M. Singh

Lesson 15: How can we proceed with linear regression analysis when the scatter is not 'straight enough'?

Q: Answer the following questions:

1. Are the variables quantitative? Units?

2. Identify the explanatory and response variables?

3. Describe the relationship between the variables as visible from a scatterplot. Discuss direction, strength, form, and any unusual point

4. Does the scatterplot indicate that it is appropriate to calculate the correlation? Explain.

5. Run regression and view residuals plot. Is the linear model appropriate? Explain.

| | | Population |
|-----|------|------------|
| | Year | (millions) |
| | 1800 | 5 |
| | 1825 | 11 |
| | 1850 | 23 |
| ts. | 1875 | 44 |
| | 1900 | 76 |
| ; | 1925 | 114 |
| | 1950 | 151 |
| | 1975 | 215 |
| | 2000 | 285 |

What are the reasons for re-expressing data:





M. Singh

2. To make the spread across different groups more similar



after re-expressing by logs





4. To make the scatter around the line in a scatterplot more consistent



Ladder of powers

| Power | Name | Comment |
|-------|-----------------|--------------------------------------------|
| 2 | Square of | Try with unimodal distributions that are |
| | data values | skewed to the left. |
| 1 | | Data with positive and negative values |
| | Raw data | and no bounds are less likely to benefit |
| | | from re-expression. |
| 1⁄2 | Square root of | Counts often benefit from a square root |
| | data values | re-expression. |
| "0" | We'll use | Measurements that cannot be negative |
| | logarithms here | often benefit from a log re-expression. |
| -1/2 | Reciprocal | An uncommon re-expression, but |
| | square root | sometimes useful. |
| -1 | The reciprocal | Ratios of two quantities (e.g., mph) often |
| | of the data | benefit from a reciprocal. |

Let's try re-expressing our initial scatter.

Words of Wisdom

- Don't expect your model to be perfect.
- Don't stray too far from the ladder.
- Don't choose a model based on R² alone:
- Re-expression can straighten many bent relationships, but not those that go up then down, or down then up.

