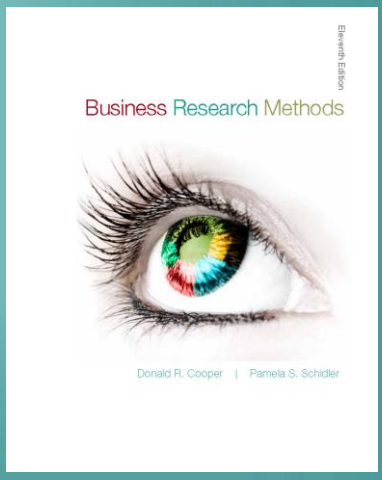
Chapter 18

Measures of Association



Learning Objectives

Understand . . .

- *How correlation analysis may be applied to study relationships between two or more variables*
- *The uses, requirements, and interpretation of the product moment correlation coefficient.*
- *How predictions are made with regression analysis using the method of least squares to minimize errors in drawing a line of best fit.*

Learning Objectives

Understand . . .

- *How to test regression models for linearity and whether the equation is effective in fitting the data.*
- *Nonparametric measures of association and the alternatives they offer when key assumptions and requirements for parametric techniques cannot be met.*

Invalid Assumptions

"The invalid assumption that correlation implies cause is probably among the two or three most serious and common errors of human reasoning."

Stephen Jay Gould
paleontologist and science writer

PulsePoint: Research Revelation

25

The percent of students using a credit card for college costs due to convenience.

Measures of Association: Interval/Ratio Data

Pearson correlation coefficient	*For continuous linearly related variables*
Correlation ratio (eta)	*For nonlinear data or relating a main effect to a continuous dependent variable*
Biserial	*One continuous and one dichotomous variable with an underlying normal distribution*
Partial correlation	*Three variables; relating two with the third's effect taken out*
Multiple correlation	*Three variables; relating one variable with two others*
Bivariate linear regression	*Predicting one variable from another's scores*

Measures of Association: Ordinal Data

Gamma	*Based on concordant-discordant pairs; proportional reduction in error (PRE) interpretation*
Kendall's tau b	*P-Q based; adjustment for tied ranks*
Kendall's tau c	*P-Q based; adjustment for table dimensions*
Somers's d	*P-Q based; asymmetrical extension of gamma*
Spearman's rho	*Product moment correlation for ranked data*

18-7

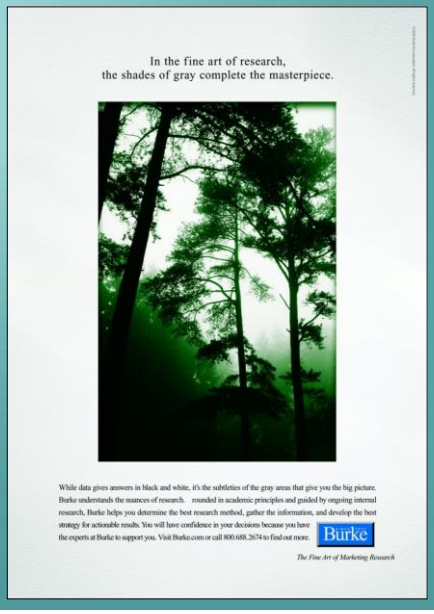
Measures of Association: Nominal Data

Phi	*Chi-square based for 2*2 tables*
Cramer's V	*CS based; adjustment when one table dimension >2*
Contingency coefficient C	*CS based; flexible data and distribution assumptions*
Lambda	*PRE based interpretation*
Goodman & Kruskal's tau	*PRE based with table marginals emphasis*
Uncertainty coefficient	*Useful for multidimensional tables*
Kappa	*Agreement measure*

18-8

Researchers Search for Insights

Burke, one of the world's leading research companies, claims researchers add the most value to a project when they look beyond the raw numbers to the shades of gray…what the data really mean.



Pearson's Product Moment Correlation *r*

Is there a relationship between X and Y?

What is the magnitude of the relationship?

