

Caitlin Daughy

AMST

1/13/09

(1) Lecture #3 Histograms, Mean, SD

next time: normal curve

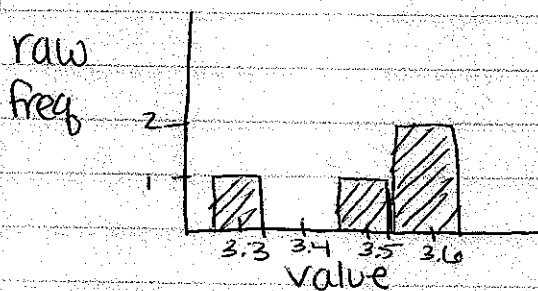
690 pgs: 300 pg book, 300 pg lecture notes, 3 100 pg reader now available in back of room: \$100 production cost

#1 Homework due Jan 20th

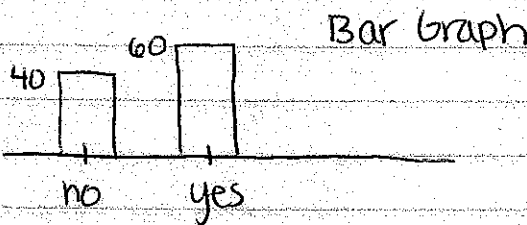
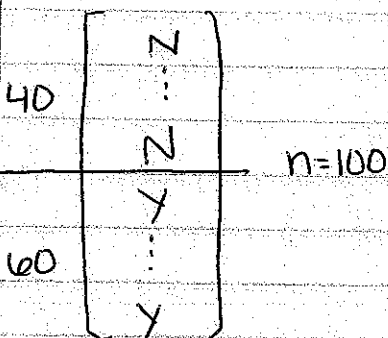
butterfly wing length (cm)

| | | | | | |
|------|---|--------|---|-------|------------------------|
| n=24 | $\begin{bmatrix} 4.4 \\ 3.6 \\ \vdots \\ 3.9 \end{bmatrix}$ | sort → | $\begin{bmatrix} 3.3 \\ 3.5 \\ 3.6 \\ 3.6 \\ \vdots \\ 4.5 \end{bmatrix}$ | Value | raw frequency Count |
| | | | | 3.3 | 1 |
| | | | 3.4 | 0 | |
| | | | 3.5 | 1 | |
| | | | 3.6 | 2 | |
| | | | ⋮ | ⋮ | |
| | | | 4.5 | 1 | |
| | | | | n=24 | |

Raw Freq. Histogram



JMP does this backwards as its default



* One secret to good statistical work: Visualize the raw data.

Min Daughy

AMST
1/13/09

(2) Lecture #3

1 row for
each nesting site
 $n=251$

location
[B
V
T
⋮
C]

1 row for each
sunfish
 $n=154$

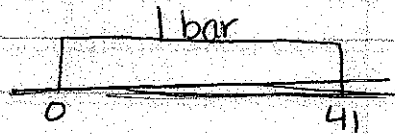
pigmentation
[faintly speckled
solid black
⋮]

1 row for
each clover
plant

aphids
[17
36
12
⋮
3]

histograms for aphid data

too few
bars



no shape
info

too many
bars



42 bars =
difficult

to see shape

wing length

[4.4
3.6
⋮
3.9]

y
[y₁
y₂
⋮
y_n] = y₂₄

→ "y sub 1" or "y 1"

