

26th Annual ASQ Audit Conference
Blueprint for a Successful Audit

October 9 - 11, 2017 • Pre-conference Courses
October 12 - 13, 2017 • Conference

ASQ
The Global Voice of Quality™ The InterContinental Hotel | Dallas, Texas



Audit
Division
The Global Voice of Quality™

Basic Statistics Made Easy

Victor R. Prybutok, Ph.D., CQE, CQA, CMQ/OE, PSTAT®
Regents Professor of Decision Sciences, UNT
Dean and Vice Provost, Toulouse Graduate School, UNT

13 October 2017

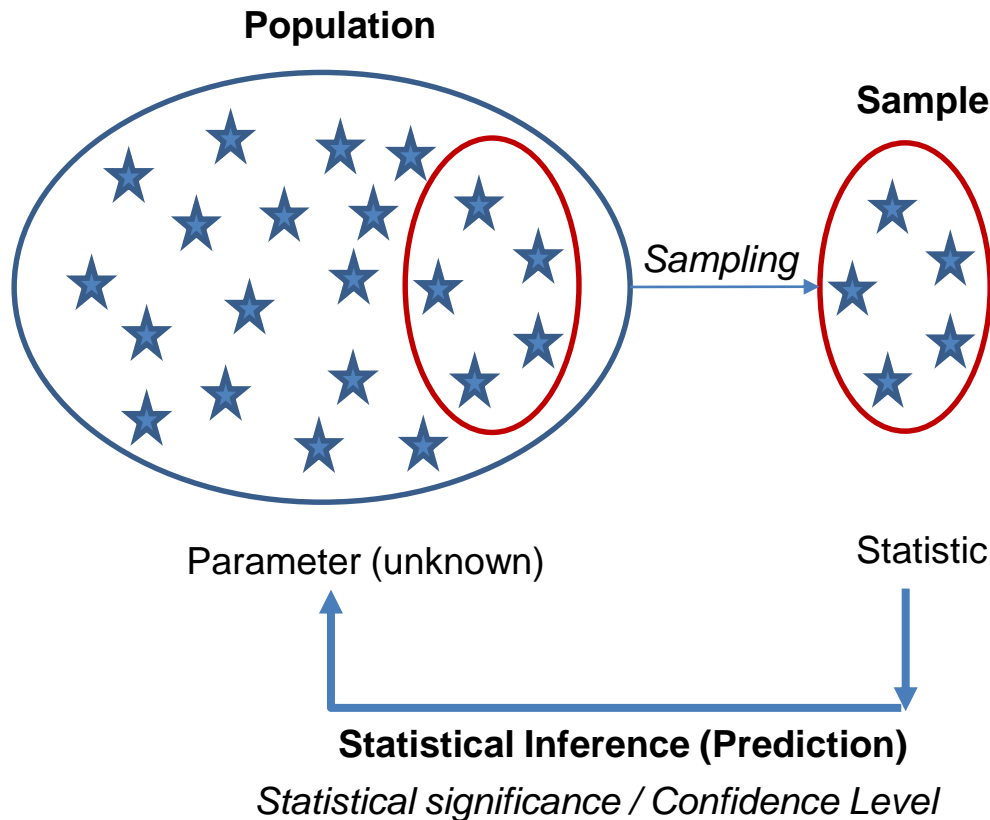
Agenda

- Statistics
- Statistical Measures
- Distributions
- Repeatability and Reproducibility (R&R)
- Process Capability
- Statistical Process Control

Statistics

- **Statistics** deals with the collection, analysis, presentation and use of data to solve problems, design and develop products and processes, and make decisions
- Descriptive Statistics – organize, summarize and present data in an informative way
 - Describe and control **Variability**
- Inferential Statistics
 - **Reasoning and Generalization** of results from a sample to a population
- **Examples**
- What is the percentage of products not meeting specifications?
- What is the tolerance of this product?
- Is this a stable process, and if not, how can we stabilize it?
- What is the probability that a lot of products will be rejected?
- Should the entire lot of 100,000 items be rejected if 10 out of 100 inspected items were found defective?

Population vs. Sample



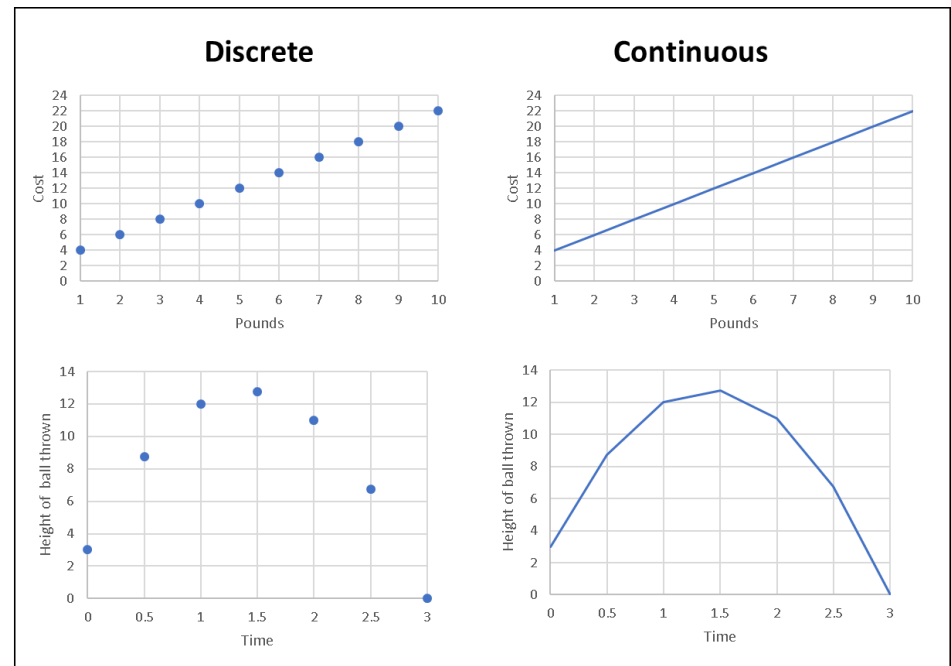
- Population = collection of objects of interest
- Sample = portion of the population of interest
- Sampling = sample selection methods
 - Random sampling most commonly used
- Statistics = key measurement calculations using sample data (observations): e.g. Z statistic; expressed as numbers
- Parameters = population parameters to draw conclusions about; can be expressed as:
 - Hypotheses: e.g. the average length is > 10 ft.
 - Confidence intervals (CI): e.g. the average length is between 9.45 ft. and 10.55 ft.
- Statistical significance (α) = probability of making a (Type I) prediction error (e.g. $\alpha = 5\%$)
 - Acceptable margin of error in prediction
- Confidence Level (CL) = $1 - \alpha$
 - E.g. I am 95% confident that the average length is > 10 ft.
 - E.g. I am 95% confident that the average length is between 9.45 ft. and 10.55 ft.

Sampling

- The sample reflects the characteristics of the population from which it is drawn, otherwise a *sampling error* occurs
 - **Increase sample size** to reduce sampling error
- *Random sampling*: each item in the population has an equal probability of being selected; most commonly used
- *Stratified sampling*: population partitioned into groups, and a sample is selected from each group
- *Systematic sampling*: every *n*th (e.g. 4th, 5th) item is selected
- *Cluster sampling*: population partitioned into groups (clusters), and a sample of clusters is selected
 - Either all elements in the chosen clusters are selected, or a random sample is taken from each cluster selected
- *Judgment sampling*: expert opinion is used to determine the sample

Data Types

- *Variables* are measurements of characteristics of interest
- *Qualitative or Attribute Variables* are not numeric
- *Quantitative Variables* are:
 - *Discrete Variables* = can only take particular, distinct values
 - No other values in-between
 - Counting / enumerating
 - *Continuous Variables* = can assume any value over a continuous range
 - Infinite values in-between
 - Measuring



