Section 6.2 - Transforming and Combining Random Variables (pp. 358-382)

In Chapter 2, we studied the effects of transformations on the shape, center, and spread of a distribution of data.

  a. *Adding (or subtracting) a constant **a** to each observation:*

- ADDS a TO MEASURES OF CENTER + LOC ($\bar{x}$, MED, QUARTILES, %ILES)
- DOES NOT Δ SHAPE OR MEASURES OF SPREAD (RG, IQR, STD DEV.)

  b. *Multiplying (or dividing) each observation by a constant **b**:*

- MULTS (DIV'S) MEASURES OF CTR + LOCATION (MEAN, MED, Q's, %ILES) BY b
- MULTS (DIVS) MEASURES OF SPREAD (IQR, RG, STD DEV) BY |b|
- DOES NOT CHANGE SHAPE OF DIST.
-

---

1. Linear Transformations of Random Variables

Example. El Dorado Community College considers a student to be full-time if he or she is taking between 12 and 18 units. The number of units X that a randomly selected EDCC full-time student is taking in the fall semester has the following distribution:

Number of Units (X):	12	13	14	15	16	17	18
Probability:	0.25	0.10	0.05	0.30	0.10	0.05	0.15

Here is a histogram of the probability distribution:
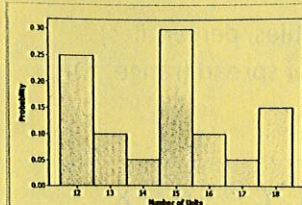
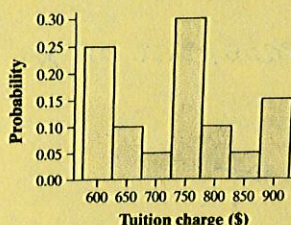μ_x = 14.65 UNITS

σ_x = 2.06 UNITS

At EDCC, the tuition for full-time students is $50 per unit. If T = tuition for a randomly selected full-time student then T = 50X . Here is the probability distribution for T and a histogram of the probability distribution:

Tuition Charge (T):	600	650	700	750	800	850	900
Probability:	0.25	0.10	0.05	0.30	0.10	0.05	0.15

μ_T = \$732.50

σ_T = \$103

What happened to the shape of the distribution?

SHAPES ARE THE SAME.

What happened to the mean and standard deviation?

$M_T = 50 M_x$, $\sigma_T = 50 \sigma_x$

Effect on a Random Variable of Multiplying (Dividing) by a Constant

Multiplying (or dividing) each value of a random variable by a number b:
- Multiplies (divides) measures of center and location (mean, median, quartiles, percentiles) by b;
- Multiplies (divides) measures of spread (range, IQR, standard deviation) by $|b|$.
- Does not change the shape of the distribution.

Example (cont). In addition to tuition charges, each full-time student at EDCC is assessed student fees of $100 per semester. If C = overall cost for a randomly selected full-time student, $C = 100 + T$

$T = 50X$

Here is the probability distribution for C and the histogram of the probability distribution:

Overall Cost (C):	700	750	800	850	900	950	1000
Probability	0.25	0.10	0.05	0.30	0.10	0.05	0.15
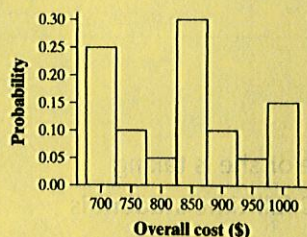
$\mu_C = \$832.50$

$\sigma_C = \$103$

What happened to the shape of the distribution?

SHAPES ARE THE SAME

What happened to the mean?

$M_C = M_T + 100$

What happened to the standard deviation?

$\sigma_C = \sigma_T$

Effect on a Random Variable of Adding (Subtracting) a Constant

Adding (or subtracting) the same number a to each value of a random variable:
- Adds a to measures of center and location (mean, median, quartiles, percentiles);
- Does not change the shape of the distribution or the measures of spread (range, IQR, standard deviation).

Check Your Understanding - Complete CYU on p. 362.

