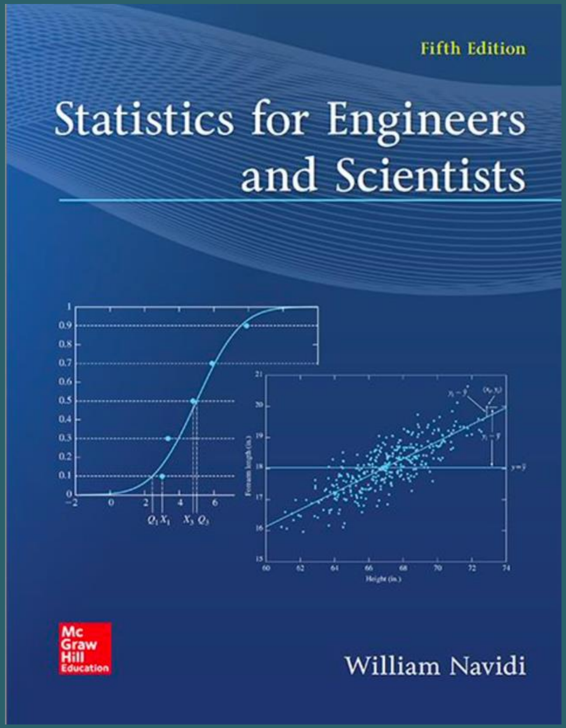


Ch 4: Commonly Used Distributions

(not all sections are required)

Ch 4: Overview (Required Sections)

4-1 The Bernoulli Distribution

4-2 The Binomial Distribution

4-3 The Poisson Distribution

4-5 The Normal Distribution (cont.)

~~*4-9 Some Principles of Point Estimation*~~ *(later)*

4-10 Probability Plots

4-11 Central Limit Theorem

Probability Distribution Curve

- The **area** under the standard normal distribution curve can also be thought of as a **probability** or as the proportion of the population with a given characteristic.
- If you select a z value at random, the probability of choosing one, say, between 0 and 2.00 would be the same as the area under the curve between 0 and 2.00. In this case, the area is 0.4772.
- Therefore, the probability of randomly selecting a z value between 0 and 2.00 is 0.4772.
- The problems involving probability are solved in the same manner as the previous examples involving areas under the standard normal distribution curve.

Probability Notation

$$P(a < z < b)$$

denotes the probability that the *z*-score is between a and b.

$$P(z > a)$$

denotes the probability that the *z*-score is greater than a.

$$P(z < a)$$

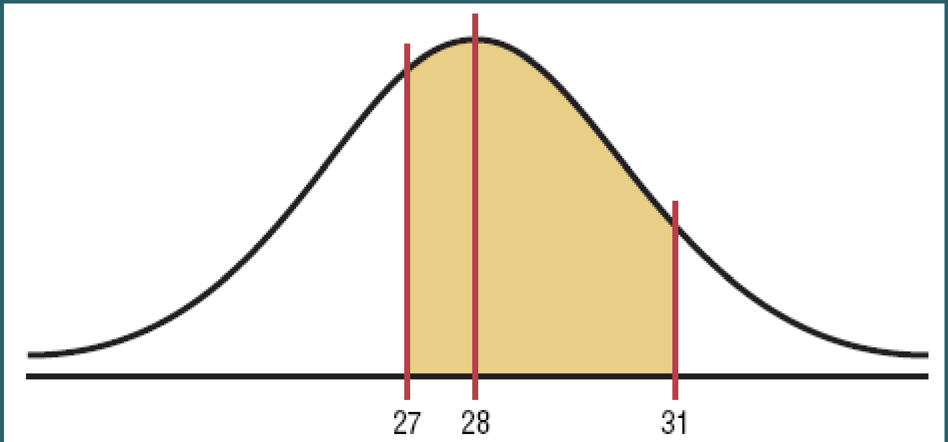
denotes the probability that the *z*-score is less than a.

Example: Newspaper Recycling

Each month, an American household generates an average of 28 pounds of newspaper for garbage or recycling. Assume the variable is approximately normally distributed and the standard deviation is 2 pounds. If a household is selected at random, find the probability of its generating **between 27 and 31 pounds per month.**

Example: Newspaper Recycling…

Step 1: Draw the normal distribution curve.



