

---

# Statistical Testing (and MS Excel)

© 2007 Duke Okes

---

## Testing Hypotheses

- **Statistical hypothesis testing is used to check for one of two conditions**
  - *Is there a difference between sets of data*
  - *Is there a relationship between sets of data*
- **The test does not prove that there is or is not, but allows one to know (and to set) the statistical probability on which the decision was made**

---

© 2007 Duke Okes

# Statistical Theory

---

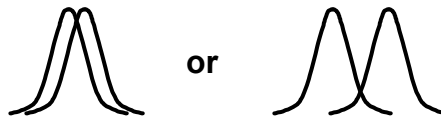
- Known average and std deviation of a population
- Take samples from the population and calculate average of the sample
- Distribution of sample averages is predictable
- Can establish confidence intervals around the population average
- Pull another sample, if it falls outside the confidence interval, the population may have changed

© 2007 Duke Okes

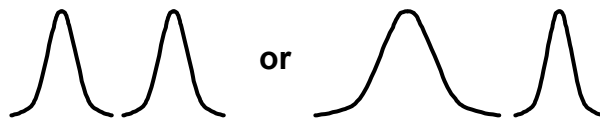
# Tests for Differences

---

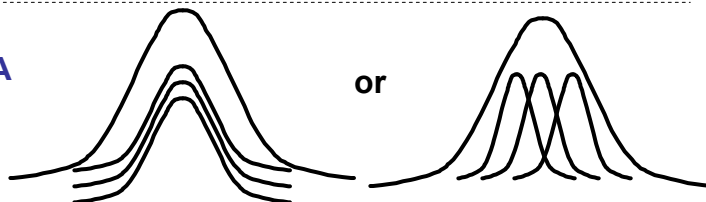
- z test
- t test



- F test



- ANOVA



© 2007 Duke Okes

## Using MS Excel for Statistical Tests

---

- If Data Analysis is not an option under Tools, click on Add-Ins, then select and install Analysis ToolPak
- Click on Tools, Data Analysis, then the type of test desired
- Indicate where the data is located
- Find the output and interpret it

*Minitab or SPSS are highly recommended for these as well as more advanced techniques.*

---

© 2007 Duke Okes

## Situation 1

---

- Two sets of data are taken from the same process, but with one variable (e.g., operator batch, raw material, or measurement device) different between the two sets
- Desire is to know whether there is a statistical difference due to that one variable

---

© 2007 Duke Okes

## Results 1a - F test

F-Test Two-Sample for Variances		
	<i>Variable 1</i>	<i>Variable 2</i>
Mean	226.375	215.375
Variance	13.98214286	9.696428571
Observations	8	8
df	7	7
F	1.44198895	
P(F<=f) one-tail	0.320598562	
F Critical one-tail	3.787050673	

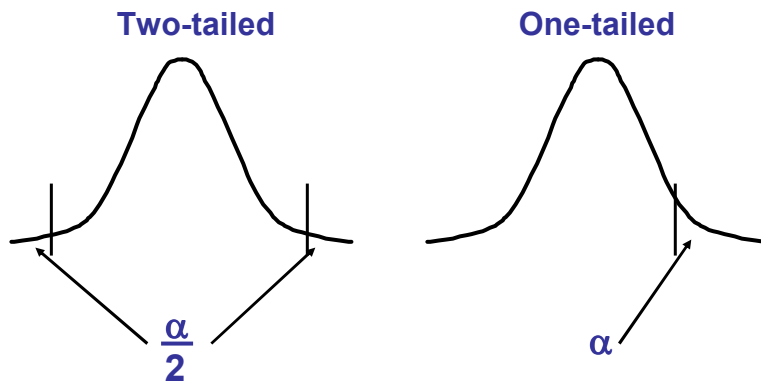
© 2007 Duke Okes

## Results 1b - t test

t-Test: Two-Sample Assuming Equal Variances		
	<i>Variable 1</i>	<i>Variable 2</i>
Mean	226.375	226.125
Variance	13.98214286	58.41071429
Observations	8	8
Pooled Variance	36.19642857	
Hypothesized Mean Difference	0	
df	14	
t Stat	0.083106912	
P(T<=t) one-tail	0.467471562	
t Critical one-tail	1.76130925	
P(T<=t) two-tail	0.934943124	
t Critical two-tail	2.144788596	

