GREEN SHEET $E(x) = \mu_x = \sum x_i p_i$ $VAR(x) = \sigma_x^2 = \sum (x_i - \mu_x)^2 p_i$

AP Statistics – 6.1	Name: KEY
Goal: Understanding Discrete and Continuous Random Variables	Date:

I. Warm Up CYU (page 349): *Discrete RV's*

- Always state what your random variable is: **X =** THE NUMBER OF CARS SOLD DURING THE 1ST HOUR OF BUSINESS ON A RANDOMLY SELECTED FRIDAY
- Calculate the mean of X by hand and interpret in context

$E(x) = \mu_x = 0(.3) + 1(.4) + 2(.2) + 3(.1) = 0 + .4 + .4 + .3 = \boxed{1.1}$

NOTE: $E(x) = \sum x_i p_i$

* THE LONG-RUN AVERAGE, OVER MANY FRIDAY MORNINGS, WILL BE ABOUT 1.1 CARS SOLD.

- Calculate the standard deviation of X by hand and interpret in context

$VAR(x) = \sigma_x^2 = \sum (x_i - \mu_x)^2 \cdot p_i = (0-1.1)^2(.3) + (1-1.1)^2(.4) + (2-1.1)^2(.2) + (3-1.1)^2(.1) = .363 + .004 + .162 + .361 = \boxed{.89}$

$SD(x) = \sigma_x = \sqrt{.89} = .943$

* ON AVERAGE, THE NUMBER OF CARS SOLD ON A RANDOMLY SELECTED FRIDAY WILL DIFFER FROM THE MEAN (1.1) BY ABOUT .94 CARS SOLD

II. Discrete Random Variables

***Example*: "National Hockey League's Goals"**

In 2010, there were 1319 games played in the regular NHL season. Imagine selecting one of these games at random and then randomly selecting one of the two teams that played in the game. Define the random variable X = number of goals scored by a randomly selected team in a randomly selected game. **The table below gives the probability distribution of X:**

Goals:	0	1	2	3	4	5	6	7	8	9
Probability:	0.061	0.154	0.228	0.229	0.173	0.094	0.041	0.015	0.004	0.001

1ST ALWAYS STATE THE RV→ X= # of goals scored by a randomly selected team in a randomly selected game.

(a) Explain that the probability distribution for X is legitimate.

① ALL THE PROBABILITIES (p_i) IS A NUMBER BETWEEN 0 AND 1

② THE SUM OF THE PROBABILITIES IS 1 ($\sum p_i = 1$)

GREEN SHEET $E(x) = \mu_x = \sum x_i p_i$ $VAR(x) = \sigma_x^2 = \sum (x_i - \mu_x)^2 p_i$

AP Statistics – 6.1	Name: KEY
Goal: Understanding Discrete and Continuous Random Variables	Date:

I. Warm Up CYU (page 349): *Discrete RV's*

- Always state what your random variable is: X = THE NUMBER OF CARS SOLD DURING THE 1ST HOUR OF BUSINESS ON A RANDOMLY SELECTED FRIDAY
- Calculate the mean of X by hand and interpret in context

$E(x) = \mu_x = 0(.3) + 1(.4) + 2(.2) + 3(.1) = 0 + .4 + .4 + .3 = \boxed{1.1}$

NOTE: $E(x) = \sum x_i p_i$

* THE LONG-RUN AVERAGE, OVER MANY FRIDAY MORNINGS, WILL BE ABOUT 1.1 CARS SOLD.

- Calculate the standard deviation of X by hand and interpret in context

$VAR(x) = \sigma_x^2 = \sum (x_i - \mu_x)^2 \cdot p_i = (0-1.1)^2(.3) + (1-1.1)^2(.4) + (2-1.1)^2(.2) + (3-1.1)^2(.1) = .363 + .004 + .162 + .361 = \boxed{.89}$

$SD(x) = \sigma_x = \sqrt{.89} = .943$

* ON AVERAGE, THE NUMBER OF CARS SOLD ON A RANDOMLY SELECTED FRIDAY WILL DIFFER FROM THE MEAN (1.1) BY ABOUT .94 CARS SOLD

II. Discrete Random Variables

Example: *"National Hockey League's Goals"*

In 2010, there were 1319 games played in the regular NHL season. Imagine selecting one of these games at random and then randomly selecting one of the two teams that played in the game. Define the random variable X = number of goals scored by a randomly selected team in a randomly selected game. **The table below gives the probability distribution of X:**

Goals:	0	1	2	3	4	5	6	7	8	9
Probability:	0.061	0.154	0.228	0.229	0.173	0.094	0.041	0.015	0.004	0.001

1ST ALWAYS STATE THE RV→ X= # of goals scored by a randomly selected team in a randomly selected game.

