

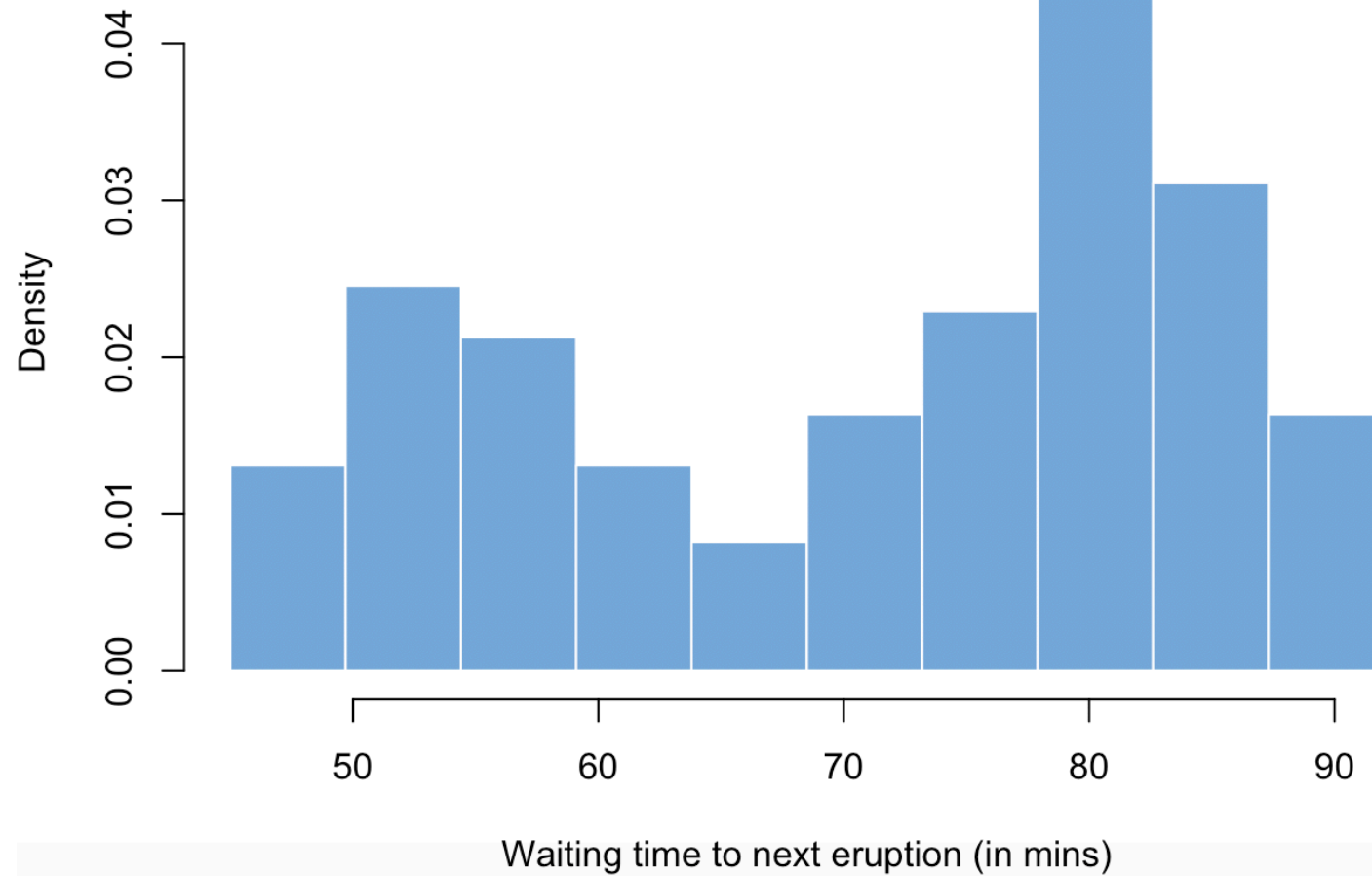
DISCUSSION 2

STAT 05 SHUANGJIE ZHANG

CHAPTER 3

STAT 05

Histogram of waiting times





CHAPTER 4

STAT 05



STATISTICS TERMS

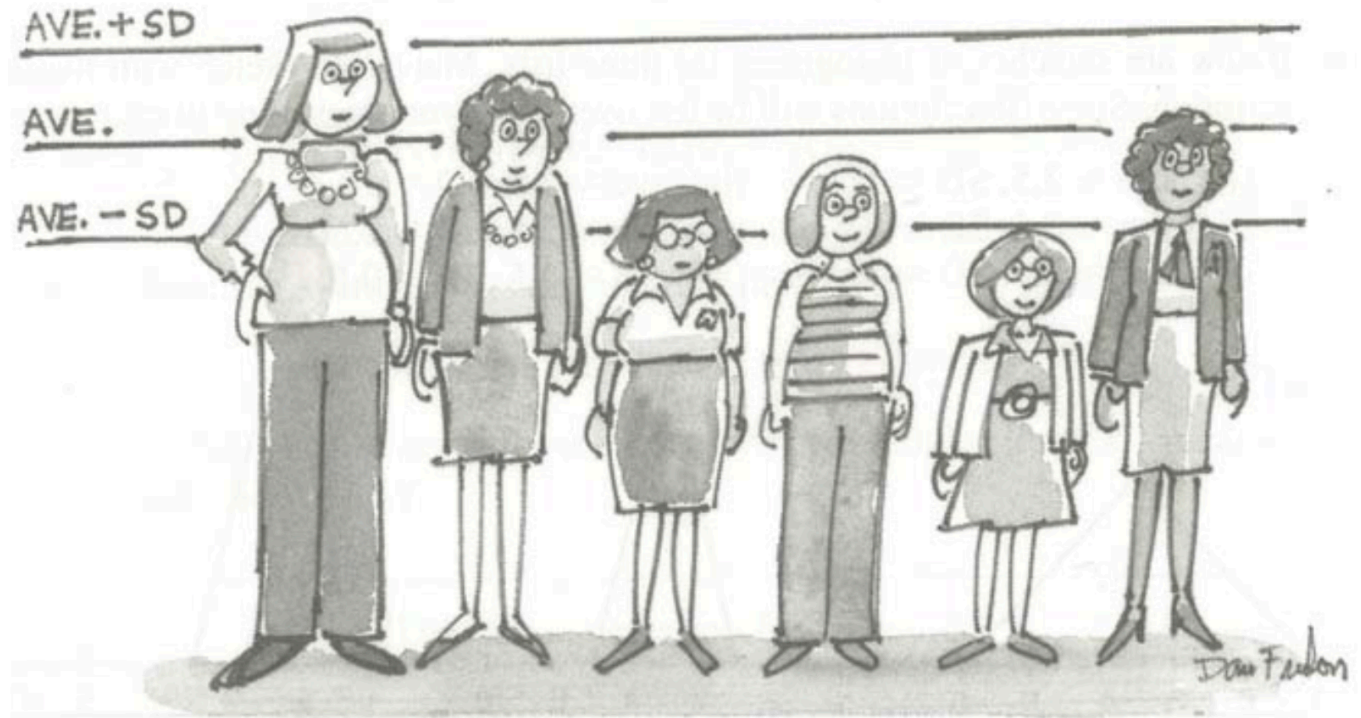
- Average
- Median
- Mode

STANDARD DEVIATION

