RIT

# STAT 614
# Applied Statistics

1



## Binomial Distribution

- The **binomial experiment** requires the following conditions:
  - The experiment consists of a sequence of n smaller experiments called **trials**, where n is fixed.
  - Each trial can result in one of the same two possible outcomes, generally referred to as **successes** and **failures**.
  - The trials are **independent**, so that the outcome on any particular trial does not influence the outcome of any other trial.
  - The probability of success from trial to trial is constant, and denoted $p$.

2

# Binomial Distribution

- The **binomial random variable** X associated with a binomial experiment consisting of n trials is defined as:

  *X* = the number of successes among the n trials
- *n* is the number of trials and *p* the probability of success
- A binomial random variable *X* has probability mass function:

$$f(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$
$$x = 0,1,2 \dots, n$$

3



# Binomial Distribution

- The mean of a binomial random variable is calculated as:

$$E(X) = np$$

- The variance of a binomial random variable is calculated as:

$$\mathrm{V}(X) = np(1-p)$$

with standard deviation calculated as the square root of the variance.

4

# Geometric Distribution

- The **geometric distribution** is similar to the binomial distribution, though the number of trials is random and there is a fixed number of successes (specifically 1).
- The random variable *X* is the number of failures until the first success.
- *p* is the probability of success.
- A geometric random variable *X* has probability mass function:

$$f(x;p) = p(1-p)^x$$
$$x = 0,1,2\ldots$$

5



# Negative Binomial Distribution

- The **negative binomial distribution** is similar to the binomial distribution, though the number of trials is random and there is a fixed number of successes *r.* The geometric distribution is a special case of the negative binomial distribution.
- The random variable *X* is the number of failures until *r* success are observed.
- *p* is the probability of success.
- A negative binomial random variable *X* has probability mass function:

$$f(x;p,r) = \binom{x+r-1}{x} p^r (1-p)^x \qquad x = 0,1,2\ldots$$

6

# Geometric and Negative Binomial Distributions

- The mean and variance of a geometric random variable is calculated as:

$$E(X) = \frac{1-p}{p} \qquad \mathrm{V}(X) = \frac{1-p}{p^2}$$

- The variance of a negative binomial random variable is calculated as:

$$E(X) = \frac{r(1-p)}{p} \qquad \mathrm{V}(X) = \frac{r(1-p)}{p^2}$$

both with standard deviation calculated as the square root of the variance.

7



# Hypergeometric Distribution

- The **hypergeometric distribution** assumes the following conditions:
  - The population *N* is finite, *M* is the number of successes, and *n* is the number of trials.
  - The trials are dependent – trials are selected without replacement.
  - The probability of success *p* therefore depends on the results of prior trials.

8

## Hypergeometric Distribution

- The random variable X associated with a hypergeometric distribution is defined as:

  *X* = the number of successes among the *n* trials
- *N* is the population size, *M* is the number of successes in the population, and *n* is the number of trials
- The random variable X following a hypergeometric distribution has a probability mass function:

$$f(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \quad \max(0, n-N+M) \leq x \leq \min(n, M)$$





## Hypergeometric Distribution

- The mean of a random variable associated with the hypergeometric distribution is calculated as:

$$E(X) = n\left(\frac{M}{N}\right)$$

- The variance of a binomial random variable is calculated as:

$$V(X) = n\left(\frac{M}{N}\right)\left(1 - \frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)$$

with standard deviation calculated as the square root of the variance.

# Poisson Distribution

- The **Poisson distribution** differs from the prior probability distributions, which are derived through an experiment consisting of trials and applications of probability laws.

- The Poisson distribution is specifically interested in understanding the probability around a number of occurrences (events) in a fixed period of time.





# Poisson Distribution

- The random variable *X* associated with a Poisson distribution is defined as:

  *X* = the number of events in an interval of fixed size
- Assuming we treat the parameter $\mu$ *as* a rate of occurrence, then the random variable X following a poisson distribution has a probability mass function:

$$f(x;\mu) = \frac{e^{-\mu}\mu^{x}}{x!} \qquad x = 0,1,2 \ldots$$

- Both the mean and the variance of a Poisson random variable are $\mu$, with a standard deviation of $\sqrt{\mu}$.

