STA 291
Lecture 13, Chap. 6

- **Describing Quantitative Data**
  - Measures of Central Location
  - Measures of Variability (spread)

Summarizing Data Numerically

- Center of the data
  - Mean (average)
  - Median
  - Mode (…will not cover)
- Spread of the data
  - Variance, Standard deviation
  - Inter-quartile range
  - Range

Mathematical Notation: Sample Mean

- Sample size *n*
- Observations $x_1, x_2, \ldots, x_n$
- Sample Mean "x-bar" --- a statistic

$$\bar{\mathrm{x}} = (x_1 + x_2 + \ldots + x_n) / n$$

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

$$\sum = \mathrm{SUM}$$

Mathematical Notation: Population Mean for a finite population of size *N*

- Population size (finite) N
- Observations $x_1 , x_2 ,\ldots, x_N$
- Population Mean “mu” --- a Parameter

$$m = (x_1 + x_2 + \ldots + x_N)/N$$

$$= \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\sum = \mathrm{SUM}$$

Percentiles

- The *p*th percentile is a number such that *p*% of the observations take values below it, and (100-*p*)% take values above it
- 50th percentile = median
- 25th percentile = lower quartile
- 75th percentile = upper quartile

Quartiles

- 25th percentile = lower quartile

  = Q1

- 75th percentile = upper quartile

  = Q3

Interquartile range = Q3 - Q1

(a measurement of variability in the data)

SAT Math scores

- Nationally (min = 210 max = 800)

  Q1 = 440

  Median = Q2 = 520

  Q3 = 610 (-- you are better than 75% of all test takers)

- Mean = 518 (SD = 115 what is that?)

SAT Percentile Ranks

Critical Reading, Mathematics, and Writing

Score	Critical Reading	Mathematics	Writing
800	99	99	99+
790	99	99	99+
780	99	99	99
770	99	99	99
760	99	98	99
750	98	98	99
740	98	97	98
730	97	97	98
720	96	96	97
710	96	95	97
700	95	93	96
690	94	92	95
680	93	91	94
670	92	89	93
660	90	88	92
650	89	86	90
640	87	83	89
630	85	81	87
620	83	79	85
610	82	76	83
600	79	74	81
590	77	71	79
580	74	68	76
570	71	66	73
560	68	63	71
550	65	60	68
540	62	56	64
530	58	53	62
520	55	50	58
510	51	47	54
500	48	43	51
490	44	40	47
480	41	36	44
470	37	33	40
460	34	30	37
450	31	27	33

Five-Number Summary

- Maximum, Upper Quartile, Median, Lower Quartile, Minimum
- Statistical Software SAS output (Murder Rate Data)

```
Quantile        Estimate


100% Max           20.30
75% Q3             10.30
50% Median          6.70
25% Q1              3.90
0% Min              1.60
```

