



UNIVERSITY OF JORDAN
FACULTY OF MEDICINE
BATCH 2013-2019



EPIDEMIOLOGY & BIOSTATISTICS

Slides Sheet Handout other.....

Number #1

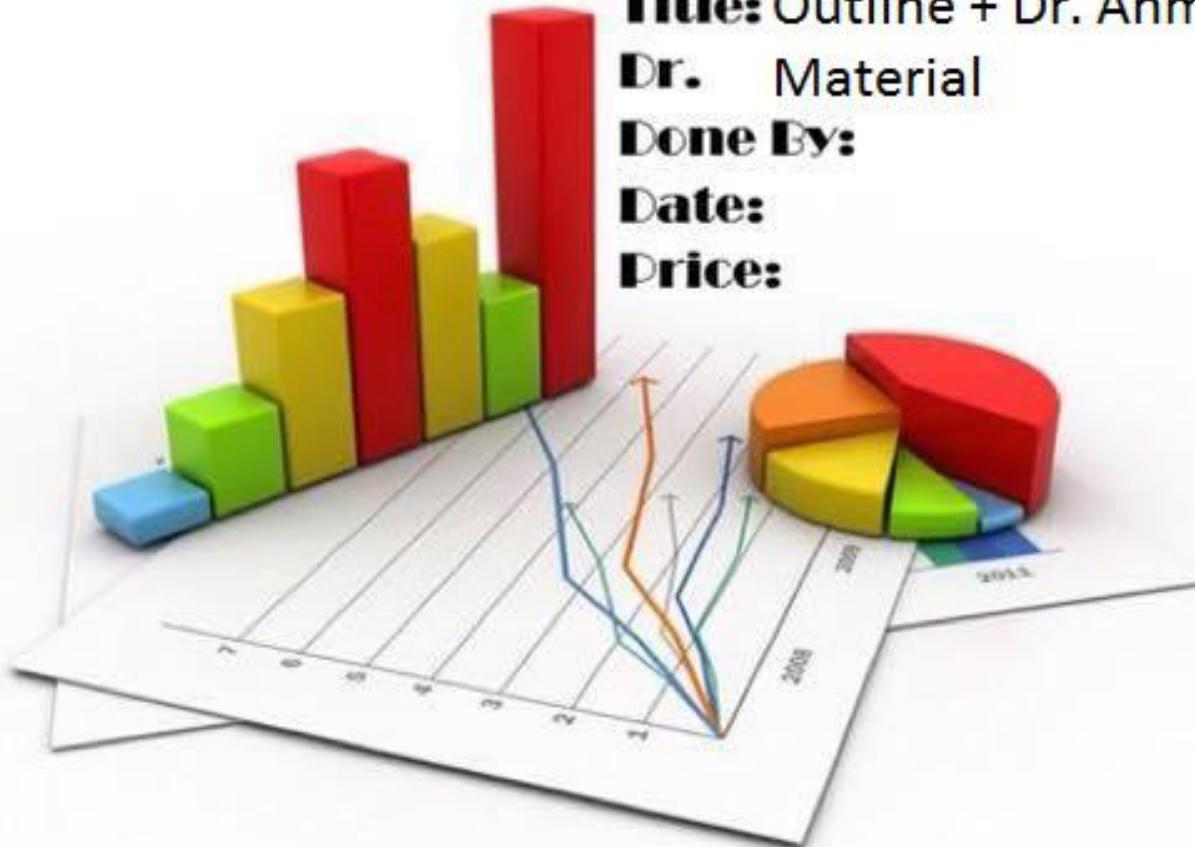
Title: Outline + Dr. Ahmad

Dr. Material

Done By:

Date:

Price:



DESIGNED BY NADEEN AL-FREIHAT

Example: In a typical medical school ,
 the mean weight of 100 fourth-year
 medical students is 140 lb, with a
 standard deviation of 28 lb. Calculate
 the Coefficient of Variation (CV) ?

* Solution:

$$CV = \frac{S}{\bar{X}} \times 100 = \frac{28}{140} \times 100 = 20\%$$

①

LEC NB

Probability Distributions

Probability Distributions of Discrete Variables

Definition: The probability distribution of a discrete

random variable is a table, graph, formula, or other device used to specify all possible values of a discrete random variable along with their respective probabilities.