# Your task…

- Given these discrete probability distributions, your task is to identify which distribution applies given the provided context.

- Consider the examples that follow.





# Chapter 3.4 Exercise 52

- Suppose 30% of all students who need to buy a textbook for a course want a new copy, whereas 70% want a used copy. Consider randomly selecting 25 purchasers.
  - What are the mean and standard deviation of the number who want a new copy of the book?
  - What is the probability that the number who want new copies is more than two standard deviations away from the mean?
  - The bookstore has 15 new copies and 15 used copies in stock. If 25 people come in one by one to purchase this textbook, what is the probability that all 25 will get the type of book they want from current stock?

# Chapter 3.4 Exercise 52

- This is a random variable following a binomial distribution with n=25 and p=0.30.
- Mean and Standard Deviation: $\mu = np$ $\quad \sigma = \sqrt{np(1-p)}$

$$\mu = 25p \qquad \sigma = \sqrt{25x0.30(1-0.30)}$$
$$\mu = 7.5 \qquad \sigma = 2.29$$

- If the mean is 7.5 and standard deviation 2.29, for *X* to be more than two standard deviations from the mean, we must consider 1 – P(3≤X≤12). Well…

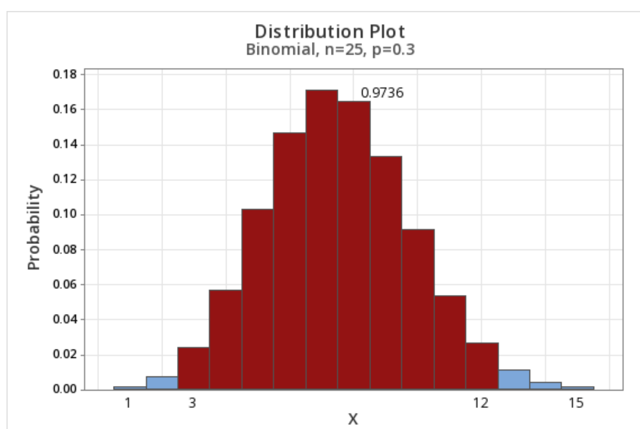
$$1 - \sum_{x=3}^{12} \binom{25}{x} (.3)^x \, (1-.3)^{25-x}$$





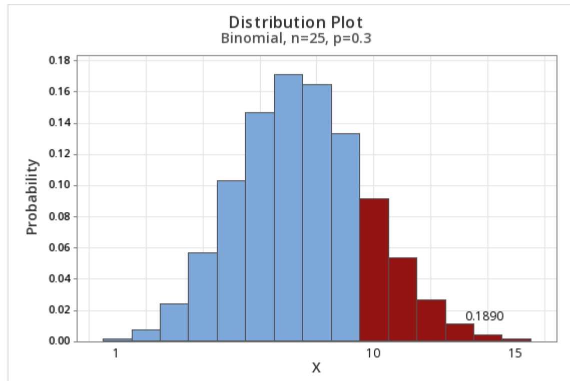
# Chapter 3.4 Exercise 52

- The previous computation is cumbersome and time-consuming.
- *MINITAB:* Graph > Probability Distribution Plot > View Probability
- Specify the distribution as binomial and within the options submenu provide the specifications from the problem.



# Chapter 3.4 Exercise 52

- What about the last question….
- Consider that if $X > 15$, the bookstore cannot meet the demand, and that given our definition of $X$, there must be $25 - X$ wanting used copies. This then also reasons that if $25 - X > 15$, the bookstore cannot meet the demand.
- We can consider $25 - X > 15$ to be equivalent to $X < 10$, and so the bookstore can meet the demand if $10 \leq X \leq 15$.








