**Stats – Chapter 6**

**Probability distributions of Discrete Random Variables**

Definitions:

* **Random variable**

A RV is a numerical measurement of the outcome of a random phenomenon

**Notation:**

* + Use a capital letter for the variable
  + Use the same letter but lowercase for the **values** of the variable
  + E.g., X is heads in 3 tosses ; x = 2 is a possible value

The probability distribution of a RV specifies:

* + **Its possible values**
  + Diagram

    Description automatically generated**Their probabilities**

Text

Description automatically generatedTable

Description automatically generated

***Mean (or Expected Value) of a Probability Distribution***

Graphical user interface, text, application

Description automatically generatedThe mean μ of the probability distribution for that random variable is the value we would get if we repeatedly observe its outcome, in the long run, for the average of those values (as seen before)

This mean is aka **weighted average**, as the outcomes are not equally distributed

***Variance (*σ^2*) and Standard Deviation (*σ*)***

The standard deviation of a probability distribution, denoted by σ, measures the variability from the mean

**Text, letter

Description automatically generated**Larger values for σ correspond to greater variability

**Probability distributions of Continuous Random Variables**

Text

Description automatically generatedA RV is **continuous** when its possible values form an interval (rather than distinct values)

Chart, line chart

Description automatically generatedText, letter

Description automatically generated

**Properties and comments**

* The probability of a single value is 0
* The interval can be infinite

f(x) is **not a probability** (i.e. it can be > 1)

**Text

Description automatically generatedProbabilities of Bell-Shaped Distributions**

The formula for the z -score is useful when we are given the value of x for some normal random variable and need to find a probability relating to that value

We convert x to a z -score and then use a normal table to find the appropriate probability

**A standard normal distribution is the normal distribution with σ = 1 and μ = 0**

When a random variable has a normal distribution and its values are converted to z-scores by subtracting the mean and dividing by the standard deviation, the z-scores have the standard normal distribution

**Finding Probabilities for the Normal Distribution**

* **Cumulative probability**

Histogram

Description automatically generated

For the standard normal distribution, the cumulative (left tail) probabilities are in a table (at the end of the book, Appendix A, p. 785)

**For any data set having approximately a bell-shaped distribution:**

* roughly **68%** of the observations lie within **one standard deviation** to either side of the mean
* roughly **95%** of the observations lie within **two standard deviations** to either side of the mean
* Text

  Description automatically generatedroughly **99.7%** of the observations lie within **three standard deviations** to either side of the mean

Text

Description automatically generated**Probabilities When Each Observation Has Two Possible Outcomes**

Graphical user interface

Description automatically generated with low confidence**Probability of a Binomial Distribution**

(probability of *x* successes)

Example : Find the probability of having 3 once rolling a die twice (using Binomial)

* n = 2 independent trials
* in each trial we have 3 (success) or not 3 (failure)
* p = 1/6 is the probability of success in a single trial
* A picture containing text, clock, watch, gauge

  Description automatically generatedx = 1 number of successes

The shape of the distribution will depend on *n* and *p*, this is because we have that:

* **Mean = n\*p**
* **Standard deviation = sqrt(n\*p (1-p))**
* **Text, letter

  Description automatically generatedVariance = n\*p (1-p)**

A picture containing timeline

Description automatically generated