Means of Random Variables

We have seen how to picture the probability distribution of random variables through the use of histograms (discrete) and density curves (continuous). However, it is also helpful to have numerical descriptions of these variables. The mean of a random variable describes where it is centered, while the variance and standard deviation describe the extent to which it spreads out about the center

The mean of a discrete random variable is also called the expected value. Why?

To calculate the mean, we must consider the probability that each outcome can occur. Since outcomes are not always equally likely, their probabilities need to be factored in when calculating the mean.

Mean of A random Variable, X

Let X be a discrete random variable with prob distribution

X	x_1	x_2	x_3	...	x_n
$P(X)$	p_1	p_2	p_3		p_n

Then $\mu_x = x_1p_1 + x_2p_2 + x_3p_3 + \dots x_np_n$

Example 20: At 1 min after birth and again at 5 min, each newborn child is given a numerical rating called an Apgar score. Possible values of this score are 0, 1, 2, ... , 9, 10. The score is determined by five factors: muscle tone, skin color, respiratory effort, strength of heartbeat, and reflex, with a high score indicating a healthy infant. Let the random variable X denote the Apgar score (at 1 min) of a randomly selected newborn infant, and suppose X has the following probability distribution:

X	0	1	2	3	4	5	6	7	8	9	10
P(X)	.002	.001	.002	.005	.02	.04	.17	.38	.25	.12	.01

Use the formula for the mean value of X. That is, find the average Apgar score that is approached as child after child is rated. Show all work:

$0(.002) + 1(.001) + 2(.002) + 3(.005) + 4(.02) + 5(.04) + 6(.17) + 7(.38) + 8(.25) + 9(.12) + 10(.01) =$

7.16

Example 21: A fair coin is flipped 3 times. Find the mean of the discrete random variable X that counts the number of heads.

$0(\frac{1}{8})+1(\frac{3}{8})+2(\frac{3}{8})+3(\frac{1}{8})$

X = # of heads	0	1	2	3
P(x)	1/8	3/8	3/8	1/8

$\frac{12}{8} = 1.5 = \mu_x$

Example 22: The daily lottery costs \$1 to play. You pick a 3 digit number. If you win, you win \$500. Find the expected value of the lottery.

X = winnings	-1	500
P(x)	999/1000	1/1000

$E(x) = -.499$

Variance and Standard Deviation of a Discrete Random Variable X:

Recall that the variance and standard deviation involve calculations based on squared deviations from the mean. That is, each observation is a certain distance from the mean. That distance is called a deviation. To calculate the variance, the deviations are squared and averaged. However, in a random variable, we must take into account the likelihood of each outcome occurring when calculating this measure. Therefore, we multiply each squared deviation by its probability to factor in its weight.

Variance of a Discrete Random Variable:

$$\sigma^2 = \sum (x_i-\mu_x)^2 p = (x_1-\mu_x)^2 p_1 + (x_2-\mu_x)^2 p_2 + \dots + (x_n-\mu_x)^2 p_n$$

$$\sigma = \sqrt{\sigma^2}$$

Example 22: Consider the distribution of Apgar scores. Calculate the standard deviation for this distribution.

X	P(X)	X · P(X)	(X − μ)	$(X-\mu)^2p(X)$
0	.002		-7.16	.1025
1	.001		-6.16	.0379
2	.002		-5.16	.0533
3	.005		-4.16	.0865
4	.02		-3.16	.1997
5	.04		-2.16	.1866
6	.17		-1.16	.2289
7	.38		-.16	.0097
8	.25		.84	.1764
9	.12		1.84	.4063
10	.01		2.84	.0807
		7.16		1.57

Interpret the mean and standard deviation of X in the context of the problem.

The average apgar of an infant is 7.16 with a standard deviation of 1.25

Example 23: A fair coin is flipped 3 times. Find the standard deviation of the discrete random variable X that counts the number of heads.

X	0	1	2	3
P(X)	1/8	3/8	3/8	1/8

.28125 .09375 .09375 .281825

$\mu_x = 1.5$ heads

$\sigma_x^2 = .75$

$\sigma = .866$ heads

Example 24: The daily lottery costs $1 to play. You pick a 3 digit number. If you win, you win $500. Find the standard deviation of the lottery.

X	-1	500
P(x)	.999	.001

.2507 250.49

$\mu_x = -.499$

$\sigma_x^2 = 250.75$

$\sigma_x = 15.84$

Example 25: Choose an American household at random and let the random variable X be the number of persons living in the household. Find the standard deviation of the average American household.

