Unit 8: Statistics and Probabilities

8-1: Sampling Techniques

Survey: - an exercise to ask a group of people about their responses to issues / products / preferences.

Population: - the group of people who will be affected by the result of the survey.

Census: - asking the survey questions to the ENTIRE population.

Sample: - asking the survey questions to a PART of the population.

Sample Size: - the size of the sample compared to the rest of the population (about between 10% to 20%).

Example 1: Identify the population and explain whether the event (survey) is conducted from a sample or as a census.

a. Filling out the Student Information Demographic Forms.

Population: All students in a school

Because all students have to complete the form, it is a **census**.

b. Determining whether the Canadian federal government should increase military spending.

Population: All Canadian citizens who can legally vote.

Because the size of the population involved, a **sample** should be used.

c. Electing the Provincial Government of Alberta

Population: All Alberta citizens who can legally vote.

Because it is an election, it should be a **census**.

d. Deciding whether Star Trek is better than Star Wars.

Population: All sci-fi fans that watches both Star Trek and Star Wars.

Because there are so many Star Wars and Star Trek fans, it can be done as a **sample**.

Probability Sampling: - sampling where a random selection of the population is used.

a. **Simple Random Sampling**: - the sample is chosen randomly with a pre-determined sample size.

b. **Systematic Sampling**: - the sample is taken from every n^{th} element (units) of the population.

c. **Stratified Sampling**: - the population is divided into appropriate groups (stratifications), and then a random set of sample is taken from each group.

d. **Clustered Sampling**: - the population is divided into groups based on the nature of the survey questions and only a specific group (cluster) relevant to the question is sampled randomly.

Non-Probability Sampling: - sampling where a NON-RANDOM selection of the population is used.
- it is important to identify these surveys as they make their results invalid.

a. **Sampling of Volunteers**: - the sample is chosen such that the favourable response will intentionally be resulted.

b. **Convenient Sampling**: - the sample is taken from a place just because it is the surveyor closest position.

Example 2: A survey needs to be designed to determine all Calgary Catholic Schools' Students should wear uniforms. Identify the population and classified the sampling techniques from the list below.

Population: All Catholic Schools Students, Parents, and Teachers.

Strategies	Sampling Techniques
The surveyor sends out a questionnaire with the Catholic School Board Quarterly Newsletter, and the results are tabulated from those completed questionnaires that have been returned.	**Sampling of Volunteers**
One Catholic Elementary, Junior High and High School are picked randomly from the list of all Catholic Schools. The surveyor then randomly selects 10% of the students, parents and teachers from those schools to ask the question to.	**Clustered**
The surveyor went to the neighbourhood Catholic School closest to her house and selected every 10^{th} person who walks in or out of the school on any school day.	**Convenient**
The list of all catholic students, parents, and school teachers and their phone number are complied separately; the surveyor then asks every 10^{th} person on each list.	**Stratified**
A list of all Catholic students, parents, and school teachers and their phone number are complied; the surveyor will ask every 10^{th} person on the list by phone.	**Systematic**
A group of Catholic parents, students and teachers are randomly chosen to represent 10% of the population.	**Simple Random**

Example 3: The student council has decided to put in a CD jukebox in the cafeteria. A decision needs to be made about the amount and the type of music to be stock into the jukebox. The council decided to conduct a survey by having students select their three most popular type of music from a list below. Identify the population and suggest a strategy for each probability sampling technique.

Cafeteria CD Juke Box Survey:

Please pick up to 3 of your favourite type of music below: **Current Grade Level: _________**

_____	Pop (Top 40)	_____	Dance	_____	Hard Rock	_____	Heavy Metal	_____	Rap
_____	Hip Hop	_____	R & B	_____	House	_____	Trance	_____	Country
_____	Independent	_____	World	_____	Sound Tracks	_____	Classical	_____	70's & 80's

Others (please specify): ______________

Population: All students who use the cafeteria frequently.

Sampling Technique	Strategy
Simple Random	100 randomly selected students in the cafeteria are taken during the lunch hour and before school starts to complete the survey.
Systematic	Every 5^{th} student that walks into the cafeteria during the lunch hour and before school starts are to complete the survey. (Unit of Arrangement - every 5^{th} student.)
Stratified	Find out from the office the percentage of Grade 10, 11 and 12 in the school. Students are randomly chosen in the cafeteria before school starts and during the lunch hour to complete the survey. Only 10 % of each group will be used for the final tabulation. (Stratifications: Grade 10, 11 and 12.)
Clustered	10 classes are selected randomly from the school during the 3^{rd} period. All students who frequent the cafeteria before school starts and during the lunch hour are to complete the survey.

