**Correlation**

**Correlation is the relationship between two variables.**

**. . .**

|  |  |  |
| --- | --- | --- |
|  | **Correlation Introduction** |  |

Consider the data from five subjects that were asked miles and minutes to arrive at a destination.

Let X=Miles and Y=Minutes.

**Does there seem to be a relationship between the two variables?**

|  |  |  |  |
| --- | --- | --- | --- |
| Subject | X | Y |  |
| 1 | 1 | 4 |
| 2 | 3 | 6 |
| 3 | 3 | 20 |
| 4 | 5 | 15 |
| 5 | 8 | 20 |

Discussion:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1. If X and Y increase (or decrease) together, then there is a positive relationship between X and Y.  If X increases as Y decreases (or X decreases as Y increases), then there is a negative relationship.  2. To represent the relationship between the two variables, consider the sum of squared differences between the two variables, SSXY = ( X –X )\*( Y –Y )  3. If X and Y increase (or decrease) together, then SSXY>0 and there is a positive relationship.  If X increases as Y decreases (or X decreases as Y increases), then SSXY<0 and there is a negative relationship.  4. Examples:   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Index | X | Y | SSXY |  | | 1 | 2 | 1 | (2-4)\*(1-3) = +4 | | 2 | 4 | 3 | (4-4)\*(3-3) = 0 | | 3 | 6 | 5 | (6-4)\*(5-3) = +4 | | Mean | 4 | 3 | SSXY = Sum = +8 | |  |  |  |  |  | | Index | X | Y | SSXY |  | | 1 | 2 | 5 | (2-4)\*(5-3) = –4 | | 2 | 4 | 3 | (4-4)\*(3-3) = 0 | | 3 | 6 | 1 | (6-4)\*(1-3) = –4 | | Mean | 4 | 3 | SSXY = Sum = –8 |   . . . |

Definitions.

|  |
| --- |
| 1. Sample Covariance is defined, **Cov(X,Y) = SSXY/(n–1)**  2. Correlation is defined: **Correlation, r = SSXY/sqrt(SSXX\*SSYY), where (–1 ≤ r ≤ +1).**  3. Covariance can be considered an absolute measure of a linear relationship.  Correlation can be considered a relative measure of a linear relationship (relative between –1 and +1 ).  . . . |

|  |  |  |
| --- | --- | --- |
|  | **Correlation Example** |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Subject | 1 | 2 | 3 | 4 | 5 | Sum | SS | S2 |  |  |
| X=Miles | 1 | 3 | 3 | 5 | 8 | 20 |  |  |  |  |
| Y=Minutes | 4 | 6 | 20 | 15 | 20 | 65 |  |  |  | SS=Sum of Squares |
| X\*X | 1 | 9 | 9 | 25 | 64 | 108 | 28 | 28/4=7 |  | SSXX = 108–20\*20/5 = 28  SSYY = 1077–65\*65/5 = 232  SSXY = 317–20\*65/5 = 57 |
| Y\*Y | 16 | 36 | 400 | 225 | 400 | 1077 | 232 | 232/4=58 |  |
| X\*Y | 4 | 18 | 60 | 75 | 160 | 317 | 57 | 57/4=14.25 |  |

|  |
| --- |
| 1. Sample Variance, S2X= SSXX/(n–1)=28/4=7. Excel: “=Var.s(Data)”  Sample Variance, S2Y= SSYY/(n–1)=232/4=58. Excel: “=Var.s(Data)”  2. Sample Covariance is defined, **Cov(X,Y) = SSXY/(n–1)**  Sample Covariance, Cov(X,Y)=SSXY/(n–1)=57/(5–1)=14.25. Excel: “=Covariance.s(Y-Data,X-Data)  3. When SSXY is compared with SSXX and SSYY, it can be shown that (SSXY)2 ≤ (SSXX)\*( SSYY).  Therefore, Correlation is defined: **Correlation, r = SSXY/sqrt(SSXX\*SSYY), where (–1 ≤ r ≤ +1).**  Correlation, r = SSXY/sqrt(SSXX\*SSYY) = 57/sqrt(28\*232) ≈ 0.707215625 ≈ 0.7072.  Correlation Coefficient. Excel: “=Correl(Y-Data,X-Data)”  4. Covariance can be considered an absolute measure of a linear relationship.  Correlation can be considered a relative measure of a linear relationship (relative between –1 and +1 ).  . . . |

{ Examples }

|  |  |  |
| --- | --- | --- |
|  | **Correlation Matrix Representations** |  |

Consider three sets of data, A, B, & C, along with their univariate and bivariate measures.

