# Today's Topics

- Percentile ranks and percentiles
- Standardized scores
- Using standardized scores to estimate percentiles
- Using $\mu$ and $\sigma_x$ to learn about percentiles

# Percentile rankings and percentiles

It is very common in our society for us to be placed in *percentile rankings*; for example, as high school students, we all took standardized tests, and our scores were compared with those of our peers.

For example, on the ACT and/or SAT, our scores on various parts of the exam were used to evaluate our ranking relative to our peers. On our score reports, they would say we fell into the "75th percentile", which is a fancy way of saying "Mary's score was better than those of 75 percent of her peers."

Percentiles tell us our relative position in a given sample or population.

# How do we figure out percentiles?

Percentiles are pretty easy to figure out; in fact, if you can figure out the median, you can figure out percentiles for any member of a data set. The procedure is pretty simple:

1. Arrange the scores from highest to lowest.
2. Look for the item we want to find the percentile of.
3. Count how many items are below that one.
4. Do a little math: $P_r = \frac{B}{N} \times 100$, where $B$ is the number of items **below** the item of interest, and $N$ is the total number of items.

# Cheap example

For example, look at Table 6.2. To figure out Anna's proficiency relative to her peers at video games, all we have to do is:

1. Arrange the scores. (Done for us already!)

2. Look for Anna. (She's number 8)

3. Count how many people are below Anna. (12)

4. Figure out $P_r = \frac{B}{N} \times 100 = \frac{12}{20} \times 100 = 60.$

This means Anna falls in the 60th percentile.

# Another example

Sometimes, percentiles alone can be misleading. Let's take a look at Table 6.3 and figure out where Sarah fits in.

1. Arrange the scores. (Done for us already!)

2. Look for Sarah. (She's number 9)

3. Count how many people are below Anna. (Just Michael: 1)

4. Figure out $P_r = \frac{B}{N} \times 100 = \frac{1}{10} \times 100 = 10$.

This makes it look like Sarah isn't a very good student. But, she got a C on the test. So maybe we'll cut her some slack.

# Another way to look at percentiles

We can also rank scores using *standardized scores* and the *standard normal distribution*.

The standard normal distribution is a theoretical shape that, given enough data ("large $N$") and good enough measurement, describes the frequency distribution of any data set. Sometimes it's called a "bell curve" or the "Gaussian distribution" for the mathematician who discovered it. We will also call this the Z-score (or Z) distribution.

Carl Freidrich Gauss was a German mathematician who lived in the late 18th and early 19th centuries. He discovered most of the fundamental concepts behind statistics, including the normal curve and linear regression. Carl was a really smart guy.

# Why we studied Chapters 4 and 5

Figuring out where observations are located in the normal curve requires two quantities: the mean ($\mu$) and the standard deviation ($\sigma$) of the data. With these quantities, we can calculate a *Z score* for every possible observation in a data set, which will help tell us its ranking.

# The Z Score Formula

The formula for calculating a Z score for any given X is actually pretty simple:

$$Z_X = \frac{X - \mu}{\sigma}$$

(Note that we are talking about *population parameters* here ($\mu$ and $\sigma$, not $\bar{x}$ and $s$). With a big enough sample, we can usually treat a sample as a population, but we'll worry about that later!)

# A Smart Example

For example, a good place to start is with IQ scores. All IQ scores are normalized to have a mean ($\mu$) of 100 and a standard deviation ($\sigma$) of 15. So, we can figure out a Z score for any IQ score, simply by plugging into our formula. Let's figure out the Z score for someone with an IQ of 130:

$$\frac{X-\mu}{\sigma} = \frac{130-100}{15} = \frac{30}{15} = +2.$$

We would say that this person has an IQ score "two standard deviations above the mean." (We can also do this for any other IQ value.)

# Converting Z scores to percentiles

Unfortunately, "two standard deviations above the mean" is gobbledygook to most people, so we need to translate that into numbers people understand. We do that by finding the likelihood of a particular Z score being observed.

So how do we do that? We locate the Z score on a normal curve, and then figure out the area under that curve to the left of that point.

This is simplified because the normal curve is symmetrical: half of the area is left of the middle, and half of it is to the right. Furthermore, the standard normal curve is really nice because the area under the curve adds up to one. (Look at Figure 6.1.)

# Workin' the normal curve

For example, let's figure out the percentage of the population that has an IQ less than 85. If we calculate the Z score, we'll find that it's -1. Let's translate that into a percentile.

Look at Figure 6.1 again. Locate -1 on the bottom axis. Now, we know the area between -1 and 0 is .3413. Since we know that half (.5) of the area is to the left of 0, we can figure out that the area to the left of -1 is $.5 - .3413 = .1587$ or 15.87%. That is, approximately 15.87 percent of the population has an IQ less than 85.

# What if it's not in Figure 6.1?

Often, our Z scores won't be nice round numbers. While we could learn calculus to figure out the weird numbers, there's an easier way. Look at Appendix 1. We can look up any value of Z from 0 to 3.7, and instantly figure out the area between the mean and Z, and the "big" part and "small" part of the area under the curve.

This table is your new best friend. You will treat it nicely.

# Let's work on an example

Let's start with the example that starts on page 88. Schacht gives us two parameters about womens' heights: $\mu = 64$ and $\sigma = 2.4$. From just these two pieces of information, we can find out all sorts of useful information.

For example, say we're recruiting for the Ole Miss Women's Basketball Team, and we need to know what percentage of women are over 70 inches (5' 10") tall for some nefarious reason. We can do this, or any other problem, in five easy steps:

# Five Easy Steps

1. Find the Z score.
2. Draw a normal curve and show where the Z score appears on it.
3. Shade the area under the curve we're interested in.
4. Look up the Z score in Appendix 1.
5. Identify which column best answers our question.

# Systematic height discrimination

The first step is to figure out the Z score:

$$Z_X = \frac{X - \mu}{\sigma} = \frac{70 - 64}{2.4} = 6/2.4 = 2.5.$$

Next, let's draw the normal curve, and shade the bit above $Z_X = 2.5$ (i.e. shade $Z_X \geq 2.5$). Now, we look up the Z score. Finally, figure out which of the three things (the distance from the mean, the "big part", or the "small part") seems to apply. Our pretty curve indicates that we want the "small part." The table says the value is 0.0062, or 0.62%. That means that just over six women per thousand are over five-ten.

# More fun with the vertically challenged

We can do this with any height. We can also do some more interesting things. For example, we can figure out what percentage of women are over 61 inches tall, or what percentage are between 62 and 64 inches tall, or even what percentage are between 63 and 67 inches tall. Let's look at the last possibility.

This one seems a bit more complex, but our handy five-step program will still do the trick.

# We can also go the other way

Not only can we find out what percentage of the population falls in a certain category; we can also find out what value corresponds to a particular percentile. For example, we can find out what height 25% of women are shorter than, or the minimum IQ to get into Mensa (98% of the population has an IQ too low).

To figure these things out, we need a second formula:

$$X = Z\sigma + \mu$$

To use this formula, we simply employ the reverse of our earlier procedure.

# Going the other way: a guide

Here are the steps:

1. Draw a normal curve, shading the area of interest.
2. Based on the shading, find the percentage value in Column 3 or Column 4 of Appendix 1.
3. Find the Z score corresponding to the percentage.
4. Plug it in, and find X.

Let's work through a couple of examples.

# Homework

Read through the end of the chapter. Note that $A - 1$ in Figure 6.11 means "Appendix 1"; it is not a new formula.

Homework: Page 100, questions 2 and 3 (all parts).

Next time: we get our gambling jones on.