# Chapter 3.5 Exercise 72

- A personnel director interviewing 11 senior engineers for four job openings has scheduled six interviews for day one and five interviews for day two. Assume the candidates are interviewed in random order.
  - What is the probability that x of the top four candidates are interviewed on day one?
  - How many of the top four candidates can be expected to be interviewed on day one?
  - What is the probability that exactly three of the top four are interviewed on day one?

## Chapter 3.5 Exercise 72

- This is a random variable following a hypergeometric distribution with a population of N=11, M=4, and n=6 for day one interviews.
- We can use the above pmf to solve for the third question. You could also explore the probability distribution plot function in Minitab.

$$f(x; 6,4,11) = \frac{\binom{4}{x}\binom{7}{6-x}}{\binom{11}{6}}$$

$$f(3; 6,4,11) = \frac{\binom{4}{3}\binom{7}{3}}{\binom{11}{6}} = 0.30303$$





## Chapter 3.5 Exercise 72

- Back to the second question…
- Recall what “expected” is equivalent to from the prior slides.

$$E(X) = n\left(\frac{M}{N}\right)$$

$$E(X) = 6\left(\frac{4}{11}\right)$$

$$E(X) = 2.18$$

## Chapter 3.5 Exercise 77

- According to "Characterizing the Severity and Risk of Drought in the Poudre River, Colorado" (Journal of Water Res. Planning and Mgmnt., 2005: 383-393), the drought length Y is the number of consecutive time intervals in which the water supply remains below a critical value $y_0$ (a deficit), preceded by and followed by periods in which the supply exceeds this critical value (i.e. surplus). The cited paper proposes a geometric distribution with *p*=0.409 for this random variable.
  - What is the probability that a drought lasts exactly 3 intervals? At most 3 intervals?
  - What is the probability that the length of a drought exceeds its mean value by at least one standard deviation?



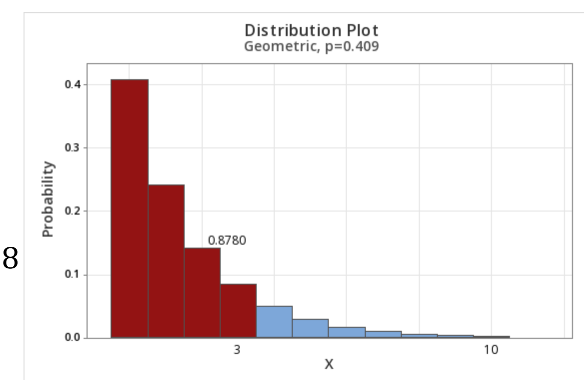


## Chapter 3.5 Exercise 78

- What is the probability that a drought lasts exactly 3 intervals? At most 3 intervals?

  $f(Y \leq 3; 0.409) = 0.409(1 - 0.409)^3 = 0.878$

- Consider the Probability Distribution Plot…



# Chapter 3.5 Exercise 78

- What is the probability that the length of a drought exceeds its mean value by at least one standard deviation?

$$E(Y) = \frac{1-p}{p} = \frac{1-0.409}{0.409} = 1.445$$

$$V(Y) = \frac{1-p}{p^2} = \frac{1-0.409}{0.409^2} = 3.533$$

$$SD(Y) = 1.88$$

$$P(Y \geq 3.325) = 1 - P(Y \leq 3)$$
$$= 1 - 0.878$$
$$= 0.122$$





# Chapter 3.6 Exercise 86 (Ed. 8)

- The number of people arriving for treatment at an ER follows a Poisson distribution with a rate parameter of five per hour.
  - What is the probability that exactly four arrivals occur during a particular hour?
  - What is the probability that at least four people arrive during a particular hour?
  - How many people do you expect to arrive during a 45 minute period?

# Chapter 3.6 Exercise 86 (Ed. 8)

- What is the probability that exactly four arrivals occur during a particular hour?

$$f(4; 5) = \frac{e^{-5}5^4}{4!} = 0.1755$$

- What is the probability that at least four people arrive during a particular hour?

Use a probability distribution plot to arrive at an answer of 0.7350.

- How many people do you expect to arrive during a 45 minute period?

With μ=5, we can revise this rate based on our knowledge of minutes in an hour…