The Normal Curve 68%, 95% and 99.7% Rules

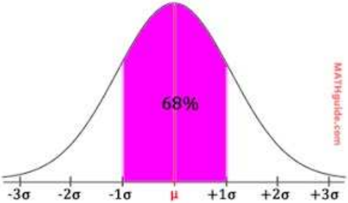
Colm Mulcahy

Math 107-03, Spring 2020, Spelman College

8 Apr 2020

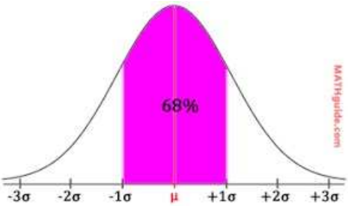
Key fact about standard deviation: the 68% rule

For large data sets, about two thirds (actually 68%) of the data is within one std dev of the mean.



Key fact about standard deviation: the 68% rule

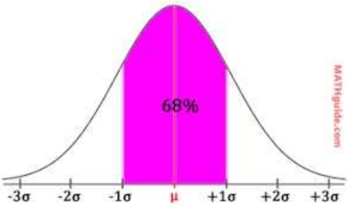
For large data sets, about two thirds (actually 68%) of the data is within one std dev of the mean.



The mean is written $\bar{x}$ or μ ("mu"); the std dev s or σ ("sigma").

Key fact about standard deviation: the 68% rule

For large data sets, about two thirds (actually 68%) of the data is within one std dev of the mean.

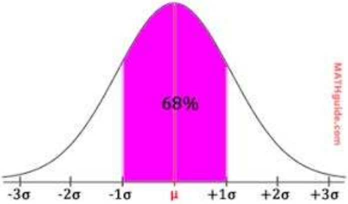


The mean is written $\bar{x}$ or μ ("mu"); the std dev s or σ ("sigma").

We're talking about the middle two-thirds here, assuming the data is symmetric and unimodal, and the mean, median and mode are equal!

Key fact about standard deviation: the 68% rule

For large data sets, about two thirds (actually 68%) of the data is within one std dev of the mean.

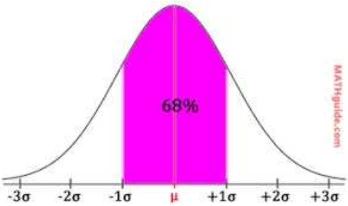


The mean is written $\bar{x}$ or μ ("mu"); the std dev s or σ ("sigma").

We're talking about the middle two-thirds here, assuming the data is symmetric and unimodal, and the mean, median and mode are equal! We speak of the Normal or Bell Curve.

Key fact about standard deviation: the 68% rule

For large data sets, about two thirds (actually 68%) of the data is within one std dev of the mean.



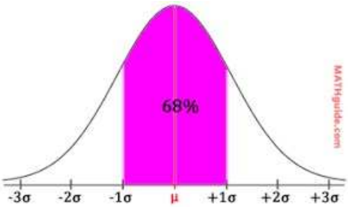
The mean is written $\bar{x}$ or μ ("mu"); the std dev s or σ ("sigma").

We're talking about the middle two-thirds here, assuming the data is symmetric and unimodal, and the mean, median and mode are equal! We speak of the Normal or Bell Curve.

Let's start thinking of the percentage of area under parts of such curves.

Key fact about standard deviation: the 68% rule

For large data sets, about two thirds (actually 68%) of the data is within one std dev of the mean.



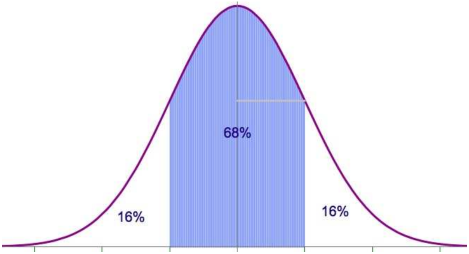
The mean is written $\bar{x}$ or μ ("mu"); the std dev s or σ ("sigma").

We're talking about the middle two-thirds here, assuming the data is symmetric and unimodal, and the mean, median and mode are equal! We speak of the Normal or Bell Curve.