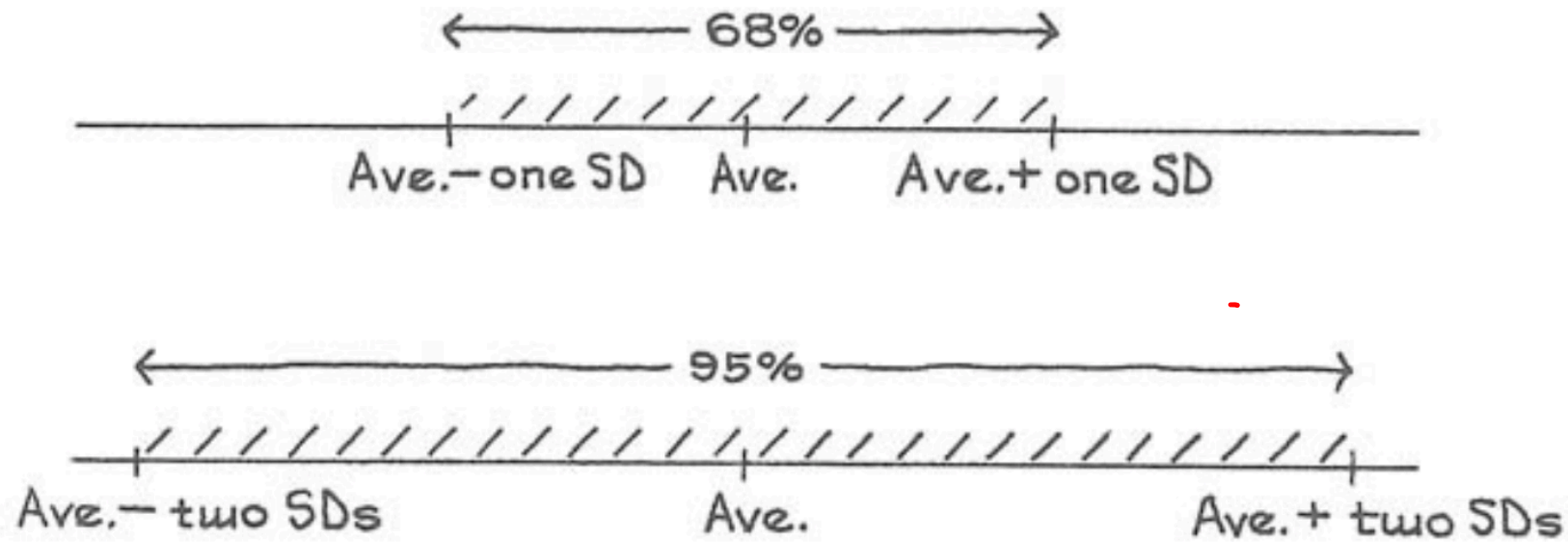
- SD = r.m.s deviation from the average= $\sqrt{\frac{\sum (x_i - \text{mean}(x_i))^2}{n}}$
- 1. Use numbers minus the mean.
- 2. Square the above result.
- 3. Mean of the above result.
- 4. Square root of the above result

