Chapter 3. The Normal Distributions

Density Curves

Definition. A (probability) *density curve* is a curve that

- is always on or above the horizontal axis, and

- has area exactly 1 underneath it (that is, the area between the curve and the $x-$axis).

A density curve describes the overall pattern of a distribution. The area under the curve and above any range of values is the proportion of all observations that fall in that range.

Example. Exercise 3.1 page 67.

Describing Density Curves

Definition. The **median** of a density curve is the equal-areas point, the point that divides the area under the curve in half. That is, half of the area under the curve is to the left of the median and the remaining half of the area is to its right. The **mean** of a density curve is the balance point, at which the curve would balance if made of solid material. The median and mean are the same for a symmetric density curve.
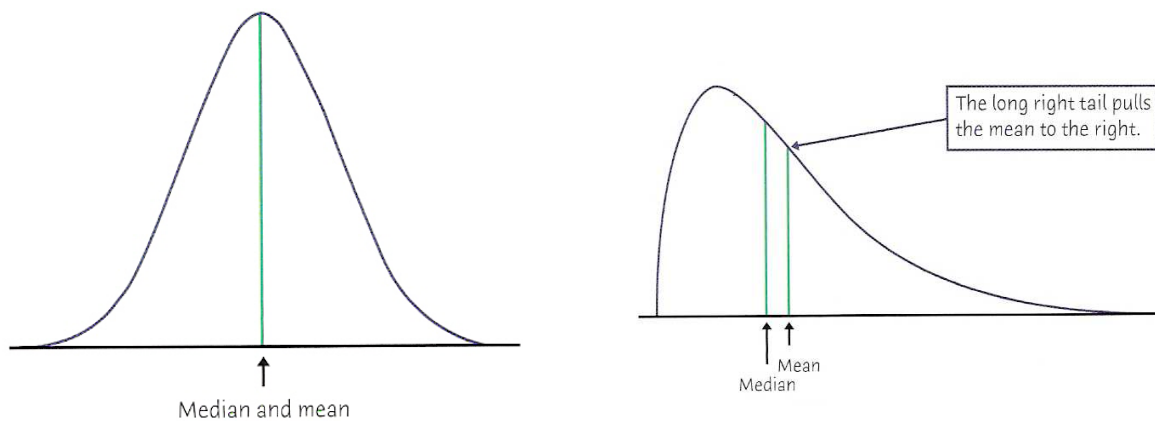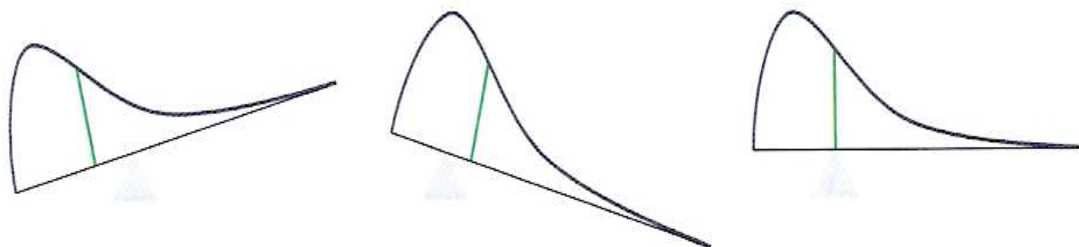

Figures 3.4(a) and 3.4(b) page 68.

Figure 3.5 page 68.

Note. The usual notation for the mean of an idealized distribution is μ (mu). The standard deviation of a density curve is denoted σ (sigma).

Example. Exercise 3.4 page 69.

Normal Distributions

Note. A VERY common class of density curves is the *normal distributions.* These curves are symmetric, single-peaked, and bell-shaped. All normal distributions have similar shapes and are determined solely by their mean μ and standard deviation σ. The points at which the curves change concavity are located a distance σ on either side of μ. We will use the area under these curves to represent a percentage of observations. (These areas correspond to integrals, for those of you with some experience with calculus.)
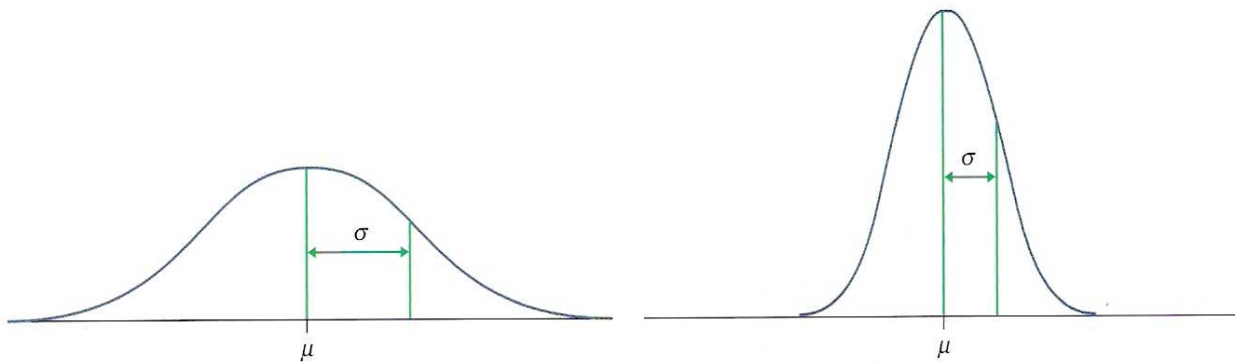
Figure 3.8 page 70.

