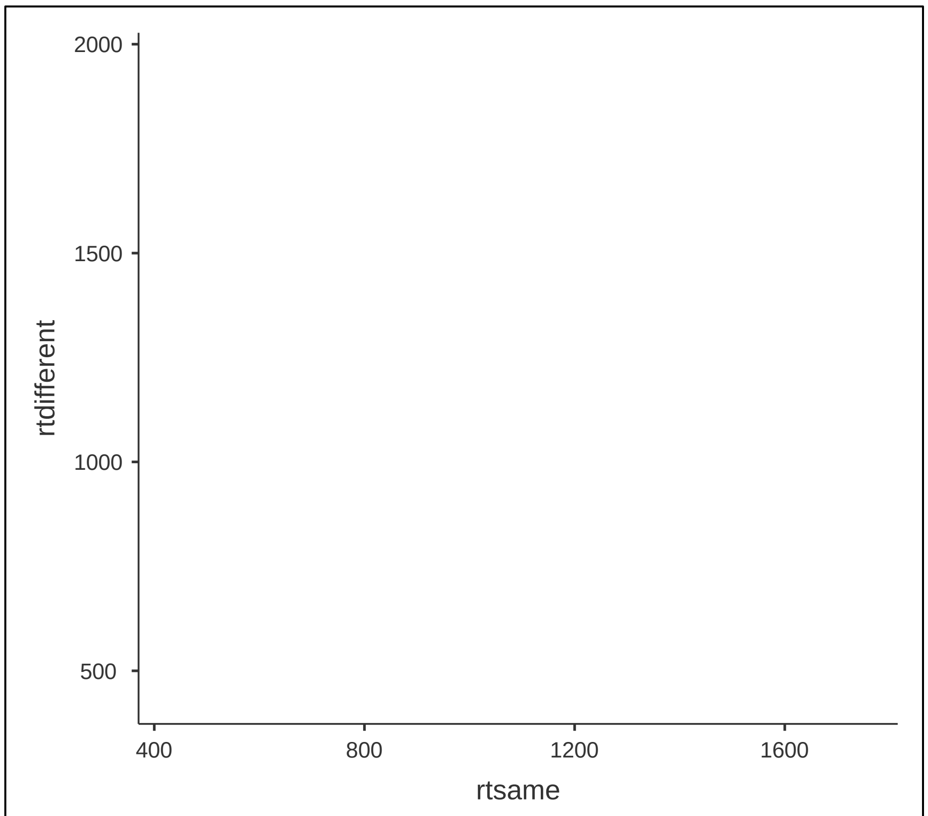
Learning Objectives

- **Describe** *correlation* in terms of direction, magnitude, and form
- **Build** intuitions about correlations based on visual *scatterplots*
- **Calculate** *correlation coefficient* and the *coefficient of determination* given sets of data
- **Describe** other correlation statistics and when we might use them

Relationships

- Thus far, we've described distributions of a single variable
- ***Relationships*** describe patterns between two variables
- Relationships are best visualized through ***scatterplots***