Let's start thinking of the percentage of area under parts of such curves. Hence, what we are saying is that 68% of the area under the Bell Curve lies between $\bar{x} - s$ and $\bar{x} + s$.

Key fact about standard deviation: the other 32%

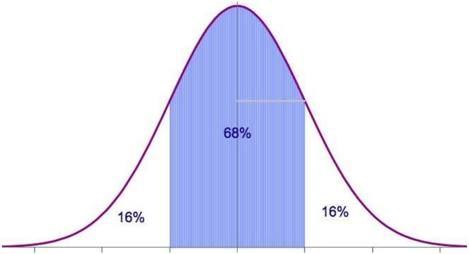
About one third (actually 32%) of the data is NOT within one std dev of the mean.



That's the data lying in one of the two 16% regions on the left and right of the central blue zone here. About 16% of the data is more than one std dev above the mean, and 16% of the data is more than one std dev below the mean.

Key fact about standard deviation: the other 32%

About one third (actually 32%) of the data is NOT within one std dev of the mean.



That's the data lying in one of the two 16% regions on the left and right of the central blue zone here. About 16% of the data is more than one std dev above the mean, and 16% of the data is more than one std dev below the mean. Note the area interpretation.

Top 16% and the bottom 84% (and vice versa)

If new born babies arrive on their due date D on average, with std dev 2 weeks, then about 16% of them arrive at least 2 weeks late.

Top 16% and the bottom 84% (and vice versa)

If new born babies arrive on their due date D on average, with std dev 2 weeks, then about 16% of them arrive at least 2 weeks late.

The others arrive early, on time, or late but no more than 2 weeks late.

Top 16% and the bottom 84% (and vice versa)

If new born babies arrive on their due date D on average, with std dev 2 weeks, then about 16% of them arrive at least 2 weeks late.

The others arrive early, on time, or late but no more than 2 weeks late. Hence 84% ($100\% - 16\%$, or $16\% + 68\%$) arrive no more than 2 weeks late.

Top 16% and the bottom 84% (and vice versa)

If new born babies arrive on their due date D on average, with std dev 2 weeks, then about 16% of them arrive at least 2 weeks late.

The others arrive early, on time, or late but no more than 2 weeks late. Hence 84% ($100\% - 16\%$, or $16\% + 68\%$) arrive no more than 2 weeks late. A baby arriving exactly 2 weeks late would be said to be "at the 84th percentile" for arrivals.

Top 16% and the bottom 84% (and vice versa)

If new born babies arrive on their due date D on average, with std dev 2 weeks, then about 16% of them arrive at least 2 weeks late.

The others arrive early, on time, or late but no more than 2 weeks late. Hence 84% ($100\% - 16\%$, or $16\% + 68\%$) arrive no more than 2 weeks late. A baby arriving exactly 2 weeks late would be said to be "at the 84th percentile" for arrivals.

A baby arriving exactly 2 weeks early would be said to be "at the 16th percentile" for arrivals.

Top 16% and the bottom 84% (and vice versa)

If new born babies arrive on their due date D on average, with std dev 2 weeks, then about 16% of them arrive at least 2 weeks late.

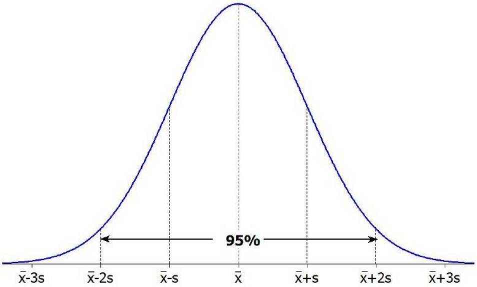
The others arrive early, on time, or late but no more than 2 weeks late. Hence 84% ($100\% - 16\%$, or $16\% + 68\%$) arrive no more than 2 weeks late. A baby arriving exactly 2 weeks late would be said to be "at the 84th percentile" for arrivals.

A baby arriving exactly 2 weeks early would be said to be "at the 16th percentile" for arrivals.