(a) Explain that the probability distribution for X is legitimate.

① ALL THE PROBABILITIES (p_i) IS A NUMBER BETWEEN 0 AND 1

② THE SUM OF THE PROBABILITIES IS 1 ($\sum p_i = 1$)

II. *Example: "National Hockey League's Goals" (continued)*

Goals:	0	1	2	3	4	5	6	7	8	9
Probability:	0.061	0.154	0.228	0.229	0.173	0.094	0.041	0.015	0.004	0.001
E(x)= $x_i \cdot p_i$	0	.154	.456	.687	.692	.470	.246	.105	.032	.009

ALWAYS DEFINE RV→ ***X = number of goals scored by a randomly selected team in a randomly selected game***

(b) Make a histogram of the probability distribution. Describe the distribution in context.

CALC COMMANDS L1 = x_i's L2 = p_i's STATPLOT XLIST : L1 FREQ : L2 WINDOW: XMIN = -1 XMAX = 10 XSCL = 1 YMIN = -.1 YMAX = .3 YSCL = .1

You must clearly label axes + scales!

Remember CUSS AND BS

- THE DISTRIBUTION OF GOALS SCORED IS SKEWED RIGHT MEANIN THE NUMBER OF GOALS SCORED IS RELATIVELY SMALL.
- THE CENTER APPEARS TO BE 2 TO 3 GOALS. WITH THE SPREAD OF GOALS IS 0 TO 9 GOALS WITH THE MAJORITY OF GOALS BETWEEN 1 – 4 SCORED.

(c) What is the probability that the number of goals scored by a randomly selected team in a randomly selected game is at least 6 ?

$$P(X \geq 6) = \underset{.041}{P(X=6)} + \underset{.015}{P(X=7)} + \underset{.004}{P(X=8)} + \underset{.001}{P(X=9)} = \boxed{.061}$$

You must Give probability stmt

NOTE FOR DISCRETE RV'S

$P(\geq 6) = P(X > 5)$ same

CONTEXT THE PROBABILITY THAT A RANDOMLY SELECTED TEAM IN A RANDOMLY SELECTED GAME WILL SCORE 6 OR MORE GOALS IS ABOUT 6%.

(d) Compute the mean of the random variable X by hand, clearly show work, and interpret this value in context.

$$E(x) = \mu_X = \sum x_i p_i = \underbrace{0(.061) + \ldots + 9(.001)}_{\text{must show this work!!}} = \boxed{\mu_X = 2.851}$$

(above in the table are #'s)

CONTEXT: THE MEAN NUMBER OF GOALS FOR A RANDOMLY SELECTED TEAM IN A RANDOMLY SELECTED GAME IS 2.851. [THAT IS ...

IF YOU WERE TO REPEAT THE RANDOM SAMPLING PROCESS OVER AND OVER AGAIN, THE MEAN # OF GOALS SCORED WOULD BE ABOUT 2.851 IN THE LONG RUN.

II. Example: "National Hockey League's Goals" (continued)

Goals:	0	1	2	3	4	5	6	7	8	9
Probability:	0.061	0.154	0.228	0.229	0.173	0.094	0.041	0.015	0.004	0.001
VAR(X) $(x_i-\mu_x)^2 \cdot p_i$	.496	.528	.165	.005	.228	.434	.407	.258	.106	.038

ALWAYS DEFINE RV→ ***X = number of goals scored by a randomly selected team in a randomly selected game***

(e) Compute the standard deviation of the random variable X by hand, clearly show work, and interpret this value in context.

$VAR(X) = \sigma_x^2 = \sum (x_i - \mu_x)^2 \cdot p_i = \underbrace{(0-2.851)^2 \cdot (.061) + \ldots + (9-2.851)^2 \cdot (.001)}$

must show this work (the above table has the numbers)

$VAR(x) = \sigma_x^2 = 2.66$ ← The variance

$SD(x) = \sigma_x = \sqrt{2.66} = \boxed{1.63}$

CONTEXT: ON AVERAGE, A RANDOMLY SELECTED TEAM'S NUMBER OF GOALS IN A RANDOMLY SELECTED GAME WILL DIFFER FROM THE MEAN (2.851) BY ABOUT 1.63 GOALS.

USE CALC FOR Mean + S.D of RV's

STAT CALC 1 VAR STAT

LIST: $\boxed{L1}$ x_i's

FREQ LIST: $\boxed{L2}$ p_i's

RV's are population NOT sample

$\bar{x}$ ← Never say $\bar{x}$!