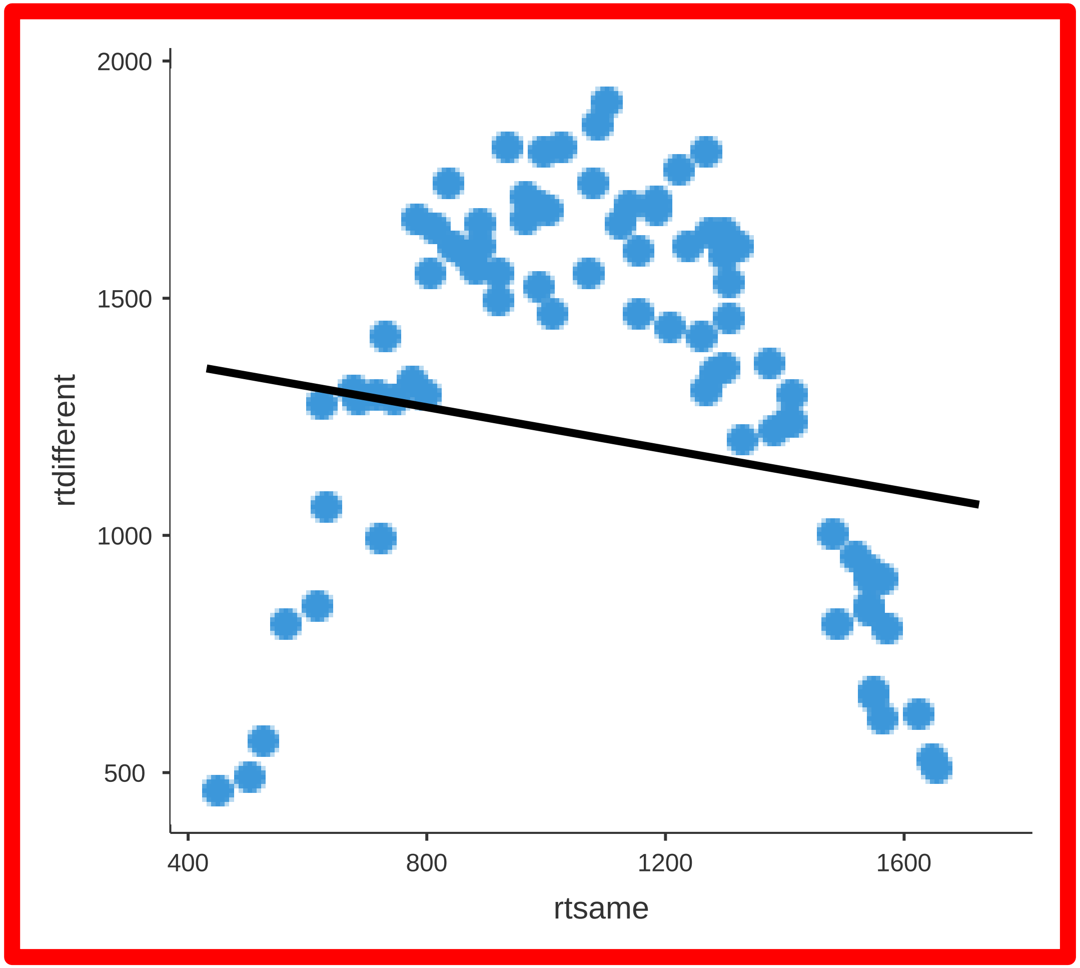


Correlation

- Quantifying the form, magnitude, and direction of relationships
  - Is the relationship linear or some other shape?
  - Is the relationship strong or weak?
  - Is the relationship positive or negative?

Linear Relationship

Form: Straight Lines



Linear Relationships

Pearson's *r*

Pearson's *r* quantifies linear relationships

- Varies between -1 and 1; where 0 indicates no relationship

- Magnitude: Larger absolute numbers = greater magnitude
  - *r* = .45 has same magnitude as *r* = -.45

- Direction: Indicated by positive or negative values

- We need to compare two different distributions (w/their own $\bar{X}$ and *s*)
  - How could we do that?
  - *r* is **standardized**! It describes how variables move together in standard deviation units
  - As variable X increases by one *s*, how many *s* does variable Y increase or decrease

Correlation, let's build intuition…

- Estimate the *magnitude* of relationships
  - http://guessthecorrelation.com/

- For more detailed practice:
  - https://www.rossmanchance.com/applets/GuessCorrelation.html

Calculating Pearson's *r*

FIRST, check the assumptions:

- Relationship is *linear*
  - Plot the scatterplot
- Variables are *normally distributed*
  - Plot a histogram, frequency polygon, or run a test of normality in SPSS
- Interval- or Ratio-level data
  - Think critically about your measurement instruments
- Absence of extreme *outliers*
  - Scrutinize the scatter plot again

$$\text{Pearson's } r = \frac{\Sigma z_x z_y}{N - 1}$$

Calculating Pearson's *r*

Method 1: *z*-scores

$$\text{Pearson's } r = \frac{\Sigma z_x z_y}{N - 1}$$

ID	rtsame (X)	rtdifferent (Y)	z_x	z_y	$z_x * z_y$
81	442	507		-1.129	
8	684	701	1.336	0.206	0.276
21	531	660	-0.399	-0.076	0.030
18	545	632	-0.240	-0.268	0.064
75	652	935	0.973	1.817	1.769
11	543	591	-0.263	-0.551	0.145
$N = 6$	$\bar{X} =$	$\bar{Y} =$			$\Sigma z_x * zy = 3.873$

Calculating Pearson's *r*

Method 2: Raw Scores

$$\text{Pearson's } r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{N}}{\sqrt{\left[\sum X^2 - \frac{(\sum X)^2}{N}\right]\left[\sum Y^2 - \frac{(\sum Y)^2}{N}\right]}}$$

ID	rtsame(x)	rtdiff(y)	X^2	Y^2	XY
81	442	507	195,364	257,049	224,094
8	684	701	467,856	491,401	479,484
21	531	660	281,961	435,600	350,460
18	545	632	297,025	399,424	344,440
75	652	935	425,104	874,225	609,620
11	543	591	294,849	349,281	320,913
N = 6	ΣX = 3397	ΣY = 4026	ΣX^2 = 1,962,159	ΣY^2 = 2,806,980	ΣXY = 2,329,011

Calculating Pearson's *r*

Method 2: Raw Scores

$$\text{Pearson's } r = \frac{\Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{N}}{\sqrt{\left[\Sigma X^2 - \frac{(\Sigma X)^2}{N}\right]\left[\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}\right]}}$$

$N = 6$

$\Sigma X = 3{,}397$

$\Sigma Y = 4{,}026$

$(\Sigma X)^2 = 11{,}539{,}609$

$(\Sigma Y)^2 = 16{,}208{,}676$

$\Sigma X^2 = 1{,}962{,}159$

$\Sigma Y^2 = 2{,}806{,}980$

$\Sigma XY = 2{,}329{,}011$

$$\frac{2{,}329{,}011 - \frac{(3{,}397)(4{,}026)}{6}}{\sqrt{([1{,}962{,}159 - \frac{11{,}539{,}609}{6}][2{,}806{,}980 - \frac{16{,}208{,}676}{6}])}}$$

Calculating Pearson's *r*

Method 2: Raw Scores

$$\text{Pearson's } r = \frac{\sum XY - \dfrac{(\sum X)(\sum Y)}{N}}{\sqrt{\left[\sum X^2 - \dfrac{(\sum X)^2}{N}\right]\left[\sum Y^2 - \dfrac{(\sum Y)^2}{N}\right]}}$$

$N = 6$

$\Sigma X = 3{,}397$

$\Sigma Y = 4{,}026$

$(\Sigma X)^2 = 11{,}539{,}609$

$(\Sigma Y)^2 = 16{,}208{,}676$

$\Sigma X^2 = 1{,}962{,}159$

$\Sigma Y^2 = 2{,}806{,}980$

$\Sigma XY = 2{,}329{,}011$

$$\frac{2{,}329{,}011 - \dfrac{(3{,}397)(4{,}026)}{6}}{\sqrt{([1{,}962{,}159 - \dfrac{11{,}539{,}609}{6}][2{,}806{,}980 - \dfrac{16{,}208{,}676}{6}])}}$$

Example pp. 134

$$\text{Pearson's } r = \frac{\Sigma XY - \dfrac{(\Sigma X)(\Sigma Y)}{N}}{\sqrt{\left[\Sigma X^2 - \dfrac{(\Sigma X)^2}{N}\right]\left[\Sigma Y^2 - \dfrac{(\Sigma Y)^2}{N}\right]}}$$

ID	X	Y	X^2	Y^2	XY
A	1	2	1	4	2
B	3	5	9	25	15
C	4	3	16	9	12
D	6	7	36	49	42
E	7	5	49	25	35
N =	ΣX =	ΣY =	ΣX^2 =	ΣY^2 =	ΣXY =

