MATH 1700

Instructor: Mehdi Maadooliat

Chapter 2B

Department of Mathematical and Statistical Sciences

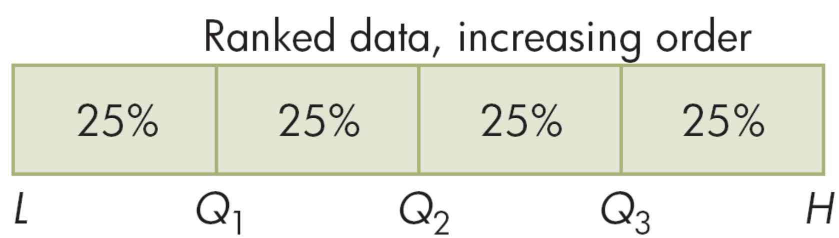
CHAPTER 2B

- Descriptive Analysis
- **Measures of Position**
  - Quartiles
  - Percentile
  - Five number summary
  - Interquartile range (IQR)
- **Box-and-whiskers display**
- **Standard score, or z-score**
- Empirical Rule
- Comparing the measures of center and spread

MEASURES OF POSITION

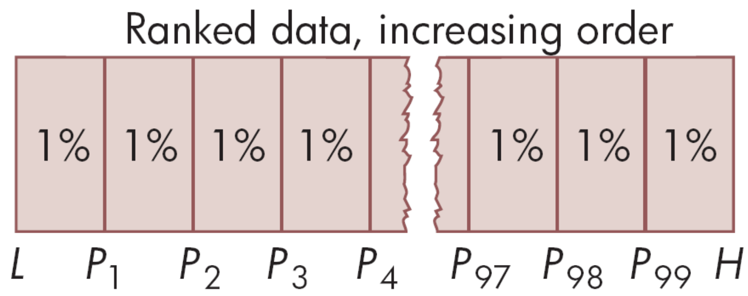
- **Measures of position** are used to describe the position a specific data value possesses in relation to the rest of the data when in ranked order. ***Quartiles*** and ***percentiles*** are two of the most popular measures of position.

- **Quartiles** Values of the variable that divide the ranked data into quarters; each set of data has three quartiles.
  - L = lowest value
  - Q_1 = data value where 25% are smaller
  - Q_2 = $\tilde{x}$ = median
  - Q_3 = data value where 75% are smaller
  - H = highest value

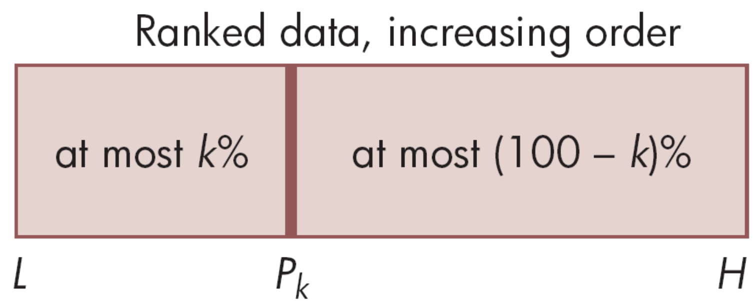


MEASURES OF POSITION

- **Percentiles** Values of the variable that divide a set of ranked data into 100 equal subsets.



- **The k^{th} percentile,**
  - P_k = value where $k\%$ are smaller



- **Quartiles are special percentiles.**

MEASURES OF POSITION

- **The Percentile Process: Finding k^{th} percentile.**

Step 1: Rank the n data, lowest to highest

Step 2: Calculate $\frac{nk}{100}$

- An integer **A** results:
  - Step 3: $d(P_k) = \mathbf{A.5}$
  - Step 4: P_k is halfway between the value of the data in the **A**th position and the value of the data in the **A + 1** position.
  - average of A^{th} and $(A+1)^{th}$ values
- A number with a fraction results:
  - Step 3: $d(P_k) = \mathbf{B}$, the next larger integer
  - Step 4: P_k is the value of the data in the **B**th position.
  - B^{th} value

$d(\cdot)$ is called the depth (indicates the location of k^{th} the percentile)

EXAMPLE 12 - FINDING QUARTILES AND PERCENTILES

- **Using the sample of 50 elementary statistics final exam scores listed in Table 2.15, find the first quartile, Q_1; the 58^{th} percentile, P_{58}.**

