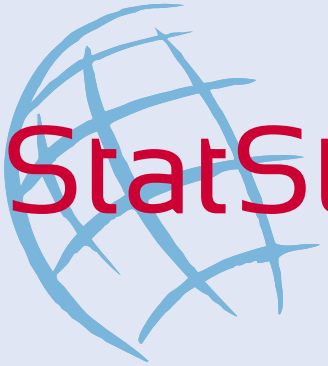
Summary Statistics

Jeffrey Michael Franc
MD, MSc, FCFP.EM, Dip Sport Med, EMDM

Medical Director, E/D Management
Alberta Health Services

Assistant Clinical Professor of Emergency Medicine
University of Alberta

Visiting Professor in Disaster Medicine
Universita' Degli Studi del Piemonte Orientale

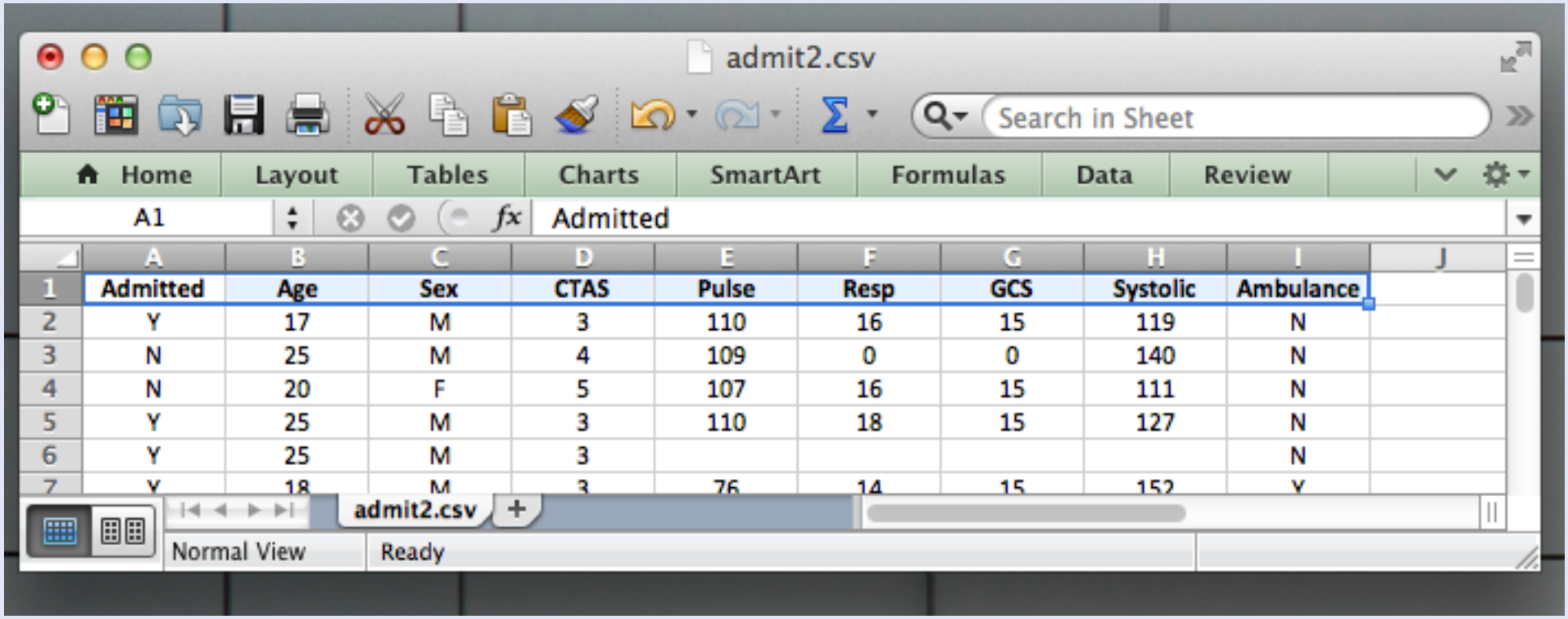


MedStatStudio

Objectives

- Define response Variable, Factor and Level and describe the difference between discrete and continuous variables
- Choose a test statistic for various univariate measures

