APPLICATIONS OF MATLAB IN ENGINEERING

Yan-Fu Kuo
Dept. of Bio-industrial Mechatronics Engineering
National Taiwan University

Fall 2015

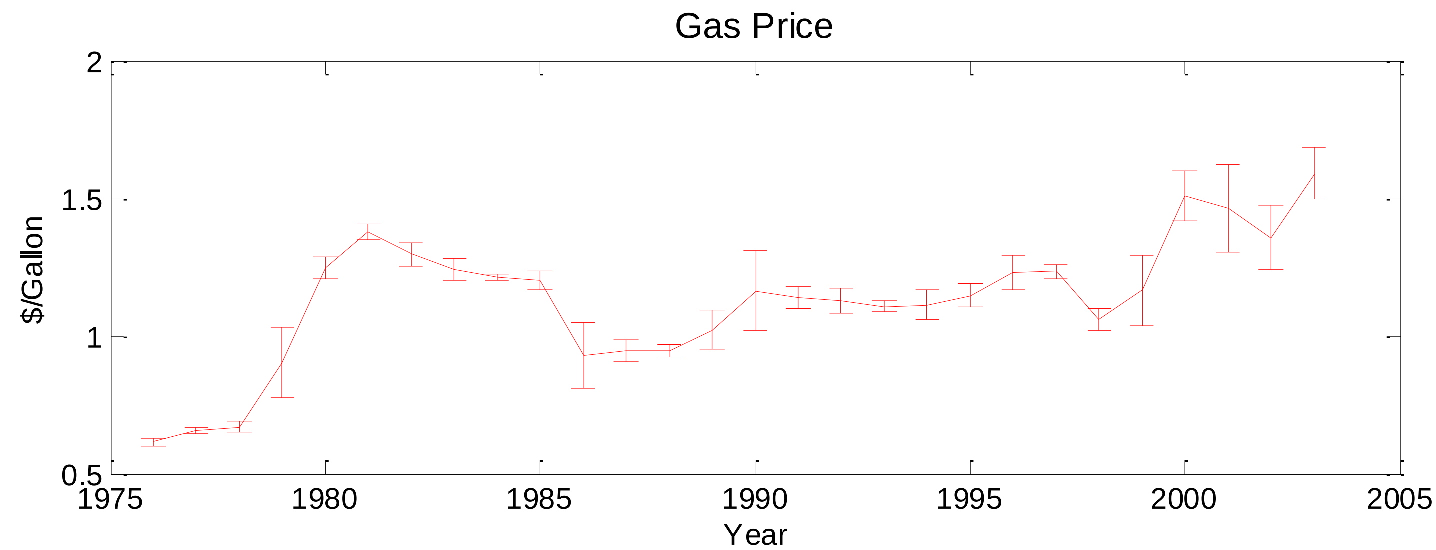
Today:

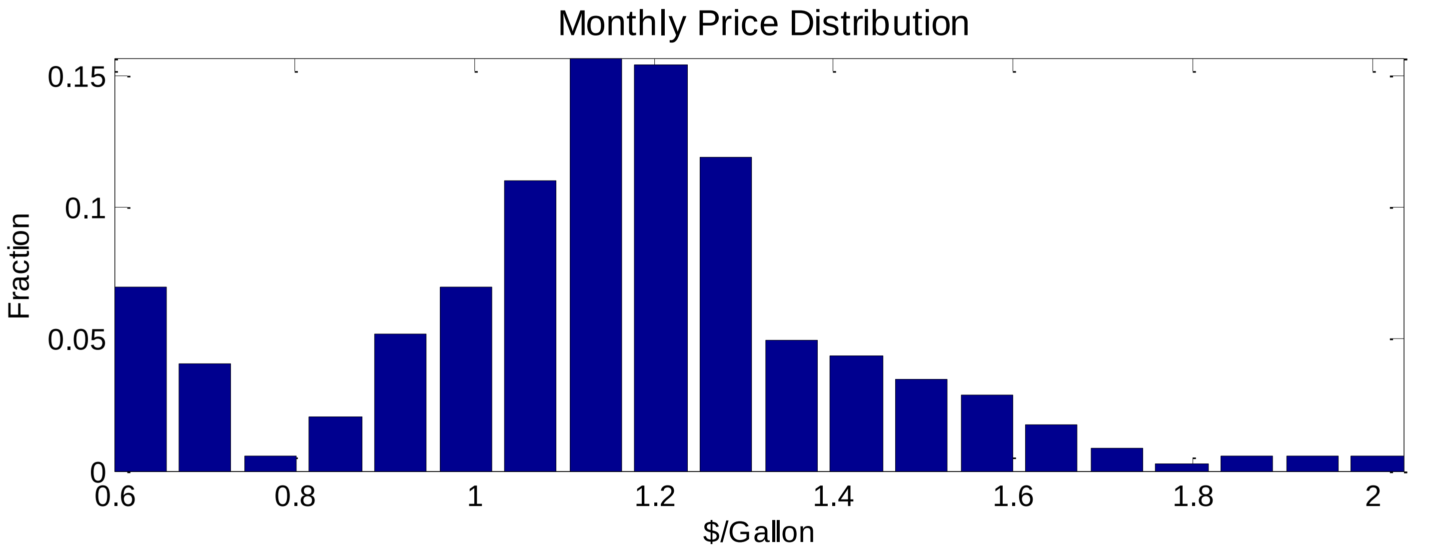
- Statistics

USA Gasoline Prices from 1984 to 2004

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1984	1.216	1.209	1.21	1.227	1.236	1.229	1.212	1.196	1.203	1.209	1.207	1.193
1985	1.148	1.131	1.159	1.205	1.231	1.241	1.242	1.229	1.216	1.204	1.207	1.208
1986	1.194	1.12	0.981	0.888	0.923	0.955	0.89	0.843	0.86	0.831	0.821	0.823
1987	0.862	0.905	0.912	0.934	0.941	0.958	0.971	0.995	0.99	0.976	0.976	0.961
1988	0.933	0.913	0.904	0.93	0.955	0.955	0.967	0.987	0.974	0.957	0.949	0.93
1989	0.918	0.926	0.94	1.065	1.119	1.114	1.092	1.057	1.029	1.027	0.999	0.98
1990	1.042	1.037	1.023	1.044	1.061	1.088	1.084	1.19	1.294	1.378	1.377	1.354
1991	1.247	1.143	1.082	1.104	1.156	1.16	1.127	1.14	1.143	1.122	1.134	1.123
1992	1.073	1.054	1.058	1.079	1.136	1.179	1.174	1.158	1.158	1.154	1.159	1.136
1993	1.117	1.108	1.098	1.112	1.129	1.13	1.109	1.097	1.085	1.127	1.113	1.07
1994	1.043	1.051	1.045	1.064	1.08	1.106	1.136	1.182	1.177	1.152	1.163	1.143
1995	1.129	1.12	1.115	1.14	1.2	1.226	1.195	1.164	1.148	1.127	1.101	1.101
1996	1.129	1.124	1.162	1.251	1.323	1.299	1.272	1.24	1.234	1.227	1.25	1.26
1997	1.261	1.255	1.235	1.231	1.226	1.229	1.205	1.253	1.277	1.242	1.213	1.177
1998	1.131	1.082	1.041	1.052	1.092	1.094	1.079	1.052	1.033	1.042	1.028	0.986
1999	0.972	0.955	0.991	1.177	1.178	1.148	1.189	1.255	1.28	1.274	1.264	1.298
2000	1.301	1.369	1.541	1.506	1.498	1.617	1.593	1.51	1.582	1.559	1.555	1.489
2001	1.472	1.484	1.447	1.564	1.729	1.64	1.482	1.427	1.531	1.362	1.263	1.131
2002	1.139	1.13	1.241	1.407	1.421	1.404	1.412	1.423	1.422	1.449	1.448	1.394
2003	1.473	1.641	1.748	1.659	1.542	1.514	1.524	1.628	1.728	1.603	1.535	1.494

