Estimating Percentiles with Linear Interpolation

Suppose we have the following really simple frequency distribution, and we wish to estimate the score that is at the 90th percentile:

X	f
81–90	8
71–80	12
61–70	12
51–60	8

Estimating, not Computing

We cannot “compute” the 90^{th} percentile, because the grouped frequency distribution does not contain all the information about the original numbers. We need to try to reconstruct the missing information, but doing so will require some assumptions, which raise the possibility that we will make a nontrivial error.

Estimating Percentiles

$$
\begin{aligned}
P_a &= X_L + (i/f_i)(\text{cum } f_P - \text{cum } f_L) \\
&= X_L + (i/f_i)(Na - \text{cum } f_L) \\
&= X_L + \left(\frac{Na - \text{cum } f_L}{f_i} \right) i
\end{aligned}
$$

Begin by computing Na, the cumulative frequency at the desired percentile point. For example, when $N = 40$ and you require the 90$^{\text{th}}$ percentile, $a = .90$ and $Na = 36$. Thus, you are estimating the value at which the cumulative frequency is 36.

$$P_a = X_L + \left(\frac{Na - \text{cum}\, f_L}{f_i} \right) i$$

The remaining quantities are:

i = Interval width in the "key interval"

cum f_L = cumulative frequency up to, but not including, the key interval

f_i = frequency in the key interval

X_L = the lower real limit of the key interval

$$P_a = X_L + \left(\frac{Na - \text{cum } f_L}{f_i} \right) i$$

X	f	cum f	cum %	Real Limits
81–90	8	40	100	80.5–90.5
71–80	12	32	80	70.5–80.5
61–70	12	20	50	60.5–70.5
51–60	8	8	20	50.5–60.5

$$P_a = 80.5 + \left(\frac{36 - 32}{8} \right) 10 = 85.5$$

The Linear Interpolation Approach

X	*f*	cum *f*	cum %	Real Limits
81–90	8	40	100	80.5–90.5
71–80	12	32	80	70.5–80.5
61–70	12	20	50	60.5–70.5
51–60	8	8	20	50.5–60.5

Simplify

X	*f*	cum *f*	cum %	Upper Real Limits
81–90	8	40	100	90.5
			90	???
71–80	12	32	80	80.5
61–70	12	20	50	60.5–70.5
51–60	8	8	20	50.5–60.5

Interpolate the missing value!

Another Try

X	*f*	cum *f*	cum %	Real Limits
81–90	8	40	100	80.5–90.5
71–80	12	32	80	70.5–80.5
61–70	12	20	50	60.5–70.5
51–60	8	8	20	50.5–60.5

Estimate P_{60}

Going the Other Way

Later in the textbook chapter, there is a formula for "computing of percentile rank" on page 50. Actually, of course, we are *estimating* percentile rank via linear interpolation. The textbook works a pair of elaborate problems on pages 50–52, using the formula on page 51. We'll try Practice Problem 3.5.

We can reproduce their solutions using a simple linear interpolation approach. On the next page, I'll reproduce *part* of Table 3.9 from page 52.

Class Interval	*f*	Cum *f*	Cum %	Upper Real Limit
65–69	7	18	25.71	69.5
60–64	4	11	15.71	64.5
55–59	4	7	10.00	59.5
			??	59.0
50–54	2	3	4.29	54.5
45–49	1	1	1.43	49.5

Our job is to estimate, via linear interpolation, the value of Cum % corresponding to a score of 59. I've highlighted in red the values that bracket the one we are looking for, and added an extra row to the table.

Completing the Interpolation

10.00	59.5
??	59.0
4.29	54.5

Note that 59 is nine-tenths of the way from 54.5 to 59.5. So the ?? value must be nine-tenths of the way from 4.29 to 10. The distance from 4.29 to 10 is 10 – 4.29 = 5.71.

So we take nine-tenths of this distance and add it to 4.29.

$$4.29 + (9/10)5.71 = 9.43$$

Another Approach

10.00	59.5
??	59.0
4.29	54.5

Alternatively, we could go one-tenth of the distance down from 10 toward 4.29. One-tenth of 5.71 is easy to compute, i.e., .571, and 10 – .571 = 9.429, the same answer.

Practice Problem 3.6

You are asked to estimate the percentile rank of a score of 117 from the data in Table 3.5. Here is the segment of that table that brackets 117. I've added the upper real limits for each interval. Note that I had to do some work to compute the cumulative frequencies and percentages, but it was made easier by remembering that the highest Cum *f* is *N*.

Class Interval	*f*	Cum *f*	Cum %	Upper Real Limit
110–119	3	86	95.56	119.5
100–109	4	83	92.22	109.5

The linear interpolation reduces to

95.56	119.5
??	117.0
92.22	109.5

The answer is

$$92.2222 + (7.5/10)(95.5555 - 92.2222)$$

$$92.2222 + (.75)(3.3333)$$

$$92.2222 + 2.5000$$

$$94.7222$$