Calculating Pearson's *r*

Method 2: Raw Scores

$$\text{Pearson's } r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{N}}{\sqrt{\left[\sum X^2 - \frac{(\sum X)^2}{N}\right]\left[\sum Y^2 - \frac{(\sum Y)^2}{N}\right]}}$$

$N = 5$

$\Sigma X = 21$

$\Sigma Y = 22$

$(\Sigma X)^2 =$

$(\Sigma Y)^2 =$

$\Sigma X^2 = 111$

$\Sigma Y^2 = 112$

$\Sigma XY = 106$

$$\frac{106 - \frac{(21)(22)}{5}}{\sqrt{([111 - \frac{???}{5}][112 - \frac{???}{5}])}}$$

Calculating Pearson's *r*

Method 2: Raw Scores

$$\text{Pearson's } r = \frac{\Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{N}}{\sqrt{\left[\Sigma X^2 - \frac{(\Sigma X)^2}{N}\right]\left[\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}\right]}}$$

$N = 5$

$\Sigma X = 21$

$\Sigma Y = 22$

$(\Sigma X)^2 = 441$

$(\Sigma Y)^2 = 484$

$\Sigma X^2 = 111$

$\Sigma Y^2 = 112$

$\Sigma XY = 106$

$$\frac{106 - \frac{462}{5}}{\sqrt{\left(\left[111 - \frac{441}{5}\right]\left[112 - \frac{484}{5}\right]\right)}}$$

Calculating Pearson's *r*

Method 2: Raw Scores

$$\text{Pearson's } r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{N}}{\sqrt{\left[\sum X^2 - \frac{(\sum X)^2}{N}\right]\left[\sum Y^2 - \frac{(\sum Y)^2}{N}\right]}}$$

$N = 5$

$\Sigma X = 21$

$\Sigma Y = 22$

$(\Sigma X)^2 = 441$

$(\Sigma Y)^2 = 484$

$\Sigma X^2 = 111$

$\Sigma Y^2 = 112$

$\Sigma XY = 106$

$$\frac{106 - 92.4}{\sqrt{([111 - 88.2][112 - 96.8])}}$$

Calculating Pearson's r

Method 2: Raw Scores

$$\text{Pearson's } r = \frac{\Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{N}}{\sqrt{\left[\Sigma X^2 - \frac{(\Sigma X)^2}{N}\right]\left[\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}\right]}}$$

$N = 5$
$\Sigma X = 21$
$\Sigma Y = 22$
$(\Sigma X)^2 = 441$
$(\Sigma Y)^2 = 484$
$\Sigma X^2 = 111$
$\Sigma Y^2 = 112$
$\Sigma XY = 106$

$$\frac{13.6}{\sqrt{([22.8][15.2])}}$$

Calculating Pearson's r

Method 2: Raw Scores

$$\text{Pearson's } r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{N}}{\sqrt{\left[\sum X^2 - \frac{(\sum X)^2}{N}\right]\left[\sum Y^2 - \frac{(\sum Y)^2}{N}\right]}}$$

$N = 5$

$\Sigma X = 21$

$\Sigma Y = 22$

$(\Sigma X)^2 = 441$

$(\Sigma Y)^2 = 484$

$\Sigma X^2 = 111$

$\Sigma Y^2 = 112$

$\Sigma XY = 106$

$$\frac{13.6}{\sqrt{346.56}} = \frac{13.6}{18.616} = .7306 = r$$

Magnitude

- Putting words to correlation coefficients

If r is...	Interpretation
Equal to 0	No relationship
Between 0 and 0.10	Trivial
Between 0.10 and 0.30	Small to medium
Between 0.30 and 0.50	Medium to large
Greater than 0.50	Large to very large