What arrival date corresponds to the 25th percentile (*aka* first quartile)? Or what percentile correponds to a baby arriving a week late? We will get to questions like that soon.

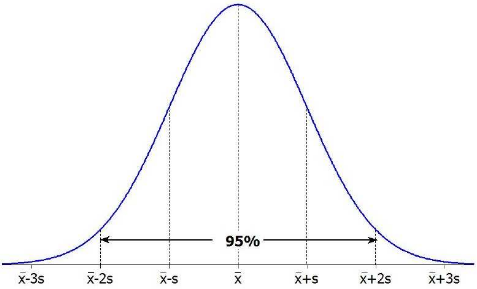
The 95% rule

About 95% of the data is within TWO std devs of the mean.



The 95% rule

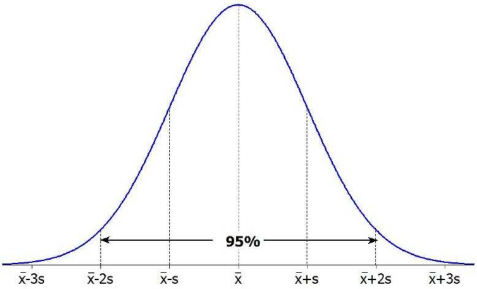
About 95% of the data is within TWO std devs of the mean.



Putting it another way, 95% of the area under the Bell Curve lies between $\bar{x} - 2s$ and $\bar{x} + 2s$.

The 95% rule

About 95% of the data is within TWO std devs of the mean.



Putting it another way, 95% of the area under the Bell Curve lies between $\bar{x} - 2s$ and $\bar{x} + 2s$.

If babies arrive on their due date D on average, with std dev 2 weeks, about 95% of them arrive within 4 weeks of their due date!

The other 5%

The other 5% arrive at least 4 weeks early or weeks late.

The other 5%

The other 5% arrive at least 4 weeks early or weeks late.

Hence 97.5% ($100\% - 2.5\%$, or $2.5\% + 95\%$) arrive early, on time, or late but no more than 4 weeks late.

The other 5%

The other 5% arrive at least 4 weeks early or weeks late.

Hence 97.5% ($100\% - 2.5\%$, or $2.5\% + 95\%$) arrive early, on time, or late but no more than 4 weeks late.

A baby arriving 4 weeks late would be said to be “at the 97 and a half-th percentile” for arrivals.

The other 5%

The other 5% arrive at least 4 weeks early or weeks late.

Hence 97.5% ($100\% - 2.5\%$, or $2.5\% + 95\%$) arrive early, on time, or late but no more than 4 weeks late.

A baby arriving 4 weeks late would be said to be "at the 97 and a half-th percentile" for arrivals.

In the modern age, of course, no pregnant woman is allowed to go over 2 weeks late, so the math model isn't perfect.

The other 5%

The other 5% arrive at least 4 weeks early or weeks late.

Hence 97.5% ($100\% - 2.5\%$, or $2.5\% + 95\%$) arrive early, on time, or late but no more than 4 weeks late.

A baby arriving 4 weeks late would be said to be "at the 97 and a half-th percentile" for arrivals.

In the modern age, of course, no pregnant woman is allowed to go over 2 weeks late, so the math model isn't perfect.

We've studied rules relating to being ONE or TWO std devs away from the mean, and explored them with several examples. There's a rule relating to being THREE std devs away from the mean.

The 99.7% rule for symmetric, unimodal Bell Curve data

About 99.7% of data is within THREE std devs of the mean.

The 99.7% rule for symmetric, unimodal Bell Curve data

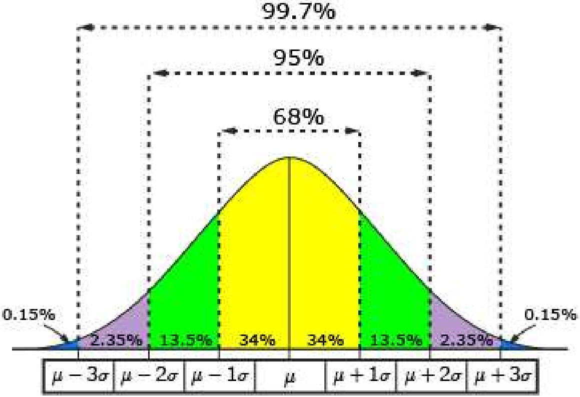
About 99.7% of data is within THREE std devs of the mean.

