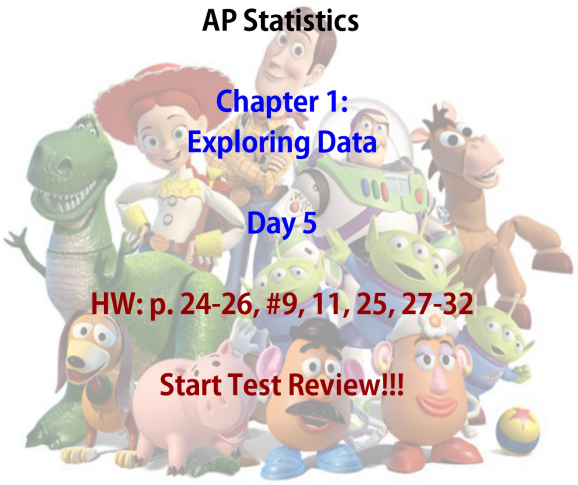
FRAPPY Review

- In your group, discuss the answers you came up with for your FRAPPY.
- Compare them to the answers given to you by the AP Board.
- Give yourself a grade (E, P, I) on each section and then add anything additional information you should have included in your response to earn an E for that section.
- Give yourself an overall score (4, 3, 2, 1) based on the rubric given by the AP Board.

AP Statistics



Chapter 1: Exploring Data

Day 5

HW: p. 24-26, #9, 11, 25, 27-32

Start Test Review!!!

p. 71-74, #89, 91, 95, 100

89.

(a) Putting the data in order we get: 74 75 76 78 80 82 84 86 87 90 91 93 96 98. There are 14 observations here so the first quartile is the median of the bottom 7 observations. This means that it is the value of the 4th observation. We find it to be 78. The third quartile is the median of the top 7 observations, so it is the value of the 11th observation. We find it to be 91. So IQR = 91 − 78 = 13. The middle 50% of the data have a spread of 13 points.

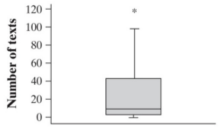
(b) Any outliers are below $Q_1 - 1.5\text{IQR}$ or above $Q_3 + 1.5\text{IQR}$. These are computed to be $78 - 1.5(13) = 58.5$ and $91 + 1.5(13) = 110.5$. There are no points outside of these bounds, so there are no outliers.

91.

(a) Using a stemplot to put the data in order:

```
0   0001133557889          1|4 represents 14 messages sent
1   4
2   5569
3
4   24
5   2
6
7   2
8
9   28
10
11  8
```

We now find that the median is 9, the first quartile is 3 and the third quartile is 43. The IQR is 40. So designate anything below $3 - 1.5(40) = -57$ or above $43 + 1.5(40) = 103$ as outliers. This means that the value of 118 is an outlier. The boxplot produced by computer software is shown below.



(b) The article claims that teens send 1742 texts a month. This works out to be about 58 texts a day (assuming a 30 day month). That seems pretty high given this data set. Twenty-one of the 25 students sent fewer than that, in fact, half of the students sent less than 10 messages (about 1/6th of the amount claimed in the article).

95.
(a) The stock fund varied between about −3.5% and 3%.

(b) The median return for the stock fund was slightly positive, about 0.1%, while the median real estate fund return appears to be close to 0%.

(c) The stock fund is much more variable. It has higher positive returns, but also higher negative returns.

100.
(a) It would appear that the distribution for the female doctors is more likely to be symmetric since the mean and median are relatively close together (19.1 and 18.5 respectively). The mean and median for the male doctors are quite far apart (41.333 and 34 respectively).

(b) The IQR measures the range of the middle 50% of the data. This does not take outliers into consideration. The standard deviation, however, uses every point and is not resistant to outliers. So, while the middle 50% of the data set may look very similar, if one data set has many more outliers, it will have a larger standard deviation.