Quiz



	A	B	C	D	E	F	G	H	I	J
1	Admitted	Age	Sex	CTAS	Pulse	Resp	GCS	Systolic	Ambulance	
2	Y	17	M	3	110	16	15	119	N	
3	N	25	M	4	109	0	0	140	N	
4	N	20	F	5	107	16	15	111	N	
5	Y	25	M	3	110	18	15	127	N	
6	Y	25	M	3					N	
7	Y	18	M	3	76	14	15	152	Y	

What is the response variable? _____

Are the observations univariate or multivariate? _____

How many Factors are present? _____

Name one discrete factor _____

Name one continuous factor _____

Which discrete factor has more than 2 levels? _____

Variables

What is the difference between discrete and continuous variables?

Discrete Values

Only certain values are allowed. May be…

- Binary
- Categorical
- Ordinal
- Counting

Discrete Values

Only certain values are allowed. May be…

- Binary (Yes/No or True/False)
- Categorical (Chemical / Biological / Radiological / Explosive)
- Ordinal (Triage: T1/T2/T3/T4)
- Counting (Number of Ambulances sent to a scene)

Continuous Variables

Have an infinite number of possibilities.

All possible values on some part of the number line

Continuous Variables

Have an infinite number of possibilities.

All possible values on some part of the number line

- Length of time from ambulance arrival to first evacuation
- Weight of the patient

Response Variable

What is a response variable?

What are some synonyms?

Response Variable

Usually this is related to the hypothesis of the study.

Also called:

- Dependent variable
- Outcome variable
- Experimental variable

Response Variables

Examples:

1. Number of ice cream scoops sold
2. Height of a corn plant
3. Pain score after surgery
4. Score on a standardized knowledge test

Factors

What is a factor?

What are some synonyms?

Factors

These are the variables that we suspect will influence the response

Also known as…

- Independent variables
- Explanatory variables
- Manipulator variables
- Risk factors

Factors

Example:

1. An antibiotic give to treat pneumonia
2. A teaching session in disaster medicine
3. Which type of triage system is used
4. Age of a patient being treated for cellulitis

Levels of Factors

Continuous Factors Have a value

Discrete factors will have 1 or more levels in the experiment

Factor Levels

Examples:

1. Antibiotic treatment levels:
   - Placebo
   - Oral Fluoroquinilone
   - Intravenous Cephalosporin
2. Triage System Used
   - CTAS
   - START
   - SALT

Variables

What is the difference between univariate and multivariate data?

Variables

What is the difference between univariate and multivariate data?

- Univariate: Observations on a single variable
- Multivariate: Observations made on more than one variable. Usually we are interested in the relationship between the two.