USA Gasoline Prices from 1984 to 2004

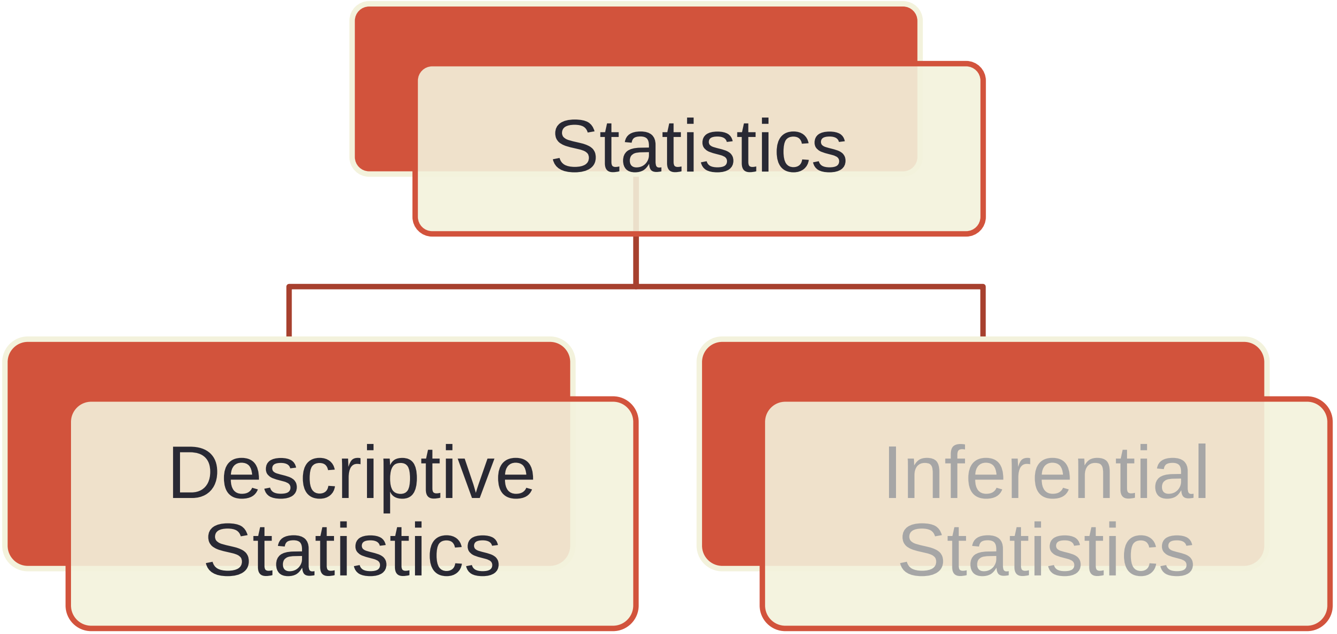




Statistics

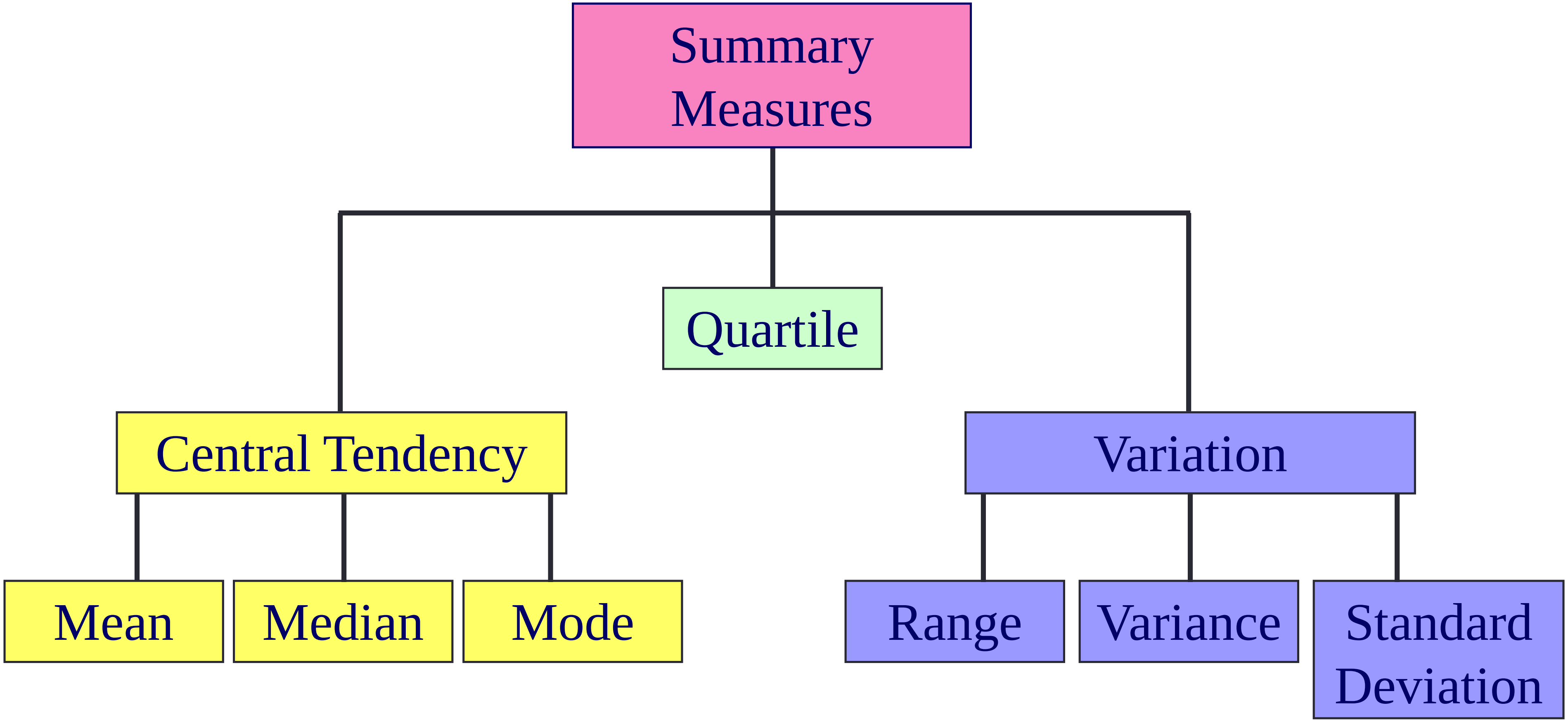
- The science of "data"
- Involving the collection, analysis, interpretation, presentation, and organization of data

