Stat 101L: Lecture 13

Prediction

- Least squares line

$\hat{y} = \mathbf{0.217} + \mathbf{0.058}x$ for $x = \mathbf{13}$

$\hat{y} = \mathbf{0.217} + \mathbf{0.058}(\mathbf{13})$

$\hat{y} = \mathbf{0.97}$ mg

1

Residual

- Tar, $x = 13$ mg
- Nicotine, $y = 0.8$ mg
- Predicted, $\hat{y} = 0.97$ mg
- Residual, $y - \hat{y} = 0.8 - 0.97$

$= -0.17$ mg

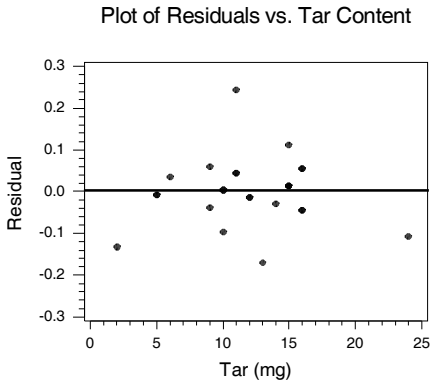
2

Residuals

- Residuals help us see if the linear model makes sense.
- Plot residuals versus the explanatory variable.
  - If the plot is a random scatter of points, then the linear model is the best we can do.

3

Stat 101L: Lecture 13



Interpretation of the Plot

- The residuals are scattered randomly. This indicates that the linear model is an appropriate model for the relationship between tar and nicotine content of cigarettes.

5

$(r)^2$ 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*.

6

Stat 101L: Lecture 13

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.

7

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