NAME: KEY

6.1 Day 0 HW

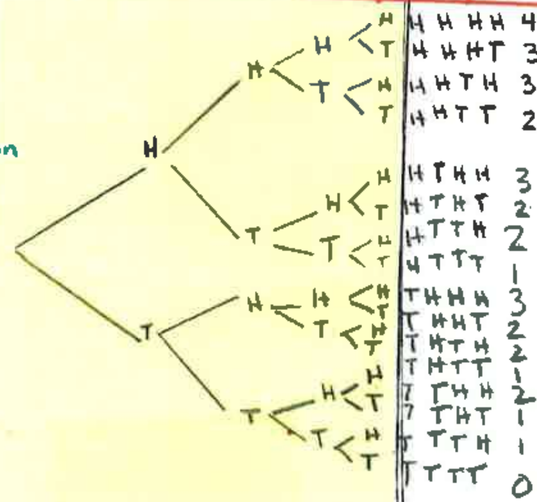
Exercises

1. **Toss 4 times** Suppose you toss a fair coin 4 times. Let X = the number of heads you get.
   (a) Find the probability distribution of X.
   (b) Make a histogram of the probability distribution. Describe what you see.
   (c) Find $P(X \leq 3)$ and interpret the result.

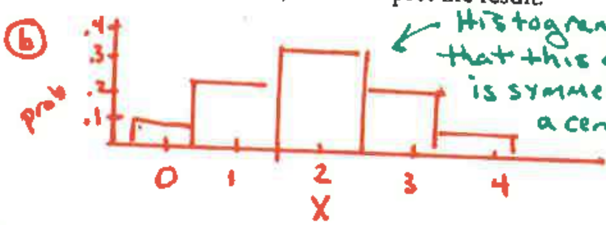
(A)

X = # of heads (4 times)

Value (x_i)	0	1	2	3	4
Probability (p_i)	1/16 .0625	4/16 .25	6/16 .375	4/16 .25	1/16 .0625



(b)



Histogram shows that this distribution is symmetric with a center at 2.

(c) $P(X \leq 3) = 1 - .0625 = .9375$

There is a 93.75% chance that you will get 3 or fewer heads on 4 tosses of a fair coin

2. **Pair-a-dice** Suppose you roll a pair of fair, six-sided dice. Let T = the sum of the spots showing on the up-faces.
   (a) Find the probability distribution of T.
   (b) Make a histogram of the probability distribution. Describe what you see.
   (c) Find $P(T \geq 5)$ and interpret the result.

(a) T = The sum of the spots on 2 dice

Value	2	3	4	5	6	7	8	9	10	11	12
prob	1/36	3/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

1 1 = 2
2 = 3
3 = 4
4 = 5
5 = 6
6 = 7

2 1 = 3
2 = 4
3 = 5
4 = 6
5 = 7
6 = 8

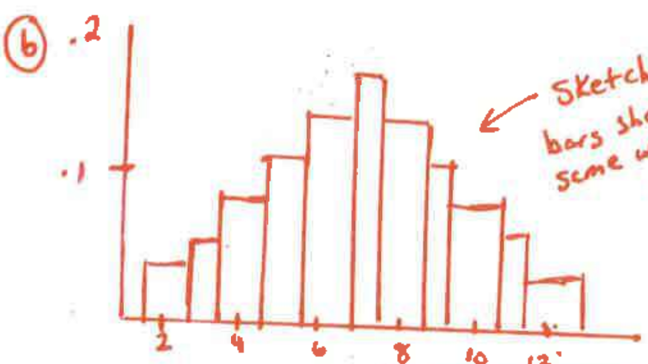
3 1 = 4
2 = 5
3 = 6
4 = 7
5 = 8
6 = 9

4 1 = 5
2 = 6
3 = 7
4 = 8
5 = 9
6 = 10

5 1 = 6
2 = 7
3 = 8
4 = 9
5 = 10
6 = 11

6 1 = 7
2 = 8
3 = 9
4 = 10
5 = 11
6 = 12

(b)



Sketch (NOT TO SCALE) bars should have same width

The histogram shows a symmetric distribution about a center of 7

(c) $P(T \geq 5) = 1 - 6/36 = 5/6$ or $\approx .83$

About 83% of the time, when you roll a pair of dice, you will have a sum of 5 or more

