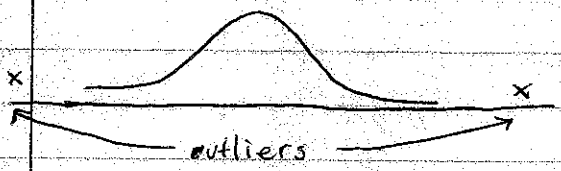
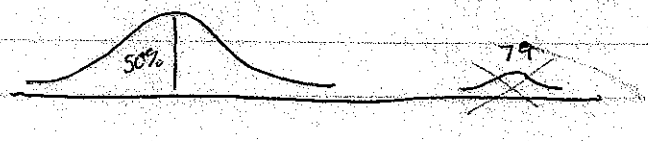
Amy Roget

- Outliers: Sensitivity Analysis



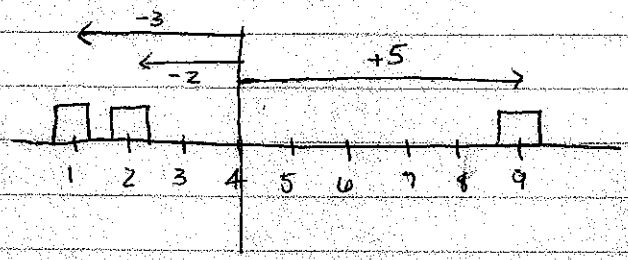
data values that are far from the bulk of the data

- Means are very sensitive to outliers
- The median, however, is not affected by outliers



- Measures of Spread

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



$\bar{y} = 4$

$$\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix} \xrightarrow[-\bar{y}]{\text{subtract}} \begin{bmatrix} 1-4 = -3 \\ 2-4 = -2 \\ 9-4 = 5 \end{bmatrix}$$

deviations from the mean

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \xrightarrow[-\bar{y}]{\text{subtract}} \begin{bmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ y_3 - \bar{y} \end{bmatrix}$$

- Need to get rid of cancellation of + and − signs
  - idea 1: use absolute values

$$\begin{bmatrix} |y_1 - \bar{y}| \\ \vdots \\ |y_n - \bar{y}| \end{bmatrix} = \begin{bmatrix} |-3| \\ |-2| \\ |5| \end{bmatrix} = \begin{matrix} 3 \\ 2 \\ 5 \end{matrix} \quad \text{mean} = \frac{10}{3} = 3.3$$

Mean Absolute Deviation (MAD) (not used much)

Amy Roget

Jan. 15, 2009 (Pg 2)

- idea 2: square the deviations

$$\begin{bmatrix}(y_1-\bar{y})^2\\(y_2-\bar{y})^2\\(y_3-\bar{y})^2\end{bmatrix} = \begin{bmatrix}-3^2\\-2^2\\5^2\end{bmatrix} = \begin{bmatrix}9\\4\\25\end{bmatrix}$$

mean = $\sqrt{12.7}$

- Standard Deviation (SD)

$$\sqrt{\frac{(y_1-\bar{y})^2+(y_2-\bar{y})^2+\ldots+(y_n-\bar{y})^2}{n-1}}$$

$$s=\sqrt{\frac{(y_1-\bar{y})^2+\ldots+(y_n-\bar{y})^2}{n-1}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

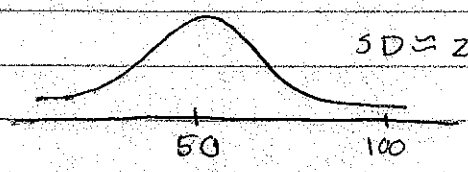
- Why divide by n-1 instead of n?

