Standard Scores for the Normal Curve

Colm Mulcahy

Math 107-03, Spring 2020, Spelman College

13 Apr 2020

Key observation about the 68%/95%/99.7% rules

The only kinds of problems we have been able to address about symmetric and unimodal dats (namely, Normal or Bell Curve data) have been those that *just happen to involve date cut offs that are either in the middle or exactly one, two, or three standard devations away from the middle*.

That's much too restrictive.

If new born babies arrive on their due date D on average, with std dev 2 weeks, then we saw how about 16% of them arrive at least 2 weeks late. But what percentage arrive at least 1 week late? Or within 3 days of their due date? What arrival date corresponds to the 25th percentile (*aka* first quartile)? Or the 70th percentile?

We will learn how to answer any such question in the next class, by using the table on page 396. Today, we take the first necessary steps, by introducing the standard "score" (or z-score) concept.

Standard “scores” (or z-scores) for normal data

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

For instance we might have test scores X = N(75,5).

The Standard “score” (or z-score) is

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

It counts “standard deviations above or below the mean”.

If Bio test scores are X = N(75,5), then a specific score of $x = 80$ results in a stardard (or z-score) of $z = \frac{x-Mean}{Std\ Dev} = \frac{80-75}{5} = 1$. (That, in turn, corresponds to the 84th percentile.)

If Chem test scores are Y = N(80,4), then a specific score of $y = 84$ results in a z-score of $z = \frac{y-Mean}{Std\ Dev} = \frac{84-80}{4} = 1$. (That's also the 84th percentile.)

Cholesterol Levels Example Revisited

If cholesterol levels X for men aged 18-24 are normally distributed with mean 180 and std dev 40, we write $X = N(180, 40)$.

So for 68% of the men, X lies between $\bar{x} - s$ and $\bar{x} + s$. Hence their cholesterol levels would be between 140 and 220. Put another way, these are the data values corresponding to z-scores between -1 and 1 (check!)

For 95% of the men, X lies between $\bar{x} - 2s$ and $\bar{x} + 2s$. Hence their cholesterol levels would be between 100 and 260. Put another way, these are the data values corresponding to z-scores between -2 and 2 (check!)

Also, for 99.7% of the men, X lies between $\bar{x} - 3s$ and $\bar{x} + 3s$. Hence their cholesterol levels would be between 60 and 300. Put another way, these are the data values corresponding to z-scores between -3 and 3 (check!)

Cholesterol Levels Example Revisited

Furthermore, for 34% of the men, X lies between $\bar{x}$ and $\bar{x} + s$. Hence their cholesterol levels would be between 180 and 220. Put another way, these are the data values corresponding to z-scores between 0 and 1 (check!)

For 13.5% of the men, X lies between $\bar{x} - 2s$ and $\bar{x} - s$. Hence their cholesterol levels would be between 100 and 140. Put another way, these are the data values corresponding to z-scores between -2 and -1 (check!)

What z-scores corerspond to cholesterol levels between 100 and 180? These are z-scores between -2 and 0. That's 47.5% of them.

What z-scores correspond to cholesterol levels above 180? These are the z-scores above 0 (so, the 50% above the median).

If Bio test scores are X = N(75,5) and Chem test scores are Y = N(80,4), and you got 78 on Bio and 82 on Chem, which is better?

Bio: a specific score of $x = 78$ results in a stardard (or z-score) for Bio of $z_B = \frac{78-75}{5} = 0.6$. (No clue what percentile!)

Chem: a specific score of $y = 82$ results in a z-score for Chem of $z_C = \frac{82-80}{4} = 0.5$. (No clue what percentile!)

Which is better and why? The key is to compare the z-scores.

Your Bio test result (though numerically lower than your Chem test result) is more impressive because it has a higher z-score!

The Bio result corresponds to a higher percentile than the Chem one, even though we don't know either of these values.

KEY POINT: *z-scores allow us to "compare apples and oranges", by "putting on a level playing field" numbers from that arise in different contexts.* We just compare the appropriate z-scores.

Again assume Bio test scores are X = N(75,5) and Chem test scores are Y = N(80,4). Your friend got 74 on Bio and 81 on Chem, which is better?

A Bio score of $x = 74$ results in $z_B = \frac{74-75}{5} = -0.2$.

A Chem score of $y = 81$ results in $z_C = \frac{81-82}{4} = 0.5$. (No clue what percentile!)

Which is better and why? The key is to compare the z-scores.

Your Bio test result (though numerically lower than your Chem test result) is more impressive because it has a higher z-score!

The Bio result corresponds to a higher percentile than the Chem one, even though we don't know either of these values.

KEY POINT: *Z-scores allow us to "compare apples and oranges", by "putting on a level playing field" numbers from that arise in different contexts.* We just compare z-scores.

If Bio test scores are X = N(75,5) and Chem test scores are Y = N(80,4), and you got 78 on Bio and 82 on Chem, which is better?

Bio: a specific score of $x = 78$ results in a stardard (or z-score) of $z = \frac{78-75}{5} = 0.6$. (No clue what percentile!)

Chem: a specific score of $y = 82$ results in a z-score of $z = \frac{84-82}{4} = 0.5$. (No clue what percentile!)

Which is better and why? The key is to compare the z-scores.

Your Bio test result (though numerically lower than your Chem test result) is more impressive because it has a higher z-score!

The Bio result corresponds to a higher percentile than the Chem one, even though we don't know either of these values.

KEY POINT: *Z-scores allow us to "compare apples and oranges", by "putting on a level playing field" numbers from that arise in different contexts.* We just compare z-scores.