Using Measures of Center and Spread

Tchebysheff's Theorem

Given a number k greater than or equal to 1 and a set of n measurements, at least $1-\frac{1}{k^2}$ of the measurement will lie within k standard deviations of the mean.
That is,

- ✓ If $k = 2$, at least $1-\frac{1}{2^2}$ = **3/4** of the measurements are within **2** standard deviations of the mean.
- ✓ If $k = 3$, at least $1-\frac{1}{3^2}$ = **8/9** of the measurements are within **3** standard deviations of the mean.

The Empirical Rule

Given a distribution of measurements that is approximately mound-shaped:

- ✓ The interval $\mu \pm \sigma$ contains approximately **68%** of the measurements.
- ✓ The interval $\mu \pm 2\sigma$ contains approximately **95%** of the measurements.
- ✓ The interval $\mu \pm 3\sigma$ contains approximately **99.7%** of the measurements.

***z*-Scores**

- From Tchebysheff's Theorem and the Empirical Rule
  - At least 3/4 and more likely 95% of measurements lie within 2 standard deviations of the mean.
  - At least 8/9 and more likely 99.7% of measurements lie within 3 standard deviations of the mean.
- *z*-scores between –2 and 2 are ***not unusual***. *z*-scores should not be more than 3 in absolute value. *z*-scores **larger than 3** in absolute value would indicate a possible **outlier.**

Example:
The length of time for a worker to complete a specified operation averages 12.8 minutes with a standard deviation of 1.7 minutes. If the distribution of times is approximately mound-shaped, what proportion of workers will take longer than 16.2 minutes to complete the task?

$\mu = 12.8 \qquad \sigma = 1.7 \qquad x = 16.2$

$z = \frac{16.2-12.8}{1.7} = \frac{3.4}{1.7} = 2$ That is, 16.2 is 2 standard deviations away from mean.

According to Empirical Rule, there are 95% of measurements are within 2 standard deviations away from mean. That is, 5% of measurements are larger than 2 standard deviations away from mean. The distribution shape is symmetric. Therefore, there are 2.5% of measurements are above 16.2.

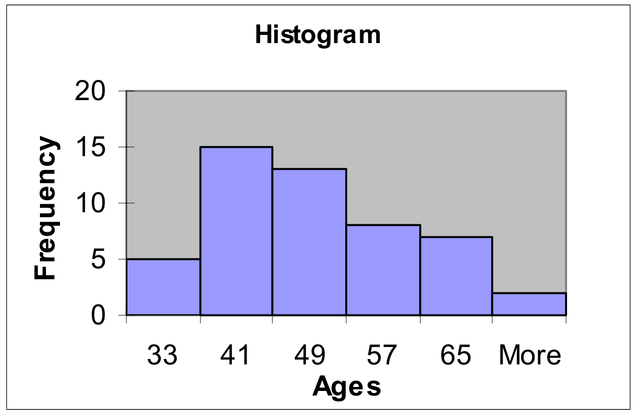
Example:

The following are the ages of 50 tenured faculties at a state university.

34	48	**70**	63	52	52	35	50	37	43	53	43	52	44
42	31	36	48	43	**26**	58	62	49	34	48	53	39	45
34	59	34	66	40	59	36	41	35	36	62	34	38	28
43	50	30	43	32	44	58	53						

Ages	
Mean	**44.9**
Standard Error	1.517046
Median	43
Mode	34
Standard Deviation	**10.72714**
Sample Variance	115.0714
Kurtosis	-0.6245
Skewness	0.39298
Range	44
Minimum	26
Maximum	70
Sum	2245
Count	50

Shape? Skewed right



k	$\pm ks$	Interval	Proportion in Interval	Tchebysheff	Empirical Rule
1	**44.9 ±10.73**	**34.17 to 55.63**	**31/50 (.62)**	**At least 0**	**≈ .68**
2	**44.9 ±21.46**	**23.44 to 66.36**	**49/50 (.98)**	**At least .75**	**≈ .95**
3	**44.9 ±32.19**	**12.71 to 77.09**	**50/50 (1.00)**	**At least .89**	**≈ .997**

Do the actual proportions in the three intervals agree with those given by Tchebysheff's Theorem?

Yes. Tchebysheff's Theorem must be true for any data set.

Do they agree with the Empirical Rule?

No. Not very well.

The data distribution is not very mound-shaped, but skewed right.