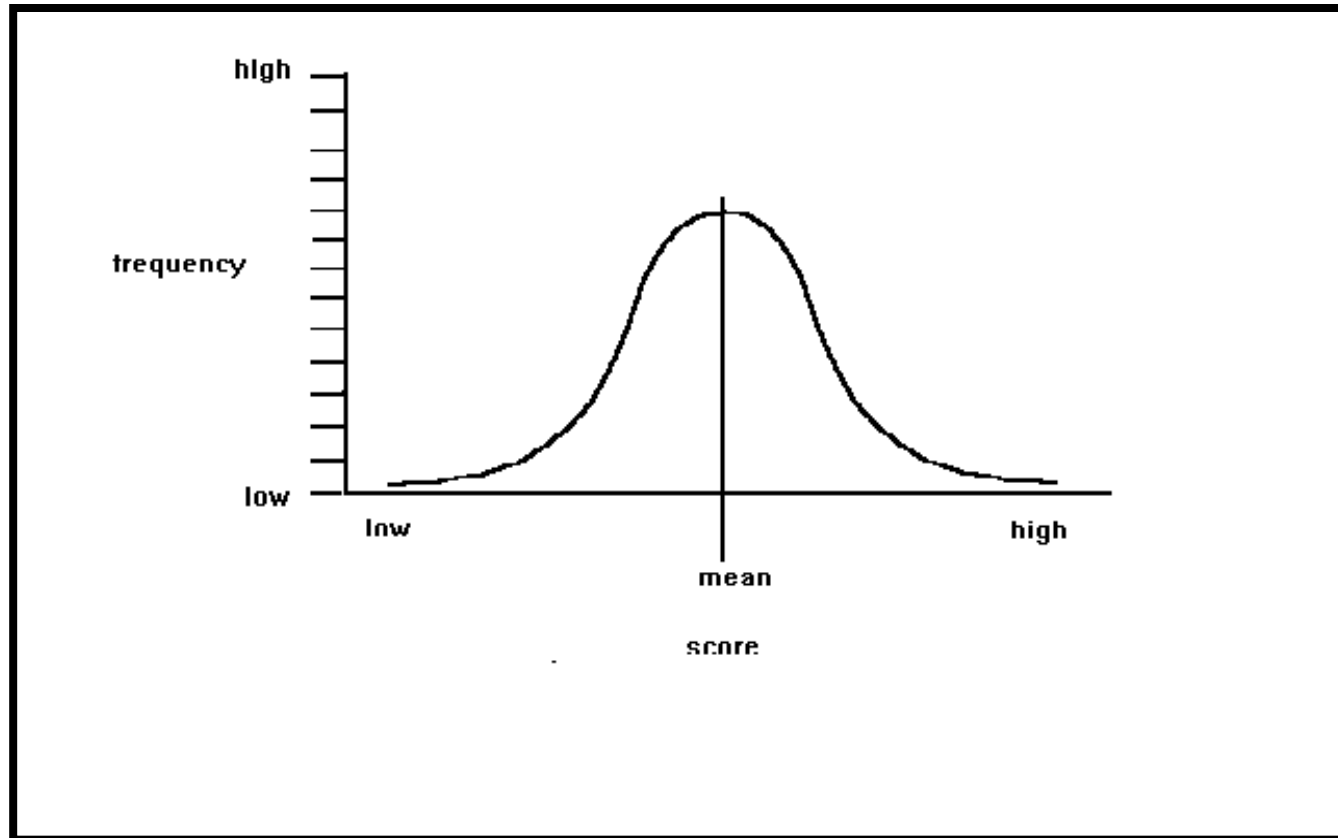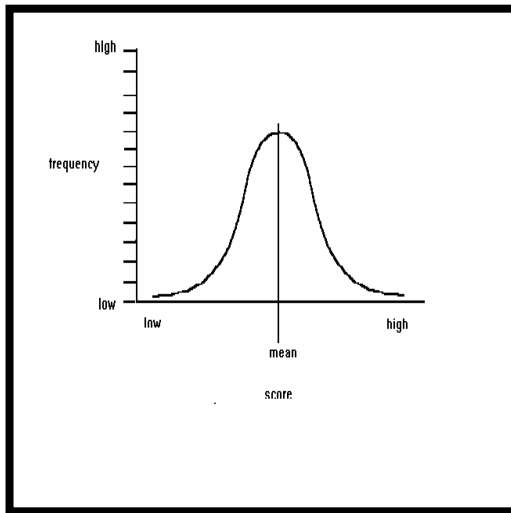# *The Normal Distribution:*

The Normal curve is a mathematical abstraction which conveniently describes ("models") many frequency distributions of scores in real-life.
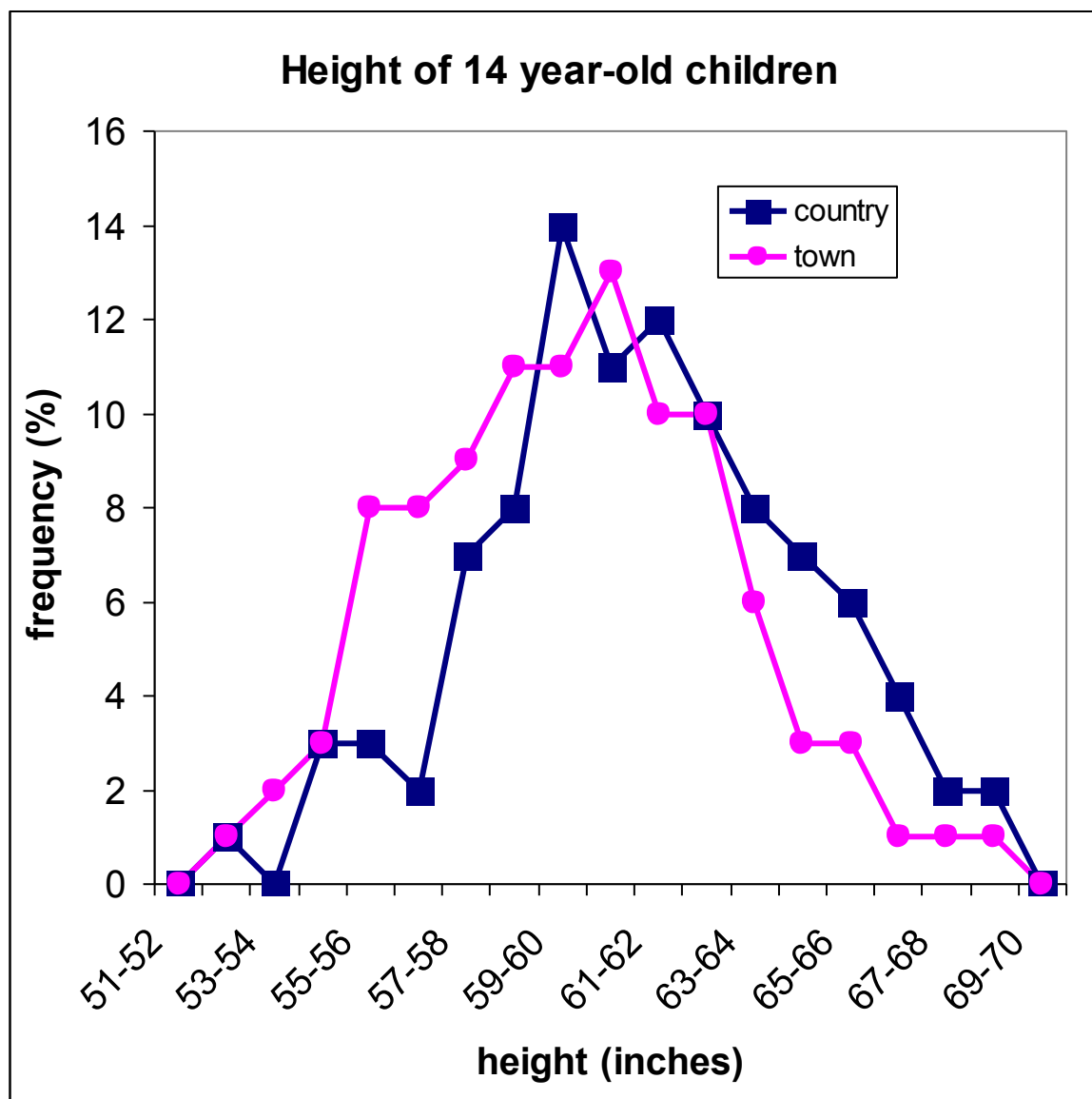
length of time before someone looks away in a staring contest:


length of pickled gherkins:

Francis Galton (1876) 'On the height and weight of boys aged 14, in town and country public schools.' *Journal of the Anthropological Institute*, 5, 174-180:

TABLE I.—NUMBER OF BOYS AGED FOURTEEN ON THEIR LAST BIRTHDAY, IN THE COUNTRY AND TOWN PUBLIC SCHOOLS, WHOSE HEIGHTS AND WEIGHTS LIE BETWEEN THE LIMITS MENTIONED BELOW.

