Yale-NUS College

YSC2239 Lecture 11

Recap

- Center and Spread
- Central Limit Theorem (CLT)

Python command: np.std, np.average

Today’s class

- Linear regression
- Method of least squares
- Residuals

- Reading: Chapter 15

Linear Regression

Linear Regression

A statement about x and y pairs

- Measured in *standard units*
- Describing the deviation of x from 0 (the average of x's)
- And the deviation of y from 0 (the average of y's)

On average, y deviates from 0 less than x deviates from 0

Regression Line

$$y_{(\mathrm{su})} = r \times x_{(\mathrm{su})}$$

Correlation

Not true for all points — a statement about averages

Slope & Intercept

Regression Line Equation

In original units, the regression line has this equation:

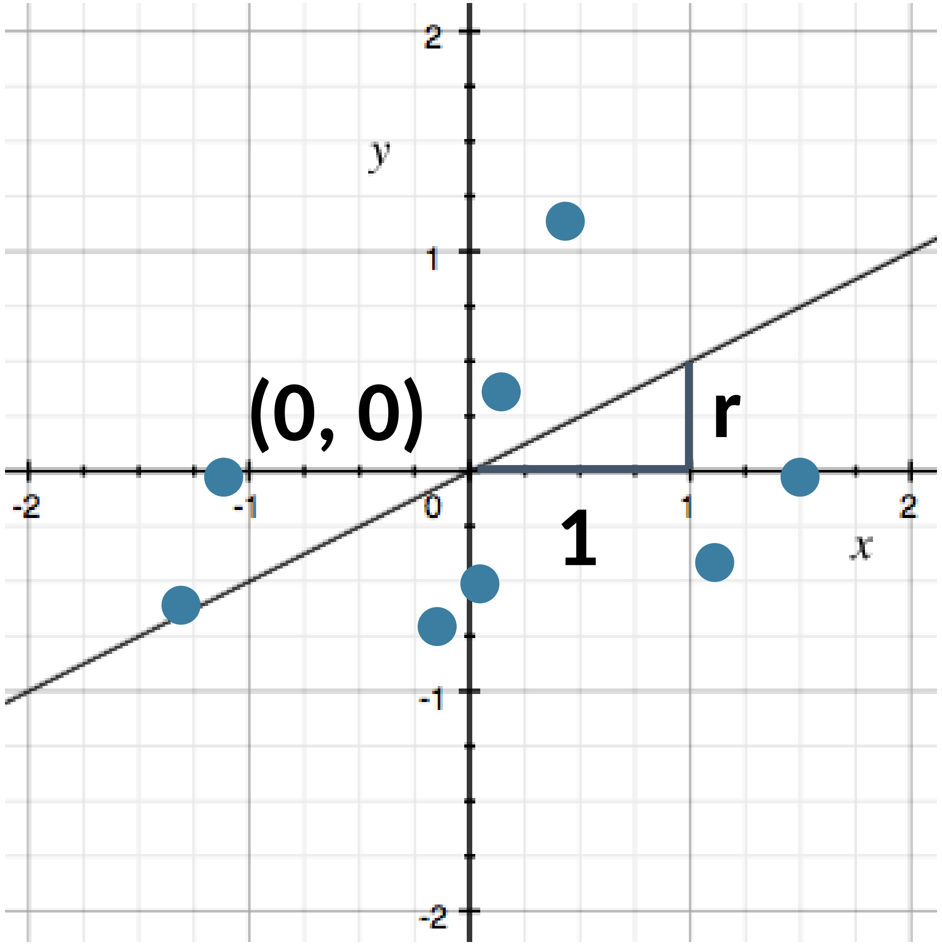
$$\underbrace{\frac{\text{estimate of } y \; - \; \text{average of } y}{\text{SD of } y}}_{\text{estimated y in standard units}} = r \times \underbrace{\frac{\text{the given } x \; - \; \text{average of } x}{\text{SD of } x}}_{\text{x in standard units}}$$

