AP Statistics – 6.2A	Name: KEY
Goal: Linear Transformations of Random Variable	Date:

I. Multiply a RV by a constant

Example: ***El Dorado Community College***

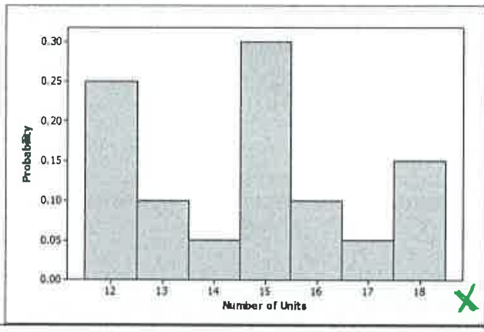
El Dorado Community College considers a student to be full-time if he or she is taking between 12 and 18 units. The number of units X that a randomly selected El Dorado Community College full-time student is taking in the fall semester has the following distribution:

	Number of Units (X)	12	13	14	15	16	17	18
L1 →								
L2 →	Probability	0.25	0.10	0.05	0.30	0.10	0.05	0.15

STATPLOT L1, L2

WINDOW
X Y
12 0
19 .3
1 .05

a) Here is a histogram of the probability distribution



b) Find the mean and standard deviation for X

1 VAR STATS
LIST (L1)
FREQLIST (L2)

$\Sigma x \Rightarrow \mu_X = 14.65$

$\sigma_X = 2.056$

c) At El Dorado Community College, the tuition for full-time students is $50 per unit.

a) **Define the Random Variable**: T = tuition charge for a randomly selected full-time student

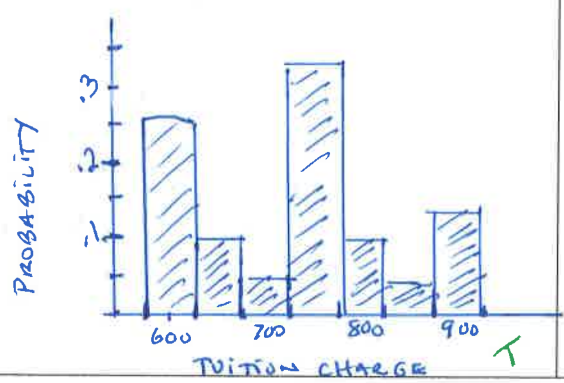
b) Use T = 50X to complete the new probability distribution for T : T = L3 = L1 * 50

	Tuition Charge (T)	$600	650	700	750	800	850	900
L3 →								
L2 →	Probability	0.25	0.10	0.05	0.30	0.10	0.05	0.15

STATPLOT L3, L2

WINDOW
X Y
600 0
950 .3
50 .05

d) Create a histogram for T.



e) Find the mean and standard deviation for T.

1 VAR STATS
LIST (L3)
FREQLIST (L2)

$\Sigma x \Rightarrow \mu_T = \$732.50$

$\sigma_T = \$102.80$

f) Compare the distributions of the random variables X and T. What do you notice?

- **Shape:** THE SHAPES OF BOTH DISTRIBUTIONS ARE THE SAME.
- **Center:** The mean (μ_T) is 50 Times bigger than μ_X : 732.5 = 50(14.65)
- **Spread:** The std dev (σ_T) is 50 times bigger than σ_X : 102.8 = 50(2.056)

II. EFFECTS OF ADDING A CONSTANT TO RV:

In addition to tuition charges, each full-time student at El Dorado Community College is assessed student fees of \$100 per semester.

a) **Define the Random Variable**: C = overall cost for a randomly selected full-time student

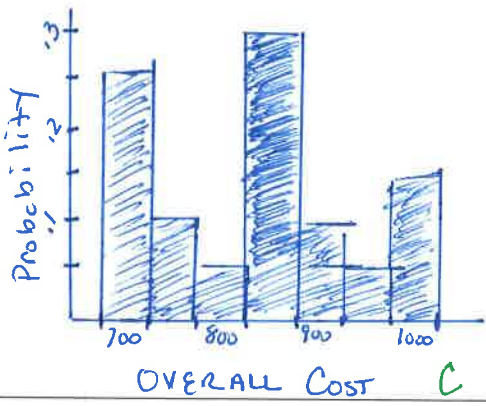
b) *Use C = 100 + T to complete the new probability distribution for C:* C = L4 = L3+100

L4 → Overall Cost (C)	\$700	750	800	850	900	950	1,000
L2 → Probability	0.25	0.10	0.05	0.30	0.10	0.05	0.15

STAT PLOT L4, L2

Window
x y
700 0
1,050 .3
50 .05

c) *Create a histogram for C*



d) *Find the mean and standard deviation for C*

1 VAR STAT
LIST (L4)
FREQLIST (L2)

$\Sigma x \longrightarrow \mu_C = \$832.50$

$\sigma_C = \$102.80$

e) Compare the distributions of the random variables C and T. What do you notice?