Variables

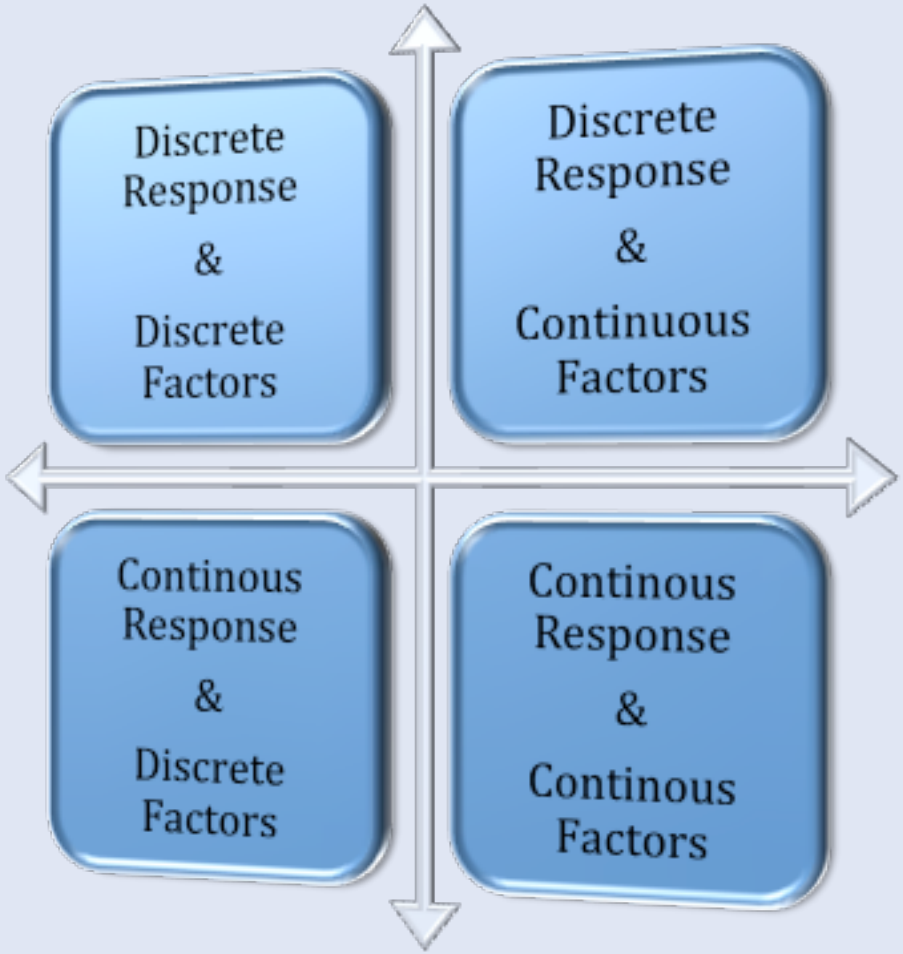
Why is the distinction of these types of variables important?

Variables

Why is the distinction of these types of variables important?

- Choice of statistic test follows directly from these variables

Choice of Statistical Tests



Variables

Questions?

Summary Statistics

We have collected some data: we wish to summarize the univariate data

What statistics are commonly used to summarize the data?

Summary Statistics

What statistics are commonly used to summarize the data?

1. Measures of Central Tendency
   - --
   - --
   - --
2. Measures of Variability
   - --
   - --

Summary Statistics

What statistics are commonly used to summarize the data?

1. Measures of Central Tendency
   - Mean
   - Median
2. Measures of Variability
   - Range
   - Standard Deviation

Mean

What is the mean?

Mean

What is the mean?

Sum of the values divided by the number of values.

$$\bar{X} = \frac{\sum x}{n}$$

Mean

What are the advantages and disadvantages of using the mean?

Mean

Advantages

- It is the Minimum Variance Best Estimator (MVBE)
- Easy to calculate

Disadvantages

- Easily influenced by outliers
- Need to have all x values to calculate

Median

What is the median?

Median

The value that separates the highest half and lowest half of the sample.

To Calculate:

- Rank observations from low to high
- Middle value if odd number of observations
- Average of two middle value if even number of observations

Median

Advantages

- Less sensitive to outliers
- Can be used when end values are not known
- Works well for skewed populations
- Gives reduced importance to outliers

Disadvantages

- By ignoring outliers, important data points may not be summarized

Range

The difference between highest and lowest values.

Often presented as the highest and lowest value:

range(14.5 to 19.8)

Standard Deviation

What is the standard deviation?

What is the variance?

Standard Deviation

Variance is the average distance from the mean, that is then squared and summed up.

Standard deviation is the square root of the mean

$$\sigma = \sqrt{\frac{\sum (x - \mu)^2}{n}}$$

Standard Deviation

Note that Standard Deviation is NOT the same as Standard Error.

