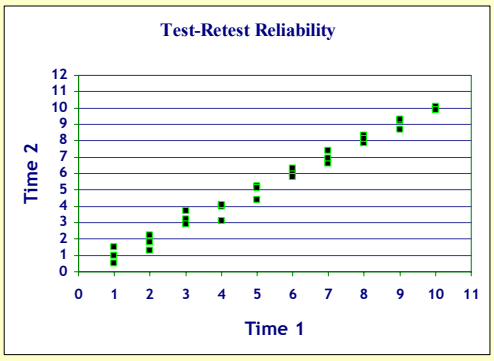
Psychological Statistics

Correlation and Regression

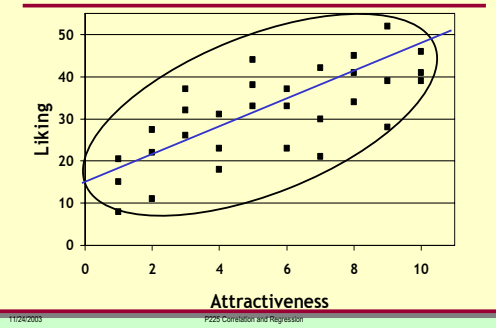
Correlation

What We Will Cover

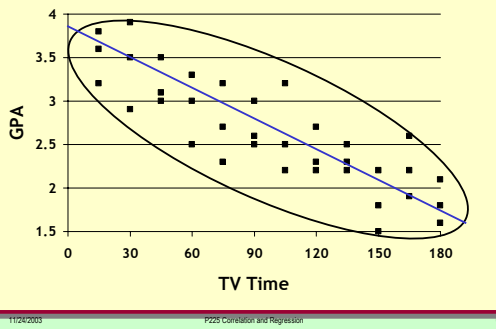
- What a correlation coefficient is.
- Assumptions about correlation calculations.
- How to calculate a correlation.
- Interpretation of the correlation.



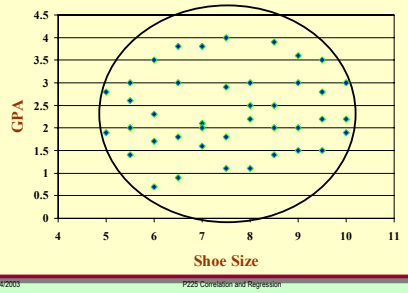
Positive Correlation



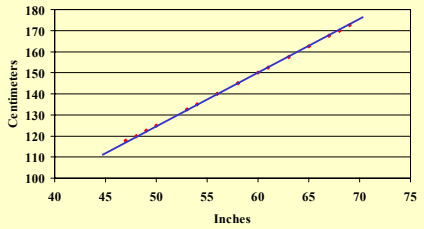
Negative Correlation

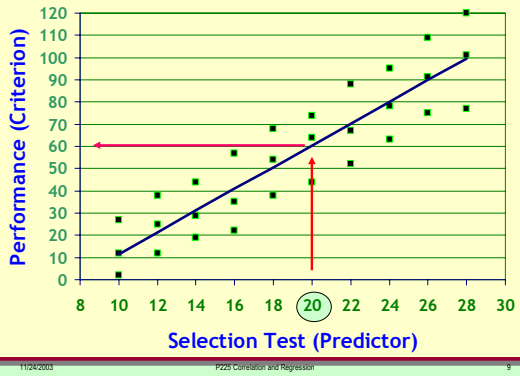


Zero Correlation Example

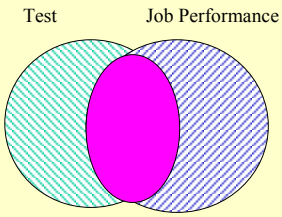


Perfect Relationship





Visual Interpretation



The Correlation Coefficient

- Indicated by r.
- Ranges from –1.00 to +1.00
  - The number indicates the strength of the relationship.
  - The closer to 0 the weaker the relationship.
  - The closer to 1.00 the stronger the relationship.
- The sign indicates whether the relationship is positive or negative.

QUESTION

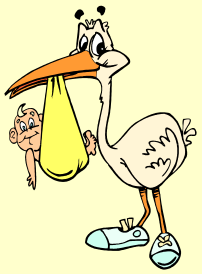
- Which of the following correlation coefficients indicates the strongest relationship?

A. +.45
B. -.33
C. +.58
D. -.67

Cannot Conclude Causality

Research Example

While conducting research in Helsinki, Finland a demographer found that the correlation between the number of stork nests on chimneys was positively correlated (r = .38) with birth rate.



CONCLUSION

A. Storks bring babies.
B. Male storks make babies in unfaithful human females.
C. Babies make storks.
D. I haven't the slightest idea.

