

Chapter 2: Statistical Models for Distributions

2.2 Normal Distributions

In Chapter 2 of YMS, we learn that distributions of data can be approximated by a mathematical model known as a **density curve**. In this section, we learn that many sets of data, probability applications, and sampling situations can be modeled by normal distributions. If a set of data is *approximately normal*, we can use *standardized normal calculations* to determine relative standing, percentiles, and probabilities. Like our data displays and numeric summaries, it is important that you understand how to perform the following calculations by hand before relying on the calculator.

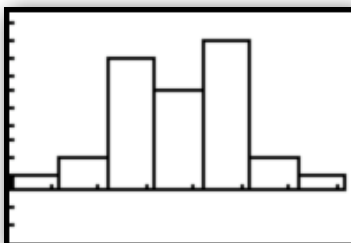
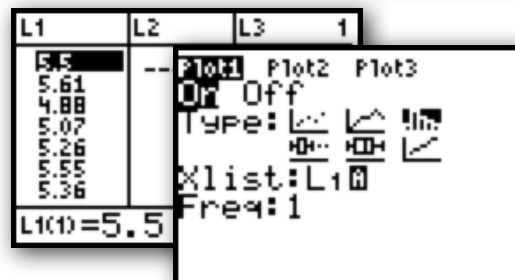
Approximating a Normal Curve

Consider the following measurements of the density of the earth taken by Henry Cavendish in 1798. Use your calculator to construct a histogram of these data.

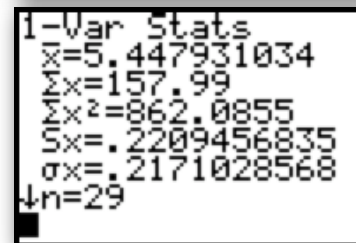
Density of the earth (as a multiple of the density of water)

5.5	5.61	4.88	5.07	5.26	5.55	5.36	5.29	5.58	5.65
5.57	5.53	5.62	5.29	5.44	5.34	5.79	5.10	5.27	5.39
5.42	5.47	5.63	5.34	5.46	5.30	5.75	5.68	5.85	

1. Enter the Density data into **L1**
2. Set up **STAT PLOT** to view a histogram
3. Use **ZOOM 9:ZoomStat** to view the distribution
4. Calculate the mean and standard deviation by using **[STAT] Calc 1:1-Var Stats**



Note the shape of this distribution. It appears it can be approximated by a **normal distribution** with mean = 5.45 and standard deviation = 0.22.







Remember, just because data appears normally distributed, that doesn't make it so. We can assess the normality of data by constructing a different stat plot on our calculator. This plot, known as a **normal quantile plot** or **normal probability plot** can be used to assess the adequacy of a normal model for a data set. If the plot appears to have a linear pattern, a normal model is reasonable. If the plot appears nonlinear, we believe the data may be from a nonnormal distribution.

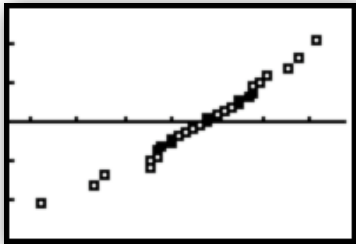
Assessing Normality

Before performing normal calculations, let's check the normality of our Density data. If you haven't done so already, enter the Cavendish Density data into **L1**.

1. Enter Density data into **L1**
2. Set up **STAT PLOT** to construct a normal probability plot of the data in **L1**. This is the last option under **Type**:
3. Use **ZOOM 9:ZoomStat** to see the plot
4. Assess the normality of the Density data

L1	L2	L3	1
5.5			
5.61			
4.88			
5.07			
5.26			
5.55			
5.36			
L1(1)=5.5			

Plot1	Plot2	Plot3
On	Off	
Type: 		
Data List: L1		
Data Axis: Y		
Mark: 		



Since this plot appears approximately linear, we can conclude that it is reasonable to assume the data are from a normal distribution.

Normal Calculations

The **Empirical (68-95-99.7) Rule** for approximately normal distributions is a handy tool for calculating relative standing, percentiles, and probabilities. However, many problems we encounter will involve observations that don't fall exactly 1, 2, or 3 standard deviations away from the mean. In these cases, we'll need to perform a standard normal calculation.

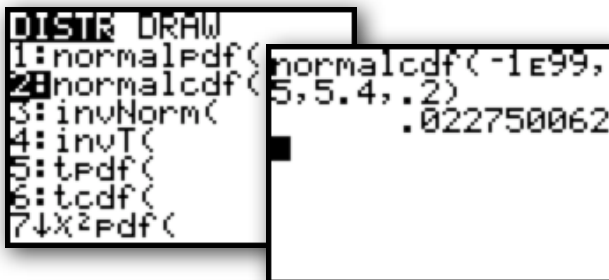
Standard normal calculations are fairly easy to perform by hand, but our calculator does have the ability to automate the process. Again, please be sure to practice these calculations by hand before using the calculator. In some cases, calculating standardized scores and looking up values on a table may be quicker than using the calculator!

Standard normal calculations are performed in the **Distribution** menu of the calculator. You will want to use the "normal cumulative density function". To use this, you must enter the following parameters in order **normalcdf(min, max, mean, standard deviation)**. If your min is not defined to the left, use **-1E99**. If your max is not defined to the right, use **1E99**. The mean and standard deviation should be the given or calculated mean and standard deviation of your approximately normal distribution.

Let's use the Density data to answer some questions about the proportion of observations that would satisfy certain conditions. We have already determined this data is approximately normally distributed with a mean=5.4 and standard deviation of .2.