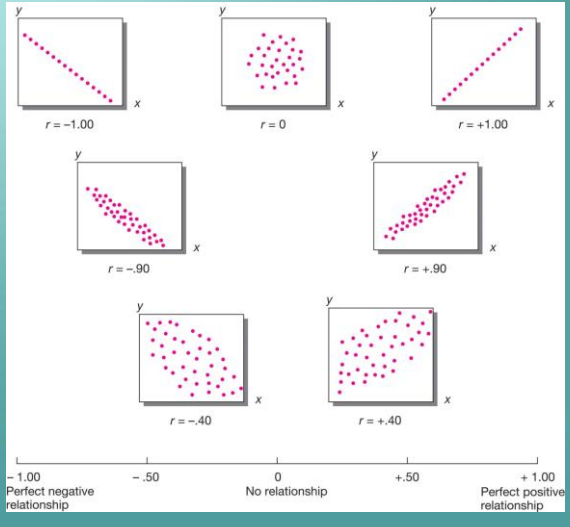
What is the direction of the relationship?

Connections and Disconnections

"To truly understand consumers' motives and actions, you must determine relationships between what they think and feel and what they actually do."

David Singleton, vp of insights
Zyman Marketing Group

Scatterplots of Relationships



Scatterplots

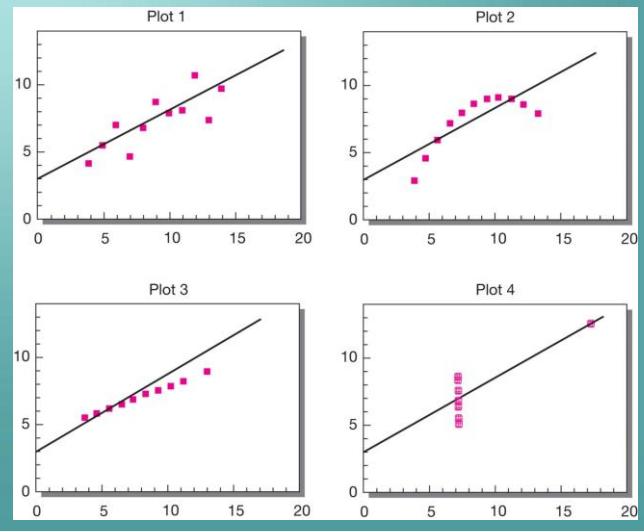
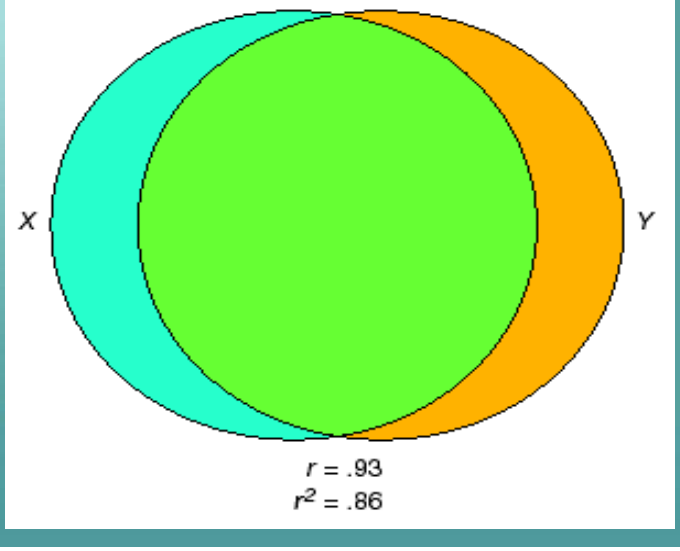


