Percentiles from z-scores (and vice versa)

Colm Mulcahy

Math 107-03, Spring 2020, Spelman College

17 Apr 2020

Review of z-scores for normal data

Assume X = N(mean, std dev), that is to say, X data is normally distributed with the given statistics.

We have been studying standard "scores" (or z-scores)

$$Z = \frac{X - Mean}{Std\ Dev}$$

which counts "standard deviations above or below the mean".

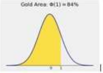
It's very useful, because it "puts things on a level playing field." It allows us to compare multiple data sets that are on different scales.

When X is normally distributed, namely X = N(mean, std dev), then $Z = \frac{X - Mean}{Std\ Dev}$ is N(0,1), i.e., the corresponding standard scores are normally distributed, with mean 0 and std dev 1.

Moreover, there is a single lookup table for percentiles for Z.

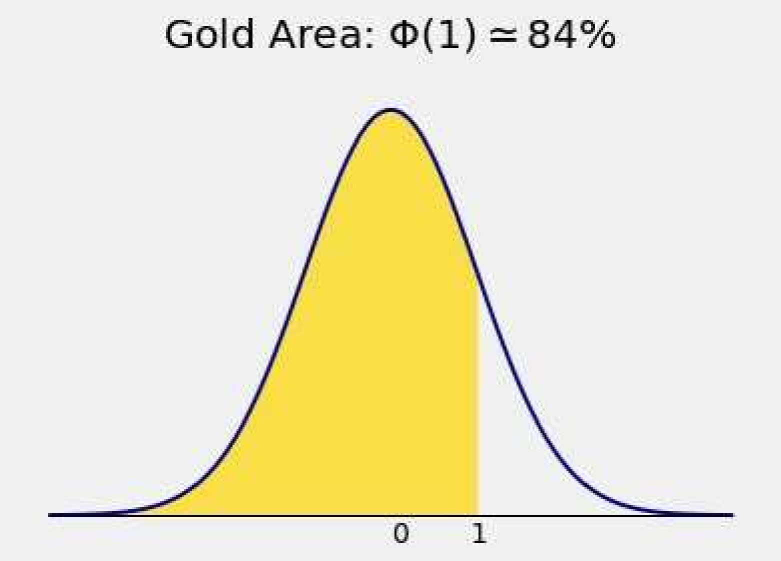
The Conversion Table

z-scores and percentiles



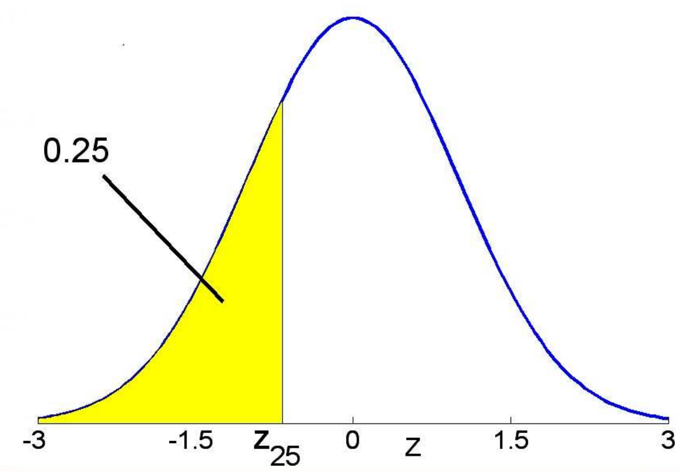
Z*score	Percentile	Z*score	Percentile	Z*score	Percentile	Z*score	Percentile
−3.5	0.02	−1.0	15.87	0.0	50.00	1.1	86.43
−3.0	0.13	−0.95	17.11	0.05	51.99	1.2	88.49
−2.9	0.19	−0.90	18.41	0.10	53.98	1.3	90.32
−2.8	0.26	−0.85	19.77	0.15	55.96	1.4	91.92
−2.7	0.35	−0.80	21.19	0.20	57.93	1.5	93.32
−2.6	0.47	−0.75	22.66	0.25	59.87	1.6	94.52
−2.5	0.62	−0.70	24.20	0.30	61.79	1.7	95.54
−2.4	0.82	−0.65	25.78	0.35	63.68	1.8	96.41
−2.3	1.07	−0.60	27.43	0.40	65.54	1.9	97.13
−2.2	1.39	−0.55	29.12	0.45	67.36	2.0	97.72
−2.1	1.79	−0.50	30.85	0.50	69.15	2.1	98.21
−2.0	2.28	−0.45	32.64	0.55	70.88	2.2	98.61
−1.9	2.87	−0.40	34.46	0.60	72.57	2.3	98.93
−1.8	3.59	−0.35	36.32	0.65	74.22	2.4	99.18
−1.7	4.46	−0.30	38.21	0.70	75.80	2.5	99.38
−1.6	5.48	−0.25	40.13	0.75	77.34	2.6	99.53
−1.5	6.68	−0.20	42.07	0.80	78.81	2.7	99.65
−1.4	8.08	−0.15	44.04	0.85	80.23	2.8	99.74
−1.3	9.68	−0.10	46.02	0.90	81.59	2.9	99.81
−1.2	11.51	−0.05	48.01	0.95	82.89	3.0	99.87
−1.1	13.57	−0.0	50.00	1.0	84.13	3.5	99.98