$\Sigma x = 8.128$ ← E(X)

$Sx = \cancel{\ }$ Not a sample stat

$\sigma x = 1.437$ ← SD(x)

III. Technology Corner (page 348) Example: "Apgar Scores (example on page 343)"

a) Define the Random Variable → X= THE APGAR SCORE OF A RANDOMLY SELECTED BABY.

b) Explain that the probability distribution for X is legitimate.

- Fill in table and place data in calculator lists (L1-L2)
- Using 1-VarStats to determine if probability distribution for X is legitimate. **EXPLAIN**:

See pg 343

x_i	p_i
0	.001
1	.006
2	.007
3	.008
4	.012
5	.020
6	.038
7	.099
8	.319
9	.437
10	.053
Totals	1.00

X is a legitimate probability distribution because

① all the individual probabilities are between 0 and 1.

② the sum of the probabilities is 1.

STAT Calc

1 Var stat

LIST: $\boxed{L2}$

$\Sigma x = 1$

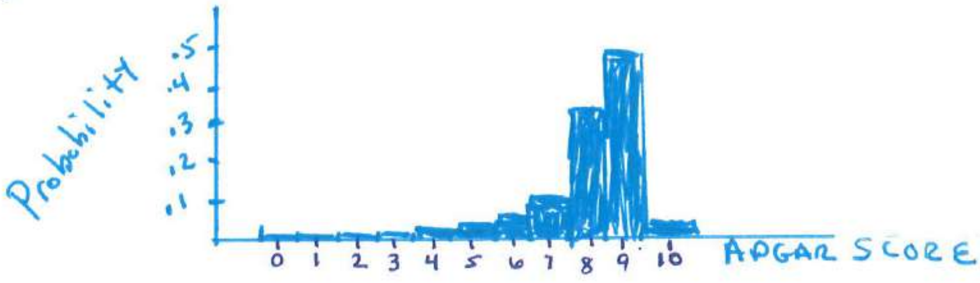
(continued) III. Technology Corner: "Apgar Scores" Example

c) ***(step 2 from the "Technology Corner")*** Make a histogram of the probability distribution.

CALC COMMANDS: L1 = x_i's, L2 = p_i's

STATPLOT: XLIST = L1, FREQ = L2

SETWINDOW: Xmin = -1, XMAX = 11, XSCL = 1, YMIN = -.1, YMAX = .5, YSCL = .1



Describe the distribution in context.

THE DISTRIBUTION OF APGAR SCORES IS HEAVILY SKEWED TO THE LEFT WITH AN EXPECTED SCORE AROUND 9 FOR A NORMAL BABY WITH MOST SCORES FALLING BETWEEN 7 and 9.

d) ***(step 3 from the "Technology Corner")*** Compute the mean of the random variable X, clearly show work, and interpret this value in context.

$E(x) = 0(.001) + \dots + 10(.053) = 8.128$

see Pg 348 FOR CALC Commands

This is the work you must show!

CONTEXT:

THE MEAN APGAR SCORE OF A RANDOMLY SELECTED NEWBORN IS 8.128, THIS IS THE LONG-RUN AVERAGE APGAR SCORE OF MANY, MANY, MANY randomly chosen babies.

e) ***(step 3 from the "Technology Corner")*** Compute the standard deviation of the random variable X, clearly show work, and interpret this value in context.

$VAR(x) = (0 - 8.128)^2(.001) + \dots + (10 - 8.128)^2(.053) =$

$SD(x) = \sqrt{VAR(x)} = \boxed{1.437}$

This is the Work you must show

CONTEXT:

ON AVERAGE, A RANDOMLY SELECTED BABY'S APGAR SCORE WILL DIFFER FROM THE MEAN (8.128) BY ABOUT 1.4 UNITS.

IV **"DISCRETE Random Variable"** VOCABULARY You need to understand these definitions for this chapter. See my web site for these definitions. Here is space for you to take your own notes:

- **Random Variable** (p313) TAKES NUMERICAL VALUES THAT DESCRIBE THE OUTCOMES OF SOME CHANCE PROCESS.
  * RV'S ARE DENOTED WITH CAPITAL LETTERS (X)

- **Probability Distribution** OF A RANDOM VARIABLE GIVES ITS POSSIBLE VALUES AND THEIR PROBABILITIES. THIS IS A DISTRIBUTION LIKE IN EARLIER CHAPTER. Remember to graph and DESRIBE WITH "CUSS and BS".

- **Discrete Random Variable and Their Probability Distribution** (p313)

  DISCRETE RV's (X) takes a fixed set of possible Values.

