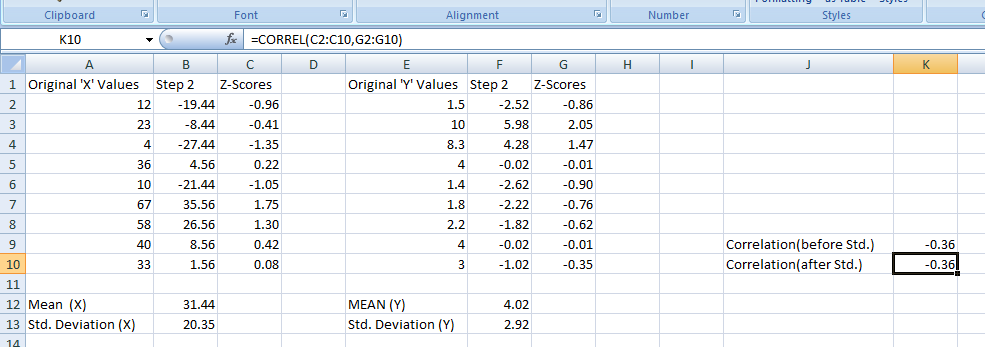
**PRACTICE QUESTIONS**

**1.** Given ordered data sets X= {12, 23, 4, 36, 10, 67, 58, 40, 33} and Y = {1.5, 10, 8.3, 4, 1.4, 1.8, 2.2, 4, 3}, **what is the correlation of their ordered pairs after standardization?**

***First Standardize the given X & Y values ; then calculate the correlation among the Z-Scores obtained.***

******

**ANSWER:- -0.36**

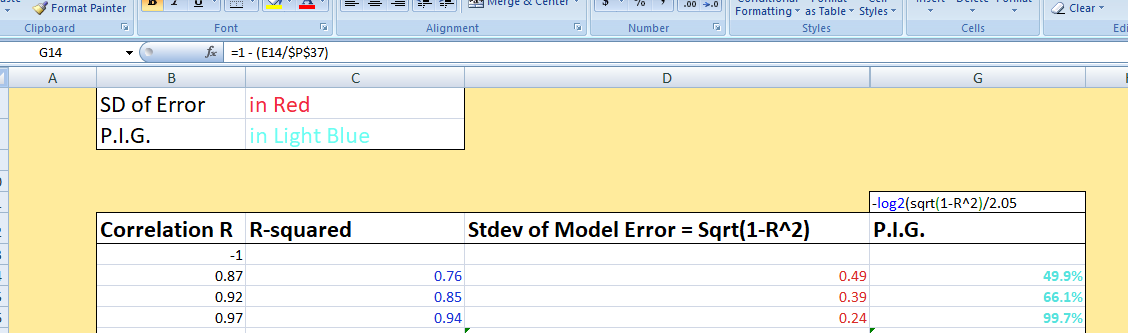
**2.** If a linear regression between standardized ordered pairs (both Gaussian distributions) has mean = 0, and root mean square error (standard deviation) = 0.5, **what is the coefficient of determination (R-squared)?**

***Coeff. Of Determination (R-squared) = 1 – (Std. Dev.)2 = 1 – (0.5)2 = =***

**ANSWER:- 0.75**

**3.** Given correlation of R = 0.87 on a standardized linear regression (with Gaussian distributions), **what is the Percentage Information Gain (P.I.G.)?**

Hint: Use the Correlation and P.I.G. Spreadsheet.



**ANSWER:- 49.8%**

**4.** True/False: After standardization, the slope of the regression line, “beta,” will equal the covariance and the correlation between the ordered pairs (x, y).

Hint: Refer to the Standardization Spreadsheet

**ANSWER:- True**

**5.** Assume a regression line on standardized data with slope = 0.7. At the point x = 1, the point estimate for y = 0.7.

**What upper bound for the confidence interval of y will ensure that 95% of all possible outcomes for y will be below it?**

Hint: Use the Correlation and Model Error Spreadsheet.

***The Excel NormSInv(probability) function gives NormSInv(0.95) = 1.644 standard deviations above the mean. Since beta = R = 0.7, we can calculate the root mean square error as sqrt(1- beta-squared) = 0.71. Then (1.644)\*(0.71) = 1.17 above the mean. Don't forget to add the mean of 0.7. to get the y-axis location.***

***Y-axis location = 1.17+0.7 = 1.87***

**ANSWER:- 1.87**

**6.** True/False: For a parameterized Gaussian model with continuous distributions, if the correlation R between random variables X and Y is increased, the percentage information gain from the linear regression also increases, until it becomes infinite at R = 1.

**ANSWER:- True**

**7.** Assume you work for a general home construction company “big box” retail chain. Your supervisor points out that there is a near-linear association between retail price of an item and how long on average it remains on the shelf. Correlation between price and time held in inventory is .82. The current model has a standard deviation of model error of 12.4 hours.

**To reduce the standard deviation of model error to 8 hours, what would the linear correlation need to be?**

***Model error on standardized data is equal to sqrt(1-R^2). At R= .82, the standard deviation of the model error is .5724. The standard deviation of the model error needed would be .5724\*(8/12.4) = .3693. Since R = sqrt(1 – error-squared), the correlation needed is .93.***

**ANSWER:- 0.93**

**8.** The latency [time delay for a signal to be processed] of a fiber optic network is closely correlated to the physical distance a signal must travel across the network. A network provider uses a linear model for expected latency in nanoseconds. The model is: Nanoseconds = (3.34\*physical distance in meters) + 25. The standard deviation of model error is 8 nanoseconds. **You are asked to give the range in nanoseconds within which 99% of all latencies will fall.**

***A 99% double-sided interval means .5% above and .5% below the interval. The Z-score p = .995 = NormSInv(.995) = 2.5758. The range of the confidence interval will be from 2.5758\*(8) nanoseconds less, to 2.5758\*(8) nanoseconds more, than the model’s point forecast. This range is 41.21 Nanoseconds.***

**ANSWER:- 41.21 nanoseconds**

**9.** Assume a linear regression model to forecast the profitability of future company clients had a standard deviation of model error of $3,885.

Assume this model error remains constant on new data. If you used the linear model to forecast the profitability for 200 completely new customers, and calculated the mean profitability for the 200, what would the standard deviation of model error for that mean profitability in that group be?

Give your answer to the nearest dollar.

***The Central Limit Theorem means that the standard deviation of sample means for a set of “n” items are related to the standard deviation of the underlying population x (in this case the “population” is of individual model errors) by x/sqrt(n). So the correct answer is ($3,885.14) / (Sqrt(200)) = $274.76 = $275***

**ANSWER:- $275**