Main Statistical Methodologies



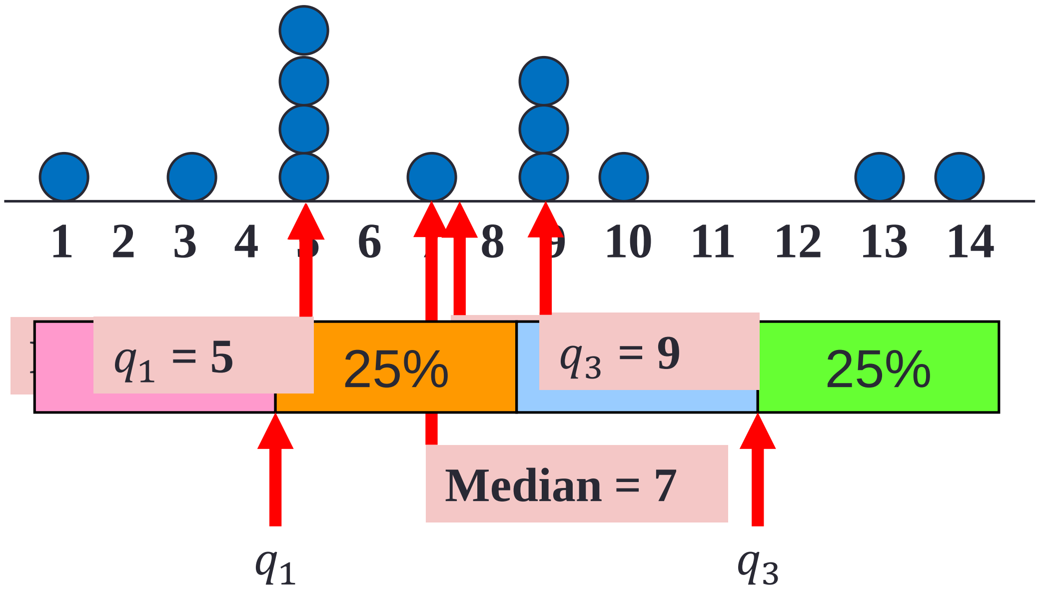
Numerical and graphical methods to look for patterns, to summarize the information in a data set

Summary Measures



Mean, Median, Mode, and Quartile

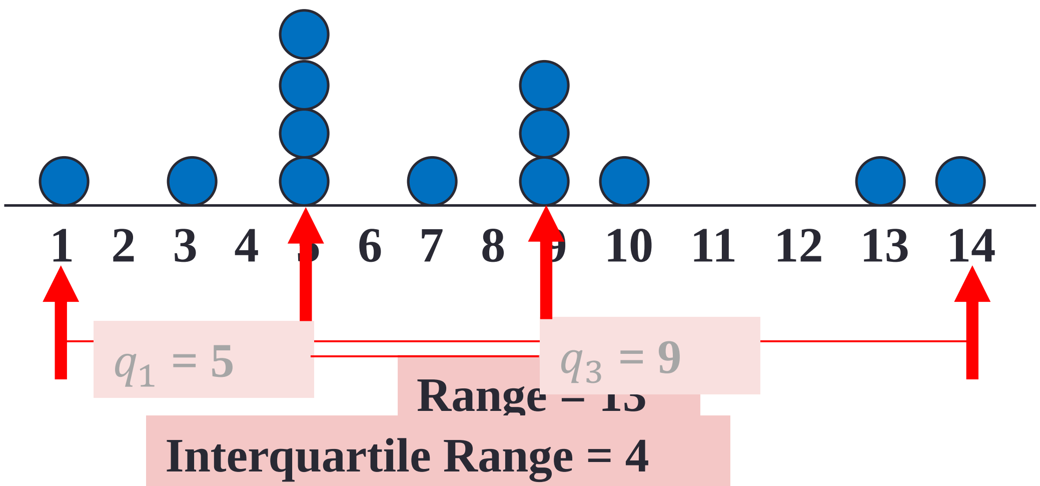
- Suppose we have samples:



mean	Average or mean value of array
median	Median value of array
mode	Most frequent values in array
prctile	Percentiles of a data set

Range and Interquartile Range

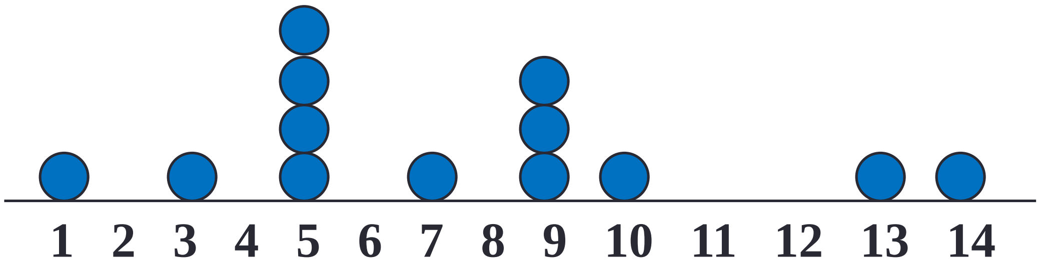
- Suppose we have samples:



max	Largest elements in array
min	Smallest elements in array

Variance and Standard Deviation

- Suppose we have samples:



- Variance: $s = \frac{\sum(x_i - \bar{x})^2}{n-1} = 14.3974$
- Standard deviation: $s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = 3.7944$

`std`	Standard deviation
`var`	Variance

Exercise

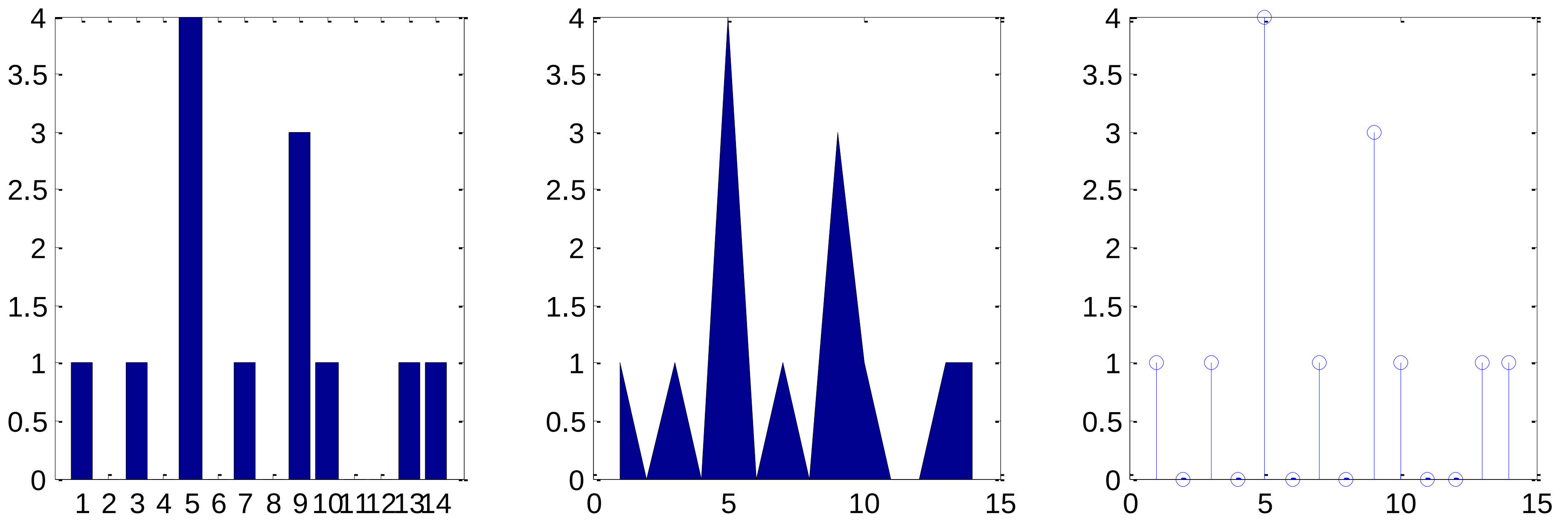
- Find the following properties of the variable `x4`
  1. Mean, median, mode, and quartile
  2. Range and interquartile range
  3. Variance and standard deviation

```
load stockreturns;
x4 = stocks(:,4);
```