HEIGHTS IN INCHES.		49 to 50	50 to 51	51 to 52	52 to 53	53 to 54	54 to 55	55 to 56	56 to 57	57 to 58	58 to 59	59 to 60	60 to 61	61 to 62	62 to 63	63 to 64	64 to 65	65 to 66	66 to 67	67 to 68	68 to 69	69 to 70	70 to 71	71 to 72	72 and above	TOTALS.
In Country Schools.	A. As observed	...	...	1	2	1	9	10	6	21	24	40	34	35	29	23	19	19	11	5	5	1	0	1	...	296
	B. As reduced to per cents.	...	...	0	1	0	3	3	2	7	8	14	11	12	10	8	7	6	4	2	2	0	0	0	...	100
	C. Sums of figures in B, from beginning.	...	...	0	1	...	4	7	9	16	24	38	49	61	71	79	86	92	96	98	100	...	...	...	...	...
In Town Schools.	A. As observed	1	0	2	6	13	14	37	39	45	54	57	60	52	48	29	17	15	4	5	5	...	...	...	...	509
	B. As reduced to per cents.	0	0	0	1	2	3	8	8	9	11	11	12	10	10	6	3	3	1	1	1	...	...	...	...	100
	C. Sums of figures in B, from beginning.	...	...	...	1	3	6	14	22	31	42	53	65	75	85	91	94	97	98	99	100	...	...	...	...	...

WEIGHTS IN POUNDS.		56 to 63	63 to 70	70 to 77	77 to 84	84 to 91	91 to 98	98 to 105	105 to 112	112 to 119	119 to 126	126 to 133	133 to 140	140 to 147	147 to 154	154 and above										TOTALS.
In Country Schools.	A. As observed	...	2	9	18	44	62	61	33	22	25	12	1	5	2	...										296
	B. As reduced to per cents.	...	1	3	6	15	21	21	11	7	8	4	0	2	1	...										100
	C. Sums of figures in B, from beginning.	...	1	4	10	25	46	67	78	85	93	97	...	99	100	...										...
In Town Schools.	A. As observed	1	18	60	83	99	89	65	39	32	9	10	2	3	...	...										510
	B. As reduced to per cents.	0	3	12	16	19	18	13	8	6	2	2	0	1	...	...										100
	C. Sums of figures in B, from beginning.	...	3	15	31	50	68	81	89	95	97	99	...	100	...	...										...

Francis Galton (1876) 'On the height and weight of boys aged 14, in town and country public schools.' *Journal of the Anthropological Institute*, 5, 174-180:

An example of a normal distribution - the length of Sooty's magic wand...
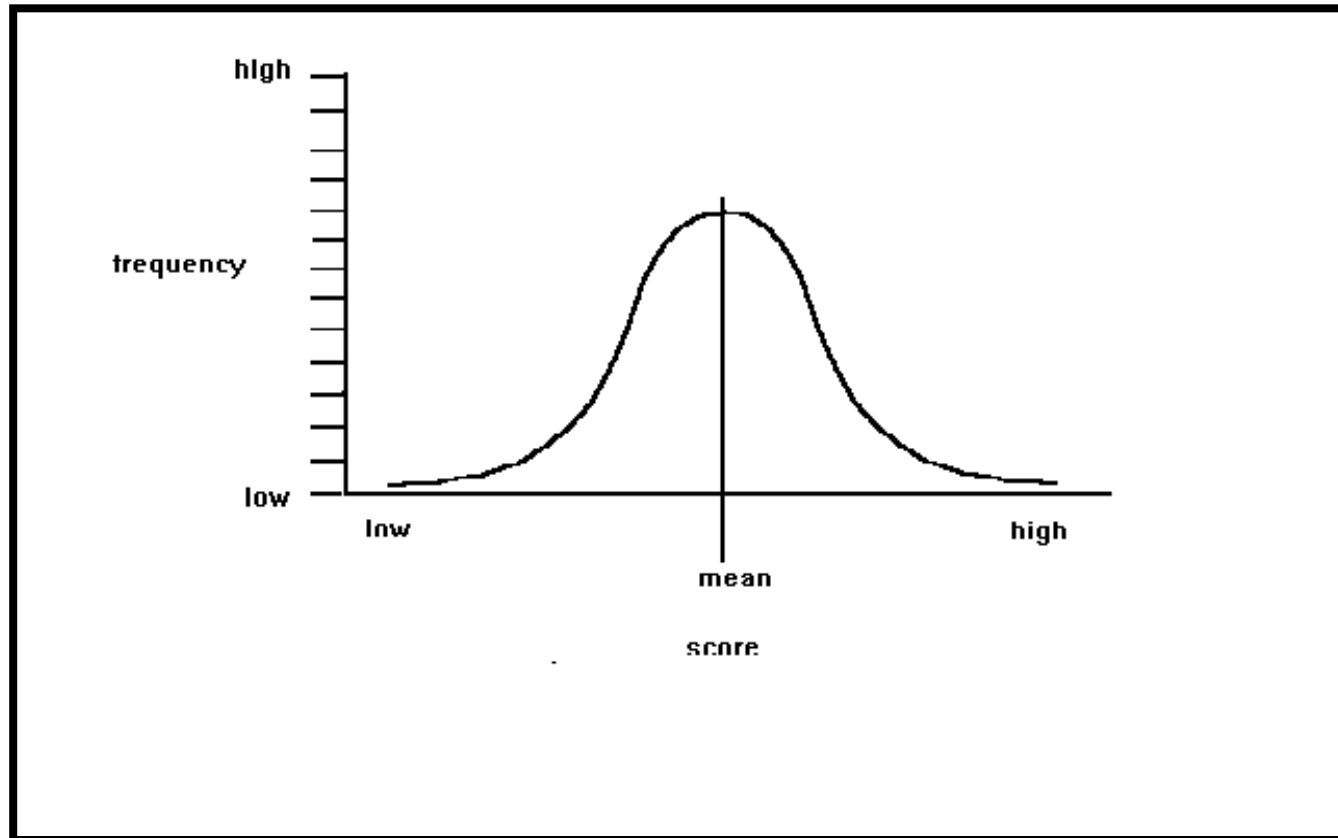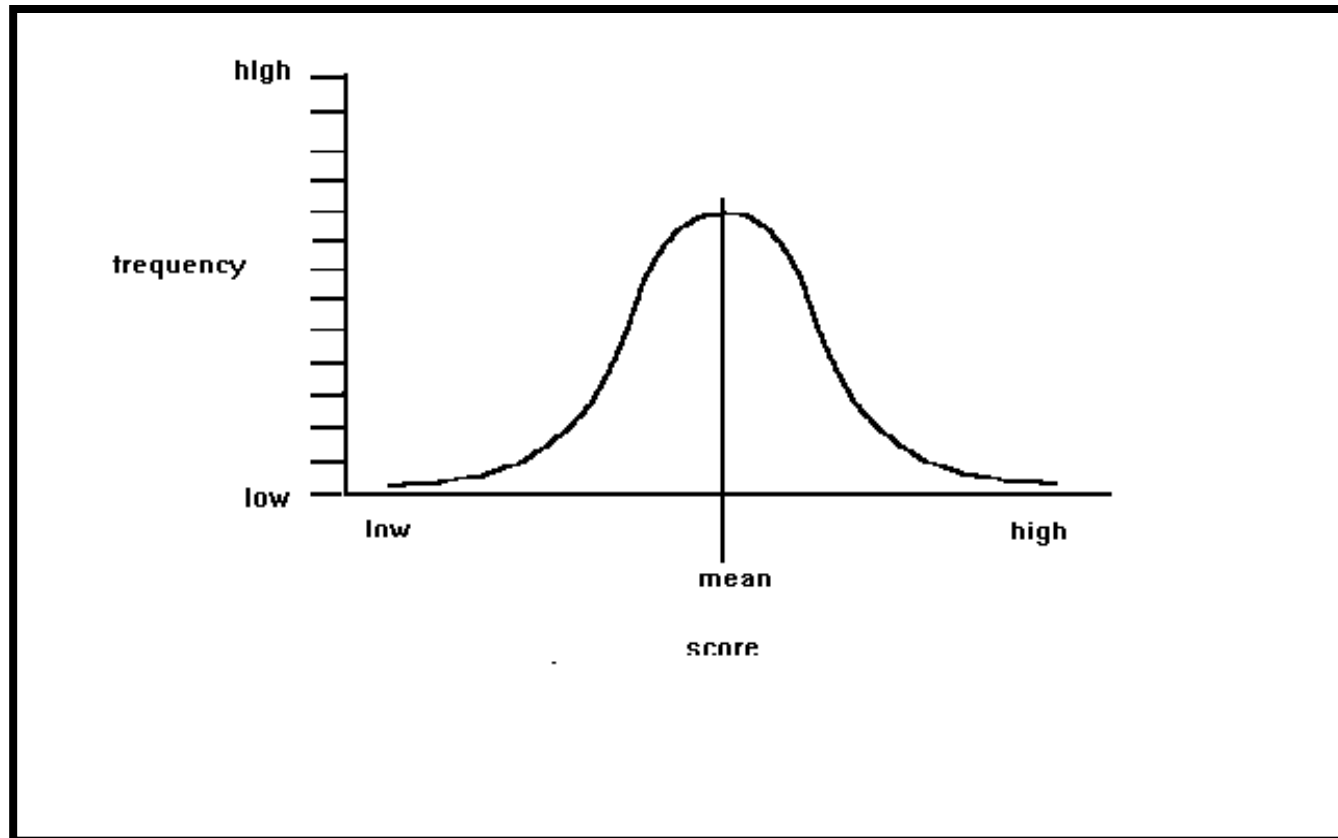

