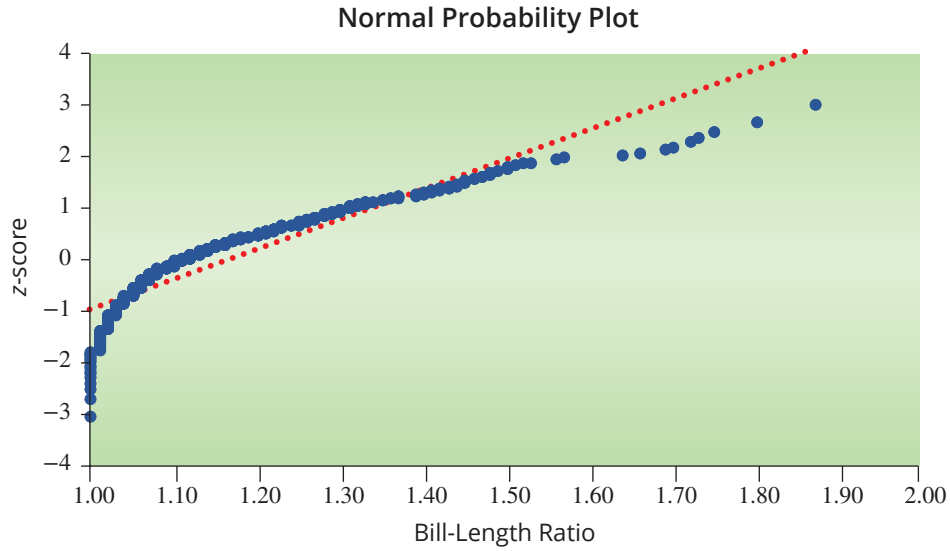

From the graph, we can see that the plotted points do not follow a linear trend. This particular pattern indicates that the data are skewed right and not normally distributed, confirming what we observed in the histogram.

Evaluating the normality of a data set is crucial in parametric statistics. We'll revisit this subject when we delve into different statistical tests and models in upcoming chapters.

8.5 Exercises

Basic Concepts

1. List two ways to graphically assess the normality of a data set. Under what conditions are each appropriate?
2. Describe the general procedure for creating a normal probability plot.
3. How should a normal probability plot look to indicate normality?

Exercises

Please use technology for all the exercises in this section.

Data

stat.hawkeslearning.com
Discovering Statistics and Data, Fourth Edition > Data Sets > Moneyball

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

Data

stat.hawkeslearning.com
Discovering Statistics and Data, Fourth Edition > Data Sets > Housefly Wing Lengths

5. Create a normal probability plot of the housefly data from Example 8.5.2. What do you observe? Does the plot lead you to the same conclusion as the histogram?

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 inches) in Asheville, North Carolina were gathered by a weather station in from 2000-2022.[7] Use a normal probability plot to determine if the data appears to come from a population that is normally distributed.

Total Annual Rainfall in Asheville, NC					
Year	Total Precipitation (in inches)	Year	Total Precipitation (in inches)	Year	Total Precipitation (in inches)
2000	35.59	2008	35.63	2016	33.40
2001	34.49	2009	62.13	2017	54.10
2002	44.47	2010	44.26	2018	79.48
2003	59.46	2011	46.04	2019	57.25
2004	52.36	2012	44.66	2020	64.71
2005	47.26	2013	75.22	2021	54.51
2006	48.29	2014	46.91	2022	45.43
2007	34.39	2015	54.35		

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	85.0	86.7
86.7	87.5	87.5	90.3	90.4	90.8

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).[8] 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

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

8.6 Approximation to the Binomial Distribution

To approximate other distributions, the normal distribution can be very useful. Although it is a continuous distribution, it is used to approximate discrete distributions, specifically the binomial.

The Binomial Distribution

Calculating binomial probabilities can be quite time consuming if n is large. For example, suppose that you intend to sample 2000 subjects for a marketing research survey. If 50 percent of the population believes your product is superior to the competition's, what is the probability of obtaining 600 or fewer subjects who believe your company's product is superior?

$$P(X \le 600) = P(X = 0) + P(X = 1) + P(X = 2) + \cdots + P(X = 599) + P(X = 600)$$

Determining the appropriate probability using the binomial distribution would require the calculation of 601 individual probabilities, many of which would have extremely large combinations such as the following.

$${}_{2000}C_{400}\, 0.5^{400}(1 - 0.5)^{1600}$$

Computing this and the other 600 similar calculations would be a formidable task. The normal distribution is useful in approximating binomial probabilities. The larger the binomial parameter, n, the more accurate the approximation. Determining the probability described above using the normal approximation is trivial in comparison to calculating the exact probability using the binomial.

Recall that the normal distribution is a function of two parameters, the mean and the standard deviation. Thus, if the normal distribution is used to approximate the binomial distribution, it seems reasonable that the mean and standard deviation of the normal should be the same as the mean and standard deviation of the binomial that is being approximated. Specifically, let

$$\mu = E(X) = np, \text{ and}$$

$$\sigma = \sqrt{V(X)} = \sqrt{np(1-p)}.$$