# Programming in MATLAB

Dr. Mihail

August 28, 2018

# The Process

## What does programming (in any language) involve?

# The Process

## What does programming (in any language) involve?

- Analysis
  - Understanding the **problem**.
  - Figure out exactly the problem to be solved.
  - Clearly define the input(s) and output(s).
  - Solve by hand first. This step is useful to generate test data later.

# The Process

## What does programming (in any language) involve?

- Analysis
    - Understanding the **problem**.
    - Figure out exactly the problem to be solved.
    - Clearly define the input(s) and output(s).
    - Solve by hand first. This step is useful to generate test data later.
- Design
    - Formulate the overall structure of the program.
    - The "how" of the program gets worked out.
    - Develop your own algorithm through pseudocode.

# The Process

## What does programming (in any language) involve?

- Analysis
  - Understanding the **problem**.
  - Figure out exactly the problem to be solved.
  - Clearly define the input(s) and output(s).
  - Solve by hand first. This step is useful to generate test data later.
- Design
  - Formulate the overall structure of the program.
  - The "how" of the program gets worked out.
  - Develop your own algorithm through pseudocode.
- Implementation
  - Translate the design (pseudocode) into a computer language

# The Process

## What does programming (in any language) involve?

- Analysis
  - Understanding the **problem**.
  - Figure out exactly the problem to be solved.
  - Clearly define the input(s) and output(s).
  - Solve by hand first. This step is useful to generate test data later.
- Design
  - Formulate the overall structure of the program.
  - The "how" of the program gets worked out.
  - Develop your own algorithm through pseudocode.
- Implementation
  - Translate the design (pseudocode) into a computer language
- Testing
  - Try various inputs and verify correctness of inputs.
  - If problems found (process called debugging), fix them.

# Sample Problem

## Problem Setting

It is assumed that achievement test scores should be correlated with student's classroom performance. One would expect that students who consistently perform well in the classroom (tests, quizes, etc.) would also perform well on a standardized achievement test (0 - 100 with 100 indicating high achievement). A teacher decides to examine this hypothesis. At the end of the academic year, she computes a correlation between the students achievement test scores (she purposefully did not look at this data until after she submitted students grades) and the overall g.p.a. for each student computed over the entire year. The data for her class are provided next.

# Problem Data

| Achievement | GPA |
| --- | --- |
| 98 | 3.6 |
| 96 | 2.7 |
| 94 | 3.1 |
| 88 | 4.0 |
| 91 | 3.2 |
| 77 | 3.0 |
| 86 | 3.8 |
| 71 | 2.6 |
| 59 | 3.0 |
| 63 | 2.2 |
| 84 | 1.7 |
| 79 | 3.1 |
| 75 | 2.6 |
| 72 | 2.9 |
| 86 | 2.4 |
| 85 | 3.4 |
| 71 | 2.8 |
| 93 | 3.7 |
| 90 | 3.2 |
| 62 | 1.6 |

# Problem questions

## Questions

- What is the correlation coefficient?
- What does this statistic mean concerning the relationship between achievement test prformance and g.p.a.?
- What would be the slope and $y$-intercept for a regression line based on this data?
- If a student scored a 93 on the achievement test, what would be their predicted G.P.A.? If they scored a 74? A 88?

# Analysis

## Understanding the problem

- Questions worth asking:
  - What mathematical tool(s) do I need to solve the problem?

# Analysis

## Understanding the problem

- Questions worth asking:
    - What mathematical tool(s) do I need to solve the problem?
        - Linear functions: $y = f(x) = ax + b$

# Analysis

## Understanding the problem

- Questions worth asking:
    - What mathematical tool(s) do I need to solve the problem?
        - Linear functions: $y = f(x) = ax + b$
        - Sample correlation coefficient: $r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}}$. This correlation coefficient is an indicator of how well the data fits a model, in this case we **assume** our model is a linear function.

# Analysis

## Understanding the problem

- Questions worth asking:
    - What mathematical tool(s) do I need to solve the problem?
        - Linear functions: $y = f(x) = ax + b$
        - Sample correlation coefficient: $r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}}$. This correlation coefficient is an indicator of how well the data fits a model, in this case we **assume** our model is a linear function.
    - What MATLAB features will I make use of to solve the problem?