- **Shape:** THE SHAPES OF BOTH DISTRIBUTIONS ARE THE SAME.
- **Center:** THE MEAN (μ_C) IS \$100 LARGER THAN μ_T: 832.5 = 732.5+100
- **Spread:** THE STANDARD DEVIATIONS ARE THE SAME $\sigma_C = \sigma_T = \$102.50$

CONCLUSION: SUMMARIZE THE RULES FOR Random Variable LINEAR TRANSFORMATIONS

① LINEAR TRANSFORMATIONS – ADD/SUBTRACT CONSTANTS:

SHAPE: The shape of the distribution does NOT change

CENTER: ADD/SUBTRACT CONSTANTS TO ALL MEASURES OF CENTER (Mean, Median)

SPREAD: DOES NOT CHANGE INCLUDING: STD DEV, VARIANCE, RANGE, IQR.

② LINEAR TRANSFORMATIONS – MULTIPLY/DIVIDE CONSTANTS:

SHAPE: THE SHAPE OF THE DISTRIBUTION DOES NOT CHANGE

CENTER: MULTIPLY MEASURES OF CENTER BY CONSTANT

SPREAD: MULTIPLY MEASURES OF SPREAD BY THE CONSTANT

③ CONCLUSION:

IF $Y = a + bX$ is a linear transformation of the RV X THEN:

- The probability distribution of Y has the same probability distribution of X.
- $\mu_Y = a + b\mu_X$
- $\sigma_Y = |b|\sigma_X$ (since b could be a negative number)

III LINEAR TRANSFORMATION OF A NORMALLY DISTRIBUTED RV

Example: *"Scaling a Test"*

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 variable Y to be the scaled score of a randomly selected student from this class. Find the mean and standard deviation of Y.

X = raw score of a test N(17.2, 3.8)

Y = Scaled test score of a randomly selected STUDENT FROM THE CLASS

$Y = 4X + 10$

$E(Y) = \mu_Y = 4(17.2) + 10 = 78.8$

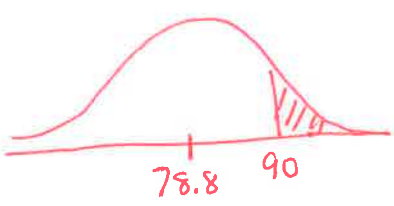
$SD(Y) = \sigma_Y = 4(3.8) = 15.2$

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

* Since linear transformations do NOT CHANGE THE SHAPE, Y HAS the N(78.8, 15.2)

* SKETCH GRAPH. STANDARDIZE SCORE.

$z = \frac{90 - 78.8}{15.2} = .737$



* STATE THE PROBABILITY. FIND THE AREA

$P(Z \geq .737) = .2306$

normalcdf(.737, E99, 0, 1)

normalcdf(.74, E99, 0, 1) = .2296

Approximately 23% of the students scored at least a 90 on the scaled test

IV. Introduction to Linear Combining Random Variables

Example 48 on page 379:

Let the random variable D=X-Y

Exercises 47 and 48 refer to the following setting. Two independent random variables X and Y have the probability distributions, means, and standard deviations shown.

X:	1	2	5
P(X):	0.2	0.5	0.3

$\mu_X = 2.7, \sigma_X = 1.55$

Y:	2	4
P(Y):	0.7	0.3

$\mu_Y = 2.6, \sigma_Y = 0.917$

IMPORTANT → Since the RV's are independent $P(X=x \text{ and } Y=y) = P(X=x) \cdot P(Y=y)$

(a) Define the sample space

X		Y	D = X − Y
1	−	2	−1
1	−	4	−3
2	−	2	0
2	−	4	−2
5	−	2	3
5	−	4	1

(b) Compute the probabilities and summarize in the probability table below:

IMPORTANT: TO CALCULATE P(D) − RV's X and Y must be independent

D	P(D)	Compute probabilities
−3	.06	1 + (−4) = −3 → (.2)(.3)
−2	.15	2 + (−4) = −2 → (.5)(.3)
−1	.14	1 + (−2) = −1 → (.2)(.7)
0	.35	2 + (−2) = 0 → (.5)(.7)
1	.09	5 + (−4) = 1 → (.3)(.3)
3	.21	5 + (−2) = 3 → (.3)(.7)
		↑ X ↑ Y
Total	1.00	

(c) Find the mean and variance of D

TI 84: L1 = D, L2 = P(D), 1-VAR STAT, LIST: L1, FREQLIST: L2

$E(D) = \mu_D = .1$

$SD(D) = \sigma_D = 1.80$

$VAR(D) = \sigma_D^2 = (1.80)^2 = 3.24$

(d) Show the mean of D is equal to $\mu_X - \mu_Y$

$\mu_D = \mu_X - \mu_Y$

$.1 = 2.7 - 2.6 = .1$ ← These are EQUAL.

(e) Show the variance of D is equal to $\sigma^2_X + \sigma^2_Y$

$\sigma_D^2 = \sigma_X^2 + \sigma_Y^2$

$3.24 = (1.55)^2 + (.917)^2 = 3.24$ ← THESE ARE EQUAL OR MAY BE SLIGHTLY OFF DUE TO ROUNDING

(f) Why are the variances added and not subtracted?

- Remember Variance is the actual deviation from the mean (so we add σ^2's and not σ's)
- The more RV's we add, means the more variability we will have

Why do they call this the "Pythagorean Theorem of Statistics"?

Geometry about area SQUARED
⇕
Statistics about SQUARED DEVIATIONS FROM THE MEAN

MUST BE A RT Δ
⇕
MUST BE INDEPENDENT