8-1 Homework Assignments

Regular: pg. 368 to 369 #1 to 26, 28

AP: pg. 368 to 369 #1 to 26, 28, 29a

8-2: Inferences and Bias

Inference: - a conclusion made based on the result of a survey.

Bias: - when the survey will result in invalid (data that does not represent the population) or unreliable (same survey could not be repeated with the same result)

a. **Selection Bias**: - when the population is not represented properly in the sample.

b. **Response Bias**: - when the phrasing of the question will lead to a one-sided response.

c. **Non-Response Bias**: - when a large part of the sample did not respond to a survey.

Destructive Sampling: - when the sample will be destroyed during the process of the survey.

Example: Vehicle Crash Tests, Average Life-Time of a Manufactured Product.

Example 1: Determine the type of bias and the inference that it will likely occur with the statement or statement below. Suggest a correction for each case.

a. 9 out of 10 Canadians agree that Tim Horton donuts are way better than the American Krispy Kreme Donuts. Which donuts do you prefer?

(Response Bias) The question leads people to believe that most of the population has already preferred Tim Horton donuts. Therefore, the likely inference is that more people will answer Tim Horton rather than Christy Cream.

Correction: Which donuts do you prefer, Tim Horton or Christy Cream?

b. 100 people were randomly selected at the hockey arena after a hockey game to determine the percentage of Canadians that watches hockey on a regular basis.

(Selection Bias) Most people coming out of a hockey game tend to like hockey. Therefore, the likely inference is that a large population will respond that they watch hockey on a regular basis.

Correction: Move the survey location to a place where the population is better represented (movie theatre or systematic telephone survey).

c. To survey satisfaction levels of customers, a restaurant owner leave comment cards on the tables of his establishment for customers to voice their opinions.

(Non-Response Bias) The survey method will likely generate small number of response. And those who response most likely has negative comments about the service.

Correction: Have every 5th customers complete a small survey as they were about to pay for their bill.

8-2 Homework Assignments

Regular: pg. 372 to 373 #1 to 23 AP: pg. 372 to 373 #1 to 23

8-3: Expected Values

Payoff: - the amount of won or lost for each possible outcome of an event.

Expected Value (Mathematical Expectation) *E*: - the average winning for each time an event occurs.

When probability of each outcome is different:

$$E = [P(A) \times (\text{Payoff for } A)] + [P(B) \times (\text{Payoff for } B)] + [P(C) \times (\text{Payoff for } C)] + \ldots$$

When the probability of each event is the same:

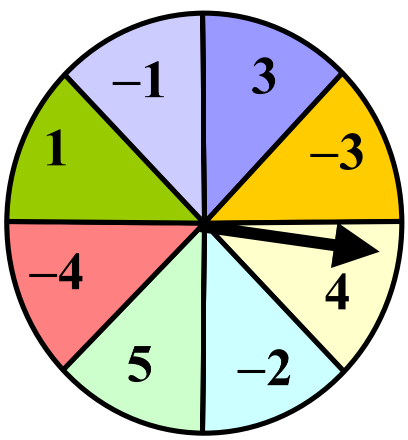
$$E = \frac{\text{Sum of Payoffs of ALL Outcomes}}{\text{Total Number of Outcomes}}$$

In a gaming scenario, a **FAIR GAME** is defined when **All Players have EQUAL Expected Value** **or** in a game **with a Single Player, the Expected Value is 0**.

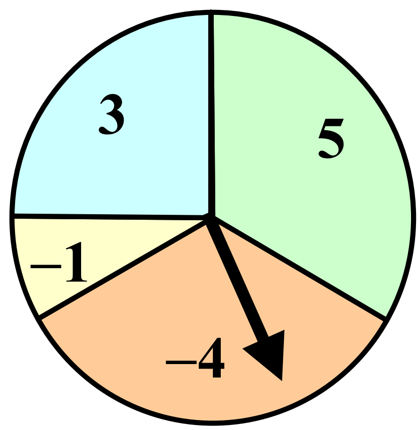
All Games involving MONEY are considered Multi-Player Games!

