

histograms, mean, standard deviation

how to describe single variables

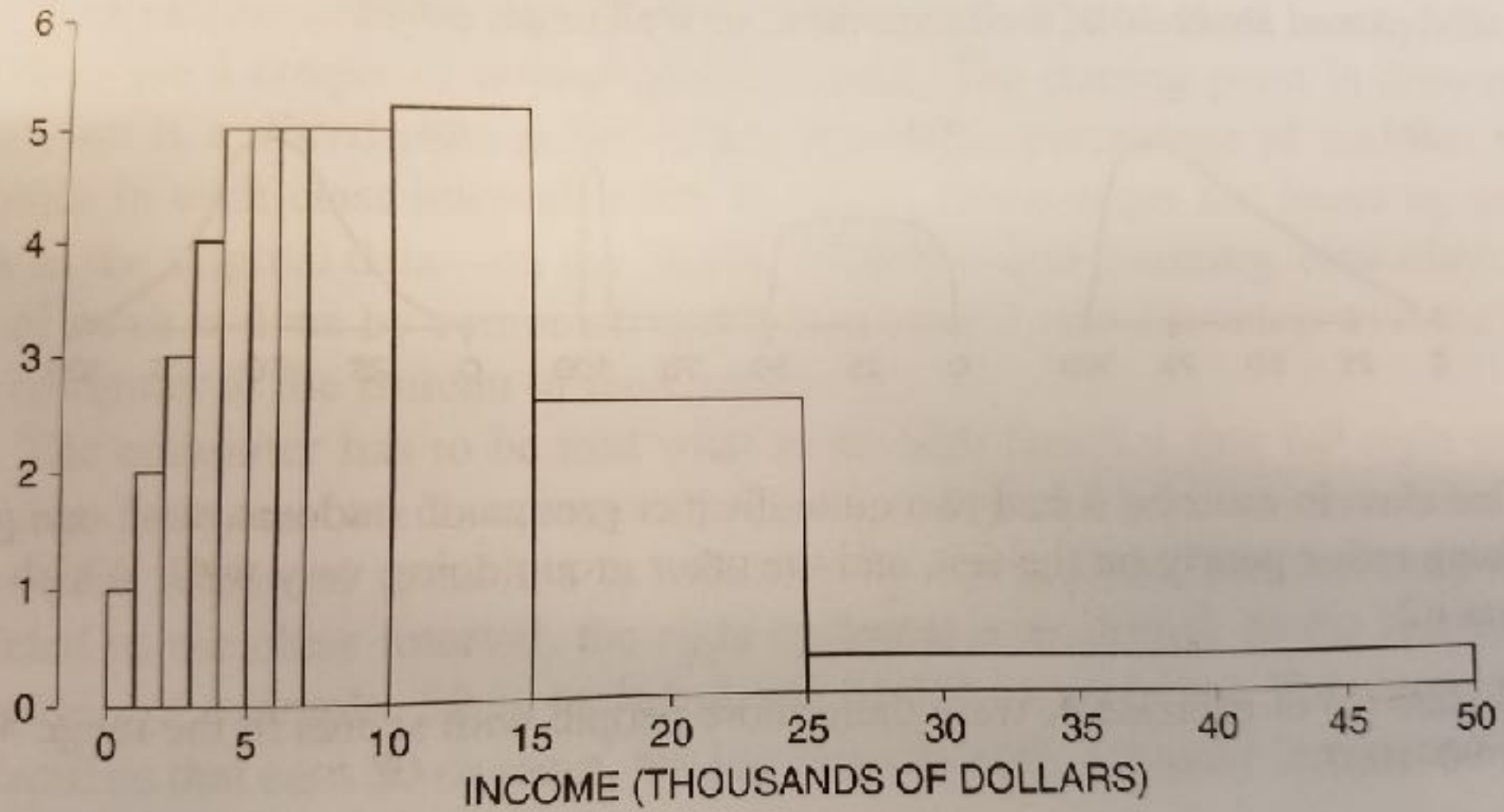
Histogram

- Histogram describes the distribution of a single variable
- How many observations are in each block (class intervals) of data?

Histograms

- We can also use histograms to compare distribution of a variable for subgroups
- Area of each block represents the percentage of that group in the data
- Size of intervals often differ, so height of blocks not necessarily indicative of size of group

Figure 2. The histogram from figure 1, with a vertical scale supplied.



