

Stat 101 – Lecture 14

Residual Standard Deviation

$$s_e = \sqrt{\frac{\sum (y - \hat{y})^2}{n - 2}}$$

- We divide by $n - 2$ because we have estimated two quantities, the slope and the intercept.

1

$(r)^2$ or R^2

- The square of the correlation coefficient gives the amount of variation in y , that is accounted for or explained by the linear relationship with x .

2

Tar and Nicotine

- $r = +0.956$
- $(r)^2 = (0.956)^2 = 0.914$ or 91.4%
- 91.4% of the variation in nicotine content can be explained by the linear relationship with tar content.

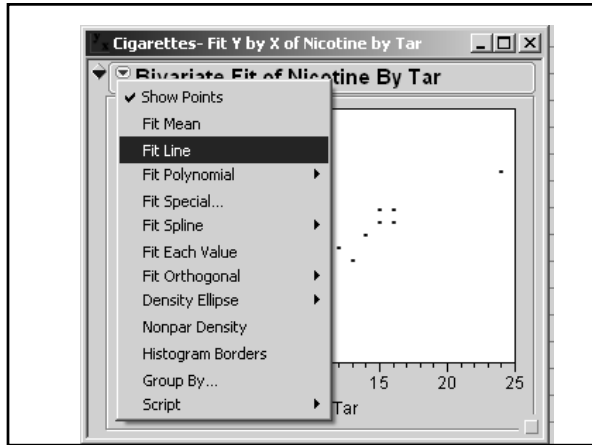
3

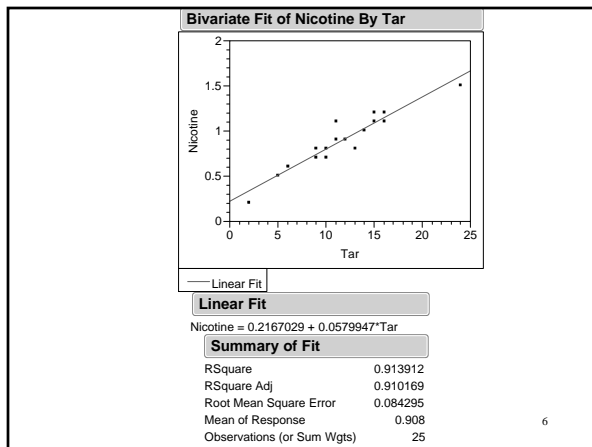
Stat 101 – Lecture 14

JMP

- Analyze – Fit Y by X
- Y, Response
 - Nicotine
- X, Factor
 - Tar

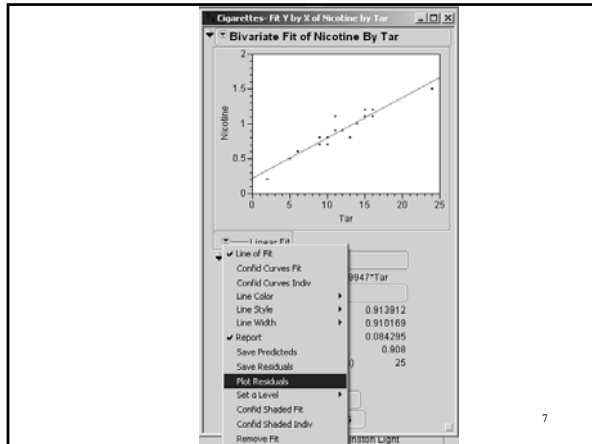
4





6

Stat 101 – Lecture 14



Regression Conditions

- Quantitative variables – both variables should be quantitative.
- Linear model – does the scatter plot show a reasonably straight line?
- Outliers – watch out for outliers as they can be very influential.

8

Regression Cautions

- Beware of extraordinary points.
- Don't extrapolate beyond the data.
- Don't infer x causes y just because there is a good linear model relating the two variables.
- Don't choose a model based on R^2 alone.

9
