Math 243 Sections 2.2.2, 2.2.7 Standard Deviation and Comparing Distributions

Standard Deviation: A measure of spread for symmetric data. The "average deviation" or distance from the mean. The symbol for the population standard deviation is $\boldsymbol{\sigma}$. We use **s** to represent the standard deviation of a sample. $s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$

sigma population

sample standard deviation

Example 1. The table below shows a sample of 10 MTH 243 student's heart rates, measured in beats per minute (bpm). The mean of this set is 72. Calculate the deviation and squared deviation by hand. Then calculate the variance and the standard deviation.

Difference

Heart Rate (in bpm)	Deviation from the mean	Squared Deviation
52 −72	−20	$(-20)^2 = 400$
65 −72	−7	$(-7)^2 = 49$
67 −72	−5	$5^2 = 25$
68 −72	−4	$4^2 = 16$
70 −72	−2	$2^2 = 4$
73 −72	1	$1^2 = 1$
74 −72	2	$2^2 = 4$
78 −72	6	$6^2 = 36$
81 −72	9	$9^2 = 81$
92 −72	20	$20^2 = 400$
	Sum of the Squared Deviations	1016

Variance: $\mathbf{s}^2 = \frac{1016}{10-1} = \frac{1016}{9} = 112.8$ bpm^2

Standard Deviation: **s** =

$\sqrt{112.89} = 10.62$ bpm

Example 2. Comparing Standard Deviations

For a, b, and c, examine the two data sets. Without doing any calculations, which set has the higher standard deviation and why?

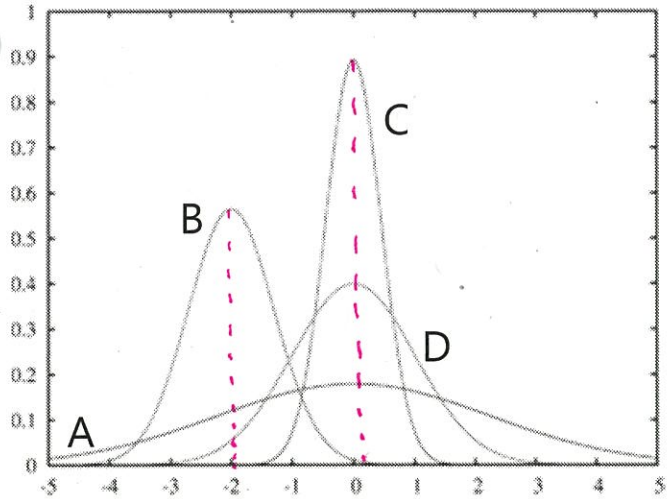
Set 1	Set 2
a) 4, 7, 7, 7, 10	4, 6, 7, 8, 10
b) 100, 140, 150, 160, 200	10, 50, 60, 70, 110
c) 10, 16, 18, 20, 22, 28	48, 56, 58, 60, 62, 70

a. Set 2 has a larger standard deviation

b. The sets have the same standard deviation

c. Set 2 has a larger standard deviation

Example 3. Center and spread. Comparing Means and Standard Deviations graphically



a. Which distributions have the same mean?

C, D, A

b. Put the distributions in order from smallest to largest standard deviation.

C, B, D, A

Graph Source: http://people.stern.nyu.edu/adamodar/New_Home_Page/StatFile/statdistns.htm

