Homework (due Monday, Nov 2)

Chapter 8: #11, 18

Announcements:

- Remember that individual quiz results (which questions you missed and the correct answer) will be available later today for the quiz that just ended.

Examples given last time (solve on board):

1. Suppose there is no relationship between two variables, e.g. listening to Mozart and increased IQ. Suppose 3 independent experiments are done, each using the 0.05 criterion for statistical significance.

   What is the probability that *at least one* experiment results in a statistically significant relationship just by chance?

2. Suppose you drive on a certain freeway daily. Speed limit is 65. You drive over 65 all the time, but over 75 about 30% of the time.

$P(\text{ticket} \mid \text{over } 75) = 1/50 = .02$

$P(\text{ticket} \mid 65 \text{ to } 75) = 1/200 = .005.$

What is the probability you get a ticket on a randomly selected day?

Some definitions from Chapter 7 Probability of accurate medical test

Define the events:

D = person has the disease

D^C = person does not have the disease

T = test is positive

T^C = test is negative

Sensitivity of a test = P(T | D), i.e., correct outcome if person *has* the disease.

Specificity of a test = $P(T^C \mid D^C)$, i.e. correct outcome if person *does not have* the disease.

Random Variables

Chapter 8

We will cover Chapter 8 this week:

Today: 8.1 to 8.3

Wednesday: 8.4, 8.5

Friday: 8.6, 8.7

8.1 What is a Random Variable?

Random Variable: assigns a number to each outcome of a random circumstance, or, equivalently, to each unit in a population.

Two different broad classes of random variables:

1. A **continuous random variable** can take any value in an interval or collection of intervals.
2. A **discrete random variable** can take one of a countable list of distinct values.

Notation for either type: X, Y, Z, W, etc.

Examples of Discrete Random Variables

Assigns a number to each outcome of a random circumstance, or to each unit in a population.

1. Couple plans to have 3 children.

The *random circumstance* includes the 3 births, specifically the sexes of the 3 children.

Possible outcomes include BBB, BBG, etc.

X = number of girls

X is *discrete* and can be 0, 1, 2, 3

For example, the number assigned to BBB is X=0

2. Population consists of UC students (*unit* = student)

Y = number of siblings a student has

Y is *discrete* and can be 0, 1, 2, …??

Examples of Continuous Random Variables

Assigns a number to each outcome of a random circumstance, or to each unit in a population.

1. Population consists of UC female students

   Unit = female student

   W = height

   W is *continuous* – can be anything in an interval, even if we report it to nearest inch or half inch

2. You are waiting at a bus stop for the next bus

   Random circumstance = when the bus arrives

   Y = time you have to wait

   Y is *continuous* – anything in an interval

Today: Discrete Random Variables

X = the random variable.

k = a number the discrete r.v. could assume.

$P(X = k)$ is the probability that X equals k.

***Probability* distribution function (pdf)** for a discrete r.v. X is a table or rule that assigns probabilities to possible values of X.

NOTE: Sometimes the probabilities are given or observed, and sometimes you have to compute them using rules from Ch. 7.

***Cumulative* distribution function (cdf)** is a rule or table that provides $P(X \leq k)$ for every real number k. (More useful for continuous random variables than for discrete, as we will see.)

Conditions for Probabilities for Discrete Random Variables

Condition 1

The *sum of the probabilities* over all possible values of a discrete random variable must equal 1.

Condition 2

The *probability of any specific outcome* for a discrete random variable, $P(X = k)$, must be between 0 and 1.

Note: The possible values *k* are *mutually exclusive*

Example of Computing the PDF and CDF from Chapter 7 Rules

Example: You buy 2 tickets for the Daily 3 lottery

X = number of winning tickets you have, could be 0, 1, 2.

$P(X = 0) = (.999)^2 = .998001$ (Rule 3b) (998,001 in a million)

$P(X = 2) = (.001)^2 = .000001$ (Rule 3b) (1 in a million)

$P(X = 1) = 1 – (.998001 + .000001)$

$= .001998$ (Rule 1) (1998 in a million)

k	pdf *P(X=k)*	cdf *P(X ≤ k)*
0	.998001	.998001
1	.001998	.999999
2	.000001	1.0

Example of *Observed* Probabilities

Survey of 173 students in introductory statistics:

k	Number with k siblings	pdf $P(X=k)$	cdf $P(X \leq k)$
0	14	14/173 = .08	.08
1	68	68/173 = .39	.39 + .08 = .47
2	53	.31	.47 + .31 = .78
3	21	.12	.90
4	8	.05	.95
5	6	.03	.98
6	3	.02	1.00

Clicker data collection

How many siblings (brothers and sisters) do you have? Count half-siblings (share one parent), but not step siblings.

