Lecture Notes & Examples 6.2

Section 6.2 - Transforming and Combining Random Variables (pp. 358-377)

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 and location (mean, median, quartiles, percentiles)
- Does not change shape or measures of spread (IQR, Range, Standard Deviation)

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

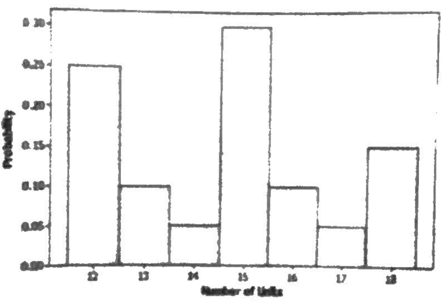
- Multiplies (Divides) measure of center and location by "b". (mean, median, quartiles, %tiles)
- Multiplies (Divides) measures of spread by |b| (Range, IQR, Standard Deviation)
- Does not change the shape of the distribution.

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:



$\mu_x =$ 14.65 units

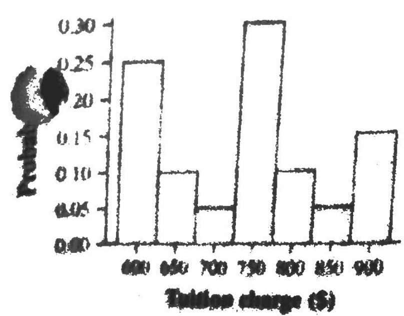
$\sigma_x =$ 2.06 units

The average # of units is 14.65 for any randomly selected full time student in the long run.

On average, the # of credits of any randomly selected student will vary/differ from the mean 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



$\mu_T =$ 732.5 units

$\sigma_T =$ 102.8 units

What happened to the shape of the distribution?
Shape Stayed The Same

What happened to the mean and standard deviation?
$\mu_T = 50\mu_x$ $\sigma_T = 50\sigma_x$

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

Multiplying (or dividing) each value of a random variable by a number *b*:

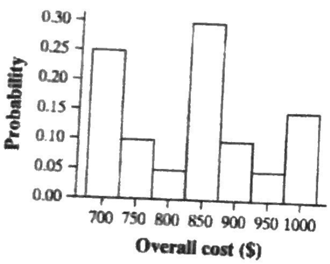
- Multiplies (divides) measures of center and location (mean, median, quartiles, percentiles) by *b*;
- Multiplies (divides) measures of spread (range, IQR, standard deviation) by $|b|$.
- Does not change the shape of the distribution.

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 = T+100

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



μ_C = \$832.50

σ_C = \$102.80

What happened to the shape of the distribution?

shape stayed the same

What happened to the mean?

$\mu_C = \mu_T + 100$

What happened to the standard deviation?

$\sigma_T = \sigma_C$ stayed the same

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

A large auto dealership keeps track of sales made during each hour of the day. Let X = the number of cars sold during the first hour of business on a randomly selected Friday. Based on previous records, the probability distribution of X is as follows:

X: Cars sold:	0	1	2	3
Probability:	0.3	0.4	0.2	0.1

The random variable X has mean µX = 1.1 and standard deviation σX = 0.943.

1. Suppose the dealership's manager receives a $500 bonus from the company for each car sold. Let Y = the bonus received from car sales during the first hour on a randomly selected Friday. Find the mean and standard deviation of Y.

Y = bonus received $Y = 500X$

so $\mu_y = 500\mu_x$ so $\sigma_y = 500\sigma_x$

$\mu_y = 500(1.1)$ $\sigma_y = 500(.943)$

$\mu_y = \$550$ $\sigma_y = \$471.50$

Multiplying each X by 500 will also multiply the μ_x and σ_x by 500 too.

2. To encourage customers to buy cars on Friday mornings, the manager spends $75 to provide coffee and doughnuts. The manager's net profit T on a randomly selected Friday is the bonus earned minus this $75. Find the mean and standard deviation of T.

T = net profit

$T = 500X - 75$

$T = Y - 75$

$\mu_T = 500\mu_x - 75$

$= \mu_y - 75$

$= 550 - 75$

$\mu_T = \$475$

$\sigma_T = \sigma_y$

$= \$471.50$ stays the same

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 the same shape as Prob Dist of X
- $\mu_Y = b\mu_X + a$ or $\mu_Y = a + b\mu_X$
- $\sigma_Y = |b|\sigma_X$

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.

$N_X(17.2, 3.8)$

(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.

