AMS 7
Class 3
6|24|16
Lecture 3

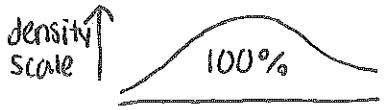
Raw Frequency (↑ how many) vs Relative Frequency (%) (↑ (how many ÷ n) · 100%)

L-16

Raw/Rel → histograms will be identical!

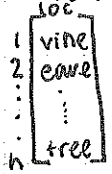
Density Scale: relative freq = area under histogram

Convention: all hist. sketches will be on density scale in this class



L-18 ex. 1 row for each nest (subject)

1.1 1 col for each (variable) location



Variable: qual, nominal, not dich
hist? no bar? yes

n = 56 + 60 + 46 + 49

1.2 1 row for each fish
1 col for each pigmentation class
n = sum fish

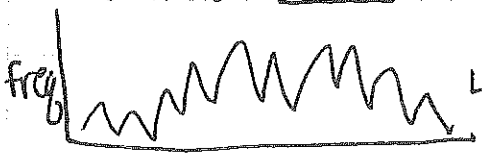
Variable: qual, ordinal, not dic
hist? no bar? yes

1.3 1 row for each litter
1 col for each size
n = sum of freq.

variable: quant, discrete, ratio
hist? yes

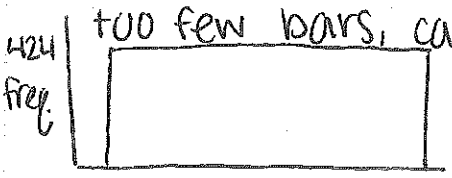
1.4a 1 row for each plant
1 col for each # of aphids
n = 424 Variable: quant, discrete, ratio

too much noise to get clear picture of shape

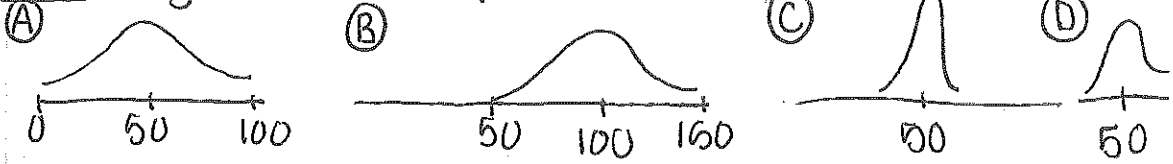


42 bars is too many

too few bars, can't see shape

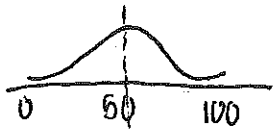


Histogram Shape

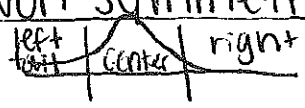


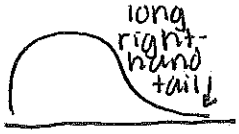
A and B different center, same spread + shape
A and C same center, same shape, different sprea
A and D same center + spread, different shape

Symmetric

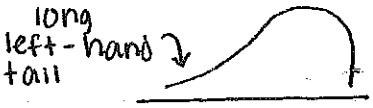


Non-Symmetric



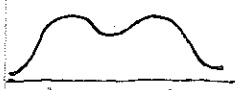


right skewed
positive skewed



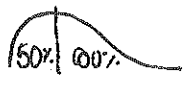
left skewed
negative skewed

Modes



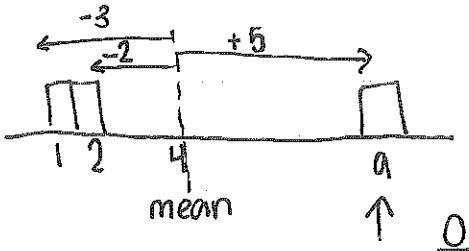
Mode = point of highest frequency

if symmetric - mode = median = pt of symmetry
↑ IF UNIMODAL

Median: 

$\tilde{y}$ = 50th percentile, 2$\frac{1}{2}$th quantile

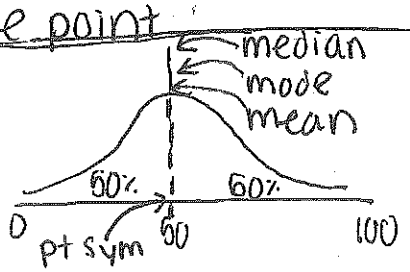
$$\begin{matrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{matrix}\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix}$$ n = 3



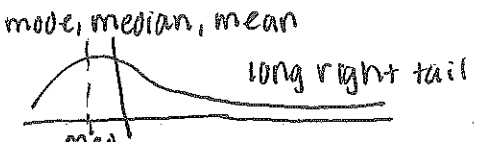
↑ outlier

Mean: balance point

Symmetric Unimodal



Move 1 end



Asymmetric Distribution: ask if mean or median

Spread measured by deviation

Deviations from mean: $y_n - \bar{y}$

$\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix}$ n = 3 sub. mean $\begin{bmatrix} -3 \\ -2 \\ +5 \end{bmatrix}$ abs value $\begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}$ mean = 16/3 $|y_n - \bar{y}|$

$$M.A.D. = \frac{1}{n}\sum_{i=1}^{n} |y_i - \bar{y}|$$
mean absolute deviation

$\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix}$ subt $\bar{y}$ $\begin{bmatrix} -3 \\ -2 \\ +5 \end{bmatrix}$ square $\begin{bmatrix} 9 \\ 4 \\ 25 \end{bmatrix}$ mean = 38/3 $(y_n - \bar{y})^2$

mean of squares $= \frac{1}{n}\sum_{i=1}^{n} (y_i - \bar{y})^2$ ← but... this squares units...

sample variance $= \frac{1}{n-1}\sum_{i=1}^{n} (y_i - \bar{y})^2$

$SD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} (y_i - \bar{y})^2} = s$

Sample Standard Deviation