Class 3

Daniel B. Rowe, Ph.D.

Department of Mathematical and Statistical Sciences

Agenda:

Recap Chapter 2.1 – 2.4

Lecture Chapter 2.5

Lecture Chapter 3.1

Recap Chapter 2.1 – 2.4

Chapter 2: Descriptive Analysis and Presentation of Single-Variable Data

Daniel B. Rowe, Ph.D.

Department of Mathematical and Statistical Sciences

MARQUETTE UNIVERSITY
Be The Difference.

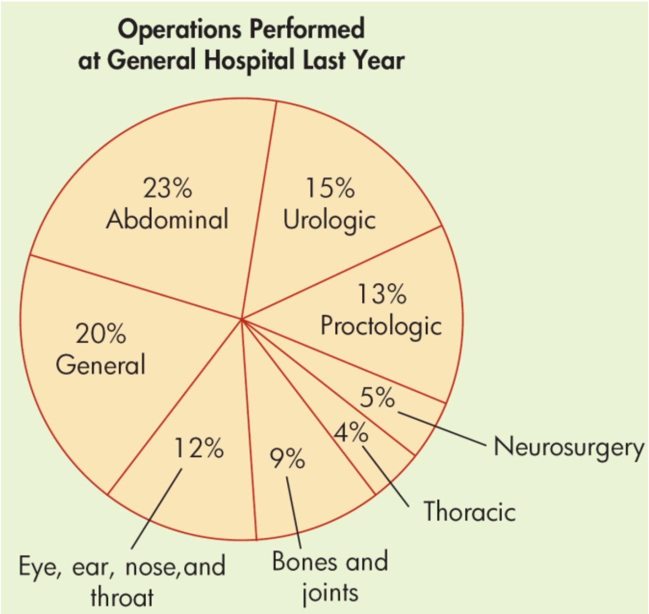
2: Descriptive Analysis and Single Variable Data

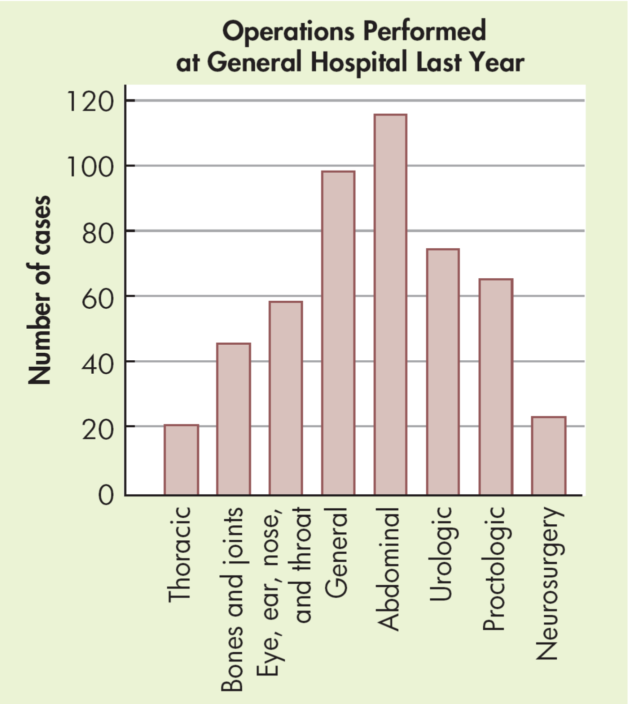
2.1 Graphs - Qualitative Data

Circle (pie) graphs and bar graphs:

Circle is parts to whole as angle.

Bar graph is amount in each category as rectangular areas.





Figures from Johnson & Kuby, 2012.

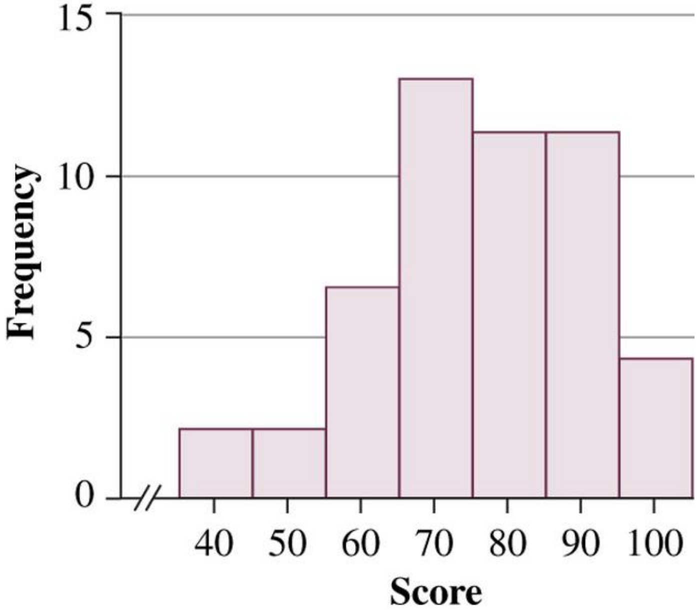
2: Descriptive Analysis and Single Variable Data

2.2 Frequency Distributions and Histograms

Statistics Exam Scores

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	

Boundaries	Frequency
$35 \leq x < 45$	2
$45 \leq x < 55$	2
$55 \leq x < 65$	7
$65 \leq x < 75$	13
$75 \leq x < 85$	11
$85 \leq x < 95$	11
$95 \leq x \leq 105$	4
	50



Figures from Johnson & Kuby, 2012.