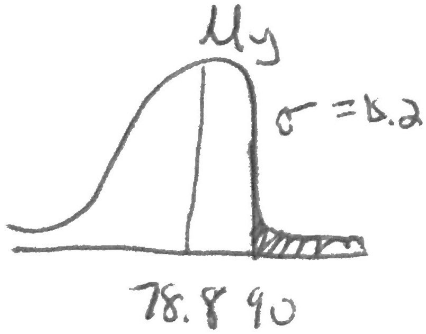
$Y = 4X + 10$ or $Y = 10 + 4X$

$\mu_Y = 4\mu_X + 10$
$= 4(17.2) + 10$
$= 78.8$

$\sigma_Y = 4(\sigma_X)$
$= 4(3.8)$
$= 15.2$

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

$P(Y \geq 90)$ Distribution is $N(78.8, 15.2)$
transformation does not change shape



normalcdf(90, 10,000, 78.8, 15.2)
$= .2306$

The prob that randomly selected student has a scaled score of at least 90 is 23.06%

2. Combining Random Variables

from previous example $\mu_X = 14.65$

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
Probability:	0.3	0.4	0.3

$\mu_Y = 15$ units

$\sigma_Y = 2.3$ 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

$$\mu_S = \mu_X + \mu_Y = 14.65 + 15 = 29.65 \text{ units}$$

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.

In the examples above its reasonable to treat random variables X and Y as independent

$\sigma_x = 2.06$ $\sigma_y = 2.3$

Example (cont). Let S = X + Y as before. 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

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

$\sigma^2_S = \sigma^2_x + \sigma^2_y$
$= (2.06)^2 + (2.3)^2$
$= 9.5336 \approx 9.6$

$\mu_S = 29.65$

$\sigma^2_S = (3.1028)^2 \approx 9.63$

variance

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

$\sigma^2_T = \sigma^2_x + \sigma^2_y$ $\quad \sigma_T = \sqrt{\sigma^2_x + \sigma^2_y}$

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

cannot be calculated if they are not Independent

* *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.

$\mu_B = 153$ $\quad \mu_C = 832.50$ $\quad \mu_{C+B} = \mu_C + \mu_B$

$\sigma_B = 32$ $\quad \sigma_C = 103$ $\quad = 832.50 + 153$

$\boxed{\mu_{C+B} = \$985.50}$

σ_{C+B} cannot be calculated because cost for tuition and fees (C) and cost for books (B) are not independent. (More credit hours typically buy more books)

CHECK YOUR UNDERSTANDING

A large auto dealership keeps track of sales and lease agreements made during each hour of the day. Let X = the number of cars sold and Y = the number of cars leased during the first hour of business on a randomly selected Friday. Based on previous records, the probability distributions of X and Y are as follows:

X

Cars sold x_i:	0	1	2	3
Probability p_i:	0.3	0.4	0.2	0.1

Mean: $\mu_X = 1.1$ Standard deviation: $\sigma_X = 0.943$

Y

Cars leased y_i:	0	1	2
Probability p_i:	0.4	0.5	0.1

Mean: $\mu_Y = 0.7$ Standard deviation: $\sigma_Y = 0.64$

Define T = X + Y.

1. Find and interpret μ_T,

$\mu_T = \mu_x + \mu_y = 1.1 + 0.7 = 1.8$

On average, this dealership sells or leases 1.8 cars in the 1st hours of business on Fridays.

2. Compute σT assuming that X and Y are independent. Show your work.

$\sigma^2_T = \sigma^2_x + \sigma^2_y$ $\sigma_T = \sqrt{\sigma^2_x + \sigma^2_y}$

$\sigma_T = \sqrt{(.943)^2 + (.64)^2} \approx 1.14$

3. The dealership's manager receives a $500 bonus for each car sold and a $300 bonus for each car leased. Find the mean and standard deviation of the manager's total bonus B. Show your work.

$B = \$500X + \$300Y$

$\mu_B = 500\mu_x + 300\mu_y$

$\mu_B = 500(1.1) + 300(0.7) = \$760$

$\sigma_B = \sqrt{(500 \cdot \sigma_x)^2 + (300 \cdot \sigma_y)^2}$

$\sigma_B = \sqrt{(500 \cdot .943)^2 + (300 \cdot .64)^2} \approx \$509.09$

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

$$E(D) = \mu_D = \mu_X - \mu_Y$$

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

$$\sigma^2_D = \sigma^2_X + \sigma^2_Y \qquad \sigma_D = \sqrt{\sigma^2_X + \sigma^2_Y}$$

CHECK YOUR UNDERSTANDING

A large auto dealership keeps track of sales and lease agreements made during each hour of the day. Let X = the number of cars sold and Y = the number of cars leased during the first hour of business on a randomly selected Friday. Based on revious records, the probability distributions of X and Y are as follows:

Cars sold x_i:	0	1	2	3
Probability p_i:	0.3	0.4	0.2	0.1

Mean: $\mu_X = 1.1$ Standard deviation: $\sigma_X = 0.943$

Cars leased y_i:	0	1	2
Probability p_i:	0.4	0.5	0.1

Mean: $\mu_Y = 0.7$ Standard deviation: $\sigma_Y = 0.64$

Define D = X – Y.