Probabilities

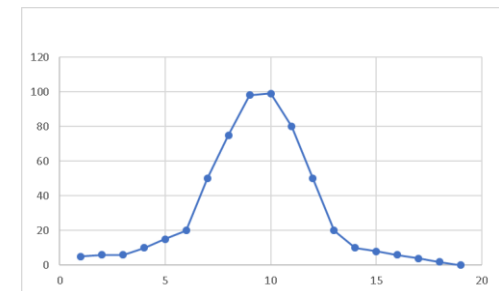
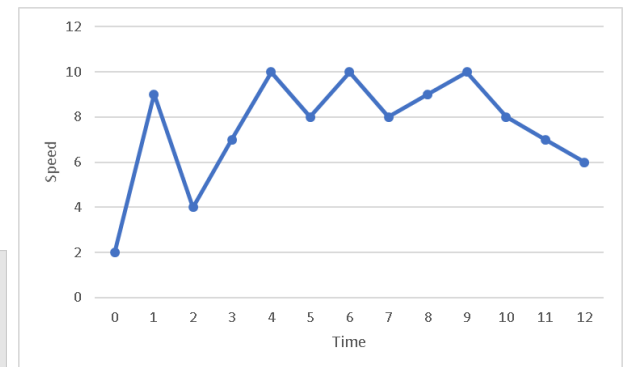
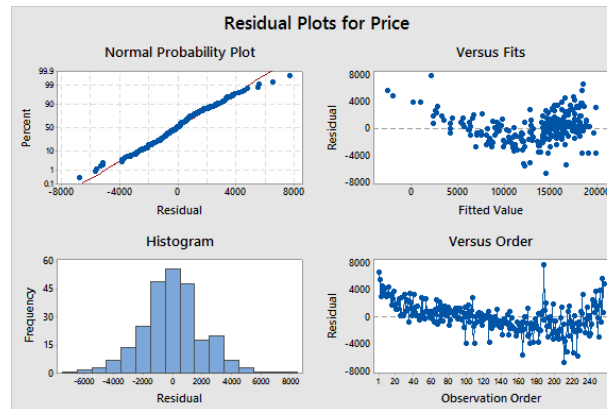
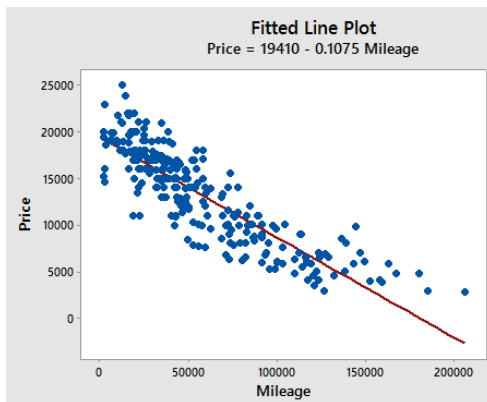
- *Probability* is the likelihood that an event outcome occurs
 - $P(A) = \frac{\text{Count of event outcome } A \text{ occurring}}{\text{Total number of occurrences}}$
 - $P(A)$ is a number between 0 and 1
- *Conditional Probability* is the probability of event A given that event B has already occurred
 - $P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$
- *Multiplication Rule:*
 - $P(A \text{ and } B) = P(A|B) * P(B) = P(B|A) * P(A)$
 - *Independent Events:* $P(A \text{ and } B) = P(A) * P(B)$
- *Addition Rule:*
 - $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
 - *Mutually Exclusive Events:* $P(A \text{ and } B) = 0$
- *Example:*
 - $P(\text{test indicates defective and product is not defective}) = P(\text{test indicates defective} | \text{product is not defective}) * P(\text{product is not defective})$

Data Visualization

- Tables
- Charts / Graphs
- Scatter Plots
- Frequency Histograms and Polygons
- Probability Distributions
- Others
 - Stem and leaf, box plots, decision trees

Multivariate Tests^a

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.939	1028.410 ^b	6.000	400.000	.000
	Wilks' Lambda	.061	1028.410 ^b	6.000	400.000	.000
	Hotelling's Trace	15.426	1028.410 ^b	6.000	400.000	.000
	Roy's Largest Root	15.426	1028.410 ^b	6.000	400.000	.000
q7	Pillai's Trace	.114	2.637	18.000	1206.000	.000
	Wilks' Lambda	.888	2.684	18.000	1131.856	.000
	Hotelling's Trace	.123	2.727	18.000	1196.000	.000
	Roy's Largest Root	.100	6.733 ^c	6.000	402.000	.000

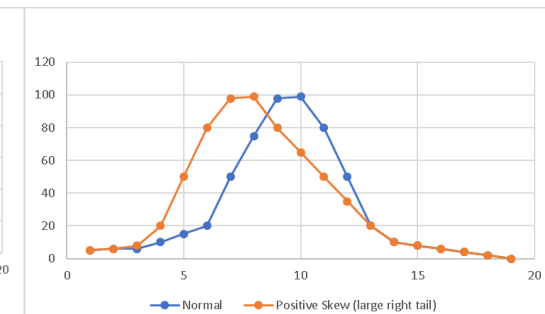
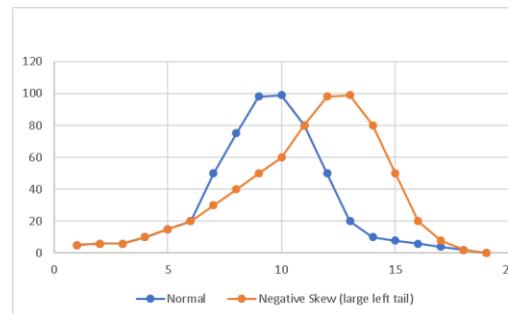
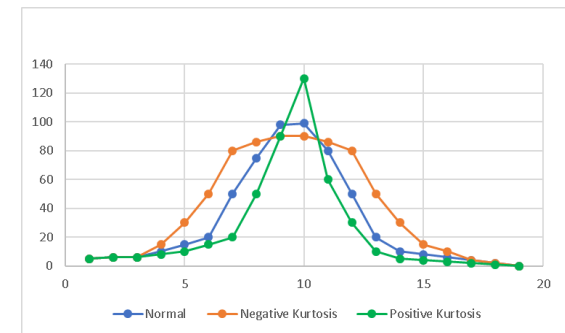
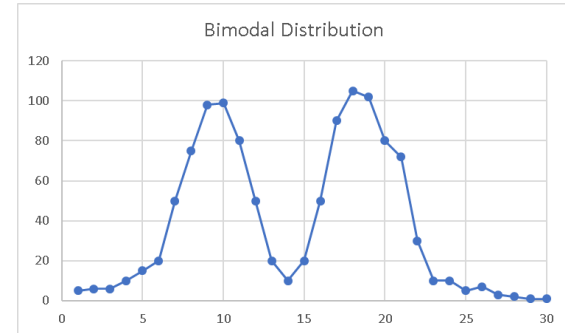


Agenda

- Statistics
- **Statistical Measures**
- Distributions
- Repeatability and Reproducibility (R&R) Analysis
- Process Capability
- Statistical Process Control