How to draw histogram by hand

1. Draw distribution table -> percentage of group in each interval
2. Label axis, make sure breaks are equidistant on x-axis
3. Draw blocks, but height is not equal to their percentage!
4. Height of block is the percentage divided by interval length

Average and Standard Deviation Practice

- Which of these has the larger SD? Guess!
 - 9, 9, 10, 10, 10, 12
 - 7, 8, 10, 11, 11, 13

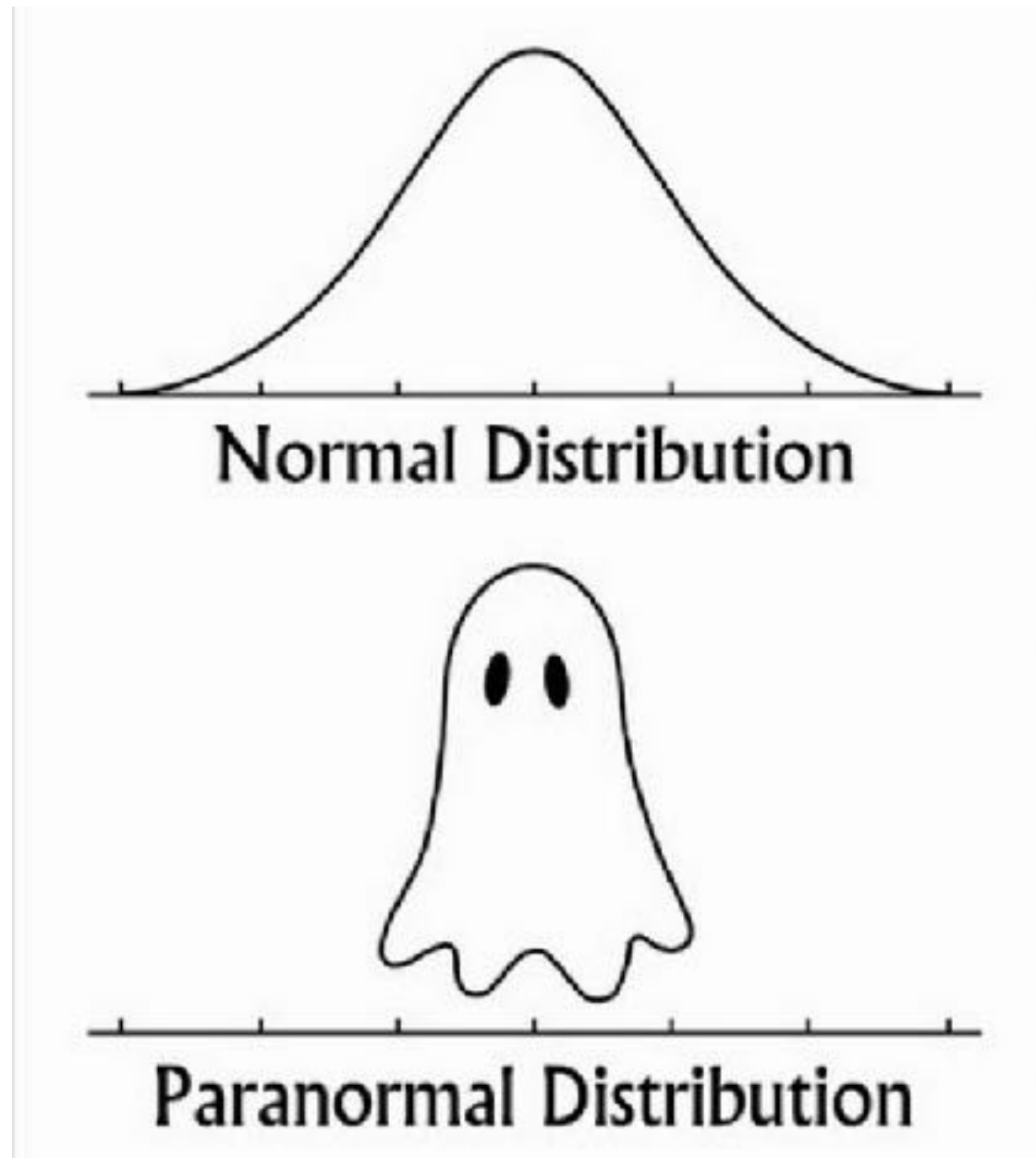
Average and Standard Deviation Practice

- Which of these has the larger SD? Now calculate the SD for both
 - 9, 9, 10, 10, 10, 12
 - 7, 8, 10, 11, 11, 13