2: Descriptive Analysis and Single Variable Data

2.3 Measures of Central Tendency

Sample Mean: Usual average, p. 63 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Sample Median: Middle value, p. 64 n odd, $\tilde{x} = \frac{n+1}{2}$ value

n even, avg $\frac{n}{2}$ & $\frac{n}{2}+1$ values

Sample Mode: Most often, p. 66 $\hat{x} = \text{most often}$

Measures of central tendency characterize center of distribution.

Measures of dispersion characterize the variability in the data.

2: Descriptive Analysis and Single Variable Data

2.4 Measures of Dispersion

Range: $H - L$, p. 74

Deviation from mean: value minus sample mean, p. 74

$$i^{th}\ deviation\ \ from\ mean = x_i - \bar{x}$$

Sample Variance: avg. squared dev using n-1 in den, p. 76

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n}x_i^2 - \left[\left(\sum_{i=1}^{n}x_i\right)^2 / n\right]\right\}$$

Sample Standard Deviation: $s = \sqrt{s^2}$

2: Descriptive Analysis and Single Variable Data

2.3, 2.4 Measures of Central Tendency and Dispersion

Example: Data values: 1,2,2,3,4

$\bar{x} = 2.4$ $\tilde{x} = 2$ $\hat{x} = 2$

$s^2 = 1.3$ $s = 1.1$

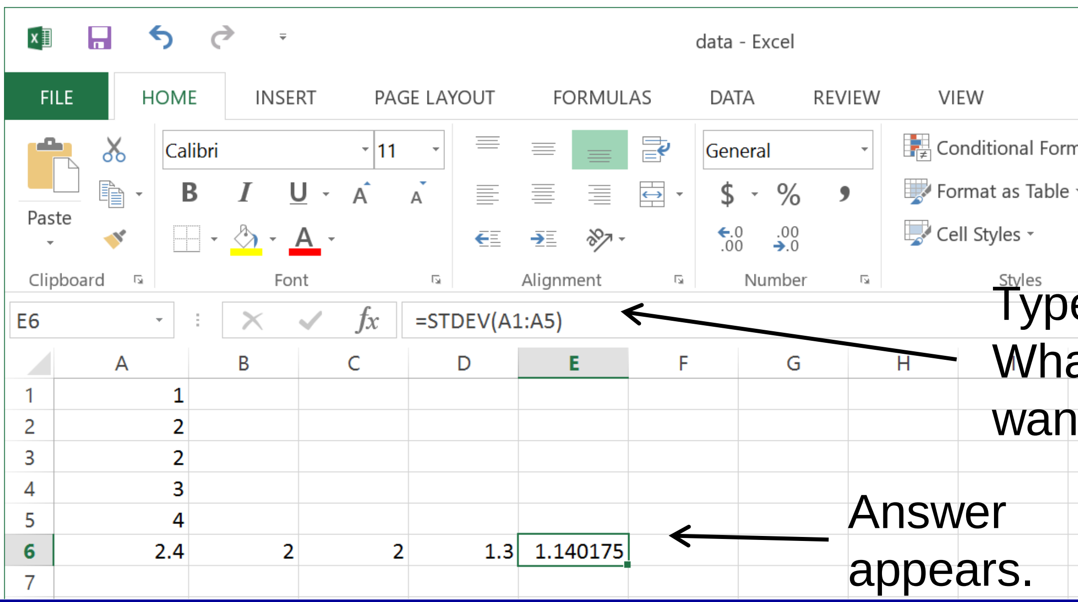
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$\tilde{x} = \text{middle value}$

$\hat{x} = \text{most often value}$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

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



Type in What you want.

Answer appears.

=AVERAGE(A1:A5)
=MEDIAN(A1:A5)
=MODE(A1:A5)
=VAR(A1:A5)
=STDEV(A1:A5)

Chapter 2: Descriptive Analysis and Presentation of Single-Variable Data Continued

Daniel B. Rowe, Ph.D.

Department of Mathematical and Statistical Sciences

2: Descriptive Analysis and Single Variable Data

2.5 Measures of Position

Measures of Position: Describe the relative position a specific data value possesses in relation to rest of data when in ranked order.

Quartiles: Values of the variable that divide ranked data into quarters.

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

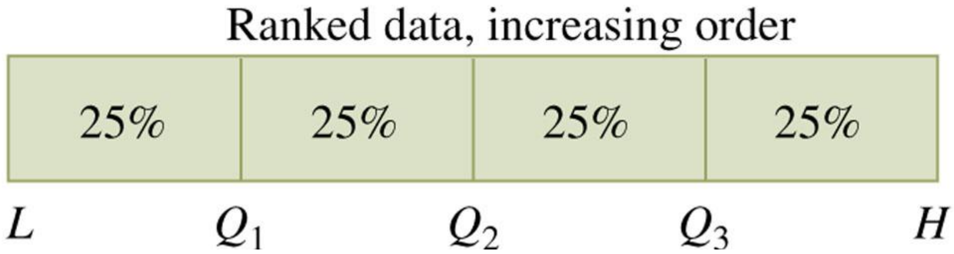


Figure from Johnson & Kuby, 2012.

