MAXIMIZING DATA'S POTENTIAL

Orthogonal Regression

Ted Ellefson
Seagate Technology
Updated: Feb 2021

SEAGATE

1

Orthogonal Regression - Outline

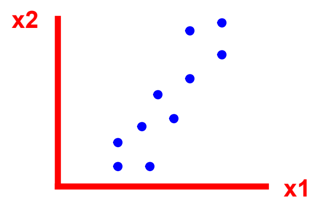
- **Why would we want to do Orthogonal Regression?**
- **What are the computational differences between ordinary least squares (OLS) and Orthogonal Regression**
- **Perform an Orthogonal Regression study using JMP**

 Page 2 Seagate

2

What Does Correlation Involve?

One variable (e.g. “x2”) is expected to be numerically related to another variable (e.g. “x1”) in a linear fashion.



Relationship may or may not be causal

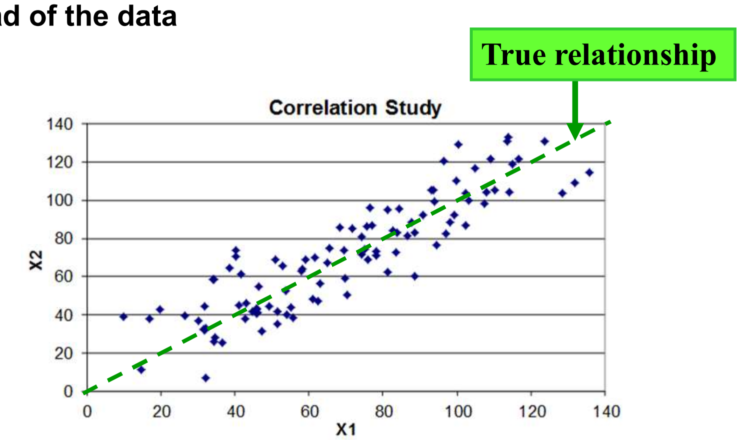
- **Two machines measuring same types of parts**
- **Number of storks versus births**
- **Number of traffic fatalities versus number of speeding cars**

3

Problems with Ordinary Least Squared Regression

Example: assume

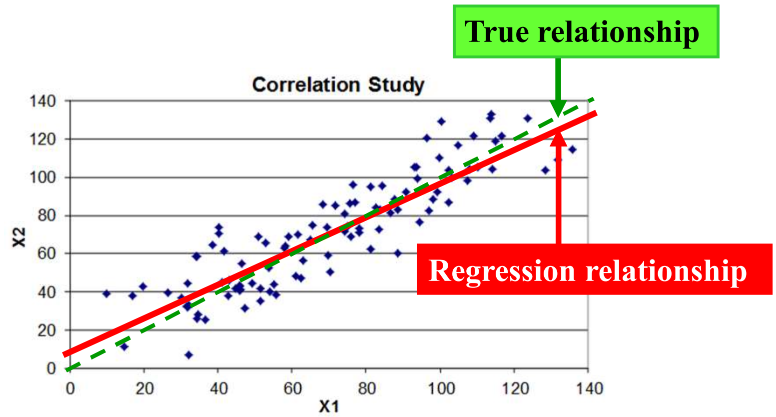
- **True relationship is x2 = x1**
- **Both x1 and x2 have a high measurement error relative to the spread of the data**



4

OLS Regression: x2 versus x1

Regression relationship is x2 = 0.895x1 + 8.675

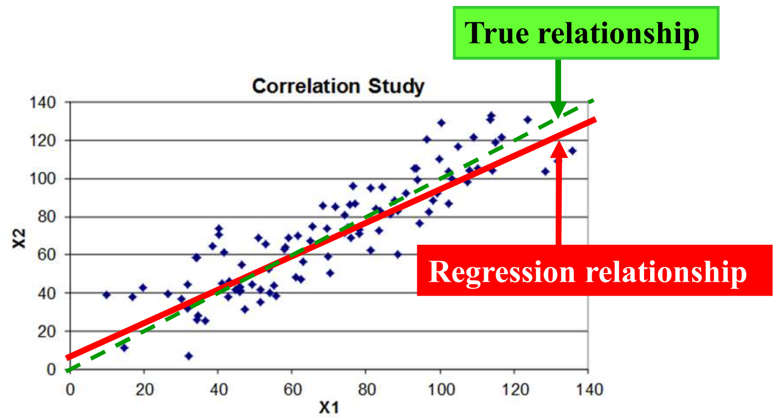


5

OLS Regression: x1 versus x2

Reverse the approach to x1 = b'x2 + a'

