Random Variables

Example: For insurance purposes, Professor Kupe's house has a replacement value of \$275,000, even though the house is not worth nearly that much. Each year, he must pay his home insurance premium, which currently runs about \$800.

Pretend we live in a simplified world where there are only three outcomes each year for Kupe's home – ***Total Destruction*** (burn to the ground, tornado sweeps it away), ***Medium Damages*** (tree falls through roof, hurricane floods basement), or ***Neither***. A table with the outcomes, payouts, and probabilities are listed next. The random variable X = the payment.

Kupe's Outcome	**Payment x**	**Probability $P(X = x)$**
Total Destruction	\$275,000	1 / 1000
Medium Damages	\$10,000	40 / 1000
Neither	\$0	959 / 1000

- This probability model is discrete, because there are a fixed number of outcomes that we can list.
- Capital letters denote the random variable. Lowercase letters denote the values the random variable can take on.
- The expected value of this home owner's insurance policy is just a fancy name for the mean. What is the **expected value** of this policy from Kupe's perspective?

$$E(X) = \mu = \sum x P(x)$$

$$= 275000\left(\frac{1}{1000}\right) + 10000\left(\frac{40}{1000}\right) + 0\left(\frac{959}{1000}\right)$$

$$= \$675 \text{ expected payment.}$$

- Since the insurance company collects \$800 up front, what is the expected value of the policy from their perspective?

$$\$800 - \$675 = \$125 \text{ expected profit}$$

- Say the insurance company sells 1000 of these policies each year. They would expect:

  1 home to reach ***Total Destruction***, for a total payout of $ 275,000

  40 homes to have ***Medium Damages***, for a total payout of $ 400,000

  959 homes to have ***Neither***, for a total payout of $ 0

- The insurance company spreads the risk. They collect $ 800,000 each year, expecting to pay out $ 675,000 in total, for a grand profit of $ 125,000. Some years will be better and some years will be worse, depending on how many random fires, random hurricanes, and random tornados hit Cecil County any given year.

- How much might the insurance company expect their payouts to vary year-to-year? They need to know the random variable's standard deviation: $\mu = 675$

X	P(x)	$(x-\mu)$ Deviation	Deviation Squared
275,000	1/1000	274,325	$(274,325)^2$
10,000	40/1000	9,325	86,955,625
0	959/1000	-675	455,625

$$VAR(x) = \sigma^2 = \sum (x-\mu)^2 P(x)$$

$$= (274325)^2 \cdot \frac{1}{1000} + (86,955,625) \cdot \frac{40}{1000} + (455,625) \cdot \frac{959}{1000} = 79,169,375$$

$$SD(x) = \sqrt{\sigma^2} = \sigma = \sqrt{79,169,375}$$

$$\approx \$8,897.72$$

Per Policy.

Working With Means and Variances for Random Variables

- The insurance company charges \$800 per policy, so the expected profit was \$675.

- As they seem to do every year, what will happen to the company's profit if next year they decide to raise the premium by \$100 but keep the coverage the same? Give formulas and values for the shifted means and variances:

$E(X \pm c) = E(X) \pm c$

$VAR(X \pm c) = VAR(X)$

Customer

$E(X+c)$ is still \$675, but we pay \$900 now.

Company

$E(X+c)$ is now \$900 − \$675 = \$225 expected profit.

Same $VAR(X) = \sigma^2$ and $SD(X) = \sigma$ as before.

Variation in payouts doesn't change.

- What if all payouts increased by 10% due to the complaining by customers? What would happen to the expected value of the policy from the customer's point of view? What about the expected profit from the company's point of view?

$E(aX) = aE(X)$

So if all payouts get multipliedy by 1.1,

$E(1.1X) = 1.1\,E(X) = 1.1(675) = \$742.5$

$VAR(aX) = a^2 VAR(X)$

So new variation is $(1.1)^2(79{,}169{,}375) =$ $95{,}794{,}943.75$

new SD is $\sigma = \$9{,}787.49$

10% increase in payouts increased SD by 10%

- Since the insurance company sells policies to many customers, we'd need to know the expected value of a sum or difference of many random variables:

X = first policy Y = second policy

$$E(X+Y) = E(X) + E(Y)$$

original values: $E(X+Y) = \$675 + \$675$
$= \$1350$

And this idea would extend to 1000 customers or even more.