Example 1: Find the expected value of the following spinners. The payoff of event is indicated on the sector of the circle.

a.



b.



Since the probability of each sector is the same,

$$E = \frac{\text{Sum of Payoffs of ALL Outcomes}}{\text{Total Number of Outcomes}}$$

$$E = \frac{3+(-3)+4+(-2)+5+(-4)+1+(-1)}{8}$$

$$E = \frac{3}{8}$$

We can expect that the average payoff is 0.375 point per spin.

Since the probability of each sector is different,

$$E = P(A)\times(\text{Payoff } A)+P(B)\times(\text{Payoff } B)+P(C)\times(\text{Payoff } C) + P(D)\times(\text{Payoff } D)$$

The sectors have the fractions $\frac{1}{4}, \frac{1}{3}, \frac{1}{3}$ and $\frac{1}{12}$ that add up to 1.

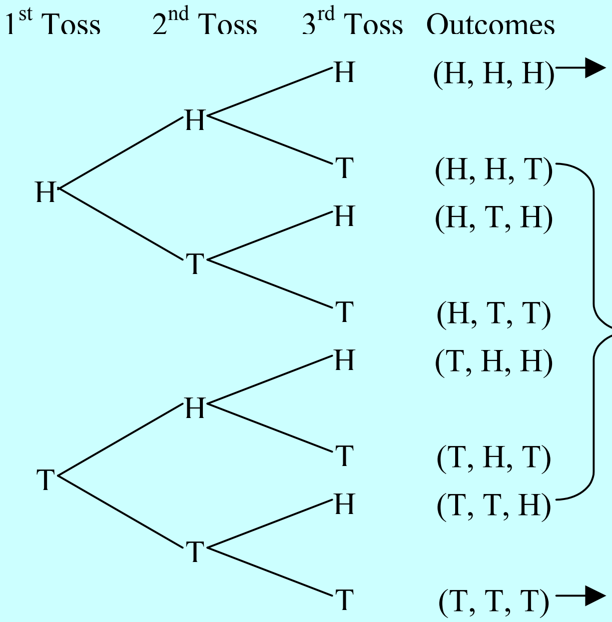
$$E = \left(\frac{1}{4}\right)(3)+\left(\frac{1}{3}\right)(5)+\left(\frac{1}{3}\right)(-4)+\left(\frac{1}{12}\right)(-1)$$

$$E = \frac{3}{4}+\frac{5}{3}+\left(\frac{-4}{3}\right)+\left(\frac{-1}{12}\right)$$

$$E = 1$$

We can expect that the average payoff is 1 point per spin.

Example 2: Three coins are tossed. He or she wins 5 points if all of the same kind appears; otherwise, the player losses 3 points. Using a tree diagram and expected value, determine whether the game is fair.



$P(\text{all H}) = \frac{1}{8}$

Payoff of all H = 5

$P(\text{2H or 2T}) = \frac{6}{8} = \frac{3}{4}$

Payoff of 2H or 2T = −3

$P(\text{all T}) = \frac{1}{8}$

Payoff of all T = 5

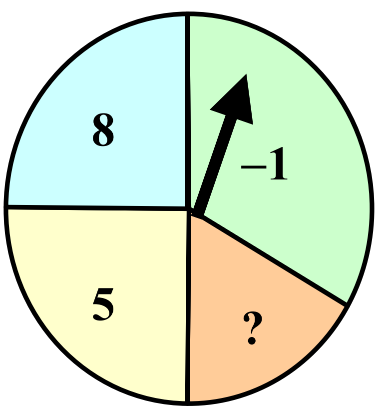
$$E = \left(\frac{1}{8}\right)(5) + \left(\frac{3}{4}\right)(-3) + \left(\frac{1}{8}\right)(5)$$

$$E = \frac{5}{8} + \left(\frac{-9}{4}\right) + \frac{5}{8}$$

$$E = -1$$

Because this single-player game has an expected number of −1, the player will lose an average of 1 point every time the three coins are tossed. **Therefore, it is NOT a fair game.**

Example 3: If the expected value for a spin on the spinner below is 2, how many points should be awarded in the last sector?