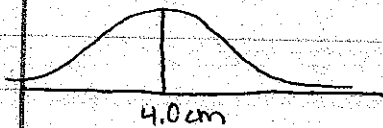
mean $\frac{95}{24} = 3.96$ cm

mean $\bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n}$

$= \frac{1}{n} (y_1 + \dots + y_n)$

$= \frac{1}{n} \sum_{i=1}^n y_i$

Σ = capital sigma
i: index of
summation

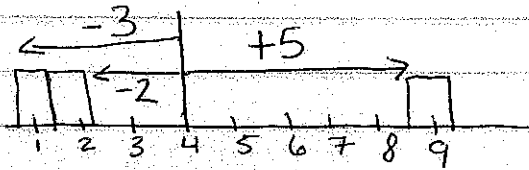


pt of highest
frequency = mode

(3) Lecture #3

$n=3$ $\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix}$

mean $\bar{y} = 4$



* mean = balance point of data set

another useful measure of center: median

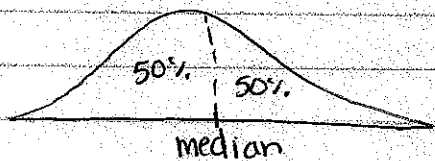
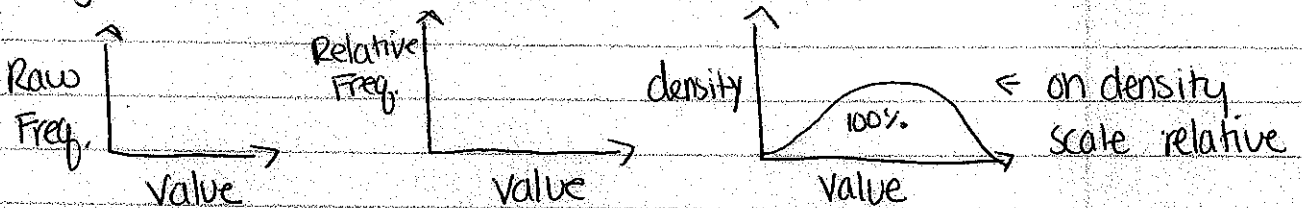
middle of data, sorted from smallest to largest

$\begin{bmatrix} 1 \\ \textcircled{2} \\ 9 \end{bmatrix}$ median

$\begin{bmatrix} 1 \\ \textcircled{2} \\ \textcircled{3} \\ 9 \end{bmatrix}$ median $2.5 = \tilde{y}$

\tilde{y} : "y tilda"

graphical interpretation of median:

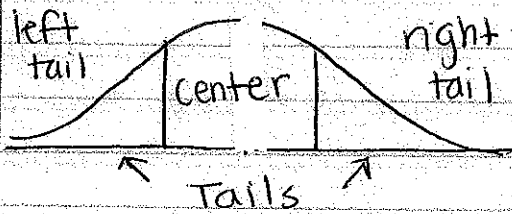


default: sketch hist. are always on density scale

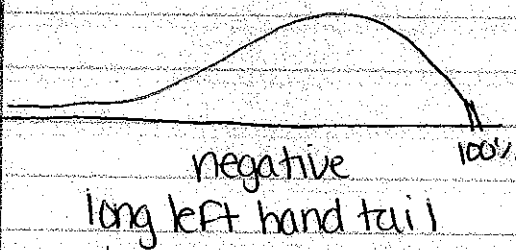
density scale: relative freq. info = area under the curve that approximates histogram.

(4) Lecture #3

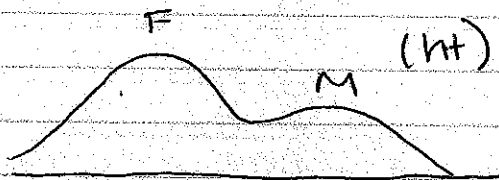
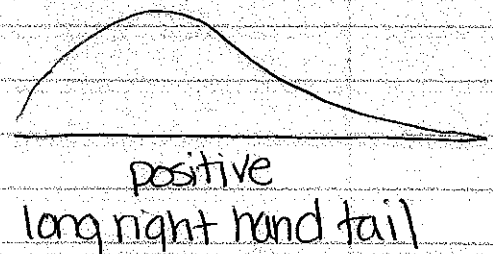
AMST
1/13/09



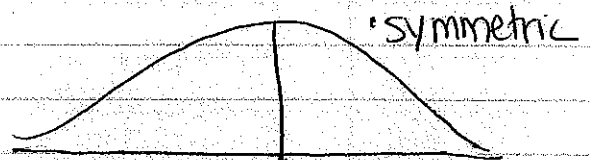
skewed, unimodal



skewed, unimodal



bimodal dist.
(multimodal)
modals: 2



unimodal
Folding Point
- point of symmetry
= mean, mode,
median