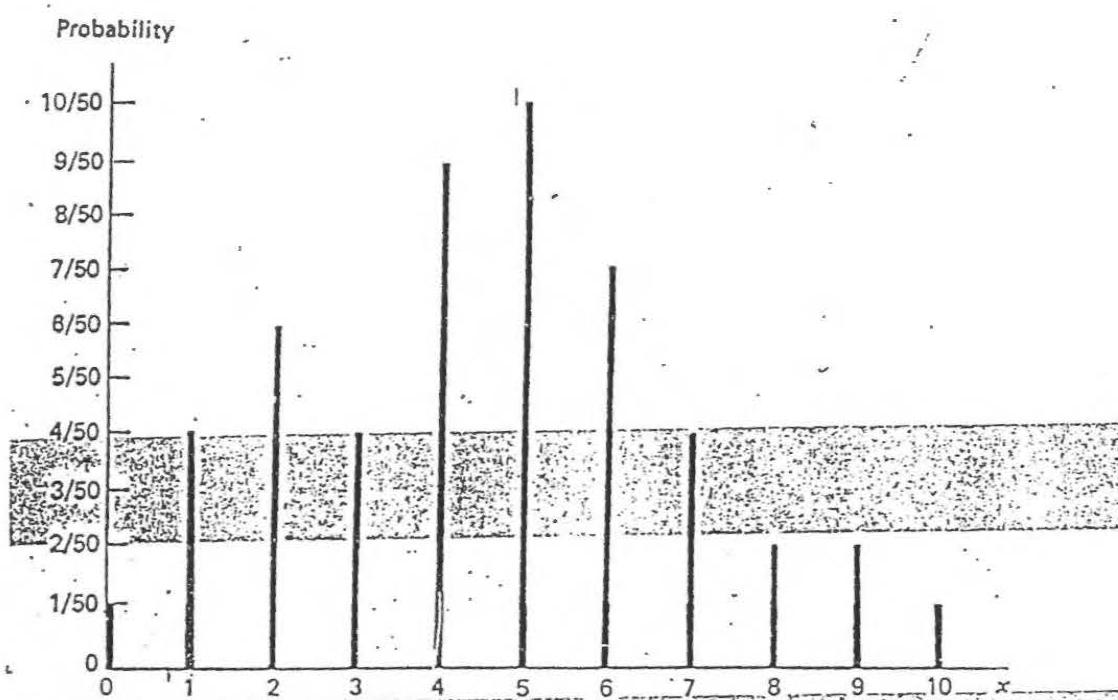
Example 3.2.1: A public health nurse has a case load of 50 families. Let us construct the probability distribution of X , the number of children per family for this population. We can do this with a table in which we list in one column, X , the possible values that X assumes, and in another column, $P(X=X)$, the probability with which X assumes a particular value, X (Table 3.2.1).

Alternatively, we can present this probability distribution in the form of a graph (Figure 3.2.1). The length of each vertical bar indicates the probability for the corresponding value of X .

(c)

TABLE 3.2.1
Probability Distribution of Number of Children per Family in
a Population of 50 Families

x	Frequency of occurrence of x	$P(X = x)$
0	1	1/50
1	4	4/50
2	6	6/50
3	4	4/50
4	9	9/50
5	10	10/50
6	7	7/50
7	4	4/50
8	2	2/50
9	2	2/50
10	1	1/50
	50	50/50



Graphical Representation of the Probability Distribution of Number of Children per Family for Population of 50 Families

(3)

From this example:

- * the values of $P(X=x)$ are all positive.
- * the values are less than 1.
- * Their sum is equal to 1.

The Probability Distribution of a discrete Variable has the following two essential properties:

$$(1) \quad 0 \leq P(X=x) \leq 1$$

$$(2) \quad \sum P(X=x) = 1$$

- * Suppose the nurse, with the case load of 50 families, randomly picks a family to visit?
- * What is the probability that the family will have three children?
From Table 3.2.1
- * What is the probability that a family chosen at random will have either three or four children?
Solution We use the addition rule.

$$P(X=3 \text{ or } X=4) = P(X=3) + P(X=4)$$

$$P(X=3 \text{ or } X=4) = 0.08 + 0.18 = 0.26$$

Cumulative Probability Distribution of a Random Variable (Table 3.2-2)

(4)

TABLE 3.2.2
Cumulative Probability Distribution of Number of Children
per Family in a Population of 50 Families

x	<i>Frequency of occurrence of x</i>	$P(X = x)$	$P(X \leq x)$
0	1	1/50	1/50
1	4	4/50	5/50
2	6	6/50	11/50
3	4	4/50	15/50
4	9	9/50	24/50
5	10	10/50	34/50
6	7	7/50	41/50
7	4	4/50	45/50
8	2	2/50	47/50
9	2	2/50	49/50
10	1	1/50	50/50
	50	50/50	

(5)

1. What is the probability that a family picked at random from the 50 will have fewer than five children? What is needed to answer the question is $P(X < 5)$, and this is obtained by determining the value of the cumulative probability for values of $X = 0$ through $X = 4$. From the table we see that this is $24/50 = .48$.
2. What is the probability that a randomly picked family will have five or more children? We may answer this question by means of the concept of the complement. The set of families with five or more children is the complement set of the set of families with fewer than five children. Their sum is equal to the universal set of 50 families. Since the total probability is 1, and we have found that $P(X < 5) = .48$, the desired probability, $P(X \geq 5)$, is equal to $1 - P(X < 5) = 1 - .48 = .52$.
3. What is the probability that a randomly selected family will have between three and six children, inclusive? What is needed is $P(3 \leq X \leq 6)$, which is equal to $P(X \leq 6) - P(X < 3)$. The probability that X is less than or equal to 6 is the cumulative probability through $X = 6$, or $41/50 = .82$; and the probability that X is less than 3 is the cumulative probability through $X = 2$, or $11/50 = .22$. Thus, we see that $P(3 \leq X \leq 6) = .82 - .22 = .60$.

