IV. BASIC STATISTICS

A STATE OF STATISTICAL CONTROL IS NOT A NATURAL STATE FOR A MANUFACTURING PROCESS. IT IS INSTEAD AN ACHIEVEMENT, ARRIVED AT BY ELIMINATING ONE BY ONE, BY DETERMINED EFFORT, THE SPECIAL CAUSES OF EXCESSIVE VARIATION.

W. EDWARDS DEMING

Introduction

Basic Statistics is presented in two major category areas:

- General Concepts
- Calculations

General Concepts is reviewed in the following topic areas:

- Terminology
- Frequency Distributions

Terminology

Parameter The true population value, often unknown, estimated by a statistic.

Population All possible observations of similar items from which a sample is drawn.

Sample A group of units or observations taken from a larger collection of units or observations that serves to provide an information basis for making a decision concerning the larger quantity.

Statistic A numerical data measurement taken from a sample that may be used to make an inference about a population.

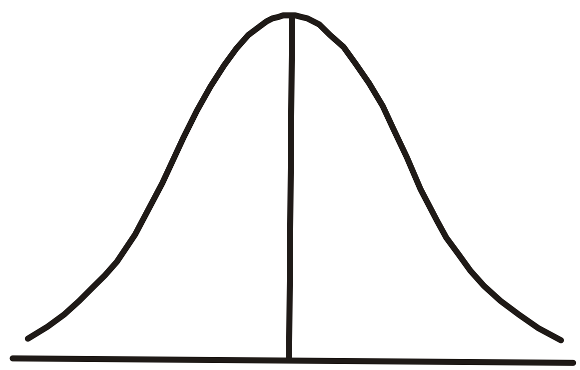
Random Variable A random variable is any observation that can vary. It can represent either discrete or continuous data.

Continuous Frequency Distributions

Normal (Gaussian)

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

μ = Mean
σ = Standard deviation
e = 2.718



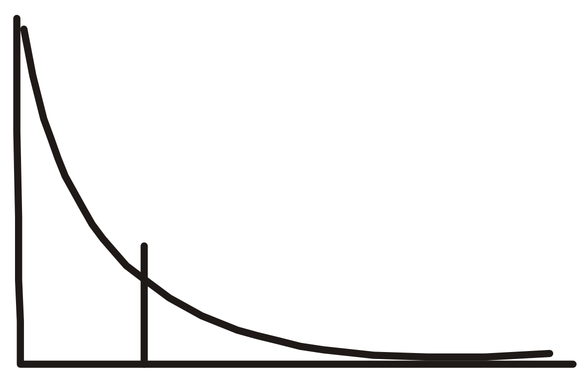
Exponential

$$P(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}}$$

or

$$P(x) = \lambda e^{-\lambda x}$$

μ = θ = Mean
X = X axis reading
λ = failure rate



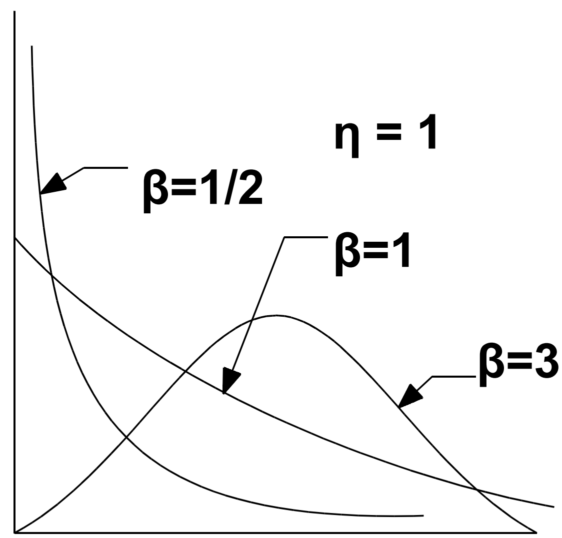
Comparison of Continuous Distributions

Continuous Distributions (Continued)

Weibull

$$P(x) = \frac{\beta}{\eta}(x - \gamma)^{\beta - 1} e^{\frac{-(x - \gamma)^{\beta}}{\eta}}$$