Source: TPS (CH6)

5 Benford's law Faked numbers in tax returns, invoices, or expense account claims often display patterns that aren't present in legitimate records. Some patterns, like too many round numbers, are obvious and easily avoided by a clever crook. Others are more subtle. It is a striking fact that the first digits of numbers in legitimate records often follow a model known as Benford's law.[5] Call the first digit of a randomly chosen record X for short. Benford's law gives this probability model for X (note that a first digit can't be 0):

First digit X:	1	2	3	4	5	6	7	8	9
Probability:	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

(a) Show that this is a legitimate probability distribution.

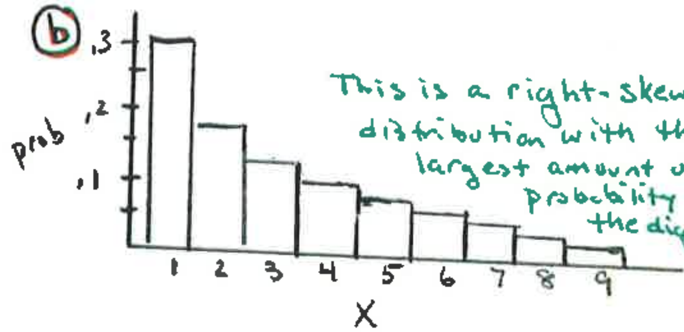
(b) Make a histogram of the probability distribution. Describe what you see.

(c) Describe the event $X \geq 6$ in words. What is $P(X \geq 6)$?

(d) Express the event "first digit is at most 5" in terms of X. What is the probability of this event?

(a) This is a legitimate probability distribution because
① all the probabilities are between 0 and 1
② the probabilities sum to 1.

(b)



This is a right-skewed distribution with the largest amount of probability on the digit one.

(c) The 1st digit in a randomly chosen record is a 6 or higher

$P(X \geq 6) = .067 + .058 + .051 + .046$

$\boxed{P(X \geq 6) = .222}$

(d) The event:

$P(X \leq 5) = 1 - .222 = .778$

7 Benford's law Refer to Exercise 5. The first digit of a randomly chosen expense account claim follows Benford's law. Consider the events A = first digit is 7 or greater and B = first digit is odd.

(a) What outcomes make up the event A? What is $P(A)$?

(b) What outcomes make up the event B? What is $P(B)$?

(c) What outcomes make up the event "A or B"? What is $P(A \text{ or } B)$? Why is this probability not equal to $P(A) + P(B)$?

(a) EVENT $A = (7, 8, 9)$ $P(A) = .058 + .051 + .046 = \boxed{.155}$

(b) EVENT $B = (1, 3, 5, 7, 9)$ $P(B) = .301 + .125 + .079 + .058 + .046 = \boxed{.609}$

(c) EVANT $A \text{ or } B = (1, 3, 5, 7, 8, 9)$ $P(A \cup B) = P(A) + P(B) - P(A \cap B = 7, 9)$
$= .155 + .609 - .104 = \boxed{.66}$

$P(A \text{ or } B) \neq P(A) + P(B)$ because event A and B are NOT mutually exclusive.

9 **Keno** Keno is a favorite game in casinos, and similar games are popular with the states that operate lotteries. Balls numbered 1 to 80 are tumbled in a machine as the bets are placed, then 20 of the balls are chosen at random. Players select numbers by marking a card. The simplest of the many wagers available is "Mark 1 Number." Your payoff is \$3 on a \$1 bet if the number you select is one of those chosen. Because 20 of 80 numbers are chosen, your probability of winning is 20/80, or 0.25. Let X = the amount you gain on a single play of the game.

(a) Make a table that shows the probability distribution of X.

(b) Compute the expected value of X. Explain what this result means for the player.

(a) X = amount you gain on a single play of Keno

VALUE	\$0	\$3
Probability	.75	.25

(b) $E(x) = 0(.75) + 3(.25) = .75$

$E(x) = \mu_x = \$.75$

IN THE LONG RUN, FOR EVERY \$1 THE PLAYER BETS, HE GETS ONLY 75¢ BACK.

