Midterm Study Guide

Regulated Domains in Law (Module 1 – Pg. 31)

- **Purpose:** To achieve certain publicly desired goals that the market may fail to realize
- **Types:**
    - **Credit (Equal Credit Opportunity Act)**
        - Prohibits discrimination on the basis of race, color, religion, national origin, sex, marital status, age, receipt of public assistance, or good faith exercise of any rights under the Consumer Credit Protection Act.
    - **Education (Education Amendments of 1972; Civil Rights Act of 1964)**
        - **Education Amendments of 1972:** Prohibits sex (including pregnancy, sexual orientation, and gender identity) discrimination in any education program or activity receiving federal financial assistance
    - **Employment (Civil Rights Act of 1964)**
        - Provisions of this civil rights act forbade discrimination on the basis of sex, as well as, race in hiring, promoting, and firing.
    - **Housing and 'Public Accommodation' (Fair Housing Act; Civil Rights Act of 1964)**
        - Prohibits discrimination in the sale, rental or financing of housing or to otherwise interfere with someone's housing rights based on race, color, religion, sex, familial status, ancestry, disability, national origin or military status

Protected Classes (Module 1 – Pg. 33)

- **Race** (Civil Rights Act of 1964, 1991)
- **Color** (Civil Rights Act of 1964, 1991)
- **Sex** (Equal Pay Act of 1963; Civil Rights Act of 1964, 1991)
- **Religion** (Civil Rights Act of 1964, 1991)
- **National Origin** (Civil Rights Act of 1964, 1991)
- **Citizenship** (Immigration Reform and Control Act)
- **Age** (Age Discrimination in Employment Act of 1967) (over 40)
- **Pregnancy** (Pregnancy Discrimination Act)
- **Familial Status** (Civil Rights Act of 1968)
- **Disability status** (Rehabilitation Act of 1973; Americans with Disabilities Act of 1990)
- **Veteran status** (Vietnam Era Veterans' Readjustment Assistance Act of
- 1974; Uniformed Services Employment and Reemployment Rights Act)
- **Genetic information** (Genetic Information Nondiscrimination Act)

Ethics Framework – Consequence-Based vs. Rule-Based (Module 1 – Pg. 45-46)

- **Consequence-Based Ethics**
  - Priority given to choices that lead to a "good" outcome
  - Outcome outweighs the selected method
  - **Utilitarian View:** the "right choice" delivers the greatest good to the most people
  - **Individualism View:** the "right choice" is best for long-term self-interest
- **Rule-Based/Justice View**
  - Priority given to following the rules without undue regard to the outcome
  - Rules are often thought to codify principles like: truthfulness, right to freedom, justice, etc.
  - **Moral-Rights View:** the "right choice" is impartial, fair, and equitable treatment of people. Exists for the benefit of society and should be followed

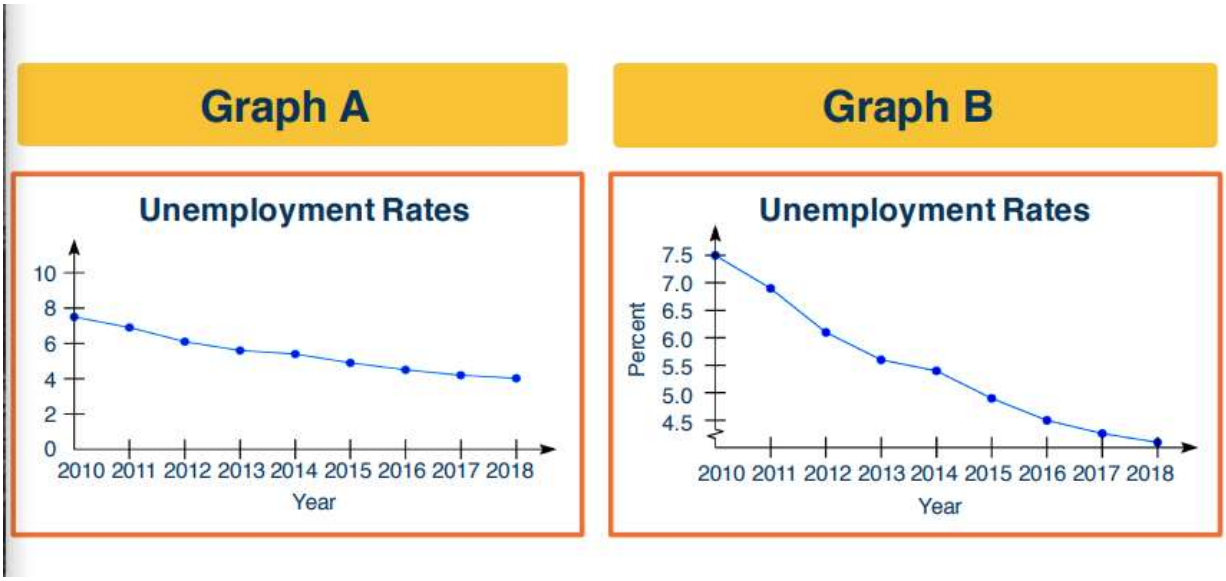
Information Privacy – Acts (Module 1 – Pg. 54-56)