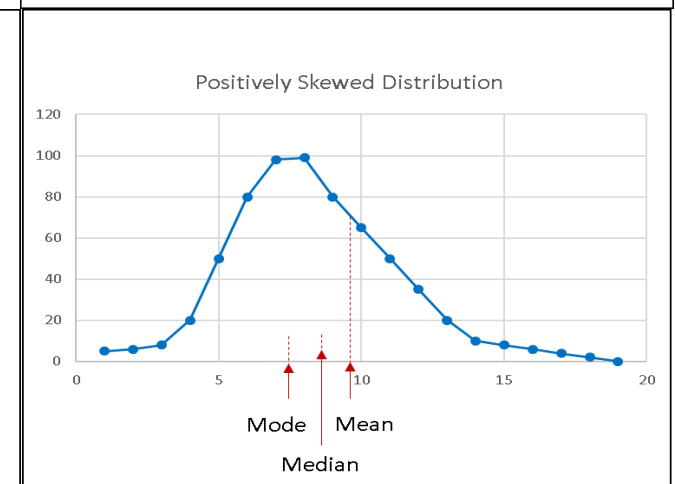
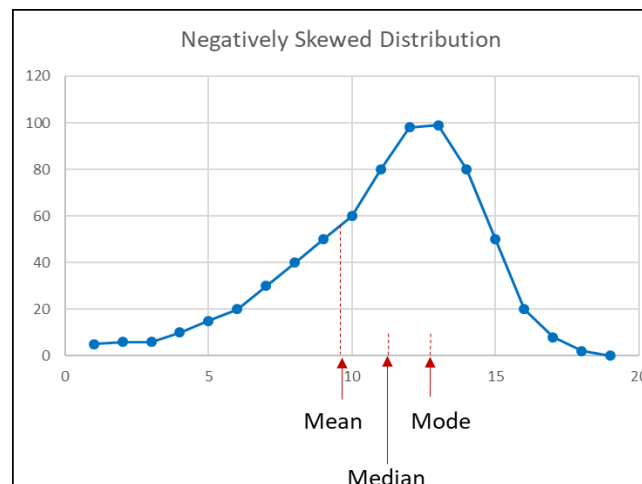
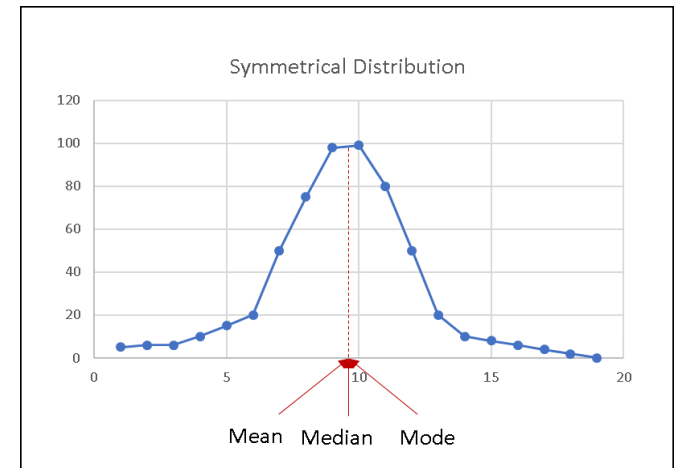
Distributional Shape

- Symmetry
- Skewness
 - Negatively Skewed: tail points to the lower end of the x axis
 - Positively Skewed: tail points to the upper end of the x axis
 - Coefficient of Skewness : $-1 < CS < 1$; $CS = 0 \Rightarrow$ no skewness
- Modality
 - Bimodal, trimodal distributions, etc.
- Kurtosis
 - More or less peaked: leptokurtic, platykurtic
 - Coefficient of Kurtosis: $CK > 3$ more peaked; $CK < 3$ more flat



Measures of Central Tendency

- *Mode*: the most frequently occurring score
- *Median*: midpoint of the distribution of scores; divides the distribution into two equally large parts
- *Mean* (\bar{x} , μ): the average of all scores
- *Midrange*: the average of lowest (L) and highest (H) scores



Measures of Variability

- Variability: degree of dispersion among scores
 - Homogeneous (low variability)
 - Heterogeneous (high variability)
- *Range (R)*: difference between the highest (H) and lowest (L) scores
- *Variance (s^2, σ^2)*
- *Standard Deviation (s, σ)*
- *Coefficient of Variation*: $CV = \frac{s}{\bar{x}} * 100$

Sample	Population
$\bar{X} = \frac{\sum x}{n}$	$\mu = \frac{\sum x}{N}$
$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$	$\sigma = \sqrt{\frac{\sum(x - \mu)^2}{N}}$

Measures of Position

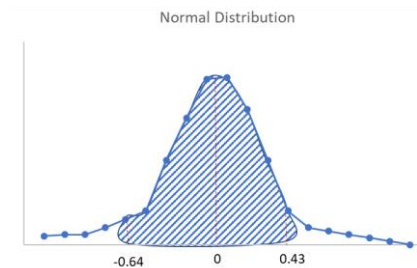
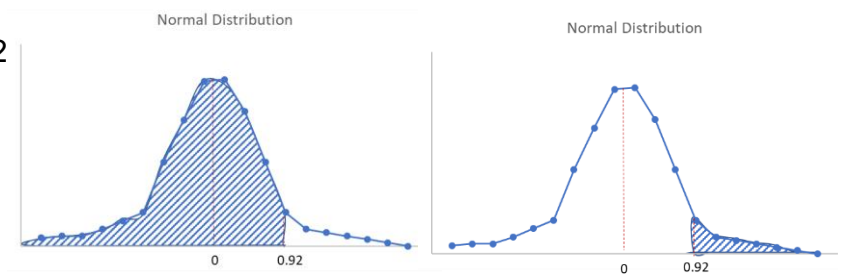
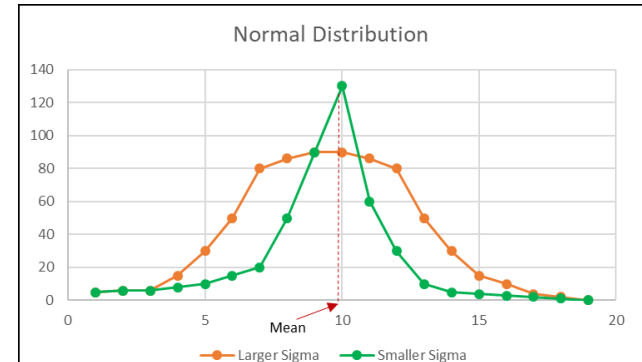
- *Percentiles*: degree of dispersion among scores
 - Q1 = the 25th percentile value
 - Q2 = the 50th percentile value = the median
 - Q3 = the 75th percentile value
- *Interquartile Range*: distance between the low and the high score for the middle half of the data
 - IQR = Q3 - Q1
- *Standard Scores*: **z-score** and *t-score* are the most common
 - Every sample observation x has a corresponding $z\text{-score} = \frac{x - \bar{x}}{s}$
 - Indicates how many standard deviations (s) a particular raw score x lies above or below the sample mean \bar{x}

Agenda

- Statistics
- Statistical Measures
- **Distributions**
- Repeatability and Reproducibility (R&R) Analysis
- Process Capability
- Statistical Process Control

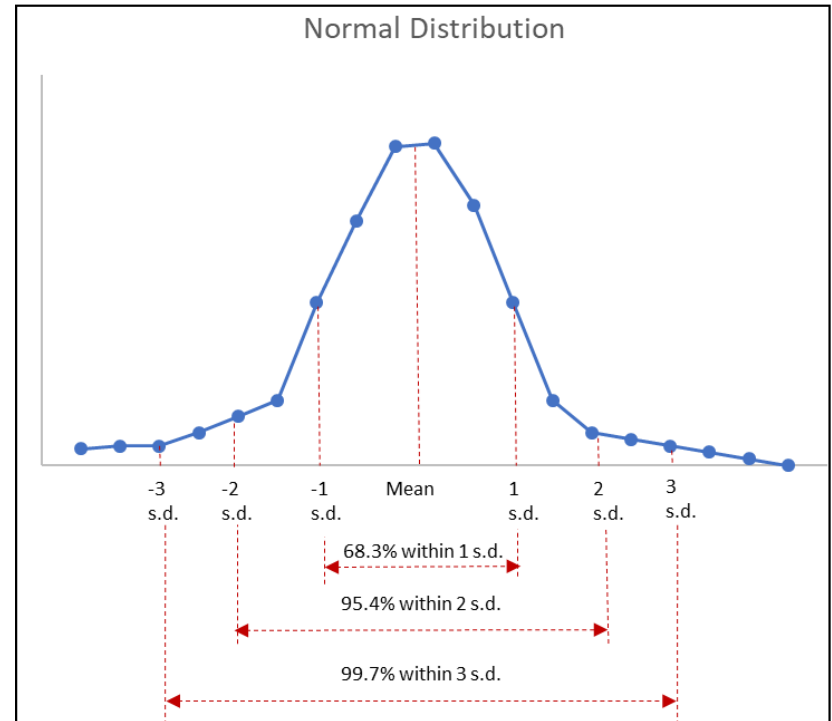
Normal (Gaussian) Distribution

- X is Normally distributed with mean μ and standard deviation $\sigma \Rightarrow X \sim N(\mu, \sigma^2)$
 - The $N(0,1)$ distribution is called Standard Normal Distribution
- Standard Normal Distribution
 - Z is a random variable taking values z (z-scores)
 - Enable easy calculations (using tables, Excel, calculator, etc.)
 - $P(Z < 0.92) = 0.8212$ (from tables)
 - Excel: $P(Z < 0.92) = \text{NORM.S.DIST}(0.92, \text{TRUE}) = 0.8212$
 - $P(-\infty < z < +\infty) = 1$
 - $P(Z > 0.92) = 1 - P(Z < 0.92) = 1 - 0.8212 = 0.1788$
 - $P(Z < 0) = 0.5 = P(z > 0)$
 - $P(-0.64 < Z < 0.43) = P(Z < 0.43) - P(z < -0.64) = 0.6664 - (1 - 0.7389) = 0.4053$