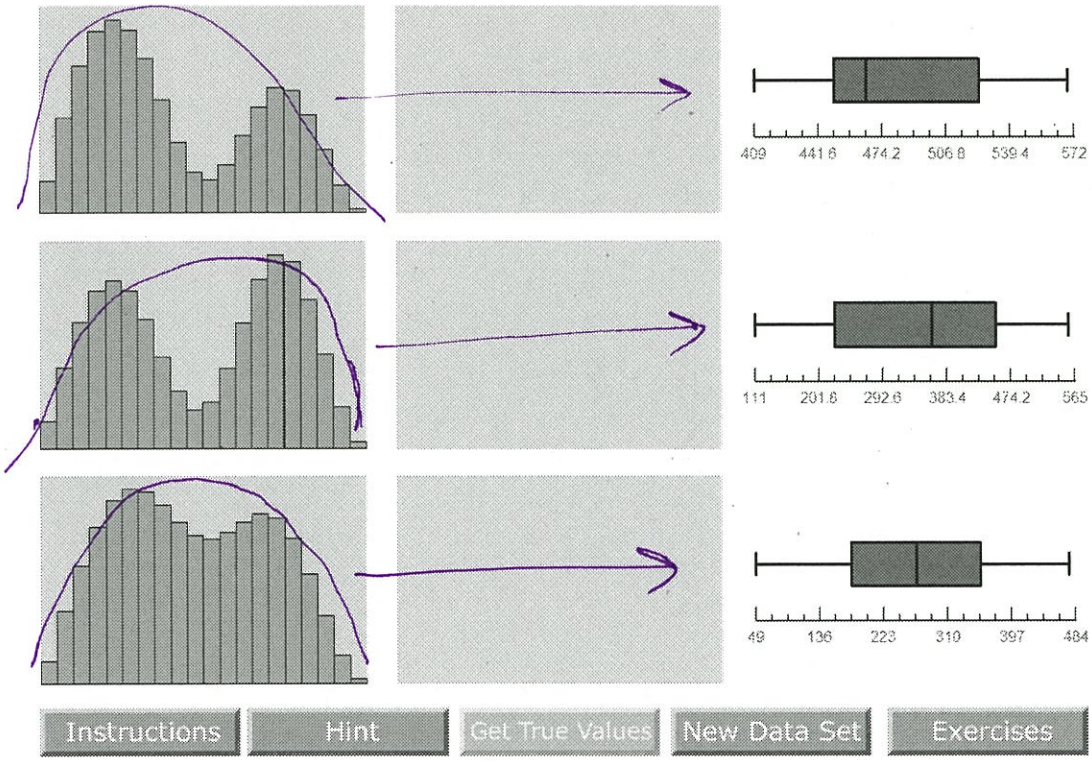
Relating Histograms and Boxplots for the Same Distribution

Class Activity 1. Matching histograms to boxplots

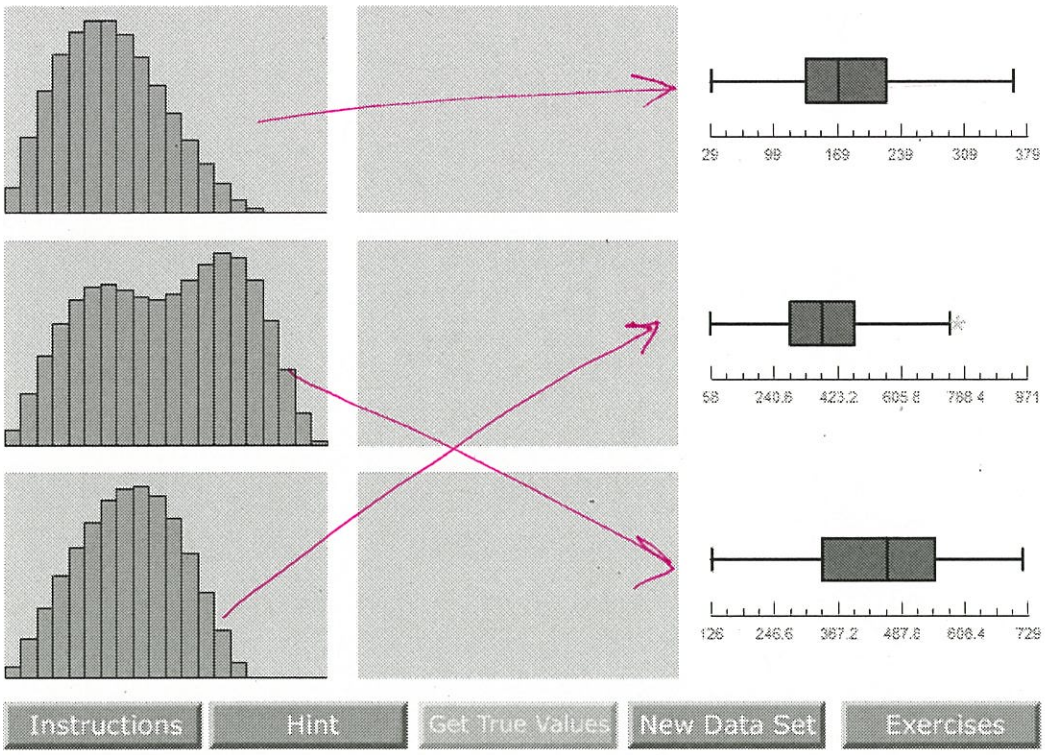
http://higheredbcs.wiley.com/legacy/college/mann/0470444665/applets/applet_01_v4.html