STANDARD DEVIATION

- The SD says how far away numbers on a list are from their average.
- Most number on the list will be somewhere around one SD away from the average. Very few will be more than two or three SDs away.



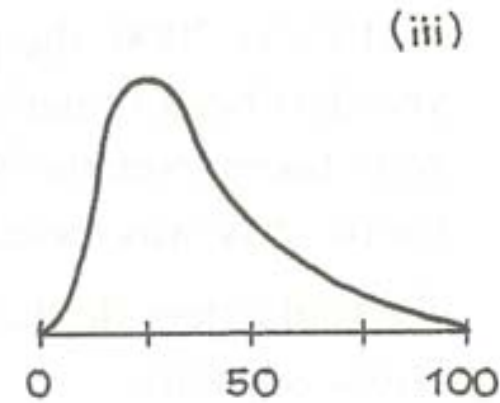
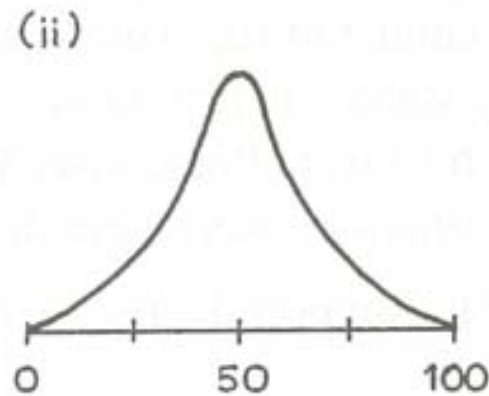
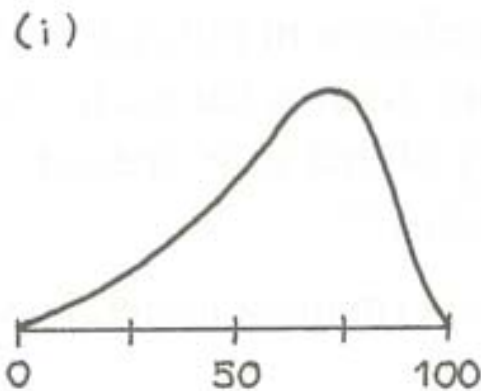
Roughly 68% of the entries on a list (two in three) are within one SD of the average, the other 32% are further away. Roughly 95% (19 in 20) are within two SDs of the average, the other 5% are further away. This is so for many lists, but not all.



CH 4 Q 6

6. Below are sketches of histograms for three lists.

- (a) In scrambled order, the averages are 40, 50, 60. Match the histograms with the averages.
- (b) Match the histogram with the description:
 - the median is less than the average
 - the median is about equal to the average
 - the median is bigger than the average
- (c) Is the SD of histogram (iii) around 5, 15, or 50?
- (d) True or false, and explain: the SD for histogram (i) is a lot smaller than that for histogram (iii).



