0825 Statistics and Graphs

Note Title 8/25/2010

Note: Most of today was using Excel & R
View the R script (a text file)
for some clarification.

§1.2 Summary statistics

Note:

- statistics are summaries of the sample
  parameters are summaries of the population

Notation

n = sample size
N = population size

Central Tendency

mode: most frequent
median: middle
mean: $\frac{1}{n}\sum x_i$
trimmed mean: mean after
removing upper & lower p%

The three means:

(Arithmetic) mean $\frac{1}{n}\sum x_i = \bar{x}$

Rarely used: Geometric mean $(\prod x_i)^{1/n}$

Harmonic mean $\frac{n}{\sum 1/x_i}$

Measures of Spread

mean-based:

Standard deviation $= s = \sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2}$

Variance $= s^2$

median-based:

Interquartile Range $= IQR = Q_3 - Q_1$

Other Summaries

Quartiles $= Q_n := P_{n/4}$

Percentiles $= P_p$

$P_p := x_j$ such that $\frac{1}{n}\left(\left|\{x_i \mid x_i < x_j\}\right|\right) = p$

Five-number Summary $\{\min, Q_1, Q_2, Q_3, \max\}$

Statistical summary	program fctn Excel	R
sample size		length (x)
mean	AVERAGE( )	mean (x)
trimmed mean		mean (x, trim = p)
geometric mean		prod (x) ^ (1 / length (x))
harmonic mean		length(x) / (sum (1/x))
median	MEDIAN( )	median (x)
minimum	MIN( )	min (x)
maximum	MAX( )	max (x)
IQR		IQR (x)
std deviation	STDEV( )	sd (x)
variance	VAR( )	var (x)
Quantiles		quantile (x)
Percentiles		quantile (x, p)
5num summary		fivenum (x)

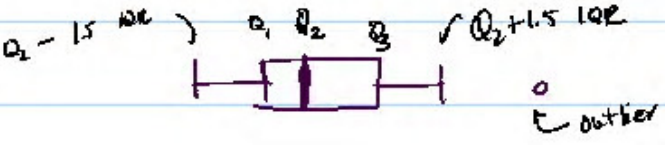
Graphing

Why? Numbers are great, but graphs help to give us a "feel" for the data.

stem and leaf — frequencies

histograms — frequencies

boxplots



scatterplot (bivariate) next class