- **Information Privacy:** the relationship between the collection and dissemination of data, technology, the public expectation of privacy, legal, and political issues surrounding them
- **Financial Services Modernization Act, or Gramm-Leach-Bliley Act of 1999**
  - Requires due notice by financial organizations to customers so that they can request that their information not be shared with third parties
  - Ensures that the organization's privacy policies are fully disclosed (and distributed annually) when a customer initiates a business relationship
- **Federal Privacy Act of 1974**
  - Regulates the government in the protection of individual privacy
- **Electronic Communications Privacy Act of 1986**
  - Regulates the interception of wire, electronic, and oral communications
- **Privacy of Customer Information Section of the common carrier regulation**
  - Specifies that any proprietary information shall be used explicitly for providing services, and not for any marketing purposes
  - Stipulates that carriers cannot disclose this information except when necessary to provide their services
- **Health Insurance Portability and Accountability Act of 1996 (HIPAA)**
  - Severely restrict the dissemination and distribution of private health information by organizations without documented consent.
  - Provides patients the right to know who has access to their information, and who has accessed it

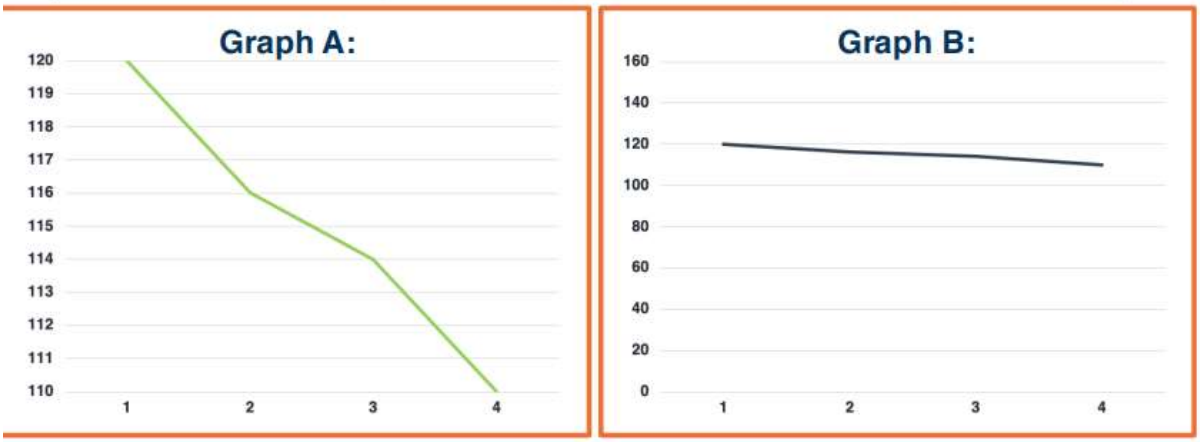
Sources of Data Bias / Fairness (Module 1 – Pg. 90-102)

- **Data Inputs:**
  - Incomplete, incorrect, or outdated
  - Poorly Selected and Selected with bias
  - Promotes historical biases
- **Algorithmic Processing:**
  - Decision making systems that assume correlation implies causation
  - Algorithms that don't compensate for datasets that disproportionately represent populations
  - Output models that are hard to understand or explain
- **Collection of Data**
  - Demographic, geographic, behavioral, temporal biases
- **Measurement of Data**
  - What do we choose to measure? How do we measure data?
- **Pre-existing Biases in Data**
  - Gender roles in text and images, racial stereotypes in language
  - **Unintentional**
    - Limited and coarse features
    - Sample size disparity
      - Less data (by definition) about minority populations
    - Skewed sample
      - Feedback loops
    - Tainted examples
    - Features that act as proxies
  - **Intentional**
    - Conscious prejudice
- **Algorithmic Fairness**
  - **Related to Accountability:** How do we supervise/audit machines which have large impact?
  - **Related to Transparency:** Why does an algorithm behave in a certain way? Can we understand its decisions? Can it explain itself?
  - **Related to AI safety:** How can we make AI without unintended negative consequences? Aligns with our values?