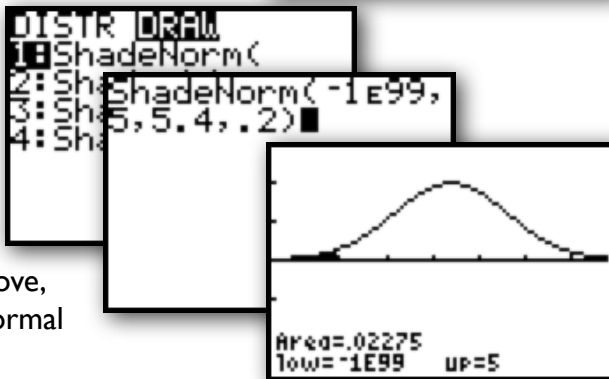
What proportion of density observations fall below 5?

1. Press **2ND** **[VARS]** (**DISTR**)
2. Choose **2:normalcdf**(
3. Enter **normalcdf(-1E99, 5, 5.4, .2)**



If you would like to see the curve:

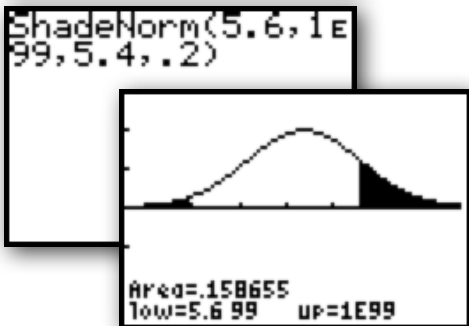
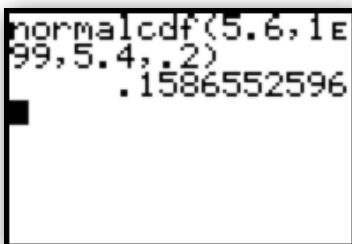
1. **2ND** **[VARS]** (**DISTR**)
2. Choose **DRAW 1:ShadeNorm**(
3. Enter **ShadeNorm(-1E99, 5, 5.4, .2)**



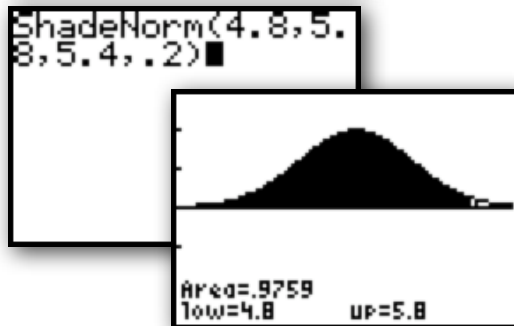
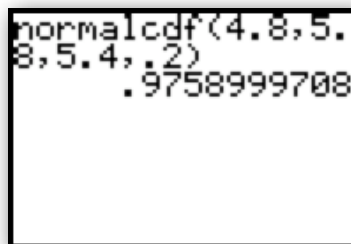
Notice in both cases we find approximately 2.3% of observations will fall below a measurement of 5.

This process can be used to calculate the proportion of observations falling above, below, or between any two points in a normal distribution.

Find the proportion of density observations above 5.6:



Find the proportion of density observations between 4.8 and 5.8:



Inverse Normal Calculations

A number of times throughout the course, you will be asked to determine what point cuts off a desired proportion in a normal distribution. For example, we may want to know what value determines the 80th percentile in our density data. That is, what value falls at a point such that 80% of observations fall below it?

To answer this question on our calculator, we need to use the “inverse normal” function with the following parameters **invNorm(left area, mean, standard deviation)**.

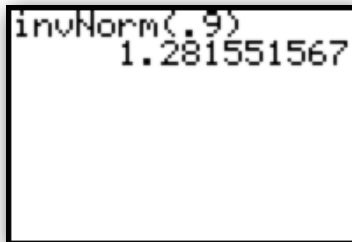
1. Press **2ND VARS (DISTR)**
2. Select **3:invNorm(**
3. Enter **invNorm(.8, 5.4, .2)**

We can see that 80% of density observations will fall below 5.57.



Finding z-scores with invNorm

If you would like to find a z-score corresponding to a desired percentile, you can simply enter **invNorm(percentile)**. If you don't enter a mean or standard deviation in the **invNorm, normcdf, or ShadeNorm** functions, the calculator assumes you are referring to a standard normal distribution with mean = 0 and standard deviation = 1.



For example, the z-score that determines the 90th percentile in a standard normal distribution is **invNorm(.9) = 1.281551567**

AP[®] Examination Tips

When taking the Advanced Placement Statistics Exam, you will most likely be asked to perform standard normal calculations. Remember, the calculator can be used to automate your calculations and provide basic data displays...however, it is your job to provide the contextual interpretation! Never answer a question by just copying a calculator plot or by using “calculator” speak.

When assessing normality

- Be careful inputting your data
- Choose the appropriate plot(s) (histogram and normal probability plot)
- Sketch the plot(s) and LABEL the axes!
- Interpret the graphs in the context of the problem
- Remember, linearity in the normal probability plot suggests normality

When performing normal calculations

- Sketch your normal distribution
- Label the mean and indicate the standard deviation
- Plot your point(s) of interest
- Shade the area you wish to calculate
- Make sure you choose **normalcdf**(NOT **normalpdf**(
- Interpret the area in the context of the problem
- DO NOT USE calculator speak (eg: Don't write “*normalcdf = .023*” ...
Instead, say something like “*2.3% of observations fall below...*”)
- Remember, sometimes it may be easier/faster to calculate a z-score by hand and use the table to find the appropriate area.

