5.1- In each situation below, is it reasonable to use a binomial distribution for the random variable \( X \)? Give reasons for your answer in each case. If a binomial distribution applies, give the values of \( n \) and \( p \).

(a) Most calls made at random by sample surveys don’t succeed in talking with a live person. Of calls to New York City, only 1/12 succeed. A survey calls 500 randomly selected numbers in New York City. Define the random variable \( X \), to be the number that reach a live person out of 500 attempted calls.

Answer – We need to go through the list of requirements and see if our situation meets the conditions. Personally I like to start with what I am measuring. Since we want to see if the binomial situation is applicable our random variable must be counting the number of successes.

- Our random variable counts the number of successes; in this case a success is defined as a person is reached on the phone.
- We are limiting the number of attempts to 500, thus the number of trials is fixed.
- The probability of success is 1/12. We don’t see a reason why this number would change throughout our attempts at calling people. So we have a fixed probability of success.
- Do we have independence throughout our attempt to reach people? What this question is really asking is, after contacting a person, does the probability of success changing? No, there is no reason to believe that this would be the case.

Thus we have a situation in which a binomial distribution is reasonable for our random variable.

(b) At peak periods, 25% of attempted logins to an Internet service provider fail. Login attempts are independent and each has the same probability of failing. Darci logs in repeatedly until she succeeds. Define the random variable \( X \) as the number of login attempts before finally succeeding.

Answer-

- Our random variable \( X \) counts the number of attempts before succeeding. A binomial random variable counts the number of successes out of \( n \) attempts. The two descriptions obviously do not match. In a binomial situation the number of trials is fixed, while in this scenario we do not know how many attempts we will try.

Thus this situation is not a binomial one.
On a bright October day, Canada geese arrive to foul the pond at an apartment complex at the average rate of 12 geese per hour; define the random variable $X$ as the number of geese that arrive in the next three hours.

Answer – The random variable $X$ counts the number of geese that arrive in a fixed three hour period. While the time unit is fixed, this is not what is required for a binomial situation. For a binomial situation what needs to be fixed is the number of trials. The binomial random variable then counts how many successes we have out of a fixed number of attempts. Thus this situation is not a binomial one.

Another way to see this is that our sample space for a binomial random variable, $X$, is always

$$\{X \mid 0, 1, 2, \ldots, n\}$$

while in the scenario for part c the random variable is

$$\{X \mid 0, 1, 2, 3, \ldots\}$$

3. Typographic errors in a text are either non-word errors (as when “the” is typed as “the”) or word errors that result in a real but incorrect word. Spell-checking software will catch non-word errors but not word errors. Human proofreaders catch 70% of word errors. You ask a fellow student to proofread an essay in which you have deliberately made 20 word errors.

(a) If the student matches the usual 70% rate, what is the distribution of the number of errors caught? What is the distribution of the number of errors missed?

We need to ask ourselves if this situation matches a binomial distribution. There are 20 words that are classified as word errors. During the reading when the proofreader gets to this word they will either catch the mistake or not. There is then a fixed number of trials 20; there are twenty occasions in which the proofreader can make the correct decision. We will assume that the 70% success rate stays constant through out. And we will also assume independence; if a reader misses a misspelled word the chances of correctly identifying a mistake next time is still 70%. The binomial model seems to be a good one for this case.

Thus the random variable $X$, counts the number of times a person correctly identifies a mistake.

The values this random variable can take is

$$\{X \mid 0, 1, 2, 3,\ldots, 20\}$$

Number of errors caught: has a binomial distribution with $p = 0.7$, $n = 20$.

Number of errors missed; has a binomial distribution with $p = 0.4$, $n = 20$.

(b) Missing 9 or more out of 20 errors seems a poor performance. What is the probability that a proofreader who catches 70% of word errors misses 9 or more out of 20?

I will use a computer program to calculate the possibilities.

Thus the random variable $X$, counts the number of times a person correctly identifies a mistake.

I want $P(X \leq 11)$ that is I want to calculate the probabilities for the counts in bold:

$$\{X \mid 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20\}$$

Corresponds to 10 misses

$$= \text{binomdist}(11,20,0.7,\text{true})$$

$$= 0.1133$$

Corresponds to 9 misses

So $P(X \leq 11) = 0.1133$. 

$$= 0.1133$$
I could have answered the question directly by naming the random variable Y, the number of misses out of 20 attempts. In this case the distribution of Y is still binomial, and n = 20 as before but the probability of success, an error is missed, is 0.3.

So I would want to calculate P(Y ≥ 9).

{Y | 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20}

The numbers in bold indicated what I want to find the probabilities of.

Know I am ready to ask how am I going to do this? If I use Excel, the command will only allow me to calculate the probability of the value I enter and everything below it. Thus, I will calculate the complement and subtract from one to get what I want.

P(Y ≥ 9) = 1 – P(Y ≤ 8)
= 1 – binomdist(8, 20, 0.3, true)
= 0.1133

5. Return to the proofreading setting of Exercise 5.3.

(a) What is the mean number of errors caught? What is the mean number of errors missed? You see that these two means must add to 20, the total number of errors.

The mean number of errors caught is 0.7(20) = 14. We would expect to catch 14 errors out of 20 word errors if our probability of success is 70%.

The mean number of errors missed is the exact opposite 6, or 0.3(20) = 6

(b) What is the standard deviation σ of the number of errors caught?

I will use the formula \( \sigma = \sqrt{np(1-p)} \)

\[ \sigma = \sqrt{20(0.7)(0.3)} = 2.05 \text{ errors} \]

(c) Suppose the proof reader catches 90% of word errors, so that \( p = 0.9 \). What is σ in this case? What is σ if \( p = 0.99 \)? What happens to the standard deviation of a binomial distribution as the probability of a success gets close to 1?

Case 1: \( p = 0.9 \)

\[ \sigma = \sqrt{20(0.9)(0.1)} = 1.3416 \]

Case 2: \( p = 0.99 \)

\[ \sigma = \sqrt{20(0.99)(0.01)} = 0.4450 \]
In general our function is $\sigma = \sqrt{20p(1-p)}$. When we graph this function we can see as $p$ approaches 1 or 0, the standard deviation, $\sigma$, goes to zero.

![Graph of $\sigma = \sqrt{20p(1-p)}$](image)

**5.6** Suppose that 50% of male Internet users aged 18 to 34 have visited an auction site at least once in the past month.

If you read the questions below, and given what we have discussed as far as the story goes, it is clear that the population we are speaking of consist of the male users 18 to 34, of which half visited the auction site at least once a month. The 50% is assumed to be a parameter, $p = 0.5$

(a) If you interview 12 at random *(notice now we are taking a random sample from the mentioned population)*, what is the mean of the count $X$ who have visited an auction site? What is the mean of the proportion $\hat{p}$ in your sample who have visited an auction site?

mean of the count $X \ \mu_X = 0.5(12) = 6$, \ mean of the proportion $\hat{p} \ \ldots \mu_p = 0.5$

(b) Repeat the calculations in (a) for samples of size 120 and 1200. What happens to the mean count of successes as the sample size increases? What happens to the mean proportion of successes?

For $n = 120$

mean of the count $X \ \mu_X = 0.5(120) = 60$, \ mean of the proportion $\hat{p} \ \ldots \mu_p = 0.5$

For $n = 1200$

mean of the count $X \ \mu_X = 0.5(1200) = 600$, \ mean of the proportion $\hat{p} \ \ldots \mu_p = 0.5$

The mean for the proportions stays the same regardless of sample size.