(	1	2	3	4	5	6	7
P(x)	.25	.32	.17	.15	.07	.03	.01

$\mu_x = 1(.25) + 2(.32) + 3(.17) + 4(.15) + 5(.07) + 6(.03) + 7(.01) = 2.6$

$\sigma_x^2 = 2.02$

$\sigma_x = 1.42$

When a problem has many values of X like this one, this can be done using your list feature of your calculator:

1. Enter the data into L1 and L2.
2. Find μ_x: Multiply L1 and L2 in L3. Then find the sum of L3.
3. We no longer need L3. In L3, we will put our formula for variance: L3 = $(L1 - \mu_x)^2 \cdot L2$.
4. To find the variance, Sum L3.
5. For the standard deviation, square root the variance.

Example 26: Choose an American suburban household at random and let the random variable X be the number of cars that people in the house own. Find the number of cars owned by the average suburban American household.

X	0	1	2	3	4	5	6
P(X)	0.08	0.29	0.32	0.17	0.08	0.04	0.02

$\mu_x = 2.09$ $\qquad \sigma = 1.33$

$\sigma_x^2 = 1.77$

Example 27: A huge cookie jar has 60% chocolate cookies and 40% vanilla cookies. Sam chooses 3 cookies blindfolded. Let X be the number of chocolate cookies he chooses. Construct the probability distribution of X and the average number of chocolate cookies he chooses.

X = # of choc	0	1	2	3
P(x)	.064	.288	.432	.216

$\mu_x = 1.8$

Example 28: A contractor is bidding on a road construction job that promises a profit of $20,000 with a probability of 0.7 and a loss, due to strikes, weather conditions, late arrival of building materials, and so on, of $40,000 with a probability of 0.3. What is the contractor's mean expectation?

X= income	20 000	-40000
P(x)	.7	.3

Example 29: An oil company will only invest in an oil well if they can expect to make at least $1 million in profit. They find a possible area in which to drill in Canada. It will cost the company $3 million to make the attempt. If they find oil, they can expect to clear $7 million in profits. Geologists have estimated that the probability of striking oil is 0.35. Should the company make the attempt? Why or why not?

X= outcome	-3	7
P(x)	.65	.35

$\mu_x = .5$

Law of Large Numbers

Draw independent observations from any population with finite mean. As the number of observations increases, the mean x of the observed values eventually approaches the mean of that population.

Law of Large Numbers:

Draw Independent observations at random from a population with mean, μ. Decide how accurate you want to estimate μ. As the number of observations increases, the sample mean increasingly approaches the population mean.

i.e. the more times you flip a coin the closer the proportion of heads is to 50%

Law of Small Numbers:

We expect short sequences of random events to show the same average behavior that really only appears in the long run.

Example 30: If I flip a fair coin, over a long period of time, the percentage of heads will approach 50%. That does not mean that the number of heads will approach the number of tails. Here is a typical experiment in tossing coins.

Heads	7	42	463	4875	49660
Tails	3	58	537	5125	50340
Percentage of heads	70%	42%	46.3%	48.75%	49.66%

Example 31: If you play blackjack perfectly at a casino, the expected value for every dollar played is 98.5 cents. That means on the average of hand you play, you lose 1.5 cents. Can an individual player win playing blackjack at a casino? The "law of small numbers" says that over a short run, a player can certainly win. But over the long run, the law of large numbers says that a player will eventually lose. And given the huge number of players who play daily, the law of large numbers states that the casino has to win. There is a reason that casinos look the way they do - it is guaranteed money. Still, players can win over the short haul. The question that crops up is what is considered "large?" There is no answer to that. What you have to know is that the larger n (the number of trials) gets, the closer the average win (or loss) per play approaches the true average of losing 1.5 cents a play.

Rules for Means

Often in statistics, we need to consider the sum, difference, or linear combination of multiple random variables. We can use several formulas to determine the mean and variance of these situations:

Rules for Means:

1. If X is a random variable and a and b are fixed numbers, then

$\mu_{a+bx} = a + b\mu_x$

2. If X and Y are random variables, then

$\mu_{x+y} = \mu_x + \mu_y$

also true for subtraction

Rules for Variances:

1. If X is a random variable and a and b are fixed numbers, then

$\sigma^2_{a+bx} = b^2\sigma^2_x$

2. If X and Y are INDEPENDENT random variables, then

$\sigma^2_{x+y} = \sigma^2_x + \sigma^2_y$, $\sigma^2_{x-y} = \sigma^2_x + \sigma^2_y$

Combine 2 rules

$\sigma^2_{ax+by} = a^2\sigma^2_x + b^2\sigma^2_y$ $\mu_{ax+By} = a\mu_x + b\mu_y$

Example 32: Suppose the average SAT-Math score is 625 and the standard deviation of SAT-Math is 90.The average SAT-Verbal score is 590 and the standard deviation of SAT-Verbal is 100. The composite SAT score is determined by adding the Math and Verbal portions. What is the mean and standard deviation of the composite SAT score?

$$\mu_{M+V} = 625 + 590 = 1215$$

$$\sigma^2_{M+V} = 90^2 + 100^2 = 18100$$

$$\sigma_{M+V} = 134.54$$

Example 32: Suppose two pro bowlers Adam (A) and Bart (B) have the following distribution of scores.