(Standard error is the standard deviation of an estimator)

Other Point Estimators

Central Tendency

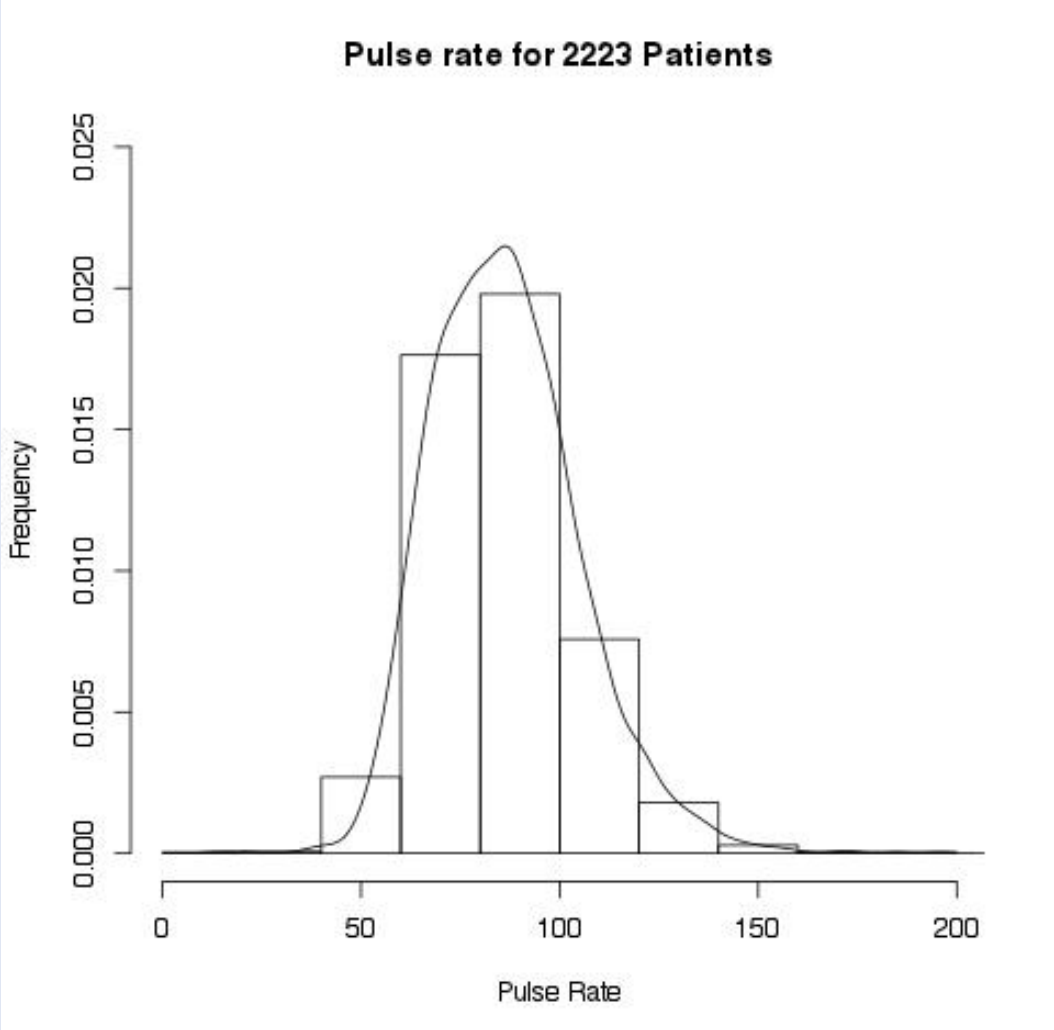
- Mode
- Trimmed Mean

Variation

- Intra-quartile Range

PLOTS!!!!