6.1

11 To find expected value

15 To find std dev.

Work For #11 + 15

x_i	p_i	$x_i p_i$	$x_i - \mu_x$	$(x_i - \mu_x)^2$	$(x_i - \mu_x)^2 p_i$
0	.1	0	-2.1	4.41	.441
1	.2	.2	-1.1	1.21	.242
2	.3	.6	-0.1	.01	.003
3	.3	.9	0.9	.81	.243
4	.1	.4	1.9	3.61	.361
	1.0	2.1			$\Sigma = 1.290$

$E(x) = \mu_x = \Sigma x_i p_i = 2.1$

$VAR(x) = \sigma_x^2 = \Sigma (x_i - \mu_x)^2 \cdot p_i = 1.29$

$\sigma_x = 1.1358$

On average, undergraduates make 2.1 nonword errors per 250-word essay.

ON average, the number of nonword errors in a randomly selected essay will differ from the mean (2.1) by about 1.14 words

These formula's are on your AP Green sheet

19. Housing in San Jose How do rented housing units differ from units occupied by their owners? Here are the distributions of the number of rooms for owner-occupied units and renter-occupied units in San Jose, California:[7]

	Number of Rooms									
	1	2	3	4	5	6	7	8	9	10
Owned	0.003	0.002	0.023	0.104	0.210	0.224	0.197	0.149	0.053	0.035
Rented	0.008	0.027	0.287	0.363	0.164	0.093	0.039	0.013	0.003	0.003

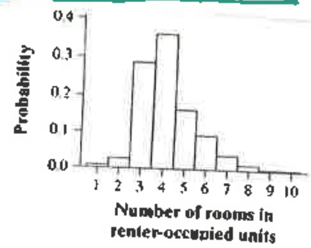
Let X = the number of rooms in a randomly selected owner-occupied unit and Y = the number of rooms in a randomly chosen renter-occupied unit.

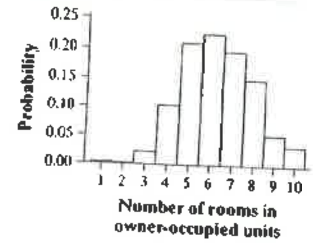
(a) Make histograms suitable for comparing the probability distributions of X and Y. Describe any differences that you observe.

(b) Find the mean number of rooms for both types of housing unit. Explain why this difference makes sense.

(c) Find the standard deviations of both X and Y. Explain why this difference makes sense.

(A) HISTOGRAMS





DESCRIBE: THE DISTRIBUTION OF THE NUMBER OF ROOMS IS ROUGHLY SYMMETRIC FOR OWNERS AND SKEWED TO THE RIGHT FOR RENTERS. OVERALL, RENTER-OCCUPIED UNITS TEND TO HAVE FEWER ROOMS THAN OWNER OCCUPIED UNITS

(B) $\mu_O = 6.284$ $\mu_R = 4.187$

USE CALC

L1 = # ROOMS 1–10
L2 = P(owned)
L3 = L1 * L2
STAT CALC 1 L3 $\boxed{\Sigma x = 6.284}$
L4 = P(rent)
L5 = L1 * L4
STAT CALC 1 L5 $\boxed{\Sigma x = 4.187}$

The mean for renters is about 4.2 rooms, compared to the mean for owners is about 6.3 rooms; which matches our observations from the histograms

(C) See BELOW on how to calculate

$\boxed{\sigma_O = 1.64 \text{ rooms}}$

$\boxed{\sigma_R = 1.31 \text{ ROOMS}}$

We would expect the owner distribution to have a slightly wider spread than the renter distribution. Even though the distribution of renter-occupied units is skewed to the right, it is more concentrated (contains less variability) about the "peak" than the symmetric distribution for owner-occupied units.