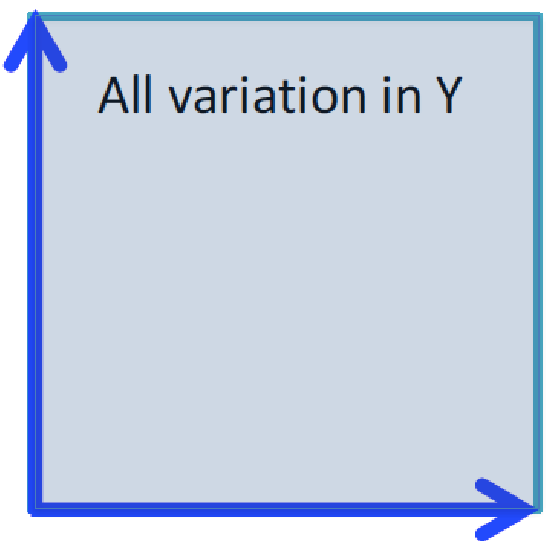
Coefficient of Determination (r^2)

- How are these two variables related? *Use r*
  - "relatedness", "covariance"

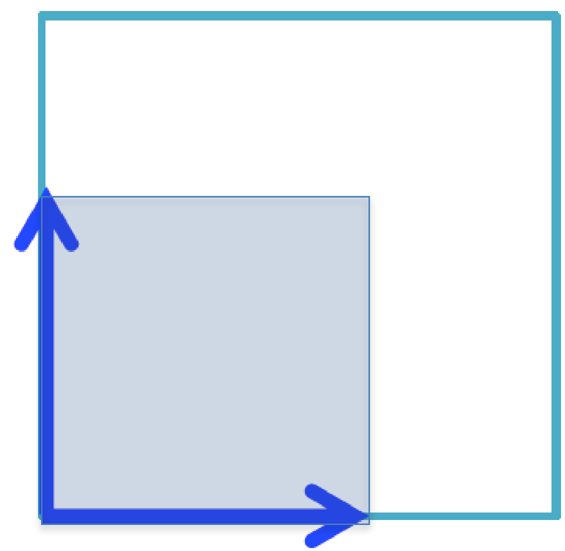
- How much variability in Y is accounted for by knowing X? *Use r^2*
  - *"explained variance"*
  - If $r = 1$, then $r^2 = 1$
    - "<u>All</u> (or 100%) of the variability in Y can be accounted for by variability in X"
  - When $r < 1$, then $r^2 < r$

Coefficient of Determination (r^2)

- Calculation is simple!
  - r^2 ranges from 0 to +1
    - Why not negative??



