

Association: Contingency, Correlation, and Regression

3

3.2 The Association between Two Quantitative Variables

Aug 8-9:21 AM

Learning Objectives:

- Constructing scatterplots
- Interpreting a scatterplot
- Correlation
- Calculating correlation

Aug 8-9:21 AM

Internet Usage & GDP Data Set

INTERNET	GDP	INTERNET	GDP
Algeria	0.05	0.98	38.42
Argentina	10.08	11.32	27.31
Australia	37.14	25.37	5.62
Austria	39.7	20.72	48.05
Belgium	31.04	25.52	45.12
Brazil	4.88	7.36	0.1
Canada	46.95	27.15	46.38
Chile	20.14	9.19	0.34
China	2.57	4.02	2.86
Denmark	42.95	20	2.94
Egypt	0.93	3.02	1.94
Ireland	49.02	24.42	6.42
France	26.38	23.99	19.27
Germany	39.26	25.35	15.80
Greece	13.21	17.44	35.7
India	0.85	2.84	6.04
Iran	1.58	0	32.98
Ireland	23.21	32.41	50.15
Israel	27.68	19.79	1.24
Japan	38.42	25.13	0.92
Malaysia	27.31	9.75	0.40
Mexico	5.62	8.40	27.93
Netherlands	48.05	27.93	9.35
New Zealand	45.12	9.35	0.89
Nigeria	0.1	0.89	59.52
Norway	46.38	59.52	1.89
Pakistan	0.34	1.89	2.86
Philippines	2.86	2.84	7.1
Russia	2.86	7.1	1.94
Saudi Arabia	1.94	1.94	6.42
South Africa	6.42	19.27	15.80
Spain	19.27	15.80	35.7
Sweden	15.80	35.7	6.04
Switzerland	35.7	6.04	32.98
Turkey	6.04	32.98	50.15
United Kingdom	32.98	50.15	1.24
United States	50.15	1.24	0.92
Vietnam	1.24	0.92	0.92
Yemen	0.92	0.92	0.92

Aug 8-9:21 AM

Scatterplot

Graph of two quantitative variables:

- Horizontal Axis: Explanatory, x
- Vertical Axis: Response, y

Aug 8-9:21 AM

Interpreting Scatterplots

- The overall pattern includes trend, direction, and strength of the relationship
- **Trend:** linear, curved, clusters, no pattern
- **Direction:** positive, negative, no direction
- **Strength:** how closely the points fit the trend
- Also look for **outliers** from the overall trend

Aug 8-9:21 AM

Used-car Dealership

What association would we expect between the age of the car and mileage?

- Positive
- Negative
- No association

Aug 8-9:21 AM

Linear Correlation, r

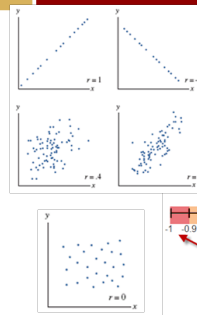
Measures the strength and direction of the linear association between x and y

Calculating the Correlation r

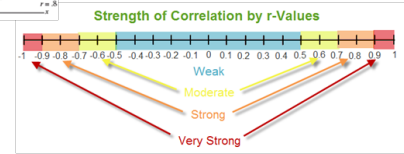
$$r = \frac{1}{n-1} \sum z_x z_y = \frac{1}{n-1} \sum \left(\frac{x - \bar{x}}{s_x} \right) \left(\frac{y - \bar{y}}{s_y} \right)$$

where n is the number of points, \bar{x} and \bar{y} are means, and s_x and s_y are standard deviations for x and y . The sum is taken over all n observations.

Correlation coefficient: Measuring Strength & Direction of a Linear Relationship



Positive $r \Rightarrow$ positive association
 Negative $r \Rightarrow$ negative association
 r close to $+1$ or -1 indicates

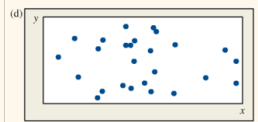
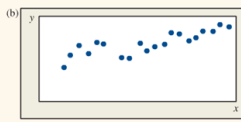
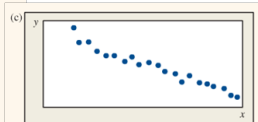
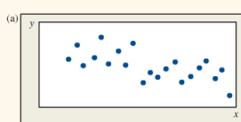


Aug 8-9:21 AM

Aug 8-9:21 AM

3.16 Match the scatterplot with r . Match the scatterplots below with the correlation values.

1. $r = -0.9$
2. $r = -0.5$
3. $r = 0$
4. $r = 0.6$



Aug 8-9:21 AM

Aug 14-3:18 PM