So, 99.7% of the area under the Bell Curve lies between $\bar{x} - 3s$ and $\bar{x} + 3s$.

The 99.7% rule for symmetric, unimodal Bell Curve data

About 99.7% of data is within THREE std devs of the mean.

So, 99.7% of the area under the Bell Curve lies between $\bar{x} - 3s$ and $\bar{x} + 3s$. All three rules are reflected in a single graphic:



The other 0.3% (about a third of 1%) rule

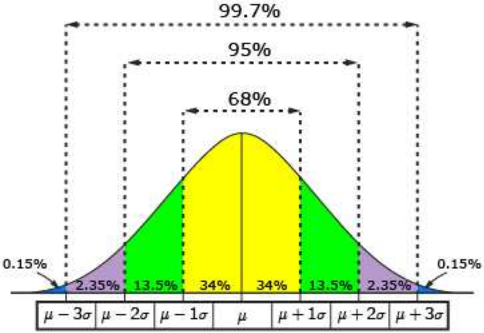
About 0.3% of data is NOT within THREE std devs of the mean.

The other 0.3% (about a third of 1%) rule

About 0.3% of data is NOT within THREE std devs of the mean. That data is "more extreme" and is in two small zones, each representing about 0.15% of the area under the curve, shown in blue in the image.

The other 0.3% (about a third of 1%) rule

About 0.3% of data is NOT within THREE std devs of the mean. That data is "more extreme" and is in two small zones, each representing about 0.15% of the area under the curve, shown in blue in the image.

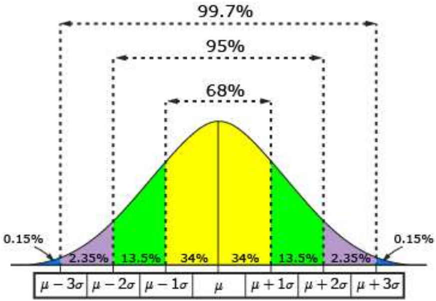


The new number: 13.5%

Look at the two zones shaded in green in the image.

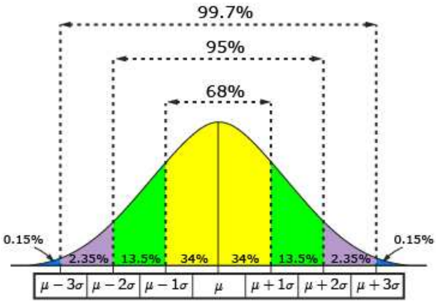
The new number: 13.5%

Look at the two zones shaded in green in the image.



The new number: 13.5%

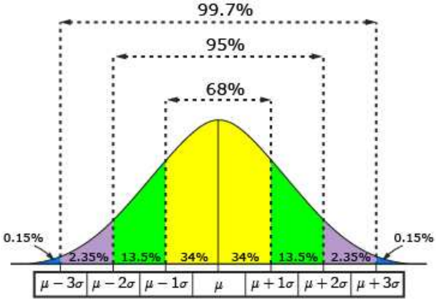
Look at the two zones shaded in green in the image.



The one on the left represents the data between $\bar{x} - 2s$ and $\bar{x} - s$.

The new number: 13.5%

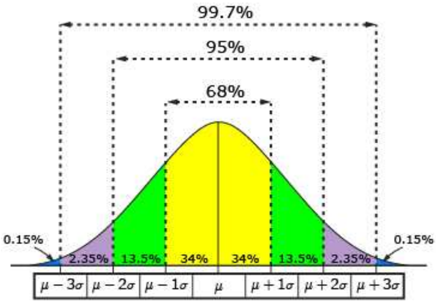
Look at the two zones shaded in green in the image.



The one on the left represents the data between $\bar{x} - 2s$ and $\bar{x} - s$. That's the left half of the data which is at least one std dev away from the mean but no more than two std devs away.