(c) It does appear that males perform more C-sections. Each of the numbers in the 5-number summary was larger for the males than for the females.

Exploring Categorical Data

- Data that are labels, names, or other nonnumerical outcomes.
  - Gender (male, female)
  - Car color (white, red, blue, etc.)
  - One's response to a survey question (yes, no, maybe)
- Typically organized into frequency tables or bar charts.

One-Way Table

Flavor		
vanilla	strawberry	vanilla
chocolate	vanilla	strawberry
vanilla	chocolate	chocolate
vanilla	chocolate	vanilla

Counts

Flavor	Chocolate	Strawberry	Vanilla	Total
Frequency	4	2	6	12

Relative Frequency (%)

Flavor	Chocolate	Strawberry	Vanilla	Total
Frequency	33.3%	16.7%	50%	100%

Two-Way Table

Flavor	Gender	Flavor	Gender
vanilla	male	chocolate	female
chocolate	female	chocolate	male
vanilla	male	vanilla	female
vanilla	male	strawberry	male
strawberry	female	chocolate	female
vanilla	male	vanilla	male

Counts

		Flavor			
		Chocolate	Strawberry	Vanilla	Total
Gender	Female	3	1	1	5
	Male	1	1	5	7
	Total	4	2	6	12

The totals are called marginal distributions.

Two-Way Tables

Two-way tables can also display relative frequencies.

These relative frequencies can be of the entire table, of rows, or of columns.

Relative Frequency of the Row

Relative Frequency of the row		Flavor			
		Chocolate	Strawberry	Vanilla	Total
Gender	Female	60.0%	20.0%	20.0%	100%
	Male	14.3%	14.3%	71.4%	100%
	Total	33.3%	16.7%	50.0%	100%

The relative frequencies for each cell are called conditional distributions.

Example

Create a two-way table of relative frequencies, with each cell being the relative frequency of the column. Round each relative frequency to three decimal places.

		Flavor			
		Chocolate	Strawberry	Vanilla	Total
Gender	Female	3	1	1	5
	Male	1	1	5	7
	Total	4	2	6	12

Relative Frequency of the column		Flavor			
		Chocolate	Strawberry	Vanilla	Total
Gender	Female	75%	50%	16.7%	41.7%
	Male	25%	50%	83.3%	58.3%
	Total	100%	100%	100%	100%

Two-Way Tables

The purpose of creating two-way tables with conditional distributions is to look for an association between the variables.

In the previous example, one may be assessing whether there is an association between gender and preferred ice cream flavor.

Relative Frequency of the row		Flavor			
		Chocolate	Strawberry	Vanilla	Total
Gender	Female	60.0%	20.0%	20.0%	100%
	Male	14.3%	14.3%	71.4%	100%
	Total	33.3%	16.7%	50.0%	100%

Two-Way Tables

Does there appear to be an association between gender and preferred ice cream flavor?

Yes. 60% of females prefer chocolate compared to 14.3% of males. 71.4% of males prefer vanilla compared 20% of females. Neither males nor females prefer strawberry.

Relative Frequency of the row		Flavor			
		Chocolate	Strawberry	Vanilla	Total
Gender	Female	60.0%	20.0%	20.0%	100%
	Male	14.3%	14.3%	71.4%	100%
	Total	33.3%	16.7%	50.0%	100%

Two-Way Tables

A common method of graphing a conditional distribution is with a relative frequency segmented bar chart.

One bar is drawn for each category of the conditional variable.

Each bar is segmented into parts whose length is proportional to the percentage of categories of the other second variable.

Example

Make a relative frequency segmented bar chart for the conditional distribution shown below.

Relative Frequency of the row

		Flavor			
		Chocolate	Strawberry	Vanilla	Total
Gender	Female	60.0%	20.0%	20.0%	100%
	Male	14.3%	14.3%	71.4%	100%
	Total	33.3%	16.7%	50.0%	100%