Math 122
Introduction to Statistics
Binomial Distributions

Binomial Random Variable: A random variable x has a binomial distribution if:

1. A procedure is repeated for a fixed number n trials.
2. The trials are independent.
3. Each trial ends in one of two ways – usually called success or failure.
4. The probability of success is the same in all trials.
5. The value of x is the number of successes in n trials.

Examples: These are examples of binomial random variables:

1. The number of female children in random sets of 10 children.
2. The number of correct questions when you guess on a TF test.
3. The number of green peas in sets of 5 offspring peas.
4. The number of Republicans among 1000 voters selected WITH REPLACEMENT.

Five-Percent Rule: If a sample of size n is selected from a population of size N, and if the sample size is less than 5% of the population, then we can calculate as if the sample is selected with replacement (even if it is selected without replacement).

Example: If 10 students are randomly selected without replacement from a college campus of 1500, and their genders are recorded, then their genders are technically not independent. By the 5% rule, we can treat the genders as if they are independent. Then, if x is the number of females among these 10 students, we can treat x as having a binomial distribution.

NOTATION AND FORMULAS

- n=number of independent trials.
- p=probability of success in each trial.
- q=probability of failure in each trial=$1 - p$.
- $P(x)$=probability of exactly x successes in n trials.

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

WE WILL USE TECHNOLOGY TO DO ALL OF THESE CALCULATIONS. WE WILL NOT USE THIS FORMULA!

$$P(x = a) = \mathit{binompdf}(n, p, a)$$
$$P(x \leq a) = \mathit{binomcdf}(n, p, a)$$

Example: A multiply choice quiz has 100 questions. Each question has four options, only one of which is correct. A student guesses on all 100 questions. What is the probability the student gets exactly 30 questions correct?

Solution: Let x be the number of correct questions. Then x has a binomial distribution with $n = 100$ and $p = 0.25$.

$$P(x = 30) = \mathit{binompdf}(100, 0.25, 30) = .0458$$

Example: A multiply choice quiz has 100 questions. Each question has four options, only one of which is correct. A student guesses on all 100 questions. What is the probability the student gets 30 or fewer questions correct?

Solution: Let x be the number of correct questions. Then x has a binomial disribution with $n = 100$ and $p = 0.25$.

$$P(x \leq 30) = \mathit{binomcdf}(100, 0.25, 30) = .8962$$

VARIATIONS ON $\leq$

To find	Use
$P(x = a)$	$\mathit{binompdf}(\ldots, a)$
$P(x \leq a)$	$\mathit{binomcdf}(\ldots, a)$
$P(x < a)$	$\mathit{binomcdf}(\ldots, a - 1)$
$P(x > a)$	$1 - \mathit{binomcdf}(\ldots, a)$
$P(x \geq a)$	$1 - \mathit{binomcdf}(\ldots, a - 1)$

This	means this
at most	less than or equal
at least	greater than or equal
no more than	less than or equal
no less than	greater than or equal
up to	less than or equal

SUMMARY STATISTICS

For a binomial distribution with n independent trials, probability of success p, and probability of failure $q = 1 - p$,

- $\mu = np$

- $\sigma = \sqrt{npq}$

Example: Gregor Mendel Estimated that the probability that a pea pod with green/yellow genes turns out to be green is 0.75. To test his claim, he bred 580 pea pods. Of these, 428 were green. Use the Range Rule of Thumb to find the usual numbers of green pea pods among 580 assuming that Mendel was correct.

Solution: We will assume that the probability that a pea pod with green/yellow genes is green is 0.75. Let x be the number of green pods out of 580. Then x has a binomial distribution with $n = 580$, $p = 0.75$, and $q = 0.25$. The summary statistics for x are

- $\mu = np = 580 \cdot 0.75 = 435$
- $\sigma = \sqrt{npq} = \sqrt{580 \cdot 0.75 \cdot 0.25} = 10.43$

The minimum and maximum usual values for x are

- Minimum usual value $= \mu - 2\sigma = 435 - 2 \cdot 10.43 = 414.14$.
- Maximum usual value $= \mu + 2\sigma = 435 + 2 \cdot 10.43 = 455.86$.

The usual numbers of green pea pods if Mendel is correct range from 415 to 455. This is consistent with his experiment.

Example: In Mendel's experiment, if the proportion of green pods really is .75, then is 428 an unusually low number of green pods out of 580?

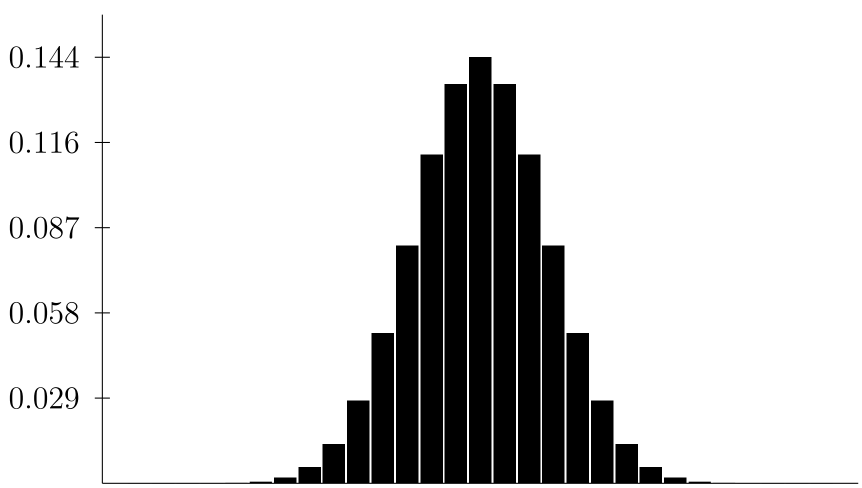
Solution: To see if 428 is an unusually low value of x, we find $P(x \leq 428)$.

$$P(x \leq 428) = \textit{binomcdf}(580, 0.75, 428) = 0.265.$$

This is not unusual.

HISTOGRAM

Binomial random variables with $p = 0.5$ have bell-shaped histograms. Here is the histogram for a binomial random variable in which $n = 30$ and $p = 0.5$.



If $p > 0.5$, then the histogram is slightly skew left. If $p < 0.5$, then the histogram is slightly skew right.