Figures Are Always More Powerful

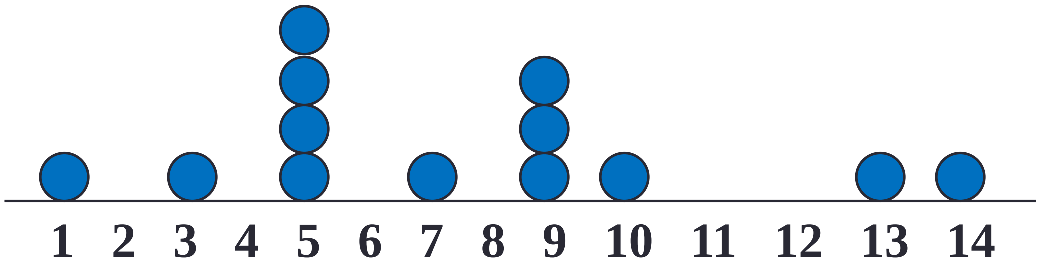
- Suppose we have samples:

```
x = 1:14;
freqy = [1 0 1 0 4 0 1 0 3 1 0 0 1 1];
subplot(1,3,1);  bar(x,freqy);  xlim([0 15]);
subplot(1,3,2);  area(x,freqy);  xlim([0 15]);
subplot(1,3,3);  stem(x,freqy);  xlim([0 15]);
```



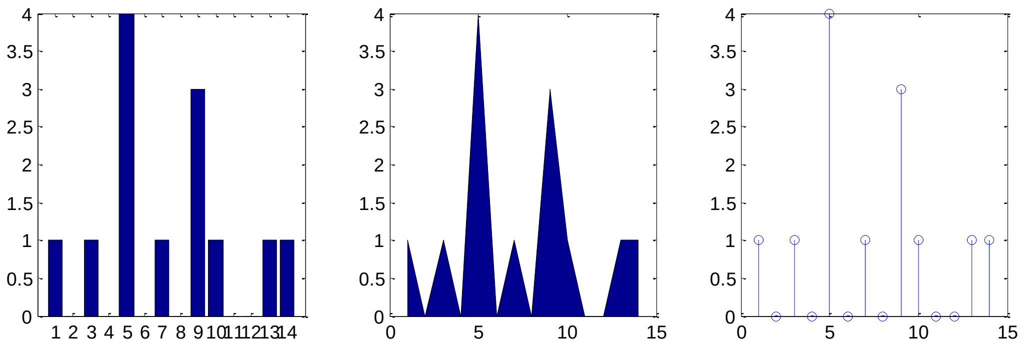
Exercise

- Suppose we are given the samples:



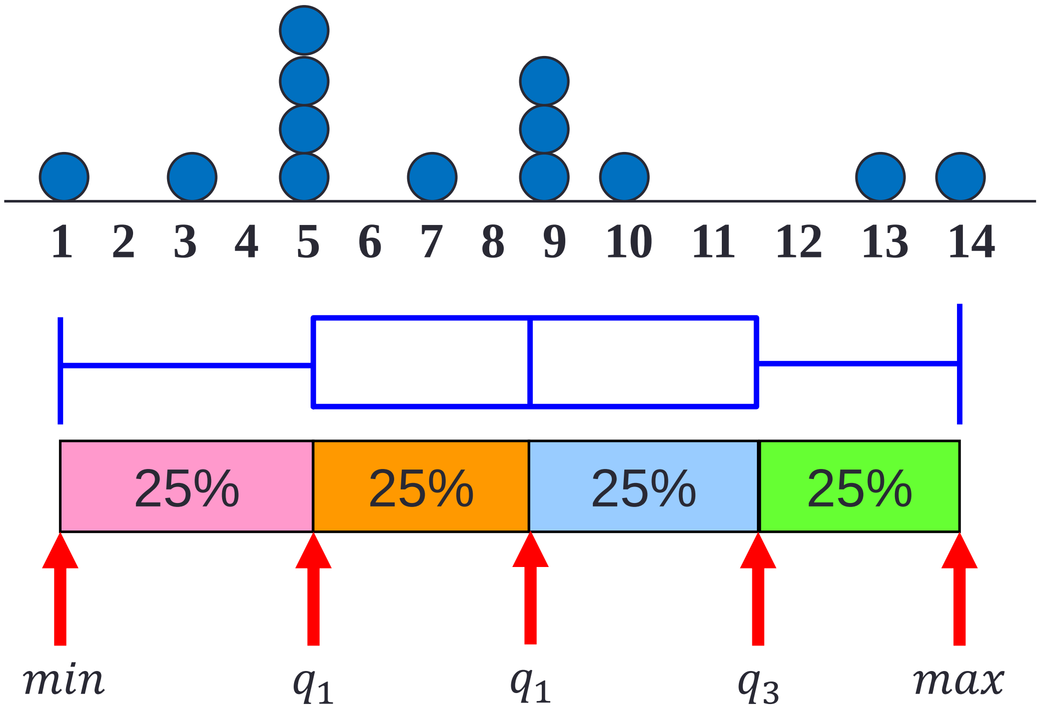

```
x = [1 3 5 5 5 5 7 9 9 9 10 13 14];
```

- Plot the histograms:



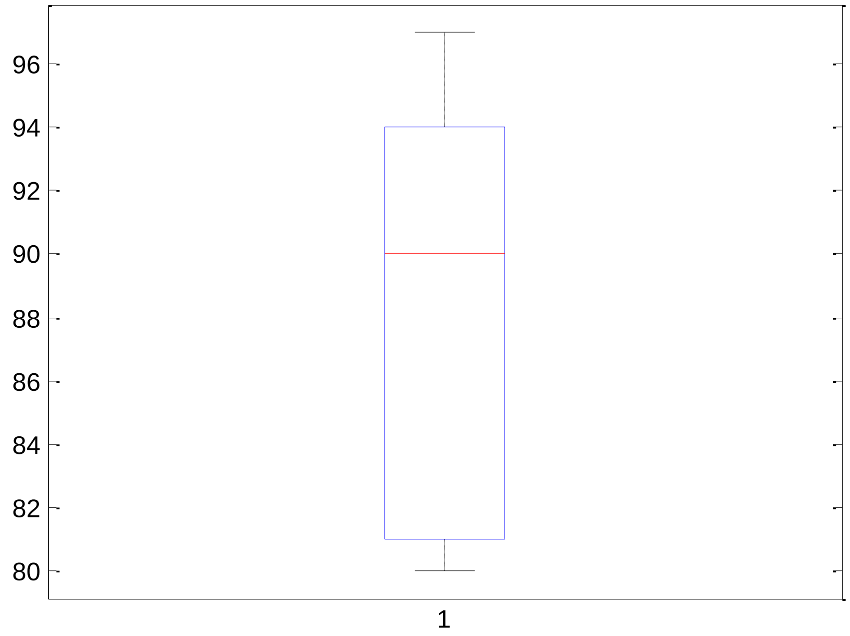
Boxplot

- Suppose we are given the samples:



Boxplot Example

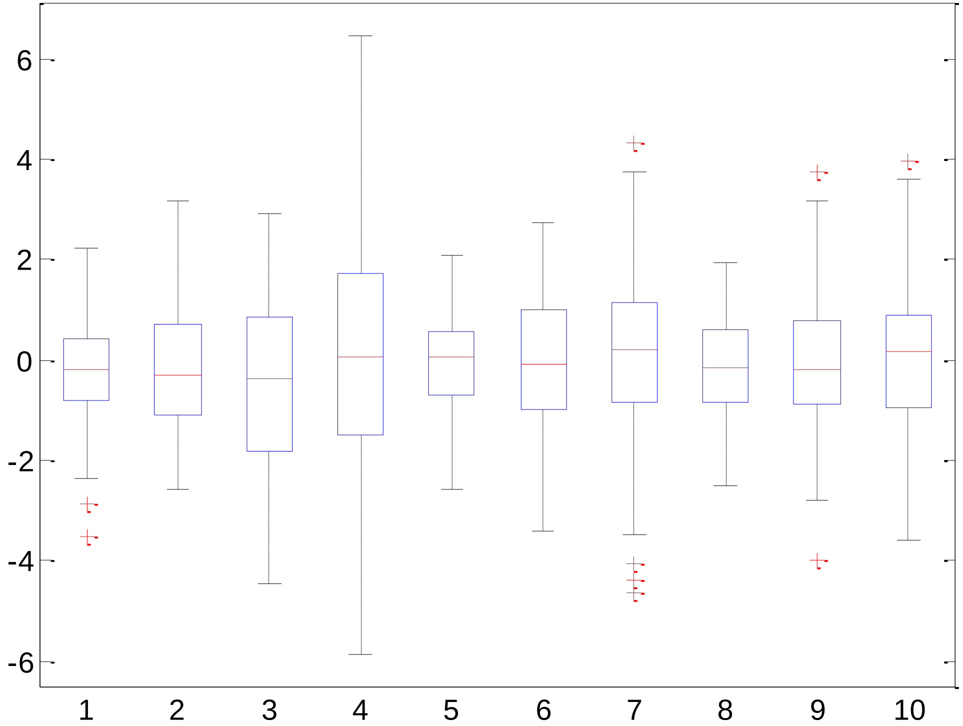
```
marks = [80 81 81 84 88 92 92 94 96 97];
boxplot(marks)
prctile(marks, [25 50 75])
```



Exercise

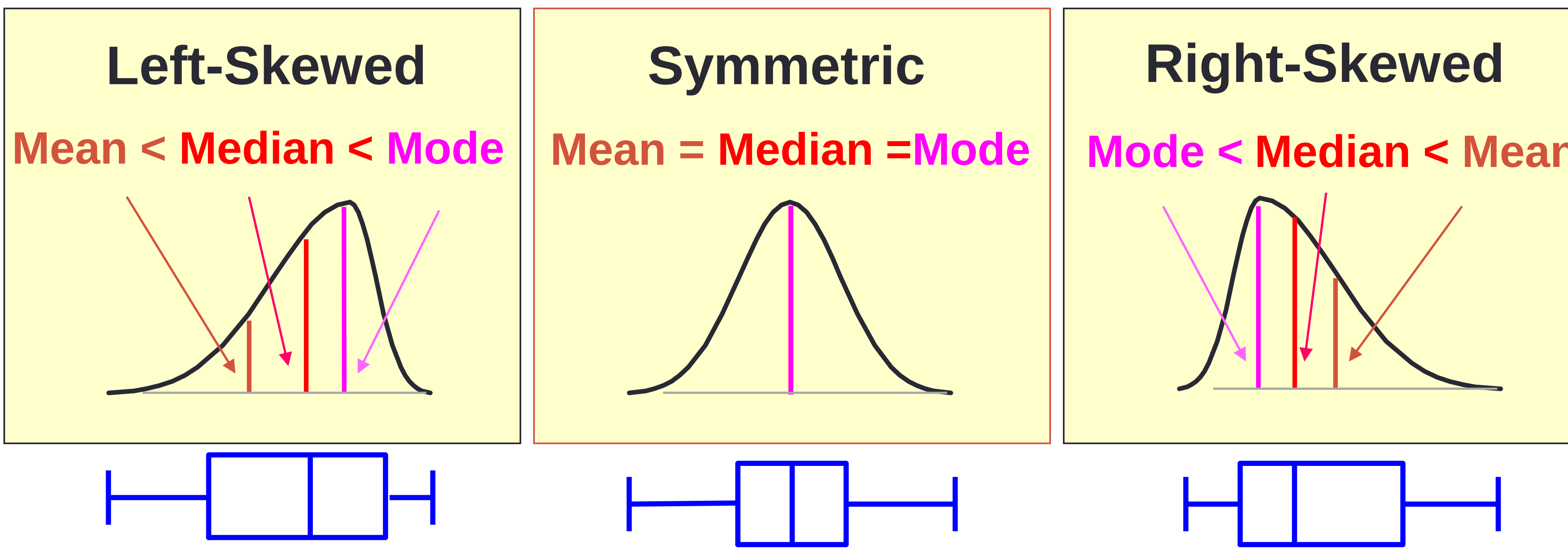
- Plot the boxplot of the variable `stocks`

```
load stockreturns;
```