(6)

The Binomial Distribution

- * One of the most widely used probability distribution in applied statistics.
- * The distribution is derived from a process known as a Bernoulli trial, named in honor of the Swiss mathematician James Bernoulli (1654-1705).
- * When a single trial of some process or experiment can result in only one of two mutually exclusive outcomes, such as dead or alive, sick or well, male or female, the trial is called a Bernoulli trial.
- * In general, if we let n equal the total number of objects, X the number of objects of one type, and $(n-X)$ the number of objects of the other type, the number of sequences is equal to $\binom{n}{X} = n! / X!(n-X)!$ which is equal to the number of combinations of n things taken X at a time. We may write the probability of obtaining exactly X successes in n trials as:

$$f(X) = \binom{n}{X} p^X q^{n-X} \quad \text{for } X=0, 1, 2, \dots, n$$

(7)

✓ 70

PROBABILITY DISTRIBUTIONS

A sequence of Bernoulli trials forms a *Bernoulli process* under the following conditions.

1. Each trial results in one of two possible, mutually exclusive, outcomes. One of the possible outcomes is denoted (arbitrarily) as a success, and the other is denoted a failure.
2. The probability of a success, denoted by p , remains constant from trial to trial. The probability of a failure, $1 - p$, is denoted by q .
3. The trials are independent; that is, the outcome of any particular trial is not affected by the outcome of any other trial.

(8)

TABLE 3.3.1
The Binomial Distribution

<i>Number of successes; x</i>	<i>Probability, f(x)</i>
0	$\binom{n}{0} q^{n-0} p^0$
1	$\binom{n}{1} q^{n-1} p^1$
2	$\binom{n}{2} q^{n-2} p^2$
:	:
x	$\binom{n}{x} q^{n-x} p^x$
:	:
n	$\binom{n}{n} q^{n-n} p^n$
Total	1

sequences is equal to $\binom{n}{x} = n! / x!(n - x)!$ which is equal to the number of combinations of n things taken x at a time.

In our example we may let $x = 3$, the number of successes, so that $n - x = 2$, the number of failures. We then may write the probability of obtaining exactly x successes in n trials as

(9)

$$f(x) = \binom{n}{x} p^x q^{n-x} \quad \text{for } x=0,1,2,\dots,n$$

This expression is called the binomial distribution

$f(x) = P(X=x)$ where X is the random variable, number of successes in n trials.

We use $f(x)$ rather than $P(X=x)$ because of its universal use.

From Table 3.3.1 we see the following facts:

1) $f(x) \geq 0$ for all real values of x .

2) $\sum f(x) = 1$

Example 3.3.2. Suppose that it is known that 30 percent of a certain population are immune to some disease. If a random sample of size 10 is selected from this population, what is the probability that it will contain exactly four immune persons?

(10)

Solution Example 3.3.2.

- We take the probability of an immune person to be 0.3, $n=10$, $X=4$, $P=0.3$, $q=0.7$

$$f(x) = \binom{n}{x} p^x q^{n-x}$$

$$f(4) = \binom{10}{4} (0.3)^4 (0.7)^6 = \frac{10!}{4!6!} (0.0081)(0.117649)$$

$$\underline{f(4) = 0.2001}$$

Example: Suppose that in a certain population 52 percent of all recorded births are males. If we randomly select five birth records from this population, what is the probability that exactly three of the records will be for male births?

Solution given: $P=0.52$, $q=1-p=1-0.52=0.48$

$$f(x) = \binom{n}{x} p^x q^{n-x} \quad n=5, X=3$$

$$f(3) = \binom{5}{3} (0.52)^3 (0.48)^2 = \frac{5!}{3!2!} (0.52)^3 (0.48)^2$$

$$\underline{f(3) = \frac{4 \times 5 \times 3!}{3!2!} \times (0.140608)(0.2304) = 0.32}$$

(11)

* What is the probability of observing exactly two records for male birth?

$$f(2) = \binom{5}{2} (0.52)^2 (0.48)^3 = \frac{5!}{2!3!} (0.52)^2 (0.48)^3$$

$$f(2) = \frac{5 \times 4 \times 3!}{2! \times 3!} \times 0.2704 \times 0.110592 = \underline{0.29}$$

If the sample size (n) is large, then the calculation of probability using the equation

or formula

$$f(x) = \binom{n}{x} p^x q^{n-x}$$

can be very difficult.

