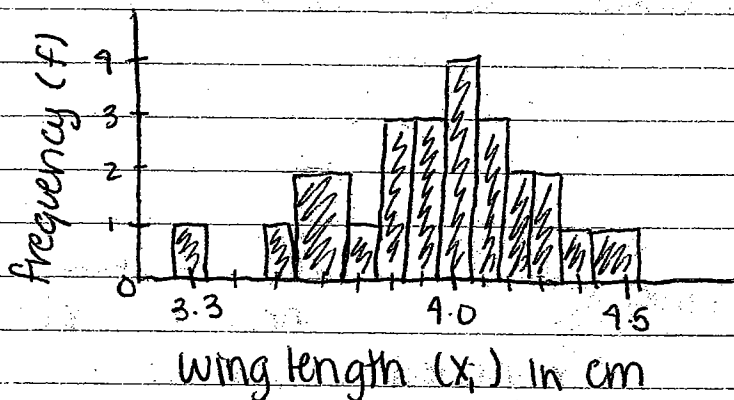


Histograms

10/2/08

- Histogram - a bar plot with locations along the horizontal axis that correspond to values of the variable, with bars over those locations indicating frequency

Ex. Monarch Butterfly Wing length (cont'd)



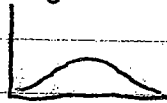
- quantitative
- continuous
- ratio

- used for quantitative variables only
 - not dichotomous
 - either continuous or discrete
 - values must have a unique place on the number line
- secret to good statistical work: visualize the data as a table with rows for subjects and columns for variables
- Grouping
 - by regrouping the data, more or less bars can be created on the histogram (butterfly wing lengths could be regrouped so there was a bar for 3.0-3.9, 3.5-3.9, 4.0-4.9, 4.5-4.9)
 - too few bars - not enough detail (information) to make accurate statements about the shape →

- too many bars - shape cannot be seen clearly (too much noise)

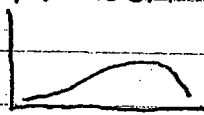
- Shape

- Symmetric - histogram looks the same to the right + left of center.



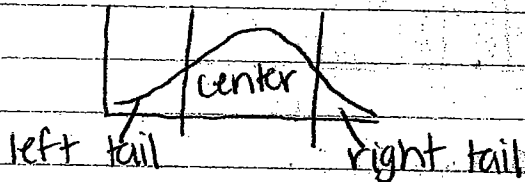
- mode generally at center.

- Asymmetric (skew Skew) - mode is not at the center



- long left tail = negative skew

- long right tail = positive skew



- Scaling

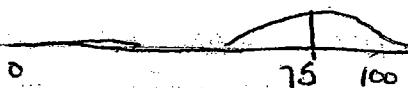
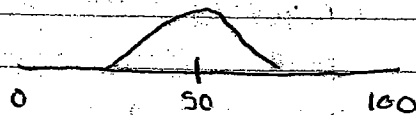
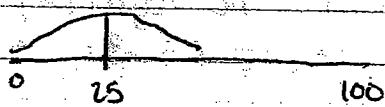
- Raw frequency scale - like histogram in butterfly example

- Relative frequency - divide all frequencies by n to obtain the relative frequencies, or proportions

- Density scale - drawn so the total area under the curve is equal to one

- area under curve = relative frequency

Numerical Descriptive Methods



different center

different spread

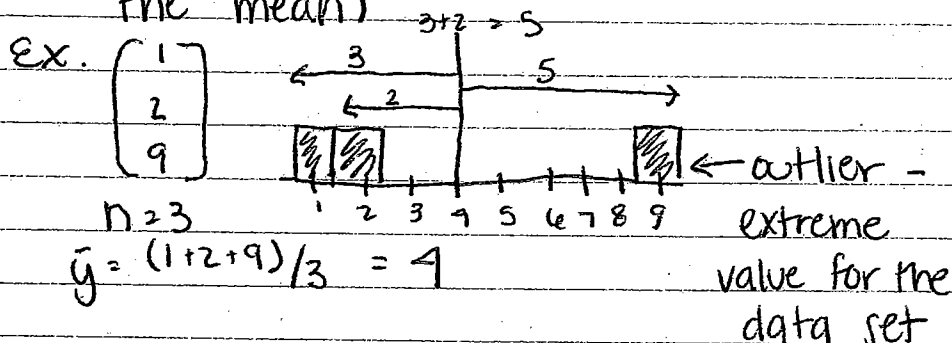


- Measures of center

- Mean - numerical average of data
- add all observations and divide

$$\text{by } n \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

- graphically, the mean is where the data would balance (if it was placed on a fulcrum at the mean)



- Median - the physical middle of a data set (i.e. the 3rd term when $n=5$)

- if n is an even number, the median is the average of the two center most numbers.

- on a histogram, median is where 50% of the data is to the left of the point

- median = 50th percentile (50% to the left)

- $\frac{1}{4}$ of data to left = 1st quartile

- $\frac{3}{4}$ of data to left = 3rd quartile

- Mode - term with the highest frequency

- graphically, the highest point(s) on the histogram

- can be unimodal, bimodal, or multimodal