$\mu_A = 209$ $\sigma_A = 14$ $\mu_B = 221$ $\sigma_B = 22$

Find the following:

a. μ_{B+A} b. μ_{B-A} c. $3\,\mu_A$ d. σ_{B+A} e. σ_{B-A}

$\mu = 221 + 209 = 430$

$\mu = 221 - 209 = 12$

$3(209) = 627$

$\sqrt{\sigma_B^2 + \sigma_A^2}$ ← Same

$\sqrt{22^2 + 14^2} = \sqrt{680} = 26.08$

Example 33: Depending on the attendance of a minor league baseball team, the number of hot dogs and sodas sold at a game is given by the following table.

Hot Dogs Sold	100	200	500	1000	2500
Probability	0.05	0.1	0.3	0.35	0.2

$\mu_H = 1025$ $\sigma^2 = 628875$ $\sigma = 793$

Sodas Sold	100	200	400	800	1500
Probability	0.1	0.15	0.25	0.30	0.20

Find the standard deviation of the number of hot dogs plus the number of sodas sold.

Mean $\mu_S = 680$ $\sigma_S^2 = 226600$ $\sigma_S = 476$

$$\mu_{H+S} = 1025 + 680 = 1705$$

$$\sigma^2_{H+S} = 793^2 + 476^2 = 855475$$

$$\sigma_{H+S} = 924.89$$

Profit $= .75S + 1.50H$

$\mu_P = .75(680) + 1.50(1025)$

$= \$2047.50$

$\sigma_P^2 = .75^2(476)^2 + 1.50^2(793)^2$

$= 1542359.25$

$\sigma_P = \$1241.95$

Combining Normal Random Variables

Example 34: Mr. Molesky and Mr. Liberty are avid golfers. Suppose Mr. Molesky's scores average 110 with a standard deviation of 10. Mr. Liberty's scores average 100 with a standard deviation of 8. Find the mean and standard deviation of the difference of their scores (Molesky - Liberty). Assuming their scores are normally distributed, find the probability that Mr. M will win on any given day.

$\mu_{M-L} = 110 - 100 = 10$

$\sigma_{M-L} = \sqrt{10^2 + 8^2} = \sqrt{164} = 12.81$

Scores are independent

$P(M < L)$

$P(M - L < 0)$

Norm cdf (-100, 0, 10, 12.81) = .2173

Example 35: Bottle caps are manufactured so that their inside diameters have a distribution that is approximately N(36mm, 1mm). The distribution of the outside diameters of bottles is approximately N(35mm, 1.2mm). If a bottle cap and a bottle are selected at random, what is the probability the cap will fit on the bottle?

A = dia of inside   B = diameter of outside bottle

$P(A > B)$

$P(A - B > 0) = .7392$

cdf(0, 100, 1, 1.56)

Example 36: W.J. Youden (Australian, 1900-1971) weighed many new pennies and found the distribution of weights to be approximately N(3.11g, 0.43g). What are the reasonably likely mean and standard deviation of weights of rolls of 50 pennies?

a roll weight is 50 W

$\mu_{50W} = 50(3.11) = 155.5g$

$\sigma^2_{50W} = 50^2(.43)^2 = 462.25$

$\sigma_{50W} = 21.5g$

Example 37: Suppose the amount of propane needed to fill a customer's tank is a random variable with a mean of 318 gallons and a standard deviation of 42 gallons. Hank Hill is considering two pricing plans for propane. Plan A would charge \$2 per gallon. Plan B would charge a flat rate of \$50 plus \$1.80 per gallon. Calculate the mean and standard deviation of the distributions of money earned under each plan. Assuming the distributions are normal, calculate the probability that Plan B would charge more than Plan A.

$A = 2G$

$B = 50 + 1.80G$

$\mu_A = 2(318) = 636$

$\sigma_A = \sqrt{2^2 42^2} = 84$

$\mu_B = 50 + 1.8(318)$

$= 622.40$

$\sigma_A = \sqrt{1.8^2(42^2)} = 75.6$

$P(B > A) = P(B - A > 0) = .4521$

cdf(0, 10000, -13.6, 113.01)