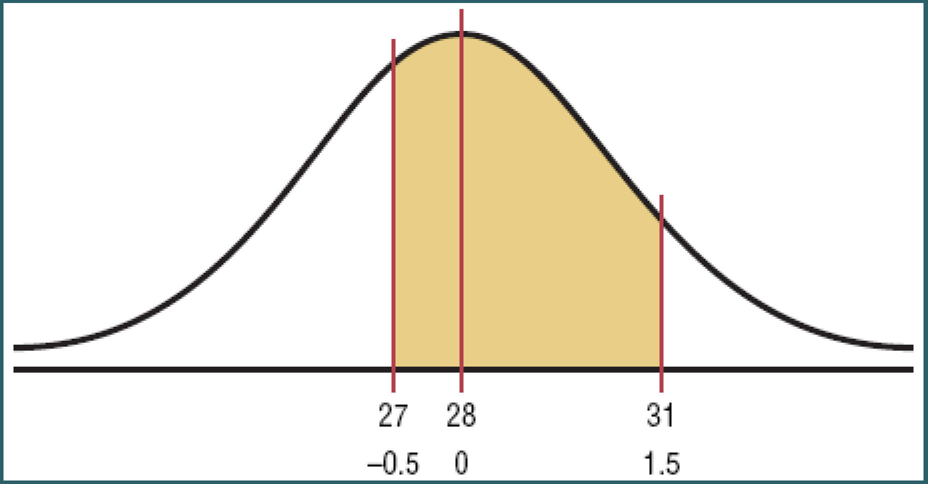
Note: *This step is not necessary if you are confident in using Table A.2*

Example: Newspaper Recycling…

Step 2: Find z values corresponding to 27 and 31

$$z = \frac{27 - 28}{2} = -0.5 \qquad z = \frac{31 - 28}{2} = 1.5$$

Step 3: Find the area between z = -0.5 and z = 1.5



Example: Newspaper Recycling…

Step 3: Find the area between z = -0.5 and z = 1.5

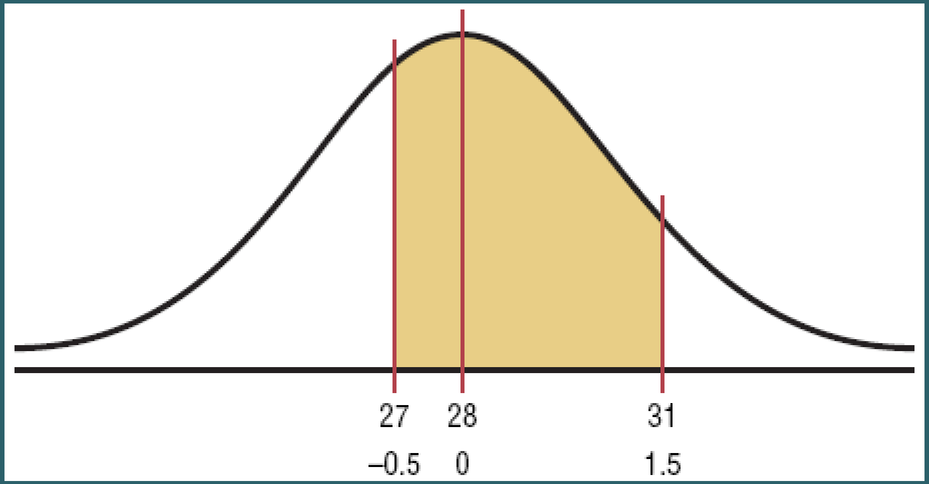


Table E gives us an area of 0.9332 - 0.3085 = **0.6247**

Hence, the probability that a randomly selected household generates between 27 and 31 pounds of newspapers per month is **62.47%**.

Answer

The probability that the square footage of a randomly selected home in the United States would be less than 3651 is equal to **0.9382**.

This can be stated in another way.

Assuming that the square footage of US homes is a normally distributed variable with a mean of 2687 square feet with a standard deviation of 626 square feet, **93.82%** of the homes in the United States would be smaller than 3651 square feet.

Finding Data Values for Specific Probabilities

A normal distribution can also be used to find specific data values for given percentages. In this case, you are given a probability or percentage and need to **find the corresponding data value X**.

Formula for Finding the Value of a Normal Variable *X*

$$X = z \cdot \sigma + \mu$$

The complete procedure for finding an X value is summarized in the Procedure Table shown.

Procedure Table

Finding Data Values for Specific Probabilities

Step 1 Draw a normal curve and shade the desired area that represents the probability, proportion, or percentile.

Step 2 Find the z value from the table that corresponds to the desired area.