Empirical Rule

- 68.3% of the distribution lies within 1σ from the mean μ
- 95.4% of the distribution lies within 2σ from the mean μ
- 99.7% of the distribution lies within 3σ from the mean μ

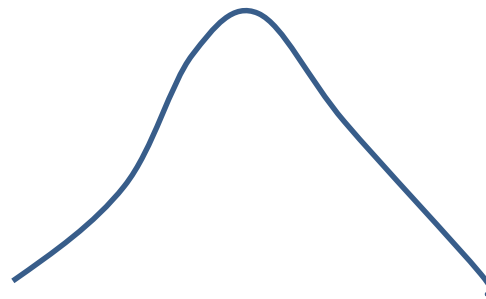
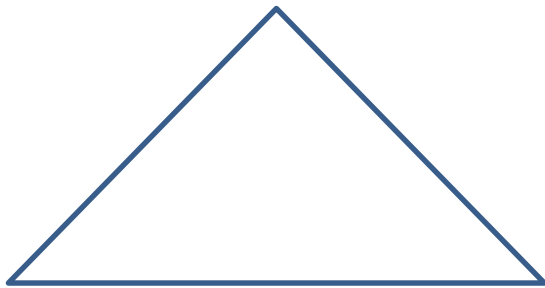


Standardization and the Central Limit Theorem (CLT)

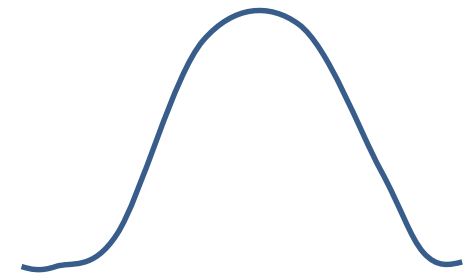
- Standardization: converts any normal distribution value to a standard normal distribution value using the transformation $Z = \frac{X-\mu}{\sigma}$
 - If $X \sim N(\mu, \sigma^2)$ and $Z = (X-\mu) / \sigma$ then $Z \sim N(0, 1)$
 - If $X \sim N(3500, 500)$ => calculate $P(X < 3100)$
 - $P(X < 3100) = P[(X-\mu) / \sigma < (3100-\mu) / \sigma] = P[Z < (3100-3500) / 500]$, where $Z \sim N(0,1)$
== $P(Z < -0.8) = 0.2119$
- Central Limit Theorem (CLT)
 - If simple random samples of size n are taken from *any* population having a mean μ and standard deviation σ , the probability distribution of the *sample mean* approaches a *normal distribution*
 - As $n \rightarrow \infty$ the distribution of the variable $Z = \frac{\bar{X}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$ approaches standard normal, $Z \sim N(0, 1)$
 - Pillar of statistical inference: allows using the sample mean distribution and its particular properties to make inferences about the population

CLT Illustration

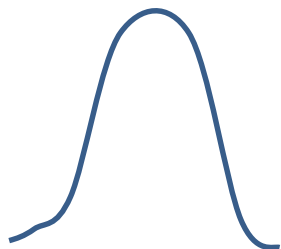
Theoretical Distribution



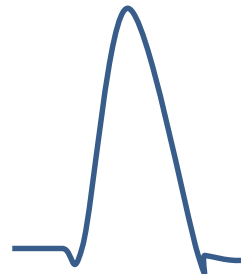
Sample Size = 1



Sample Size = 2



Sample Size = 10



Sample Size = 20

Importance of CLT

- In certain conditions (e.g. large sample size), one can approximate any distribution with a Normal distribution although the distribution is not Normally distributed
 - Through sampling, distributions that are prohibitively difficult to define are approximated by the sampling distribution of the mean which is a standard normal distribution
 - Results can be generalized from sample to the population
- Allows inference from sample to population
- With a large enough sample, most of the sample means will be close to the population mean
- Can determine probability that a certain sample mean falls within a certain distance from the population mean

Normal Distribution Example 1

- The average length of steel bars produced by a construction company has historically been 75 inches, with a standard deviation of 0.25 inch. If a sample of 49 steel bars is taken, what is the probability that the sample mean is at least 75.05 inches?
- Solution
 - $\mu = 75$ in.; $\sigma = 0.25$ in.; $n = 49$
 - $P(\bar{X} \geq 75.05) = ?$
 - CLT $\Rightarrow Z = \frac{\bar{X} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} \sim N(0, 1)$
 - $P(\bar{X} \geq 75.05) = P\left(\frac{\bar{X} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} \geq \frac{75.05 - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right) = P\left(Z \geq \frac{75.05 - 75}{\left(\frac{0.25}{\sqrt{49}}\right)}\right) = P(Z \geq 1.4) = 1 - P(Z < 1.4) = 1 - \text{NORM.S.DIST}(1.4, \text{true}) = 0.0808$

Normal Distribution Example 2

- A manufacturer of MRI scanners has data that indicates that the number of days between scanner malfunctions is normally distributed, with a mean of 1020 days and a standard deviation of 20 days. What is the number of days for which the probability of scanner malfunction is 0.8.
- Solution
 - $\mu = 1020$ days; $\sigma = 20$ days; $P(X < x) = 0.8$
 - Standardization of x values using $z = \frac{x - \mu}{\sigma} \Rightarrow Z \sim N(0,1)$ and $P(Z < z) = 0.8 \Rightarrow z = \text{NORM.S.INV}(0.8) = 0.84 \Rightarrow x = z * \sigma + \mu = 0.84 * 20 + 1020 = 1036.8$

Normal Distribution Example 3

- Assume that the noise in a digital transmission system is normally distributed with a mean of 0 Volts and a standard deviation of 0.45 Volts. A digital “1” is transmitted when voltage exceeds 0.9 Volts. What is the probability of false detection at the receiving end? Determine symmetric bounds around 0 that include 99% of all noise readings.
- Solution
 - $\mu = 0; \sigma = 0.45$
 - False detection = detect a digital “1” when none was sent => noise > 0.9 Volts will wrongly be interpreted as digital “1”
 - $P(X > 0.9) = P\left(\frac{x-\mu}{\sigma} > \frac{0.9-\mu}{\sigma}\right) = P\left(Z > \frac{0.9}{0.45}\right) = P(Z > 2) = 1 - P(Z < 2) = 1 - \text{NORM.S.DIST}(2, \text{true}) = 0.02275$
 - $X = \text{noise} \Rightarrow P(-x < X < x) = 99\% \Rightarrow \text{standardization} \Rightarrow P(-z < Z < z) = 99\% \Rightarrow P(Z < z) - P(Z < -z) = 0.99 \Rightarrow 1 - P(Z > z) - P(Z < -z) = 0.99 \Rightarrow 1 - 2 * P(Z < -z) = 0.99 \Rightarrow P(Z < -z) = (1 - 0.99) / 2 = 0.005 \Rightarrow -z = \text{NORM.S.INV}(0.005) = -2.58 \Rightarrow x = z * \sigma + \mu = 2.58 * 0.45 + 0 = 1.161 \Rightarrow P(-1.161 < X < 1.161) = 99\%$

Determining Sample Size (σ unknown)

- A telecommunications company claims that 90% of their voice & data switches still function after major data center floods when they were under water for up to 8 hours. How many flooded switches have to be tested to determine the true proportion with a 95% confidence interval 6% wide (i.e. margin of error 3%)? What should the sample size be if we decrease the margin of error from 3% to 2%?