Now you try matching these histograms to the box-and-whisker plots:

Relating Histograms and Box-and-Whisker Plots

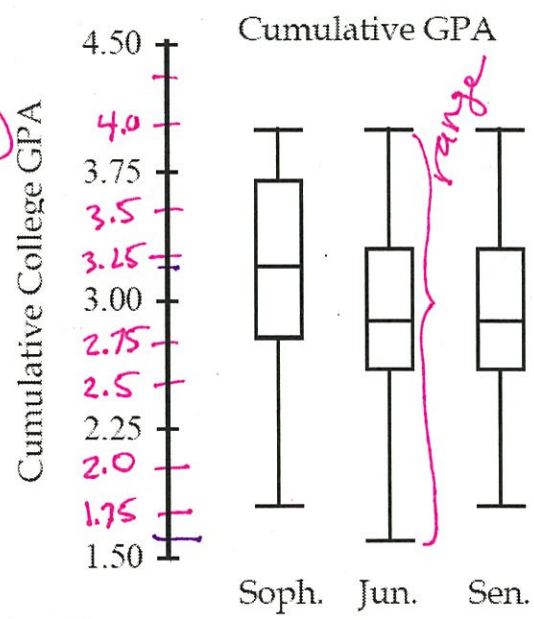


Relating Histograms and Box-and-Whisker Plots



Comparing Different Groups with Side-By-Side Boxplots

Example 4. The side-by-side boxplots show the cumulative college GPAs for sophomores, juniors and seniors taking an intro stats course.



a. Which class (sophomore, junior, or senior) had the lowest cumulative college GPA? What is the approximate value of that GPA?

The juniors (one or more students) had the lowest GPA at about 1.6.

b. Which class has the highest median GPA, and what is that GPA?

Sophomores have the highest median at about 3.2.

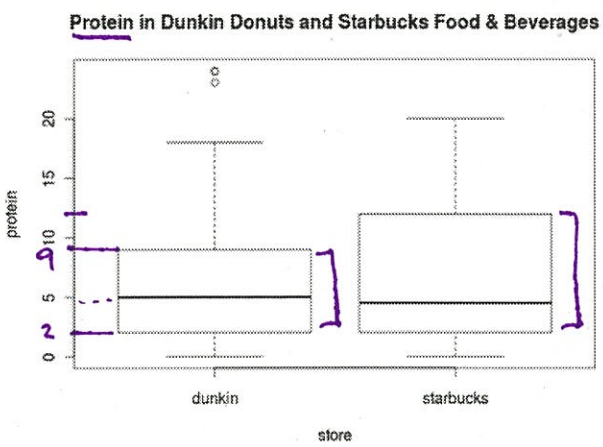
c. Which class has the largest range for GPA, and what is it? The juniors have the largest range. max − min = 4.0 − 1.6 = 2.4