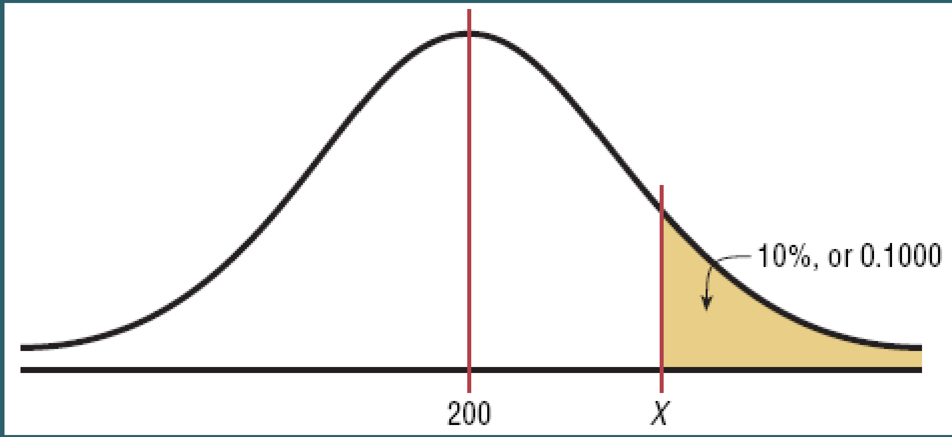
Step 3 Calculate the X value by using the formula $X = z\sigma + \mu$.

Example: Sport Academy Qualifications

To qualify for a sport academy, candidates must score in the top 10% on a general abilities test. Assume the test scores are normally distributed and the test has a mean of 200 and a standard deviation of 20. Find the lowest possible score to qualify.

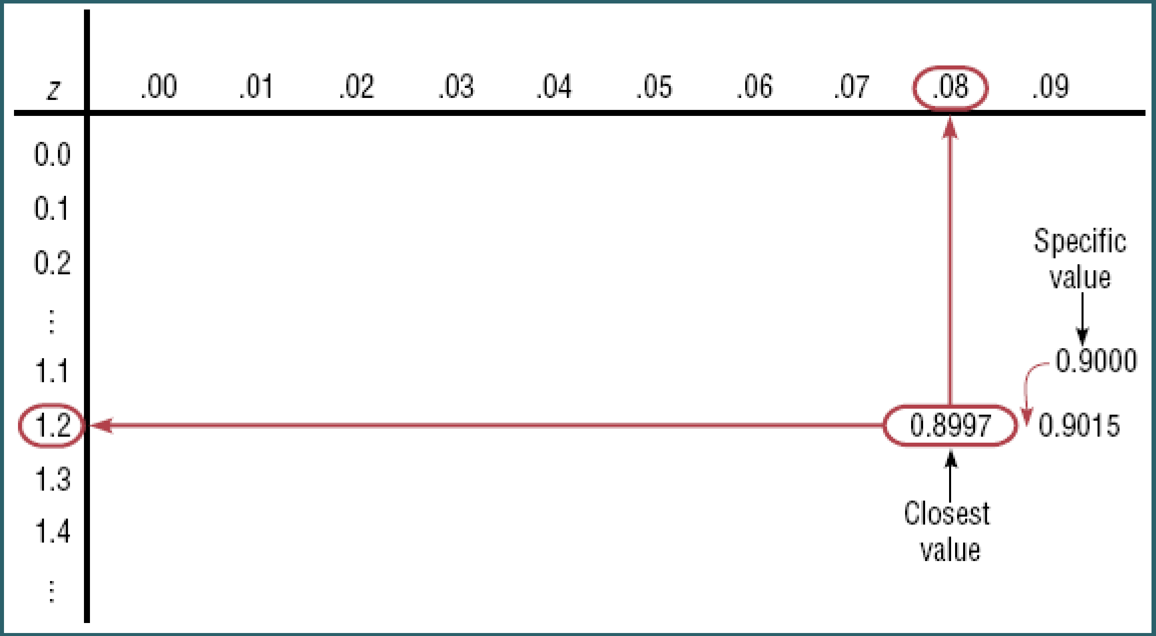
Example: Sport Academy Qualifications…

Step 1: Draw the normal distribution curve.



Example: Sport Academy Qualifications…

<u>Step 2:</u> Subtract 1 - 0.1000 to find area to the left, 0.9000. Look for the closest value to that in Table E.



Example: Sport Academy Qualifications…

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625

Example: Sport Academy Qualifications…

Step 2: Find *X*.