2: Descriptive Analysis and Single Variable Data

2.5 Measures of Position

5-number summary

1. L = lowest value
2. Q_1 = data value where 25% are smaller
3. $Q_2 = \tilde{x}$ = median (where 50% are smaller)
4. Q_3 = data value where 75% are smaller
5. H = highest value

Interquartile range: The difference between the first and third quartiles. It is the range of the middle 50% of the data.

$$IQR = Q_3 - Q_1$$

2: Descriptive Analysis and Single Variable Data

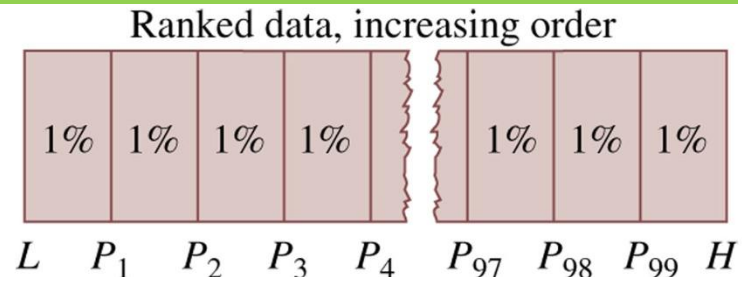
2.5 Measures of Position

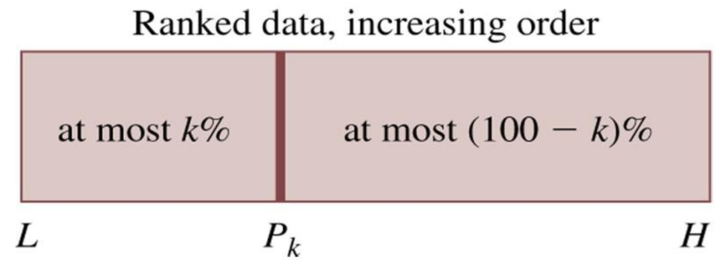
More generally, percentiles. Quartiles are special percentiles.

Percentile: Values of the variable that divide ranked data into 100 equal subsets.

L = lowest value
H = highest value
P_k = value where k% are smaller

You've taken standardized exams and received a %ile.





Figures from Johnson & Kuby, 2012.

2: Descriptive Analysis and Single Variable Data

2.5 Measures of Position

The Percentile Process: Four basic steps for k^{th} percentile.

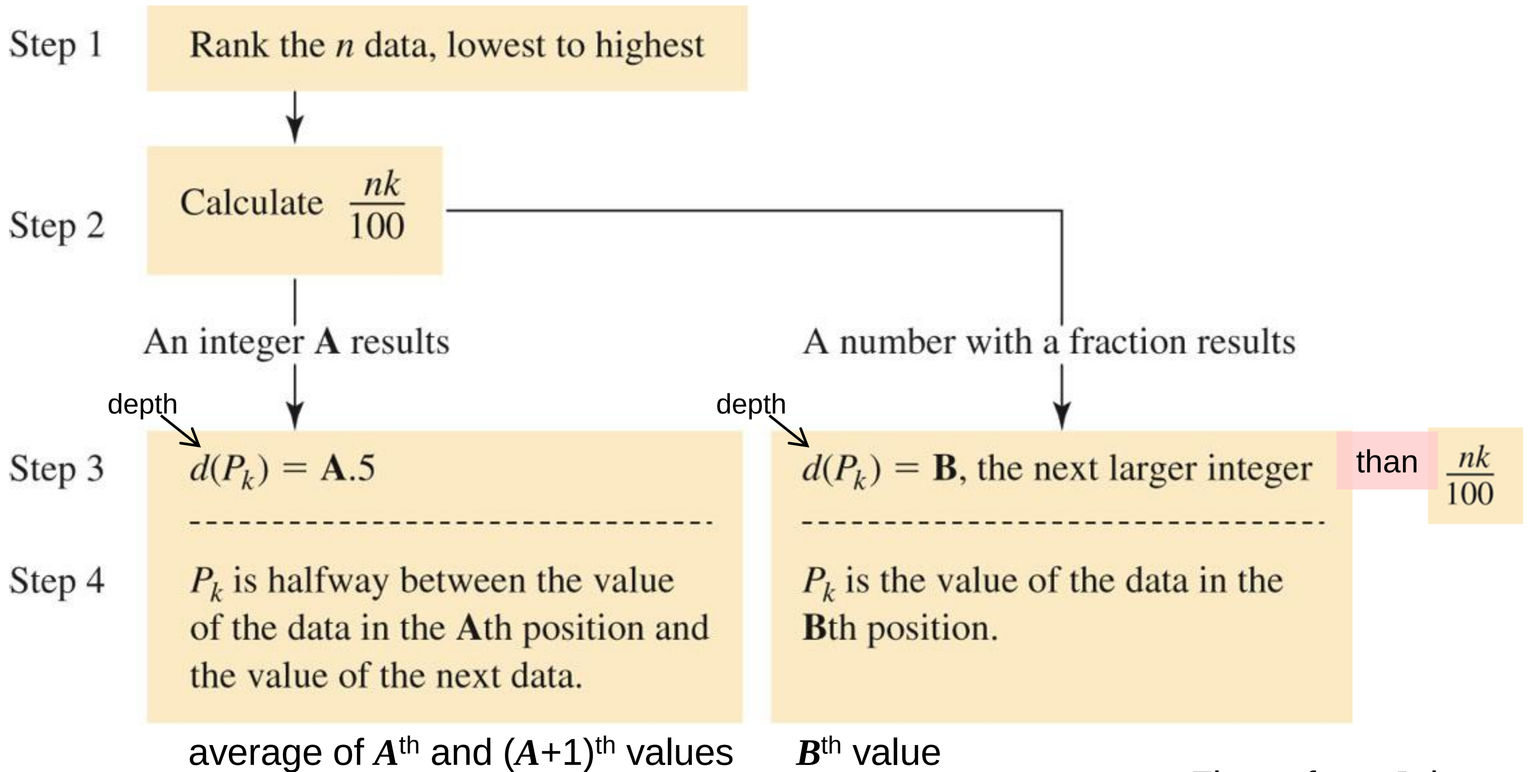


Figure from Johnson & Kuby, 2012.