Frequency of different wand lengths

Length of wand
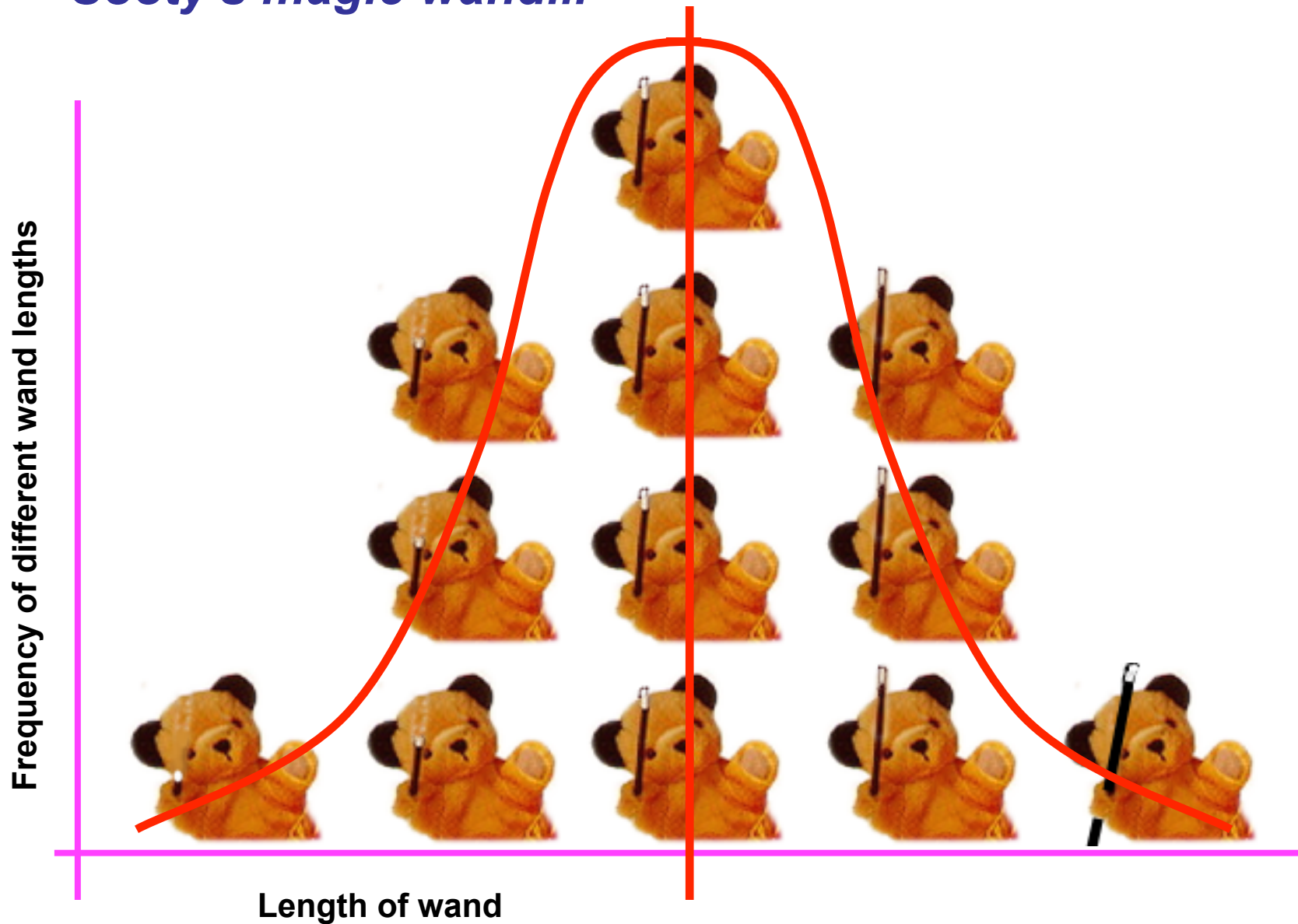
Properties of the Normal Distribution:

1. It is bell-shaped and asymptotic at the extremes.
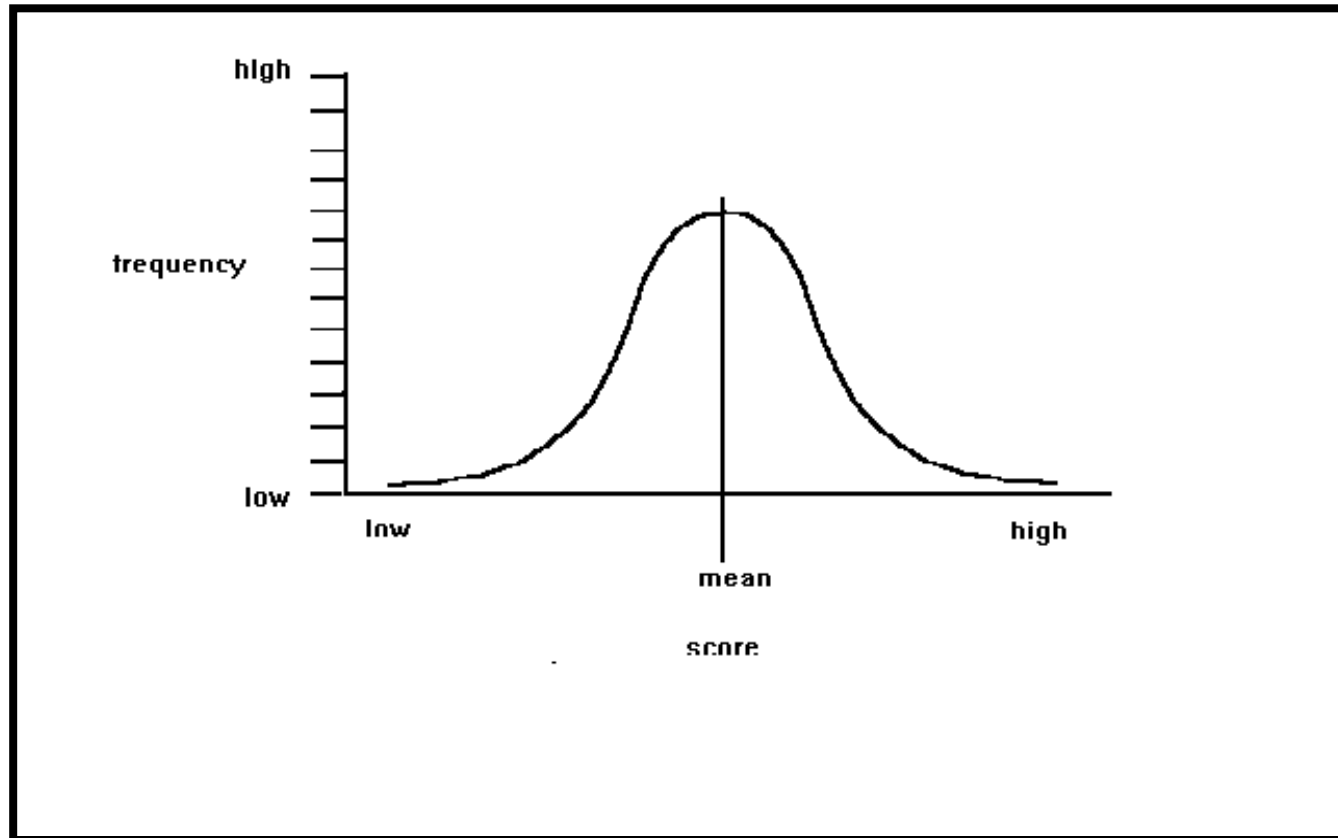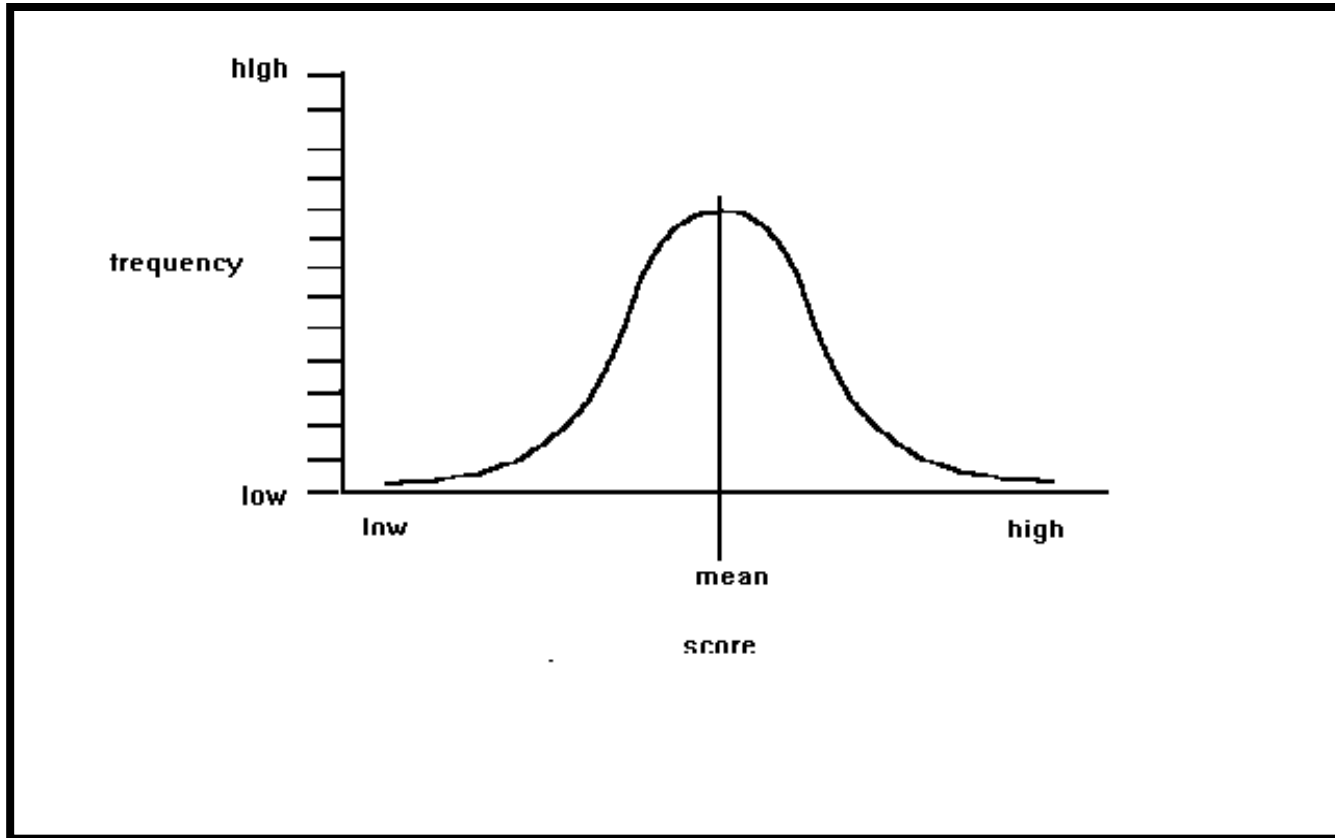
2. It's symmetrical around the mean.