Regression relationship is x1 = 0.889x2 + 6.549



6

OLS Regression Question

Results

- Equation 1: $x2 = 0.895x1 + 8.675$
- Equation 2: $x1 = 0.889x2 + 6.549$

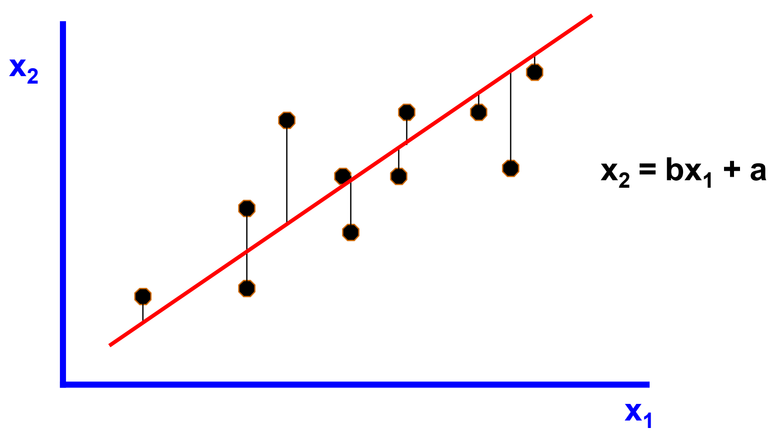
Invert Equation 1

- $x1 = (1/0.895)(x2 - 8.675) = 1.117x2 - 9.627$
- Why isn't this equal to Equation 2?

OLS Regression uses an approach that assumes all error is in the "Y" variable, and so tries to minimize deviations of "Y" from the line.

7

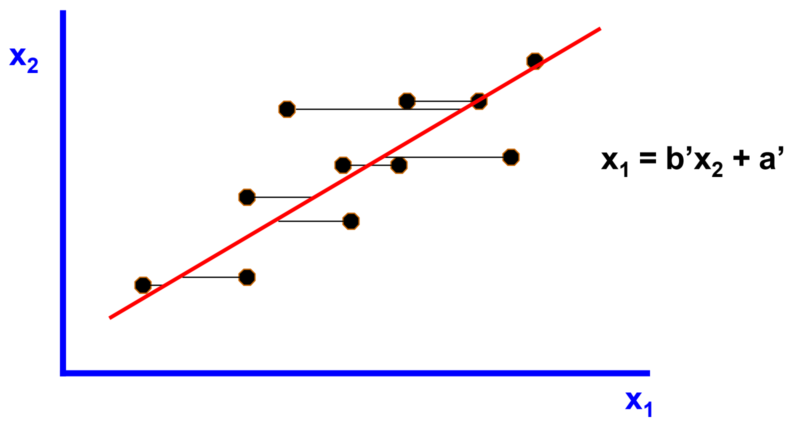
Mechanics of Ordinary Least Squared Regression



Regression uses an approach that assumes all error is in the "Y" variable, and so tries to minimize deviations of "Y" from the line.

8

Mechanics of OLS Regression - Reversed



The regression approach thus changes for the different relationship of $x_1 = b'x_2 + a'$.