60	47	82	95	88	72	67	66	68	98	90	77	86
58	64	95	74	72	88	74	77	39	90	63	68	97
70	64	70	70	58	78	89	44	55	85	82	83	
72	77	72	86	50	94	92	80	91	75	76	78	

Raw Scores for Elementary Statistics Exam **[TA02-06]**

Table 2.15

- **Step 1:**
  - **Rank the data from lowest to highest**

39	64	72	78	89
44	66	72	80	90
47	67	74	82	90
50	68	74	82	91
55	68	75	83	92
58	70	76	85	94
58	70	77	86	95
60	70	77	86	95
63	72	77	88	97
64	72	78	88	98

EXAMPLE 12 – *SOLUTION*

- **Find Q_1:**
  - **Step 2: Find** $\frac{nk}{100}$: $\frac{nk}{100} = \frac{(50)(25)}{100} = \mathbf{12.5}$
  - **Step 3: B is the next larger integer, 13.**
  - **Step 4: Find Q_1: Q_1 is the 13^{th} value, $Q_1 = \mathbf{67}$**

- **Find P_{58}:**
  - **Step 2: Find** $\frac{nk}{100}$: $\frac{nk}{100} = \frac{(50)(58)}{100} = \mathbf{29}$
  - **Step 3: Since $A = 29$, an integer, add 0.5 and use 29.5.**
  - **Step 4: Find P_{58} : P_{58} is the average of 29^{th} and 30^{th} values, $P_{58} = \mathbf{77.5}$**

39	64	72	78	89
44	66	72	80	90
47	67	74	82	90
50	68	74	82	91
55	68	75	83	92
58	70	76	85	94
58	70	77	86	95
60	70	77	86	95
63	72	77	88	97
64	72	78	88	98

13th position from L

29th and 30th positions from L

13th position from H

Stem-and-leaf of score $N = 50$

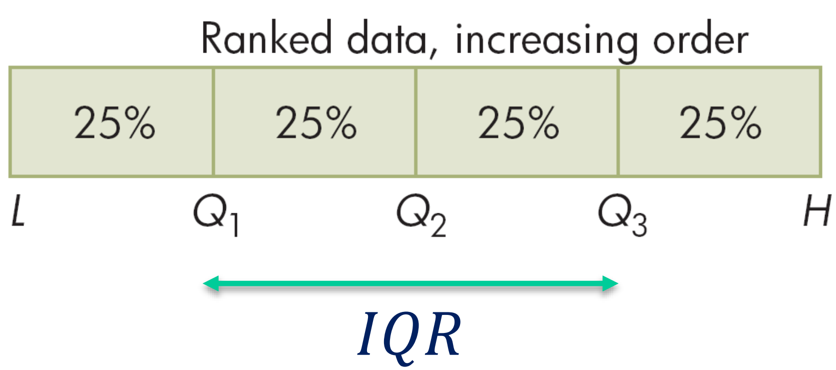
Leaf Unit = 1.0

1	3	9
2	4	4
3	4	7
4	5	0
7	5	5 8 8
11	6	0 3 4 4
15	6	6 7 8 8
24	7	0 0 0 2 2 2 2 4 4
(7)	7	5 6 7 7 7 8 8
19	8	0 2 2 3
15	8	5 6 6 8 8 9
9	9	0 0 1 2 4
4	9	5 5 7 8

MEASURES OF POSITION

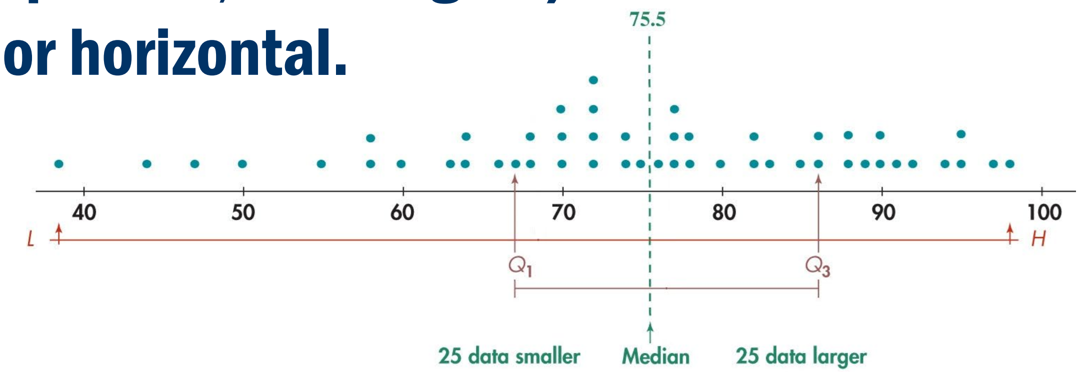
- **Five Number Summary**
  - L = lowest value
  - Q_1= data value where 25% are smaller
  - Q_2 = $\tilde{x}$ =median
  - Q_3 = data value where 75% are smaller
  - H = highest value

