Dec 10
Rachel

Lecture 15:
Lesson and Activity Packet

MATH 232: Introduction to Statistics

November 21, 2016

Homework and Announcements

- Homework 15 due in class today
- Post on book discussion forum on Canvas before 11:59 p.m. Wednesday
- Submit book summary on Canvas before 11:59 p.m. Wednesday
- No class on Wednesday or Friday this week
- 10 a.m. section: your final exam is W, Dec 14 10:30 am
- 12 p.m. section: your final exam is F, Dec 16 1:00 pm

Last time:

- Descriptive vs. Inferential Statistics
- Random Variables (Discrete and Continuous)
- Probability Distributions
- Mean, Variance, and Standard Deviation for a Probability Distribution

Questions?
Today

- Identifying unusual results
- Expected value
- Binomial distribution

Recall that in Chapter 2, we called a data point "unusual" if it did not lie within 2 standard deviations of the mean. This is true for random variables and probability distributions, too. A general guideline is therefore:

Theorem 1 (*Range Rule of Thumb*)

A general guideline is that "most" values a random variable can take on lie within 2 standard deviations of the mean. That is, the minimum usual value of a random variable that is distributed with mean μ and standard deviation σ is $\mu - 2\sigma$ and the maximum usual value is $\mu + 2\sigma$.

Another way of identifying unusual events is the **rare event rule**:

Theorem 2 (*Rare Event Rule for Inferential Statistics*)

Generally speaking:

- ***Unusually* high *number of successes:*** *x successes among n trials is an* unusually high *number of successes if $P(x$ or more$) \leq 0.05$.*
- ***Unusually* low *number of successes:*** *x successes among n trials is an* unusually low *number of successes if $P(x$ or fewer$) \leq 0.05$.*

Note: The value 0.05 is not absolutely rigid, and depending on the problem, other values for the threshold might also be informative, too—but 0.05 is the most important and widely used.

Example 1

Suppose you were tossing a coin with the goal of determining whether it is fair, and suppose that 1000 tosses resulted in 501 heads. This is **not** *evidence that the coin favors heads, because it is very easy to get a result like 501 heads in 1000 tosses just by chance. Yet, the probability of getting* exactly *501 heads in 1000 tosses is quite small: 0.0252 (we will learn later how to compute this probability). However, the probability of* 501 or more *heads is high: 0.487.*

Group Exercise 1

Use probabilities to determine whether 1 is an unusually low number of peas with green pods when 5 offspring are generated from parents both having the green/yellow combination of genes. Recall that the relevant probability distribution is

x	0	1	2	3	4	5
$P(x)$	0.001	0.015	0.088	0.264	0.396	0.237

The mean of a discrete random variable is the theoretical mean outcome for infinitely many trials. We can think of that mean as the **expected value** in the sense that it is the average value that we would expect to get if the trials could continue indefinitely.

Definition 1 (*Expected value*)

The expected value of a discrete random variable is denoted E, and it is the mean value of the outcomes:

$$E = \sum [x \cdot P(x)].$$

Note: Even if x can take on only integer values, E doesn't need to be.

Example 2

When generating groups of five offspring peas, the mean number of peas with green pods is 3.8—therefore, the expected number of peas with green pods is also 3.8.

Example 3

In the Illinois Pick 3 lottery game, you pay 50 cents to select a sequence of 3 digits, such as 314. If you select the same sequence of three digits that are drawn by the organizer, then you win and collect $250.

- *How many different selections are possible?* $\underline{10} \cdot \underline{10} \cdot \underline{10} = 10^3 = 1{,}000$
- *What is the probability of winning?* $P(\text{win}) = \frac{1}{1{,}000}$
- *If you win, what is your net profit?* Net profit is $\underbrace{\$250.00}_{\text{jackpot}} - \underbrace{\$0.50}_{\text{ticket cost}} = \$249.50$
- *Find the expected value.*

x	$-0.50	$249.50
$P(x)$	999/1,000	1/1,000

$$E(x) = \frac{-0.5 \cdot 999}{1000} + \frac{249.5 \cdot 1}{1000} = \frac{-250}{1000} = \$-0.25$$

Group Exercise 2

In the New Jersey Pick 4 lottery game, you pay 50 cents to select a sequence of 4 digits, such as 3141. If you select the same sequence of four digits that are drawn by the organizer, then you win and collect $2788.

- *How many different selections are possible?*
- *What is the probability of winning?*
- *If you win, what is your net profit?*
- *Find the expected value.*
- ***Which is better: A 50 cent ticket in the Illinois Pick 3, or a 50 cent ticket in the New Jersey Pick 4?***

- 10,000 selections
- $P(\text{win}) = \frac{1}{10{,}000}$
- Net profit if win = \$2,788 − \$0.50 = \$2,787.50
- Net profit if lose = − \$0.50

x	-0.50	$2{,}787.50$
$P(x)$	$\frac{9{,}999}{10{,}000}$	$\frac{1}{10{,}000}$

$$E(x) = \$\frac{-0.5\,(9{,}999) + 1 \cdot 2{,}787.5}{10{,}000}$$

$$= \$\frac{-2{,}212}{10{,}000}$$

$$= -\$0.2212 \approx -22¢$$

The NJ ticket is a better value — but you still lose.

Binomial distributions are very commonly seen in inferential statistics—they allow us to deal with circumstances in which outcomes belong to only two relevant categories, such as *acceptable/defective*, *survived/died* or *pass/fail.*

Definition 2

A binomial probability distribution *results from a procedure that meets all of the following requirements:*

1. *The procedure has a fixed number of trials;*
2. *The trials are independent events (the outcome of any individual trial does not affect the probabilities of the outcomes for the other trials);*
3. *Each trial has all outcomes classified into two categories (commonly called* success *and* failure*);*
4. *The probability of a success remains the same in all trials.*

The canonical example of a procedure that follows a binomial distribution is coin flipping.