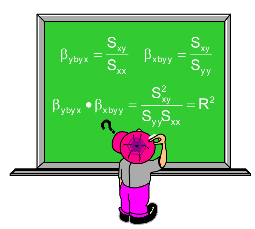
(This graph did not reverse the X and Y axis variables)

9

Interesting Regression Fact

Given

- **Regular regression: x2 = bx1 + a**
- **Reversed regression : x1 = b'x2 + a'**



Then $bb' = R^2$

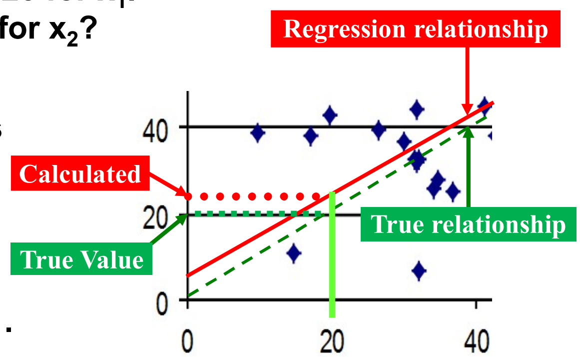
This means that regression will (incorrectly) reduce the true slope by a factor approximately equal to R.

Thus, if your correlation has an R^2 of 0.81, the slope of the regression line will be about 10% too low!

10

Example of OLS Regression Error: Setting Specs

Specs are at 20 and 120 for x_1.
What should they be for x_2?



OLS Regression says that the lower x_2 spec should be 26.6.

The upper spec is calculated to be 116.1.

The estimated x_2 spec range is [26.6, 116.1], whereas it should be [20, 120]. We're needlessly decreasing the acceptable range by more than 10%, throwing away a lot of good material!

11

Impact of Using OLS Regression for Gage to Gage Correlation

Regression may be used for correlation only for the following two situations:

- **The R^2 value exceeds 0.95.**
- **One of the systems represents the "gold" standard; that is, ALL other systems are tied directly and primarily (not secondarily) to that standard**

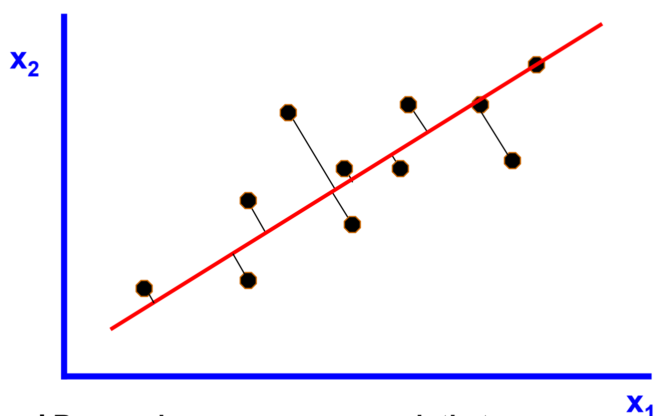
12

Solution for Performing Correlation

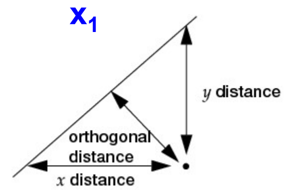
- A method originally attributed to W. Edwards Deming takes into account the error in both variables.
- The result is completely invertible! It does not matter which is the "x" variable and which is the "y" variable.
  - The method is referred to as "***orthogonal regression***" (regression when both input and output have variability).
- Creates a simple linear relationship only (no higher order terms): x2 = bx1 + a
- Requires only a knowledge (or estimate) of the ratio of the error in y to the error in x.

13

Principle of Orthogonal Regression



Orthogonal Regression uses an approach that minimizes the sum of the squared perpendicular differences



Requires that you specify the ratio of the variance of the error in X and Y

14

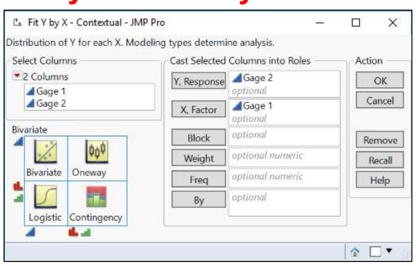
Orthogonal Regression Procedure Step 1 and 2

File: Correlation Example.jmp

We'll assume the measurement errors are equal (similar testers).

