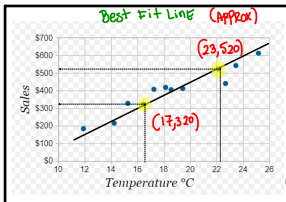
Lesson #9 (new) Week 4: Line of Best Fit

Project 1: Select two sport personalities from two different sports. Compare one specific statistic from each personality to get their corresponding z score. Reach a conclusion based on their z score.

ex | Soccer vs. Basketball
- Messi ✓ — M. Jordan ✓
- Avg. # of goals per game per season — Avg # of points per game per season
- $z = \frac{x - \mu}{\sigma}$ — $z =$
- μ: population
- x: Avg # of goals per game per season considering all players

Due: FRIDAY October 26, 2018

Oct 18-1:57 PM



Best Fit Line (Approx)

Step1 $y = mx + b$ find the equation of the Best Fit Line

Step2 Select two point. (17,320) and (23,520)
ex: x_1 y_1

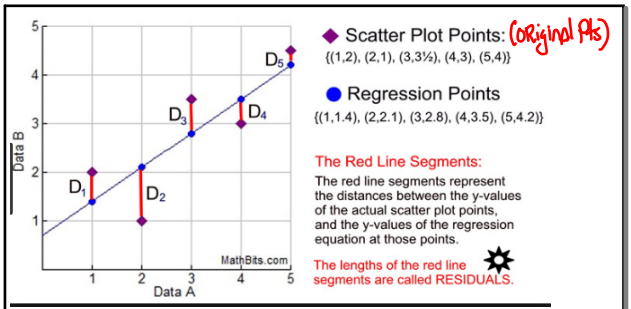
Step3 find m (slope) = $\frac{y_2 - y_1}{x_2 - x_1} = \frac{520 - 320}{23 - 17} = \frac{200}{6} = 33.\overline{3}$ ✓
b (y intercept)

$y = mx + b$ (slope/y int form)
$y - y_1 = m(x - x_1)$ (two point/slope form) ✓
$y - 320 = 33.\overline{3}(x - 17)$
$y = 33.\overline{3}(x - 17) + 320$ Best Fit Line (NO)
↳ What's the purpose of the equation? to predict values

ex | temp = 30° C (x)
Sales = 753 (y)

- Correlation Coefficient is based on Residuals

Oct 18-2:25 PM



◆ Scatter Plot Points: (original pts)
{(1,2), (2,1), (3,3½), (4,3), (5,4)}

● Regression Points
{(1,1.4), (2,2.1), (3,2.8), (4,3.5), (5,4.2)}

The Red Line Segments:
The red line segments represent the distances between the y-values of the actual scatter plot points, and the y-values of the regression equation at those points.

The lengths of the red line segments are called RESIDUALS.

Oct 18-2:51 PM

Facebook tracked how long users watched videos in order to set pricing for advertisers, but simply did not count views that lasted fewer than three seconds — effectively skewing its results to achieve higher average watch times.

} Deleting the smaller values

Oct 18-3:01 PM