Graphical



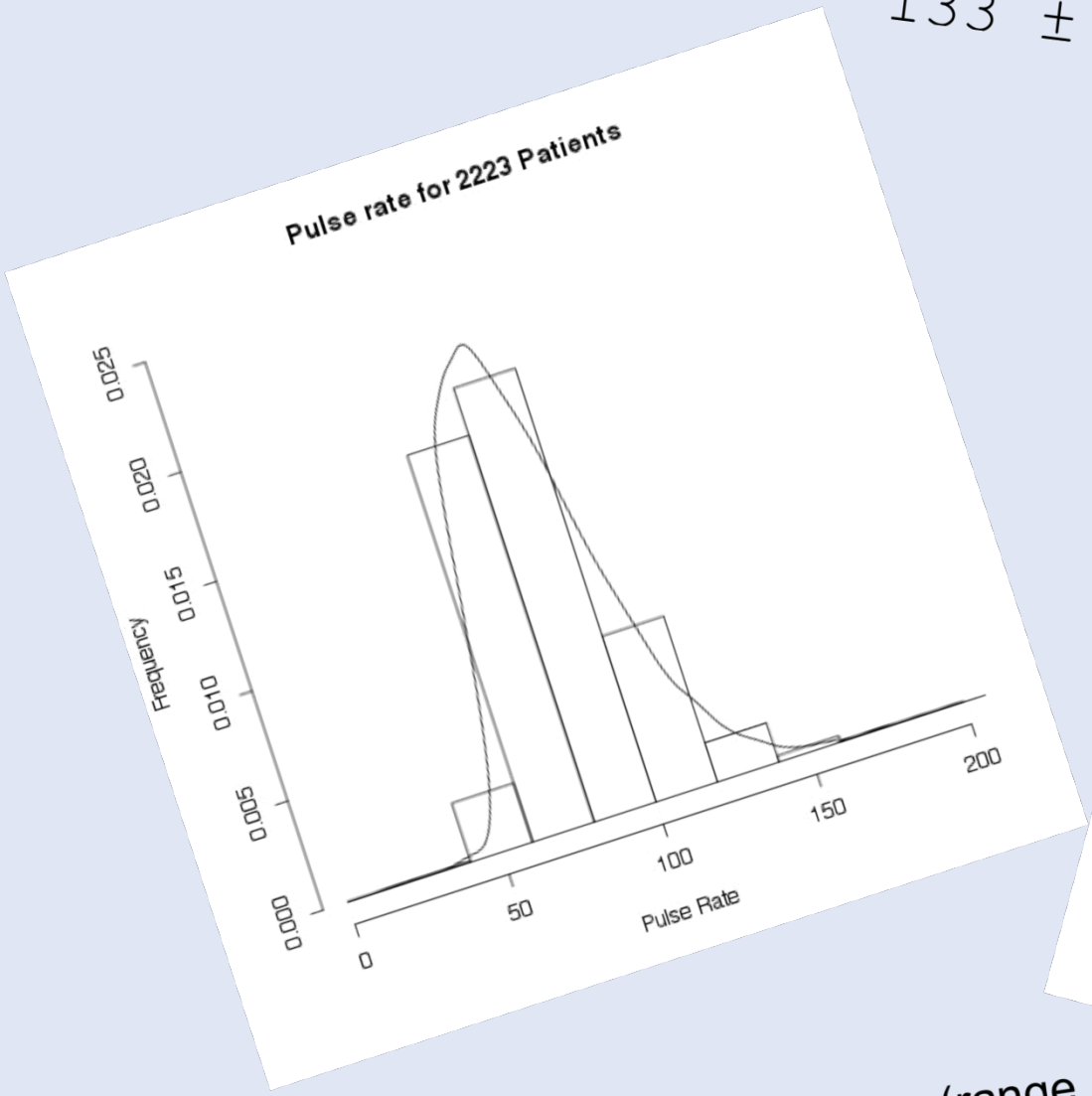
Summary Statistics

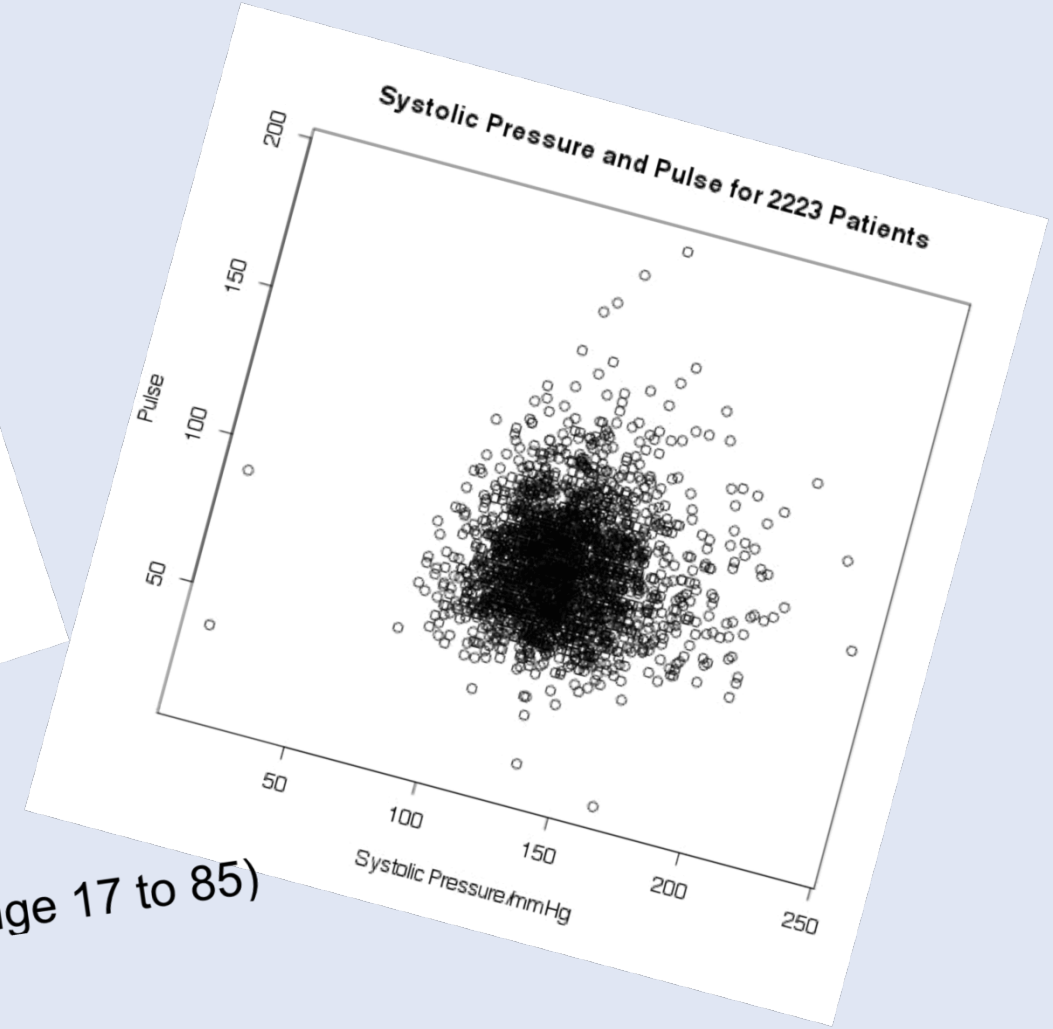
The mean pulse rate for 2223 patients was 86 with a standard deviation of 19. Minimum pulse was 13, and maximum was 196.

Exploratory Data Analysis

133 ± 20

We found overall accuracy to be 87% for CTAS triage.





