6.1 Summarizing Outcomes and Their Probabilities

- <u>random variable</u> : a numerical measurement of the outcome of a random phenomenon.

- <u>notation</u> : use lower case letters at the end of the alphabet for possible values (i.e. x) and upper case for the random variable (X).

- <u>ex</u> : $X = \#$ of heads in 4 flips of a fair coin. $x = 2$ and example of a possible value.

- <u>probability distribution</u> : of a random variable specifies its possible values and their probabilities.

- <u>Discrete Random Variables</u>

  1. take on discrete (separate and distinct) values.
  2. probability distribution specifies a probability for each value.
  3. $0 \leq P\{X = x\} \leq 1$ for all x
  4. $\displaystyle\sum_x P\{X = x\} = 1$
  5. <u>ex</u> : $X = $ a random digit selected using software (use in experimental design). possible values are $0, 1, 2, \ldots, 9$ w/ probability distribution

  $$P\{X = 0\} = P\{X = 1\} = \cdots = P\{X = 9\} = 0.10$$

  . is X a random variable?
  6. <u>ex</u> : $X = \#$ of home runs in a game by the Red Sox.

i	$P\{X = i\}$
0	0.23
1	0.38
2	0.22
3	0.13
4	0.03
5	0.01
6+	0.00

  (a) is X a random variable?
  (b) what is the probability of at least 3 home runs?

- the mean of a probability distribution

  1. <u>recall</u> : numerical summaries of a population are called parameters.
  2. numerical summaries of probability distributions are also called parameters.
  3. we can think of probability distributions representing populations.
  4. we use μ to represent the population mean.
  5. we use σ to represent the population standard deviation.

6. suppose we repeatedly observe values of a random variable (e.g. rolling a die), the mean (μ) of the probability distribution for that random variable is the value we would get in the long run.

7. population mean or expected value

$$\mu = \sum_x xP\{X = x\} = E[X]$$

8. μ also called a weighted average.

9. <u>ex</u> : $X = \#$ of home runs (Red Sox). find $E[X]$.

10. <u>ex</u> : Risk Taking.
   (a) we are given $1000 to invest
      i. sure gain of $500
      ii. gain of $1000 w/prob 0.5, gain of $0 w/prob 0.5
   (b) we are given $2000 to invest
      i. sure loss of $500
      ii. loss of $1000 w/prob 0.5, loss of $0 w/prob 0.5
   (c) find the expected gain/loss in each scenario.

• summarizing spread

1. <u>standard deviation</u> : describes how far the random variable falls, on average, from the mean (μ) of the distribution.

2. why do we care? e.g. investment strategies may have the same expected value (payoff), variability makes a big difference.

3. <u>ex</u> : risk taking redux
   (a) what is the standard deviation of the first scenario?

• probability distribution for categorical variables

1. remember that X is numerical, therefore it does not translate well to categorical variables.

2. later we'll see that assigning values 0 and 1 to a categorical variable w/ two categories can be helpful.

• Probability Distributions for continuous r.v.

1. a random variable is continuous when its values form an interval.

2. in the case of continuous random variables, we assign values to intervals. e.g. $X =$ travel time on the bus, then we have $P\{X < 15\}$ or $P\{10 < X < 15\}$.

3. take the histogram as an example. we separate the range of values for a continuous r.v. into intervals, for wait time we might have (0 to 30, 30 to 60, 60 to 90, ...). these are wide intervals, we could take the intervals (0 to 1, 1 to 2, 2 to 3, ...). as the intervals get smaller, the histogram approaches a smooth curve. these smooth functions are what we will use as probability distributions for continuous r.v.'s

6.2 Bell Shaped Distributions

- <u>normal distribution</u> : commonly used distribution used for continuous random variables. characterized by a particular bell shaped curve w/ mean μ and standard deviation σ. We usually use the notation $N(\mu, \sigma)$.

- take a look at density plots for male/female height.

  1. the peak of the plot appears at the mean.
  2. entire plot for women appears more pinched, making the peak taller.
  3. area under the curve represents probability within that range.

- next picture. explination of empirical rule.

- <u>cumulative probability</u> : represents area under the curve below the point $\mu + z\sigma$.

- explination of cummulative probability table. table represents the cummulative probabilities for the standard normal distribution.

- <u>standard normal distribution</u> has $\mu = 0$ and $\sigma = 1$.

- examples

  1. <u>ex 7</u> Mensa is a society of high IQ people. members score at or above the 98th percentile of a specific IQ test. if $X =$ the IQ score, then $X \sim N(100, 16)$
     (a) how many st dev above the mean is the 98th percentile?
     (b) what is the IQ score for the 98th percentile?

  2. <u>ex 8</u> find relative standing on SAT. let $X =$ SAT score. then we are told that $X \sim N(500, 100)$.
     (a) you scored 650 on the SAT, how many st deviations was it from the mean?
     (b) what percentage of scores was higher than yours?

  3. <u>ex 9</u> let $X =$ score on the midterm. the instructor gives B's for students scoring between 80 and 90. if $X \sim N(83, 5)$, what percentage of the class gets a B on the exam?

  4. <u>ex 10</u> we would like to compare scores on two different exams. we are given that $SAT \sim N(500, 100)$ and $ACT \sim N(21, 4.7)$. you scored a 650 on the SAT, and your friend scored a 30 on the ACT.
     (a) who is smarter?
     (b) what percentiles did you each score in?