do after recieving data is to view the data. After having viewed your data, you might start to make assumptions about your distribution. For example, you might start to think your data is normal or possibly even uniform.

This assistments activity deals with assessing if your data is normal or not. It is very risky to just assume your data takes on a certain distribution. Therefore, a variety of techniques can be used to assess normality.

Lets go ahead and use R to view a hypothetical data set of 100 numbers. Use the following code with R:

```
### Start Code
set.seed(100)
x=rnorm(100,70,10)
hist(x, main="Histogram of 100 Numbers", breaks=20)
### End Code
```

After viewing the code, do you feel that your data may possibly be normal?

Multiple choice:

✓ Yes

✗ No

Hints:

- Data does not have to be perfectly symmetric to be normal. Remember, data is random, and we expect to see some imperfections in the histogram.

B) We can then go ahead and overlay the histogram with a density plot.

```
### Start Code
set.seed(100)
mu=70
sd=10
n=1000
x=rnorm(n,mu,sd)
m=mean(x)
s=sd(x)
h=hist(x, main="Histogram of 100 Numbers", col="red", breaks=20)
xfit<-seq(min(x),max(x),length=40)
yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))
yfit <- yfit*diff(h$mids[1:2])*length(x)
lines(xfit, yfit, col="blue", lwd=2)
```

```
### Percentage of random numbers within +/- 1 standard deviation
length(x[x>m-1*s & x<m+1*s])/n*100
### Percentage of random numbers within +/- 2 standard deviation
length(x[x>m-2*s & x<m+2*s])/n*100
### Percentage of random numbers within +/- 3 standard deviation
length(x[x>m-3*s & x<m+3*s])/n*100
### End Code
```

Question: This data set had 1000 random variables. Does it appear as if this data set is normal? Be sure to check the histogram and the percentages in the output to the 68-95-99.7 rule.

Multiple choice:

- Yes
- No

The data follows the 68-95-99.7 rule. Therefore, we can assume the data is normal.

C) We can then go ahead and overlay the histogram with a density plot for comparison. In this example, we will use a much smaller data set of only 50.

```
### Start Code
set.seed(100)
mu=70
sd=10
n=50
x=rnorm(n,mu,sd)
m=mean(x)
s=sd(x)
h=hist(x, main="Histogram of 100 Numbers", col="red", breaks=20)
xfit<-seq(min(x),max(x),length=40)
yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))
yfit <- yfit*diff(h$mids[1:2])*length(x)
lines(xfit, yfit, col="blue", lwd=2)
### Percentage of random numbers within +/- 1 standard deviation
length(x[x>m-1*s & x<m+1*s])/n*100
### Percentage of random numbers within +/- 2 standard deviation
length(x[x>m-2*s & x<m+2*s])/n*100
### Percentage of random numbers within +/- 3 standard deviation
length(x[x>m-3*s & x<m+3*s])/n*100
### End Code
```

Question: This data set has only 50 numbers. Does it appear as if this data set is normal? Be sure to check the histogram and the percentages in the output to the 68-95-99.7 rule.

Multiple choice:

- Yes
- No

D) We can then go ahead and overlay the histogram with a density plot for comparison. In this example, we will use an even smaller data set of only 25.

```
### Start Code
set.seed(100)
mu=70
sd=10
n=25
x=rnorm(n,mu,sd)
m=mean(x)
s=sd(x)
h=hist(x, main="Histogram of 25 Random Normal Numbers", col="red",
breaks=20)
### Percentage of random numbers within +/- 1 standard deviation
length(x[x>m-1*s & x<m+1*s])/n*100
### Percentage of random numbers within +/- 2 standard deviation
length(x[x>m-2*s & x<m+2*s])/n*100
### Percentage of random numbers within +/- 3 standard deviation
length(x[x>m-3*s & x<m+3*s])/n*100
### End Code
```

Question: This data set has only 25 numbers. This small data set is drawn from the normal distribution. What conclusions can you draw from this example?