Lines can be expressed by *slope* & *intercept*

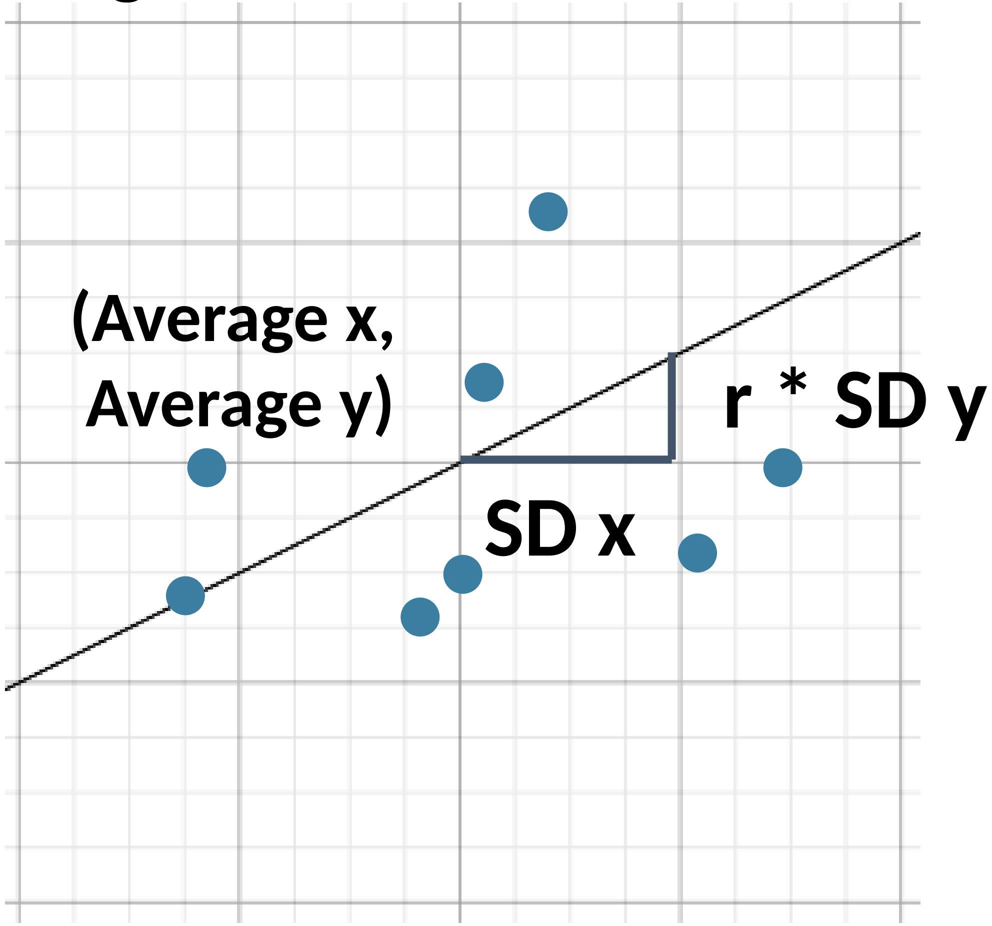
$$y = \text{slope} \times x + \text{intercept}$$

Regression Line

Standard Units



Original Units



Slope and Intercept

estimate of y = slope * x + intercept

$$\textbf{slope of the regression line} = r \cdot \frac{\text{SD of } y}{\text{SD of } x}$$

$$\textbf{intercept of the regression line} = \text{average of } y - \text{slope} \cdot \text{average of } x$$

(Demo)

Least Squares

Error in Estimation

- **error = actual value − estimate**
- Typically, some errors are positive and some negative
- To measure the rough size of the errors
  - **square** the **errors** to eliminate cancellation
  - take the **mean** of the squared errors
  - take the square **root** to fix the units
  - **root mean square error** (rmse)

(Demo)

Least Squares Line

- Minimizes the root mean squared error (rmse) among all lines
- Equivalently, minimizes the mean squared error (mse) among all lines
- Names:
  - “Best fit” line
  - Least squares line
  - Regression line

(Demo)

Numerical Optimization

- Numerical minimization is approximate but effective
- Lots of machine learning uses numerical minimization
- If the function `mse(a, b)` returns the mse of estimation using the line “estimate = ax + b”,
  - then `minimize(mse)` returns array $[a_0, b_0]$
  - a_0 is the slope and b_0 the intercept of the line that *minimizes* the mse among lines with arbitrary slope a and arbitrary intercept b (that is, among all lines)

(Demo)

Regression Diagnostics

Residuals

- Error in regression estimate
- One residual corresponding to each point (x, y)
- **residual**
  = observed *y* - regression estimate of *y*
  = observed y - height of regression line at x
  = vertical distance between the point and the best line

(Demo)

Residual Plot

A scatter diagram of residuals

- Should look like an unassociated blob for linear relations
- But will show patterns for non-linear relations
- Used to check whether linear regression is appropriate
- Look for curves, trends, changes in spread, outliers, or any other patterns

Properties of residuals

- Residuals from a linear regression **always** have
  - **Zero** mean
    - (so **rmse = SD of residuals**)
  - **Zero** correlation with x
  - **Zero** correlation with the fitted values

- These are all true **no matter what the data look like**
  - Just like deviations from mean are zero on average

(Demo)

A Measure of Clustering

Correlation, Revisited

- “The correlation coefficient measures how clustered the points are about a straight line.”

- We can now quantify this statement.

(Demo)

SD of Fitted Values

- $$\frac{\text{SD of fitted values}}{\text{SD of } y} = |r|$$

- SD of fitted values = $|r|$ * (SD of y)

Variance of Fitted Values

- Variance = Square of the SD

  = Mean Square of the Deviations
- Variance has weird units, but good math properties

- Variance of fitted values

$$\frac{\text{Variance of fitted values}}{\text{Variance of } y} = r^2$$

A Variance Decomposition

By definition,

$$y = \text{fitted values} + \text{residuals}$$

Tempting (but wrong) to think that:

$$\cancel{\mathrm{SD}(y) = \mathrm{SD}(\text{fitted values}) + \mathrm{SD}(\text{residuals})}$$

But it **is** true that:

$$\mathrm{Var}(y) = \mathrm{Var}(\text{fitted values}) + \mathrm{Var}(\text{residuals})$$

(a result of the **Pythagorean theorem!**)

A Variance Decomposition

Var(y) = Var(fitted values) + Var(residuals)

- $$\frac{\text{Variance of fitted values}}{\text{Variance of } y} = r^2$$

- $$\frac{\text{Variance of residuals}}{\text{Variance of } y} = 1 - r^2$$

Residual Average and SD

- The average of residuals is always 0

- $$\frac{\text{Variance of residuals}}{\text{Variance of } y} = 1 - r^2$$

- SD of residuals $= \sqrt{(1 - r^2)}$ SD of y

(Demo)

End of midterm coverage

To-do

- Assignment 5