η = Scale parameter
β = Shape parameter
γ = Location parameter

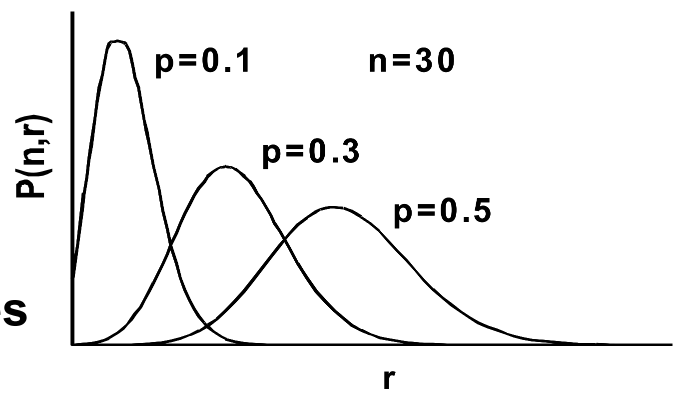


Comparison of Continuous Distributions

Discrete Frequency Distributions

Poisson

$$P(r) = \frac{(np)^r e^{-np}}{r!}$$

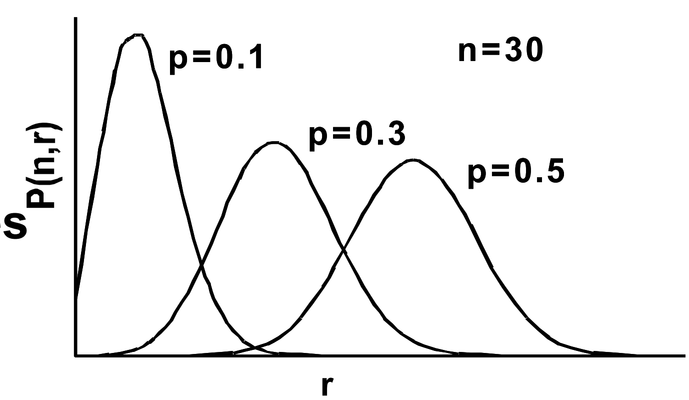


n = Sample size
r = Number of occurrences
p = Probability
np = μ = Average

Binomial

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

n = Sample size
r = Number of occurrences
p = Probability
q = 1 - p

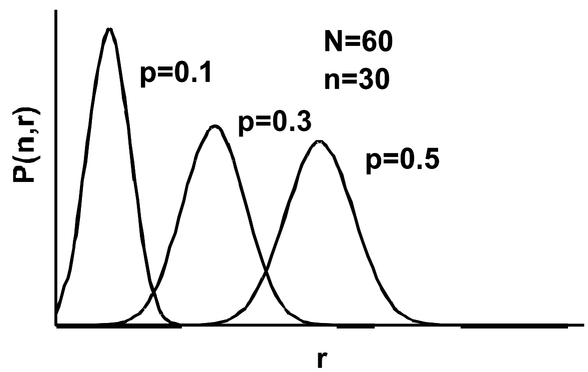


Comparison of Discrete Distributions

Discrete Frequency Distributions (Cont.)

Hypergeometric

$$P(r) = \binom{d}{r}\frac{\binom{N-d}{n-r}}{\binom{N}{n}}$$



n = Sample size
r = Number of occurrences
d = Occurrences in population
N = Population size

Comparison of Discrete Distributions

Hypergeometric Frequency Distribution

The hypergeometric distribution will not be on the CQT exam. The hypergeometric distribution applies when the population is small compared to the sample size. Sampling is done without replacement.

The number of occurrences (r) in the sample follows the hypergeometric function:

$$P(r) = \frac{C_r^d C_{n-r}^{N-d}}{C_n^N}$$

- **N = Population size**
- **n = Sample size**
- **d = Number of occurrences in the population**
- **N - d = Number of non occurrences in the population**
- **r = Number of occurrences* in the sample**

Hypergeometric Distribution (Cont.)

Example: From a group of 20 products, containing 5 defectives, 10 are selected at random. What is the probability that these 10 contain the 5 defectives?

$$N = 20,\ n = 10,\ d = 5,\ (N\text{-}d) = 15 \text{ and } r = 5$$

$$P(r) = \frac{C_5^5\ C_5^{15}}{C_{10}^{20}} \qquad \text{note that } C_r^n = \frac{n!}{r!(n-r)!}$$

$$P(r) = \frac{\left(\frac{5!}{5!0!}\right)\left(\frac{15!}{5!10!}\right)}{\left(\frac{20!}{10!10!}\right)} = \left(\frac{15!}{5!10!}\right)\left(\frac{10!10!}{20!}\right)$$

Answer $P(r) = 0.0163 = 1.63\ \%$

Binomial Frequency Distribution

The binomial distribution applies when the population is large (N > 50) and the sample size is small compared to the population. Generally, n is less than 10 % of N. It is most appropriate to use when the proportion defective is equal to or greater than (0.1).

$$P(r) = C_r^n p^r (1 - p)^{n - r} = \frac{n!}{r!(n - r)!} p^r (1 - p)^{n - r}$$

n = Sample size
r = Number of defectives
p = Proportion defective

Binomial Distribution (Continued)

Example: A random sample of 10 units is taken from a steady stream of product from a press. Past experience has shown 10 % defective parts. Find the probability of exactly one bad part.