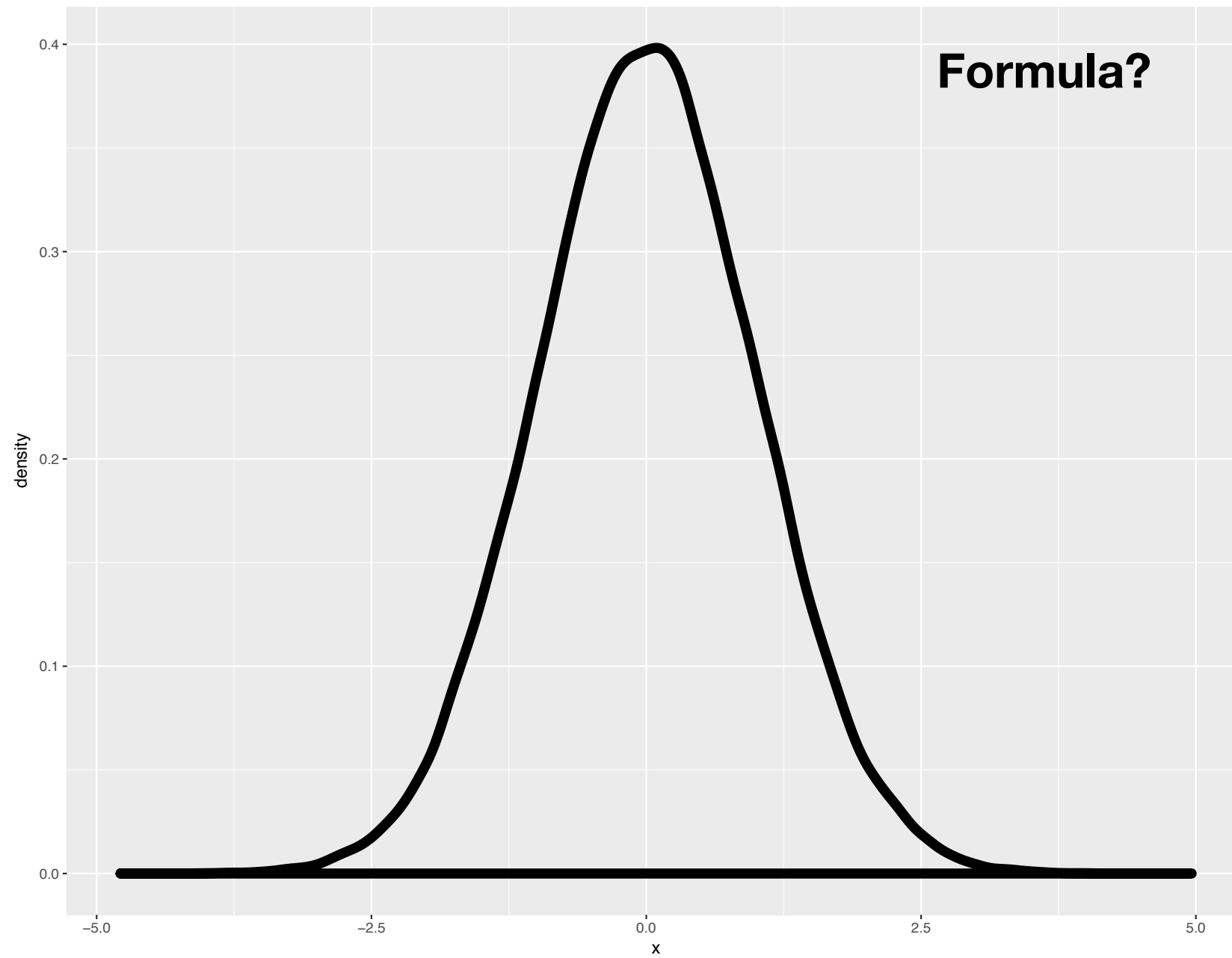
Average and Standard Deviation Practice

- Calculate the average, deviation from average, and SD for each of the two lists:
 - 5, -4, 3, -1, 7
 - -5, 4, -3, 1, -7

Normal Approximation



The normal (Gaussian) distribution (curve)