	Height	Hand	Zheight	Zhand	
	70	9.25	1.59	1.81	2.88
	65	8	-0.02	0.32	-0.01
	66	7.75	0.30	0.02	0.01
	70	8.25	1.59	0.62	0.98
	63	8	-0.66	0.32	-0.21
	62	6	-0.98	-2.08	2.04
	62.5	7.75	-0.82	0.02	-0.01
	61	7	-1.30	-0.88	1.15
	66	7.625	0.30	-0.13	-0.04
Mean	65.06	7.74			0.75
SD	3.11	0.83			

Defining Formula: Correlation Coefficient

$$r_{xy} = \frac{\sum (Z_x \, x \, Z_y)}{N}$$

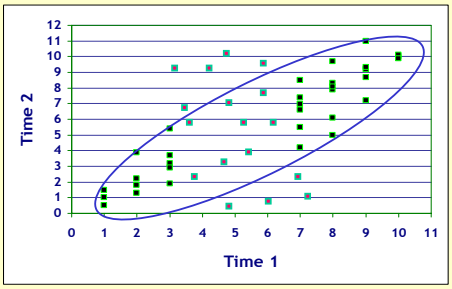
$$r = \frac{N(\sum_{i=1}^{n} X_i Y_i) - (\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y)_i}{\sqrt{N\left(\sum_{i=1}^{n} X_i^2\right) - \left(\sum_{i=1}^{n} X_i\right)^2 \, x \, N\left(\sum_{i=1}^{n} Y_i^2\right) - \left(\sum_{i=1}^{n} Y_i\right)^2}}$$

An equation you don't need to know.

Correlation Assumptions

1. The relationship is linear.
2. The variables are measured on a interval or ratio scale.
3. The distribution of the variables in normal.
4. The distribution is continuous.

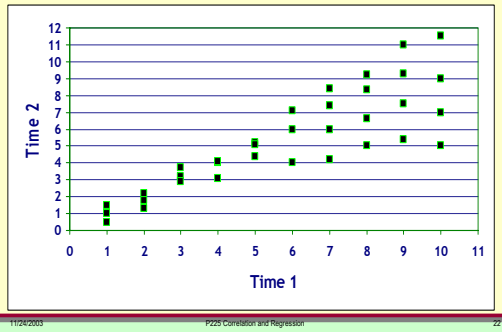
Non-continuous Distribution



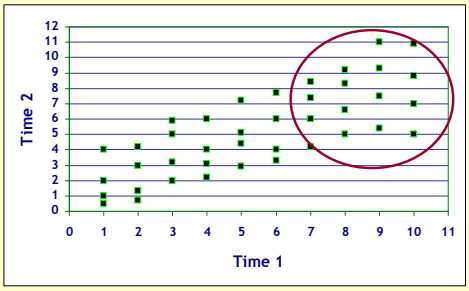
Assumption #5

- HOMOSCEDASTICITY
  - The variance at each data point is equal.
- HETEROSCEDASTICITY
  - The variance at each data point is not equal.

Violation of Homoscedasticity



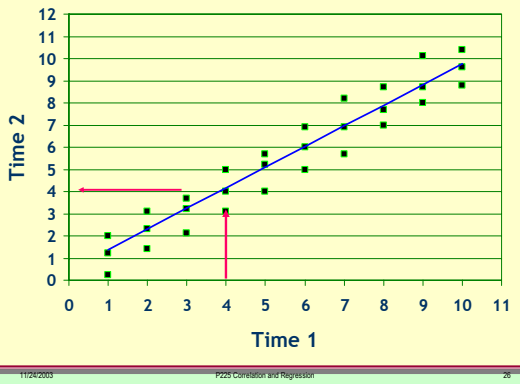
Assumption #6 Restriction of Range



Statistical Interpretation

r_{xy}^2	Coefficient of Determination	Percent variance that is shared between variable x and variable y.
$1 - r_{xy}^2$	Coefficient of Alienation	Percent variance that is NOT shared between variable x and variable y.

Regression



Regression Equation

$$Z_{y'} = r_{xy} \; x \; Z_x$$

Predicted Z-score for Y.

Correlation between X and Y

Z- score for X.

Regression Example

- The relationship between number of green M&Ms eaten a day and sex appeal is r = .58.
  - Sex Appeal
    - M = 22
    - S = 4
  - M&Ms
    - M = 12.3
    - S = 3.1
- What is the predicted sex appeal for a person who eats 20 M&Ms?