CARS SOLD(X)	0	1	2	3
PROB	0.3	0.4	0.2	0.1

$M_X = 1.1 \quad \sigma_X = 0.943$

① $Y = BONUS$ $Y = 500X$

$M_Y = 500 M_X = 500(1.1) = \$550$

② $\sigma_Y = 500\sigma_X = 500(0.943) = \$471.50$

$T = 500X - 75$

$M_T = 500 M_X - 75 = 550 - 75 = \$475$

$\sigma_T = 500\sigma_X = \$471.50$

Effects of a Linear Transformation on the Mean and Standard Deviation

If $Y = a + bX$ is a linear transformation of the random variable X, then:

- The probability distribution of Y HAS SAME SHAPE AS PROB. DIST. OF X

- $\mu_Y = a + b M_X$

- $\sigma_Y = |b|\sigma_X$

Problem: In a large introductory statistics class, the distribution of X = raw scores on a test was approximately Normally distributed with a mean of 17.2 and a standard deviation of 3.8. The professor decides to scale the scores by multiplying the raw scores by 4 and adding 10.
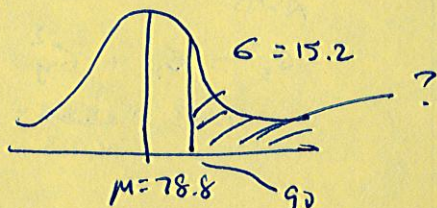
(a) Define the random variable Y to be the scaled score of a randomly selected student from the class. Find the mean and standard deviation of Y.

$$Y = 10 + 4x \implies \mu_y = 10 + 4\mu_x =$$
$$\cancel{\text{for use}} \; 10 + 4(17.2) = \boxed{78.8}$$
$$\sigma_y = 4\sigma_x = 4(3.8) = \boxed{15.2}$$

(b) What is the probability that a randomly selected student has a scaled test score of at least 90?

$\sigma = 15.2$

$M = 78.8$ 90

DIST IS $N(78.8, 15.2)$ SINCE TRANS. DOES NOT Δ SHAPE.

$$P(Y \geq 90) = \boxed{0.2306}$$

---

2. Combining Random Variables

Example (cont). EDCC also has a campus downtown, specializing in just a few fields of study. Full-time students at the downtown campus take only 3-unit classes. Let Y = number of units taken in the fall semester by a randomly selected full-time student at the downtown campus. Here is the probability distribution of Y:

Number of units (Y):	12	15	18		μ_Y = 15 units
Probability:	0.3	0.4	0.3		σ_Y = 2.3 units

If you were to randomly select one full-time student from the main campus and one full-time student from the downtown campus and add their number of units, the expected value of the sum (S = X + Y) would be

$$\mu_S = \mu_X + \mu_y = 14.65 + 15 = 29.65$$

Mean of the Sum of Random Variables

For any two random variables X and Y, if T = X + Y, then the expected value of T is

$$E(T) = \mu_T = \mu_x + \mu_y$$

In general, the mean of the sum of several random variables is the sum of their means.

Definition: If knowing whether any event involving X alone has occurred tells us nothing about the occurrence of any other event involving Y alone, and vice versa, then X and Y are **independent random variables**.

Probability models often assume *independence* when the random variables describe outcomes that appear unrelated to each other. You should always ask whether the assumption of independence is reasonable.

DISCUSS

Example (cont). Let S = X + Y as before. ← TYPO Assume that X and Y are independent, which is reasonable since each student was selected at random. Here are the possible combinations of X and Y and the probability distribution of S:

X	P(X)	Y	P(Y)	S = X + Y	P(S) = P(X)P(Y)
12	0.25	12	0.3	24	0.075
12	0.25	15	0.4	27	0.10
12	0.25	18	0.3	30	0.075
13	0.10	12	0.3	25	0.03
13	0.10	15	0.4	28	0.04
13	0.10	18	0.3	31	0.03
14	0.05	12	0.3	26	0.015
14	0.05	15	0.4	29	0.02
14	0.05	18	0.3	32	0.015
15	0.30	12	0.3	27	0.09
15	0.30	15	0.4	30	0.12
15	0.30	18	0.3	33	0.09
16	0.10	12	0.3	28	0.03
16	0.10	15	0.4	31	0.04
16	0.10	18	0.3	34	0.03
17	0.05	12	0.3	29	0.015
17	0.05	15	0.4	32	0.02
17	0.05	18	0.3	35	0.015
18	0.15	12	0.3	30	0.045
18	0.15	15	0.4	33	0.06
18	0.15	18	0.3	36	0.045