$$n = 10 \quad r = 1 \quad p = 0.1$$

$$P(r) = C_r^n p^r (1 - p)^{n - r}$$

$$P(r) = \frac{10!}{1!9!}(0.1)^1(0.9)^9$$

$$P(r) = (10)(0.1)(0.3874)$$

$$\text{Answer } P(r) = 0.3874 = 38.74\ \%$$

Solve for 2 bad parts (answer = 19.37 %). Solve for 0 bad parts (answer = 34.87 %)

Binomial Distribution (Continued)

The binomial distribution average and sigma can be obtained from the following calculations:

The binomial average = $\mu = n\bar{p}$

The binomial sigma = $\sigma = \sqrt{np(1 - p)}$

Example: If one tosses an honest coin 100 times, what is expected to be the average number of heads? What will be the 3 sigma variation?

$n = 100 \; \bar{p} = 0.5$

$n = 100 \quad \bar{p} = 0.5$

Answer: $\mu = n\bar{p} = (100)(0.5) = 50$ **heads**

Sigma = $\sqrt{n\bar{p}(1 - \bar{p})} = \sqrt{(50)(1 - 0.5)}$

$= \sqrt{25} = 5$

Answer: $\mu \pm 3s = 50 \pm 15$ **heads**

Poisson Frequency Distribution

The Poisson is an approximation to the binomial distribution when p is equal to or less than 0.1, and the sample size n is fairly large.

$$P(r) = \frac{\mu^r e^{-\mu}}{r!}$$

$\mu = n\bar{p}$ the population mean
r = number of defectives
e = 2.71828 the base of natural logarithms

Poisson Distribution (Continued)

Example: A continuous process is running a 2 % defective rate. What is the probability that a 100 piece sample will contain exactly 2 defectives?

$$\mu = n\bar{p} = (100)(0.02) = 2 \qquad r = 2$$

$$P(r) = \frac{\mu^r e^{-\mu}}{r!} = \frac{2^2 e^{-2}}{2!} = \frac{2^2}{2! e^2}$$

$$\text{Answer } P(r) = \frac{4}{(2)(7.389)} = 0.27 = 27\ \%$$

Solve for r = 0 Answer 0.135 (13.5 %)
Solve for r = 1 Answer 0.27 (27 %)
Solve for r = 3 Answer 0.18 (18 %)

The Poisson distribution average and sigma can be obtained from the following calculations:

The Poisson average = $\mu = n\bar{p} = \bar{c}$ *

The Poisson sigma = $\sigma = \sqrt{\mu} = \sqrt{n\bar{p}} = \sqrt{\bar{c}}$ *

Calculations

Calculations are presented in the following topic areas:

- **Measures of Central Tendency**
- **Measures of Dispersion**
- **Statistical Inference**
- **Confidence Intervals**
- **Probability**

Measures of Central Tendency

The Mean (X-bar, $\bar{X}$)

The mean is the sum total of all data values divided by the number of data points.

Formula: $\bar{X} = \frac{\sum X}{n}$

$\bar{X}$ is the mean

X represents each number

$\sum$ means summation

n is the sample size

The arithmetic mean is the most widely used measure of central tendency.

Measures of Central Tendency (Cont.)

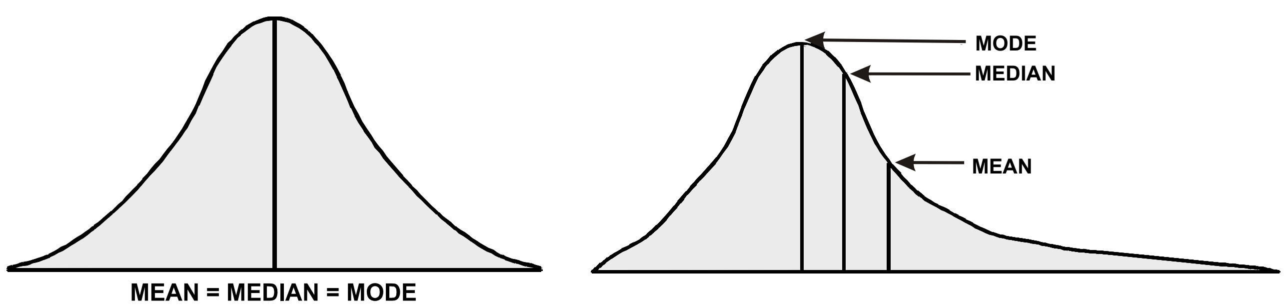
The Mode

The mode is the most frequently occurring number in a data set. It is possible for groups of data to have more than one mode.

The Median (Midpoint)

The median is the middle value when the data is arranged in ascending or descending order. For an even set of data, the median is the average of the middle two values.

For a Normal Distribution Right Skewed Distribution



Comparison of Central Tendency in Normal and Skewed Distributions