SAS Tutorial for STAT 350 Lab 3

Author: Leonore Findsen, Cheng Li

1. Generate random samples from a normal distribution

We are going to generate random samples from a number of different distributions in this laboratory. The following code is for the normal distribution. I will also be providing similar code for the other distributions that we will be using in part three of this tutorial. The function that is used in SAS is `RAND('Normal',mu,sigma)`.

a) Generate 20 random numbers from a normal distribution with $\mu = 572$ and $\sigma = 51$ and calculate the mean and standard deviation of the data set.

Solution:

```
*random number generation for normal is RAND('Normal',mu,sigma);

/*The %Let statement allows you to define a "macro variable" at the
  beginning of the code (a placeholder for some text - it is NOT the
  same as a dataset variable). This is useful if you want to make
  multiple runs of the same code, but with slight modifications to the
  analysis each time. This way, you don't have to search through the
  whole code to make all the necessary changes; you just update the
  %Let statement. For example, you could use code like the below to
  generate different normal distributions with several different values
  of sigma to compare, just by changing %Let.

  To access this variable's value later, put an & in front of the
  variable name. For example, the first line is %Let points = 20. To
  access this later, use &points */

%Let points = 20; *this is the number of data points in the sample;
%Let mu = 572;
%Let sigma = 51;
%Let Random = rand('normal', &mu, &sigma);
%Let title = Random Normal;
data RandomData;
   do x=1 to &points; /*When I use &points, I don’t need to search to
                       change the number of points. The change is now
                       only done in the beginning of the code; */
      answer=&Random; /*When you want to use a different distribution,
                       you replace 'normal' in the rand() function with
                       the name of the desired distribution in &Random
                       above. See below for more details; */
      output;
   end;
run;

*You are required to print out your data in this lab;

proc print data=RandomData;
   var answer; *only answer (which holds the random data) will be
                written out, all other variables will be excluded;
run;
```

SAS Tutorial for STAT 350 Lab 3

Author: Leonore Findsen, Cheng Li

```
/*"proc means" outputs just the number of data points, mean, standard
    deviation, minimum and maximum. This outputs much less material
    than "proc univariate".
  You will not be using this in this lab, but it is useful information
    for lab 4.*/

proc means data=RandomData;
   var answer; *I want the information just for the variable answer,
                not x;
run;
```

The SAS System

Obs	answer
1	574.651
2	683.140
3	535.007
4	583.070
5	584.365
6	643.306
7	592.253
8	496.166
9	587.682
10	526.566
11	666.887
12	598.203
13	488.300
14	554.402
15	608.596
16	528.734
17	550.497
18	623.820
19	598.920
20	448.061

The SAS System

The MEANS Procedure

Analysis Variable : answer				
N	Mean	Std Dev	Minimum	Maximum
20	573.6311805	59.3630140	448.0607594	683.1396371

Note: Each time that the program is run, you will get different values, means, and standard deviations.

2. Determine if a distribution is normal

SAS Tutorial for STAT 350 Lab 3

Author: Leonore Findsen, Cheng Li

b) Make a histogram of the data in part (a) and visually assess if the normal density curve and the histogram density estimate (i.e., the kernel) are similar.
c) Make a normal probability plot of the data in part (a) and visually assess if the plotted points are randomly scattered below and above the line without a discernable pattern.

Solution:

I am doing the problem with the data from part (a), but it doesn't matter what data is used. I am adding titles to the plots, making it easier for you to identify which graph belongs to which part.

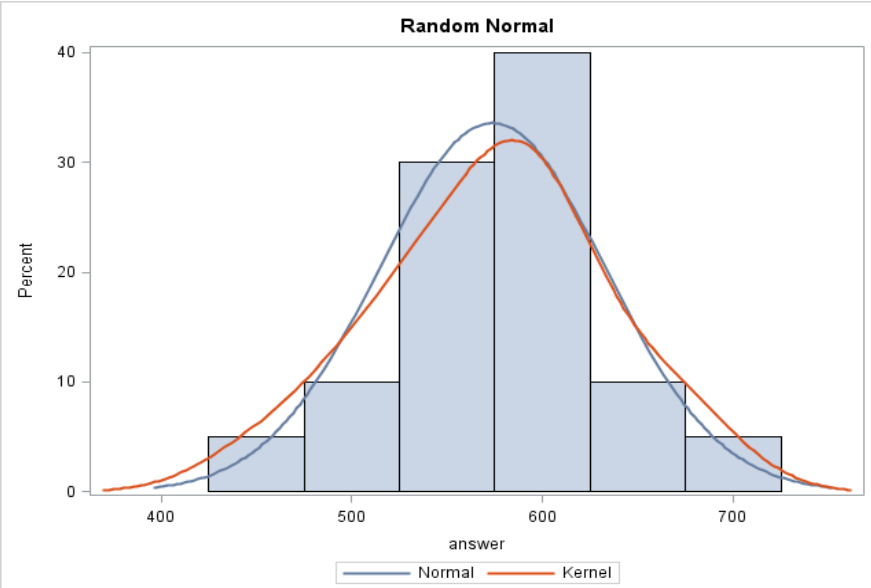
```
proc sgplot data=RandomData;
  title "&title";
  histogram answer;
  density answer; *this adds the theoretical normal density curve using
                   the mean and standard deviation from the data;
  density answer/type=kernel; *this adds the smoothed kernel density
                   curve, which is like a smooth version to the
                   histogram to represent the overall shape of the
                   actual data;
run;

proc univariate data=RandomData noprint;
     *noprint prevents all of the tables from being printed out. This
      option is often used when only graphs are wanted from
      "proc univariate";
  qqplot answer/normal (mu=est sigma=est);
   *explanation of the keywords after the /
    I am comparing this to a normal distribution (so SAS is looking up
      the z values).
    The line will be drawn using the data points to estimate mu and
      sigma;
run;
```