The new number: 13.5%

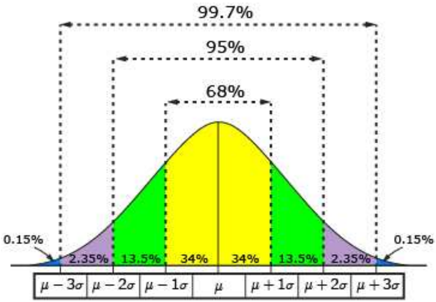
Look at the two zones shaded in green in the image.



The one on the left represents the data between $\bar{x} - 2s$ and $\bar{x} - s$. That's the left half of the data which is at least one std dev away from the mean but no more than two std devs away. Numerically, we are speaking of half of $95\% - 68\% = 27\%$, hence 13.5%.

The new number: 13.5%

Look at the two zones shaded in green in the image.



The one on the left represents the data between $\bar{x} - 2s$ and $\bar{x} - s$. That's the left half of the data which is at least one std dev away from the mean but no more than two std devs away. Numerically, we are speaking of half of $95\% - 68\% = 27\%$, hence 13.5%.

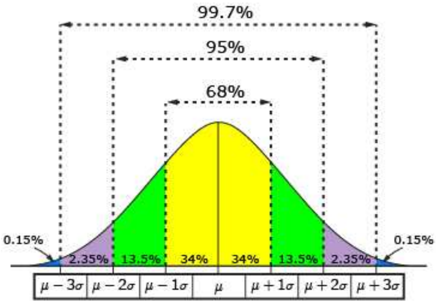
Similarly, the green zone on the right represents the data between $\bar{x} + s$ and $\bar{x} + 2s$, and accounts for another 13.5% of the area.

The new number: 2.35%

Look at the two zones shaded in violet in the image.

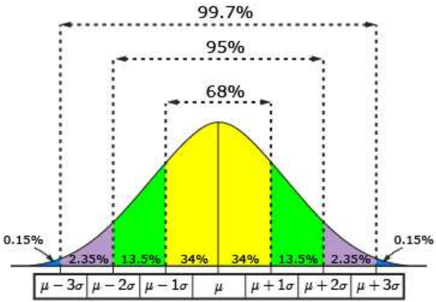
The new number: 2.35%

Look at the two zones shaded in violet in the image.



The new number: 2.35%

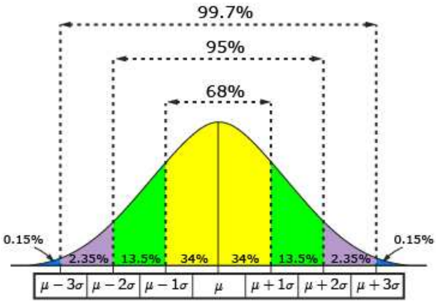
Look at the two zones shaded in violet in the image.



The one on the left represents the data between $\bar{x} - 3s$ and $\bar{x} - 2s$.

The new number: 2.35%

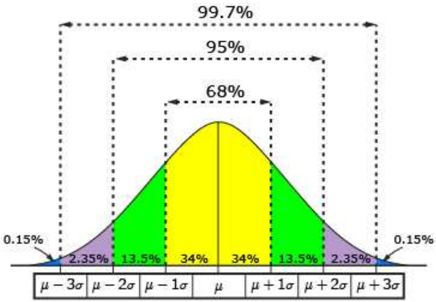
Look at the two zones shaded in violet in the image.



The one on the left represents the data between $\bar{x} - 3s$ and $\bar{x} - 2s$. That's the left half of the data which is at least two std devs away from the mean but no more than three std devs away.

The new number: 2.35%

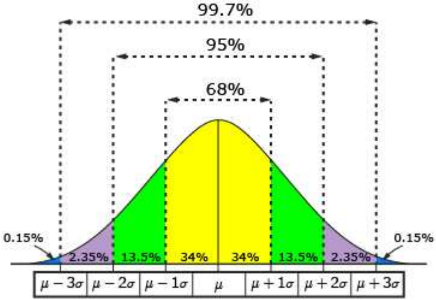
Look at the two zones shaded in violet in the image.