Note. The text mentions three reasons we are interested in normal distributions:

1. Normal distributions are good descriptions for some distributions of real data. Examples include test scores and characteristics of biological populations (such as height or weight).

2. Normal distributions are good approximations to the results of many kinds of chance outcomes. An example is the proportion of heads in a repeatedly tossed coin experiment.

3. Many statistical inference procedures based on normal distributions work well for other roughly symmetric distributions.

The 68–95–99.7 Rule

Note. In the normal distribution with mean μ and standard deviation σ:

- Approximately 68% of the observations fall within σ of the mean μ.

- Approximately 95% of the observations fall within 2σ of μ.

- Approximately 99.7% of the observations fall within 3σ of μ.
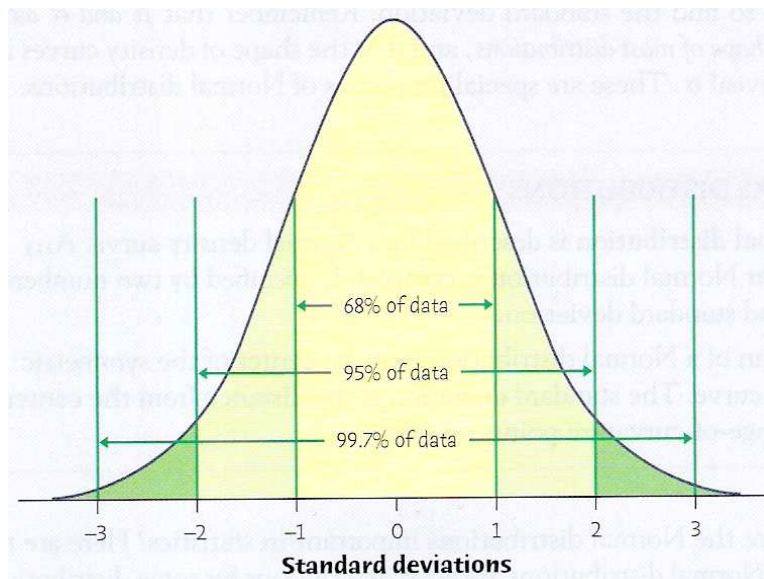
This is called the "68-95-99.7 Rule."

Figure 3.9 page 72.

Example. Exercise 3.7 page 74.

Example S.3.1. (Ab)Normal Stooges.

Suppose the the number of slaps per film in the Three Stooges 25 years of shorts is a normally distributed variable with mean $\mu = 12.95$ slaps per films and standard deviation $\sigma = 4.50$ per film. (These are the mean and standard deviation we would compute from the formulas of Chapter 2, if we remove the two outliers from the annual slaps per film averages of problem S.1.2. However, the supposition of a normal distribution might be questionable.) Then:

(a) 84% of the films include at most how many slaps per film?

(b) The top 16% of the films include at least how many slaps per film?

(c) The top 2.5% of the films include at least how many slaps per film?

(d) What is the range of the number of slaps per film for the center 95% of the films?

Note. We denote the normal distribution with mean μ and standard deviation σ as $N(\mu, \sigma)$. For example, the supposed distribution of Example S.3.1 is $N(12.95, 4.50)$.

The Standard Normal Distribution

Definition. If x is an observation from a distribution that has mean μ and standard deviation σ, the **standard value** of x is

$$z = \frac{x - \mu}{\sigma}.$$

This value is sometimes called a $z-$**score**.

Example. S.3.2. Stooge Percentiles.

In the Three Stooges short "Three Hams on Rye" (short number 125), there are 22 slaps (according to Fleming's *The Three Stooges—An Illustrated History*). What is the z-score for this film if the number of slaps per film has distribution $N(12.95, 4.50)$? Approximately what percentage of the Stooges' films has less slaps than "Three Hams on Rye" (this is the **percentile** ranking of this film)? Approximately what percentage has more?

Definition. The *standard normal distribution* is the normal distribution $N(0, 1)$ with mean 0 and standard deviation 1. If a variable x has any normal distribution $N(\mu, \sigma)$ with mean μ and standard deviation σ, then the standardized variable $z = \dfrac{x - \mu}{\sigma}$ has the standard normal distribution.

Note. The standard normal distribution allows to compare numbers from two different populations in terms of a common unit (the standard deviation). For example, Exercise 3.8 involves comparing SAT and ACT scores.

Example. Exercise 3.9 page 76.