2: Descriptive Analysis and Single Variable Data

2.5 Measures of Position

Example: 1,2,3,4,5.

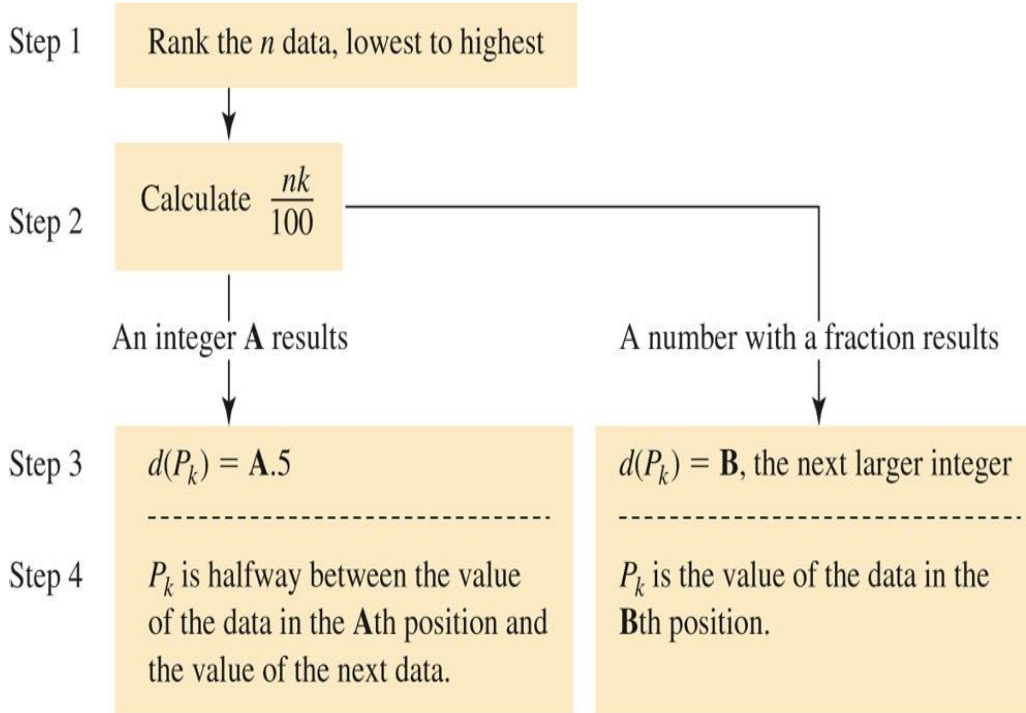
L = low value

H = high value

Q_2 = median

Q_1 = value where 25% smaller

Q_3 = value where 75% smaller



L = ?

H = ?

Q_2 = ?

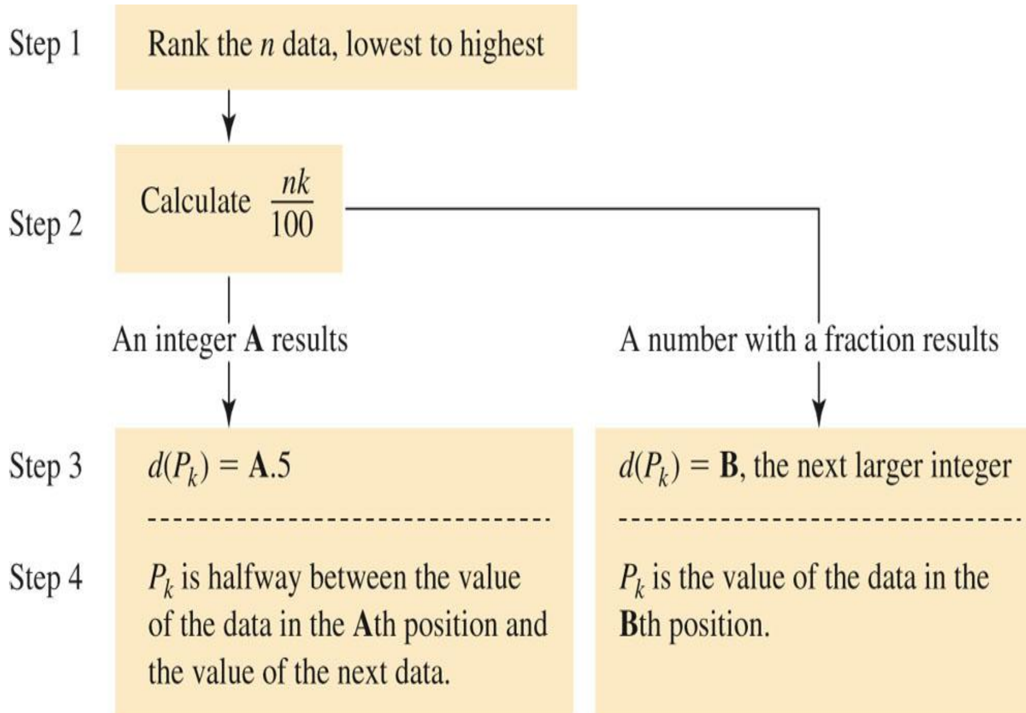
$$Q_1\text{: } \frac{nk}{100} = \qquad \rightarrow ? \quad \text{value} \qquad Q_3\text{: } \frac{nk}{100} = \qquad \rightarrow ? \quad \text{value}$$

Q_1 =?, Q_3 =?

