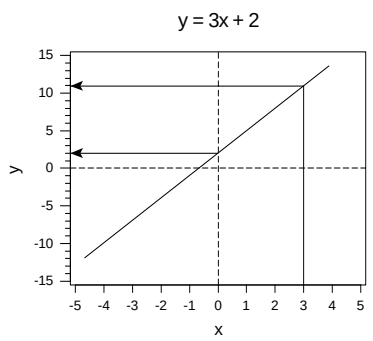
Stat 101 – Lecture 12

Algebra Review

- The equation of a straight line
- $y = mx + b$
  - m is the slope – the change in y over the change in x – or rise over run.
  - b is the y-intercept – the value where the line cuts the y axis.

1



2

Algebra Review

- $y = 3x + 2$
  - $x = 0$ ⇨ $y = 2$ (y-intercept)
  - $x = 3$ ⇨ $y = 11$
  - Change in y (+9) divided by the change in x (+3) gives the slope, 3.

3

Stat 101 – Lecture 12

Linear Regression

- Example: Tar (mg) and nicotine (mg) in cigarettes.
  - y, Response: Nicotine (mg).
  - x, Explanatory: Tar (mg).
  - Cases: 25 brands of cigarettes.

4

Correlation Coefficient

- Tar and nicotine

$$r = \frac{\sum z_x z_y}{n-1} = \frac{+22.9437}{24}$$

- $r = +0.956$

5

Linear Regression

- There is a very strong positive linear association between tar and nicotine.
- What is the equation of the line that models the relationship between tar and nicotine?

6

Stat 101 – Lecture 12

Linear Model

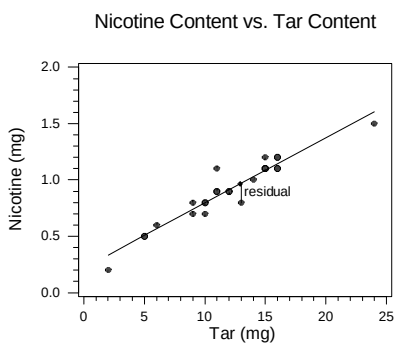
- The linear model is the equation of a straight line through the data.
- A point on the straight line through the data gives a predicted value of y, denoted $\hat{y}$.

7

Residual

- The difference between the observed value of y and the predicted value of y, $\hat{y}$, is called the residual.
- Residual = $y - \hat{y}$

8



9

Stat 101 – Lecture 12

Line of "Best Fit"

- There are lots of straight lines that go through the data.
- The line of "best fit" is the line for which the sum of squared residuals is the smallest – the least squares line.

10

Line of "Best Fit"

$$\hat{y} = b_0 + b_1 x$$

Least squares slope: $b_1 = r\dfrac{s_y}{s_x}$

intercept: $b_0 = \bar{y} - b_1\bar{x}$

11

Least Squares Estimates

Tar, x	Nicotine, y
$\bar{x} = 11.92\,\text{mg}$	$\bar{y} = 0.908\,\text{mg}$
$s_x = 4.636\,\text{mg}$	$s_y = 0.2812\,\text{mg}$
$r = 0.956$	

12