The one on the left represents the data between $\bar{x} - 3s$ and $\bar{x} - 2s$. That's the left half of the data which is at least two std devs away from the mean but no more than three std devs away. Numerically, we have half of $99.7\% - 95\% = 4.7\%$, hence 2.35%.

The new number: 2.35%

Look at the two zones shaded in violet in the image.



The one on the left represents the data between $\bar{x} - 3s$ and $\bar{x} - 2s$. That's the left half of the data which is at least two std devs away from the mean but no more than three std devs away. Numerically, we have half of $99.7\% - 95\% = 4.7\%$, hence 2.35%.

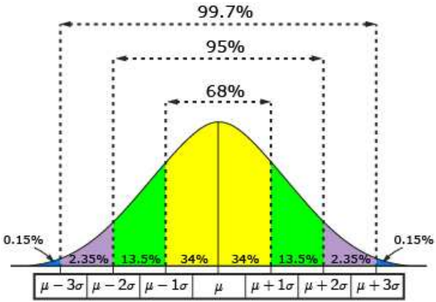
Similarly, the violet zone on the right represents the data between $\bar{x} + 2s$ and $\bar{x} + 3s$, and accounts for another 2.35% of the area.

The number: 0.15% (which is not so new!)

Look at the two tiny zones shaded in blue in the image.

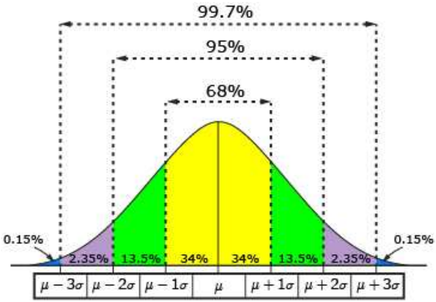
The number: 0.15% (which is not so new!)

Look at the two tiny zones shaded in blue in the image.



The number: 0.15% (which is not so new!)

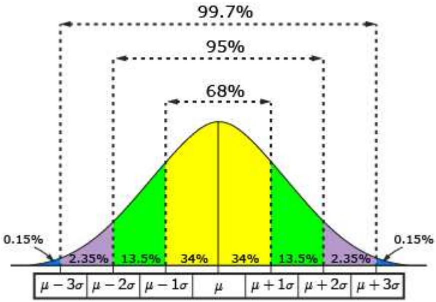
Look at the two tiny zones shaded in blue in the image.



The one on the left represents the data less than $\bar{x} - 3s$.

The number: 0.15% (which is not so new!)

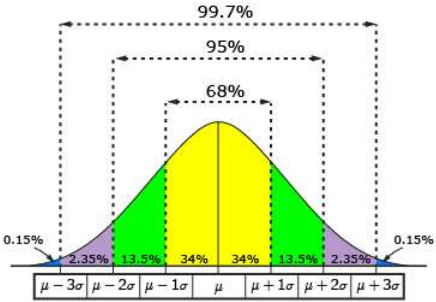
Look at the two tiny zones shaded in blue in the image.



The one on the left represents the data less than $\bar{x} - 3s$. That's the left half of the data which is at least three std devs away from the mean.

The number: 0.15% (which is not so new!)

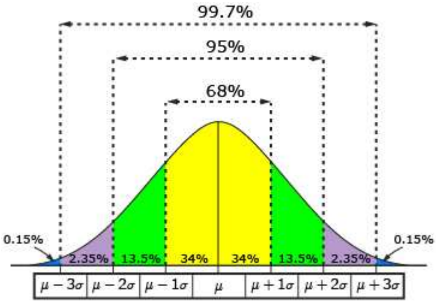
Look at the two tiny zones shaded in blue in the image.



The one on the left represents the data less than $\bar{x} - 3s$. That's the left half of the data which is at least three std devs away from the mean. It's half of $100\% - 99.7\% = 0.3\%$, hence 0.15%.

The number: 0.15% (which is not so new!)

Look at the two tiny zones shaded in blue in the image.