1. Find and interpret μD

$\mu_D = \mu_x - \mu_y = 1.1 - 0.7 = 0.4$

On average, this dealership sells 0.4 cars more than it leases during 1st hours of business on Fridays.

. Compute σD assuming that X and Y are independent. Show your work.

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

$\sigma_D = \sqrt{(.943)^2 + (.64)^2} \approx 1.14$

e dealership's manager receives a \$500 bonus for each car sold and a \$300 bonus for each car leased. Find the and standard deviation of the difference in the manager's bonus for cars sold and leased. Show your work.

$B = \$500X - \$300Y = \$500\mu_x - \$300\mu_y$ $\sigma_B = \sqrt{(500\sigma_x)^2 + (300\sigma_y)^2}$

$\mu_B = 500(1.1) - \$300(.7) = \$340$

$\sigma_B = \sqrt{(500 \cdot .943)^2 + (300 \cdot .64)^2}$

$\sigma_B = \$509.09$

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 the prob that the sum of 12 r.s. apples is less than 100 oz?

Plan: X = weight of randomly selected apple $N_x(9, 1.5)$

Let x_1 = weight of 1st apple, x_2 = weight of 2nd apple, ... x_{12}....

$T = x_1 + x_2 + x_3 + \dots x_{12}$

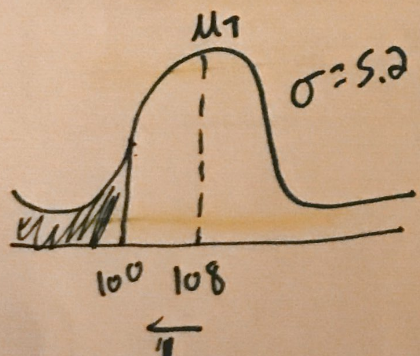
Do:

$P(T < 100)$ T is normally distributed

$N_T(108, 5.2)$ $\mu_T = \mu_{x_1+x_2+x_3\dots} = 9+9+9+\dots \frac{9}{12} = 9(12) = 108$

$\sigma_T = \sqrt{(\sigma_{x_1})^2 + (\sigma_{x_2})^2 + \dots \underline{(\sigma_{x_{12}})^2}_{12}}$

Conclude:



$\sigma_T = \sqrt{(1.5)^2 + (1.5)^2 + \dots \frac{(1.5)^2}{12}} = \sqrt{12(1.5)^2}$

$\sigma_T \approx 5.2$ oz

normalcdf(-10,000, 100, 108, 5.2) = .06

Conclude: There is about a 6% chance that the 12 r.s. apples will have a total weight less than 100 oz.

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?

State: What is the prob that a r.s. male speed dater is taller than the r.s. female speed dator with whom he is paired?

Plan: M = height of male speed dater $N(70, 3.5)$

F: height of female speed dater $N(65, 3)$

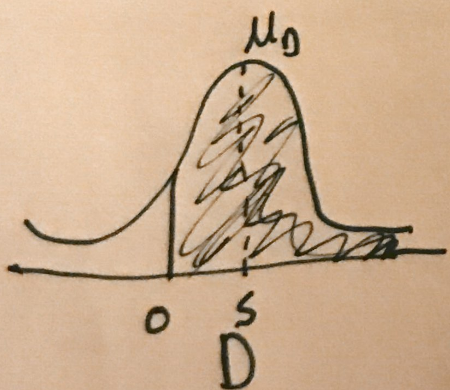
$\underline{D} = M - F$ difference between male's and female's height

Do: $P(M > F) = \underbrace{P(D > 0)}_{\text{want}}$ $N_D(\mu_D, \sigma_D)$

$N_D(5, 4.61)$

$\mu_D = \mu_m - \mu_F = 70 - 65 = 5$ in

$\sigma_D = \sqrt{(\sigma_m)^2 + (\sigma_F)^2} = \sqrt{(3.5)^2 + (3)^2} = \sqrt{21.25} \approx 4.61$



normalcdf$(0, 10{,}000, 5, 4.61) \approx .86$

Conclude: There is about 86% chance that the r.s. male speed dater is taller than the female is randomly paired with.