- Solution:

- CL = 95% $\Rightarrow \alpha = 1 - 95\% = 0.05$

(1- α)*100% confidence interval for p using a large sample is

$$\hat{p} - Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}} \quad \text{to} \quad \hat{p} + Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}$$

- Margin of Error: $E = 6\%/2 = 0.03 = Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}$ (a)

- $Z_{\alpha/2} = \text{NORM.S.INV}(1-\alpha/2) = 1.96$ (or from Z tables)

- $\hat{p} = 90\% \Rightarrow$ from eq. (a) $\Rightarrow n = 1 + (Z_{\alpha/2}/E)^2 [\hat{p}(\hat{p} - 1)] = 1 + (1.96/0.03)^2 (0.9*0.1) = 385.16 \Rightarrow 386$ switches to be tested

- $n' = 1 + (Z_{\alpha/2}/E')^2 [\hat{p}(\hat{p} - 1)] = 1 + (1.96/0.02)^2 (0.9*0.1) = 865.36 \Rightarrow 866$ switches to be tested

- Precision increases 1% (from 3% to 2%) \Rightarrow larger sample size is needed (from 386 to 866)

Determining Sample Size (σ unknown)

- A semiconductor company manufactures military specification resistors. We want to know the average resistance of these special resistors, with a margin of error of 10%, and be 99% confident about our result. How many special resistors should be inspected and measured? From previous studies, we know σ is 3. What should the sample size be if we relax the confidence level to 95%?

- Solution:

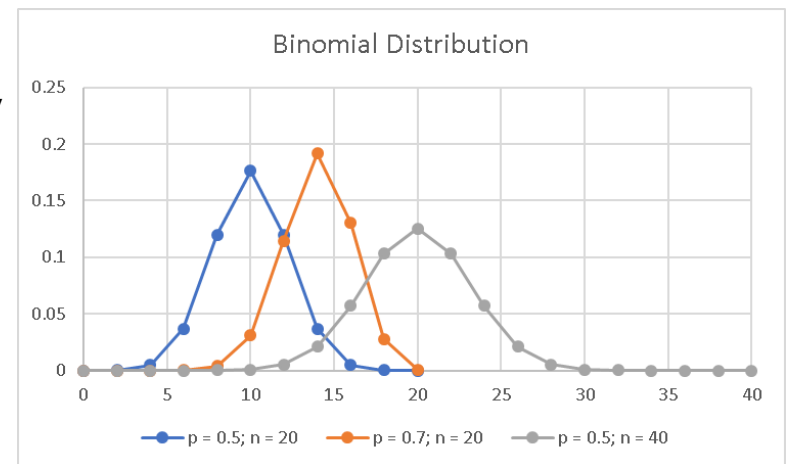
- CL = 99% $\Rightarrow \alpha = 1 - 99\% = 0.01$
- Margin of Error: $E = 10\% = 0.1 = k \left(\frac{\sigma}{\sqrt{n}} \right)$ (a)
- $k = Z_{\alpha/2} = \text{NORM.S.INV}(1 - \alpha/2) = 2.576$ (or from Z tables)
- From eq. (a) $\Rightarrow n = [k * (\sigma / E)]^2 = (2.576 * 3 / 0.1)^2 = 5972.2 \Rightarrow 5973$ resistors to be measured
- High confidence (99%) \Rightarrow large sample size (5973)
- CL' = 95% $\Rightarrow \alpha' = 1 - 95\% = 0.05 \Rightarrow k' = Z_{\alpha'/2} = \text{NORM.S.INV}(1 - \alpha'/2) = 1.96$ (or from Z tables)
 \Rightarrow
- $n' = [k' * (\sigma / E)]^2 = (1.96 * 3 / 0.1)^2 = 3457.44 \Rightarrow 3458$ resistors to be measured
- Lower confidence (95%) \Rightarrow smaller sample size needed (3458)

(1- α)*100% confidence interval for μ is:

$$\bar{X} - Z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}} \right) \quad \text{to} \quad \bar{X} + Z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}} \right)$$

Binomial Distribution

- Discrete probability of obtaining exactly x “successes” in a sequence of n trials
- $p(x) = \binom{n}{x} p^x (1 - p)^{n-x}, x = 0, 1, 2, \dots, n = \frac{n!}{x!(n-x)!} p^x (1 - p)^{n-x}, x = 0, 1, 2, \dots, n$
- $\mu = np, \sigma^2 = np(1 - p)$
- “Success” = any one of two possible outcomes
 - E.g. Defective / non-defective , male / female, etc.
- p = probability of “success”
- Excel: BINOM.DIST (x, n, p, TRUE)
- TRUE => result is cumulative probability

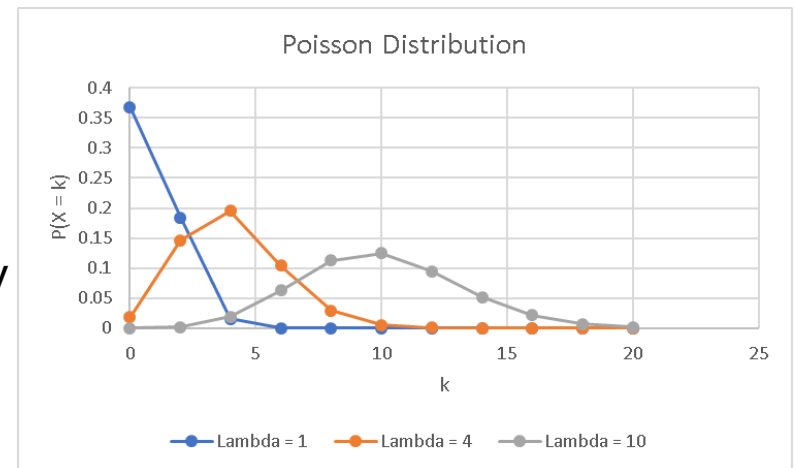


Binomial Distribution Example

- The probability that a process produces a non-defective part is 0.8. What is the probability that 4 parts among a sample of 20 will be defective? What is the probability that at least 1 part will be defective? What is the expected (average) number of defective parts if the sample size is increased to 50?
- Solution:
 - “Success” = defective part; $p = P(\text{“success”}) = P(\text{defective}) = 1 - P(\text{non-defective}) = 1 - 0.8 = 0.2$
 - $x = 4, n = 20 \Rightarrow P(4) = \frac{20!}{4! \cdot 16!} * 0.2^4 * 0.8^{16} = (17 \cdot 18 \cdot 19 \cdot 20 / 24) * 0.002 * 0.028 = 0.218$
 - EXCEL: $P(4) = \text{BINOM.DIST}(4, 20, 0.2, \text{false}) = 0.218$
 - $P(X \geq 1) = 1 - P(X < 1) = 1 - P(0) = 1 - \text{BINOM.DIST}(0, 20, 0.2, \text{false}) = 1 - 0.012 = 0.988$
 - $E(\text{defective}) = \mu = n * p = 50 * 0.2 = 10$

Poisson Distribution

- Discrete probability of exactly x events occurring in a fixed interval (can be time, space, area, volume, etc.) if these events occur with a known average rate λ (per interval) and independently of the time since the last event
- $$p(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$
- $\mu = \lambda, \sigma^2 = \lambda$
- λ = average rate or expected number of occurrences
- Excel: POISSON.DIST (x, λ, TRUE)
- TRUE => result is cumulative probability



Poisson Distribution: Example 1

- The average number of non-conforming products found during inspection is 12. What is the probability that exactly 5 non-conforming products are found? What is the probability that a maximum of 3 non-conforming units are found? What is the probability that between 2 and 6 non-conforming units are found?
- Solution:
 - $\lambda = 12, x = 5 \Rightarrow P(5) = \text{POISSON.DIST}(5, 12, \text{false}) = 0.0127$
 - $P(X < 4) = \text{POISSON.DIST}(4, 12, \text{true}) = 0.0076$
 - $P(2 < X < 6) = P(X < 6) - P(x < 2) = \text{POISSON.DIST}(6, 12, \text{true}) - \text{POISSON.DIST}(2, 12, \text{true}) = 0.045822 - 0.000522 = 0.0453$