The one on the left represents the data less than $\bar{x} - 3s$. That's the left half of the data which is at least three std devs away from the mean. It's half of $100\% - 99.7\% = 0.3\%$, hence 0.15%.

Similarly, the blue zone on the right represents the data greater than $\bar{x} + 3s$, and accounts for another 0.15% of the area.

Cholesterol Levels Example: quoting the rules

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

Cholesterol Levels Example: quoting the rules

If men aged 18-24 cholesterol levels X 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.

Cholesterol Levels Example: quoting the rules

If men aged 18-24 cholesterol levels X 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.

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.

Cholesterol Levels Example: quoting the rules

If men aged 18-24 cholesterol levels X 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.

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.

Also, for 99.75% of the men, X lies between $\bar{x} - 3s$ and $\bar{x} + 3s$. Hence their cholesterol levels would be between 60 and 300.

Cholesterol Levels Example: quoting the rules

If men aged 18-24 cholesterol levels X 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.

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.

Also, for 99.75% of the men, X lies between $\bar{x} - 3s$ and $\bar{x} + 3s$. Hence their cholesterol levels would be between 60 and 300.

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.

Cholesterol Levels Example: quoting the rules

If men aged 18-24 cholesterol levels X 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.

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.

Also, for 99.75% of the men, X lies between $\bar{x} - 3s$ and $\bar{x} + 3s$. Hence their cholesterol levels would be between 60 and 300.

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.

Finally, 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 180.

Test Scores Examples: mastering the rules

Assume test scores X are $N(80, 6)$, so are are normally distributed with mean 80 and std dev 6.

Test Scores Examples: mastering the rules

Assume test scores X are $N(80, 6)$, so are are normally distributed with mean 80 and std dev 6.

What percentage of the test scores are over 80?

Test Scores Examples: mastering the rules

Assume test scores X are $N(80, 6)$, so are are normally distributed with mean 80 and std dev 6.

What percentage of the test scores are over 80? Answer: 50% (since 80 is the median as well as the mean and mode)

Test Scores Examples: mastering the rules

Assume test scores X are $N(80, 6)$, so are are normally distributed with mean 80 and std dev 6.

What percentage of the test scores are over 80? Answer: 50% (since 80 is the median as well as the mean and mode)

What percentage are beteeen 74 and 80?

Test Scores Examples: mastering the rules

Assume test scores X are $N(80, 6)$, so are are normally distributed with mean 80 and std dev 6.

What percentage of the test scores are over 80? Answer: 50% (since 80 is the median as well as the mean and mode)

What percentage are beteeen 74 and 80? Answer: 34%

Test Scores Examples: mastering the rules

Assume test scores X are $N(80, 6)$, so are are normally distributed with mean 80 and std dev 6.

What percentage of the test scores are over 80? Answer: 50% (since 80 is the median as well as the mean and mode)

What percentage are beteeen 74 and 80? Answer: 34%

What percentage are beteeen 80 and 86?

Test Scores Examples: mastering the rules

Assume test scores X are $N(80, 6)$, so are are normally distributed with mean 80 and std dev 6.

What percentage of the test scores are over 80? Answer: 50% (since 80 is the median as well as the mean and mode)

What percentage are beteeen 74 and 80? Answer: 34%

What percentage are beteeen 80 and 86? Answer: 34%

Test Scores Examples: mastering the rules

Assume test scores X are $N(80, 6)$, so are are normally distributed with mean 80 and std dev 6.

What percentage of the test scores are over 80? Answer: 50% (since 80 is the median as well as the mean and mode)

What percentage are beteeen 74 and 80? Answer: 34%

What percentage are beteeen 80 and 86? Answer: 34%

What percentage are beteeen 80 and 92?

Test Scores Examples: mastering the rules

Assume test scores X are $N(80, 6)$, so are are normally distributed with mean 80 and std dev 6.

What percentage of the test scores are over 80? Answer: 50% (since 80 is the median as well as the mean and mode)

What percentage are beteeen 74 and 80? Answer: 34%

What percentage are beteeen 80 and 86? Answer: 34%

What percentage are beteeen 80 and 92? Answer: 47.5% (that's half of 95%, and it's also 34% plus 13.5%)

Test Scores Examples: mastering the rules

Assume test scores X are $N(80, 6)$, so are are normally distributed with mean 80 and std dev 6.