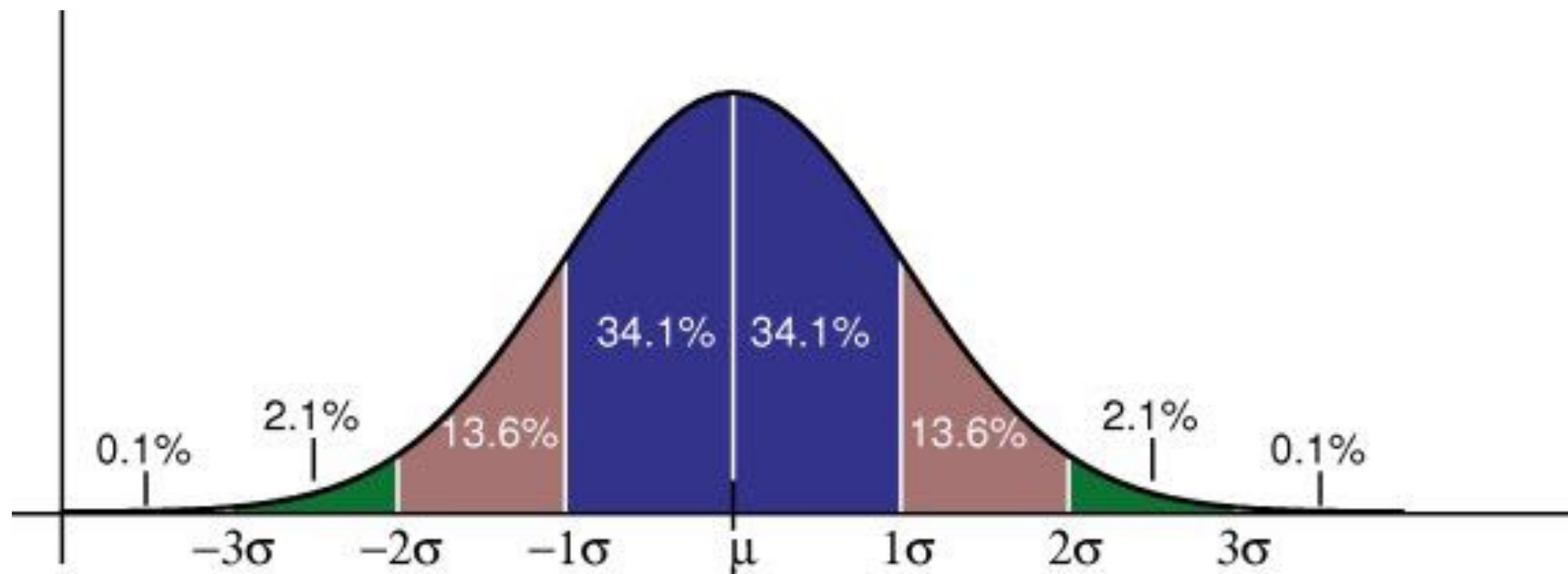
Normal curve

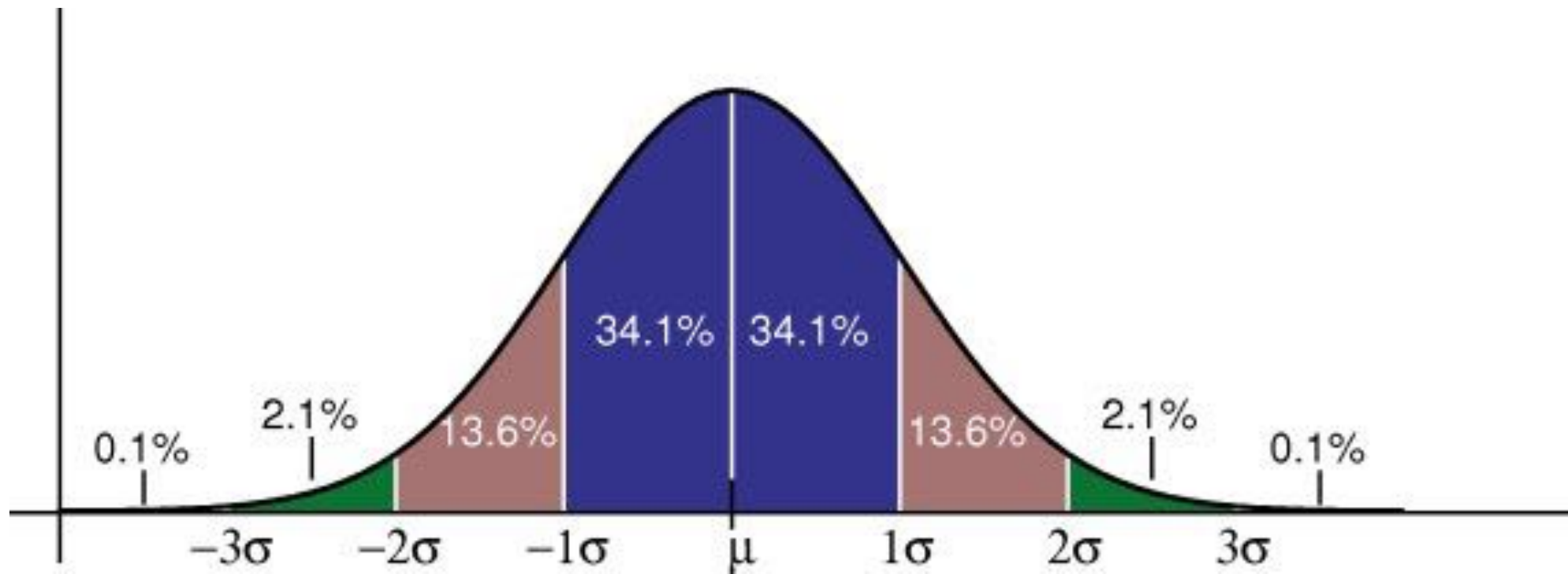
- Symmetric around 0
- Total area under the curve is 100%
- Area between -1 and 1 is ~68%
- Area between -2 and 2 is ~95%
- Area between -3 and 3 is ~99.7%