Poisson Distribution: Example 2

- 5 sheets of polished metal are examined for scratches and the results are given in the table below. What is the probability of finding no scratches per square inch? What is the probability of choosing a sheet at random that contains 4 or more scratches?
- Solution:
 - $\lambda = \text{average \# of scratches per sq. in.} = (4+3+5+2+4) / (25+30+40+15+20) = 0.138$
 - $P(0) = \text{POISSON.DIST}(0, 0.138, \text{false}) = 0.87109$
 - $P(4 \text{ or more scratches}) = \text{Number of sheets with 4 or more scratches} / \text{Total number of sheets} = 3/5 = 0.6$

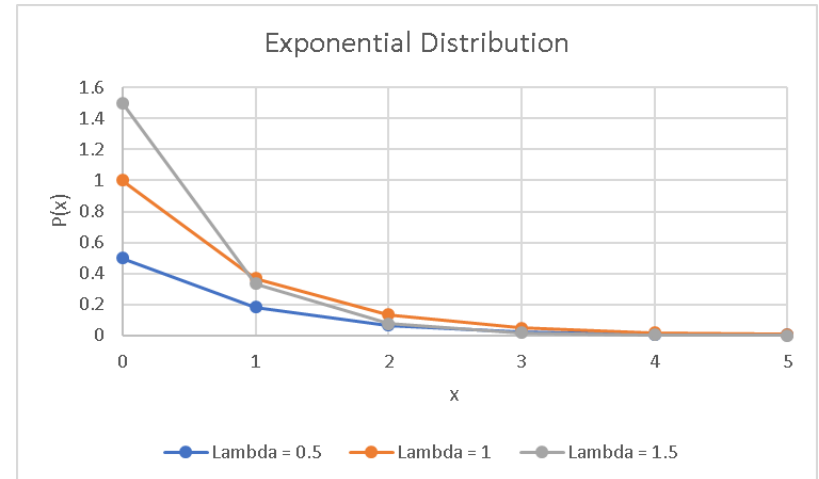
Sheet #	Surface Area (sq. in.)	# of Scratches
1	25	4
2	30	3
3	40	5
4	15	2
5	20	4

Poisson Distribution: Example 3

- Electricity power failures occur with an average of 3 failures every 20 weeks. What is the probability that there will not be more than one failure during a particular week?
- Solution:
 - $\lambda = 3/20 = 0.15$
 - $P(\text{not more than 1 failure}) = P(0)+P(1) = \text{POISSON.DIST}(0, 0.15, \text{false}) + \text{POISSON.DIST}(1, 0.15, \text{false}) = 0.860708+0.129106 = 0.989814$

Exponential Distribution

- Models the time between randomly occurring events in a Poisson process i.e. where events occur continuously and independently at a constant average rate λ
- Example: time between failures
- $f(x) = \lambda e^{-\lambda x}, x > 0$
 $\mu = \frac{1}{\lambda}, \sigma^2 = \frac{1}{\lambda^2}$
 $F(x) = 1 - e^{-\lambda x}, x > 0$
- F(x) calculates probability of failure within x hours
- Excel: F(x) = EXPON.DIST (x, λ , TRUE)
- TRUE => result is cumulative probability



Product Reliability

- *Failure rate:*

$$\lambda = \frac{\text{\# of failures}}{\text{Total unit operating hours}} = \frac{\text{\# of failures}}{(\text{Units tested}) * (\text{number of hours tested})}$$

- *Mean Time to Failure: MTTF = $\theta = 1 / \lambda$; used for replaceable products*
- *Probability of failure by time T : $F(T) = 1 - e^{-\lambda T} = 1 - e^{-T/\theta}$*
- *Probability of failure during a time interval: $F(T_2) - F(T_1) = e^{-\lambda(T_2 - T_1)}$*
- *Mean Time Between Failures (MTBF): sum of the lengths of the operational periods divided by the number of observed failures; used for reparable products*

$$MTBF = \frac{\sum(\text{start of downtime} - \text{start of uptime})}{\text{\# of failures}}$$

- *Reliability Function: probability of survival: $R(T) = 1 - F(T) = e^{-\lambda T} = e^{-T/\theta}$*

Exponential Distribution: Example 1

- A large number of electronic system components is tested and the average time to failure is found to be 4000 hours. What is the probability that a component will fail within 500 hours?
- Solution
 - $\lambda = \text{average rate} \Rightarrow 1/\lambda = \text{average time (in this case - to failure)} \Rightarrow 1/\lambda = 4000 \text{ hrs.} \Rightarrow \lambda = 1/(4000 \text{ hrs.}) = 0.00025 \text{ failures/hr.}$
 - $P(\text{failure within } 500 \text{ hrs.}) = F(500) = \text{EXPON.DIST}(500, 0.00025, \text{TRUE}) = 0.1175$

Exponential Distribution: Example 2

- Assume that the average time to failure of a particular make of a car cooling fan is 3333 hours. Find the proportion of fans that will give at least 10000 hours service. If the fan is redesigned so that the average time to failure is 4000 hours, would you expect more fans or less to give at least 10000 hours of service?
- Solution
 - $1/\lambda = \text{average time to failure} = 3333 \text{ hrs.} \Rightarrow \lambda = 1/(3333 \text{ hrs.}) = 0.0003$ failures/hr.
 - $P(X \geq 10000) = 1 - P(X < 10000) = 1 - F(10000) = 1 - \text{EXPON.DIST}(10000, 0.0003, \text{TRUE}) = 0.0497 \Rightarrow$ approx. 5% of the fans will give at least 10000 hours of service
 - $1/\lambda = \text{average time to failure} = 4000 \text{ hrs.} \Rightarrow \lambda = 1/(4000 \text{ hrs.}) = 0.00025$ failures/hr.
 - $P(X \geq 10000) = 1 - P(X < 10000) = 1 - F(10000) = 1 - \text{EXPON.DIST}(10000, 0.00025, \text{TRUE}) = 0.0821 \Rightarrow$ approx. 8.2 % of the fans will give at least 10000 hours of service

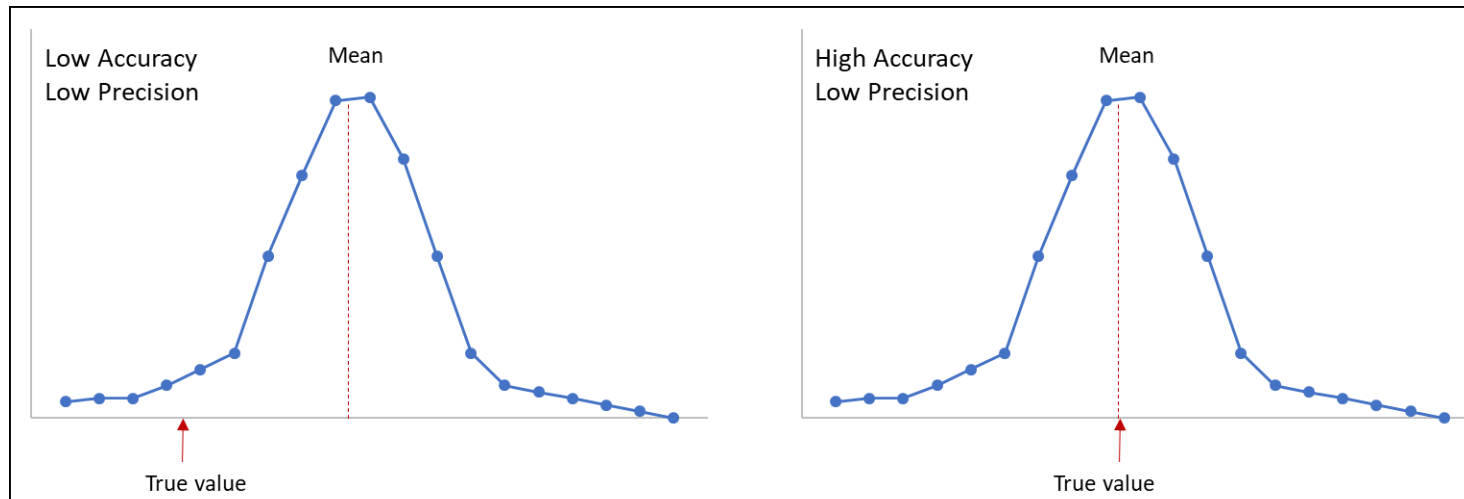
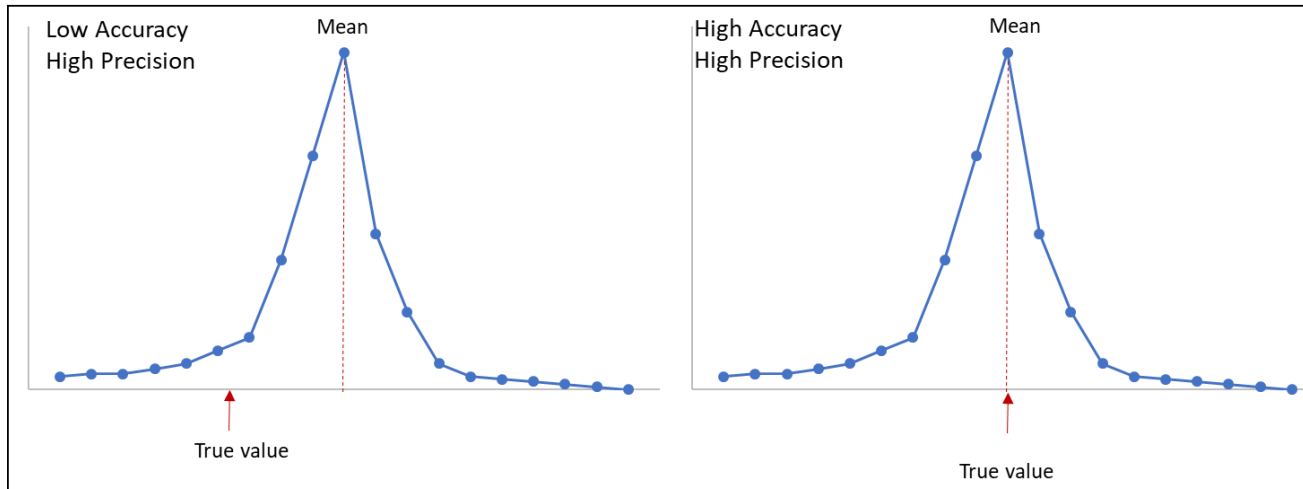
Exponential Distribution: Example 3