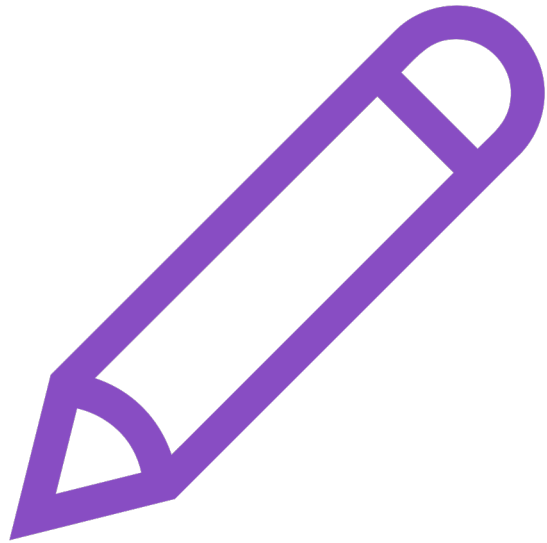
CROSS SECTIONAL/LONGITUDINAL DATA

- Cross sectional data
 - – picture of a population at a given point in time, and one of the variables measured is age/ year.
 - In a cross-sectional study, different subjects are compared to each other at one point in time.
- Longitudinal data
 - – each individual is followed as time goes by.
 - In a longitudinal study, subjects are **tracked** over time, and **compared with themselves** at different points in time.



EXAMPLE

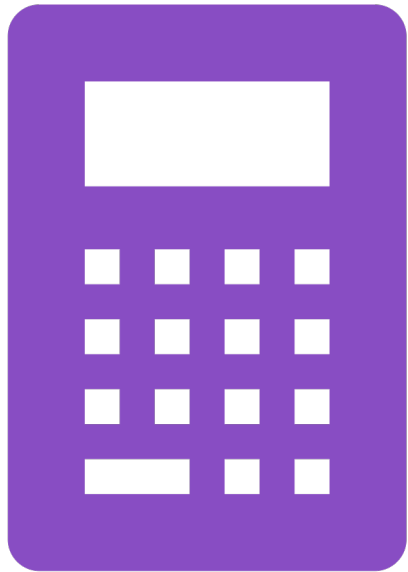
- Whether average height of men appears to decrease after age 50?
- 1. Take a sample of people from different ages and calculate the average height in each age.
- 2. Take a sample of people at their 20s. And record height data each year from their 20s to 60s. Calculate the average height in each age.



CHAPTER 5

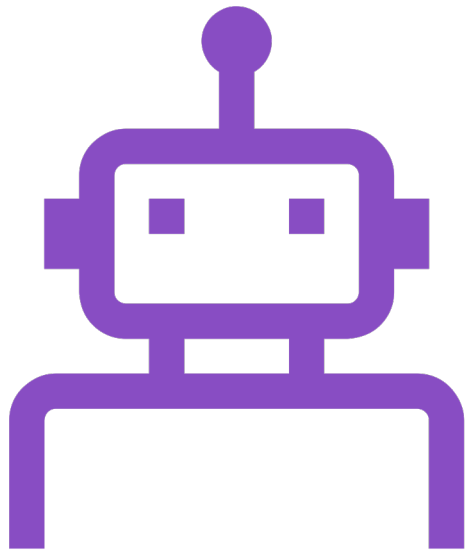
STANDARD NORMAL DISTRIBUTION(CURVE)

- 1. Bell shape
- 2. Mean = 0
- 3. SD = 1
- 4. Variance = 1



Z SCORE

- $Z \text{ score} = (\text{number} - \text{average}) / \text{SD}$
- Z score says how many SDs a value is above (+) or below (-) the average.
- Standardization: Convert all number into z score

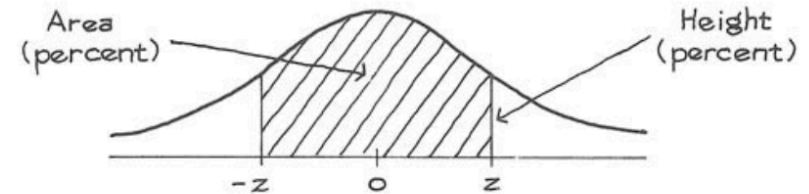


PERCENTILE

- A number related to rank
- Under standard normal curve
- -1 is 16th percentile means that numbers less than -1 accounts for 16%
- 2 is 97.5th percentile means that numbers less than 2 accounts for 97.5% or only 2.5% numbers are larger than 2.
- 1 is ? % percentile means.... 84%
- -2 is ? % percentile means.... 2.5%

PRACTICE

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	38.67	19.74	1.75	8.63	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	36.83	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

- A normal data with Mean = 50, SD =10
- 1. Find 20th percentile for this data

CH 5 Q 11

10. For women age 25–34 with full time jobs, the average income in 2004 was \$32,000. The SD was \$26,000, and 1/4 of 1% had incomes above \$150,000. Was the percentage with incomes in the range from \$32,000 to \$150,000 about 40%, 50%, or 60%? Choose one option and explain briefly.⁵