Calculations

1. Convert number of M&Ms to z-score.

$$Z_X = \frac{20 - 12.3}{3.1} = 2.48$$

2. Multiply this z score by the correlation coefficient to get the predicted z-score for Sex Appeal.

$$Z'_Y = 2.48 \; x \; .58 = 1.438$$

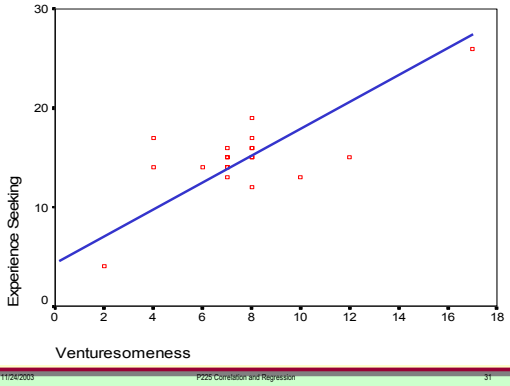
3. Convert the Z-score for Y into the raw score.

$$Y' = (1.438 \text{ x } 4) + 22 = 27.75$$

Research Example

Descriptive Statistics

	Mean	Std. Deviation	N
Experience Seeking	15.0000	3.9068	20
Venturesomeness	7.6500	3.0483	20
Sociability	6.1000	2.2455	20



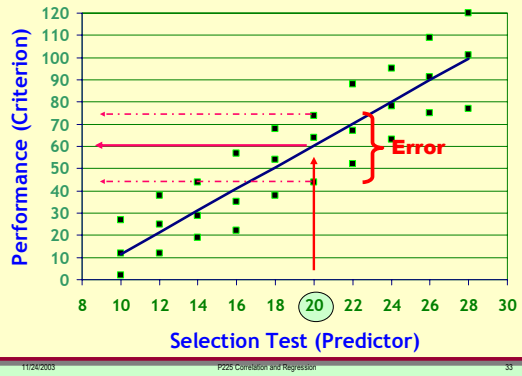
Correlation Matrix

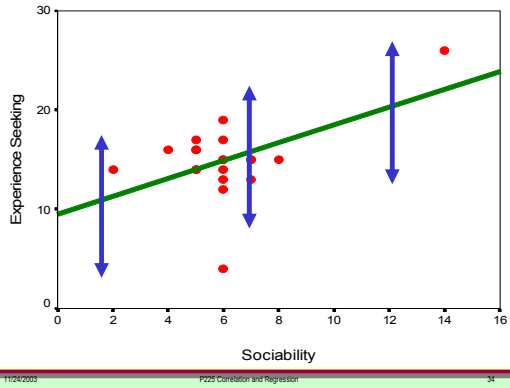
Correlations

		Experience Seeking	Ventures omeness	Sociability
Experience Seeking	Pearson Correlation	1.000	.716**	.516*
	Sig. (2-tailed)	.	.000	.020
	N	20	20	20
Venturesomeness	Pearson Correlation	.716**	1.000	.774**
	Sig. (2-tailed)	.000	.	.000
	N	20	20	20
Sociability	Pearson Correlation	.516*	.774**	1.000
	Sig. (2-tailed)	.020	.000	.
	N	20	20	20

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).





Errors in Prediction

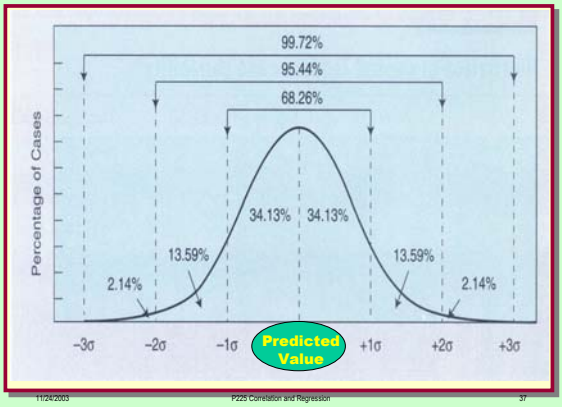
- Assume.
  - The errors are normally distributed.
  - Homoscedasticity.
    - The variance of the errors is equal at all points along the regression line.

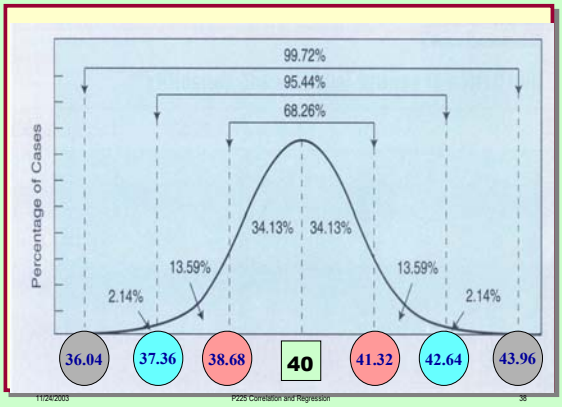
Standard Error of Estimate

$$SEst_y = S_y\sqrt{1 - r^2_{xy}}$$

Interpretation

This is the standard deviation of the error in predicting Y from X.


99.72%
95.44%
68.26%
Percentage of Cases
34.13%
34.13%
13.59%
13.59%
2.14%
2.14%
−3σ
−2σ
−1σ
Predicted Value
+1σ
+2σ
+3σ


99.72%
95.44%
68.26%
Percentage of Cases
34.13%
34.13%
13.59%
13.59%
2.14%
2.14%
36.04
37.36
38.68
40
41.32
42.64
43.96


THE END!