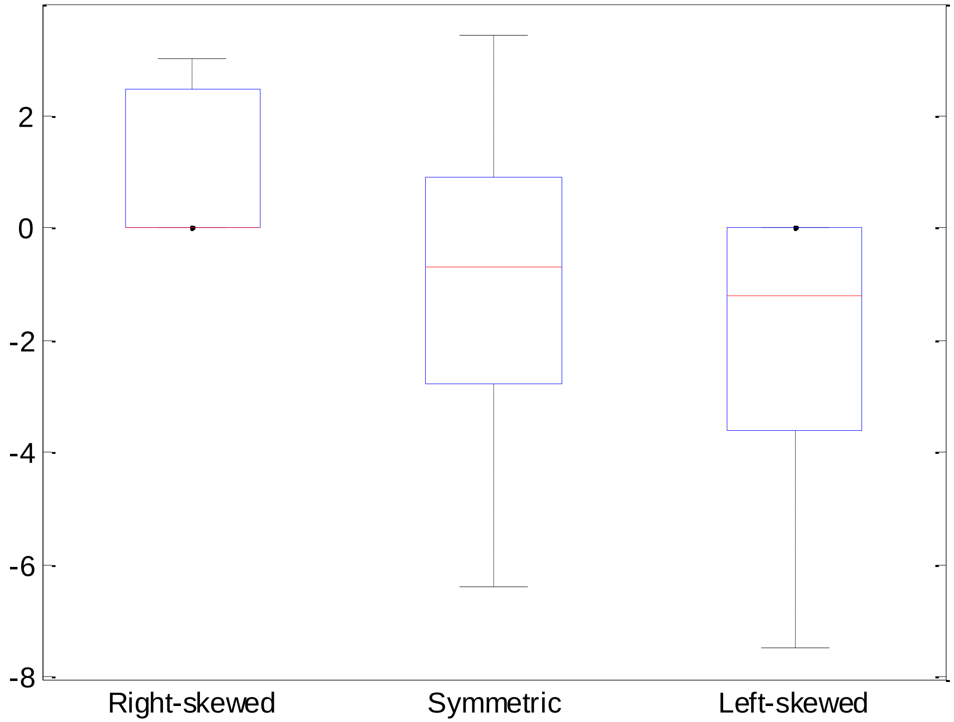
Skewness

- A measure of distribution skewness
  - Left-skewed: skewness < 0
  - Right-skewed: skewness > 0



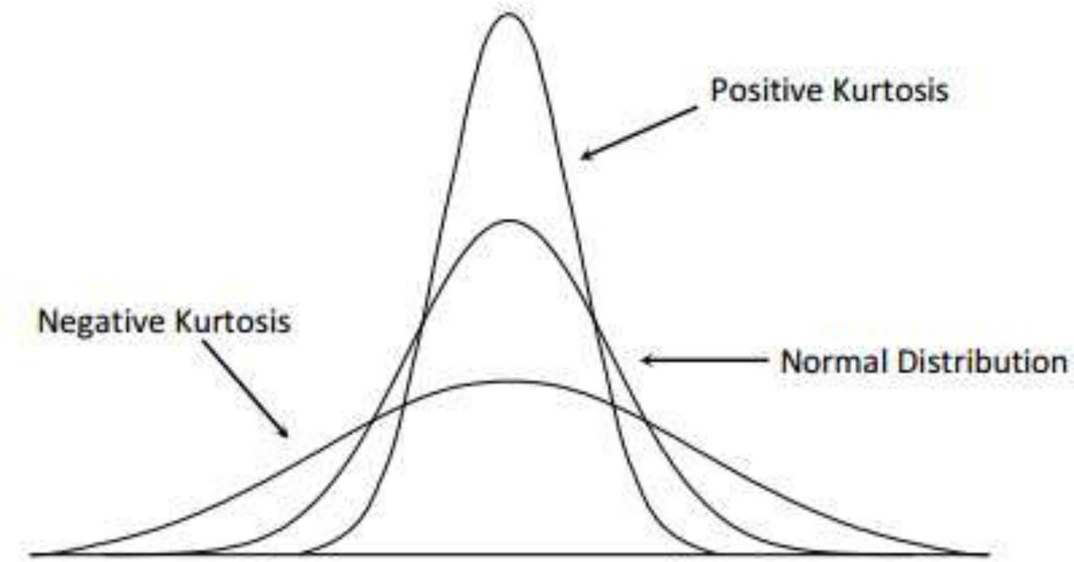
Skewness: `skewness()`

```
X = randn([10 3])*3;
X(X(:,1)<0, 1) = 0;  X(X(:,3)>0, 3) = 0;
boxplot(X, {'Right-skewed', 'Symmetric', 'Left-skewed'});
y = skewness(X)
```



Kurtosis

- A measure of distribution flatness
- A kurtosis of a normal distribution is zero
  - Positive Kurtosis: more acute peak
  - Negative Kurtosis: more flat peak

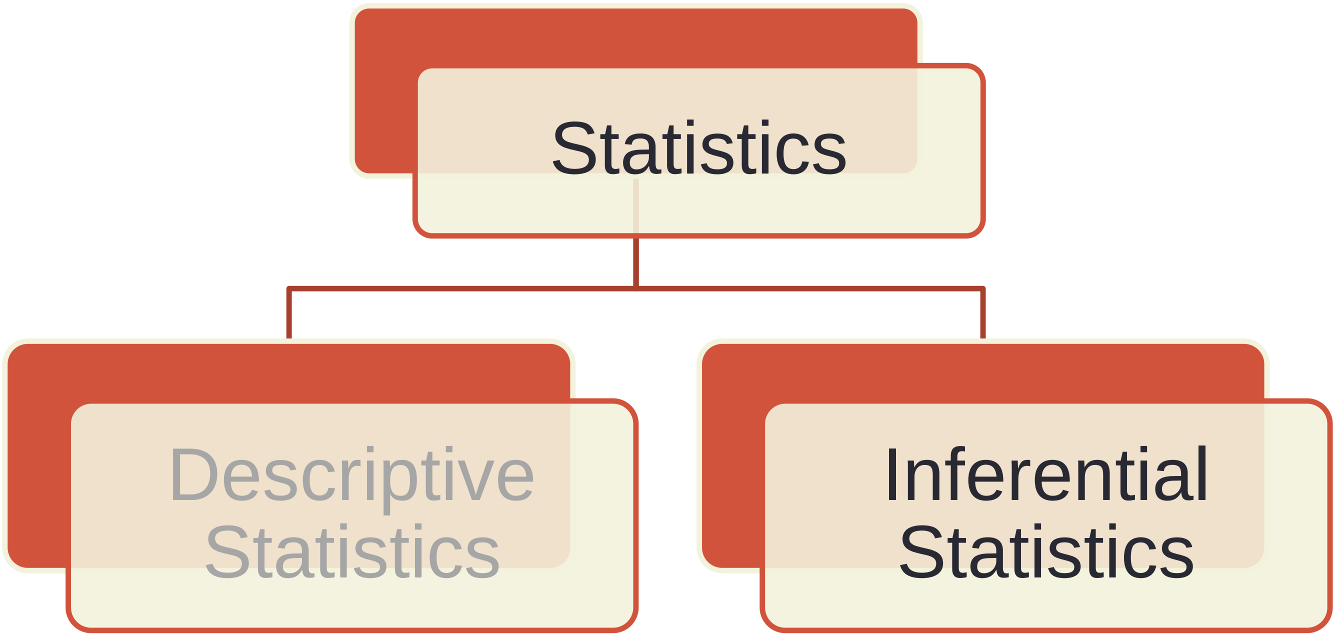


Exercise

- Find the skewness and kurtosis for each column of the variable `stocks`

```
load stockreturns;
```