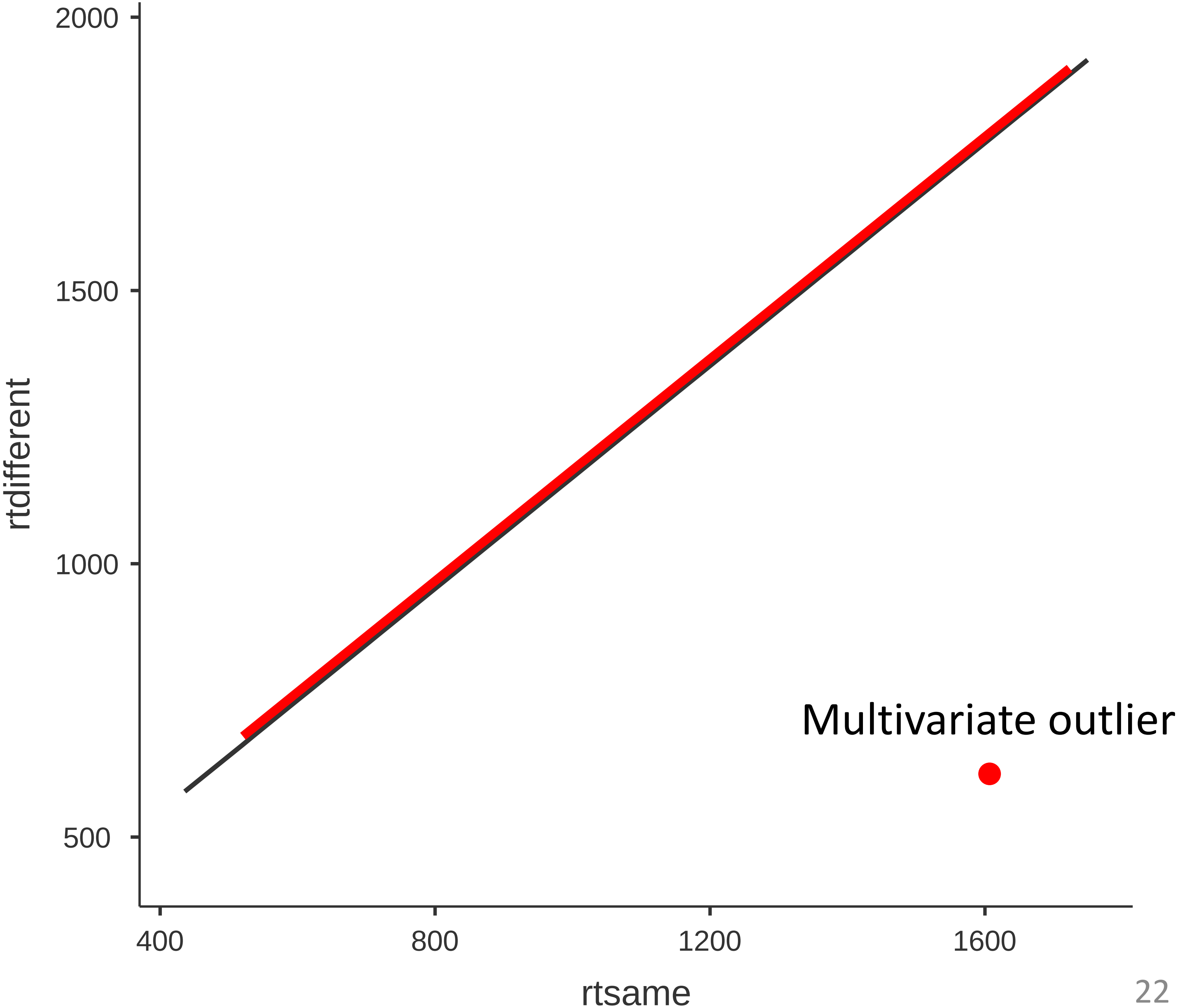
$r = 1.00$
$r^2 = 1.00$



$r = 0.65$
$r^2 = 0.42$

Other Correlation Coefficients

- Pearson's *r* (linear, normal dist, interval+ variable, no extreme outliers)
- Spearman's rho (r_s, ρ):
  - Use for ordinal data, or non-normal distributed variables
  - Only assumes variables are ranked
  - Also see Kendall's Tau (τ) & Goodman's Gamma (γ)
- Point biserial r_b
  - Use when 1 var is interval/ratio, 1 var is dichotomous
- Pearson's Phi (ϕ)
  - Use when both vars are dichotomous
- Eta correlation ratio (η)
  - Use for simple non-linear relationships
  - Common for ANOVA
  - Also see Omega correlation ratio (ω)