Step 1

Analyze > Fit Y by X

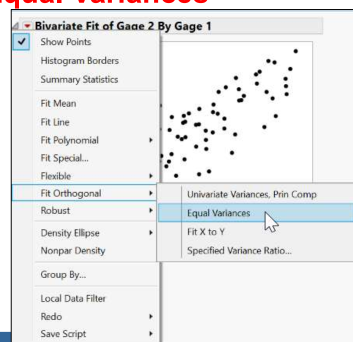


Step 2

▼ Bivariate Fit of Gage 2 by Gage 1

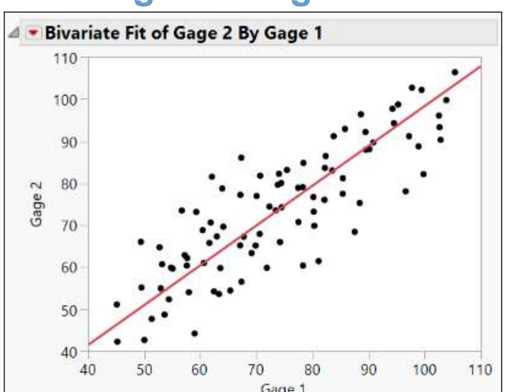
> Fit Orthogonal

> Equal Variances



15

Orthogonal Regression Procedure



Orthogonal Fit Ratio=1.000

Orthogonal Fit Ratio=1.000

Variable	Mean	Std Dev	Variance Ratio	Correlation
Gage 1	73.88944	15.82984	1	0.8426
Gage 2	73.64089	15.13755		

Intercept	Slope	LowerCL	UpperCL	Alpha
3.570232	0.948318	0.827121	1.086201	0.05000

Correlation = R

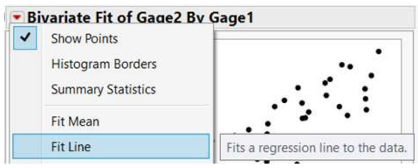
$R^2 = (0.8426)^2 = 0.71$

16

Orthogonal Regression Procedure Step 3

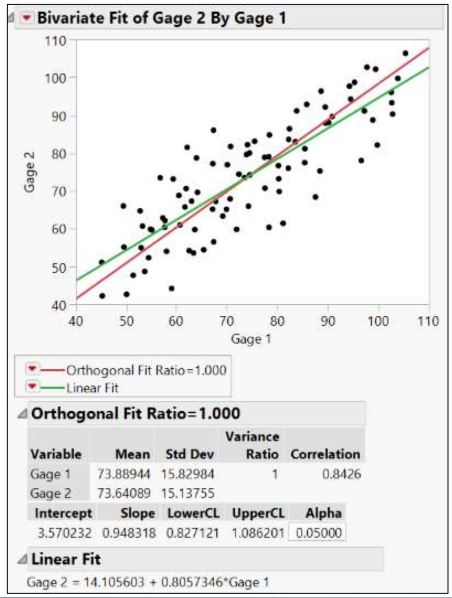
Compare to the linear fit:

▼ Bivariate Fit of Gage 2 by Gage 1
> Fit Line



Slope of the

- **OLS line = 0.806**
- **OR line = 0.948**

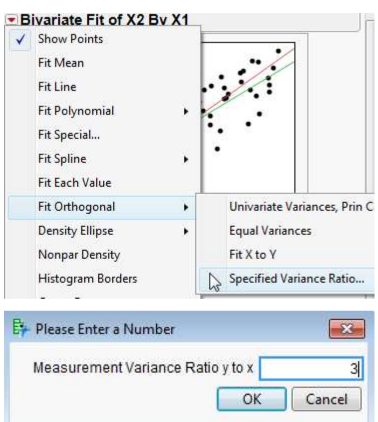


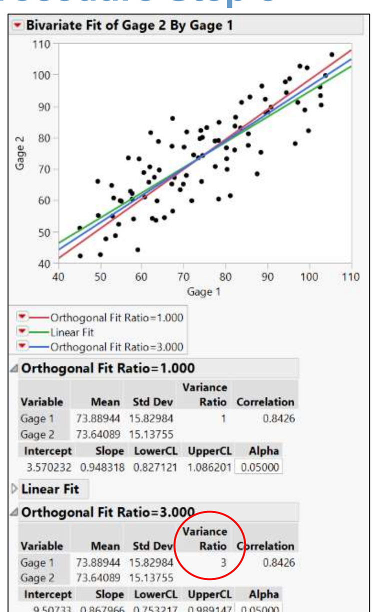
17

Orthogonal Regression Procedure Step 3

Unequal variances

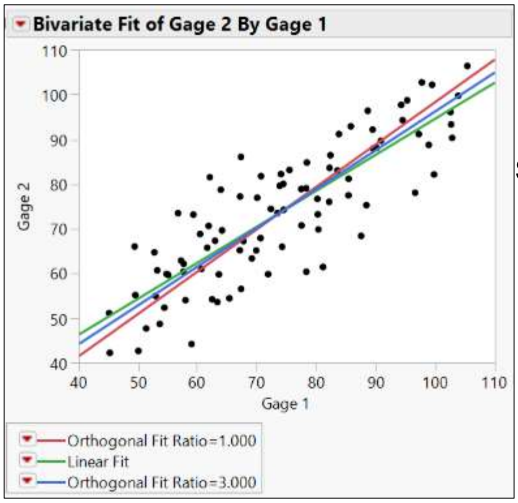
Try it with the variance in X is 2 and in Y is 6: Ratio of 6/2=3





18

Orthogonal Regression Procedure Step 3



Slope of the

- OLS line ratio 1) = 0.806
- OR line = 0.948
- OR line (ratio 3) = 0.867

19

Orthogonal Regression Procedure Step 4

Even though we have a more “correct” equation for the relationship between Gage 1 and Gage 2, practically, we really cared whether we could say that the testers are “equal,” that is Gage 1 = Gage 2.

To do this, we need to do two more steps:

H_O: Slope = 1
H_A: Slope $\neq$ 1

1. Could the slope = 1.0?

Look at the Confidence Interval for the slope (b) in the previous output – if it contains the value 1.0, statistically the slope could be 1.0. Continue to Step II.

H_O: Offset = 0
H_A: Offset $\neq$ 0

2. Is there an offset in the paired data?

Using the paired t comparison to calculate the confidence interval on the difference: if it contains 0, then we can “assume” there is no offset; otherwise we accept the offset given.

20

Orthogonal Regression Procedure Step 4

1. Could the slope be 1.0?

Yes – Confid Int contains value 1. Assume Y = X

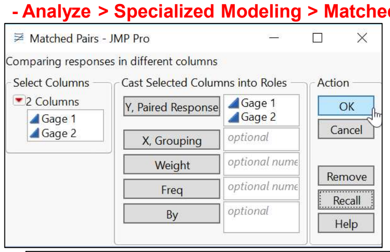
Orthogonal Fit Ratio=1.000

Variable	Mean	Std Dev	Variance Ratio	Correlation
Gage 1	73.88944	15.82984	1	0.8426
Gage 2	73.64089	15.13755		

Intercept	Slope	LowerCL	UpperCL	Alpha
3.570232	0.948318	0.827121	1.086201	0.05000

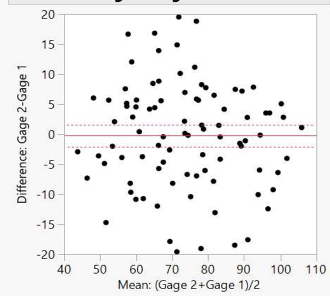
2. Is there an offset in the paired data?

- Analyze > Specialized Modeling > Matched Pairs



Matched Pairs

Difference: Gage 2-Gage 1



Gage 2	73.6409	t-Ratio	-0.27062
Gage 1	73.8894	DF	89
Mean Difference	-0.2486	Prob > |t|	0.7873
Std Error	0.91846	Prob > t	0.6063
Upper 95%	1.5764	Prob < t	0.3937
Lower 95%	-2.0735		
N	90		
Correlation	0.84258		

No.

Final answer: line is Gage 2 = Gage 1

21

Summary

- **Orthogonal Regression is preferable over Ordinary Least Squared Regression for gage to gage correlations**
- **JMP has a simple method as part of the Fit Y by X**

22