$\mathbf{X = z * \sigma + \mu}$ = 1.28 (20) + 200 = 225.6

= 226 (rounded)

Answer: A score of 226 should be used as a cutoff. Anybody scoring 226 or higher qualifies for the academy.

Determining Normality

- A normally shaped or bell-shaped distribution is only one of many shapes that a distribution can assume.
- However, it is very important since many statistical methods require that the distribution of values (shown in subsequent chapters) be normally or approximately normally shaped.
- There are several ways statisticians use to check for normality.

Determining Normality

A few basic methods are introduced here:

- ✓ Histogram
- ✓ Pearson's Index (PI) of Skewness also called Pearson's Coefficient (PC)
- ✓ Outliers
- ✓ Probability Plots

Example: Technology Inventories

A survey of 18 high-technology firms showed the number of days' inventory they had on hand.

Determine if the data are approximately normally distributed.

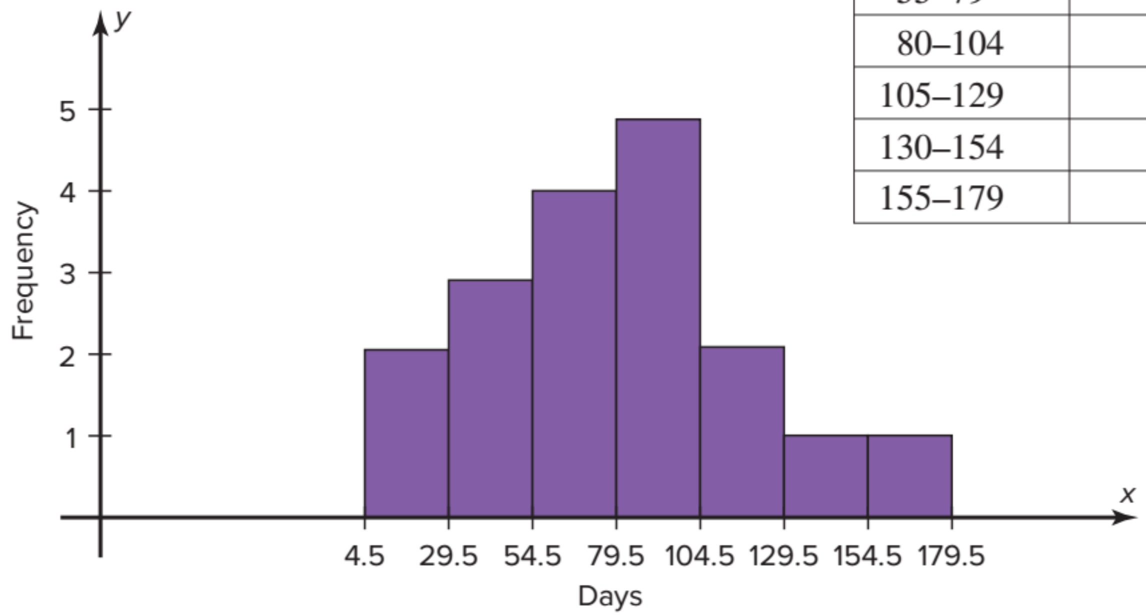
5	29	34	44	45	63	68	74	74
81	88	91	97	98	113	118	151	158

Example: Technology Inventories…

Method 1: Construct a histogram

Class	Frequency
5–29	2
30–54	3
55–79	4
80–104	5
105–129	2
130–154	1
155–179	1

FIGURE 6–29
Histogram for Example 6–11



Since the histogram is approximately bell-shaped, we can say that the distribution is approximately normal.

Example: Technology Inventories…

Method 2: Check for Skewness

Skewness can be checked by using the Pearson coefficient (PC) of skewness also called Pearson's index of skewness. The formula is

$$\boxed{PC = \frac{3(\bar{X} - \text{median})}{s}}$$

Using the Pearson coefficient of skewness gives

$$PC = \frac{3(79.5 - 77.5)}{40.5}$$
$$= 0.148$$

The PI is not greater than +1 or less than -1, so it can be concluded that the distribution is not significantly skewed.

Example: Technology Inventories…

Method 3: Check for Outliers

Lowest –	Q1 –	MD –	Q3 –	Highest
5	45	77.5	98	158

Q1 – 1.5(IQR) = 45 – 1.5(53) = -34.5

Q3 + 1.5(IQR) = 98 + 1.5(53) = 177.5

No data below -34.5 or above 177.5, so **no outliers**.

Example: Technology Inventories…

Conclusion:

- The histogram is approximately bell-shaped.
- The data are not significantly skewed.
- There are no outliers.