3. The mean, median and mode all have same value.

4. It can be specified completely, once mean and SD are known.
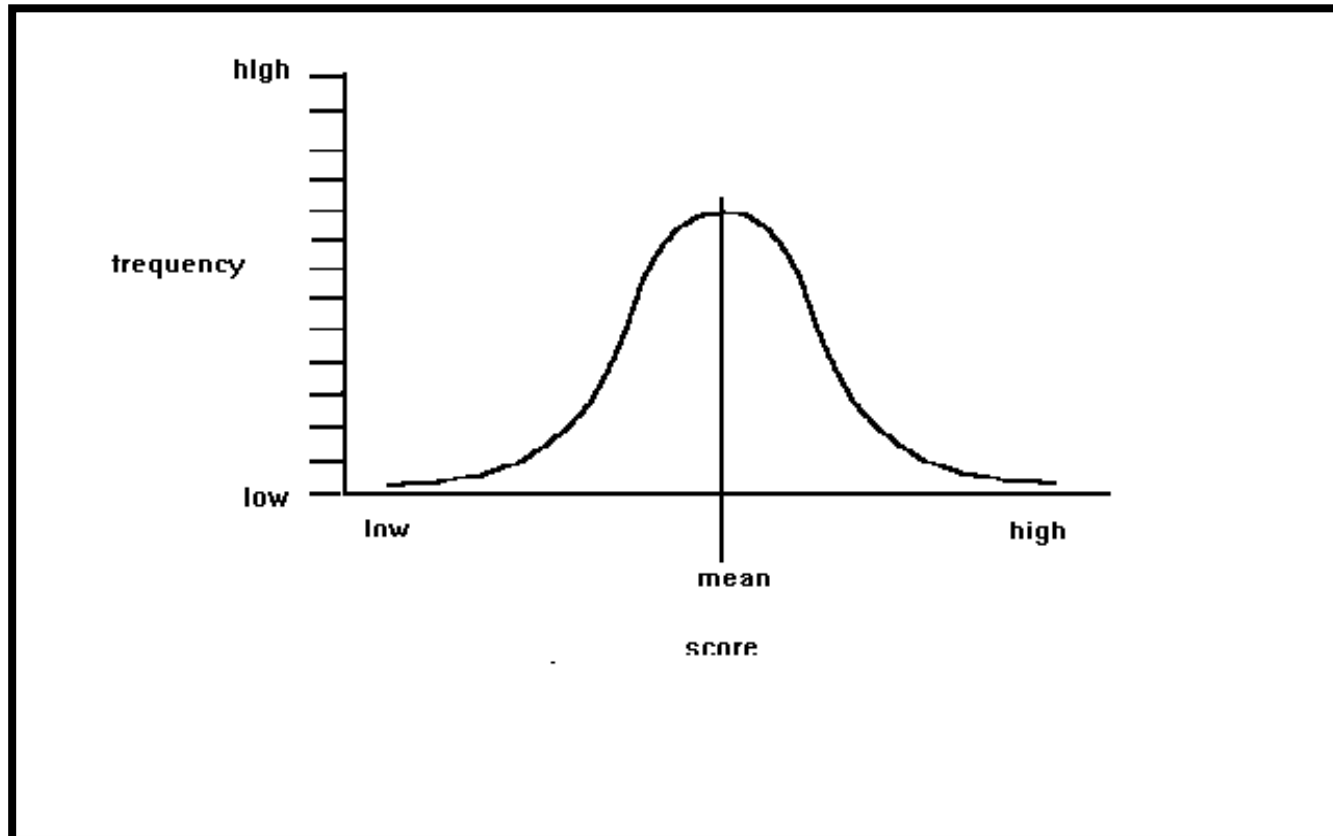
5. The area under the curve is directly proportional to the relative frequency of observations.
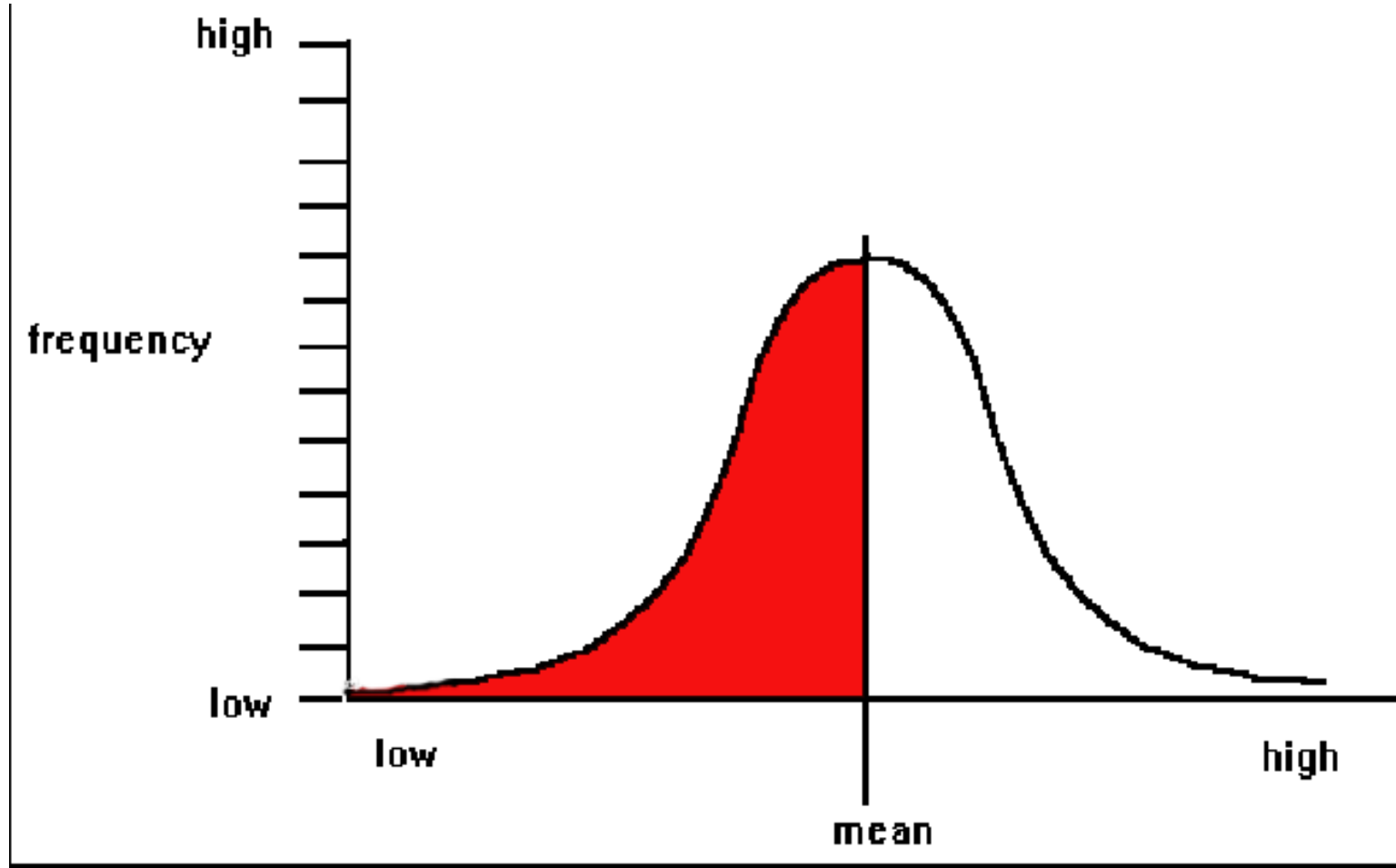
e.g. here, 50% of scores fall below the mean, as does 50% of the area under the curve.
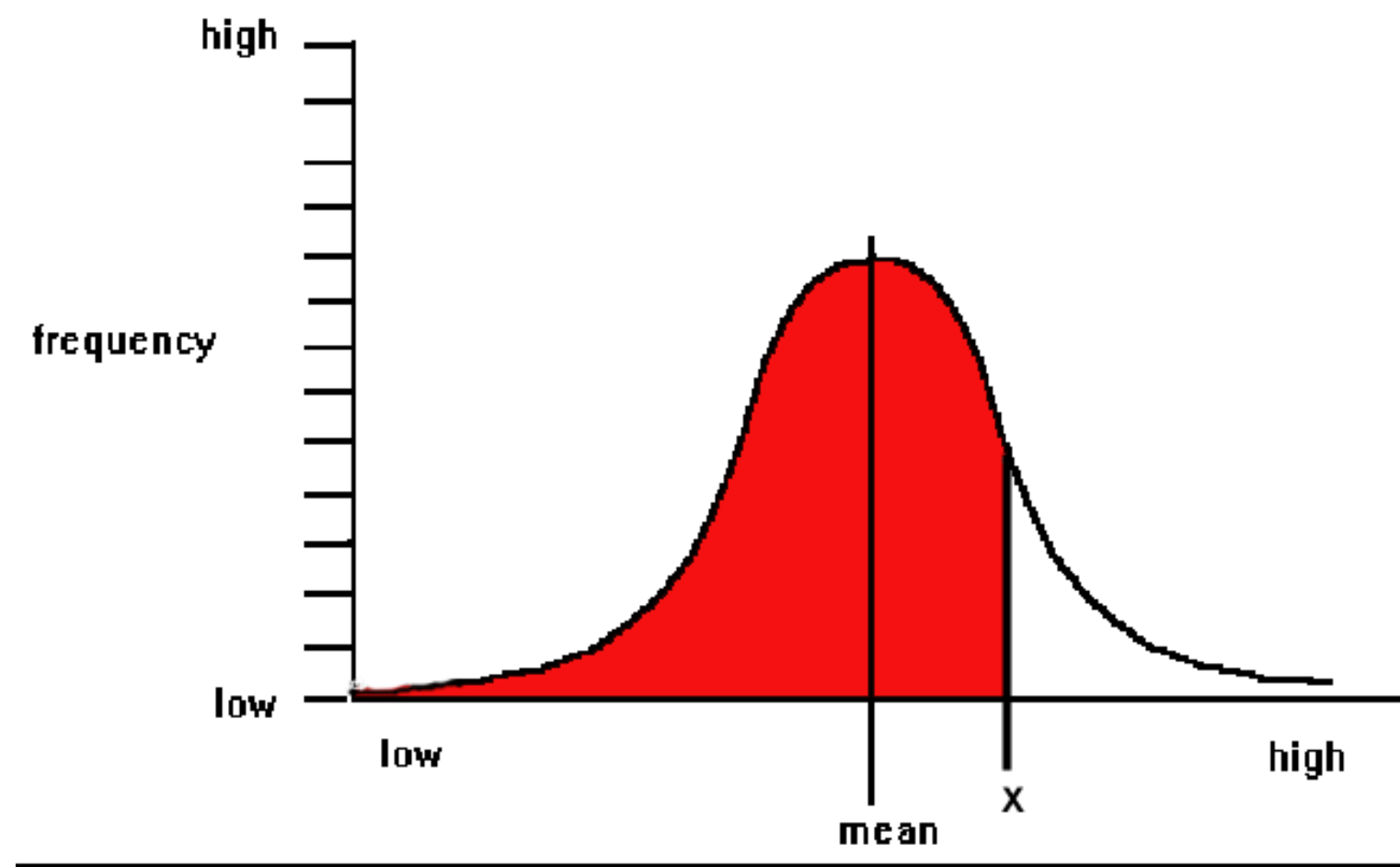
e.g. here, 85% of scores fall below score X, corresponding to 85% of the area under the curve.