Main Statistical Methodologies



Methods to make estimates, decisions, and predictions using sample data

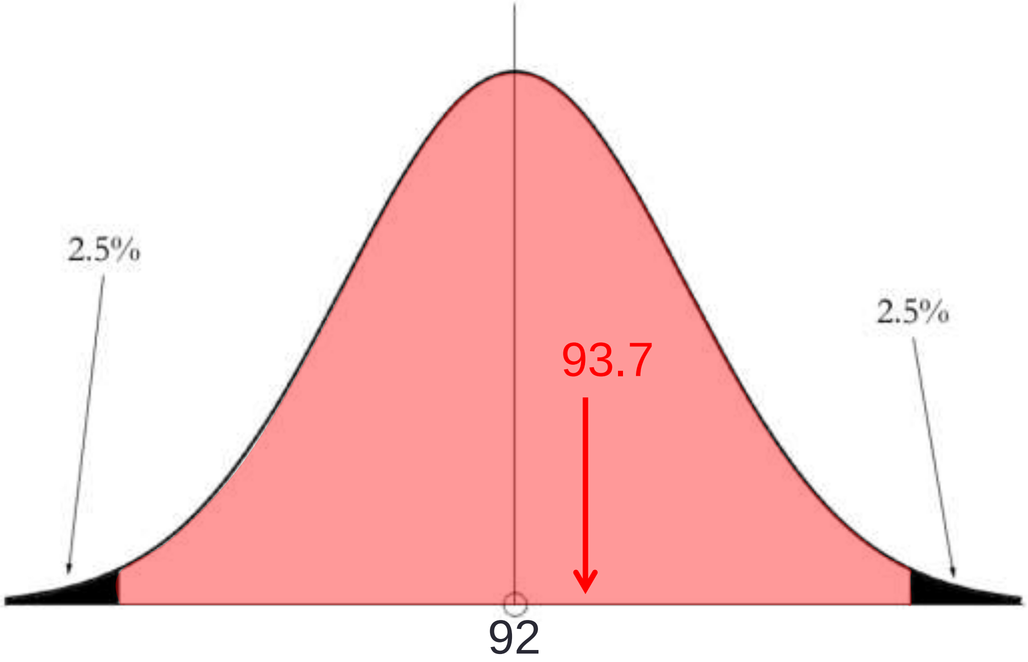
Statistical Hypothesis Testing

- A method of making decisions using data
- Example: Am I going to get grade A in this class?
- Typical hypothesis:
  - $H_0\colon \theta = \theta_0 \quad v.s. \quad H_1 : \theta \neq \theta_0$
  - $H_0\colon \theta \geq \theta_0 \quad v.s. \quad H_1 : \theta < \theta_0$
  - $H_0\colon \theta \leq \theta_0 \quad v.s. \quad H_1 : \theta > \theta_0$

where H_0 is null hypothesis, and H_1 is alternative hypothesis

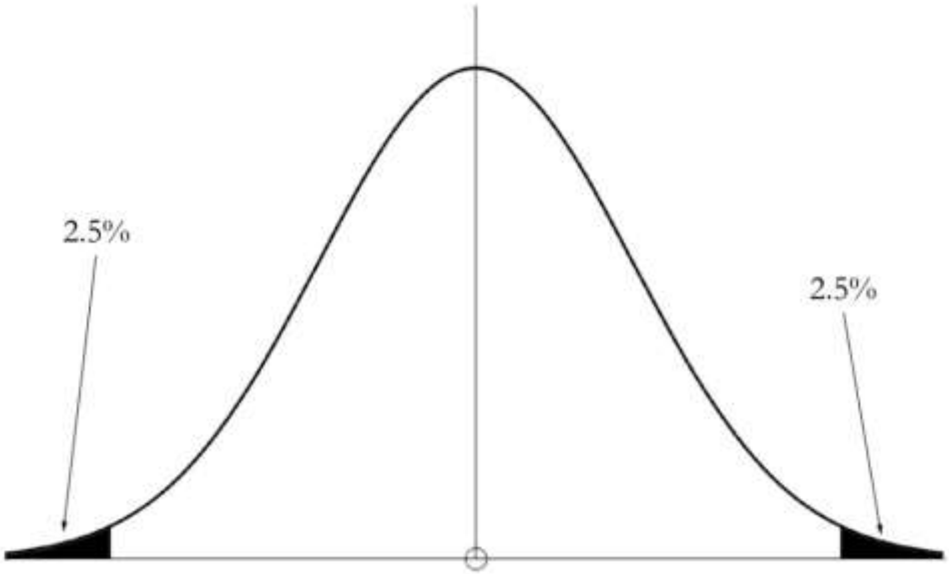
Hypothesis Testing Procedure

- Determine a probability, say 0.95, for the hypothesis test
- Find the 95% “confidence Interval” of the H_0
- Check if your score falls into the interval



Terminology in Hypothesis Testing

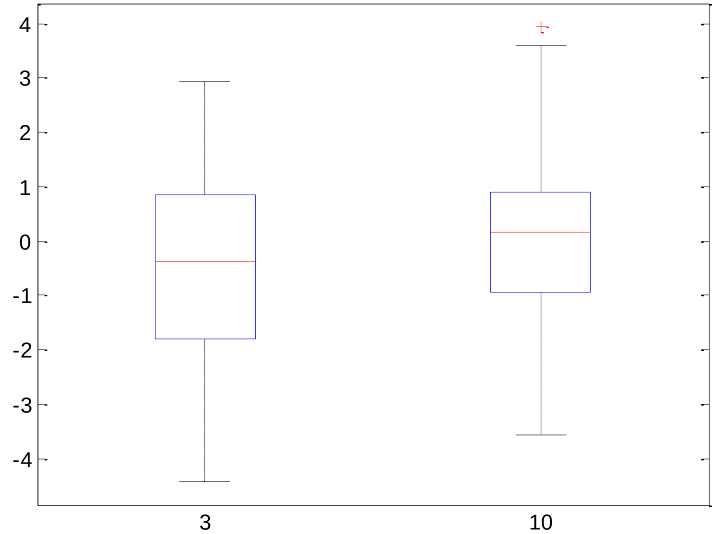
- Determine a probability, say 0.95, for the hypothesis test
- Find the 95% "confidence Interval" of the H_0
- Check if your score falls into the interval
- Terminology:
    - Confidence interval
    - Confidence level $(1-\alpha)$
    - Significance level α
    - p-value