- Assume that an electronic component has a failure rate of 0.0001 failures per hour. What is the mean time to failure? Calculate the probability that the component will not fail in 15000 hours.
- Solution
 - $\lambda = 0.0001$ failures / hr.
 - $MTTF = \theta = 1/\lambda = 1/0.0001 = 10000$ hrs.
 - Probability that a component will not fail in 15000 hrs. = $R(15000) = e^{-\frac{15000}{10000}} = 0.223$

Agenda

- Statistics
- Statistical Measures
- Distributions
- **Repeatability and Reproducibility (R&R) Analysis**
- Process Capability
- Statistical Process Control

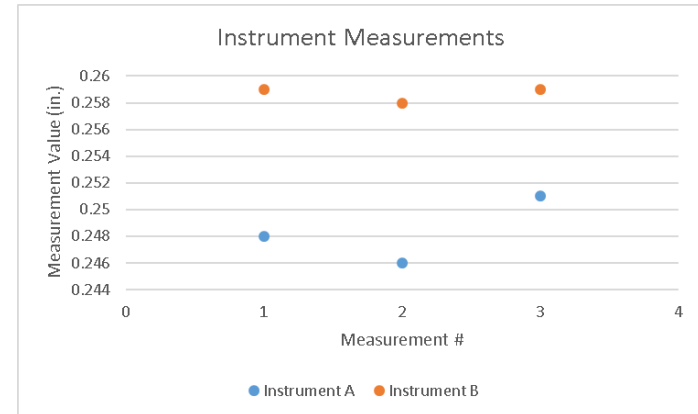
Measurement Systems Evaluation



Measurement Systems Evaluation Example

- Two instruments measure an attribute whose true value is 0.250 in, with the results given in the table below. Which instrument is more precise, and more accurate?

Meas. #	Instrument A	Instrument B
1	0.248	0.259
2	0.246	0.258
3	0.251	0.259



- Solution**

- $\text{Relative_Error}_A = |(0.250 - 0.248) / 0.250| = 0.8\%$
- $\text{Relative_Error}_B = |(0.250 - 0.259) / 0.250| = 3.6\% \Rightarrow$ Instrument A is more accurate
- Instrument B values are more clustered together than instrument A \Rightarrow Instrument B is more precise

Repeatability and Reproducibility (R&R) Analysis

- $\sigma_{total}^2 = \sigma_{process}^2 + \sigma_{measurement}^2$
- An R&R study is a study of variation in a measurement system using statistics
 - Select m operators and n parts
 - Calibrate the measuring instrument
 - Randomly measure each part by each operator for r trials
 - Compute key statistics to quantify repeatability and reproducibility
- *Repeatability (equipment variation, EV)*: variation in multiple measurements by an individual using the same instrument
- *Reproducibility (appraiser variation, AV)*: variation in the same measuring instrument used by different individuals
- *Part Variation (PV)*: measures variation among different parts
- *Total Variation(TV)*: $TV^2 = R\&R^2 + PV^2 = EV^2 + AV^2 + PV^2$

Repeatability and Reproducibility (R&R) Analysis (cont.)

- A measurement system is adequate if R&R is low relative to the total variation, or equivalently, the PV is much greater than the measurement system variation

$$\%EV = 100 \frac{EV}{TV}$$

$$\%AV = 100 \frac{AV}{TV}$$

$$\%R\&R = 100 \frac{R\&R}{TV}$$

$$\%PV = 100 \frac{PV}{TV}$$



Under 10% error – OK
 10-30% error – may be OK
 Over 30% error – unacceptable

$$EV\% \text{ of Total Variance} = 100 \frac{EV^2}{TV^2}$$

$$AV\% \text{ of Total Variance} = 100 \frac{AV^2}{TV^2}$$

$$R\&R\% \text{ of Total Variance} = 100 \frac{R\&R^2}{TV^2}$$

$$PV\% \text{ of Total Variance} = 100 \frac{PV^2}{TV^2}$$

Number of operators		3	Upper specification limit		1	Average range		0.042								
Number of trials		2	Lower specification limit		0.5	Range of operator average		0.050								
Number of parts		10				Range of part averages		0.525								
Data	Operator 1				Operator 2				Operator 3				Part			
Part #	Trial			Average	Range	Trial			Average	Range	Trial			Average	Range	Average
1	0.590	0.570		0.580	0.020	0.580	0.570		0.575	0.010	0.540	0.530		0.535	0.010	0.563
2	0.960	0.980		0.970	0.020	1.060	0.970		1.015	0.090	1.080	1.000		1.040	0.080	1.008
3	0.790	0.750		0.770	0.040	0.820	0.770		0.795	0.050	0.840	0.800		0.820	0.040	0.795
4	0.820	0.920		0.870	0.100	0.840	0.790		0.815	0.050	0.840	0.800		0.820	0.040	0.835
5	0.550	0.490		0.520	0.060	0.450	0.440		0.445	0.010	0.490	0.480		0.485	0.010	0.483
6	0.940	0.960		0.950	0.020	1.020	1.050		1.035	0.030	1.070	0.990		1.030	0.060	1.005
7	0.920	0.940		0.930	0.020	0.960	0.910		0.935	0.050	0.980	0.940		0.960	0.040	0.942
8	0.820	0.810		0.815	0.010	0.740	0.750		0.745	0.010	0.840	0.800		0.820	0.040	0.793
9	0.930	0.950		0.940	0.020	1.000	0.950		0.975	0.050	1.060	1.020		1.040	0.040	0.985
10	0.600	0.700		0.650	0.100	0.580	0.530		0.555	0.050	0.870	0.800		0.835	0.070	0.680
Average				0.800	0.041				0.789	0.040				0.839	0.045	
				% of Total Variati		% of Tolerance		Variance Ratios								
Repeatability (Equipment Variation) - EV				0.192	EV	21.74%	EV	38.30%	EV	4.73%						
Reproducibility (Appraiser Variation) - AV				0.127	AV	14.37%	AV	25.32%	AV	2.07%						
Repeatability and Reproducibility - R&R				0.230	R&R	26.06%	R&R	45.92%	R&R	6.79%						
Part Variation - PV				0.851	PV	96.54%	PV	170.10%	PV	93.21%						
Total Variation - TV				0.881												
Upper control limit for individual ranges				0.137												