- Variability is another matter though. The risk of insuring two houses with the same risk is different than insuring one house for twice the risk. Insurance companies spread the risk. As long as House A is independent from House B, their payouts or lack of payout will not affect each other. The rule goes like this:

"The variance of the sum of two indep. random variables is the sum of their variances"

$$VAR(X \pm Y) = VAR(X) + VAR(Y)$$

$$SD(X \pm Y) = \sqrt{VAR(X) + VAR(Y)}$$

VARIANCES ALWAYS ADD!

- The variance for one house for twice as much would be twice as big as the variance for two independent policies:

$$VAR(2X) = 2^2 VAR(X) = 4(79{,}169{,}375) = 316{,}677{,}500$$

$$SD(2X) = \sqrt{316{,}677{,}500} \approx \$17{,}795.43$$

one house twice as much

$$VAR(X+Y) = VAR(X) + VAR(Y) = 79{,}169{,}375 + \ldots$$
$$= 158{,}338{,}750$$

$$SD(X+Y) = \sqrt{158{,}338{,}750} \approx \$12{,}583.27$$

Example: Draw a card from a standard deck. If you get a red card, you win nothing. If you get a club, you win \$100. If you get a spade, you get \$100 plus a bonus of \$400 for the ace of spades.

a. Set up a probability model for the amount you win.
b. Find the expected amount won.
c. How much should you be willing to pay to play this game?
d. Find the standard deviation of the random variable.

a.)

X	$P(X=x)$
\$0	26/52
\$100	25/52
\$500	1/52

b.) $E(X) = \mu = 0(26/52) + 100(25/52) + 500(1/52)$
$= \$57.69$

c.) Breakeven at \$57.69.
If you gamble anything less to play, you come out ahead in the long run.

d.) $VAR(X) = (0-57.69)^2 \cdot \frac{26}{52} + (100-57.69)^2 \cdot \frac{25}{52} + (500-57.69)^2 \cdot \frac{1}{52} = 6286.98$

$SD(X) = \sqrt{VAR(X)} \approx \$79.29$

e.) If you played 100 times, what would SD(X) be?

$VAR(X_1 + X_2 + \cdots + X_{100}) = 100(6286.98) = 628698$

$SD(X_1 + X_2 + \cdots + X_{100}) = \sqrt{628698} \approx \$792.90$

You'd expect to win $100(57.69) = \$5,769$.

Example: Given these independent random variables X and Y, along with their means and standard deviations, find the mean and standard deviation of:

$SD(aX) = a\,SD(X)$

a. $2Y - 20$
b. $4X$
c. $0.35X + Y$
d. $X - 4Y$
e. $Y_1 + Y_2 + Y_3$

	Mean	Standard Deviation
X	70	15
Y	10	10

a.) $E(2Y-20) = 2E(Y) - 20 = 2(10) - 20 = 0$

$SD(2Y-20) = 2\,SD(Y) = 2(10) = 20$

b.) $E(4X) = 4\,E(X) = 4(70) = 280$

$SD(4X) = 4\,SD(X) = 4(15) = 60$

c.) $E(0.35X + Y) = 0.35\,E(X) + E(Y)$

$= 0.35(70) + 10 = 34.50$

$SD(0.35X + Y) = \sqrt{VAR(0.35X + Y)}$

$= \sqrt{VAR(0.35X) + VAR(Y)}$

$= \sqrt{0.35^2\,VAR(X) + VAR(Y)}$

$= \sqrt{0.35^2(15)^2 + (10)^2}$

$= \sqrt{127.5625}$

≈ 11.29

d.) $E(X-4Y) = E(X) - 4E(Y) = 70 - 4(10) = 30$

$$SD(X-4Y) = \sqrt{VAR(X-4Y)}$$
$$= \sqrt{VAR(X) + 4^2 VAR(Y)}$$
$$= \sqrt{15^2 + 16(10)^2}$$
$$= \sqrt{1825} \approx 42.72$$

