

Agenda

- Warm Up 10 min
- Coefficient of variation 10 min
 - *Change examples, add more*
- Calculating regression lines 15 min
- Residuals and residual plots 20 min
 - *Bring up video*
- Exit Pass 5 min
- *Add → Bring course request forms Wed. 2/5*

Warm Up

Let's say your Math 3 grade and your grade in this class are positively correlated, with $r = \mathbf{0.72}$, and defined by the equation $\mathbf{y = 0.596x + 0.4113}$.

Briefly interpret the following:

1. Correlation coefficient
2. Coefficient of variation
3. Slope
4. Y-intercept

Course Request sheets

- Counselors available both lunches today, tomorrow, Thursday
- Important links:
 - Summary of each department → tinyurl.com/RCHS-catalog-summary
 - Full course descriptions → tinyurl.com/RCHS-catalog-2020
 - A-G catalog → tinyurl.com/RCHS-A-G-2020
- Make sure to:
 1. Sign, and get your parent to sign.
 2. Read prerequisites.
 3. Choose backups carefully. Don't leave them blank.
- Bring to class TOMORROW, so I can check it.
- Due to your period 2 teacher this Friday 2/7.
- Consider suggesting this class (Statistics) to people.
- I strongly recommend you take Senior Seminar (or AVID 12).

Coefficient of variation “ r^2 ”

- Also “coefficient of determination”
 - *Proportion of the variation in y that can be explained by x .*
 - *Example. Housing prices*
-

Interpret the coefficient of variation:

1. DOOR. When using nicotine consumption in milligrams per day to predict life expectancy.... $r^2=0.6182$
2. WINDOW. When using students' Math 3 grades to predict their final grades in AP Statistics.... $r^2=0.2833$
3. DOOR. When using tea consumption (in gallons per year) to predict the number of people killed annually in the United States by misusing a lawnmower.... $r^2=0.93$
4. WINDOW. When using goals scored by Lionel Messi for Argentina annually, to predict the earnings of each year's top-grossing Marvel movie.... $r^2=0.907$

Equation from slope and 1 point

- Warm Up #5

5. In the equation $y = mx + b$, if $m=3$, $x=4$, and $y=17$, what is b ?

Regression Line

The least-squares regression line is

$$y = ax + b$$

with slope

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

that passes through the point

$$(\bar{x}, \bar{y})$$

What's the regression line?

$$\bar{x} = 2482.5$$

$$s_x = 871.03$$

$$\bar{y} = 479692.86$$

$$s_y = 144797.13$$

$$r = 0.9414$$

| Size | Cost |
|------|---------|
| 1289 | 319000 |
| 1569 | 349000 |
| 1600 | 1300000 |
| 1600 | 1300000 |
| 1804 | 369000 |
| 1842 | 399000 |
| 1870 | 399800 |
| 2230 | 450000 |
| 2230 | 425000 |
| 2393 | 400000 |
| 2404 | 429900 |
| 2859 | 522000 |
| 3088 | 475000 |
| 3303 | 689000 |
| 3303 | 690000 |
| 4571 | 799000 |
| 5488 | 250000 |

What's the regression line? (EXAMPLE)

Given

$$\bar{x} = 2482.5$$

$$s_x = 871.03$$

$$\bar{y} = 479692.86$$

$$s_y = 144797.13$$

$$r = 0.9414$$

Slope

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

$$a = 0.9414 \frac{144797.13}{871.03}$$

$$= 156.49$$

Intercept

$$(\bar{x}, \bar{y}) \quad (2482.5, 479692.86)$$

$$y = ax + b \quad 479692.86 = (156.49)(2482.5) + b$$

$$b = 91193.5$$

Equation

$$y = 156.49x + 91193.5$$

Try it!

$$y = ax + b$$

Eighteen of you submitted data on your parents' heights. The mean height of parents was 65.868, with standard deviation 4.862. The mean height of students was 65.823, with standard deviation 4.618. The relationship was linear, positive and fairly strong, with $r=0.8237$. There were no outliers.

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

$$(\bar{x}, \bar{y})$$

You try! ANSWERS

Given

$$\begin{aligned}\bar{x} &= 65.868 \\ &= 4.862\end{aligned}$$

$$\begin{aligned}\bar{y} &= 65.823 \\ &= 4.618\end{aligned}$$

$$= 0.8237$$

Slope

$$\begin{aligned}a &= r \frac{s_y}{s_x} \\ &= 0.8237 \frac{4.618}{4.862} \\ &= 0.78236\end{aligned}$$