© 2007 Duke Okes

## One-tailed versus Two-tailed Test



© 2007 Duke Okes

## Results 1c - t test

t-Test: Two-Sample Assuming Equal Variances		
	Variable 1	Variable 2
Mean	226.375	215.375
Variance	13.98214286	9.696428571
Observations	8	8
Pooled Variance	11.83928571	
Hypothesized Mean Difference	0	
df	14	
t Stat	6.393812997	
P(T<=t) one-tail	8.35799E-06	
t Critical one-tail	1.76130925	
P(T<=t) two-tail	1.6716E-05	
t Critical two-tail	2.144788596	

© 2007 Duke Okes

## Situation 2

---

- A similar situation, but there are more than two sets of data (e.g., more than two operators, batches of material, or measurement devices were being evaluated)

© 2007 Duke Okes

## Situation 2 - ANOVA (1-way)

---

Anova: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
Column 1	8	1811	226.375	13.98214286		
Column 2	8	1809	226.125	58.41071429		
Column 3	8	1815	226.875	10.125		
Column 4	8	1793	224.125	52.69642857		
Column 5	8	1723	215.375	9.696428571		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	740.6	4	185.15	6.388416513	0.00057179	2.641463936
Within Groups	1014.375	35	28.98214286			
Total	1754.975	39				

© 2007 Duke Okes

## Situation 3 - Correlation

---

- Have paired data, and want to know if there is any relationship between the two variables

	<i>Column 1</i>	<i>Column 2</i>
<i>Column 1</i>	1	
<i>Column 2</i>	-0.95225	1

© 2007 Duke Okes

## Situation 4 - $\chi^2$ test (1-way)

---

- Is the production out increasing during the week on a new machine?

	<u>Machine Output (1st week)</u>			
	<u>Day1</u>	<u>Day2</u>	<u>Day3</u>	<u>Day4</u>
Observed -	49	53	65	73
Expected -	60	60	60	60
(O-E) <sup>2</sup> /E	2.02	.82	.42	2.82

Total  $\chi^2 = 6.08$ , df = 3

Critical  $\chi^2$  (from table) = 7.815

© 2007 Duke Okes

## Situation 5 - $\chi^2$ test (2-way)

---

Use  $\chi^2$  to analyze a cross-tabulated table

	Machine 1	Machine 2	Totals
Shift 1	323	401	724
Shift 2	375	368	743
Shift 3	295	435	730
Totals	993	1204	2197

Calculated  $\chi^2 = 15.05$ ,  $df = 2$

Critical value of  $\chi^2 = 12.59$

---

© 2007 Duke Okes

## Some Cautions

---

- z, t, F, and ANOVA tests are for interval data (as opposed to ordinal or nominal), and assume that they are normally distributed
- for  $\chi^2$  tests using frequencies, Expected values should be  $>5$  in each cell. Combine rows or columns if possible. In a 2x2 table if Expected  $<10$  use Yates correction (reduce O-E for that cell by .5)
- It is assumed that the data are randomly selected and that sample groups (factors) are independent

---

© 2007 Duke Okes

## **Statistical versus Practical Significance**

---

**Just because something is statistically significant doesn't mean that action should be taken in the direction indicated by the test.**

**Other factors, such as the need of the application, economics, safety, etc. should be taken into account.**

**And remember, significance only says the probability is in our favor ... it doesn't say that it's actually true.**

**Never allow statistics to overrule common sense!**

---

© 2007 Duke Okes

## **Instructor Contact Info**

---

**Duke Okes**

**423-323-7576**

**dokes@earthlink.net**

**www.aplomet.com**

---

© 2007 Duke Okes