e.) $E(Y_1+Y_2+Y_3) = E(Y_1) + E(Y_2) + E(Y_3)$
$$= 3(10) = 30$$

$$VAR(Y_1+Y_2+Y_3) = VAR(Y_1) + VAR(Y_2) + VAR(Y_3)$$
$$= 100 + 100 + 100$$
$$= 300$$
$$SD(Y_1+Y_2+Y_3) = \sqrt{300} \approx 17.32$$

The Probability Model for Sums of Random Variables (Be Careful)

- If the insurance company would like to display a probability model for the payouts when insuring two homes, it is not as simple as just doubling the payouts and probabilities:

One Home	Payment x	Probability $P(X = x)$
Total Destruction	\$275,000	1 / 1000
Medium Damages	\$10,000	40 / 1000
Neither	\$0	959 / 1000

- Assuming independence, write out a probability model for insuring two homes. List the possible outcomes, payouts and probabilities.

Outcome	Payment	Probability
Both Destroyed	\$550,000	$\left(\frac{1}{1000}\right)^2 = \frac{1}{1,000,000}$
One Destroyed, One Medium	\$285,000	$\left(\frac{1}{1000}\right)\left(\frac{40}{1000}\right)(2) = \frac{80}{1,000,000}$
One Destroyed, One Fine	\$275,000	$\left(\frac{1}{1000}\right)\left(\frac{959}{1000}\right)(2) = \frac{1918}{1,000,000}$
Both Medium	\$20,000	$\left(\frac{40}{1000}\right)^2 = \frac{1600}{1,000,000}$
One Medium, One Fine	\$10,000	$\left(\frac{40}{1000}\right)\left(\frac{959}{1000}\right)(2) = \frac{76720}{1,000,000}$
Both Fine	\$0	$\left(\frac{959}{1000}\right)^2 = \frac{919681}{1,000,000}$

$$E(X_1 + X_2) = (\$550,000)\left(\frac{1}{1,000,000}\right) + \cdots + (0)\left(\frac{919681}{1000000}\right)$$

$$= \$1350 = 2 \times \$675 \text{ from before}$$

Combining Normal Random Variables

- The probability model for a Normal random variable was covered in Math 127 and used quite frequently in Math 128.
- If we take the sum of independent Normal random variables, we know the mean of the sum will be the sum of the means and that the variance of the sum will be the sum of the variances.
- **Bonus Fact:** Sums of Normal random variables follow a Normal Model!

Let $X_1 \sim \text{Normal}(\mu_1, \sigma_1)$ and

$X_2 \sim \text{Normal}(\mu_2, \sigma_2)$

Then $X_1 + X_2$ will follow a Normal Model too! (This idea does not generalize to all random variables!)

Mean of $X_1 + X_2$ is $\mu_1 + \mu_2$

VARIANCE OF $X_1 + X_2$ IS $\sigma_1^2 + \sigma_1^2$

SD OF $X_1 + X_2$ IS $\sqrt{\sigma_1^2 + \sigma_2^2}$

Difference of two Normals has mean $\mu_1 - \mu_2$ but SD $\sqrt{\sigma_1^2 + \sigma_2^2}$

Example: An exercise equipment company in Lancaster, PA sells sets of free-weight dumbbells. A commonly purchased set contains two 20-pounders, two 15-pounders, and two 10-pounders. The 20-pounders have a standard deviation of 0.2 pounds, the 15s have SD = 0.15, and the 10s have SD = 0.10. The weight of the weights follow a Normal model quite well.

a. Describe the probability model weight of the box to be shipped for the set of six weights. Assume the box's weight is negligible.

$X_1 \sim \text{Normal}(20, 0.2)$ $X_3 \sim N(15, 0.15)$ $X_5 \sim N(10, 0.1)$

$X_2 \sim \text{Normal}(20, 0.2)$ $X_4 \sim N(15, 0.15)$ $X_6 \sim N(10, 0.1)$

Let S = Sum of all six, assuming indep.