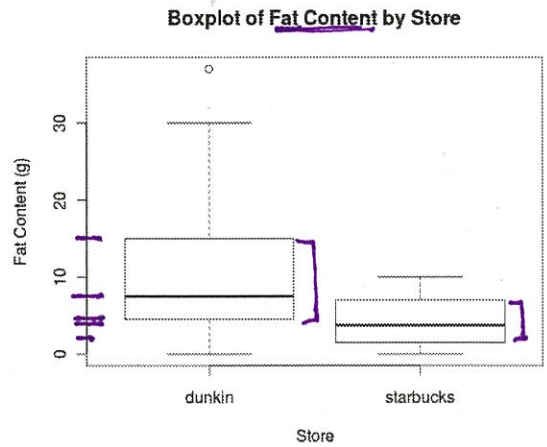
d. Which class has the most symmetric set of GPAs? — seniors.
The most skewed set of GPAs? — Sophomores.

Example 5. Comparing side-by-side box plots

Data was collected on the nutritional content in the food and beverages served at Starbucks and Dunkin Donuts. The study shows data collected from the menus of each establishment. Data about the calorie, fat, carbohydrate, protein and fiber content of 20 items from each store was collected. Ten beverages and ten food items were chosen and recorded.

Source: http://statsbls.weebly.com/nutrition-content-in-starbucks-and-dunkin-donuts.html





a. Compare the distributions (shape, center and spread, and unusual features) of the protein in food and beverages at Starbucks and Dunkin Donuts.

The shape of the distribution of protein content at both stores is skewed to the right. The medians are the same at about 5 grams of protein. The IQR of the starbucks items is larger at about 10 grams. The IQR at Dunkin is about 7 grams. Dunkin has 2 items with the highest protein 23+24 g.

b. Compare the distributions (shape, center and spread, and unusual features) of the fat content in the food and beverages at Starbucks and Dunkin Donuts.

The shape of both distributions is skewed to the right, but the fat content at Dunkin is much more skewed. The median fat content at Starbucks is about 4 grams compared with 8 grams at Dunkin. The IQR at Starbucks is about 6 grams of fat compared with 10 grams at Dunkin. Overall there is much more spread in the Dunkin items and they have one item that is an outlier at about 36 g. of fat.

Summarizing Distributions

Practice. The number of chocolate chips was counted in each of 34 Keebler Chocolate Chip cookies. The data are listed below. Use GeoGebra to calculate the summary statistics and make a histogram and boxplot. Write a paragraph to describe the distribution including all four characteristics.

29	31	25	32	27	31	30	29	31	26	32	33	32	30
33	29	30	28	32	35	37	31	24	30	30	34	29	27
24	38	37	32	26	30								

Summary Statistics

$\bar{x}$ = 30.4 chips (1 more decimal place than the data)

S = 3.46 chips (2 more decimal places than the data)