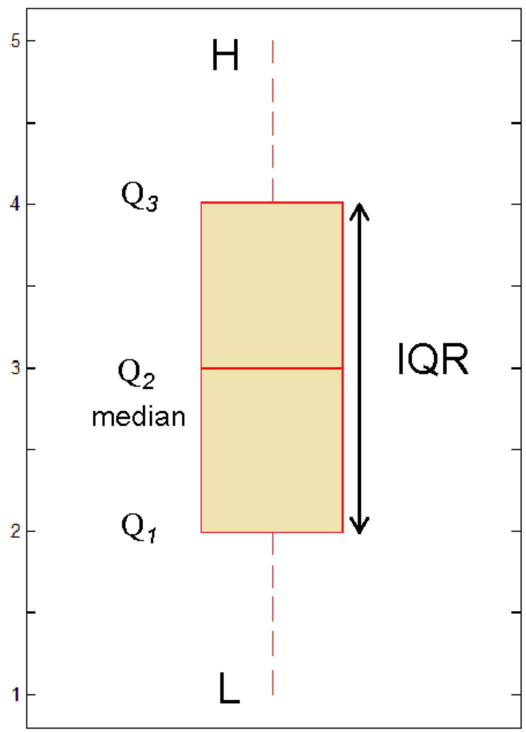
- **Interquartile range: The difference between the first and third quartiles. It is the range of the middle 50% of the data.**
  - $IQR = Q_3 - Q_1$

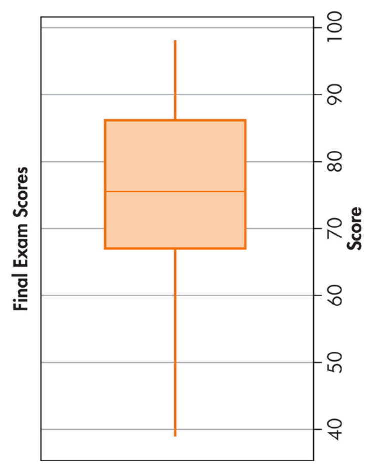


MEASURES OF POSITION

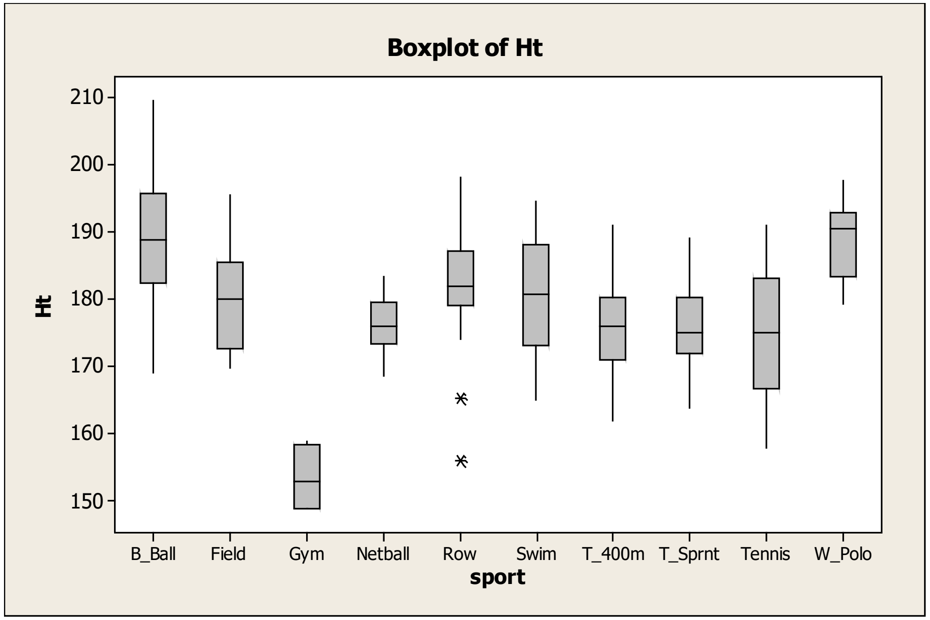
Box-and-whiskers display A graphic representation of the 5-number summary. The five numerical values (smallest, first quartile, median, third quartile, and largest) are located on a scale, either vertical or horizontal.