Figure from Johnson & Kuby, 2012.

2: Descriptive Analysis and Single Variable Data

2.5 Measures of Position



Example: 1,2,3,4,5.

L = low value

H = high value

Q_2 = median

Q_1 = value where 25% smaller

Q_3 = value where 75% smaller

L = ? $\rightarrow$ 1

H = ? $\rightarrow$ 5

Q_2 = ? $\rightarrow$ 3

$$Q_1: \frac{nk}{100} = \frac{5(25)}{100} = 1.25 \rightarrow 2^{nd} \text{ value} \qquad Q_3: \frac{nk}{100} = \frac{5(75)}{100} = 3.75 \rightarrow 4^{th} \text{ value}$$

Q_1 =2, Q_3 =4

Figure from Johnson & Kuby, 2012.

2: Descriptive Analysis and Single Variable Data

2.5 Measures of Position

Box-and-whiskers display: A graphic representation of the 5-number summary. L, Q_1, Q_2, Q_3, H.

Example: 1,2,3,4,5.

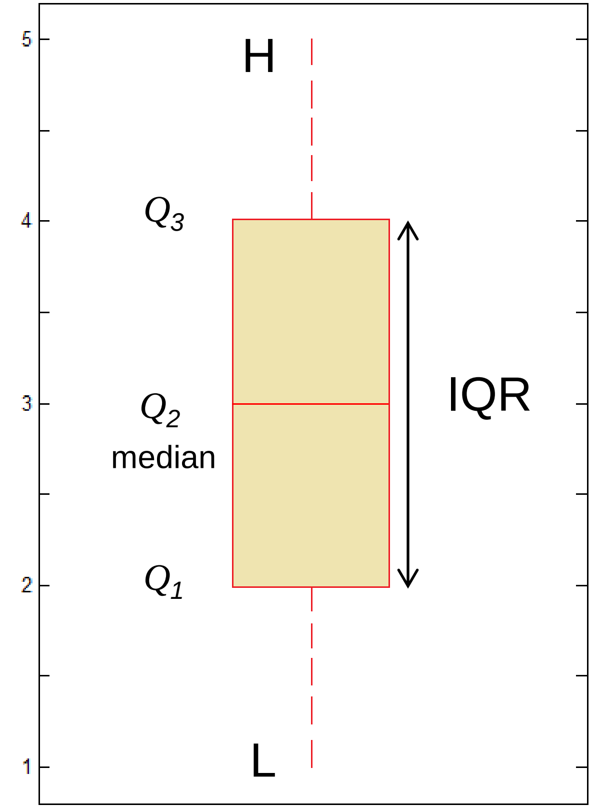
$L = 1$

$Q_1 = 2$

$Q_2 = 3$

$Q_3 = 4$

$H = 5$



2: Descriptive Analysis and Single Variable Data

2.5 Measures of Position

Standard score, or z-score: The position a particular value of x has relative to the mean, measured in standard deviations.

$$z_i = \frac{i^{\text{th}} \text{ value - mean}}{\text{std. dev.}} = \frac{x_i - \bar{x}}{s} \qquad (2.11)$$

There can be n of these because we have $x_1, x_2, \ldots, x_n$.

2: Descriptive Analysis and Single Variable Data

2.5 Measures of Position

Standard score, or z-score:

Example: 1, 2, 3, 4, 5

$$z_i = \frac{x_i - \overline{x}}{s}$$

$$\overline{x} = 3$$

$$s = 1.58$$

$$z_1 = \frac{x_1 - \overline{x}}{s} = ?$$

2: Descriptive Analysis and Single Variable Data

2.5 Measures of Position

Standard score, or z-score:

Example: 1, 2, 3, 4, 5

$$z_i = \frac{x_i - \bar{x}}{s}$$

$$\bar{x} = 3$$

$$s = 1.58$$

$$z_1 = \frac{x_1 - \bar{x}}{s} = \frac{1 - 3.00}{1.58} = \text{-}1.3$$

$$z_1 = -1.2649$$

$$z_2 = -0.6325$$

$$z_3 = 0$$

$$z_4 = 0.6325$$

$$z_5 = 1.2649$$

2: Descriptive Analysis and Single Variable Data

2.5 Measures of Position

$n = 100$

mean	stdev
67.4	3.6

Example: What is your z score?

