

Lastly, the normal probability plot is displayed in Figure 7.3.6. The solid line represents the theoretical normal data and we can see that the actual data do not stray too far from the line. Again, the normal probability plot supports that the data follow a normal distribution.

### Normal Probability Plot of the Sales Revenue Data

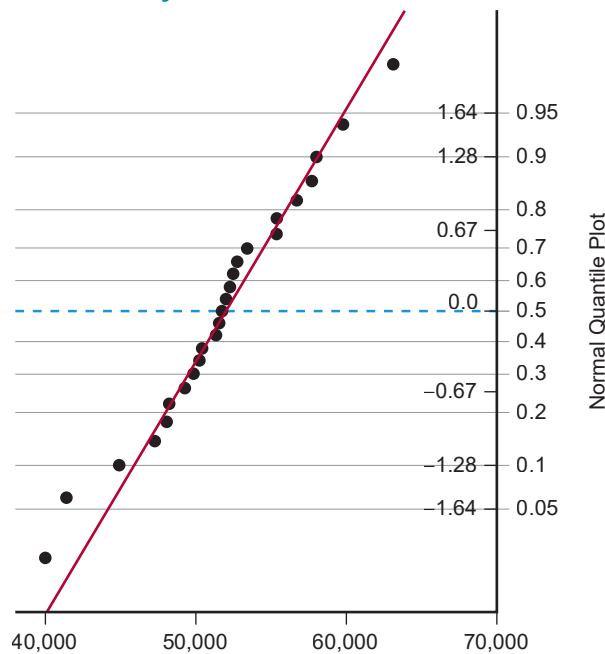


Figure 7.3.6

Given that all three graphical displays used in the example tend to support that the data follow a normal distribution, it is safe to continue with any analysis based on the assumption that the data are normally distributed.

More formal and precise tests are available which calculate the probability that a sample is collected from a normal population. To name a few, these tests are the Anderson-Darling Test, Kolmogorov-Smirnov Test, and Shapiro-Wilk W Test. These tests have the advantage of allowing the analyst to make an objective judgment of normality. However, the primary disadvantage is that these tests are sensitive to small sample sizes.

We have discussed using graphical techniques to determine if the data are collected from a normal distribution. What if the data do not follow a normal distribution? You can still analyze the data but first you must perform some type of transformation on the data or use nonparametric procedures. Keep in mind that if you use nonparametric procedures, they have less power than the classical parametric procedures. Lastly, parametric tests are robust to violations of normality when you have large sample sizes.

## 7.3 Exercises

### Basic Concepts

1. List three ways to graphically assess the normality of a data set.
2. Describe the general procedure for creating a normal probability plot.
3. How should a normal probability plot look to indicate normality?

### Exercises

4. Construct a histogram using the “BA” (batting average) column of the Moneyball data set. Can we assume batting averages have a normal distribution?

©HAWKES LEARNING

### Data

The Moneyball data set can be found by visiting [stat.hawkeslearning.com](http://stat.hawkeslearning.com) and navigating to **Discovering Business Statistics, Second Edition > Data Sets > Moneyball**.

**Data**

The Housefly Wing Lengths data set can be found by visiting [stat.hawkeslearning.com](http://stat.hawkeslearning.com) and navigating to **Discovering Business Statistics, Second Edition > Data Sets > Housefly Wing Lengths**.

5. Create a normal probability plot of the housefly wing lengths data. What do you observe?
6. A pharmaceutical company wants to test whether a new cold medication will perform better than an existing medication. Laboratory technicians observe a sample of 25 patients and record the number of hours it takes for each patient to feel symptom relief after taking the medicine. Before the company performs a test of the new medication against the current one, they need to know if the data are normally distributed. Use a normal probability plot to determine if the data appear to come from a population that is normally distributed.

3.00	1.50	0.20	1.62	1.06
3.01	2.45	0.66	1.94	0.21
1.51	3.08	5.37	6.96	1.32
0.79	7.20	1.36	4.45	3.29
1.74	3.87	1.90	3.50	3.09

7. Data on the total annual rainfall (in millimeters) in South Carolina was gathered by a weather station in Aiken, South Carolina from 2001-2015. Use a normal probability plot to determine if the data appear to come from a population that is normally distributed.

Total Annual Rainfall in South Carolina					
Year	Total Precipitation (in millimeters)	Year	Total Precipitation (in millimeters)	Year	Total Precipitation (in millimeters)
2001	895.7	2006	1031.3	2011	991.8
2002	1106.9	2007	1002.7	2012	1089.5
2003	1681.3	2008	1321.6	2013	1584.0
2004	1003.6	2009	1434.0	2014	1070.2
2005	1166.1	2010	946.2	2015	1537.4

Source: The United States Historical Climatology Network.

8. A professor is interested in examining the distribution of the grades his students received on the midterm exam. There are 18 students in the class, and no time limit was given for the exam. Use a normal probability plot to determine if the students' grades are normally distributed.

80.8	81.7	81.7	81.7	81.7	82.5
83.3	83.3	84.2	84.2	85	86.7
86.7	87.5	87.5	90.3	90.4	90.8

Source: <https://openmv.net/info/unlimited-time-test>

9. A group of students and professors are studying conifers in the Pacific Northwest United States. They take a sample of 25 Douglas Fir trees and record several metrics, including the circumference of the trunks (in meters). Use a normal probability plot to determine if the trunk circumference values are normally distributed.

4.97	0.45	0.40	0.15	2.84
6.65	0.62	0.39	0.86	1.24
4.93	0.64	0.62	2.22	2.23
0.29	0.18	0.27	1.97	2.45
0.19	0.55	0.41	2.85	9.09

Source: Biometrics of Douglas firs. <http://seattlecentral.edu/qelp/sets/076/076.html>. White River Valley, Washington, 20-Apr-01. Students: Ingrid McNeely and Dylan Morgan, Seattle Central Community College.

10. A group of friends decide to run a marathon together. There are 16 runners in the group, and they are all in relatively good shape. Use a normal probability plot to determine if their marathon times are normally distributed.

4:07:58	4:18:34	4:21:15	4:24:23
4:08:07	4:18:40	4:22:17	4:25:12
4:16:28	4:19:39	4:23:52	4:25:14
4:17:30	4:19:45	4:23:55	4:26:34

11. The underwriters for a new type of auto insurance policy gathered monthly mileage data from 18 city drivers. What conclusions can be drawn about the distribution of the data?

- Using the mileage data provided below, construct a box plot to assess the normality of the data set.
- Using the mileage data provided below, construct a histogram with eight classes to assess the normality of the data set.

385	410	416.5	421	433.5	451.5
408.5	411	416.5	425.2	437.5	452
408.5	412.5	421	433.5	437.5	460

12. Billings Marketing is asked to develop a recruiting campaign for ABC University. Using the age data from recent college applications shown below, construct a box plot and a histogram with eight classes. Describe the normality of the data set. The results will aid in the development of a marketing strategy.

18	19	20	26	22	20
18	19	20	27	22	19
18	28	20	19	23	20
19	19	21	29	19	30
24	19	21	18	24	19
19	19	21	19	25	18

13. Using the data from Exercise 8, construct a box plot and a histogram with 8 columns. The results seem to suggest that the data are not normally distributed. A normal probability plot indicated that the data was normally distributed. Why do these graphical methods for accessing normality seem to contradict one another?

80.8	81.7	81.7	81.7	81.7	82.5
83.3	83.3	84.2	84.2	85	86.7
86.7	87.5	87.5	90.3	90.4	90.8