- We can match up other data that is in similar shape to the normal curve
- To do some histogram has to be on same scale, this is often done by using standard units on the x-axis
- One Standard Unit is the difference to mean, divided by one standard deviation

- Average score on exam was 50 and standard deviation was 10
- What are these scores in standard units: 60, 45, 75
 - What are the scores for the following standard units: 0, +1.5, -2.8

- It is quite easy to figure out areas under the normal curve, the important thing is to remember symmetry!





Find the area:

- 1. To the right of 1.25**
- 2. To the left of 0.80**
- 3. To the left of -0.4**
- 4. Between 0.4 and 1.30**
- 5. Between -0.3 and 0.9**
- 6. Outside of -1.5 to 1.5**

Tables



A NORMAL TABLE

<i>z</i>	<i>Height</i>	<i>Area</i>	<i>z</i>	<i>Height</i>	<i>Area</i>	<i>z</i>	<i>Height</i>	<i>Area</i>
0.00	39.89	0	1.50	12.95	86.64	3.00	0.443	99.730
0.05	39.84	3.99	1.55	12.00	87.89	3.05	0.381	99.771
0.10	39.69	7.97	1.60	11.09	89.04	3.10	0.327	99.806
0.15	39.45	11.92	1.65	10.23	90.11	3.15	0.279	99.837
0.20	39.10	15.85	1.70	9.40	91.09	3.20	0.238	99.863
0.25	38.61	19.74	1.75	8.53	91.99	3.25	0.203	99.885
0.30	38.14	23.58	1.80	7.90	92.81	3.30	0.172	99.903
0.35	37.52	27.37	1.85	7.21	93.57	3.35	0.146	99.919
0.40	36.85	31.08	1.90	6.56	94.26	3.40	0.123	99.933
0.45	36.05	34.73	1.95	5.96	94.88	3.45	0.104	99.944
0.50	35.21	38.29	2.00	5.40	95.45	3.50	0.087	99.953
0.55	34.29	41.77	2.05	4.88	95.96	3.55	0.073	99.961
0.60	33.32	45.15	2.10	4.40	96.43	3.60	0.061	99.968
0.65	32.30	48.43	2.15	3.96	96.84	3.65	0.051	99.974
0.70	31.23	51.61	2.20	3.55	97.22	3.70	0.042	99.978
0.75	30.11	54.67	2.25	3.17	97.56	3.75	0.035	99.982
0.80	28.97	57.63	2.30	2.83	97.86	3.80	0.029	99.986
0.85	27.80	60.47	2.35	2.52	98.12	3.85	0.024	99.988
0.90	26.61	63.19	2.40	2.24	98.36	3.90	0.020	99.990
0.95	25.41	65.79	2.45	1.98	98.57	3.95	0.016	99.992
1.00	24.20	68.27	2.50	1.75	98.76	4.00	0.013	99.9937
1.05	22.99	70.63	2.55	1.54	98.92	4.05	0.011	99.9949
1.10	21.79	72.87	2.60	1.36	99.07	4.10	0.009	99.9959
1.15	20.59	74.99	2.65	1.19	99.20	4.15	0.007	99.9967
1.20	19.42	76.99	2.70	1.04	99.31	4.20	0.006	99.9973
1.25	18.26	78.87	2.75	0.91	99.40	4.25	0.005	99.9979
1.30	17.14	80.64	2.80	0.79	99.49	4.30	0.004	99.9983
1.35	16.04	82.30	2.85	0.69	99.56	4.35	0.003	99.9986
1.40	14.97	83.85	2.90	0.60	99.63	4.40	0.002	99.9989
1.45	13.94	85.29	2.95	0.51	99.68	4.45	0.002	99.9991