A. 0

B. 1

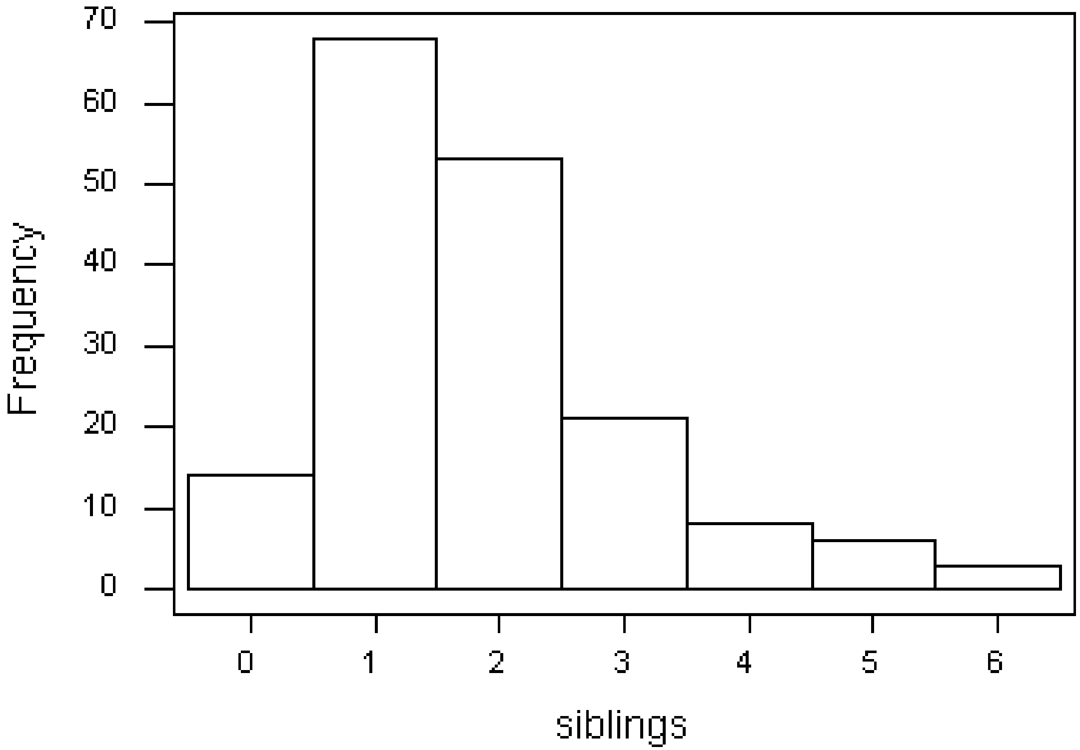
C. 2

D. 3

E. 4 or more

Graph of pdf for number of siblings (with frequency instead of relative frequency)

Compare class results on board.



More Complicated Examples for Discrete R.V.s

Probability distribution function (pdf) X is a table or rule that assigns probabilities to possible values of X.

Using the sample space to find probabilities:

Step 1: List all simple events in sample space.

Step 2: Find probability for each simple event.

Step 3: List possible values for random variable X and identify the value for each simple event.

Step 4: Find all simple events for which $X = k$, for each possible value k.

Step 5: $P(X = k)$ is the sum of the probabilities for all simple events for which $X = k$.

Example 8.6 *How Many Girls are Likely?*

Family has 3 children. Probability of a girl is ½ (simpler!). What are the probabilities of having 0, 1, 2, or 3 girls?

Sample Space: For each birth, write either B or G. There are eight possible arrangements of B and G for three births. These are the *simple events*.

S = {BBB, BBG, BGB, GBB, BGG, GBG, GGB, GGG}

Sample Space and Probabilities: The eight simple events are equally likely. (Could use the .512, .488 probabilities to find exact, but we'll round off by using ½ for each B and G.)

Random Variable X: number of girls in three births. For each simple event, the value of X is the number of G's listed.

Example 8.6 *How Many Girls? (continued)*

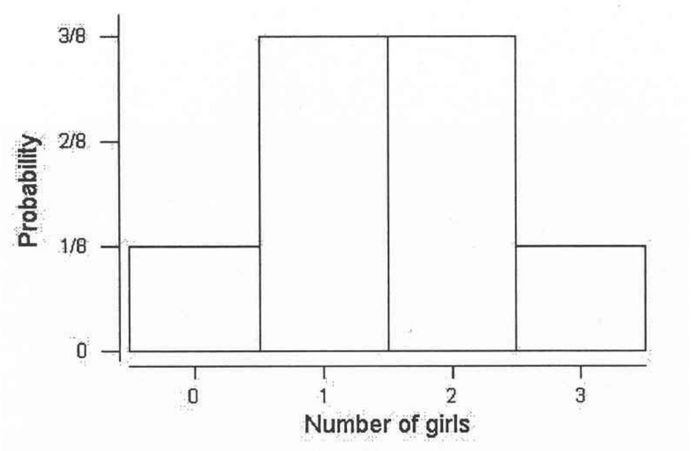
Value of X for each simple event:

Simple Event	BBB	BBG	BGB	GBB	BGG	GBG	GGB	GGG
Probability	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
X = # of Girls	0	1	1	1	2	2	2	3

Probability distribution function for Number of Girls X:

k	0	1	2	3
$P(X = k)$	1/8	3/8	3/8	1/8

Graph of the pdf of X:



Cumulative Distribution Function for number of girls:

Cumulative distribution function (cdf) provides the probabilities $P(X \leq k)$ for any real number k.

Cumulative probability = probability that X is less than or equal to a particular value.

Example 8.6 ***Cumulative Distribution Function for the Number of Girls***

k	0	1	2	3
$P(X \leq k)$	1/8	4/8	7/8	1

For example, the probability is 7/8 that family has ≤ 2 girls.

8.3 Expected Value (Mean) for Random Variables

The **expected value** of a random variable is the **mean** value of the variable *X* in the sample space, or population, of possible outcomes.

If *X* is a random variable with possible values $x_1, x_2, x_3, \ldots,$ occurring with probabilities $p_1, p_2, p_3, \ldots,$
then the **expected value** of *X* is calculated as

$$\mu = E(X) = \sum x_i p_i$$

Example of expected value

Number of siblings for intro stat students:

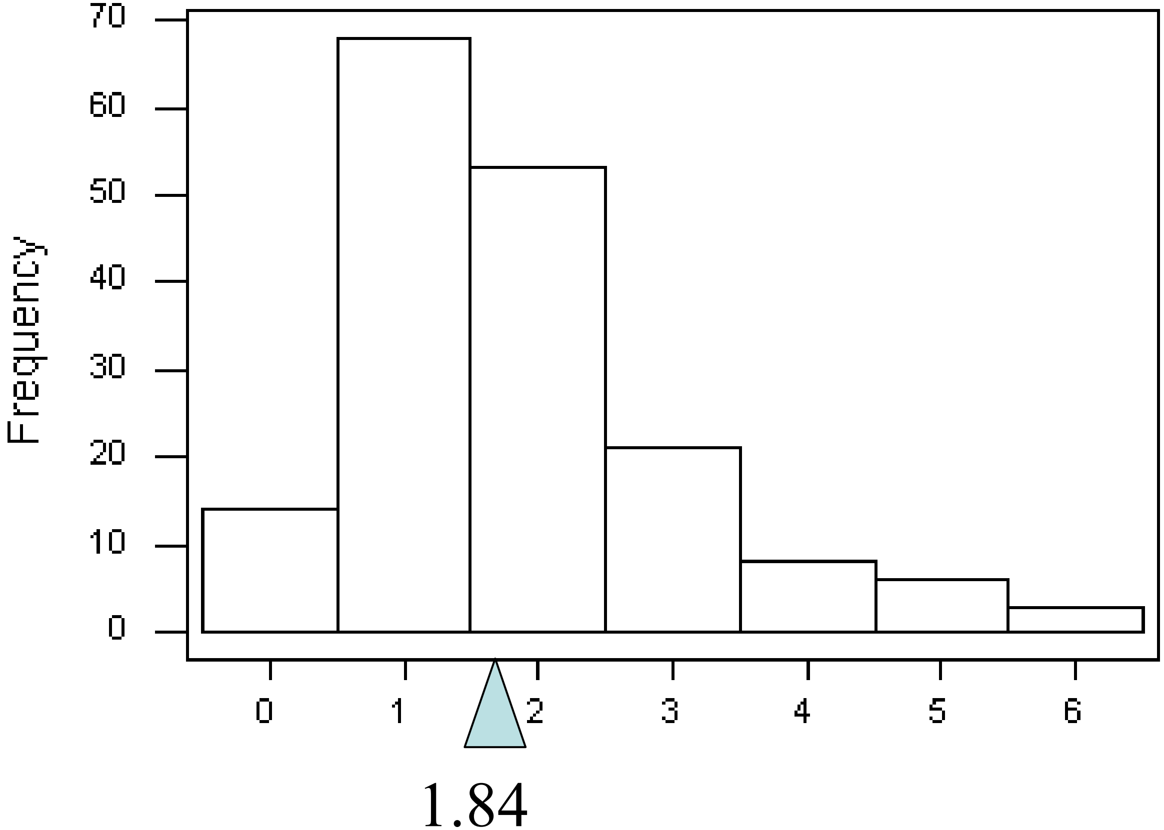
x_i	p_i	$x_i p_i$
0	14/173 = .08	.00
1	68/173 = .39	.39
2	.31	.62
3	.12	.36
4	.05	.20
5	.03	.15
6	.02	.12

$$\mu = E(X) = \sum x_i p_i$$

$$= 1.84$$

= mean number of siblings

Expected value = mean value is where the picture of the pdf "balances"