  * GAPS MAY OCCUR
  * PROBABILITY DISTRIBUTION:

X	all possible outcomes
Value	x_1 to x_i
Probabilities	p_1 to p_i

- *The probabilities must satisfy 2 requirements*
  1) p_i's are betweel 0 and 1
  2) The sum of probabilities is 1 $\Sigma p_i = 1$

  To find the probability of an EVENT, add the p_i's for the x_i's IN THE EVENT.

- **Expected Value of a Discrete Random Variable** (p345)

  The EXPECTED VALUE is the mean of a RV.

$$\boxed{\mu_X = E(x) = \sum x_i p_i}^{*}$$

CONTEXT: IF YOU WERE TO REPEAT THE RANDOM SAMPLING PROCESS OVER AND OVER AGAIN, THE MEAN VALUE would be about ______ in the long run.

(random variable have population means. NEVER use $\overline{X}$. No!!

- **Standard Deviation of a Discrete Random Variable**(p347)

  To find the SD(x), you must first find the variance:

$$\boxed{\sigma_X^2 = VAR(x) = \sum (x_i - \mu_x)^2 \cdot p_i}^{**}$$

CONTEXT OF SD

ON AVERAGE, A RANDOMLY SELECTED PROCESS WILL DIFFER FROM THE MEAN BY ABOUT ______.

* The formulas are on YOUR AP GREEN SHEET

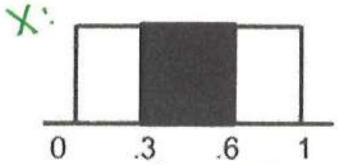
V. Continuous Random Variables

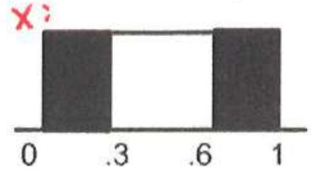
A. VOCABULARY (IMPORTANT)-- "CONTINUOUS Random Variable"

- ***Continuous Probability Distribution*:** described by the area under a density curve
- ***A CONTINUOUS probability distribution differs from a DISCRETE probability distribution in several ways:***

1. The probability that a continuous random variable will assume an exact value is **zero**.
2. As a result, a continuous probability distribution cannot be expressed in tabular form.
3. Instead, an equation or formula is used to describe a continuous probability distribution.
4. We assign probabilities to **intervals** of outcomes rather than to individual outcomes.

EXAMPLE: Find the areas (shaded region) for the following density curves (*uniform distribution*).





$P(.3 \le X \le .6) = .3$

area = $l \cdot w$ =

$1 \cdot .3 = .3$

$1 - P(.3 \le X \le .6) =$

$1 - .3 = .70$

Your Notes: **Continuous Random Variables** (p350) SEE ABOVE NOTES.

IN MANY CASES ... DISCRETE RANDOM VARIABLES ARISE FROM COUNTING something

CONTINUOUS RV'S ARISE FROM MEASURING SOMETHING

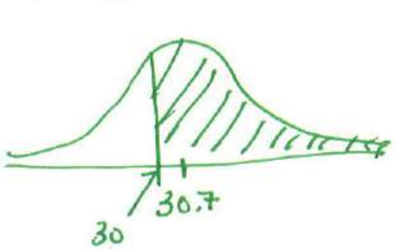
NOTE: $P(-1 \le X \le 2)$ is really the same as $P(-1 < X < 2)$

SINCE THERE IS NO AREA DIRECTLY ABOVE -1 OR 2.

***Example*: "Weights of Three-Year-Old Females"**

The weights of three-year-old females closely follow a Normal distribution with a mean of $\mu = 30.7$ pounds and a standard deviation of 3.6 pounds. Randomly choose one three-year-old female and call her weight X. Find the probability that the randomly selected three-year-old female weighs at least 30 pounds.

1. DEFINE RV: X = THE WEIGHT OF A RANDOMLY CHOSEN 3 YEAR OLD FEMALE
2. STATE DISTRIBUTION: $N(30.7, 3.6)$
3. STATE PROBABILITY OF INTEREST: $P(X \geq 30)$
4. DRAW A PICTURE AND STANDARDIZE THE WEIGHT TO FIND PROBABILITY



$$Z = \frac{30 - 30.7}{3.6} = -.1944$$

$$P(Z \geq -.1944) = .5770$$

normalcdf(-.1944, E99, 0, 1)
DO NOT NEED TO GIVE THIS.

5. CONCLUDE:
THERE IS ABOUT A 58% CHANCE THAT THE RANDOMLY SELECTED 3 YEAR OLD FEMALE WILL WEIGH AT LEAST 30 pounds