Misleading Graphs (Module 2 – Pg. 12-24)



Year	1	2	3	4
# of students matriculating into CS	120	116	114	110



Average – Mean, Median, Mode (Module 2 – Pg. 61-65)

- **Average/Mean**
  - The arithmetic average (sum of scores / # of scores)
  - **Example:** Mean of 20, 30, 40 is (20+30+40)/3 = 30
- **Median**
  - The middle value in an ordered sequence of observations
  - Examples:

        - Median of {9, 3, 6, 7, 5}
            - Sort data -> {3, 5, 6, 7, 9}
            - The median is 6
        - If the number of observations is even, e.g., {9, 3, 6, 7, 5, 2}, then the median is the average of the two middle values from the sorted sequence, in this case, (5 + 6) / 2 = 5.5
- **Mode**
    - The most frequently occurring number (score, measurement, value, cost)

Frequency Distribution (Module 2 – Pg. 71-72)

- Number of times a data item occurs
- **Cumulative Frequency Distribution:** finding the "running total" of frequencies.
    - Helps show the total number of data items at different stages in the data set

Distribution of Age:

Age	1	2	3	4	5	6
Frequency	5	3	7	5	4	2
Cumulative Frequency	5	8	15	20	24	26

Grouped Distribution of Age:

Age Group	1-2	3-4	5-6
Frequency	8	12	6
Cumulative Frequency	8	20	26

https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch10/5214862-eng.htm

Standard Deviation / Quartiles (Module 2 – Pg. 76-77)

- **Standard Deviation:** Square root of the variance. For the example below the answer is √4 = **2**

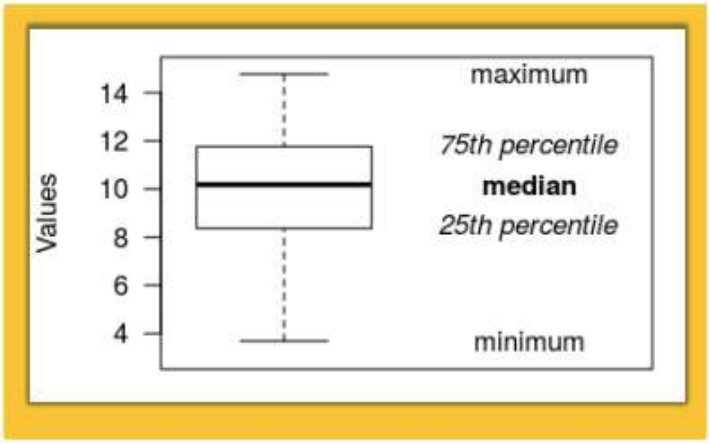
Variance: The variance of a set of observations is the average of the squares of the deviations of the observations from their mean. In symbols, the variance of the n observations x1, x2,…,xn is

$$S^2 = \frac{(x_1 - \bar{x})^2 + ... + (x_n - \bar{x})^2}{n-1}$$

Variance of 5, 7, 3? **Mean** is (5+7+3)/3 = 5 and the **variance** is

$$\frac{(5-5)^2 + (3-5)^2 + (7-5)^2}{3-1} = 4$$

- **Quartiles:** Data can be divided into four regions that cover the total range of observed values. Cut points for these regions are known as quartiles
    - The first quartile (Q1) is the first 25% of the data. The second quartile (Q2) is between the 25th and 50th percentage points in the data. The upper bound of Q2 is the median. The third quartile (Q3) is the 25% of the data lying between the median and the 75% cut point in the data.



Other Notes:

Autonomous Vehicles Algorithms

- **Humanist Algorithm:** The humanist algorithm shares the damage among all the people involved in the crash. If possible, deaths are avoided. Less damage is inflicted to the weakest group of a given population, which is the most vulnerable to physical trauma.
- **Protectionist Algorithm:** Attempts to protect the vehicle and its driver at all costs. This algorithm is safe from a user perspective, but unsafe for everyone else because it dismisses everything and everyone to preserve the user's safety.
- **Profit-based Algorithm:** New insurances policies are emerging for autonomous vehicle crashes. Chooses the best possible outcome depending on the user's insurance policy. Also favors famous people.