S	P(X)
24	0.075
25	0.03
26	0.015
27	0.19
28	0.07
29	0.035
30	0.24
31	0.07
32	0.035
33	0.15
34	0.03
35	0.015
36	0.045

NOTICE:

$$\mu_S = \mu_X + \mu_Y$$
$$= 29.65 = 14.65 + 15$$

AND

$$\sigma_S^2 = \sigma_X^2 + \sigma_Y^2$$
$$9.63 = \cancel{} + 5.40$$
$$4.23$$

$$\mu_S = (24)(0.075) + (25)(0.03) + \ldots = \boxed{29.65}$$

$$\sigma^2_S = (24 - 29.65)^2(0.075) + (25 - 29.65)^2(0.03) + \ldots =$$
$$\boxed{9.63}$$

★ VARIANCE (NOT STD. DEV.

Variance of the Sum of Independent Random Variables

For any two (independent) random variables X and Y, if T = X + Y, then the variance of T is

$$\sigma^2_T = \sigma_X^2 + \sigma_Y^2$$

(LOOKS LIKE PYTHAG. THM.)

In general, the variance of the sum of several *independent* random variables is the sum of their variances.

Note: On the AP Exam, many students lose credit when combining two or more random variables because they add the standard deviations instead of adding the variances. ★ ★

Problem: Let B = the amount spent on books in the fall semester for a randomly selected full-time student at EDCC. Suppose that $\mu_B = 153$ and $\sigma_B = 32$. Recall from earlier that C = overall cost for tuition and fees for a randomly selected full-time student at EDCC and that $\mu_C = 832.50$ and $\sigma_C = 103$. Find the mean and standard deviation of the cost of tuition, fees and books (C + B) for a randomly selected full-time student at EDCC.

$$\mu_B = 153 \qquad \mu_C = 832.50 \Bigg\} \quad \mu_{C+B} = 153 + 832.50 = \boxed{\$985.50}$$
$$\sigma_B = 32 \qquad \sigma_C = 103$$

$$\sigma^2_{C+B} = \sigma_C^2 + \sigma_B^2$$
$$=$$

★ σ_{C+B} CANNOT BE COMPUTED SINCE COSTS NOT INDEPENDENT

Check Your Understanding - Complete CYU on p. ~~370~~.

SOLD (X): 0   1   2   3 $M_x = 1.1$

PROB: 0.3   0.4   0.2   0.1 $\sigma_x = 0.943$

LEASE (Y): 0   1   2 $M_y = 0.7$

PROB: 0.4   0.5   0.1 $\sigma_y = 0.64$

INDEPENDENT ✓

① $M_T = 1.1 + 0.7 = \boxed{1.8 \text{ CARS}}$

ON AVG, DLRSHP SELLS OR LEASES 1.8 CARS IN 1st HR ON FRIDAYS

② $\sigma_T = \sqrt{\sigma_x{}^2 + \sigma_y{}^2} = \boxed{1.14}$

③ $B = 500X + 300Y$

$M_B = 500 M_x + 300 M_y = \boxed{\$760}$

$\sigma_B = \sqrt{\left((500)(.943)\right)^2 + \left((300)(.64)\right)^2} = \boxed{\$509.09}$

Mean of the Difference of Random Variables

For any two random variables X and Y, if D = X - Y, then the expected value of D is

$$E(D) = \mu_D = M_x - M_y$$

In general, the mean of the difference of several random variables is the difference of their means.

Note: the order of subtraction is important.

Variance of the Difference of Random Variables

For any two *independent* random variables X and Y, if D = X ─ Y, then the variance of D is

$$\sigma^2_D = \sigma_x{}^2 + \sigma_y{}^2$$

Check Your Understanding - Complete CYU on p. ~~372~~.