Finding Normal Proportions

Note. Since the normal distribution is a probability distribution and since areas under a probability distribution represent probabilities, the total area under a normal distribution must be 1. In general, areas under the normal distribution represent proportions of a population. For example, about 95% of the population lies within 2 standard deviations of the mean in a normal distribution, so the area under a normal distribution between $\mu - 2\sigma$ and $\mu + 2\sigma$ is 0.95.

Definition. The **cumulative proportion** for a value x in a distribution is the proportion of observations in the distribution that lie at or below x.
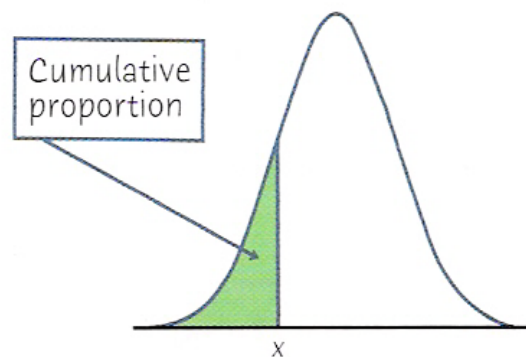

Figure from page 76.

Note. Unfortunately, it is impossible to algebraically calculate cumulative proportions under the normal distribution. Therefore we will have to rely on calculators, software, or tables.

Example S.3.3. More Stooge Percentiles.

In the Three Stooges short "Restless Knights" (short number 6), there are 20 slaps (according to Fleming's *The Three Stooges—An Illustrated History*). Assume that the number of slaps per film has distribution $N(12.95, 4.50)$. Use Minitab's Cumulative Distribution Function to answer the following. Approximately what percentage of the Stooges' films has less slaps than "Restless Knights"? Approximately what percentage has more?

Solution. The Minitab commands to access this are:

1. Click on the `Calc` pulldown menu.

2. Click on `Probability Distributions`.

3. Click on `Normal`

4. To answer the "less than" part of the question, select the `cumulative probability` option.

5. Enter the mean and standard deviation as given.

6. Select **Input constant** and enter 20.

The Minitab output will include the following:

```
   x  P( X<= x )
  20     0.941404
```

So the answer to the "less than" part is $0.9414 = 94.14\%$. The answer to the "greater than" part is $100\% - 94.14\% = 5.86\%$.

Using the Standard Normal Table

Note. We can use z-scores and a standard normal distribution table to find cumulative proportions.

Example S.3.4. Stooge Percentiles—Tables.

In the second-to-the-last Three Stooges short "Triple Crossed" (short number 189), there are 11 slaps (according to Fleming's *The Three Stooges—An Illustrated History*). What is the z-score for this film if the number of slaps per film has distribution $N(12.95, 4.50)$? Use the z-score and the Standard Normal Distribution (Table A page 684) to answer the following. Approximately what percentage of the Stooges' films has less slaps than "Triple Crossed"? Approximately what percentage has more?

Solution. The z-score for $x = 11$ is

$$z = \frac{x - \mu}{\sigma} = \frac{11 - 12.95}{4.50} = -0.4333.$$

Table A has z-scores to the nearest 0.01. So we round our z-score to $z = -0.43$. The percentage of the films with less slaps than "Triple Crossed" is the cumulative proportion of the standard normal distribution corresponding to $z = -0.433$. This is precisely the entry given in Table A corresponding to

$z = -0.43$. To find the appropriate entry, we read *down the first column* of Table A to the entry -0.4 and *over the first row* to .03. The relevant entry is 0.3336. So the percentage of films with less slaps that "Triple Crossed" is $100 \times 0.3336 = 33.36\%$. The percentage of films with more slaps than "Triple Crossed" is $100\% - 33.36\% = 66.64\%$ (notice that this number is not directly in Table A, but can be computed from numbers in the table).

Note. The protocol for finding normal proportions with Table A is:

- State the problem in terms of the observed variable x. **Draw a picture** that shows the proportion you want in terms of cumulative proportions.

- **Standardize** x to restate the problem in terms of a standard normal variable z.

- **Use Table A** and the fact that the total area under the curve is 1 to find the required area under the standard normal curve.

Example. Exercise 3.45 page 87.

Finding a Value Given a Proportion

Note. Instead of calculating proportions from Table A, we might be given the proportion of a population below a certain unknown value, and asked to find that value. To carry this out, we must use Table 2 backwards. We illustrate this with an example.

Example. Exercise 3.40 page 87. Notice that the instructions say that the distribution for SAT scores is $N(1026, 209)$. Use Table A.

Example S.3.5. How Many Slaps?

If the number of slaps per film in the Three Stooges shorts has distribution $N(12.95, 4.50)$, then what are the number of slaps in the top 5% of their films?

Partial Solution. This can be solved with Minitab using the `Inverse cumulative probability` option of the `Calc`, `Probability Distributions`, `Normal` sequence mentioned above. Select `Input constant` and enter 0.95 (since the software gives cumulative probabilities, the upper 5% cor-

responds to the z-score 0.95). The software outputs a value of `20.3518`. That is, the upper 5% of the films contain at least 20.35 slaps each.