Thus, it can be concluded that the distribution is approximately normally distributed.

Probability Plots

- Sometimes, the only way to determine an appropriate distribution is to examine the sample to find a probability distribution that fits.
- Probability plots provide a good way to do this.
- Given a random sample $X_1, \ldots, X_n$, a probability plot can determine whether the sample might plausibly have come from some specified population.

Probability Plots…

- The probability plot consists of the points **(X_i, Q_i)** where **X_i** is a random variable and **Q_i** is typically a percentile.
- If the distribution that generated the Q_i is a normal distribution, it is called a normal probability plot.
- If $X_1, \ldots, X_n$ do in fact come from the distribution that generated the Q_i, the points should lie close to a straight line.
- To construct the plot, we must compute the Q_i.
- In practice, the **Q_i** are invariably calculated by a computer software package.

Example (p.288)

- A random sample of size 5 is drawn. We want to determine whether the population from which it came might have been normal.
- The sample, arranged in increasing order, is

  3.01, 3.35, 4.79, 5.96, 7.89

- $\bar{X}$ = 5.00 and s = 2.00.
- Q_i values are the **100(i – 0.5)/n percentiles** of the distribution that is suspected of generating the sample
- Thus, the Q_i are the 10th, 30th, 50th, 70th, and 90th percentiles of the N(5, 2^2) distribution

i	X_i	Q_i
1	3.01	2.44
2	3.35	3.95
3	4.79	5.00
4	5.96	6.05
5	7.89	7.56

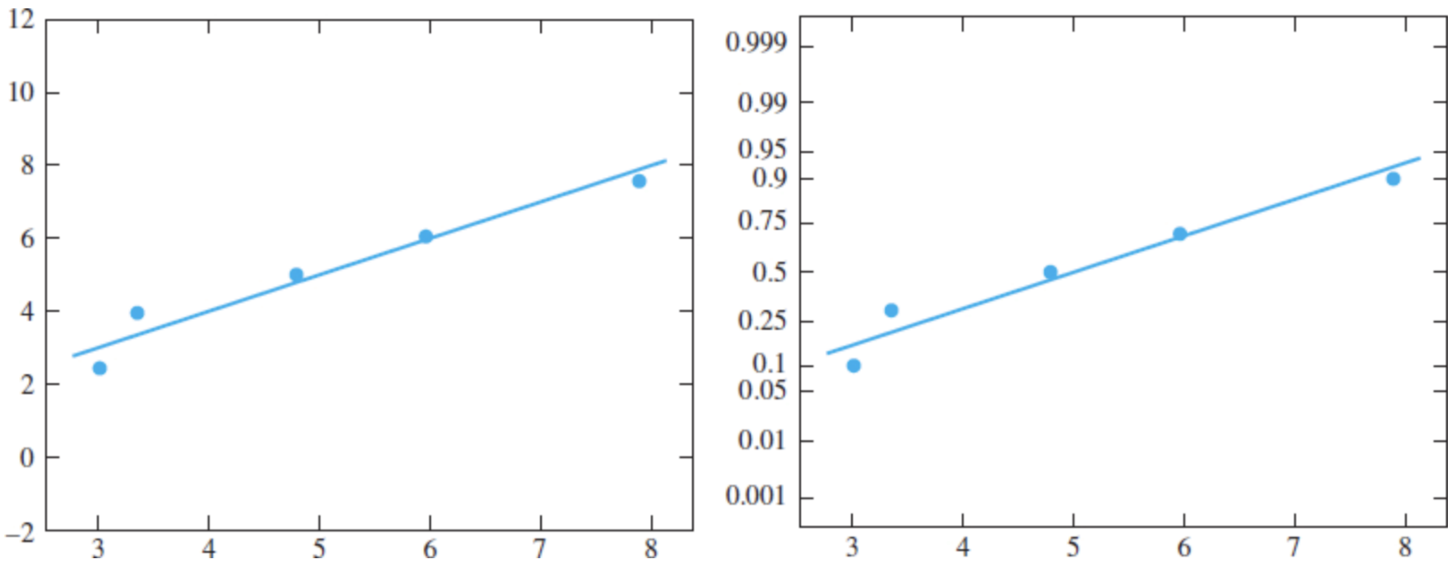


FIGURE 4.22 Normal probability plots for the sample $X_1, \ldots, X_5$. The plots are identical, except for the scaling on the vertical axis. The sample points lie approximately on a straight line, so it is plausible that they came from a normal population.