t-test Example

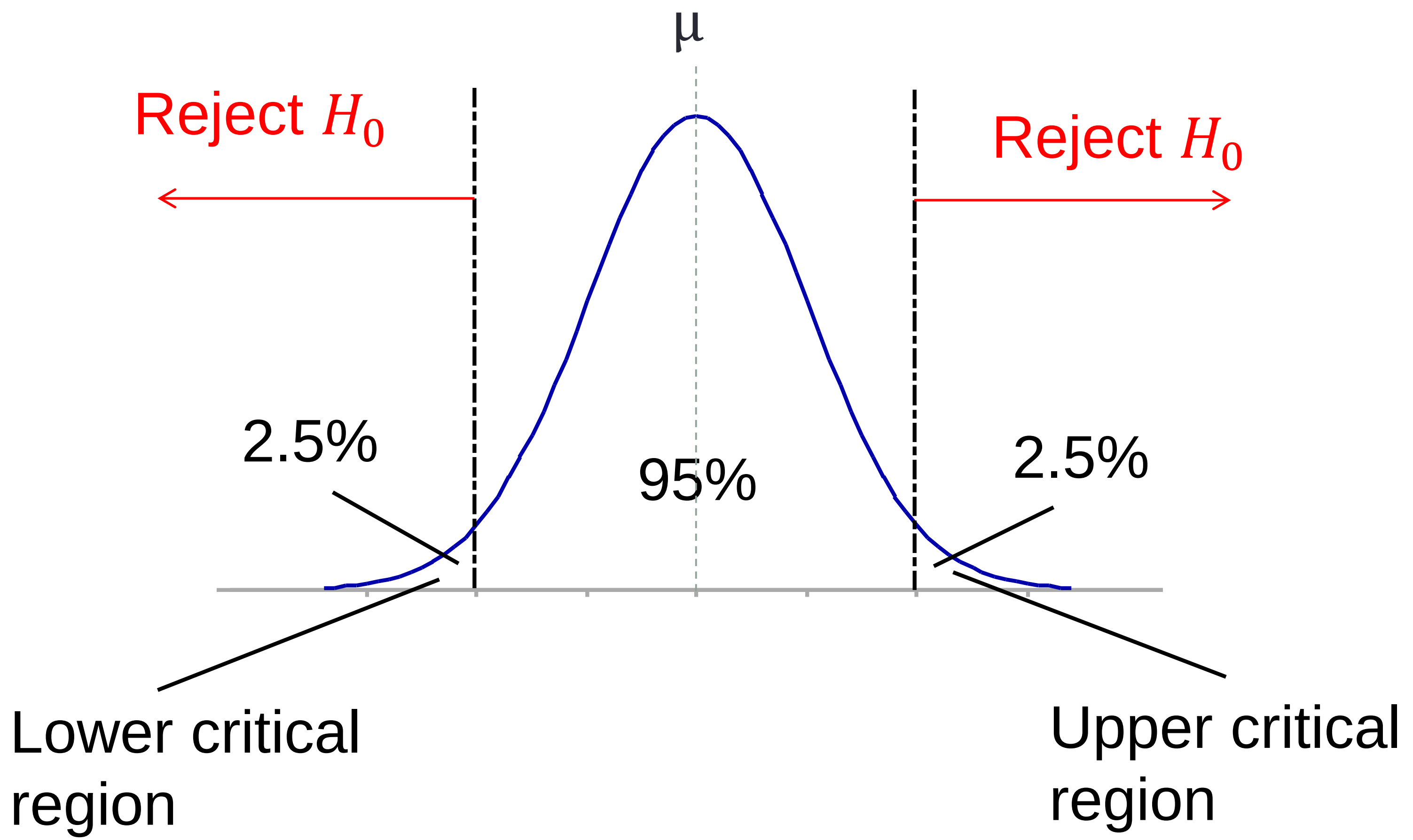
- Are means of the two stock returns (#3 and #10) the same?

```
load stockreturns;
x1 = stocks(:,3);  x2 = stocks(:,10);
boxplot([x1, x2], {'3', '10'});
[h,p] = ttest2(x1, x2)
```



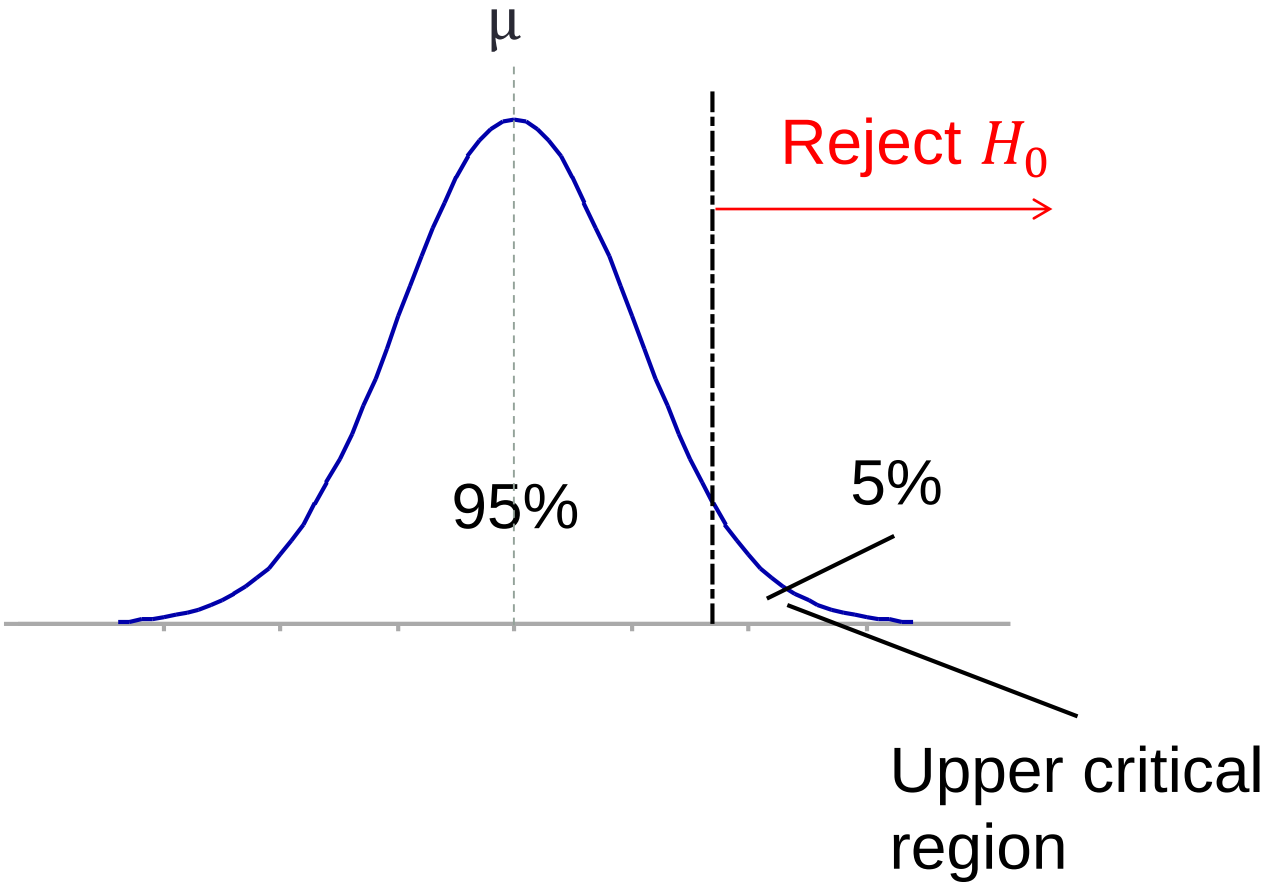
Two-tailed Significance Test

- Using a 5% significance level



One-tailed Significance Test

- Using a 5% significance level



Common Hypothesis Tests

	Paired data	Unpaired data	More than two groups
Parametric	• z-test • t-test	• two-sample t-test	• Analysis of variance (ANOVA)
Non-parametric	• Sign test • Wilcoxon signed-rank test	• Wilcoxon rank-sum test	

`ranksum()`	Wilcoxon rank sum test
`signrank()`	Wilcoxon signed rank test
`ttest()`	One-sample and paired-sample t-test
`ttest2()`	Two-sample t-test
`ztest()`	z-test

End of Class