Relationship between the normal curve and the standard deviation:

All normal curves share this property: the SD cuts off a constant proportion of the distribution of scores:-

frequency

68%

95%

99.7%

-3   -2   -1   mean   +1   +2   +3

Number of standard deviations either side of mean
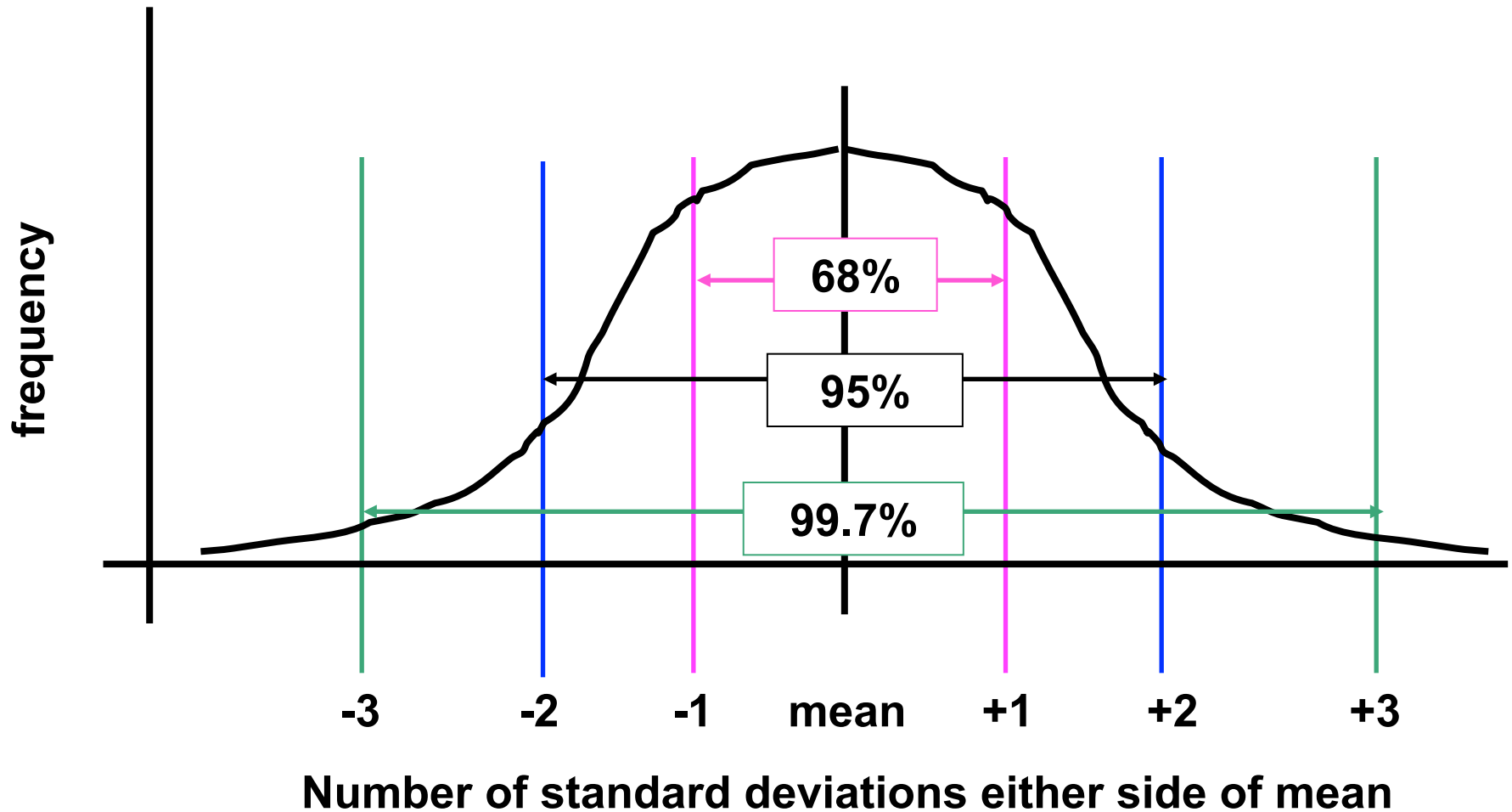
About 68% of scores fall in the range of the mean plus and minus 1 SD;

95% in the range of the mean +/- 2 SDs;

99.7% in the range of the mean +/- 3 SDs.

e.g. IQ is normally distributed (mean = 100, SD = 15).

68% of people have IQs between 85 and 115 (100 +/- 15).

95% have IQs between 70 and 130 (100 +/- (2*15).

99.7% have IQs between 55 and 145 (100 +/- (3*15).

68%

85 (mean - 1 SD)   115 (mean + 1 SD)
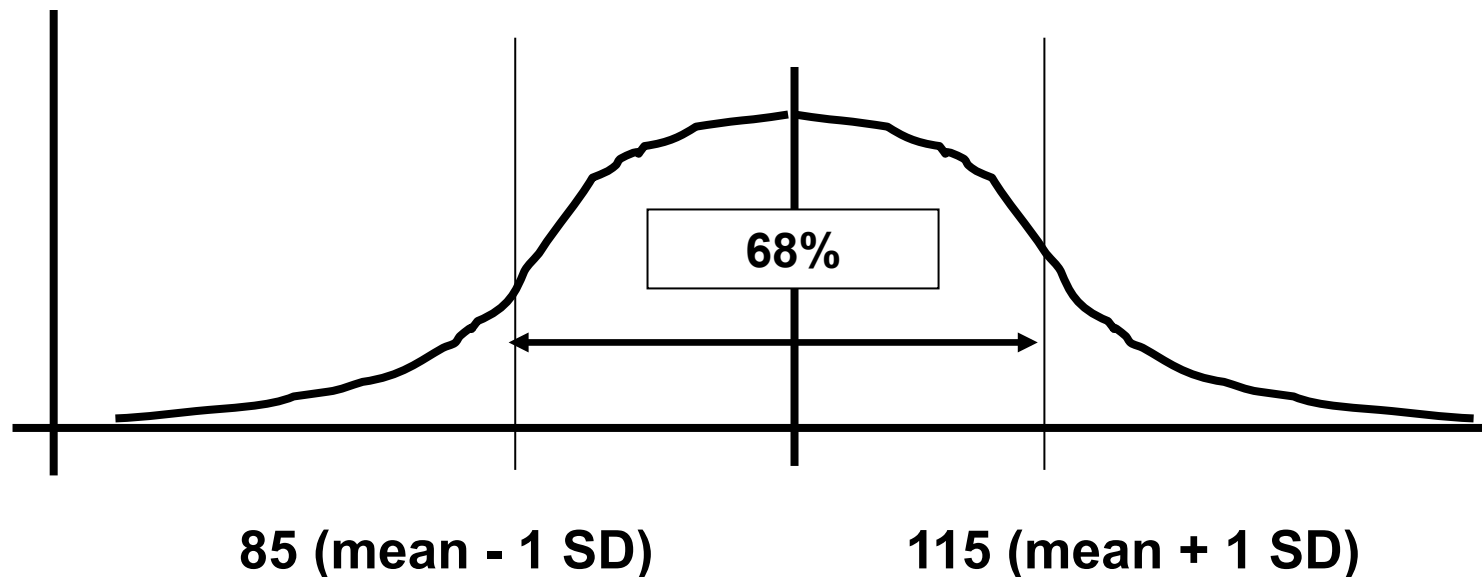
We can tell a lot about a population just from knowing the mean, SD, and that scores are normally distributed.