min = 24 chips

Q_1 = 29 chips

med = 30 chips

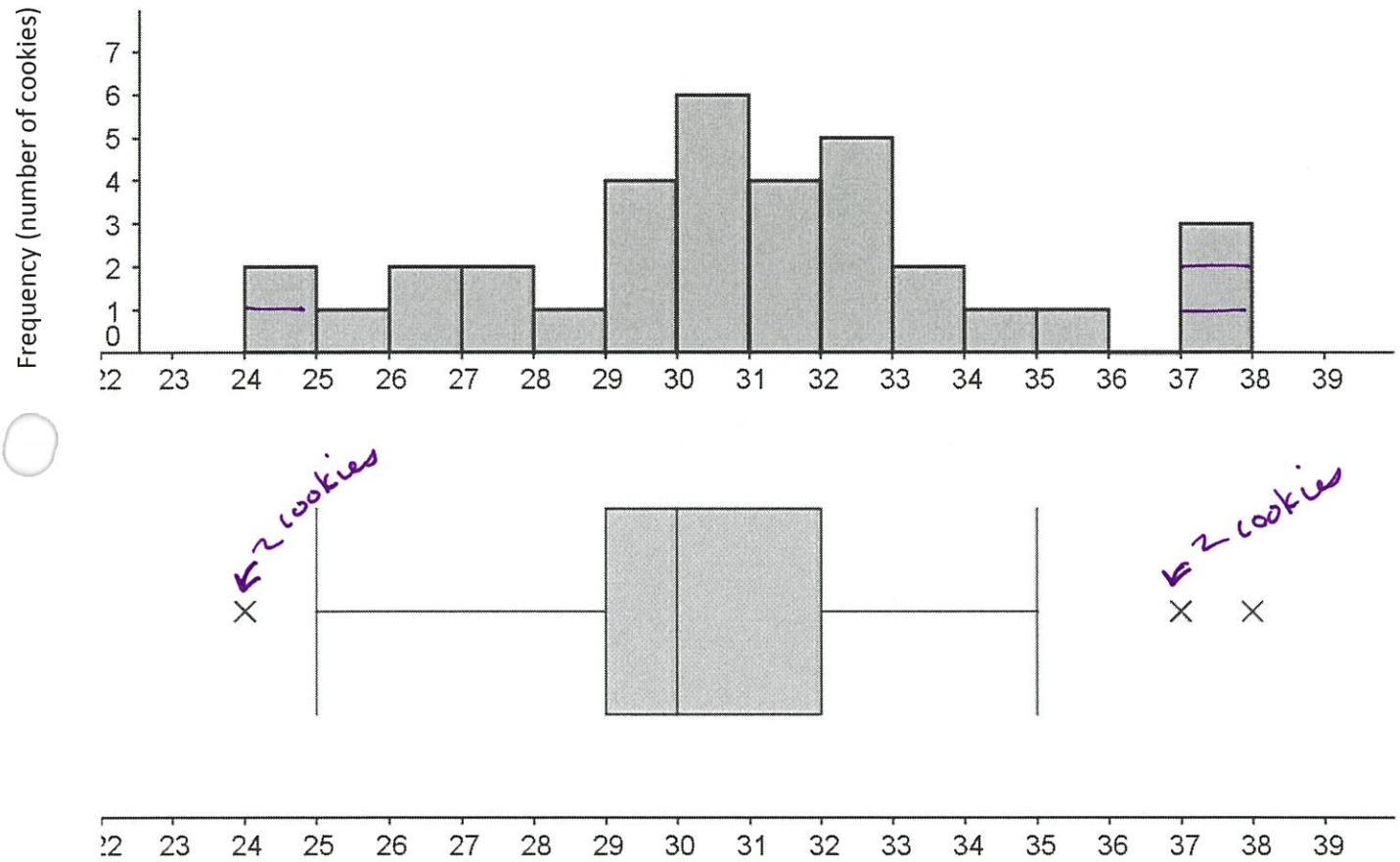
Q_3 = 32 chips

max = 38 chips

See key on the next page

Number of Chocolate Chips in Keebler Chocolate Chip Cookies

Summary Statistics:
$\bar{x}$ = 30.4 chips
s = 3.46 chips
minimum = 24 chips
Q_1 = 29 chips
median = 30 chips
Q_3 = 32 chips
maximum = 38 chips



The distribution of the number of chocolate chips per cookie in the Chips Ahoy brand is unimodal and symmetric. The mode is 30-31 chips per cookie, the mean is 30.4 chips and the median is 30 chips. This also indicates symmetry because all three measures of center are very close. The standard deviation is 3.46 chips. There are ~~four~~ 5 outliers, which are cookies with 24, 24, 37 and 38 chips. There is a gap in the data with no cookies having 36 chips. 137