Warm Up:

Given \( f(x) = 4x + 1 \) and \( g(x) = x^2 - 2 \), which of the following is an expression for \( f(g(x)) \)?

- \( f(x^2 + 4x + 1) \)
- \( G. x^2 + 4x - 1 \)
- \( H. 4x^2 - 7 \)
- \( J. 4x^2 - 1 \)
- \( K. 16x^2 + 8x - 1 \)

3.2 Least-Squares Regression Day 1

Learning Goal: I will be able to interpret the slope and y-intercept of a least-squares regression line, use the least-squares regression line to predict y for a given x.

Basics:

A regression line is a line that describes how y changes when x changes.

We often use a regression line to predict y given x.

Interpreting Slope and Y-Intercept:

Slope: For every one "x" the "y" value goes up/down by "slope".

Y-Intercept: The "y" value when the "x" value is 0.

Example 1: Used Hondas

The following data shows the number of miles driven and advertised price for 11 used Honda CR-Vs from the 2002-2006 model years (prices found at www.carmax.com)

<table>
<thead>
<tr>
<th>Thousand Miles Driven</th>
<th>Cost (dollars)</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>17998</td>
</tr>
<tr>
<td>29</td>
<td>16450</td>
</tr>
<tr>
<td>35</td>
<td>14998</td>
</tr>
<tr>
<td>39</td>
<td>13998</td>
</tr>
<tr>
<td>45</td>
<td>14599</td>
</tr>
<tr>
<td>49</td>
<td>14988</td>
</tr>
<tr>
<td>55</td>
<td>13599</td>
</tr>
<tr>
<td>56</td>
<td>14599</td>
</tr>
<tr>
<td>69</td>
<td>11998</td>
</tr>
<tr>
<td>70</td>
<td>14450</td>
</tr>
<tr>
<td>86</td>
<td>10998</td>
</tr>
</tbody>
</table>

The scatterplot shows a linear negative relationship between miles and cost.

The correlation is -0.874.

The line on the plot is the regression line for predicting cost given miles.
Example 1 cont:
The regression line shown on the scatterplot is:
\[ y = 18773 - 86.18x \]
\[ \hat{y} = \text{predicted} \]
\[ y = \text{cost} \]
OR
\[ \text{cost} = 18773 - 86.18(\text{miles driven}) \]

a.) Identify the slope and y-intercept of the regression line. Interpret each value in context.
For every mile driven the cost decreases by $86.18. The cost when there are 0 miles is $18,773.

b.) Predict the asking price for a car with 50,000 miles.
\[ \hat{y} = 18773 - 86.18(50) \]
\[ \hat{y} = 14,464 \]

Example 1 cont:
c.) Should we predict the asking price for a used 2002-2006 Honda CR-V with 250,000 miles?
This is called **extrapolation**!!!!!! Extrapolation is risky business.

d.) Predict the asking price for a used 2002-2006 Honda CR-V with 45,000 miles. How far off is the prediction from the actual data?
\[ \hat{y} = 18773 - 86.18(45) \]
\[ \hat{y} = 14,849.40 \]
\[ \text{actual} = 14,599 \]
\[ \text{difference} = 249.40 \]

Example 1 cont:
e.) Show this difference on the graph:
The difference between actual and predicted is called a **residual**.

Analyzing Residuals:
You can tell a lot about the accuracy of the regression line by analyzing the residuals in the Honda CR-V data.

<table>
<thead>
<tr>
<th>Thousand Miles Driven</th>
<th>Cost (dollars)</th>
<th>Predicted Cost (dollars)</th>
<th>Residual (dollars)</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>17,998</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>16,450</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>14,998</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>13,998</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>14,599</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>14,988</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>13,999</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>14,599</td>
<td></td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>11,998</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>14,450</td>
<td></td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>10,998</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What is the mean of the residuals?
This graph is called a **residual plot**.
Two Important things:
To look for when you examine a residual plot:
1.) The residual plot should show **NO** obvious pattern. If it does than a linear prediction does **NOT** work.
2.) The residuals should be relatively small in size.

How do we know?
If the residuals are relatively small in size? Find the standard deviation of the residuals. The formula is:
\[
\sigma_x = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-2}} \quad \text{in other words:} \quad \sqrt{\frac{\sum(\text{residuals})^2}{n-2}}
\]

For the used Hondas data, square the residuals above, then complete the formulas.
The standard deviation is \( s = \sqrt{\frac{8460808}{11-2}} = \$472 \). So when we use the number of miles to estimate the asking price, we will be off by an average of \( \$472 \).

Summary
What is the linear regression formula and what does each letter stand for?
\[
\hat{y} = a + bx
\]
\[\hat{y}\] = y-int, \( b \) = slope

What does the standard deviation of the residuals tell us?
**Average amount we will be off by**

Coursework:
pg 191-192 # 35-42 all, 45 and 46