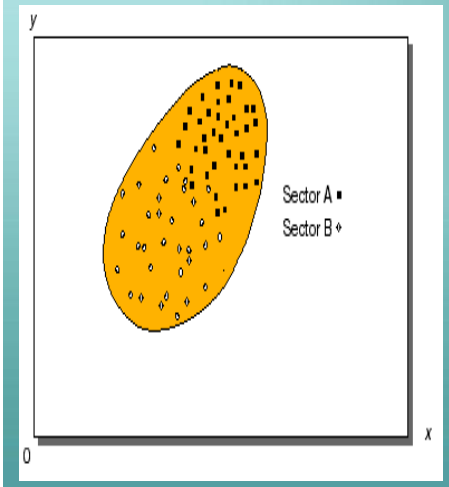
Diagram of Common Variance

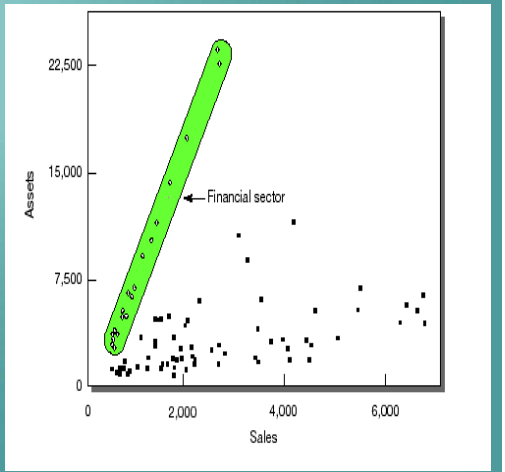


Interpretation of Correlations

- X causes Y
- Y causes X
- X and Y are activated by one or more other variables
- X and Y influence each other reciprocally

Artifact Correlations





Interpretation of Coefficients

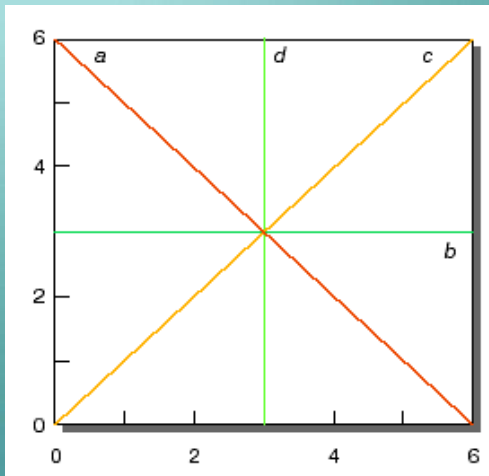
A coefficient is not remarkable simply because it is statistically significant!

It must be practically meaningful.

Comparison of Bivariate Linear Correlation and Regression

	Correlation	Regression
Measurement level	Interval or ratio scale	Interval or ratio scale
Nature of variables	Both continuous, linearly related	Both continuous, linearly related
$X - Y$ relationship	X and Y are symmetric; $r_{xy} = r_{yx}$	Y is dependent, X is independent; regression of X on Y differs from Y on X
Correlation	The correlation of x and y produces an estimate of linear association based on sampling data	Correlation of $Y - X$ is the same as the correlation between the predicted values of Y and observed values of Y
Coefficient of determination	Explains common variance of X and Y	Proportion of variability of Y explained by its least-squares regression on X