Height

x

Deviation

$$dev = x - \bar{x}$$

z-score

$$z = \frac{x - \bar{x}}{s}$$

2: Descriptive Analysis and Single Variable Data

2.5 Measures of Position

$n = 100$

$$\frac{nk}{100}$$

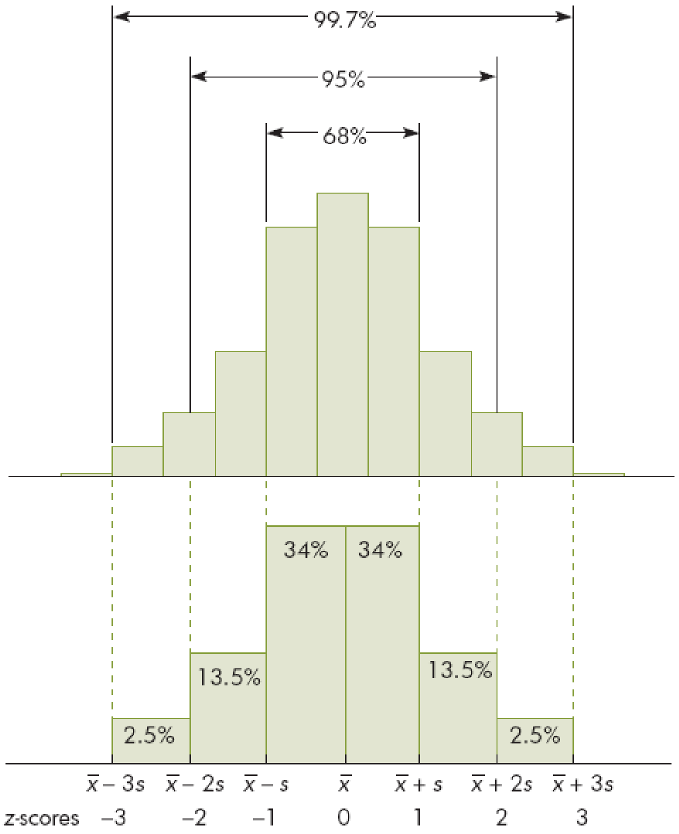
Example: Compute the 5 number summary.

2: Descriptive Analysis and Single Variable Data

2.6 Interpreting and Understanding Standard Deviation

Read this section (and 2.7) on your own.

Bell Curve
Normal Distribution
Gaussian Distribution



2: Descriptive Analysis and Single Variable Data

Questions?

Homework: Read Chapter 2.5-2.7
WebAssign
Chapter 2 # 115, 123c-d, 129, 137

Chapter 3: Descriptive Analysis and Presentation of Bivariate Data

Daniel B. Rowe, Ph.D.

Department of Mathematical and Statistical Sciences

3: Descriptive Analysis and Bivariate Data

3.1 Bivariate Data

Bivariate data: The values of two different variables that are obtained from the same population element.

Qualitative-Qualitative
Qualitative-Quantitative
Quantitative-Quantitative

When Qualitative-Qualitative
Cross-tabulation tables or **contingency tables**
Sometimes called r by c ($r \times c$)

3: Descriptive Analysis and Bivariate Data

3.1 Bivariate Data: two qualitative

Example:

M = male
F = female

LA = liberal arts
BA = business admin
T = technology

Name	Gender	Major	Name	Gender	Major	Name	Gender	Major
Adams	M	LA	Feeney	M	T	McGowan	M	BA
Argento	F	BA	Flanigan	M	LA	Mowers	F	BA
Baker	M	LA	Hodge	F	LA	Ornt	M	T
Bennett	F	LA	Holmes	M	T	Palmer	F	LA
Brand	M	T	Jopson	F	T	Pullen	M	T
Brock	M	BA	Kee	M	BA	Rattan	M	BA
Chun	F	LA	Kleeberg	M	LA	Sherman	F	LA
Crain	M	T	Light	M	BA	Small	F	T
Cross	F	BA	Linton	F	LA	Tate	M	BA
Ellis	F	BA	Lopez	M	T	Yamamoto	M	LA

Figure from Johnson & Kuby, 2012.

3: Descriptive Analysis and Bivariate Data

3.1 Bivariate Data: two qualitative

Example:

Construct a 2×3 table.

Name	Gender	Major	Name	Gender	Major	Name	Gender	Major
Adams	M	LA	Feeney	M	T	McGowan	M	BA
Argento	F	BA	Flanigan	M	LA	Mowers	F	BA
Baker	M	LA	Hodge	F	LA	Ornt	M	T
Bennett	F	LA	Holmes	M	T	Palmer	F	LA
Brand	M	T	Jopson	F	T	Pullen	M	T
Brock	M	BA	Kee	M	BA	Rattan	M	BA
Chun	F	LA	Kleeberg	M	LA	Sherman	F	LA
Crain	M	T	Light	M	BA	Small	F	T
Cross	F	BA	Linton	F	LA	Tate	M	BA
Ellis	F	BA	Lopez	M	T	Yamamoto	M	LA

	Major		
Gender	LA	BA	T
M	||||| (5)	||||| | (6)	||||| || (7)
F	||||| | (6)	|||| (4)	|| (2)

M = male
F = female
LA = liberal arts
BA = business admin
T = technology

Figures from Johnson & Kuby, 2012.

3: Descriptive Analysis and Bivariate Data

3.1 Bivariate Data: two qualitative

Example:
Percentages based on grand total (next slide).

Gender	LA		BA		T	
	Major					
M	|||||	(5)	|||||		(6)	||||| ||	(7)
F	|||||		(6)	||||	(4)	||	(2)

Gender	LA	BA	T	Row Total
	Major			
M	5	6	7	18
F	6	4	2	12
Col. Total	11	10	9	30

M = male
F = female
LA = liberal arts
BA = business admin
T = technology

Figures from Johnson & Kuby, 2012.

3: Descriptive Analysis and Bivariate Data
3.1 Bivariate Data: two qualitative

Example:
Percentages based on grand total.