What percentage of the test scores are over 80? Answer: 50% (since 80 is the median as well as the mean and mode)

What percentage are beteeen 74 and 80? Answer: 34%

What percentage are beteeen 80 and 86? Answer: 34%

What percentage are beteeen 80 and 92? Answer: 47.5% (that's half of 95%, and it's also 34% plus 13.5%)

What percentage are beteeen 86 and 92?

Test Scores Examples: mastering the rules

Assume test scores X are $N(80, 6)$, so are are normally distributed with mean 80 and std dev 6.

What percentage of the test scores are over 80? Answer: 50% (since 80 is the median as well as the mean and mode)

What percentage are beteeen 74 and 80? Answer: 34%

What percentage are beteeen 80 and 86? Answer: 34%

What percentage are beteeen 80 and 92? Answer: 47.5% (that's half of 95%, and it's also 34% plus 13.5%)

What percentage are beteeen 86 and 92? Answer: 13.5%

Test Scores Examples: mastering the rules

Assume test scores X are $N(80, 6)$, so are are normally distributed with mean 80 and std dev 6.

What percentage of the test scores are over 80? Answer: 50% (since 80 is the median as well as the mean and mode)

What percentage are beteeen 74 and 80? Answer: 34%

What percentage are beteeen 80 and 86? Answer: 34%

What percentage are beteeen 80 and 92? Answer: 47.5% (that's half of 95%, and it's also 34% plus 13.5%)

What percentage are beteeen 86 and 92? Answer: 13.5%

What percentage are over 98?

Test Scores Examples: mastering the rules

Assume test scores X are $N(80, 6)$, so are are normally distributed with mean 80 and std dev 6.

What percentage of the test scores are over 80? Answer: 50% (since 80 is the median as well as the mean and mode)

What percentage are beteeen 74 and 80? Answer: 34%

What percentage are beteeen 80 and 86? Answer: 34%

What percentage are beteeen 80 and 92? Answer: 47.5% (that's half of 95%, and it's also 34% plus 13.5%)

What percentage are beteeen 86 and 92? Answer: 13.5%

What percentage are over 98? Answer: 0.15%

Test Scores Examples: mastering the rules

Assume test scores X are $N(80, 6)$, so are are normally distributed with mean 80 and std dev 6.

What percentage of the test scores are over 80? Answer: 50% (since 80 is the median as well as the mean and mode)

What percentage are beteeen 74 and 80? Answer: 34%

What percentage are beteeen 80 and 86? Answer: 34%

What percentage are beteeen 80 and 92? Answer: 47.5% (that's half of 95%, and it's also 34% plus 13.5%)

What percentage are beteeen 86 and 92? Answer: 13.5%

What percentage are over 98? Answer: 0.15%

What percentage are between 74 and 92?

Test Scores Examples: mastering the rules

Assume test scores X are $N(80, 6)$, so are are normally distributed with mean 80 and std dev 6.

What percentage of the test scores are over 80? Answer: 50% (since 80 is the median as well as the mean and mode)

What percentage are beteeen 74 and 80? Answer: 34%

What percentage are beteeen 80 and 86? Answer: 34%

What percentage are beteeen 80 and 92? Answer: 47.5% (that's half of 95%, and it's also 34% plus 13.5%)

What percentage are beteeen 86 and 92? Answer: 13.5%

What percentage are over 98? Answer: 0.15%

What percentage are between 74 and 92? (Note that this is a non-symmetric zone straddling the mean)

The textbook

Section 6C, pages 391-395; do problems 19-20.

The textbook

Section 6C, pages 391-395; do problems 19-20.

Expect a quiz on the basics of this on Monday!

The textbook

Section 6C, pages 391-395; do problems 19-20.

Expect a quiz on the basics of this on Monday!

Read ahead, pages 395-397, “standard scores and percentiles” (via the table on page 396)