If we encounter someone with a particular score, we can assess how they stand in relation to the rest of their group.

e.g. someone with an IQ of 145 is quite unusual (3 SDs above the mean).

IQs of 3 SDs or above occur in only 0.15% of the population [(100-99.7) / 2].

z-scores:

z-scores are "standard scores".

A z-score states the position of a raw score in relation to the mean of the distribution, using the standard deviation as the unit of measurement.

$$z = \frac{raw\ score - mean}{standard\ deviation}$$

for a population :

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

for a sample :

$$z = \frac{X - \overline{X}}{s}$$

1. Find the *difference* between a score and the mean of the set of scores.

2. Divide this difference by the SD (in order to assess how big it really is).

Raw score distributions:

A score, X, is expressed in the original units of measurement:

X = 65

$\overline{X} = 50 \quad s = 10$

X = 236

$\overline{X} = 200 \quad s = 24$

z = 1.5

$\overline{X} = 0 \quad s = 1$
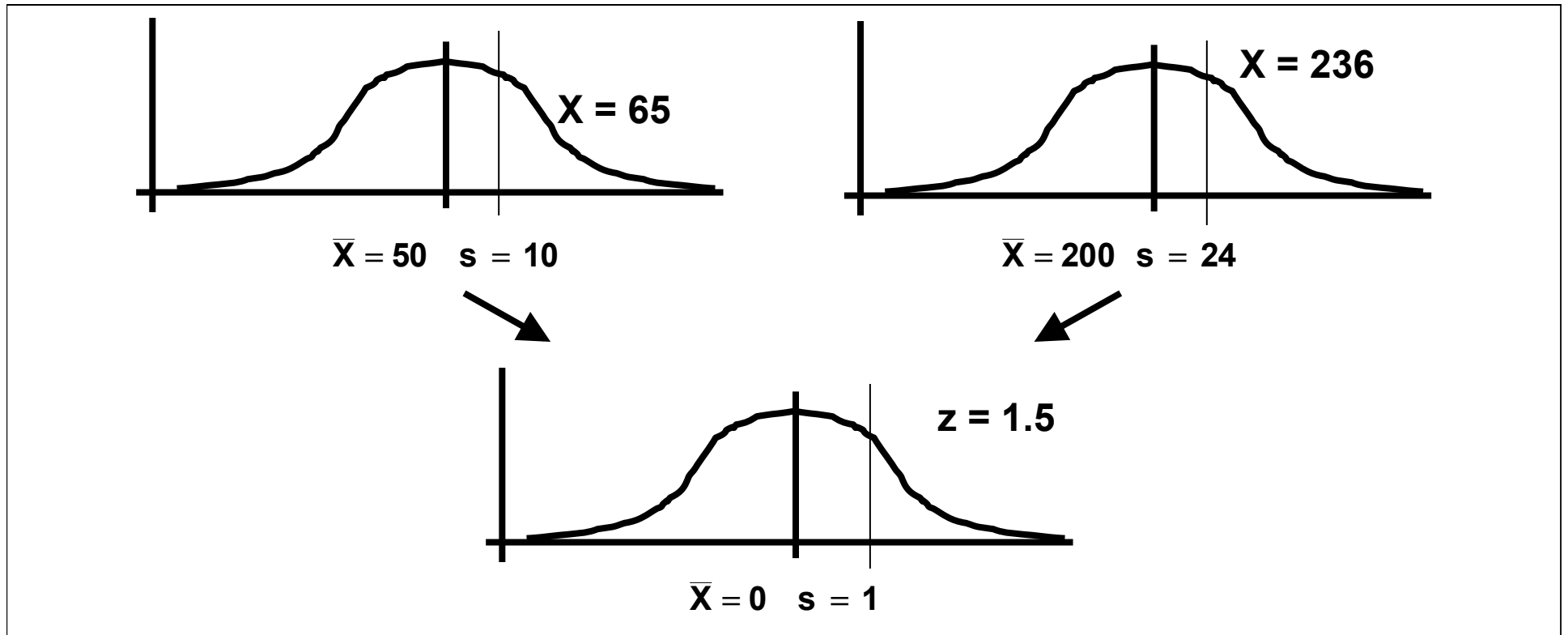
z-score distribution:

X is expressed in terms of its *deviation* from the mean (in SDs).

z-scores *transform* **our original scores into scores with a mean of 0 and an SD of 1.**

Raw IQ scores (mean = 100, SD = 15)

z for 100 = (100-100) / 15 = **0**, z for 115 = (115-100) / 15 = **1**,

z for 70 = (70-100) / **-2**, etc.

raw:	55	70	85	100	115	130	145
z-score:	-3	-2	-1	0	+1	+2	+3
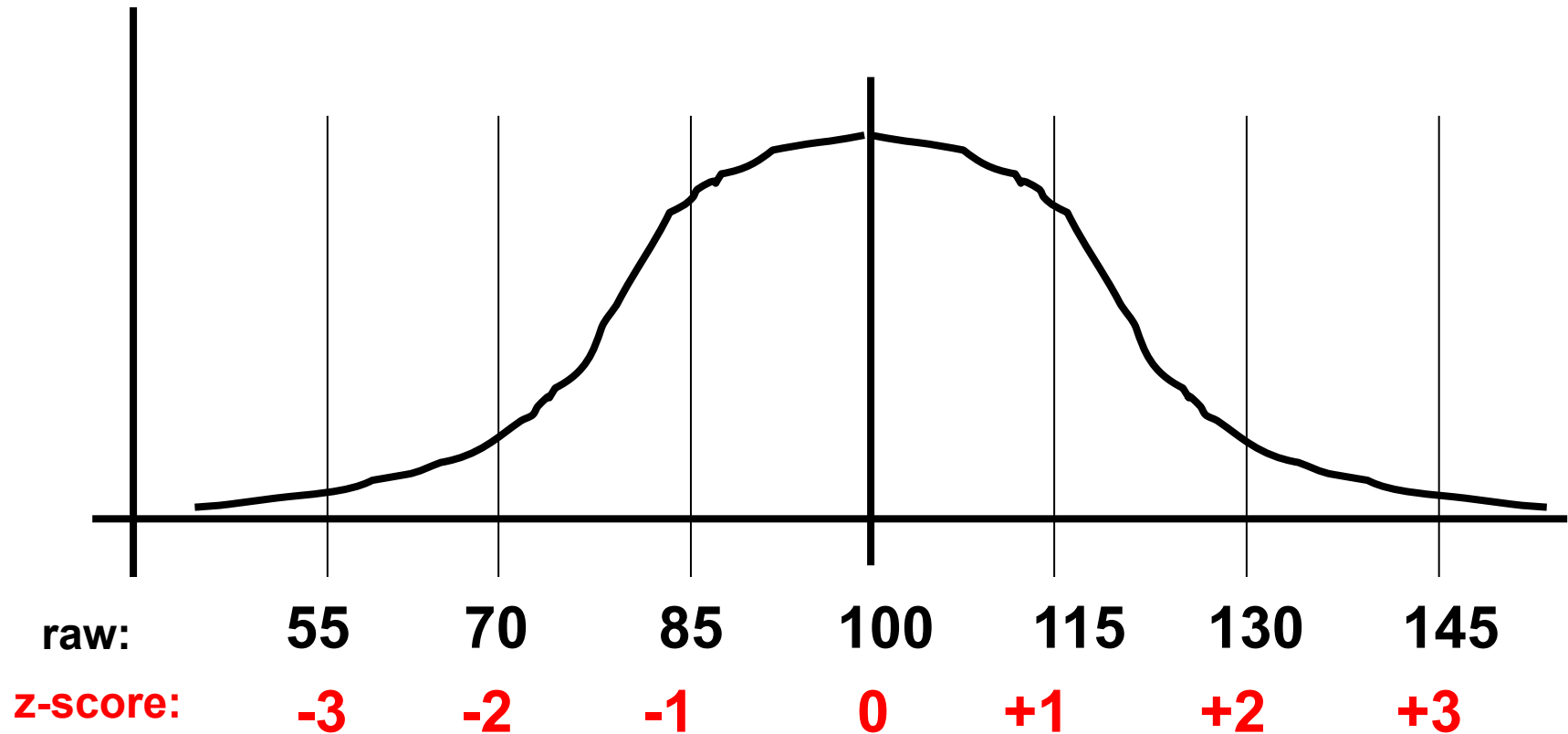
Why use z-scores?