Mean age was 47 years (range 17 to 85)

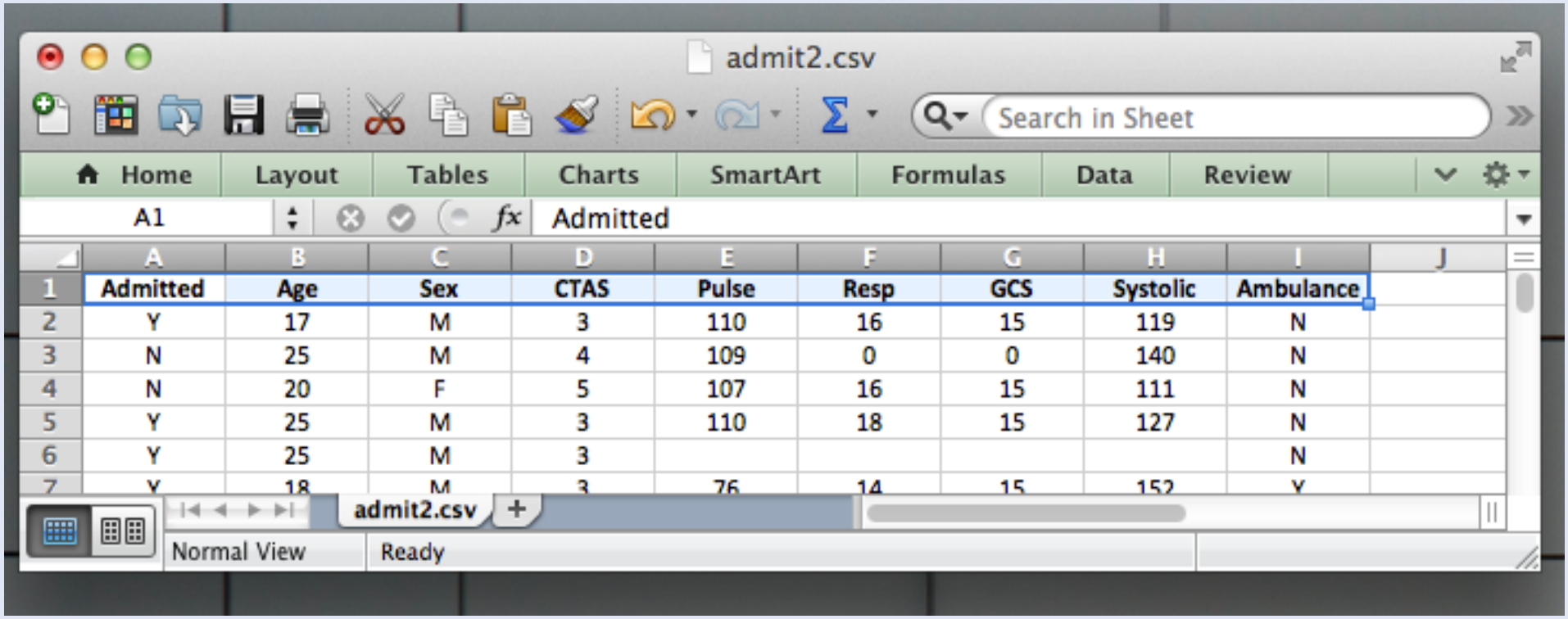
Summary Statistics

Questions?

Objectives

- Define response Variable, Factor and Level and describe the difference between discrete and continuous variables
- Choose a test statistic for various univariate measures

Quiz



	A	B	C	D	E	F	G	H	I	J
1	Admitted	Age	Sex	CTAS	Pulse	Resp	GCS	Systolic	Ambulance	
2	Y	17	M	3	110	16	15	119	N	
3	N	25	M	4	109	0	0	140	N	
4	N	20	F	5	107	16	15	111	N	
5	Y	25	M	3	110	18	15	127	N	
6	Y	25	M	3					N	
7	Y	18	M	3	76	14	15	152	Y	

What is the response variable? _____

Are the observations univariate or multivariate? _____

How many Factors are present? _____

Name one discrete factor _____

Name one continuous factor _____

Which discrete factor has more than 2 levels? _____

Math Lesson

What is the difference between standard deviation, sample standard deviation, and standard error?