**Correlation & Regression – Introduction**

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

**Regression is estimating one variable conditioned on another variable.**

**. . .**

|  |
| --- |
| **Fundamentals of Mechanics**Consider a paired random sample for random variables X and Y with a sample size of n. |
| . . . |
| Example. Five subjects were asked miles and minutes to arrive at a destination.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Subject | 1 | 2 | 3 | 4 | 5 | Sum |  |
| X=Miles | 1 | 3 | 3 | 5 | 8 | 20 | = S X |
| Y=Minutes | 4 | 6 | 20 | 15 | 20 | 65 | = S Y |
|  |  |  |  |  |  |  |  |
| X\*X | 1 | 9 | 9 | 25 | 64 | 108 | = S X2 |
| Y\*Y | 16 | 36 | 400 | 225 | 400 | 1077 | = S Y2 |
| X\*Y | 4 | 18 | 60 | 75 | 160 | 317 | = S X\*Y |

. . . |
| Sample Size, n=5Sample Mean

|  |
| --- |
| X = S X / n = 20/5 = 4Y = S Y / n = 65/5 = 13 |

. . . | SS = Sum of Squares of Error = Sum of Squared Errors

|  |
| --- |
| SSXX= ( X –X )2 = X2–(X)\*(X)/n = 108–20\*20/5 = 28SSYY= ( Y –Y )2 = Y2–(Y)\*(Y)/n = 1077–65\*65/5 = 232SSXY= ( X –X )\*( Y –Y ) = X\*Y–(X)\*(Y)/n = 317–20\*65/5 = 57 |

. . . |
| Sample Variance

|  |
| --- |
| S2X = SSXX / (n–1) = 28/(5-1) = 7 S2Y = SSYY / (n–1) = 232/(5-1) = 58 |

Sample Standard Deviation

|  |
| --- |
| SX = √( SSXX/(n–1) ) = √7 SY = √( SSYY/(n–1) ) = √58 |

. . . | Sample Covariance

|  |
| --- |
| Cov(X,Y) = SSXY / (n–1) = 57/(5–1) = 14.25 |

Sample Correlation

|  |
| --- |
| r = SSXY/sqrt(SSXX\*SSYY) = 57/sqrt(28\*232) ≈ 0.7072 where (–1 ≤ r ≤ +1)Notice, r = Cov(X,Y)/( SX \* SY ) |

. . . |
| Regression

|  |
| --- |
| Regress Y on X. Dependent variable, Y, is regressed on independent variable, X. Use X to estimate mean of Y. E[Y|X].Simple Linear Regression Model, E[Y|X] = b0+b1\*X + e  ( b0 & b1 are parameters. The error term, e~N(m,s2) )Simple Linear Regression Equation, Y=b0+b1\*X  ( b0 & b1 are estimates of the parameters b0 & b1 )b1 = SSXY / SSXX  = 57/28 ≈ 2.0357b0 =Y–b1 \*X = 13-(57/28) \* 4 ≈ 4.8571Regression Equation: Y ≈ 4.8571 + 2.0357 \* X . . . |

. . . |