Fortunately, probabilities for different values of n , p , and X have been tabulated. Table A gives the probability that X is less than or equal to some specified value. The table gives the cumulative probabilities from $X=0$ up through some specified value.

* Let us illustrate the use of the table by the desire to find the probability that $X=4$, when $n=10$, and $p=0.3$

* What is the probability of observing two records for male birth?

$$f(2) = \binom{5}{2} (0.52)^2 (0.48)^3 = \frac{5!}{2!3!} (0.52)^2 (0.48)^3$$

$$f(2) = \frac{5 \times 4 \times 3!}{2! \times 3!} \times 0.2704 \times 0.110592 =$$

If the sample size (n) is large, the calculation of probability using the formula

$$f(x) = \binom{n}{x} p^x q^{n-x}$$

can be very difficult.

Fortunately, probabilities for different values of n , p , and X have been tabulated. This gives the probability that X is less than or equal to some specified value. The cumulative probabilities from $X=0$ through some specified value.

* Let us illustrate the use of the table. We desire to find the probability that $X=10$, and $P=0.93$.

$$P(X=4) = P(X \leq 4) - P(X \leq 3)$$

$$= 0.8497 - 0.6496 = 0.2001$$

What is the probability that $X=2$, when

$$n=5, p=0.48$$

$$P(X=2) = P(X \leq 2) - P(X \leq 1)$$

$$P(X=2) = 0.5375 - 0.2135 = 0.324$$

* Frequently we are interested in determining probabilities not for specific values of X , but for intervals such as the probability that X is between 5 and 10.

Example 3.3.3

The binomial distribution has two parameters, n and p . They are parameters in the sense that they are sufficient to specify a binomial distribution. The binomial distribution is really a family of distributions with each possible value of n and p designating a different member of the family.

The mean of the binomial $M = np$

The Variance of the binomial $G^2 = np(1-p) = npq$

When n is small relative to $N \rightarrow$ the binomial model is appropriate. (If N is 10 times as large as n , then n is small)

Example 3.3.3 Suppose it is known that in a certain population 10 percent of the population is colorblind. If a random sample of 25 people is drawn from this population, use Table A in the Appendix to find the probability that:

1. Five or fewer will be colorblind.

Solution This probability is an entry in the table. No addition or subtraction is necessary. $P(X \leq 5) = .9666$.

2. Six or more will be colorblind.

Solution This set is the complement of the set specified in 1; therefore,

$$P(X \geq 6) = 1 - P(X \leq 5) = 1 - .9666 = .0334$$

3. Between six and nine inclusive will be colorblind.

Solution We find this by subtracting the probability that X is less than or equal to 5 from the probability that X is less than or equal to 9. That is,

$$P(6 \leq X \leq 9) = P(X \leq 9) - P(X \leq 5) = .9999 - .9666 = .0333$$

4. Two, three, or four will be colorblind.

Solution This is the probability that X is between 2 and 4 inclusive.

$$P(2 \leq X \leq 4) = P(X \leq 4) - P(X \leq 1) = .9020 - .2712 = .6308$$

TABLE A (*Continued*)

15

TABLE A (Continued)

16

TABLE A (*Continued*)

The Poisson Distribution

- discrete distribution
- Named for the French mathematician Denis Poisson (1781-1840)

If X is the number of occurrences of some random event in an interval of time or space (or some volume of matter), the probability that X will occur is given by:

$$f(X) = \frac{e^{-\lambda} \lambda^X}{X!} \quad \text{where, } X = 0, 1, 2, \dots$$

λ -(lambda) is called the parameter of the distribution and is the average number of occurrences of the random event in the interval (or volume).

e - is the constant 2.7183

It can be shown that $f(X) \geq 0$ for every X and that $\sum f(X) = 1$

So that the distribution $\sum_{X=0}^{\infty} f(X) = 1$ satisfies the requirements for a probability distribution.

(18)

The Poisson process. The following statements describe what is known as the poisson process:

- 1) The occurrences of the events are Independent.
- 2) Theoretically, an infinite number of occurrences of the event must be possible in the interval.
- 3) The probability of the single occurrence of the event in a given interval is proportional to the length of the interval.

* The mean and variance of the poisson distribution are equal

$$\mu = G^2$$

The poisson distribution is employed when counts are made of events or entities that are distributed at random in space or time.

For example: under the assumption that the distribution of some parasite among individual host members follows the poisson law, one may with knowledge of the parameter λ , calculate the probability that a randomly selected individual host will yield x number of parasites.