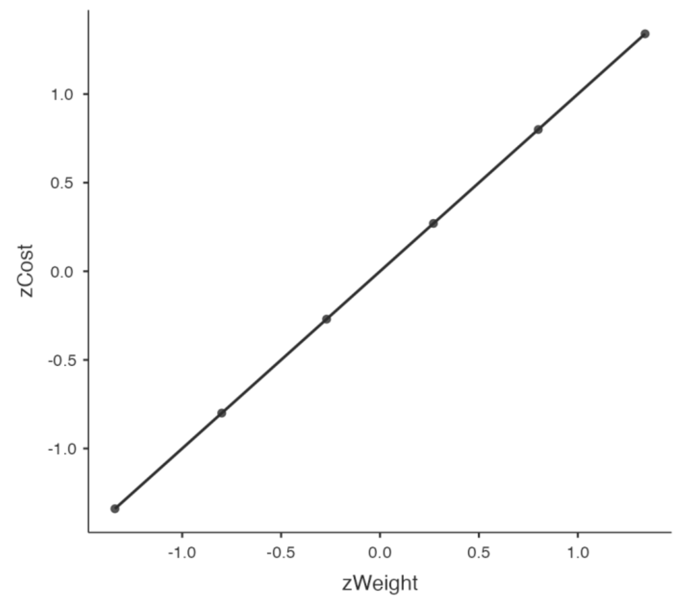

2000
1500
1000
500
rtdifferent
400
800
1200
1600
rtsame
Multivariate outlier

Pearson's *r*, Table 6.3

Original Data

$$\text{Pearson's } r = \frac{\Sigma z_x z_y}{N-1}$$

ID	z_{weight} (X)	Cost (Y)	Z_{cost}	$z_x * z_y$
A	-1.34	0.75	-1.34	
B	-0.80	1.00	-0.80	
C	-0.27	1.25	-0.27	
D	0.27	1.50	0.27	
E	0.80	1.75	0.80	
F	1.34	2.00	1.34	
N = 6				Σ =

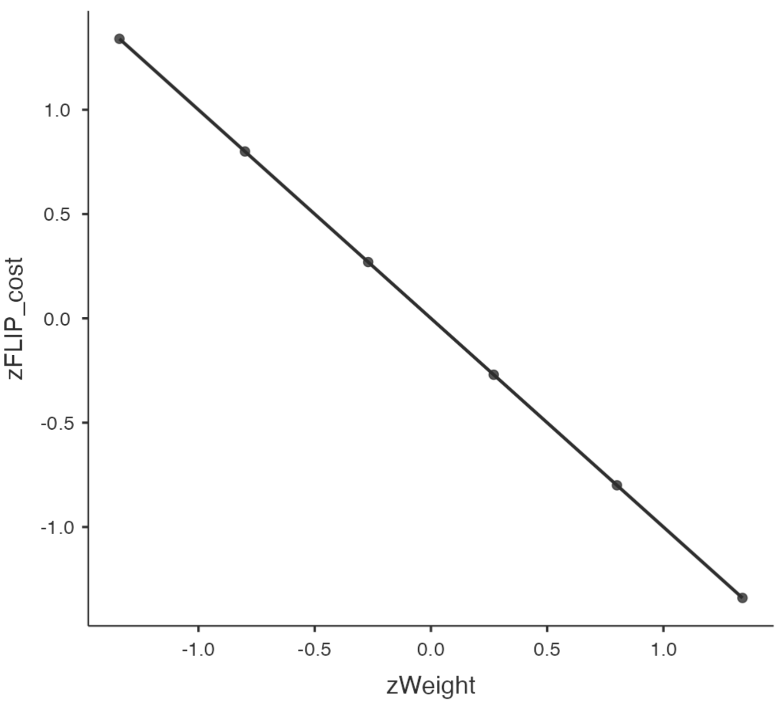


Pearson's *r*, Table 6.3

Flipped Data

$$\text{Pearson's } r = \frac{\Sigma z_x z_y}{N - 1}$$

ID	z_{weight} (*X*)	Cost (*Y*)	Z_{cost}	$z_x * z_y$
A	-1.34	2.00	1.34	
B	-0.80	1.75	0.80	
C	-0.27	1.50	0.27	
D	0.27	1.25	-0.27	
E	0.80	1.00	-0.80	
F	1.34	0.75	-1.34	
N = 6				Σ =



Pearson's *r*, Table 6.3

Randomized Data

$$\text{Pearson's } r = \frac{\Sigma z_x z_y}{N - 1}$$

ID	z_{weight} (X)	Cost (Y)	Z_{cost}	$z_x * z_y$
A	-1.34	1.00	-0.80	
B	-0.80	1.50	0.27	
C	-0.27	2.00	1.34	
D	0.27	0.75	-1.34	
E	0.80	1.25	0.27	
F	1.34	1.75	0.80	
N = 6				Σ =