Example…

- The points Q1, …, Q_n are called quantiles of the distribution from which they are generated. Sometimes the sample points X_1, …, X_n are called empirical quantiles.
- For this reason, the probability plot is sometimes called a **quantile-quantile** plot, or **QQ plot**.
- In this example, we used a sample of only five points to make the calculations clear.
- In practice, probability plots work better with larger samples.
- A good rule of thumb is to require **at least 30 points** before relying on a probability plot.

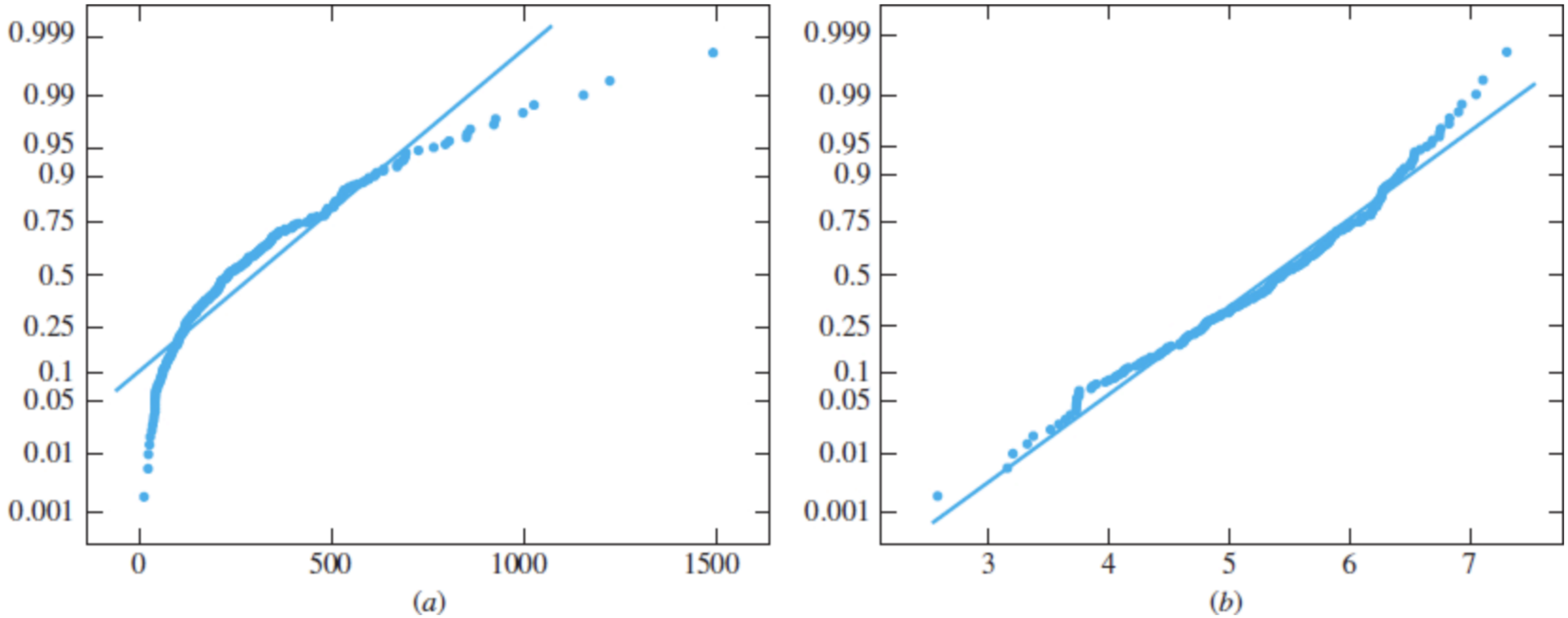


FIGURE 4.23 Two normal probability plots. *(a)* A plot of the monthly productions of 255 gas wells. These data do not lie close to a straight line, and thus do not come from a population that is close to normal. *(b)* A plot of the natural logs of the monthly productions. These data lie much closer to a straight line, although some departure from normality can be detected. See Figure 4.16 for histograms of these data.

Interpreting Probability Plots

- It is best to judge the straightness of the plot by eye.
- When deciding whether the points on a probability plot lie close to a straight line or not, do not pay too much attention to the points at the very ends (high or low) of the sample, unless they are quite far from the line.
- It is common for a few points at either end to stray from the line somewhat.
- However, a point that is very far from the line when most other points are close is an outlier, and deserves attention.

Exercise 1 (p.289)

Each of three samples has been plotted on a normal probability plot. For each, say whether the sample appears to have come from an approximately normal population.