SAS Tutorial for STAT 350 Lab 3

Author: Leonore Findsen, Cheng Li

b) Make a histogram of the data in part (a) and visually assess if the normal density curve and the histogram density estimate are similar.

Solution:

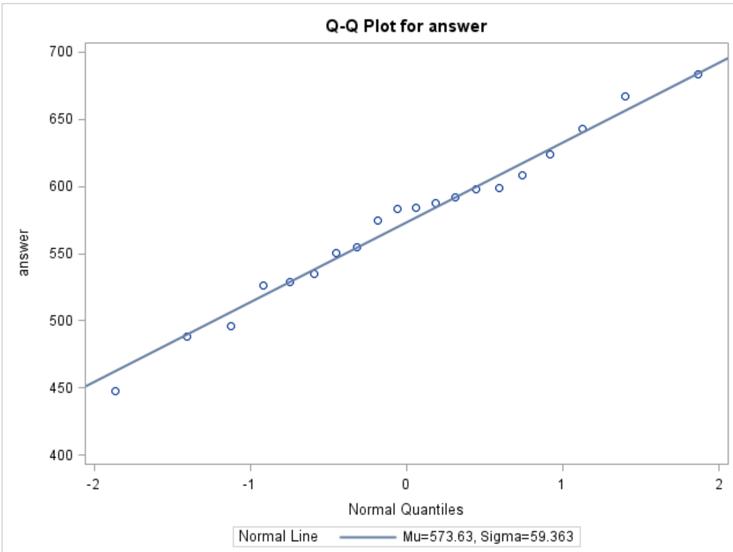


There are two complimentary ways to determine if this distribution is normal. 1) You can assess whether the blue normal curve is 'close' to the red smoothed curve. 2) You can look at the symmetry, modality and tails of the histogram and red smoothed curve to see if it is possible that the distribution is normal. When you are assessing normality from a histogram, always use both methods. In this case, the two curves are similar, the distribution is unimodal, and looks approximately symmetric. Therefore, this distribution resembles a normal distribution. When you are using a histogram to determine normality, please always include the two extra curves.

Author: Leonore Findsen, Cheng Li

c) Make a normal probability plot of the data in part (a) and visually assess if the sample quantiles are randomly scattered below and above the line without a discernable pattern.

Solution:



Since the data points are randomly scattered below and above the line without a pattern, the randomly generated data does not appear to deviate substantially from a normal distribution.

3. Generate random samples for right skewed, left skewed, short tailed, long tailed distributions

The following four distributions illustrate different types of skewness and tails. We will generate random samples from each of them respectively, and compare them with the normal distribution using the visual methods introduced in part (2).

Right skewed: Exponential distribution ($\lambda = 5$)
Left skewed: Beta distribution (on [0,1], $\alpha = 7$, $\beta = 0.8$)
Short tailed: Uniform (on [a = -3, b = 3])
Long tailed: t-distribution (df = 1)

The following commands are used for the above distributions. For each distribution, change the random number generator, change the title, and replace &Random with the appropriate macro variable name, e.g. answer = &right. Then run the commands to create a histogram and the probability plots .

Author: Leonore Findsen, Cheng Li

```
%Let points = 100; *Number of data points;
/* The four distributions
   Only one needs to be included in the code at a time. Replace &Random
   with the following rand() statement as appropriate.
     right skewed: rand('Exponential',1/5
     left skewed: rand('Beta',7,0.8)
     short tailed: 6*rand('Uniform') - 3
     long tailed: rand('T',1)
  I have typed in the right skewed below.
%Let Random = rand('Exponential',1/5);
  /*The exponential only works with version 9.4 like the version on
    ITaP computers. If you are using an older version of SAS,
    please see your instructor*/
%Let title = Right Skewed Distribution;
/* In addition, the first (and second) word(s) in the main title needs
   to be changed. I have it set to right skewed, you will need to
   change this to left skewed, long tailed, or short tailed as
   appropriate. */

/*The rest of the code is the same as before and will not be repeated*/
```

No output is provided.