Examples of Different Slopes

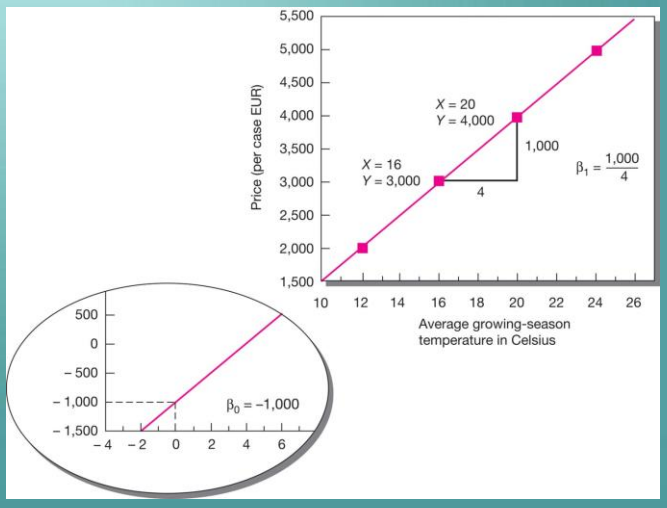


Line	Slope
a	−1
b	0
c	+1
d	∞

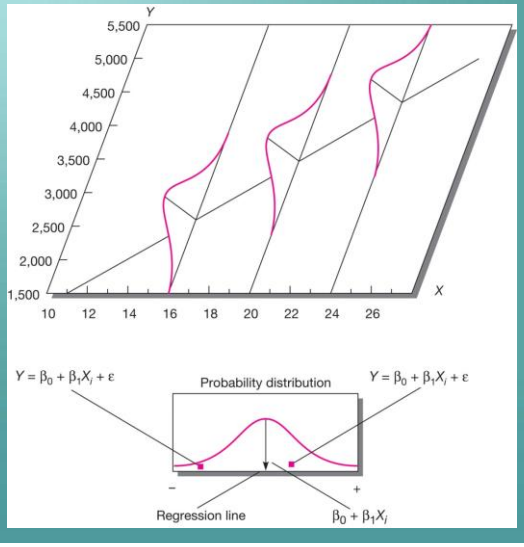
Concept Application

X *Average Temperature (Celsius)*	*Y* *Price per Case (FF)*
12	*2,000*
16	*3,000*
20	*4,000*
24	*5,000*
Mean =18	*Mean = 3,500*

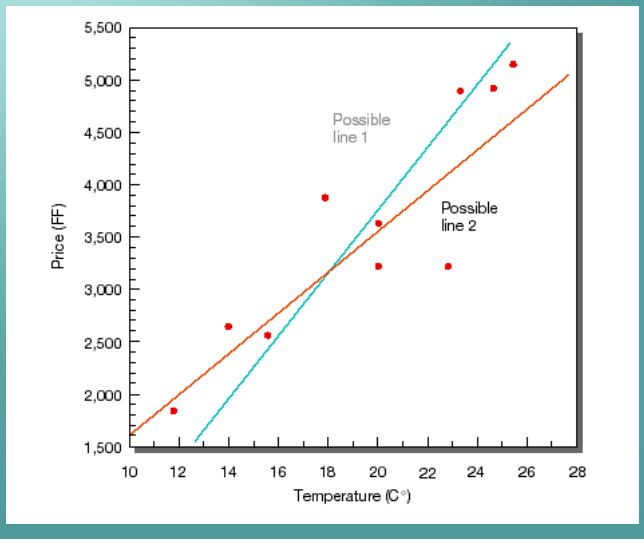
Plot of Wine Price by Average Temperature