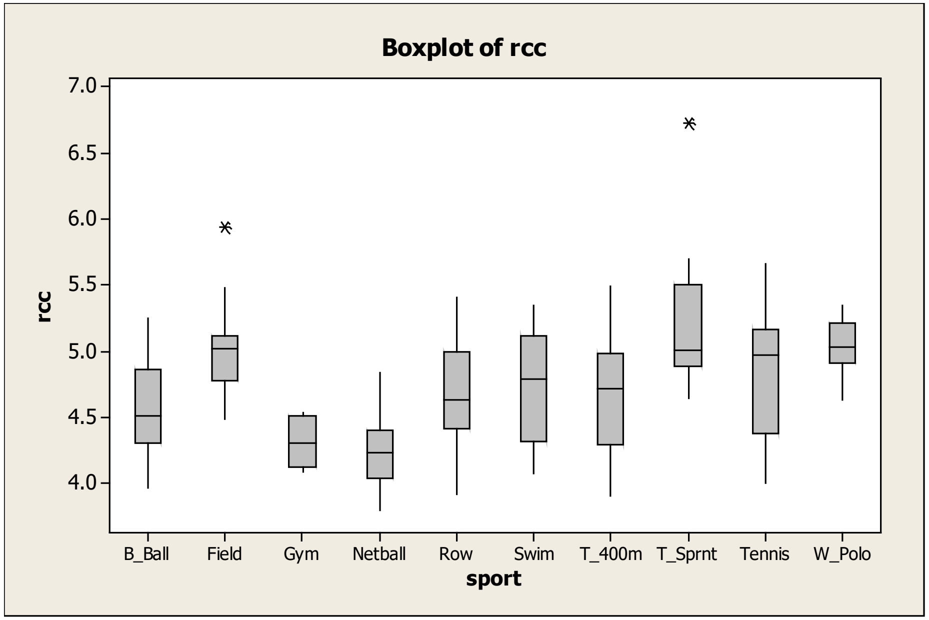




SIDE-BY-SIDE BOX PLOT FOR AIS DATA (AUSTRALIAN INSTITUTE OF SPORT)



- [JAMM: STAT-Calculator]()



MEASURES OF POSITION

- **The position of a specific value can also be measured in terms of the mean and standard deviation using the *standard score*, commonly called the *z-score*.**

- **Standard score, or *z*-score The position a particular value of *x* has relative to the mean, measured in standard deviations. The *z*-score is found by the formula**

$$z = \frac{\text{value} - \text{mean}}{\text{st.dev.}} = \frac{x - \bar{x}}{s} \qquad (2.11)$$

- **Example: Find the standard scores for (a) 92 and (b) 72 with respect to a sample of exam grades that have a mean score of 74.92 and a standard deviation of 14.20.**

EXAMPLE 14 – FINDING Z-SCORES

- **a.** $x_1 = 92,\ \bar{x} = 74.92, s = 14.20$. **Thus,**

$$z = \frac{x - \bar{x}}{s}$$
$$= \frac{92 - 74.92}{14.20} = \frac{17.08}{14.20} = 1.20.$$