$S \sim \text{Normal}(90, 0.381)$

$\mu_S = 20+20+15+15+10+10 = 90$ lbs.

$$\sigma_S = \sqrt{(0.2)^2+(0.2)^2+(0.15)^2+(0.15)^2+(0.10)^2+(0.10)^2}$$

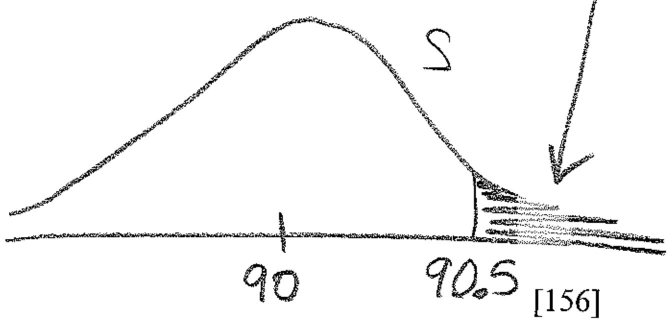
≈ 0.381

b. What is the probability that the box weight exceeds 90.5 pounds?

$P(S > 90.5) = \text{ncdf}(90.5, E99, 90, 0.381)$

$= 0.095$

$\sim 9.5\%$ chance



c. At home, you pick up the two 20s, one in each hand. What is the probability that your right hand holds more than 0.25 lb. more than your left hand?

Let $D = R - L$

$\sigma_D = \sqrt{0.2^2 + 0.2^2} = 0.283$

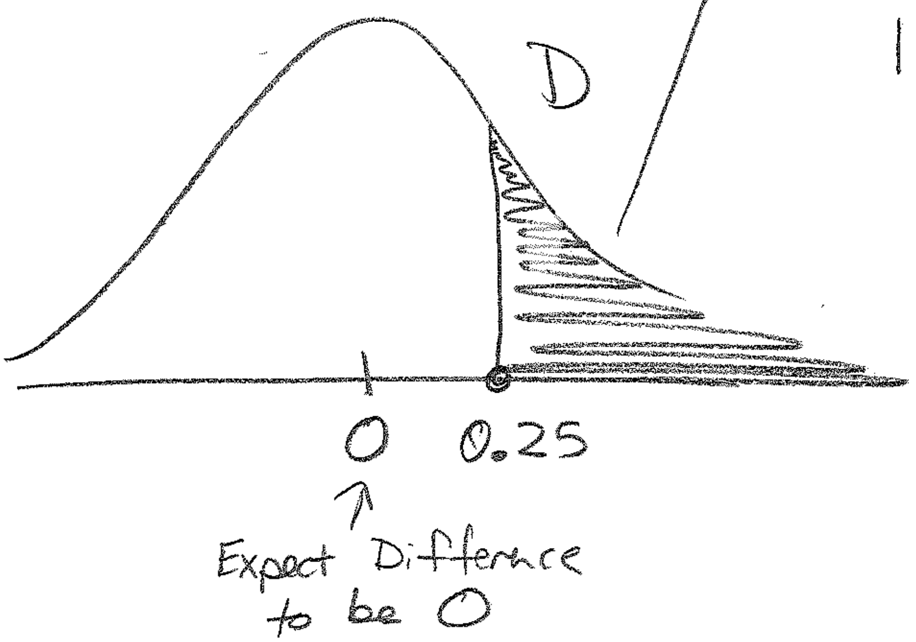
$R \sim \text{Normal}(20, 0.20)$

$L \sim \text{Normal}(20, 0.20)$

$D = R - L \sim \text{Normal}(0, 0.283)$

$P(D > 0.25 \text{ lbs}) = \text{ncdf}(0.25, E99, 0, 0.283)$

$= 0.189$



18.9% chance your right-hand holds at least 0.25 lbs more than left.

* Assuming randomly selected weights