	Major			
Gender	LA	BA	T	Row Total
M	5	6	7	18
F	6	4	2	12
Col. Total	11	10	9	30

	Major			
Gender	LA	BA	T	Row Total
M	17%	20%	23%	60%
F	20%	13%	7%	40%
Col. Total	37%	33%	30%	100%

7/30*100%=23%

M = male
F = female
LA = liberal arts
BA = business admin
T = technology

Divide all numbers by grand total.

Figures from Johnson & Kuby, 2012.

3: Descriptive Analysis and Bivariate Data

3.1 Bivariate Data: two qualitative

Example:
Percentages based on row totals.

	Major			
Gender	LA	BA	T	Row Total
M	5	6	7	18
F	6	4	2	12
Col. Total	11	10	9	30

	Major			
Gender	LA	BA	T	Row Total
M	28%	33%	39%	100%
F	50%	33%	17%	100%
Col. Total	37%	33%	30%	100%

7/18*100%=39%

M = male
F = female
LA = liberal arts
BA = business admin
T = technology

Divide all row numbers by row total.

Figures from Johnson & Kuby, 2012.

3: Descriptive Analysis and Bivariate Data

3.1 Bivariate Data: two qualitative

Example:
Percentages based on column totals.

	Major			
Gender	LA	BA	T	Row Total
M	5	6	7	18
F	6	4	2	12
Col. Total	11	10	9	30

	Major			
Gender	LA	BA	T	Row Total
M	45%	60%	78%	60%
F	55%	40%	22%	40%
Col. Total	100%	100%	100%	100%

7/9*100%=78%

M = male
F = female
LA = liberal arts
BA = business admin
T = technology

Divide all row numbers by column total.

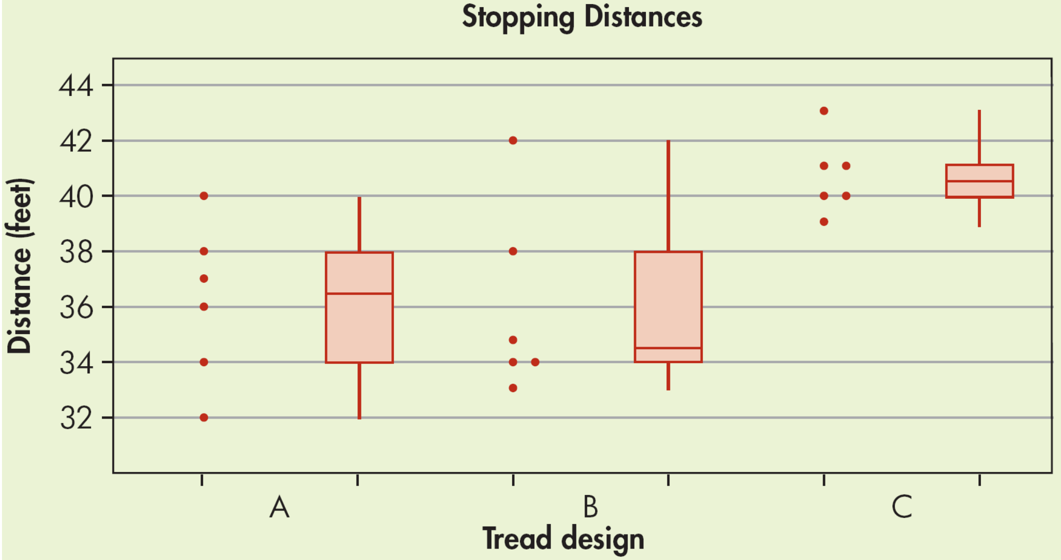
Figures from Johnson & Kuby, 2012.

3: Descriptive Analysis and Bivariate Data

3.1 Bivariate Data: one qualitative and one quantitative

Example: Stopping Distances (in feet) for three treads.

Design A ($n = 6$)			Design B ($n = 6$)			Design C ($n = 6$)		
37	36	38	33	35	38	40	39	40
34	40	32	34	42	34	41	41	43



Vertical box-and-whiskers

Figures from Johnson & Kuby, 2012.

3: Descriptive Analysis and Bivariate Data

3.1 Bivariate Data: one qualitative and one quantitative

Example:

Design A (n = 6)			Design B (n = 6)			Design C (n = 6)		
37	36	38	33	35	38	40	39	40
34	40	32	34	42	34	41	41	43

	Design A	Design B	Design C
High	40	42	43
Q_3	38	38	41
Median	36.5	34.5	40.5
Q_1	34	34	40
Low	32	33	39

	Design A	Design B	Design C
Mean	36.2	36.0	40.7
Standard deviation	2.9	3.4	1.4

Figures from Johnson & Kuby, 2012.

3: Descriptive Analysis and Bivariate Data

3.1 Bivariate Data: two quantitative

When have paired quantitative data, represent as (x,y) **ordered pairs**.

Input variable called independent variable, x.
Output variable called dependent variable, y.