- Much of the data we observe in the real world actually follows the normal distribution
- So often histograms when overlaid with the normal curve follow it quite nicely and we can approximate it with the normal curve
- If the histogram is not symmetric, the IQR may be preferred over the standard deviation as a measure of spread

Percentiles

- Percentiles: x percentile meaning x percent below this threshold
- E.G. 99th percentile in a test was a score of 95 means that 99 percent of students scores less or equal to 95

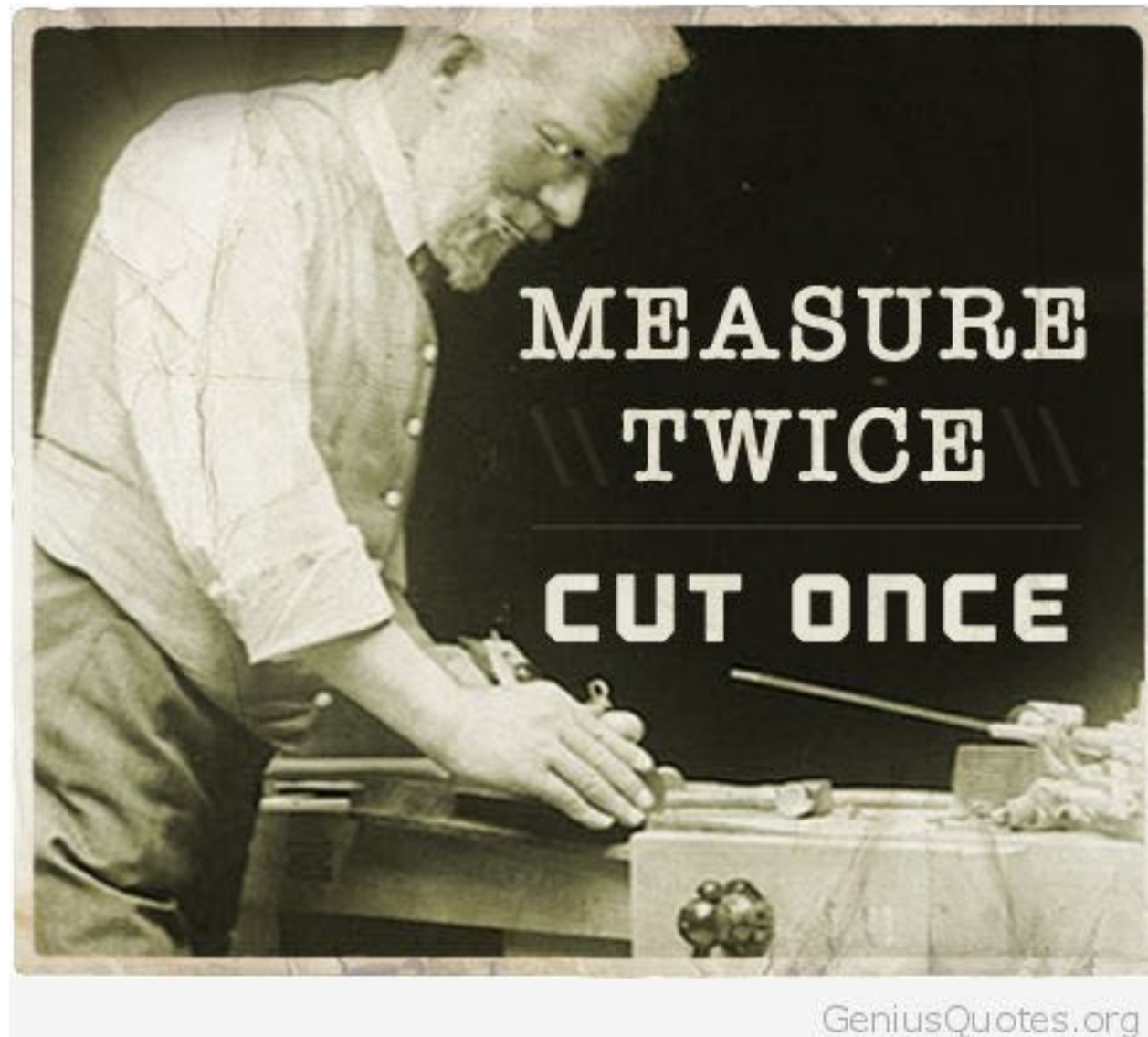
Percentile Rank

- Percentile rank is the percentile associated with a given score
- E.g. in the previous example the percentile rank for a score of 95 is 99

Interquartile Range (IQR)

- 75th percentile - 25th percentile
- Better measure of spread if distribution is not symmetric

Measurement



GeniusQuotes.org

Measurement



- One of (maybe) the most important prerequisite to good science is good measurement
- Without measurement how can we tell whether X or Y changes and by how much?