Let x = payoff of the unknown sector

$$P(\text{unknown sector}) = 1 - \frac{1}{3} - \frac{1}{4} - \frac{1}{4} = \frac{1}{6} \qquad E = 2$$

$$\left(\frac{1}{3}\right)(-1) + \left(\frac{1}{4}\right)(5) + \left(\frac{1}{4}\right)(8) + \left(\frac{1}{6}\right)(x) = 2$$

$$\left(\frac{-1}{3}\right) + \left(\frac{5}{4}\right) + 2 + \left(\frac{x}{6}\right) = 2$$

$$\frac{35}{12} + \frac{x}{6} = 2$$

$$\frac{x}{6} = 2 - \frac{35}{12}$$

$$\frac{x}{6} = \frac{-11}{12}$$

$$12x = -66$$

$$x = \frac{-66}{12}$$

$$x = -5.5$$

Example 4: In a game involving rolling two dice, a player pays \$1 per game. If the sum of a roll is 7 or if the two dice turn up with the same number, the player loses. If the sum is 3 or less, or 10 or more, the player doubles the money (\$2 back, which means \$1 gain). Any other sums mean a "push". The player gets the \$1 back, which means \$0 gain.

a. Using a table, find the probability of the win, push and lose of this game.
b. Evaluate whether this game is fair.

First Dice	Second Dice					
Sum	**1**	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$$P(\text{sum of 7 or same number dice}) = \frac{6}{36} + \frac{6}{36}$$

$$P(\text{sum} = 7 \text{ or same } \#) = \frac{1}{3}$$

Payoff of Sum of 7 and two dice with same number = –\$1

$$P(\text{sum} \le 3 \text{ or sum} \ge 10) = \frac{2}{36} + \frac{4}{36} \qquad P(\text{sum} \le 3 \text{ or sum} \ge 10) = \frac{1}{6}$$

(NOT including 1&1, 5&5, and 6&6)

Payoff for sum ≤ 3 or sum ≥ 10 = \$1

$$P(\text{other sums}) = 1 - \frac{1}{3} - \frac{1}{6} \qquad P(\text{other sums}) = \frac{1}{2}$$

Payoff for other sums = \$0

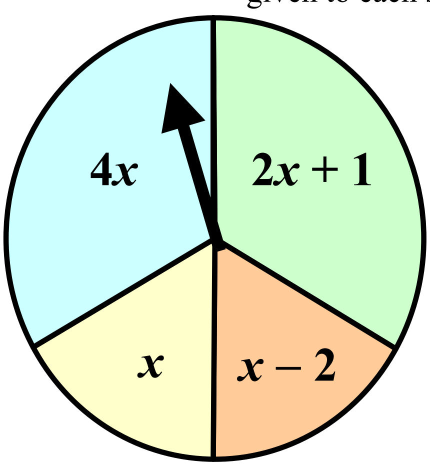
$$E = \left(\frac{1}{3}\right)(-1) + \left(\frac{1}{6}\right)(1) + \left(\frac{1}{2}\right)(0)$$

$$E = \left(\frac{-1}{3}\right) + \left(\frac{1}{6}\right)$$

$$E = \frac{-1}{6} = -0.1\overline{6}$$

Since it costs \$1 to play and the expected value is about –0.17, the player will expect to lose on average \$1 – (–\$0.16) = \$1.16 to the dealer. **Therefore, it is NOT a fair game.**

(AP) Example 5: The expected value of the spinner below is 7. Determine how many points should be given to each sector.



$$\left(\frac{x}{6}\right) + \left(\frac{x-2}{6}\right) + \left(\frac{4x}{3}\right) + \left(\frac{2x+1}{3}\right) = E$$

$$\frac{x + (x-2) + 2(4x) + 2(2x+1)}{6} = 7$$

$$\frac{x + x - 2 + 8x + 4x + 2}{6} = 7$$

$$\frac{14x}{6} = 7$$

$$14x = 42$$

$$x = \frac{42}{14}$$

$$x = 3$$

x sector:	$4x$ sector:	$2x + 1$ sector:	$x - 2$ sector:
	= 4(3)	= 2(3) + 1	= (3) – 2
3	**12**	**7**	**1**

8-3 Homework Assignments

Regular: pg. 382 to 383 #1 to 13, 14b, 15, 16

AP: pg. 382 to 383 #1 to 13, 14b, 15 to 19, 21

8-4: Making Decisions Using Probability

Example 1: In Lotto 6/49, the player picks 6 numbers out of the numbers 1 to 49. Each play costs \$1. In a typical draw, the probabilities and the payoffs for each outcome are listed below. Determine the expected value of Lotto 6/49 and how much money a player can expect to win or lose by buying one ticket.