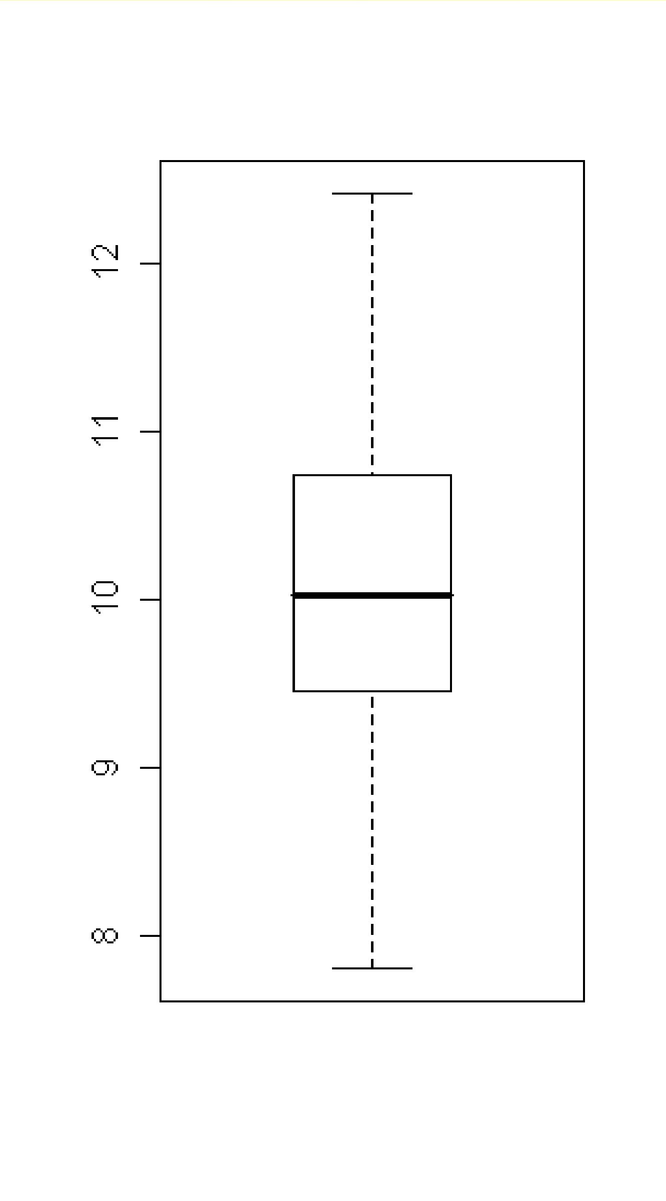
Five-Number Summary

- Maximum, Upper Quartile, Median,
  Lower Quartile, Minimum

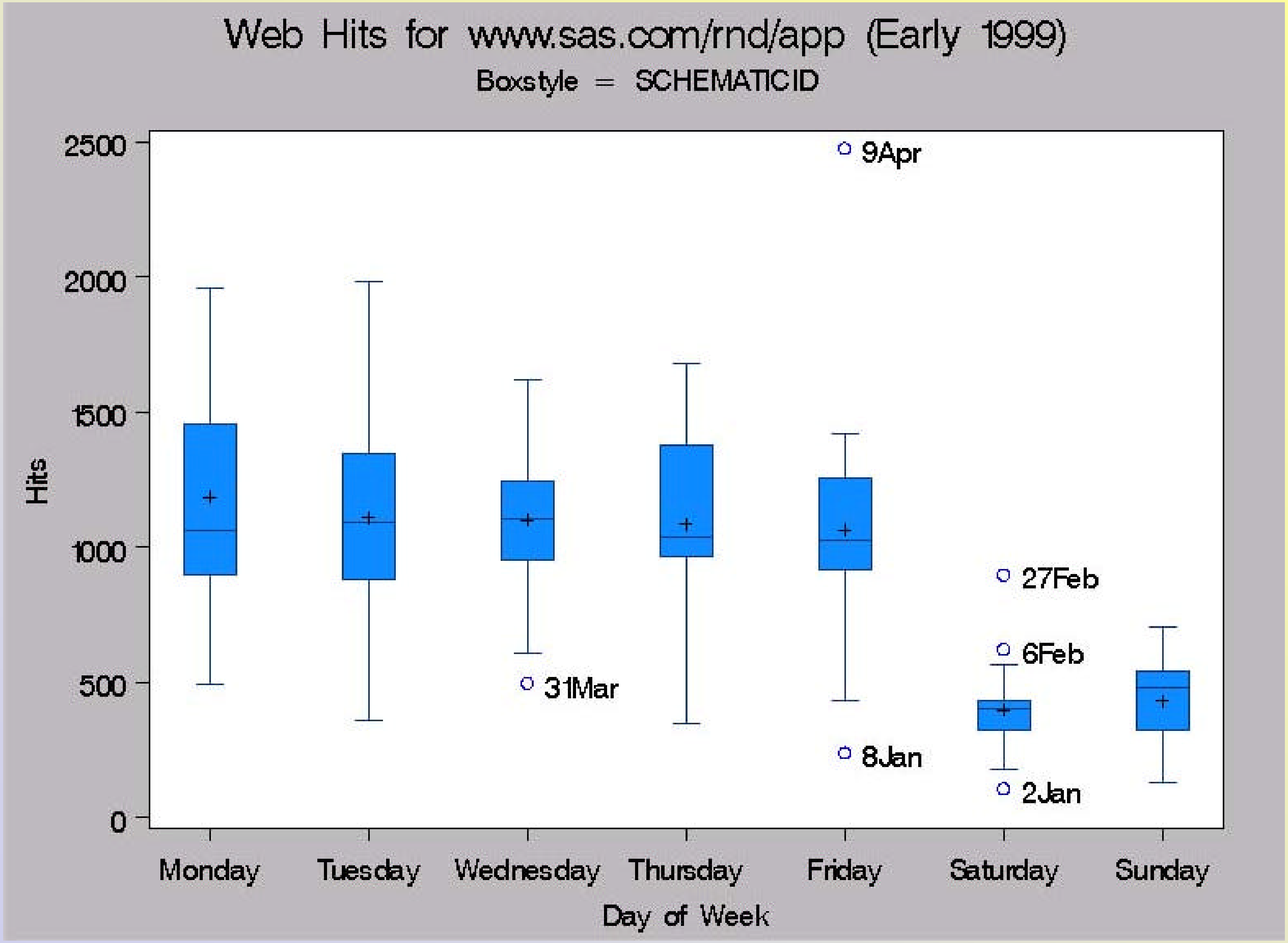
- Example: The five-number summary for a data set is min=4, Q1=256, median=530, Q3=1105, max=320,000.
- What does this suggest about the shape of the distribution?