From z-scores to percentiles, visually



A z-score of 1 corresponds to the 84th percentile, that's the percentage of z-values equal to or less than 1.

From percentiles to z-scores, visually



What z-score corresponds to the 25th percentile? This can be read off the table, "by reading it backwards". It's near -0.65.

Key Observation

Our earlier mastering of the 68% and related rules was based on symmetric ranges about the data mean, and often forced us to "think outside the box" to answer some questions.

We kept running into numbers like 50%, 34%, 84%, 16%, 2.5%, etc., as we manipulated 68%, 95% and 99.7%, divided by 2, subtracted from 100% and so on.

When looking up z-scores in the table, we need to change our approach. Now the focus is always on "left tails": namely, the (percentage of) area to the left of a particular z-score.

We need to learn how to think outside the box in a new way!

Also, we won't always find the numbers we want in the table. Generally, we will have to settle for approximate answers, and occasionally average two numbers found in the table.

Using the table to find percentiles

If Bio test scores are B = N(75,5), and Chem test scores are C = N(80,4), then $z_B = \frac{b-75}{5}$ and $z_C = \frac{c-80}{4}$ are both N(0,1).

What percentage of Bio test scores are (at or) below 78? Find $z_B = \frac{78-75}{5} = 0.6$ and the table says about 72.5%. Getting 78 on this Bio test puts one at the 75 and a half-th percentile!

What percentage of Bio test scores are (at or) above 78? In view of what we just did, the other $100\% - 72.5\% = 27.5\%$!

What percentage of Bio test scores are (at or) below 82? Find $z_B = \frac{82-75}{5} = 1.4$ and the table says about 92%. Getting 82 on this Bio test puts one at the 92nd percentile!

What percentage of Bio test scores are (at or) below 70? Find $z_B = \frac{70-75}{5} = -1$ and the table says 15.87% which agrees well with the 16% we would have said by using the 68% rule. Getting 70 on this Bio test puts one at the 16th percentile.

Using the table to find percentiles

What percentage of Bio test scores are betweeen 70 and 82? We are in luck: we just learned that 92% of the scores are at or below 82, whereas 16% are at or below 70. Hence, by subtracting, we see that $92\% - 16\% = 76\%$ of the Bio test scores are between 70 and 82. Note that there is no tie-in with a single percentile here.

What percentage of Bio test scores are betweeen 69 and 72? We have to work with these numbers separately, and then substract. No prior work helps.