Multiple choice:

- It is hard to make assumptions about data when the data set is small
- If the data set is small and it doesn't look normal, then the data is clearly not normal

E) A second method can be used to assess normality. This method requires the construction of a normal probability plot.

Probability plotting is a graphical method for determining whether sample data conform to a normal distribution, based on a subjective visual examination of the data. If the data is normal, a straight line should appear on the plot.

```
### Start Code
set.seed(100)
mu=70
sd=10
n=100
```

```
x=rnorm(n,mu,sd)
m=mean(x)
s=sd(x)
par(mfrow=c(2,2))
h=hist(x, main="Histogram of 25 Random Normal Numbers", col="red",
breaks=20)
qqnorm(x, main="Normal Probability Plot - Normal"); qqline(x)
x=runif(100, 0, 10)
h=hist(x, main="Histogram from the Uniform Distribution", col="blue",
breaks=20)
qqnorm(x, main="Normal Probability Plot - Non-Normal"); qqline(x)

### End Code
```

Question: Through inspection of the 4 graphs, do you notice how a non-normal distribution can be detected by using the normal probability plot?

Multiple choice:

✓ Yes
✗ No

F) What is the z-score at the mean of a normal distribution?

Multiple choice:

✓ 0
✗ 1
✗ 50
✗ 100

Hints:

- Remember: The z score tells you how far a number is away from the mean, in terms of the standard deviation.

G) Suppose that weights of bags of potato chips coming from a factory follow a normal distribution with mean 12.8 ounces and standard deviation .6 ounces. If the manufacturer wants to keep the mean at 12.8 ounces but adjust the standard deviation so that only 1% of the bags weigh less than 12 ounces, how small does he/she need to make that standard deviation?

Multiple choice:

✓ .343 ounces
✗ -.343 ounces
Standard deviation can never be negative
✗ -.8
✗ .8
✗ .31

Scaffold:

First, find the z-score that corresponds to an area of .01. Remember, area is found in the main body of the z-table and the z-score is found on the edges.

Multiple choice:

- ✓ -2.33
- ✗ .4960

No, .01 is NOT the z-score, it is the area.

Scaffold:

Once you have the z-score of -2.33, you can then use the formula z=(x-mu)/sigma. Remember, you are solving for sigma. Therefore, you need to solve for sigma. Sigma=(x-mu)/z . You now know x, mu, and z. Solve for sigma.

Multiple choice:

- ✓ .343
- ✗ .01
- ✗ .495
- ✗ .26

H) Which of the following statements is NOT correct according to the Empirical Rule?

Multiple choice:

- ✗ Approximately 99.7% of observations fall within three standard deviations of the average.
- ✓ It is good for all symmetric distributions.
- ✗ Approximately 95% of the observations fall within two standard deviations of the average.
- ✗ Approximately 68% of observations fall within one standard deviation of the average.

I) Suppose that combined verbal and math SAT scores follow a normal distribution with mean 896 and standard deviation 174. Suppose further that Peter finds out that he scored in the top 3% of SAT scores. Determine how high Peter's score must have been.

Multiple choice:

- ✓ 1223.12
- ✗ 1132
- ✗ 1342
- ✗ 1234

Hints:

- If I score in the top 3%, I essentially do better than 97% of the students. Remember, the z-table always displays the area to the left.

J) A large college class has 900 students, broken down into section meetings with 30 students each. On the final exam, scores followed a normal distribution with an average of 63 and a standard deviation of 20.

If you randomly select one of these students, what is the probability that the selected student scored between 56 and 70 on the final exam?

Multiple choice:

✗ 35%
✓ .2736
✗ .3632
✗ .6368

K) A standardized measure of achievement motivation is normally distributed, with a mean of 35 and a standard deviation of 14. Higher scores correspond to more achievement motivation.

Shamu scored in the top 5% of test takers. What was her actual achievement motivation score?

Multiple choice:

✓ 58
✗ 49
✗ 63
✗ 77