Type of Winning	Jackpot (all 6 numbers)	5 Numbers and Bonus Number	5 Numbers	4 Numbers
Probability	0. 000 000 071 511	0. 000 000 429 067	0. 000 018 020 832	0. 000 968 619 724
Payoff	\$2,200,000	\$100,000	\$2,000	\$100

$$E = [P(\text{Jackpot}) \times \text{Payoff}(\text{Jackpot})] + [P(5\# \,\&\, \text{Bonus}\,\#) \times \text{Payoff}(5\# \,\&\, \text{Bonus}\,\#)] + [P(5\#) \times \text{Payoff}(5\#)] + [P(4\#) \times \text{Payoff}(4\#)]$$

$$E = (0.000\,000\,071\,511)(\$2{,}200{,}000) + (0.000\,000\,429\,067)(\$100{,}000) + (0.000\,018\,020\,832)(\$2{,}000) + (0.000\,968\,619\,724)(\$100)$$

$$E = 0.1573242 + 0.0429067 + 0.036041664 + 0.0968619724$$

$$E = 0.3331345364$$

Since it costs \$1 to purchase a Lotto 6/49 ticket, the player can expect to **LOSE \$1 – \$0.33 = \$0.67** per ticket bought.

Example 2: An auto insurance company would like to determine the premium for young drivers. They found that young drivers between the ages of 16 to 25 have the following claim probabilities and amounts.

Annual Claim Amounts (\$)	0 (no claim)	1000	5000	10000	15000	20000	25000
Probability	0.57	0.16	0.12	0.08	0.04	0.02	0.01

a. Find the expected amount of annual premium the insurance company should charge young drivers.
b. If the insurance company would increase its rate with a mark-up of 30%, what will be the annual premium for young drivers?
c. Most insurance companies have deductibles (money that claimant has to pay initially to cover the damage) on their policies. What should the new premium be with a \$500 deductible and a mark up of 30%?
d. Statistically, if a new driver does not get into an accident in the first year, his or her chances to remain accident-free increases 15% each year. What is the probability of a young driver having to make an insurance claim if she has 5 straight years of accident-free driving?

a. Annual Premium = Expected Number of Annual Claim Amount per Policy

$$E = (0.57)(\$0) + (0.16)(\$1000) + (0.12)(\$5000) + (0.08)(\$10000) + (0.04)(\$15000) + (0.02)(\$20000) + (0.01)(\$25000)$$

$$E = \$0 + \$160 + \$600 + \$800 + \$600 + \$400 + \$250$$

E = $2810 (Annual Premium)

b. Mark-Up 30% = $2810 × 0.30 = $843

New Premium with mark-up = $2810 + $843

New Premium with mark-up = $3653

c.

Annual Claim Amounts with $500 deductible ($)	0 (no claim)	500	4500	9500	14500	19500	24500
Probability	0.57	0.16	0.12	0.08	0.04	0.02	0.01

Annual Premium = Expected Number of Annual Claim Amount per Policy

$$E = (0.57)(\$0) + (0.16)(\$500) + (0.12)(\$4500) + (0.08)(\$9500) + (0.04)(\$14500) + (0.02)(\$19500) + (0.01)(\$24500)$$

$$E = \$0 + \$80 + \$540 + \$760 + \$580 + \$390 + \$245$$

E = $2595 (Annual Premium with $500 deductible)

Mark-Up 30% = $2595 × 0.30 = $778.50
New Premium with mark-up = $2595 + $778.50

Premium with $500 deductible and 30% mark-up = $3373.50

d. P (no accident during first year) = 0.57
P (no accident during two years) = 0.57(1 + 0.15) = 0.6555 (× 1.15 due to 15% increase)
P (no accident during three years) = 0.57(1.15)(1.15) = 0.753825
P (no accident during four years) = 0.57(1.15)(1.15)(1.15) = 0.86689875
P (no accident during five years) = 0.57(1.15)(1.15)(1.15)(1.15) = 0.9969335625

P (making a claim with 5 years of no-accidents) = 1 − 0.9969335625

***P* (making a claim with 5 years of no-accidents) ≈ 0.0030664375 ≈ 0.31%**

8-4 Homework Assignments

Regular: pg. 387 to 389 #1 to 14 AP: pg. 387 to 389 #1 to 16