Measurement

- Once we have a theory, we have to figure out how to measure our main concepts
- How do we measure our dependent and independent variables?

- Even when measurement seems straightforward, in the social sciences it is often not
- Political science is not like physics where we can easily measure some physical quantity
- E.G. Measuring law maker ideology is really hard and complicated

- Researchers have to make many decisions on how to measure the concepts they are interested in
- Recall example about measuring income — what do we include in income?
- At a minimum these decisions need to be transparent

GDP

If the concept of interest is “economic output” (or “Gross Domestic Product”), which is commonly defined as the total sum of goods and services produced by labor and property in a given time period, then it is a relatively straightforward matter to obtain an empirical observation that is consistent with the concept of interest. (Kellstedt & Whitten p.95)

Hold on!

- “The central issue in national income accounting is deciding which economic activities and actors should and can be included in the official accounts” (Morten Jerven)

National Accounting in Zambia in 2010:

- now only one man responsible for data collection on national accounts (GDP)
- also had to do statistics for industrial stats and public finances
- Data is “awful”

Reliability

- “repeatable or consistent”
- If we repeat the measurement on identical objects, we ought to get the same results
- If we have two people independently measure something, they should derive at the same result

Validity

- “A valid measure accurately represents the concept that it is supposed to measure” (Kellstedt & Whitten, page 101)
- **face validity** - on its face, the measure appears to be measuring what it purports to be measuring
- **content validity** - What is the concept to be measured? What are all of the essential elements to that concept and the features that define it? And have you excluded all of the things that are not it?
- **construct validity** - the degree to which the measure is related to other measures that theory requires them to be related to

Measuring Democracy

Adam Przeworski:

- Democracy is a system in which parties lose elections.

Juan Linz:

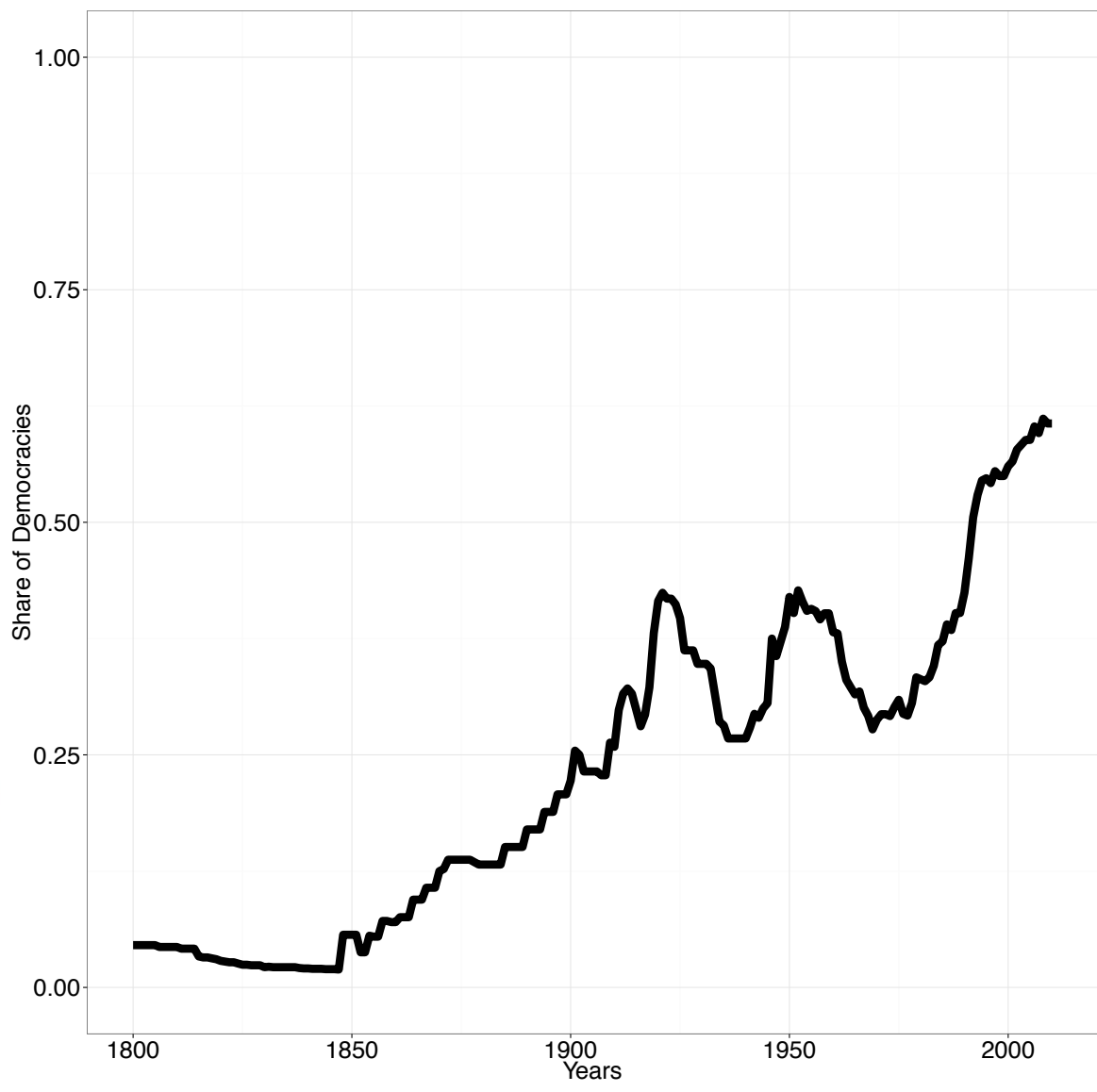
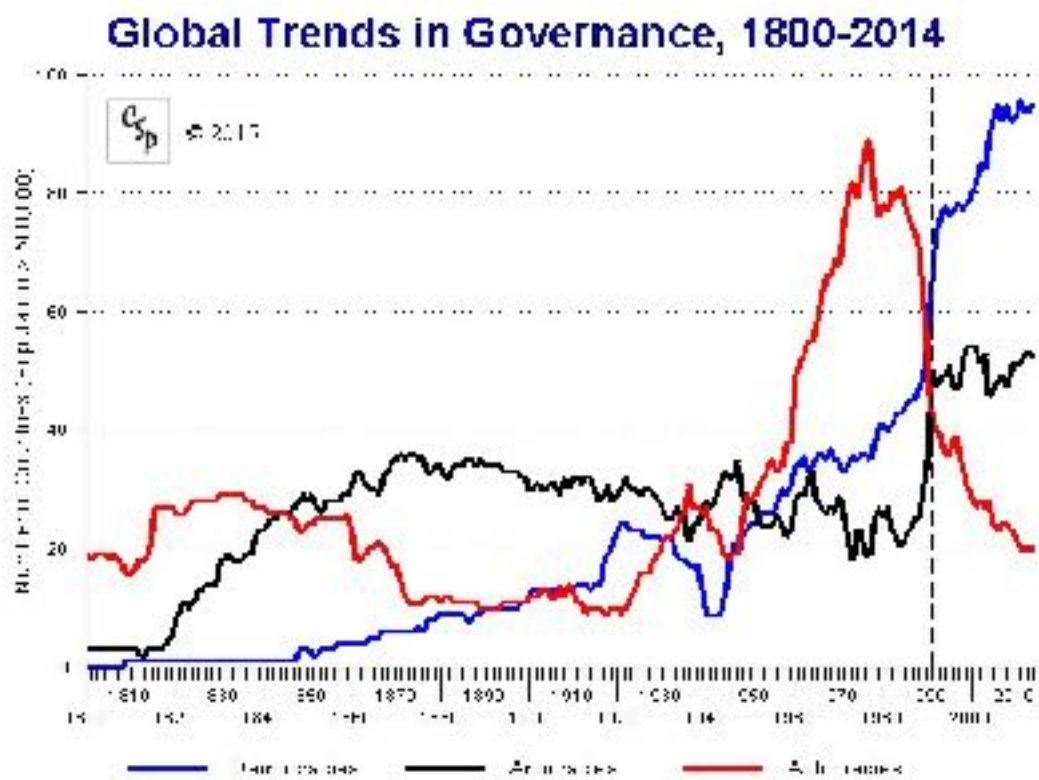
- Democracy is government pro tempore (for time being)

Kenneth A. Bollen:

- the extent to which the political power of the elite is minimized and that of the nonelite is maximized. Political rights and political liberties are two major dimensions of the concept and these encompass most of the traits usually attributed to democratic systems

Measuring Democracy

- Binary or Continuous?
- Countries are democratic or not!
- Or can we have more or less democratic countries?



Measurements can be wrong

Bias



Systematic error

Measurements can be wrong

Bias



Systematic error

i.e. the measurement is always wrong in the same way

Measurements can be wrong

Chance error  **non-systematic error**

Measurements can be wrong

Chance error  **non-systematic error**

i.e. the measurement is wrong but randomly in all directions

Examples

- **Bias**
- **Chance error**