Other examples of expected value

Ex 1: Mean number of girls, on board

Ex 2: Daily 3 lottery, X = number of wins for 2 plays

$E(X) = 0(.998001) + 1(.001998) + 2(.000001) = .002$

- So the "expected" number of wins in 2 plays is .002 wins, or 2/1000 wins.
- You would have to play 1000 times to have an "expected number of wins" of 1.
- Note that this does *not* mean you would definitely win once if you played 1000 times!

Do you benefit from extended warranties? What is your "expected" cost of repair?

You buy a new appliance, computer, etc.

- Extended warranty for a year costs \$10.
- Unknown to you, the probability you will need a repair is 1/50, and it will cost \$200 if you do.

Is the warranty a good deal??

X = your cost to repair the item (pdf on board)

E(X) = \$4.00 (calculation on board)

Your *expected* cost for repair is much less than \$10.

Notes about expected value

- It's the *average* or *mean* value of the random variable *over the long run*.
- It may not be an actual possible value for the random variable (usually it isn't; e.g. 1.84 sibs).
- In gambling, lotteries, insurance, extended warranty, etc., you can be pretty sure that your "expected" cost per event if you play or buy is more than if you don't – the house wins!
- However, for insurance, for example, you might prefer the peace of mind of knowing your fixed cost. For lottery, you might want the thrill of the possibility of winning, even though you lose on average.

Standard Deviation for a Discrete Random Variable

The **standard deviation** of a random variable is essentially the average distance the random variable falls from its mean over the long run.

If X is a random variable with possible values $x_1, x_2, x_3, \ldots,$ occurring with probabilities $p_1, p_2, p_3, \ldots,$ and **expected value** $E(X) = \mu$, then

$$\text{Variance of } X = V(X) = \sigma^2 = \sum (x_i - \mu)^2 p_i$$

$$\text{Standard Deviation of } X = \sigma = \sqrt{\sum (x_i - \mu)^2 p_i}$$

Example 8.13 *Stability or Excitement*

Two plans for investing $100 – which would you choose? (Pictures on board)

Plan 1		Plan 2	
X = Net Gain	*Probability*	*Y = Net Gain*	*Probability*
$5,000	.001	$20	.3
$1,000	.005	$10	.2
$0	.994	$4	.5

Expected Value for each plan:

Plan 1:

$E(X) = \$5{,}000 \times (.001) + \$1{,}000 \times (.005) + \$0 \times (.994) = \$10.00$

Plan 2:

$E(Y) = \$20 \times (.3) + \$10 \times (.2) + \$4 \times (.5) = \$10.00$

Example 8.13 *Stability or Excitement (cont)*

Variability for each plan:

Plan 1			Plan 2		
$(X-\mu)^2$	p	$(X-\mu)^2 p$	$(Y-\mu)^2$	p	$(Y-\mu)^2 p$
$(\$5,000 - \$10)^2 = \$24,900,100$	.001	\$24,900.1	$(\$20 - \$10)^2 = \$100$	.3	\$30
$(\$1,000 - \$10)^2 = \$980,100$	.005	\$4,900.5	$(\$10 - \$10)^2 = 0$	.2	0
$(\$0 - \$10)^2 = 100$	.994	\$99.4	$(\$4 - \$10)^2 = \$36$	.5	\$18

Plan 1: $V(X) = \$29,900.00$ and $\sigma = \$172.92$

Plan 2: $V(X) = \$48.00$ and $\sigma = \$6.93$

The possible outcomes for Plan 1 are much more variable. If you wanted to ***invest cautiously***, you would choose **Plan 2**, but if you wanted to have the ***chance to gain a large amount of money***, you would choose **Plan 1**.

Notes about standard deviation

- Similar to when we used standard deviation for data in Chapter 2, it is most useful for *normal* random variables, which we will cover on Friday.
- In general, useful for comparing two random variables to see which is more spread out. *Examples*:
  - Two cities both have average yearly temperature of 65 degrees, but one has s.d of 5 degrees and the other has s.d. of 20 degrees. Which would you prefer?
  - One investment fund has average rate of return over many years of 8%, and s.d. of 2%. The other has average of 10%, but s.d. of 20%. The second one is higher on average, but is much more volatile.