(Let X=(A,B,C) & Y=(A,B,C))

Univariate Measures:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 |  | X | X2 | SSXX = X2 – (X)2/n | S2=SS/(n–1) | s2=SS/n |
| Set A | 0 | 2 | 1 | 2 | 1 |  | 6 | 10 | 2.8 | 0.70 | 0.56 |
| Set B | 4 | 1 | 2 | 0 | 2 |  | 9 | 25 | 8.8 | 2.20 | 1.76 |
| Set C | 2 | 0 | 0 | 2 | 0 |  | 4 | 8 | 4.8 | 1.20 | 0.96 |
|  |  |  |  |  |  |  |  |  | Sum of Squares | Sample  Variance | Population  Variance |

Bivariate Measures: Sum of Squares, SSXY = XY – (X)\*(Y)/n

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  | Sum(XY) | | | SSXY = XY – (X)\*(Y)/n | | |
|  | 1 | 2 | 3 | 4 | 5 |  | X | A | B | C | A | B | C |
| Set A | 0 | 2 | 1 | 2 | 1 |  | 6 | 10 | 6 | 4 | 2.8 | –4.8 | –0.8 |
| Set B | 4 | 1 | 2 | 0 | 2 |  | 9 |  | 25 | 8 |  | 8.8 | 0.8 |
| Set C | 2 | 0 | 0 | 2 | 0 |  | 4 |  |  | 8 |  |  | 4.8 |

Bivariate Measures: Population Variance & Population Covariance = SSXY/n

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  | SSXY | | | Population Variance/  Population Covariance = SSXY/n | | |
|  | 1 | 2 | 3 | 4 | 5 |  | A | B | C | A | B | C |
| Set A | 0 | 2 | 1 | 2 | 1 |  | 2.8 | –4.8 | –0.8 | 0.56 | –0.96 | –0.16 |
| Set B | 4 | 1 | 2 | 0 | 2 |  |  | 8.8 | 0.8 |  | 1.76 | 0.16 |
| Set C | 2 | 0 | 0 | 2 | 0 |  |  |  | 4.8 |  |  | 0.96 |

Bivariate Measures: Correlation, r = SSXY/sqrt(SSXX\*SSYY), where (–1 ≤ r ≤ +1)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  | SSXY | | | Correlation, r = SSXY/sqrt(SSXX\*SSYY),  where (–1 ≤ r ≤ +1). | | |
|  | 1 | 2 | 3 | 4 | 5 |  | A | B | C | A | B | C |
| Set A | 0 | 2 | 1 | 2 | 1 |  | 2.8 | –4.8 | –0.8 |  | –0.967 | –0.218 |
| Set B | 4 | 1 | 2 | 0 | 2 |  |  | 8.8 | 0.8 |  |  | 0.123 |
| Set C | 2 | 0 | 0 | 2 | 0 |  |  |  | 4.8 |  |  |  |

Matrix Representation of Population Variance, Population Covariance, and Correlation

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Correlation Matrix** | | | |  | **Covariance Matrix** | | | |  | **Variance Matrix** | | | |
|  | A | B | C |  |  | A | B | C |  |  | A | B | C |
| A |  | –0.967 | –0.218 |  | A |  | –0.96 | –0.16 |  | A | 0.56 |  |  |
| B |  |  | 0.123 |  | B |  |  | 0.16 |  | B |  | 1.76 |  |
| C |  |  |  |  | C |  |  |  |  | C |  |  | 0.96 |

Combined Matrix Representation (variations exist)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variance/Covariance/Correlation Matrix** | | | |  |
|  | **A** | **B** | **C** |  |
| **A** | 0.56 | –0.96 | –0.16 | 🡨Covariance in Upper Triangular |
| **B** | –0.967 | 1.76 | 0.16 |
| **C** | –0.218 | 0.123 | 0.96 | 🡨Variance in Main Diagonal |
|  | **↑** Correlation in Lower Triangular | | |  |



|  |  |  |
| --- | --- | --- |
|  | **Test of Hypothesis for Correlations** |  |

A Test of Hypothesis for Significant Correlations implies that the Population Correlation, r, is Statistically Significantly Different from Zero.

|  |
| --- |
| **Consider the pair of Hypotheses about the Population Correlation, r** |
| Ho: r = 0  Ha: r ≠ 0 |

Details of Statistics omitted. Common Representations (variations exist)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Correlation Matrix** | | | |  | **t-Test Statistic** | | | |  | **p-value** | | | |
|  | A | B | C |  |  | A | B | C |  |  | A | B | C |
| A |  | –0.967 | –0.218 |  | A |  | 6.573\* | 0.387\*\* |  | A |  | 0.007 | 0.724 |
| B |  |  | +0.123 |  | B |  |  | 0.215\*\* |  | B |  |  | 0.844 |
| C |  |  |  |  | C |  |  |  |  | C |  |  |  |
| [ \*(P<0.10), \*\*(P>0.10), for Ho:=0 vs. Ha:r≠0 ] | | | | | | | | | | | | | |

{ Examples }