Another example is choosing 5 pea plants from a large pool of offspring. [Since we assume the pool of offspring is very large, it is safe to treat the dependent events as independent. Rule of thumb: if a sample size is less than 5% of the population, treat repeated sampling without replacement as independent.]

Definition 3 (***Notation for the Binomial Distribution***)

- *S and F denote the two possible categories of all outcomes*
- *$P(S) =: p$, so p is the probability of a success*
- *$P(F) = 1 - p =: q$, so q is the probability of a failure*
- *n denotes the fixed number of trials*
- *x denotes a specific number of successes in n trials, so x can be any whole number between 0 and n, inclusive*
- *$P(x)$ denotes the probability of getting exactly x many successes in the n many trials*

Make sure x and p refer to the same event as a "success".

Example 4

In the pea pod experiment, the probability of offspring of heterozygous parents having a green shell is 0.75. That is, $P(\text{green pod}) = 0.75$. Suppose we want to find the probability that exactly 3 of 5 offspring have a green pod. Does this procedure result in a binomial distribution? If yes, then identify the values of n, x, p, and q.

To compute the probabilities in a binomial distribution, consider the pea experiment.

for indep. events

Using the multiplication rule, we can compute the probability that the five peas we select can have the pattern GGGYY:

$$P(GGGYY) = P(G)P(G)P(G)P(Y)P(Y) = (0.75)^3(0.25)^2$$

But we're also interested in, for example, YGGGY, YGYGG, and so on. That is, we're interested in other permutations of the letters GGGYY. **This is very similar to the problems about permutations of letters in a word.** (Remember the *PEPPER* and *MISSISSIPPI* examples?) So we need to multiply the probability computed above by the number of possible permutations: 5 letters total

# of permutations: $\dfrac{5!}{\underbrace{3!}_{3\,G\text{'s}}\,\underbrace{2!}_{2\,Y\text{'s}}}$. thus, $P(3G, 2Y) = \dfrac{5!\,(0.75)^3(0.25)^2}{3!\,2!}$

This logic can be generalized:

Theorem 3 (*Binomial Formula*)

The probability of exactly x many successes in n many trials, if the probability of a success is ~~q~~ p and the probability of failure is ~~p~~ q, is:

$$P(x) = \frac{n!p^x q^{n-x}}{x!(n-x)!}.$$

Note that ! denotes the factorial that we saw before. By convention, $0! := 1$.

TYPOGRAPHICAL ERROR

Example 5

We can fill in the rest of the pea pod probabilities ourselves now...

$$P(0) = \frac{5!\,(0.75)^0(0.25)^5}{0!\,5!} = (0.25)^5 = \frac{1}{4^5} \approx 0.001$$

$$P(1) = \frac{5!\,(0.75)^1(0.25)^4}{1!\,4!} = 5(0.75)(0.25)^4 = \frac{5\cdot 3}{4^5} \approx 0.015$$

$$P(2) = \frac{5!\,(0.75)^2(0.25)^3}{2!\,3!} = \frac{5\cdot 4\cdot 3^2}{2\cdot 4^5} = \frac{5\cdot 9}{2\cdot 4^4} \approx 0.088$$

$$P(3) = \frac{5!\,(0.75)^3(0.25)^2}{3!\,2!} = \frac{5\cdot 4\cdot 3^3}{2\cdot 4^5} = \frac{5\cdot 27}{2\cdot 4^4} \approx 0.264$$

$$P(4) = \frac{5!\,(0.75)^4(0.25)^1}{4!\,1!} = \frac{5\cdot 3^4}{4^5} \approx 0.396$$

$$P(5) = \frac{5!\,(0.75)^5(0.25)^0}{5!\,0!} = \frac{3^5}{4^5} \approx 0.237$$

Group Exercise 3

Assume that you engage in blind guessing on a multiple-choice exam whose questions each have five possible answer choices (a,b,c,d,e). For a random selection of 3 questions on the test, what is the probability of getting 0, 1, 2, or all 3 guesses correct?

S = success := correct answer

$P(S) = p = \frac{1}{5}$ (the chance of getting a correct answer)

$n = 3$, $x \in \{0, 1, 2, 3\}$.

$$P(0) = \frac{3!\,(\frac{1}{5})^0(\frac{4}{5})^3}{0!\,3!} = \frac{4^3}{5^3} = \frac{64}{125} = 0.512$$

$$P(1) = \frac{3!\,(\frac{1}{5})^1(\frac{4}{5})^2}{1!\,2!} = \frac{3\cdot 4^2}{5^3} = \frac{48}{125} = 0.384$$

$$P(2) = \frac{3!\,(\frac{1}{5})^2(\frac{4}{5})^1}{2!\,1!} = \frac{3\cdot 4}{5^3} = \frac{12}{125} = 0.096$$

$$P(3) = \frac{3!\,(\frac{1}{5})^3(\frac{4}{5})^0}{3!\,0!} = \frac{1}{5^3} = \frac{1}{125} = 0.008$$

Reality Check: do these sum to 1? $0.512 + 0.384 + 0.096 + 0.008 \stackrel{?}{=} 1$?

$$\underbrace{\underbrace{0.512 + 0.008}_{} + \underbrace{0.384 + 0.096}_{}}_{}$$
$$= \underbrace{0.520 + 0.480}_{}$$
$$= 1.$$

So, yes, the probabilities sum to 1.