1. z-scores make it easier to compare scores from distributions using different scales.

e.g. two tests:

Test A: Fred scores 78. Mean score = 70, SD = 8.

Test B: Fred scores 78. Mean score = 66, SD = 6.

Did Fred do better or worse on the second test?

Test A: as a z-score, z = (78-70) / 8   = 1.00

Test B: as a z-score , z = (78 - 66) / 6 = 2.00
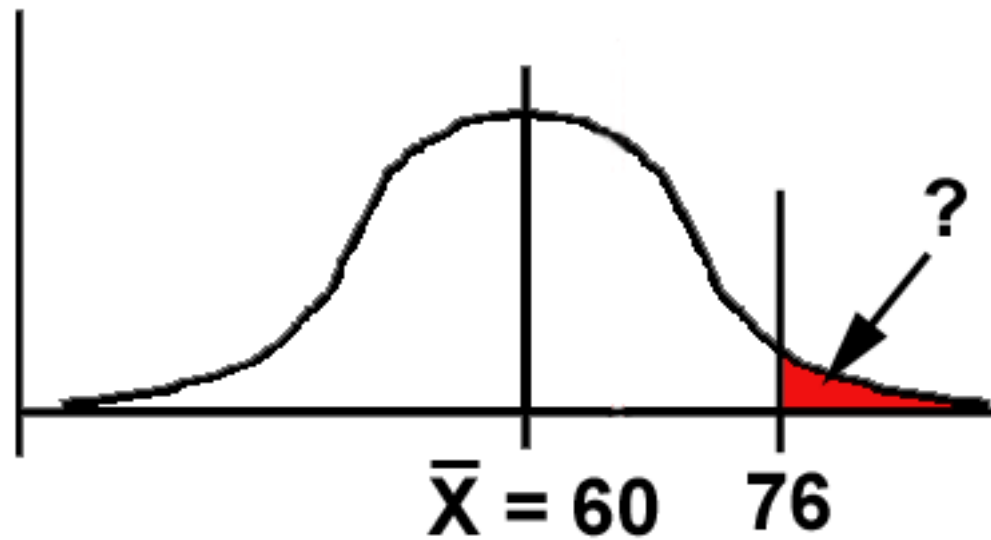
Conclusion: Fred did much better on Test B.

$\overline{X}$ = 70   X = 78

sd = 8

$\overline{X}$ = 66   X = 78

sd = 6
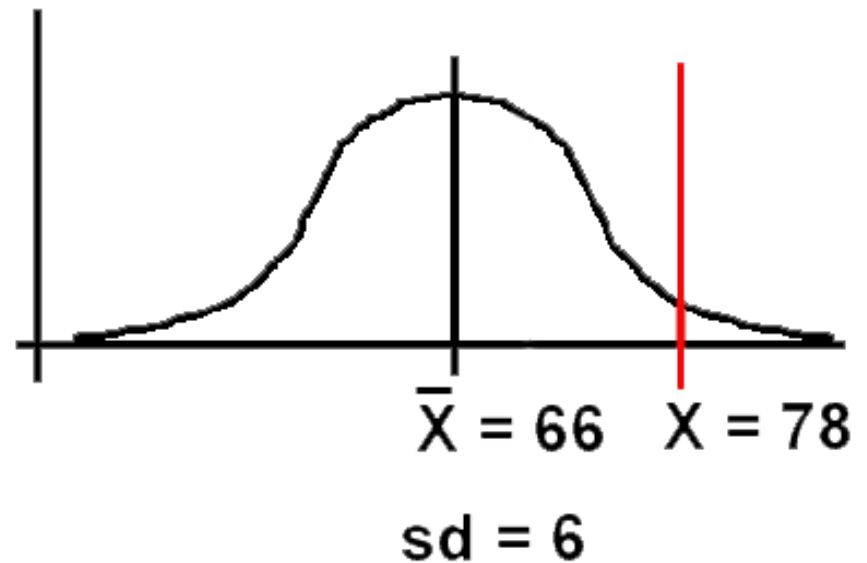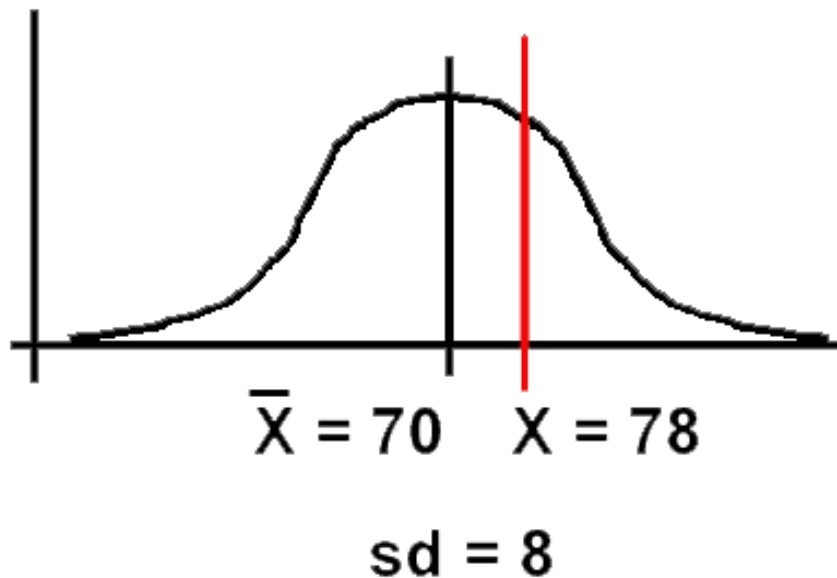
2. z-scores enable us to determine the relationship between one score and the rest of the scores, using just one table for *all* normal distributions.

e.g. If we have 480 scores, normally distributed with a mean of 60 and an SD of 8, how many would be 76 or above?
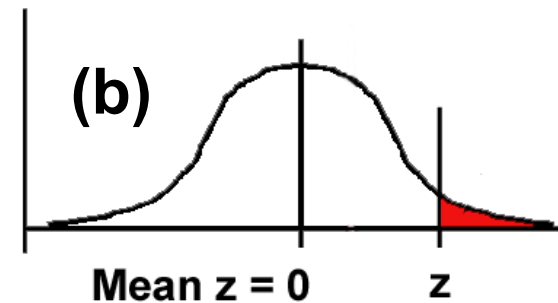
(a) Graph the problem:

$\overline{X} = 60$ 76

total = 480   s = 8

$\overline{X} = 60$ 76
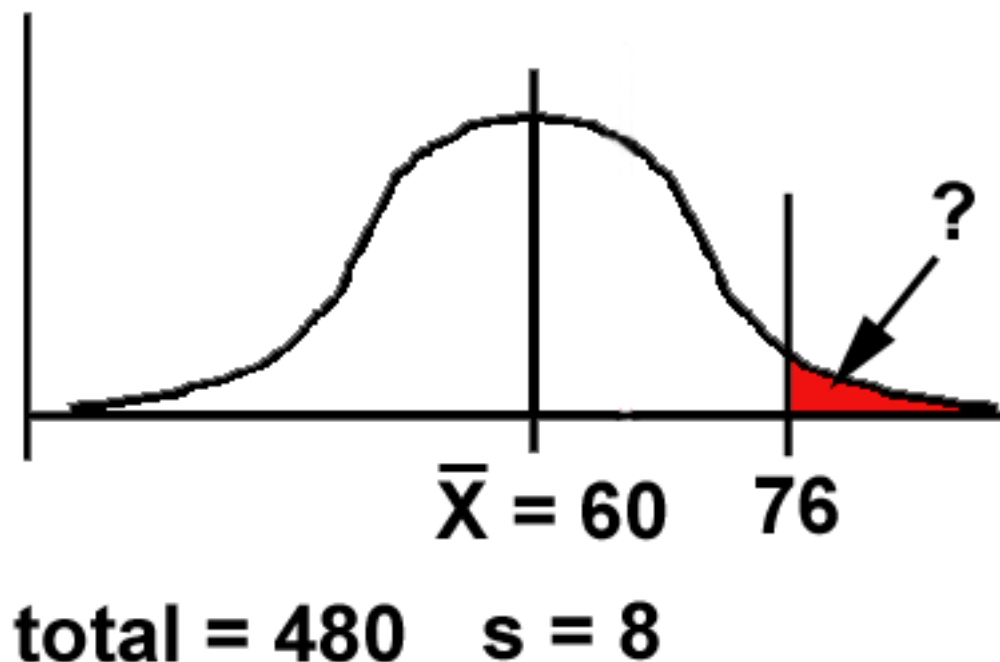
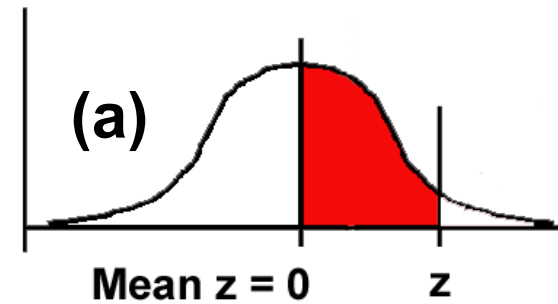total = 480   s = 8

(b) Work out the z-score for 76:

z = (X - $\overline{X}$) / s   =   (76 - 60) / 8   =   16 / 8 = 2.00

(c) We need to know the size of the **area beyond z** (remember - the area under the Normal curve corresponds directly to the proportion of scores).