$$(\bar{x}, \bar{y})$$

$$(65.868, 65.823)$$

Intercept

$$y = ax + b$$

$$\begin{aligned}65.823 &= 0.78236(65.868) + \\ &14.2903 =\end{aligned}$$

Equation

$$= 0.78236 + 14.2903$$

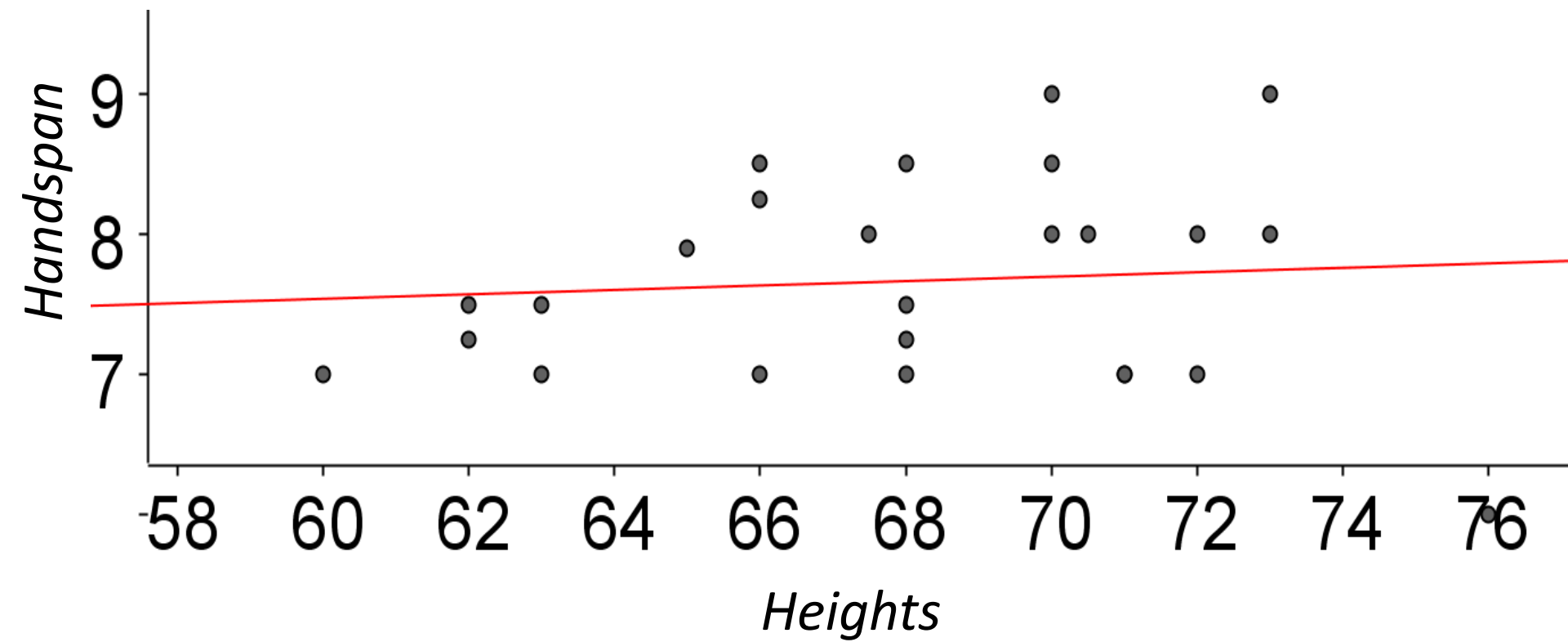
- 500 days of summer

Residuals and Residual Plots

- “Residual” = observed y – predicted y

$$= y - \hat{y}$$

- Vertical distance from point to line
- *Example: Hands and height.....*



$$y = 0.13x - 0.02$$

26. A linear regression was performing using the five following data points: A(1,11), B(5,2), C(3,7), D(7,1), E(9,-2). The residual for which of the five points has the largest absolute value?

A. A

B. B

- Enter into L1, L2

C. C

- Stat: LinReg $\rightarrow y = -1.6x + 11.8$

Residuals:

D. D

$$y = -1.6(1) + 11.8$$

$$\hat{y} = 10.2$$

$$\text{Residual} = 0.8$$

E. E

$$y = -1.6(5) + 11.8$$

$$\hat{y} = 3.8$$

$$\text{Residual} = -1.8$$

$$y = -1.6(3) + 11.8$$

$$\hat{y} = 7$$

$$\text{Residual} = 0$$

$$y = -1.6(7) + 11.8$$

$$\hat{y} = 0.6$$

$$\text{Residual} = 0.4$$

$$y = -1.6(9) + 11.8$$

$$\hat{y} = -2.6$$

$$\text{Residual} = 0.6$$

Residuals and Residual Plots

- “Residual” = observed y – predicted y
$$= y - \hat{y}$$
 - Vertical distance from point to line
 - *Example: Hands and height.....*
- Residual plot = Scatterplot (1st graph type)
 - X-axis → x-values
 - Y-axis → residuals (2nd, List, RESID)
- No pattern = good linear “fit” (no=good, yes=bad)

“Unscramble” Game (P.1)

- Groups of 4. *Dry erase marker & whiteboard.*
- I will give your group a problem.
- When everyone in your group has done the problem (*with work shown*), raise your hands.
- Each problem has a letter on the back. Unscramble all the letters to answer:

**Where am I going on vacation starting
Thursday June 4th (the first day of summer break)?**

- +5 extra credit to the winning group.

“Unscramble” Game (P.3)

- Groups of 3-4. *Dry erase marker & whiteboard.*
- I will give your group a problem.
- When everyone in your group has done the problem (*with work shown*), raise your hands.
- Each problem has a letter on the back. Unscramble all the letters to answer:

**What do I dislike more than anything
(as a personality characteristic)?**

- +5 extra credit to the winning group.

Exit Pass

Let's say that the relationship between "days you exercise per week" and "days since you last ate fast food" is positively correlated with $r=0.82$, and is defined by the least-squares equation $y = 4.82x + 3.21$.

1. Interpret the slope.
2. Interpret the y-intercept.
3. Interpret the coefficient of variation.