# Analysis

## Understanding the problem

- Questions worth asking:
  - What mathematical tool(s) do I need to solve the problem?
    - Linear functions: $y = f(x) = ax + b$
    - Sample correlation coefficient: $r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}}$. This correlation coefficient is an indicator of how well the data fits a model, in this case we **assume** our model is a linear function.
  - What MATLAB features will I make use of to solve the problem?
    - Mechanism to load and store data as floating point type vectors and scalars.

# Analysis

## Understanding the problem

- Questions worth asking:
    - What mathematical tool(s) do I need to solve the problem?
        - Linear functions: $y = f(x) = ax + b$
        - Sample correlation coefficient: $r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}}$. This correlation coefficient is an indicator of how well the data fits a model, in this case we **assume** our model is a linear function.
    - What MATLAB features will I make use of to solve the problem?
        - Mechanism to load and store data as floating point type vectors and scalars.
        - Simple arithmetic operations on vectors and scalars (exponentiation and summation)

# Analysis

## Understanding the problem

- Questions worth asking:
  - What mathematical tool(s) do I need to solve the problem?
    - Linear functions: $y = f(x) = ax + b$
    - Sample correlation coefficient: $r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}}$. This correlation coefficient is an indicator of how well the data fits a model, in this case we **assume** our model is a linear function.
  - What MATLAB features will I make use of to solve the problem?
    - Mechanism to load and store data as floating point type vectors and scalars.
    - Simple arithmetic operations on vectors and scalars (exponentiation and summation)
    - Visualization (plotting)

# Analysis

## Designing the algorithm

We seek a step by step algorithm (or recipe) to solve the problem.

# Analysis

## Designing the algorithm

We seek a step by step algorithm (or recipe) to solve the problem.

## Step 1

Where do we start?

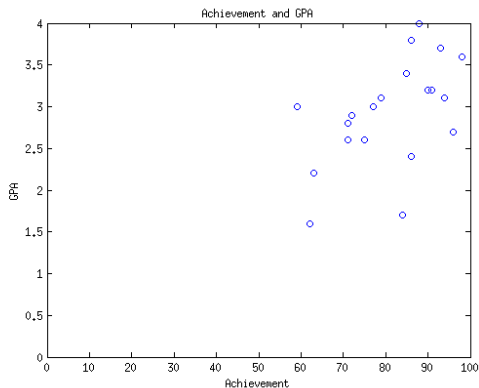# Analysis

## Designing the algorithm

We seek a step by step algorithm (or recipe) to solve the problem.

## Step 1

Where do we start?



Achievement and GPA

# Algorithm Design

## Design

1. Load data (excel spreadsheet)

# Algorithm Design

## Design

1. Load data (excel spreadsheet)

## Math

We now need to compute $r$:

$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}} \tag{1}$$

# Algorithm Design

## Design

1. Load data (excel spreadsheet)
2. Compute $r$

# Algorithm Design

## Design

1. Load data (excel spreadsheet)
2. Compute $r$

## Math

We now need to estimate $a$ and $b$ from $y = ax + b$ given the data we have. The data we have comes in pairs of $(x, y)$ values, **ASSUMED, or HYPOTHESIZED** to come from a linear generating function where $a$ is the slope and $b$ is the $y$-intercept.

# Algorithm Design

## Design

1. Load data (excel spreadsheet)
2. Compute $r$

## Math

We now need to estimate $a$ and $b$ from $y = ax + b$ given the data we have. The data we have comes in pairs of $(x, y)$ values, **ASSUMED, or HYPOTHESIZED** to come from a linear generating function where $a$ is the slope and $b$ is the $y$-intercept.

## Formulas

$$a = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{2}$$

$$b = \bar{y} - a * \bar{x} \tag{3}$$

# Algorithm Design

## Design

1. Load data (excel spreadsheet)
2. Compute $r$
3. Calculate $a$, the slope
4. Calculate $b$, the $y$-intercept
5. Plug in new values of $x$ and get the prediction:
   - $f(74)$ will predict GPA based on achievement score of 74
   - $f(88)$ will predict GPA based on achievement score of 88

# Let's write code

## Big Idea

We will write a script, which is self-contained, ordered set of MATLAB statements, in order to solve this problem.

## The first few steps

- Open MATLAB
- Create new script
- Write code! (solution is available to download)