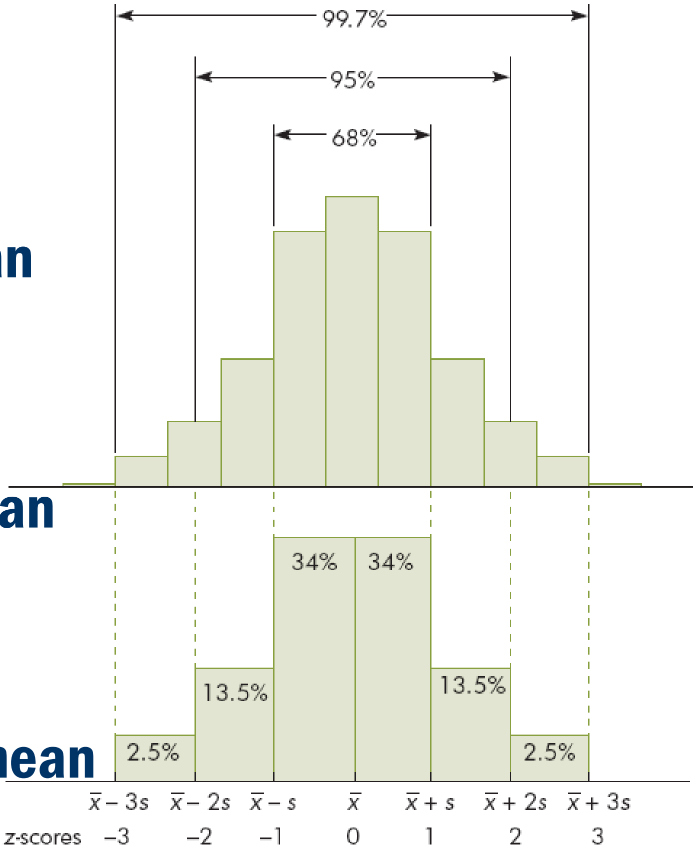
- **b.** $x_2 = 72,\ \bar{x} = 74.92, s = 14.20$. **Thus,**

$$z = \frac{x - \bar{x}}{s}$$
$$= \frac{72 - 74.92}{14.20} = \frac{-2.92}{14.20} = -0.21.$$

- **This means that the score 92 is approximately 1.2 standard deviations above the mean and**
- **that the score 72 is approximately one-fifth of a standard deviation below the mean.**

EMPIRICAL RULE (THE 68-95-99.7 RULE)

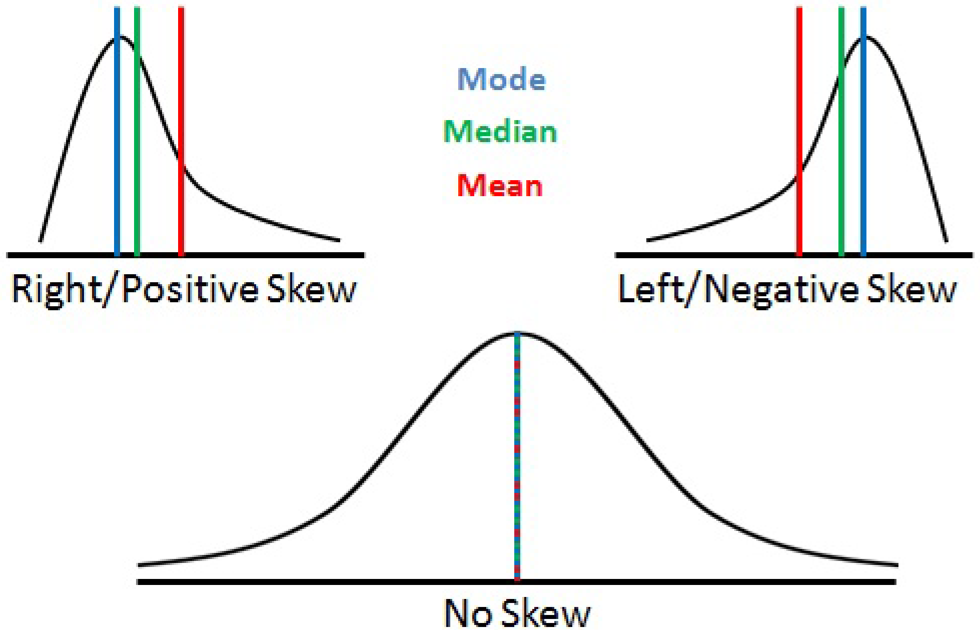
- If the distribution is <u>mound-shaped</u>, then
  - Approximately 68% of the data falls within one standard deviation of the mean
  - Approximately 95% of the data falls within two standard deviations of the mean
  - Approximately 99.7% of the data falls within three standard deviations of the mean



- Read more about this on Section 2.6 of the Book (Chebyshev's Theorem)

COMPARING
MEASURES OF CENTER AND SPREAD

- The **sample mean** and the **sample standard deviation** are good measures of center and spread, respectively, for **symmetric** data
- If the data set is **skewed** or has **outliers**, the **sample median** and the **interquartile range** are more commonly used
- Mean versus median



QUESTIONS?

- **ANY QUESTION?**