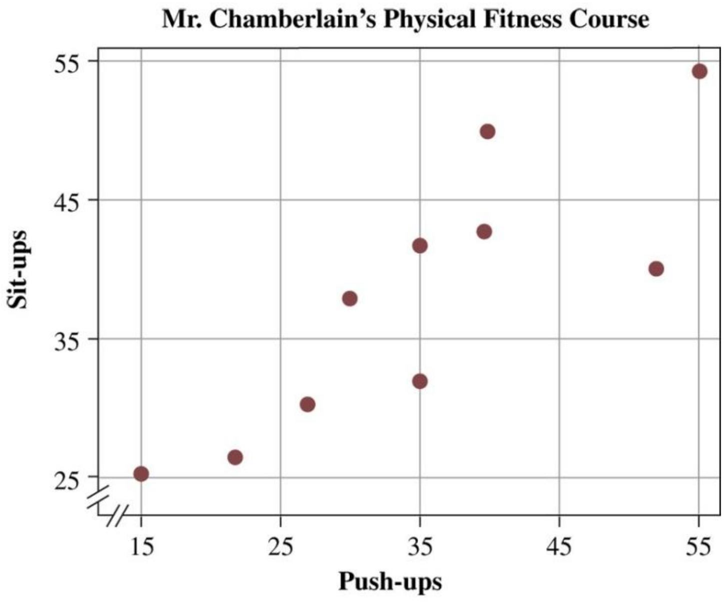
Scatter Diagram: A plot of all the ordered pairs of bivariate data on a coordinate axis system. The input variable, x, is plotted on the horizontal axis and the output variable, y, is plotted on the vertical axis.

3: Descriptive Analysis and Bivariate Data

3.1 Bivariate Data: two quantitative, Scatter Diagram

Example: Push-ups

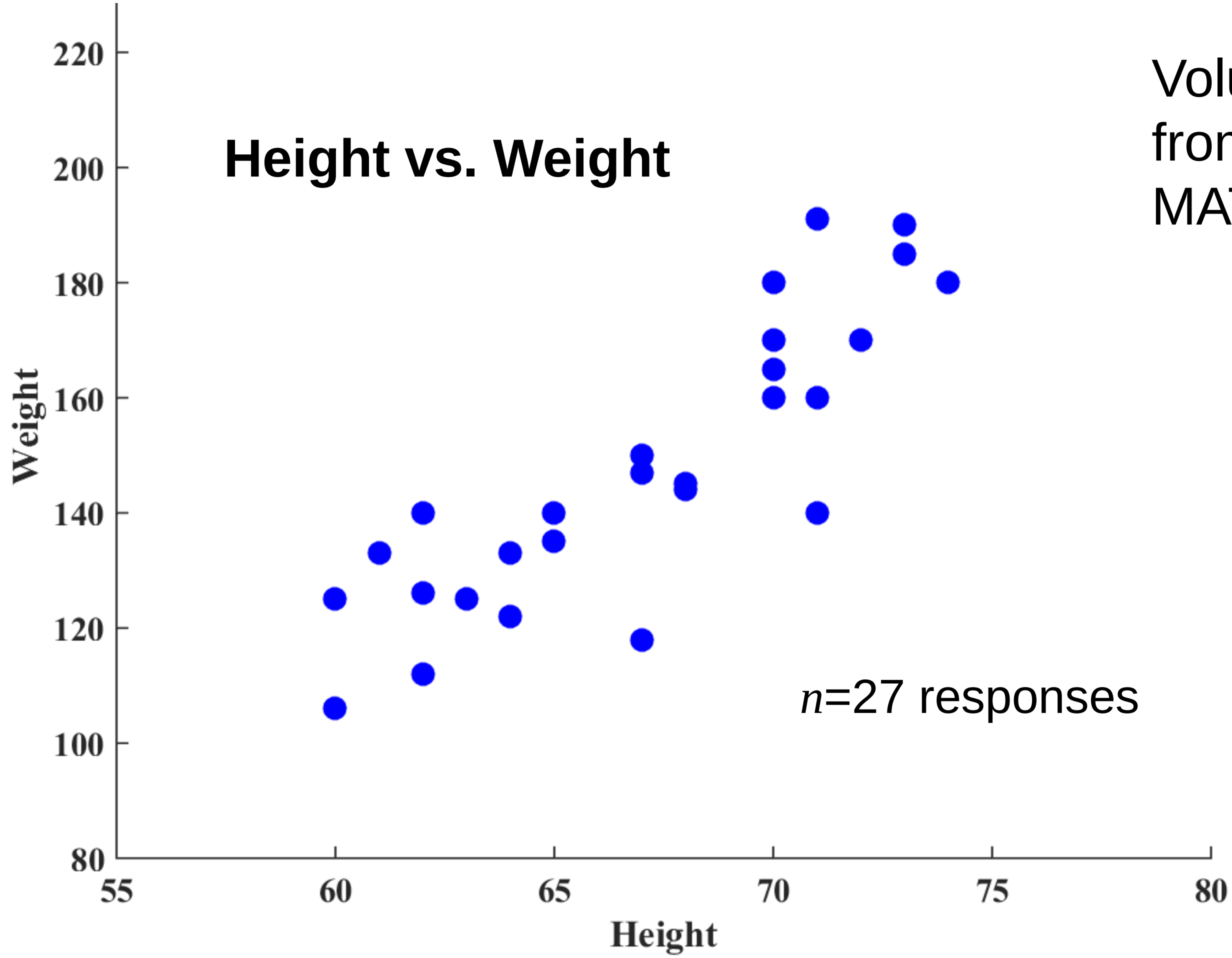
Student	1	2	3	4	5	6	7	8	9	10
Push-ups, x	27	22	15	35	30	52	35	55	40	40
Sit-ups, y	30	26	25	42	38	40	32	54	50	43



Figures from Johnson & Kuby, 2012.

3: Descriptive Analysis and Bivariate Data

3.1 Bivariate Data: Scatter Diagram Our data.



Voluntarily provided data from students in a previous MATH 1700 class.

3: Descriptive Analysis and Bivariate Data

Questions?

Homework: Read Chapter 3
WebAssign
Chapter 3 # 3, 7, 15