

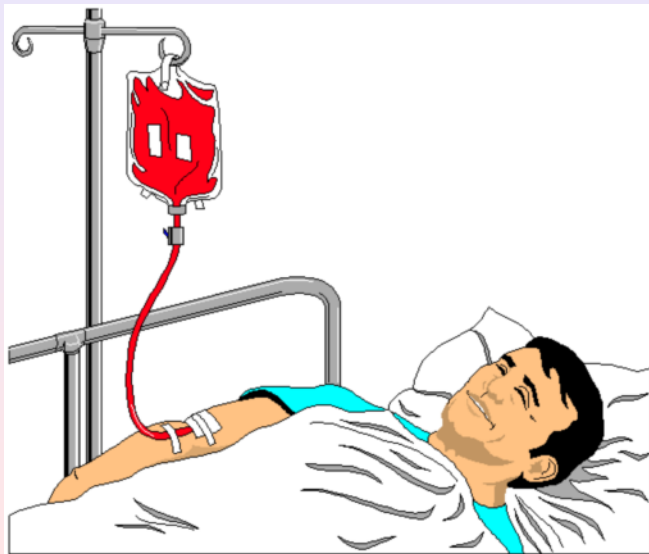
# Correlation and Regression

Dale Easley  
University of Dubuque



# Correlation does not imply causation

Towns with more doctors have more reported cases of disease.



# Independent and dependent variables

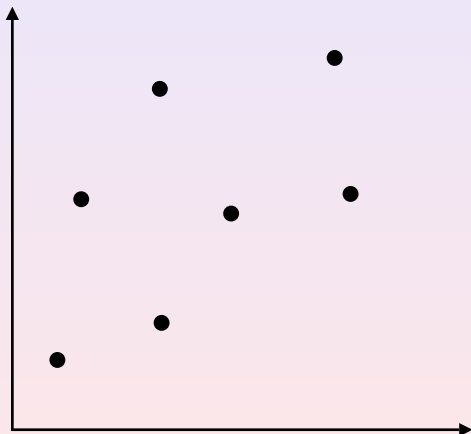
pH	germinated
6	0
6	3
6	4
6.5	1
6.5	2
6.5	4
7	2
7	4
7	6
7.5	5
7.5	6
7.5	7

Experiment: For a variety of pH values, count the number of plants germinating.

1. Design a lab experiment that gives the results shown.
  - Which is your independent variable? Why?
2. Design a field experiment that gives the results shown.
  - With what might your variables be autocorrelated?
3. Graph this with Excel.

# Pearson's product moment correlation coefficient ( $r$ )

1. Paired data
2. Measure of correlation
3. Assumes data is normal about line fit
4. Impacted by number of data
5. Degrees of freedom =  $n-2$



Try fitting a line to the data on the preceding slide.

# Is the r value significant?

	A	B	C	D	E	F	G	H	I	J
1	pH	Num								
2	6	0								
3	6	3								
4	6	4								
5	6.5	1								
6	6.5	2								
7	6.5	4								
8	7	2								
9	7	4								
10	7	6								
11	7.5	5								
12	7.5	6								
13	7.5	7								
14										
15										
16										
17										
18										
19										
20										
21										

### Regression

Input

Input Y Range:

Input X Range:

Labels  Constant is Zero

Confidence Level:  %

Output options

Output Range:

New Worksheet Ply:

New Workbook

Residuals

Residuals  Residual Plots

Standardized Residuals  Line Fit Plots

Normal Probability

Normal Probability Plots

OK Cancel Help

# Excel Output

## SUMMARY OUTPUT

### Regression Statistics

Multiple R	0.689202
R Square	0.475
Adjusted R	0.4225
Standard Error	1.630951
Observations	12

### ANOVA

	df	SS	MS	F	Significance F
Regressor	1	24.06667	24.06667	9.047619	0.013164
Residual	10	26.6	2.66		
Total	11	50.66667			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-13.43333	5.704443	-2.35489	0.040303	-26.1436	-0.72304	-26.1436	-0.72304
pH	2.533333	0.842219	3.007926	0.013164	0.656752	4.409915	0.656752	4.409915

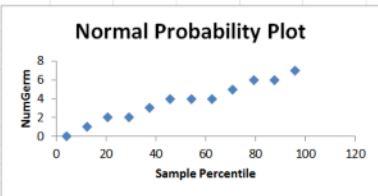
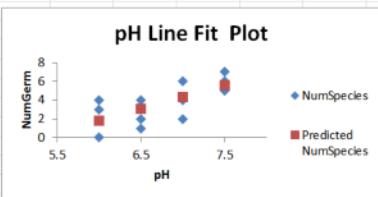
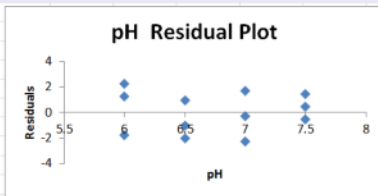


## RESIDUAL OUTPUT

Observation	redicted NumGerm	Residuals	Standard Residuals
1	1.766667	-1.76667	-1.13608
2	1.766667	1.233333	0.793115
3	1.766667	2.233333	1.436181
4	3.033333	-2.03333	-1.30757
5	3.033333	-1.03333	-0.6645
6	3.033333	0.966667	0.62163
7	4.3	-2.3	-1.47905
8	4.3	-0.3	-0.19292

## PROBABILITY OUTPUT

Percentile	NumGerm
4.166667	0
12.5	1
20.83333	2
29.16667	2
37.5	3
45.83333	4
54.16667	4
62.5	4



Now try the data in Worked Example 5.1

