Let's model the association between protein and fat of McDonald's menu items by writing an equation for the line of best fit: $\hat{y} = b_0 + b_1 x$

1. $b_1 = ?$

$= \frac{r s_y}{s_x}$

$= \frac{(0.69)(11.6\text{g})}{13.2\text{g}}$

$= 0.61\text{g}$

2. $b_0 = ?$

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

$19.9 = b_0 + (0.61)(18.3)$

$b_0 = 8.7$

3. $\hat{y} = 2.7 + 0.61x$

$\hat{fat} = 8.7 + 0.61 \cdot protein$

Just as the mean summarizes a variable and the standard deviation tells how well. The regression line (line of best fit) summarizes the response variable in term of the explanatory variable and R^2 tells how well.

- We know choosing $m = r$ minimizes the sum of the squared residuals, but how small does that sum get? Equation (4) told us that the mean of the squared residuals is $1 - 2mr + m^2$. When $m = r$, $1 - 2mr + m^2 = 1 - 2r^2 + r^2$ $= 1 - r^2$. This is the percentage of variability NOT explained by the regression line. Since $1 - r^2$ of the variability is NOT explained, the percentage of variability in y that is explained by x is r^2. This important fact will help us assess the strength of our models.

$\vec{b_1} = \frac{r s_y}{s_x}$

$\vec{b_0}$ = plug in $(\bar{x}, \bar{y})$, and solve for (b_0)

And there's still another bonus. Because r^2 is the percent of variability explained by our model, r^2 is at most 100 %. If $r^2 \le 1$, then $-1 \le r \le 1$, proving that correlations are always between -1 and $+1$.

$\frac{rs_y}{s_x}$ $\frac{rs_y}{s_x}$

pick $\bar{x}, \bar{y}$ for your point

Residual If positive If negative	observed y − predicted value $\hat{y}$ Then the model makes an Then the model makes an
Regression line Line of best fit For standardized values For actual x and y values	The unique line that minimizes variance of residuals (sum of the squared residuals). $\hat{z}_y = r z_x$ ← Day 20 $\hat{y} = b_0 + b_1 x$ ← Day 21
To calculate the regression line in real units (actual x and y values)	1. Find slope; $b_1 = \frac{r s_y}{s_x}$ $\hat{y} = b_0 + b_1 x$ 2. Find y-intercept; plug b_1 and $P(x, y)$ (usually $\bar{x}, \bar{y}$) into $\hat{y} = b_0 + b_1 x$, solve for b_0 3. Plug in slope, b_1, and y-intercept, b_0, into $\hat{y} = b_0 + b_1 x$ (use
3 conditions needed for Linear Regression Models: /* same as correlation */	1. Quantitative Variables 2. Straight Enough – check original scatterplot & residual scatterplot 3. Outlier (clusters) –points with large residuals and/or high leverage
R^2	The square of the correlation r between x and y. The success of the regression model in terms of the fraction of the variation of y accounted for by the model. (differences in x explain xx% of the variability of y)