Box plot

- A box plot is a graphic representation of the five number summary --- provided the max is within 1.5 IQR of Q3 (min is within 1.5 IQR of Q1)



- Otherwise the max (min) is suspected as an **outlier** and treated differently.


Web Hits for www.sas.com/rnd/app (Early 1999)
Boxstyle = SCHEMATICID
Hits
2500
2000
1500
1000
500
0
9Apr
27Feb
6Feb
31Mar
8Jan
2Jan
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Day of Week

- Box plot is most useful when compare several populations

Measures of Variation

- Mean and Median only describe the central location, but not the spread of the data
- Two distributions may have the same mean, but different variability
- Statistics that describe variability are called measures of spread/variation

Measures of Variation

- Range: = max - min

  Difference between maximum and minimum value

- Variance: $$s^2 = \frac{\sum (x_i - \overline{x})^2}{n-1}$$
-
- Standard Deviation: $$s = \sqrt{s^2} = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n-1}}$$

- Inter-quartile Range: = Q3 – Q1

  Difference between upper and lower quartile of the data

Deviations: Example

- Sample Data: 1, 7, 4, 3, 10
- Mean (x-bar): (1+7+4+3+10)/5 =25/5=5

data	Deviation	Dev. square
1	(1 - 5)= -4	16
3	(3 - 5)= -2	4
4	(4 - 5) = -1	1
7	(7 - 5) = 2	4
10	(10 - 5) = 5	25
Sum=25	Sum = 0	sum = 50

Sample Variance

$$s^2 = \frac{\sum (x_i - \overline{x})^2}{n-1}$$

The variance of *n* observations is the sum of the squared deviations, divided by *n-1.*

Variance: Example

Observation	Mean	Deviation	Squared Deviation
1	5		16
3	5		4
4	5		1
7	5		4
10	5		25
Sum of the Squared Deviations			50
n-1			5-1=4
Sum of the Squared Deviations / *(n-1)*			50/4=12.5

- So, sample variance of the data is 12.5

- Sample standard deviation is 3.53

$$\sqrt{12.5} = 3.53$$

- Variance/standard deviation is also more susceptible to extreme valued observations.

- We are using x-bar and variance/standard deviation mostly in the rest of this course.

Population variance/standard deviation

- Notation for Population variance/standard deviation (usually obtain only after a census)

- Sigma-square / sigma

s^{2} s

standardization

- Describe a value in a sample by
- "how much standard deviation above/below the average"

- The value 6 is one standard deviation above mean -- the value 6 corresponds to a z-score of 1
- May be negative (for below average)

Attendance Survey Question

- On a 4”x6” index card
  - write down your name and section number
  - Question: Independent or not?
  - Gender of first child and second child from same couple.