Many statistics books (and my website!) have z-score tables, giving us this information:

(a)
Mean z = 0   z

(b)
Mean z = 0   z

z	(a) Area between mean and z	(b) Area beyond z
0.00	0.0000	0.5000
0.01	0.0040	0.4960
0.02	0.0080	0.4920
:	:	:
1.00	0.3413 *	0.1587
:	:	:
2.00	0.4772 +	0.0228
:	:	:
3.00	0.4987 #	0.0013

* x 2 = 68% of scores

\+ x 2 = 95% of scores

\# x 2 = 99.7% of scores

(roughly!)

0.0228

$\overline{X} = 60$ 76
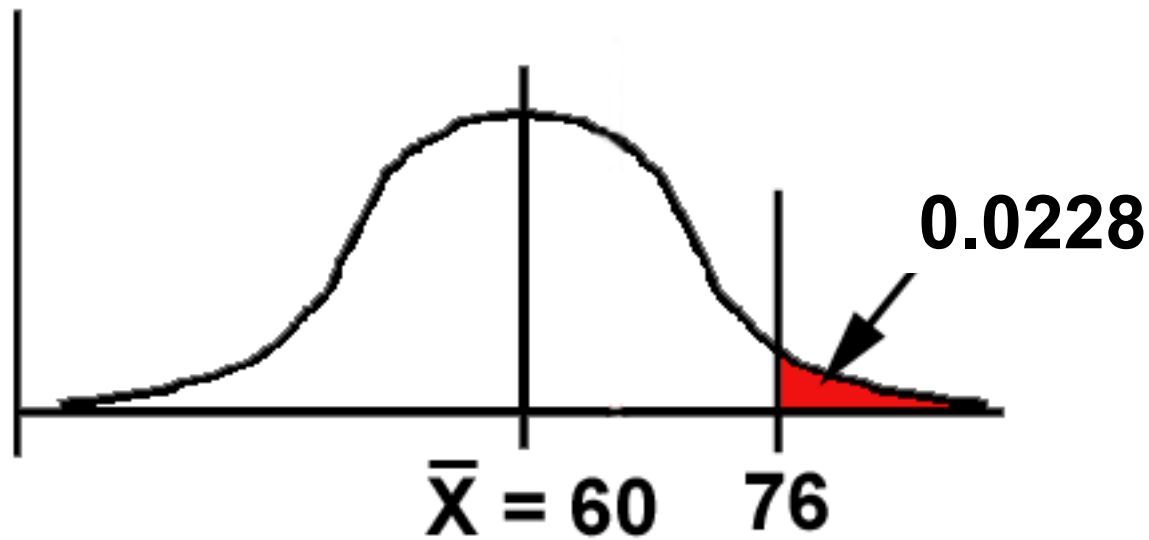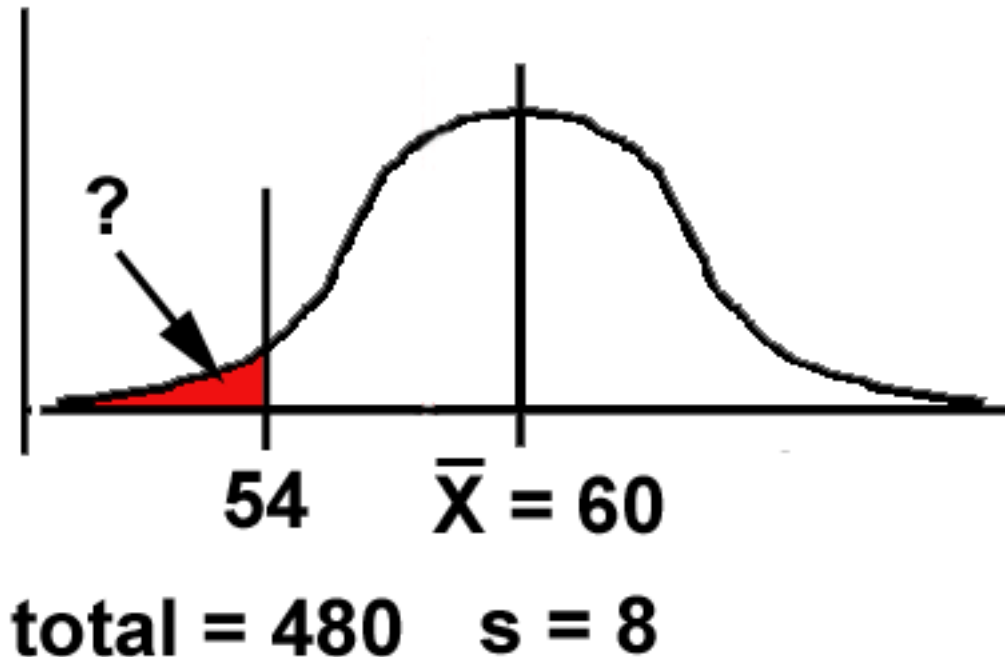
total = 480   s = 8

(d) So: as a proportion of 1, 0.0228 of scores are likely to be 76 or more.

As a percentage, = 2.28%

As a number, 0.0228 * 480 = 10.94 scores.

How many scores would be 54 or less?

Graph the problem:

54   X̄ = 60

total = 480   s = 8

z = (X - X̄) / s   =   (54 - 60) / 8   =   - 6 / 8 = - 0.75

Use table by *ignoring* the sign of z : "area beyond z" for 0.75 = 0.2266. Thus 22.7% of scores (109 scores) are 54 or less.

Word comprehension test scores:

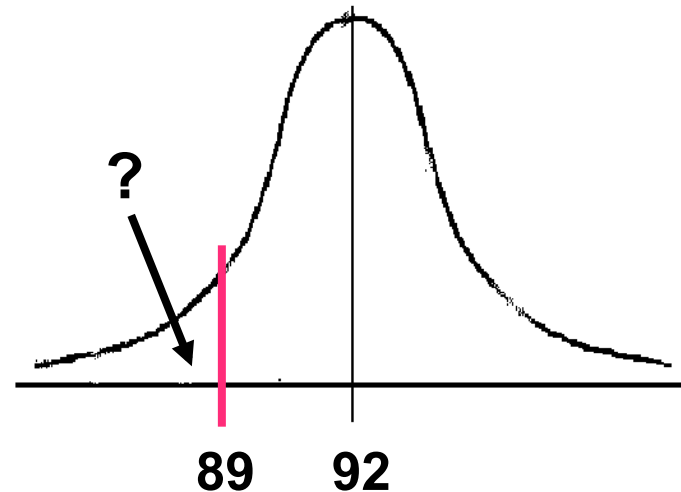**Normal no. correct: mean = 92, SD = 6 out of 100**

Brain-damaged person's no. correct: 89 out of 100.

Is this person's comprehension significantly impaired?
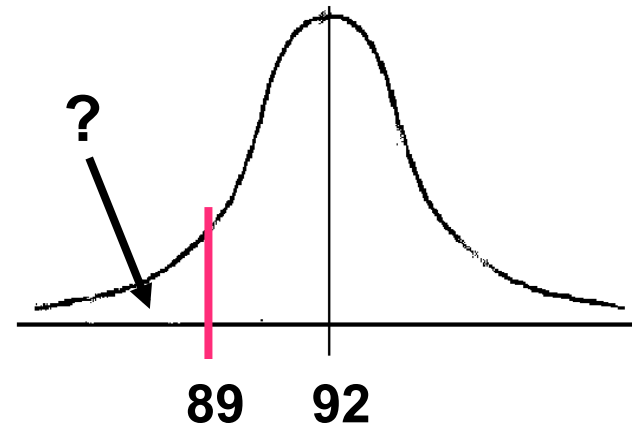
Step 1: graph the problem:

Step 2: convert 89 into a z-score:

z = (89 - 92) / 6 = - 3 / 6 =   - 0.5

?

89   92

Step 3: use the table to find the "area beyond z" for our z of - 0.5:

Area beyond z = 0.3085

89   92

Conclusion: .31 (31%) of normal people are likely to have a comprehension score this low or lower.

z-score value:	Area between the mean and z:	Area beyond z:
0.44	0.17	0.33
0.45	0.1736	0.3264
0.46	0.1772	0.3228
0.47	0.1808	0.3192
0.48	0.1844	0.3156
0.49	0.1879	0.3121
0.5	0.1915	0.3085
0.51	0.195	0.305
0.52	0.1985	0.3015
0.53	0.2019	0.2981
0.54	0.2054	0.2946
0.55	0.2088	0.2912
0.56	0.2123	0.2877
0.57	0.2157	0.2843
0.58	0.219	0.281
0.59	0.2224	0.2776
0.6	0.2257	0.2743
0.61	0.2291	0.2709

Conclusions:

Many psychological/biological properties are normally distributed.

This is very important for statistical inference (extrapolating from samples to populations - more on this in later lectures...).

z-scores provide a way of

(a) comparing scores on different raw-score scales;

(b) showing how a given score stands in relation to the overall set of scores.

Conclusions:

The logic of z-scores underlies many statistical tests.

1. Scores are normally distributed around their mean.

2. Sample means are normally distributed around the population mean.

3. Differences between sample means are normally distributed around zero ("no difference").

We can exploit these phenomena in devising tests to help us decide whether or not an observed difference between sample means is due to chance.