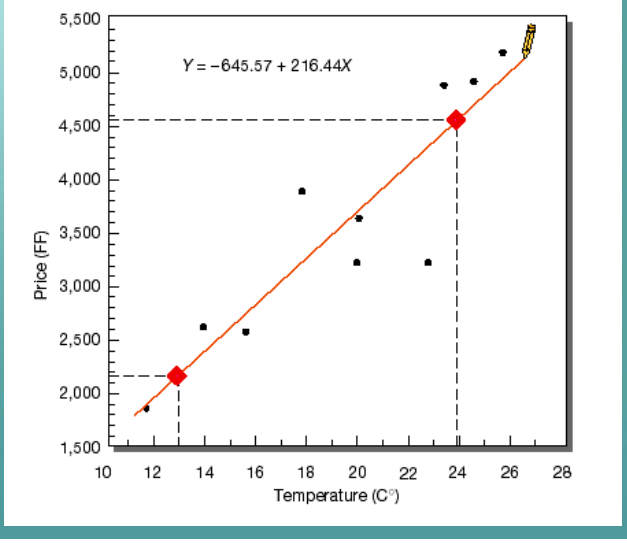
Distribution of Y for Observation of X



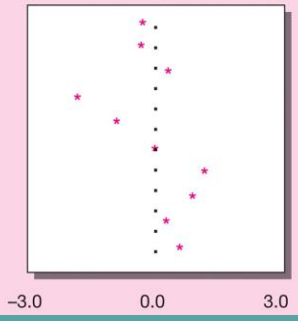
Wine Price Study Example



Least Squares Line: Wine Price Study

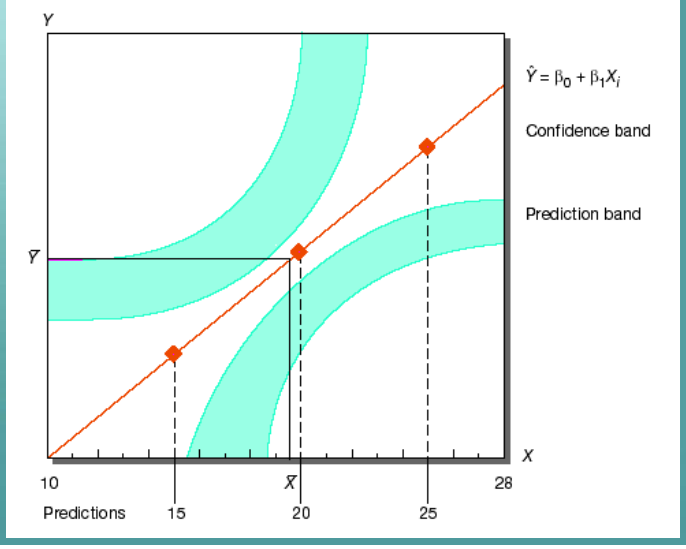


Plot of Standardized Residuals



Case	Y Price	Predicted Price	Residual
1	1,813	1,908.4112	−95.4112
2	2,558	2,752.5234	−194.5234
3	2,628	2,384.5771	243.4229
4	3,217	4,310.8844	−1,093.8844
5	3,228	3,683.2112	−455.2112
6	3,629	3,704.8551	−75.8551
7	3,886	3,228.6893	657.3107
8	4,897	4,419.1039	477.8961
9	4,933	4,678.8307	254.1693
10	5,199	4,916.9137	282.0863

Prediction and Confidence Bands



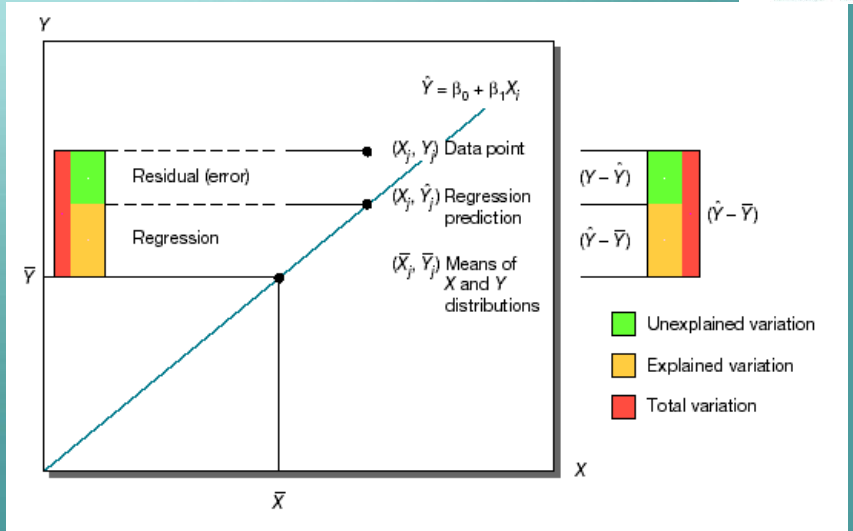
Testing Goodness of Fit

Y is completely unrelated to X
and no systematic pattern is evident

There are constant values of
Y for every value of X

The data are related but
represented by a nonlinear function

Components of Variation



F Ratio in Regression

ANOVA Summary Table: Test of Regression Model

Source	Degrees of Freedom	Sum of Squares	Mean Square	*F* Ratio
Regression	1	9,287,143.11	9,287,143.11	32.02
Residual (error)	8	2,320,368.49	290,046.06	
Total		11,607,511.60		

Significance of $F = .0005$

18-29

Coefficient of Determination: r^2

Total proportion of variance in Y explained by X

Desired r^2: 80% or more

18-30

Chi-Square Based Measures

Marketing Campaign Success

Count		Yes	No	Row Total
Direct Mail	Yes	21	10	31
	No	13	22	35
	Column Total	34	32	66

Chi-Square	*Value*	*d.f.*	*Significance*
Pearson	6.16257	1	.01305
Continuity correction	4.99836	1	.02537

Minimal expected frequency 15.030

Statistic	*Value*	*Approximate Significance*
Phi	.30557	.01305*
Cramer's *V*	.30557	.01305*
Contingency coefficient *C*	.29223	.01305*

*Pearson chi-square probability.

