Chapter 3

Descriptive Statistics: Numerical Methods

Numerical Methods

✓ Ungrouped Data
- Measures of Location
  - Mean
    - Population $\mu = \frac{\sum x_i}{N}$
    - Sample $\bar{x} = \frac{\sum x_i}{n}$

Numerical Methods

Measures of Location

Mode – X_o - the observation with the highest frequency

12	14	19	18
15	15	18	17
20	27	22	23
22	21	33	28
14	18	16	13

•Advantage: easy to calculate

•Disadvantage: solution not unique

Numerical Methods

– Measures of Location
  - Percentiles – the p^th percentile is a value such that at least p percent of the observations are less than or equal to this value.

– Steps in calculating the p^th percentile.
  - Step 1. Order the data
  - Step 2. Compute the index ***i*** $i = \left(\frac{p}{100}\right)n$
  - Step 3.

–If ***i*** is not an integer: The next integer greater than ***i*** denotes the position of the p^th percentile.
–If ***i*** is an integer: The pth percentile is the average of the values in positions i and i+1.

Numerical Methods

Find the Median value. This is the 50^th percentile. P_{50}

	Student Income
X_1	9,800
X_2	10,000
X_3	10,500
X_4	11,000
X_5	12,000

i is not an integer:
The next integer greater than ***i*** denotes the position of the p^th percentile.

Numerical Methods

Find the 20^th percentile. P_{20}

	Student Income
X_1	9,800
X_2	10,000
X_3	10,500
X_4	11,000
X_5	12,000

i is an integer: The pth percentile is the average of the values in positions i and i+1.

Numerical Methods

– Measures of Variability

- Range = highest value - lowest value

A	B
4	1
5	2
5	3
5	5
5	8
6	11

Numerical Methods

– Interquartile Range: IQR

- IQR= Q_3 - Q_1 = P_{75} – P_{25}

A	B
4	1
5	2
5	3
5	5
5	8
6	11

For P_{75}:

For P_{25}:

Numerical Methods

– Measures of Variability

- Average Deviation = the average of positive difference of each observation from the mean

A	B	C
4	1	1
5	2	3
5	3	4
5	5	5
5	8	6
6	11	11

$$AD = \frac{\sum |x_i - \mu|}{N}$$

Numerical Methods

– Measures of Variability

- Average Deviation = the average of positive difference of each observation from the mean

A	B	$\|x_i-\mu\|$	C
4	1	4	1
5	2	3	3
5	3	2	4
5	5	0	5
5	8	3	6
6	11	6	11

$$AD = \frac{\sum |x_i - \mu|}{N}$$

Numerical Methods

– Measures of Variability

- Average Deviation = the average of positive difference of each observation from the mean

A	B	$\|x_i-\mu\|$	C	$\|x_i-\mu\|$
4	1	4	1	4
5	2	3	3	2
5	3	2	4	1
5	5	0	5	0
5	8	3	6	1
6	11	6	11	6

$$AD = \frac{\sum |x_i - \mu|}{N}$$

Numerical Methods

– Measures of Variability

- Variance = the average of squared differences of each observation from the mean

A	B	C	$(x_i-\mu)^2$	D	$(x_i-\mu)^2$
4	1	1		1	
5	2	3		2	
5	3	4		5	
5	5	5		5	
5	8	6		6	
6	11	11		11	

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

Numerical Methods

– Measures of Variability
  - Variance = the average of squared difference of each observation from the mean

A	B	C	$(x_i-\mu)^2$	D	$(x_i-\mu)^2$
4	1	1	16	1	
5	2	3	4	2	
5	3	4	1	5	
5	5	5	0	5	
5	8	6	1	6	
6	11	11	36	11	

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

Numerical Methods

– Measures of Variability
  - Variance = the average of squared difference of each observation from the mean

A	B	C	$(x_i-\mu)^2$	D	$(x_i-\mu)^2$
4	1	1	16	1	16
5	2	3	4	2	9
5	3	4	1	5	0
5	5	5	0	5	0
5	8	6	1	6	1
6	11	11	36	11	36

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

Numerical Methods

– Measures of Variability
  - Standard Deviation - positive square root of the variance

A	B	C	$(x_i-\mu)^2$	D	$(x_i-\mu)^2$
4	1	1	16	1	16
5	2	3	4	2	9
5	3	4	1	5	0
5	5	5	0	5	0
5	8	6	1	6	1
6	11	11	36	11	36

$$\sigma = \sqrt{\sigma^2}$$

Variance – ungrouped quantitative data

	Conceptual	Computational
– Population	$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$	$\sigma^2 = \frac{N\sum x_i^2 - \left(\sum x_i\right)^2}{N^2}$
– Sample	$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$	$s^2 = \frac{n\sum x_i^2 - \left(\sum x_i\right)^2}{n(n-1)}$

You must use the computational form of variance on the homework and exams.

Numerical Methods

– Measures of Variability

- Coefficient of Variation - a relative measure of standard deviation. A unitless measure

$D_{(dollars)}$	$E_{(pesos)}$
1	12.29
2	24.58
5	61.45
5	61.45
6	73.74
11	135.19

population $cv = \frac{\sigma}{\mu}$

sample $cv = \frac{s}{\bar{x}}$

Numerical Methods

– Measures of Relative Location

- z score - a standardized value where the distance from the mean is measured in standard deviations
- see Table 3.5, $s = 8.0$, $\bar{x} = 44$

Students	$x_i - \bar{x}$	z score
46	2	0.25
54	10	1.25
42	-2	-0.25
46	2	0.25
32	-12	-1.5

$$z = \frac{(x_i - \bar{x})}{s}$$

Numerical Methods

- Measures of Relative Location
  - z score - a standardized value where the distance from the mean is measured in standard deviations

D(dollars)	E(pesos)	z score
1	12.29	
2	24.58	
5	61.45	
5	61.45	
6	73.74	
11	135.19	

$$z = \frac{(x_i - \bar{x})}{s}$$

Chebyshev's Theorem

- At least $(1 - 1/z^2)$ of the items in any data set must be within z standard deviations of the mean, where z is any value greater than 1.

± z	Chebyshev Proportion
-1≤z≤1	*undefined*
-2≤z≤2	0.75
-3≤z≤3	0.89
-4≤z≤4	0.94

The Empirical Rule

- If the data is nearly bell shaped (normally distributed) then the proportion of items within z standard deviations is given by the following table.

± z	Chebyshev Proportion	Bell Shaped (normal dist)
-1≤z≤1	*undefined*	0.68
-2≤z≤2	0.75	0.95
-3≤z≤3	0.89	0.9973
-4≤z≤4	0.94	nearly 1.0