(SAME DEALERSHIP AS ABOVE)

$D = X - Y$

① $M_D = M_x - M_y = \boxed{0.4} \Rightarrow$ ON AVG, THE DLRSHP SELLS 0.4 MORE CARS THAN IT LEASES DURING THE 1st HR ON FRIDAYS.

② $\sigma_D = \sqrt{\sigma_x{}^2 + \sigma_y{}^2} = \boxed{1.14}$

③ $M_B = 500(M_x) - 300(M_y)$
$= 500(1.1) - 300(0.7) = \boxed{\$340}$

$\sigma_B = \sqrt{\left((500)(.943)\right)^2 + \left((300)(.64)\right)^2} = \boxed{\$509.09}$

3. Combining Normal Random Variables

Example. Suppose that a certain variety of apples have weights that are approximately Normally distributed with a mean of 9 ounces and a standard deviation of 1.5 ounces. If bags of apples are filled by randomly selecting 12 apples, what is the probability that the sum of the weights of the 12 apples is less than 100 ounces?

State: WHAT IS PROB. THAT A RANDOM SAMPLE OF 12 APPLES HAS A TOTAL WEIGHT LESS THAN 100 OUNCES?

Plan: LET X = WT OF RANDOMLY SELECTED APPLE. X IS $N(9, 1.5)$.
LET X_1 = WT OF 1st APPLE, X_2 = WT OF 2ND, etc.
$T = X_1 + X_2 + \cdots X_{12}$

Do: SINCE T IS THE SUM OF 12 INDEPENDENT RANDOM VAR'S
T IS NORMALLY DIST WITH $M_T = M_{X_1} + M_{X_2} + \cdots M_{X_{12}} = 108$
AND VARIANCE $6_T^2 = 6_{X_1}^2 + 6_{X_2}^2 + \cdots + 6_{X_{12}}^2 = 27$.
$6_T = \sqrt{27} = 5.2$ OUNCES.

CALCULATOR: 0.0620

$6 = 5.2$

100 108
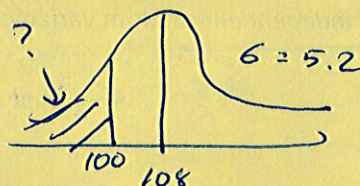
Conclude:

THERE IS ABOUT A 6.2% CHANCE THAT THE 12
✗ RANDOMLY SELECTED APPLES WILL HAVE A TOTAL WT.
LESS THAN 100 OUNCES.

Check Your Understanding - Suppose that the height M of male speed daters follows a Normal distribution with mean 70 inches and standard deviation 3.5 inches, and suppose the height F of female speed daters follows a Normal distribution with a mean of 65 inches and a standard deviation of 3 inches. What is the probability that a randomly selected male speed dater is taller than the randomly selected female speed dater with whom he is paired?
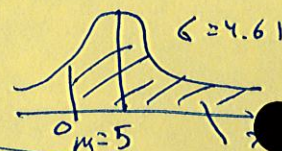
$M_M = 70$
$6_X = 3.5$
$M_F = 65$
$6_F = 3$

① STATE: WHAT IS PROB THAT RANDOMLY SELECTED MALE IS TALLER THAN RANDOMLY SELECTED FEMALE WITH WHOM HE IS PAIRED?

② PLAN: LET $D = M - F$ REPRESENT DIFF. IN M'S HT AND F'S HT.
OUR GOAL IS TO FIND $P(M > F) \Rightarrow$
$P(D > 0)$ 35b

③ DO: SINCE D IS DIFF IN 2 INDEP. NORMAL RV'S, D IS NORMALLY DIST WITH $M_D = M_M - M_F$
$70 - 65 = 5$ INCHES AND VARIANCE
$6_D^2 = 6_M^2 + 6_F^2 = 3.5^2 + 3^2 = 21.25$ INCHES.
∴ $6_D = \sqrt{21.25} = 4.61$
CALC: 0.8610

$6 = 4.61$
$O_M = 5$

④ CONCLUDE

THERE IS ABOUT 86% CHANCE THAT A RANDOMLY
✗ SELECTED MALE WILL BE TALLER.