Find $z_B = \frac{69-75}{5} = -1.2$. The table says about 11.5%. (Getting 69 on this Bio test puts one at the 11 and a half-th percentile.) Also, find $z_B = \frac{72-75}{5} = -0.6$ and the table says about 27.5%. (Getting 72 on this Bio test puts one at the 27 and a half-th percentile.)

Subtracting, we get $27.5\% - 11.5\% = 16\%$. So, about 16% of the Bio test scores are between 69 and 72. (That 16% has nothing to do with the two 16%s that arise using the 68% rule.)

What to do when the table won't coopoerate

Earlier, we saw that a specific Bio test score of $b = 78$ results in a stardard (or z-score) of $z_B = \frac{78-75}{5} = 0.6667$. Back then, we didn't know what percentile this corresponded to. Now, looking up $z = 0.67$ in the table, we don't find it exactly. However, we see that $z = 0.65$ corresponds to the percentile 74.22 and $z = 0.70$ corresponds to the percentile 75.80. So a reasonable compromise answer is the 75th percentile.

The 75th percentile (the third quartile) always corresponds to about 2/3 of a standard deviation above the mean!

If English test scores are E = N(87,3), then what English test score corresponds to the 75th percentile? One which is 2/3 std dev above the mean: namely $87 + (2/3)3 = 87 + 2 = 89$.

Question: What percentile corresponds to about 2/3 of a standard deviation below the mean? (The 25th percentile.)

Using the table to find z-scores

Assume the price of a certain type of Audi is A = N(26,4) (in thousands of dollars).

What Audi price corresponds to the 60th percentile?

The table "read backwards" reveals that $z = 0.25$ corresponds to the 60th percentile. Hence, the answer is 0.25 std dev above the mean price for the Audi, namely \$26K + \$1K = \$27K. Note that $a = 27$ yields $z_A = \frac{27-26}{4} = 0.25$ too.

Here's another way to do this problem: set the Audi z-score equal to the -0.85 found in the table (using A for the to-be-found Audi price) and solve for A using algebra. Namely, write

$$\frac{A - 26}{4} = 0.25.$$

So, $A - 26 = (0.25)(4)$ and $A = 26 + (0.25)(4) = 27$.

Using the table to find z-scores

Assume the price of a certain type of Mercedes is M = N(35,3) (in thousands of dollars).

1. What Merc price corresponds to the 20th percentile?

The table "read backwards" reveals that a z-score of -0.85 corresponds to the 20th percentile. Hence, the answer is 0.85 std dev below the mean price for the Merc, namely \$35K - (0.85)\$3K = \$32.45K. Note that $m = 32.45$ yields $z_A = \frac{32.45-35}{3} = -0.85$.

Here's another way to do this problem: set the Merci z-score equal to the 0.25 found in the table (using M for the to-be-found Merc price) and solve for M using algebra. Namely, write

$$\frac{M-35}{3} = -0.85.$$

So, $M - 35 = (-0.85)(4)$ and $M = 35 - (0.85)(3) = 32.45$.

Problems about heights for you to try

Assume that in Georgia adult men's heights (in inches) are N(69.2,3.1) and that adult women's heights are N(64.5,2.6).

1. A 6 foot tall man is at what percentile? (Convert to inches first. Hint: get a z-score of about 0.90)

2. A 5 foot tall woman is at what percentile? (Convert to inches first. Get a z-score of about -1.73, which is between two values in the table. About the 4th percentile seems reasonable.)

3. What is the 70th percentile height for women? (Hint: get a z-score of about halfway between 2.1 and 2.2, so use $z = 2.15$)

4. What is the 70th percentile height for men? (Hint: again, use $z = 2.15$, but what you do next is different.)

5. What is the 8th percentile height for men? (Hint: get a z-score of about -1.40)

HWs from the textbook

There are more worked examples in the book (and on the web).

Make sure you can do 21-47 on page 399–400.

You can also ask and answer any kind of questions (like the ones done today) for the other examples studied earlier. Baby weights, baby arrival times, etc.