$\begin{bmatrix}1\\2\\9\end{bmatrix}$ mean = 4 $\quad$ $\begin{bmatrix}\checkmark\\\checkmark\\X\end{bmatrix}$ mean = 4 $\quad$ > 1st 2 free to vary

A data set has only n-1 degrees of freedom for measuring spread

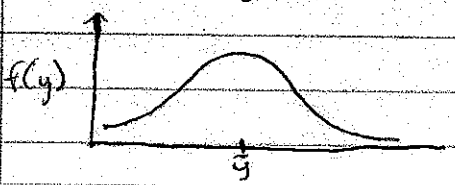
- Graphical Interpretation of SD

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



* each # in this data set is around 50, give or take around 20

1.4: Using the Normal Distribution Descriptively

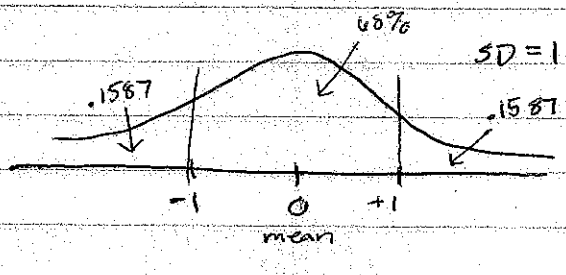


$$f(y)=\frac{1}{s\sqrt{2\pi}}e^{-\frac{(y-\bar{y})^2}{2s^2}}$$

Amy Roget

Jan. 15, 2009 (pg. 3)

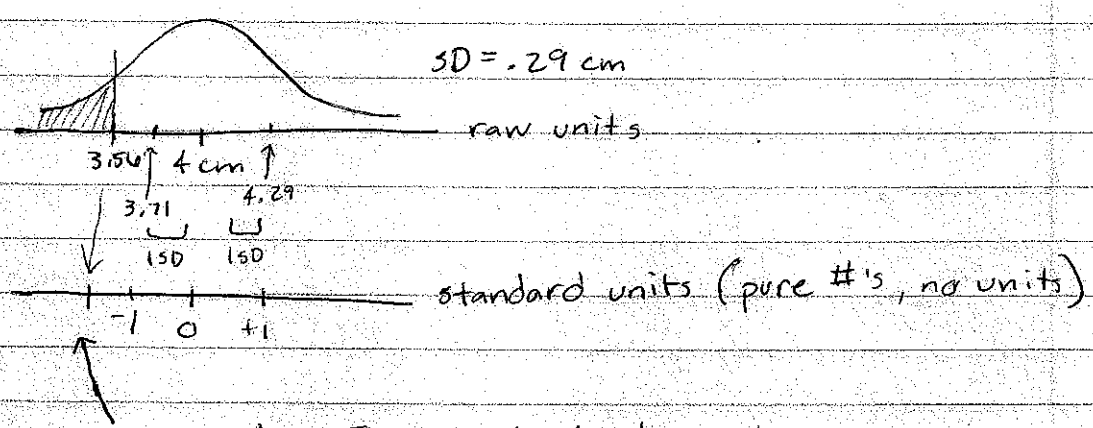
Standard Normal Curve



area = 1 - 2(.1587)

- Facts about Normal Curve
  1) symmetric
  2) total area under it is 1

Butterfly Wing Length



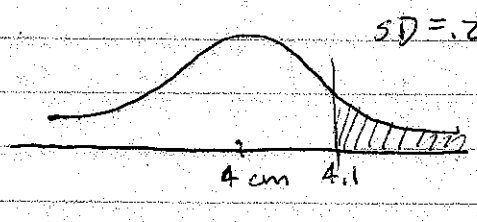
convert 3.56 to standard units

$$\frac{3.56 - 4.0 \text{ cm}}{.29 \text{ cm}} = -1.52$$

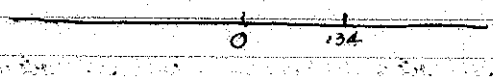
- use table to find area: = .0643 = 6.4%
  actual answer = 8.3%

* normal approximation here is good, but not great

Example:



What % of data is above 4.1 cm?



Amy Raget

Jan. 15, 2009 (Pg. 4)

$\frac{4.1 - 4.0 \text{ cm}}{.29 \text{ cm}} = \frac{.1}{.29} = .34$

- area using standard table = .3669 = 36.7% ≈ 37%
- actual % = 38%

- General formula for converting to standard units:

$$\frac{\# - \text{mean}}{SD} = z = \frac{y - \bar{y}}{s}$$