Proportional Reduction of Error Measures

What is your opinion about capping executives' salaries?

	Cell designation Count Row Pct.	Favor	Do Not Favor	Row Total
Occupational Class	Managerial	1,1 90 82.0	1,2 20 18.0	110
	White collar	2,1 60 43.0	2,2 80 57.0	140
	Blue collar	3,1 30 20.0	3,2 120 80.0	150
	Column Total	180 45.0%	220 55.0%	400 100.0%

Chi-Square	*Value*	*d.f.*	*Significance*
Pearson	98.38646	2	.00000
Likelihood ratio	104.96542	2	.00000

Minimum expected frequency 49.500

Statistic	*Value*	*ASEI*	*T Value*	*Approximate Significance*
Lambda:				
Symmetric	.30233	.03955	6.77902	
With occupation dependent	.24000	.03820	5.69495	
With opinion dependent	.38889	.04555	7.08010	
Goodman & Kruskal tau:				
With occupation dependent	.11669	.02076		.00000*
With opinion dependent	.24597	.03979		.00000*

*Based on chi-square approximation.

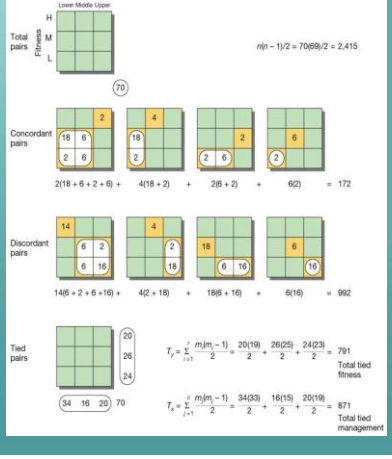
Statistical Alternatives for Ordinal Measures

		Management Level			
	Count	Lower	Middle	Upper	
Fitness	High	14	4	2	20
	Moderate	18	6	2	26
	Low	2	6	16	24
		34	16	20	70

Statistic	*Value**
Gamma	–.70
Kendall's tau *b*	–.51
Kendall's tau *c*	–.50
Somers's *d*	
Symmetric	–.51
With fitness dependent	–.53
With management-level dependent	–.50

*The *t* value for each coefficient is –5.86451.

Calculation of Concordant (*P*), Discordant (*Q*), Tied (*Tx*,*Ty*), and Total Paired Observations: KeyDesign Example



KDL Data for Spearman's Rho

	Rank By			
Applicant	Panel x	Psychologist y	d	d^2
1	3.5	6.0	-2.5	6.25
2	10.0	5.0	5.0	25.00
3	6.5	8.0	-1.5	2.52
4	2.0	1.5	.05	0.25
5	1.0	3.0	-2	4.00
6	9.0	7.0	2.0	4.00
7	3.5	1.5	2.0	4.00
8	6.5	9.0	-2.5	6.25
9	8.0	10.0	-2	4.00
10	5.0	4.0	1.0	1.00
				57.00

18-35

Key Terms

- *Artifact correlations*
- *Bivariate correlation analysis*
- *Bivariate normal distribution*
- *Chi-square-based measures*
- *Contingency coefficient C*
- *Cramer's V*
- *Phi*
- *Coefficient of determination (r2)*
- *Concordant*
- *Correlation matrix*
- *Discordant*
- *Error term*
- *Goodness of fit*
- *lambda*

18-36

Key Terms

- *Linearity*
- *Method of least squares*
- *Ordinal measures*
- *Gamma*
- *Somers's d*
- *Spearman's rho*
- *tau b*
- *tau c*

- *Pearson correlation coefficient*
- *Prediction and confidence bands*
- *Proportional reduction in error (PRE)*
- *Regression analysis*
- *Regression coefficients*

Key Terms

- *Intercept*
- *Slope*
- *Residual*

- *Scatterplot*
- *Simple prediction*
- *tau*