$\boxed{\text{VAR}(x) = \sigma_x^2 = (x_1 - \mu_x)^2 p_1 + (x_2 - \mu_x)^2 \cdot p_2 \ldots}$

CALCULATE

LISTS
L1 = # ROOMS 1–10 (x_i's)
L2 = P(OWNED)
L3 = L1 * L2
L4 = P(rent)
L5 = L1 * L4

$\sigma_O = 1.639$ rooms
$\mu_O = 6.284$

Create L6
$L6 = (L1 - 6.284)^2 \cdot L2$

STAT CALC 1 L6
$\Sigma x = 2.689 = \sigma_O^2$

FIND σ_O
$\sqrt{2.689} = 1.6398$

$\sigma_R = 1.3077$ rooms
$\mu_R = 4.187$

Create L6
$L6 = (L1 - 4.187)^2 \cdot L4$

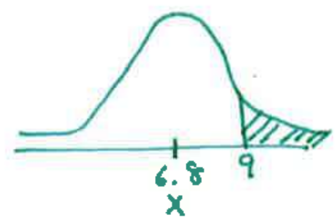
STAT Calc 1 L6
$\Sigma x = 1.710031 = \sigma_R^2$

FIND σ_R
$\sqrt{1.710031} = 1.3077$

23 **ITBS scores** The Normal distribution with mean $\mu = 6.8$ and standard deviation $\sigma = 1.6$ is a good description of the Iowa Test of Basic Skills (ITBS) vocabulary scores of seventh-grade students in Gary, Indiana. Call the score of a randomly chosen student X for short. Find $P(X \geq 9)$ and interpret the result. Follow the four-step process.

① STATE: What is the probability that a randomly chosen student scores a 9 or better on the ITBS?

② PLAN: The score X of the randomly chosen students has the N (6.8, 1.6) distribution. We want to find the $P(X \geq 9)$, we will standardize the scores and find the area under the normal curve.



③ DO: The standardized score for the test:

$$Z = \frac{9 - 6.8}{1.6} = 1.38$$

$P(Z \geq 1.38) = .0838$ 2ND DIST $\longrightarrow$ Normalcdf(1.38, 9999) = .0838

④ Conclude: there is about an 8% chance that the chosen student's score is 9 or higher.

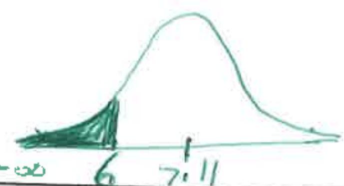
6.1 HOMEWORK

FOUR STEP PROCESS

24. **Running a mile** A study of 12,000 able-bodied male students at the University of Illinois found that their times for the mile run were approximately Normal with mean 7.11 minutes and standard deviation 0.74 minute.[8] Choose a student at random from this group and call his time for the mile Y. Find $P(Y < 6)$ and interpret the result. Follow the four-step process. (See last page for steps)

① STATE: What is the probability that a randomly chosen student runs the mile in under 6 MINUTES.

② PLAN:
- LET Y = the time of a randomly chosen student
- Normal distribution $N(7.11, .74)$
- We want to find $P(Y<6)$
- Draw the density curve



③ DO:

SINCE $N(7.11, .74)$ AND $P(Y<6)$

NORMALCDF(-EE99, 6, 7.11, .74) = .0668

④ CONCLUDE:

There is about a 7% chance that a randomly selected student will run the mile in under 6 minutes.

25. **Did you vote?** A sample survey contacted an SRS of 663 registered voters in Oregon shortly after an election and asked respondents whether they had voted. Voter records show that 56% of registered voters had actually voted. We will see later that in repeated random samples of size 663, the proportion in the sample who voted (call this proportion V) will vary according to the Normal distribution with mean $\mu = 0.56$ and standard deviation $\sigma = 0.019$.

(a) If the respondents answer truthfully, what is $P(0.52 \le V \le 0.60)$? This is the probability that the sample proportion V estimates the population proportion 0.56 within ± 0.04.

(b) In fact, 72% of the respondents said they had voted ($V = 0.72$). If respondents answer truthfully, what is $P(V \ge 0.72)$? This probability is so small that it is good evidence that some people who did not vote claimed that they did vote.

V = proportion in the sample that voted

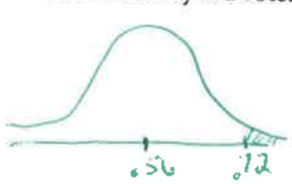
$N(.56, .019)$

(a) $P(.52 \le V \le .60) = .9647$

Normalcdf(.52, .60, .56, .019)

There is about a 96% chance that the population proportion of .56 is within $\pm .04$.

(b)



normalcdf(.72, EE99, .56, .019) = 1.8E-17 (essentially 0)

This confirms the probability that 72% of the respondents answered truthfully is zero(0).