Agenda

- Statistics
- Statistical Measures
- Distributions
- Repeatability and Reproducibility (R&R) Analysis
- **Process Capability**
- Statistical Process Control

Process Capability Studies

- *Process capability*: the ability of a process to produce output that conforms to specifications
- Typical questions include:
 - Where is the process centered?
 - How much variability exists in the process?
 - Is the performance relative to specifications acceptable?
 - What proportion of output will be expected to meet specifications?
 - What factors contribute to variability?
- Process Capability Indexes
 - Process centered on specification range:
 - $C_p > 1 \Rightarrow$ process capable of meeting specifications
 - $C_p < 1 \Rightarrow$ process produces some nonconforming output
 - Process un-centered \Rightarrow use C_{pu} , C_{pl} , C_{pk}

$$C_p = \frac{USL - LSL}{6\sigma}$$

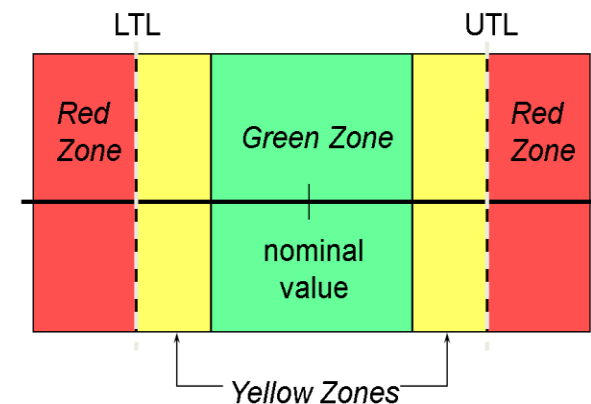
$$C_{pu} = \frac{USL - \mu}{3\sigma} \text{ (upper one - sided index)}$$

$$C_{pl} = \frac{\mu - LSL}{3\sigma} \text{ (lower one - sided index)}$$

$$C_{pk} = \min(C_{pl}, C_{pu})$$

Pre-Control

- Used for $C_p \geq 1.14$
- Divide the tolerance range into zones by setting two pre-control lines halfway between the center of the specification and the upper and lower specification limits
 - Green zone: comprises 50% of the total tolerance
 - Yellow zone: between the pre-control lines and the specification limits
 - Red Zone: outside the specification limits
- At process start: 5 consecutive parts must fall within the green zone
 - If not, the production setup must be reevaluated before the full production run can start
- During regular process operations: sample 1 part
 - If it falls within the green zone, production continues
 - If it falls in a yellow zone, a 2nd part is inspected.
 - If this falls in the green zone, production continues
 - If not, production stops and a special cause should be investigated
- If any part falls in a red zone, then action should be taken



Process Capability Example

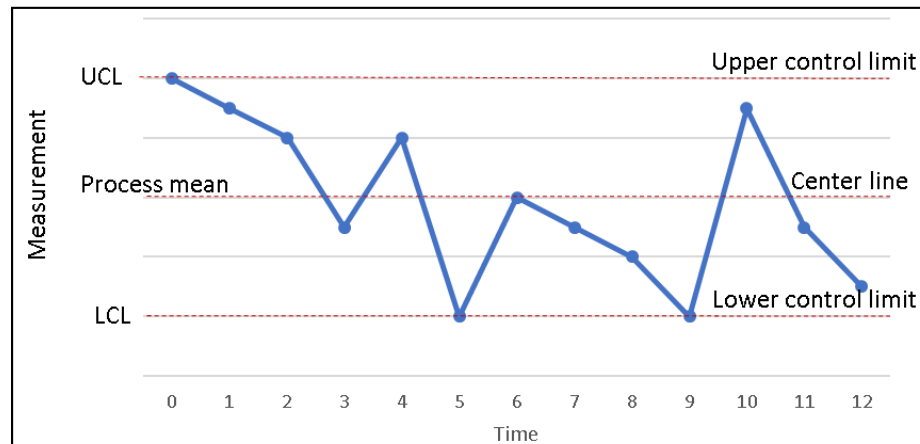
- Diameter measurements of automotive bearings in a random sample indicate an average $\bar{x} = 10.8273$, a standard deviation $s = 0.0767$, and a normal distribution. If the product design specifications are between 10.65 and 10.95, will the process produce nonconforming units? What is the proportion of units below specification; above specification? What is the probability that a part will not meet specification?
- Solution:
 - Empirical rule: virtually all dimensions are expected to fall within 3 std. dev. from the mean
 - Lower limit: $10.8273 - 3 \cdot 0.0767 = 10.597$; Upper Limit: $10.8273 + 3 \cdot 0.0767 = 11.057$
 - Expected interval: [10.597, 11.057]; Production interval: [10.65, 10.95] => Some non-conforming units are expected
 - Proportion of units below 10.65 = $\text{NORM.DIST}(10.65, 10.8273, 0.0767, \text{TRUE}) = 0.0104 = 1.04\%$
 - Proportion of units above 10.95 = $1 - \text{NORM.DIST}(10.95, 10.8273, 0.0767, \text{TRUE}) = 0.0548 = 5.48\%$
 - Probability that a unit will not meet specifications = $0.0104 + 0.0548 = 0.065 = 6.5\%$
 - Process not centered on specified range [$\bar{x} \ll$ average (10.65, 10.95)] => use C_{pu} , C_{pl} , C_{pk}
 - $C_{pu} = (\text{USL} - \bar{x}) / (3s) = (10.95 - 10.8273) / (3 \cdot 0.0767) = 0.533$
 - $C_{pl} = (\bar{x} - \text{LSL}) / (3s) = (10.8273 - 10.65) / (3 \cdot 0.0767) = 0.771$
 - $C_{pk} = \min(C_{pl}, C_{pu}) = 0.533 < 1$ => process will produce non-conforming units

Agenda

- Statistics
- Statistical Measures
- Distributions
- Repeatability and Reproducibility (R&R) Analysis
- Process Capability
- **Statistical Process Control**

Statistical Process Control (SPC)

- Statistical monitoring of a process to identify special causes of variation and signal the need to take action
 - Uses *Control Charts*
- Controlled Process:
 - No points are outside control limits
 - The number of points above and below the center line is about the same
 - The points seem to fall randomly above and below the center line
 - Most points, but not all, are near the center line, and only a few are close to the control limits



Patterns in Control Charts

- Typical out-of-control patterns:
 - One point outside control limits
 - Sudden shift in process average
 - 8 consecutive points fall on one side of the center line
 - 2 out of 3 consecutive points fall in the outer one-third region between the center line and UCL (or LCL)
 - 4 out of 5 consecutive points fall in the outer two-thirds region between the center line and UCL (or LCL)
 - Cycles – short, repeated patterns with alternating high peaks and low valleys
 - Trends – points gradually moving up or down from the center line
 - Hugging the center line – most points fall close to the center line
 - Hugging the control limits - most points are close to the control limits, with few in between

Control Charts

- For Variables Data
 - X-bar and R-charts
 - Point estimate for σ : $\hat{\sigma} = \frac{\bar{R}}{d_2}$
 - X-bar and S-charts
 - Charts for Individuals (X-charts)
- For Attributes Data
 - Fractions nonconforming: p-charts
 - Number nonconforming: np-charts
 - Nonconforming per unit: c-charts, u-charts

	R Chart	X-bar Chart
LCL	$LCL = D_3\bar{R}$ Or 0 if LCL < 0	$LCL = \bar{\bar{X}} - A_2\bar{R}$
CL	$\bar{R} = \frac{\text{sum of subgroup ranges}}{\# \text{ of subgroups}}$	$\bar{\bar{X}} = \frac{\text{sum of subgroup averages}}{\# \text{ of subgroups}}$
UCL	$UCL = D_4\bar{R}$	$UCL = \bar{\bar{X}} + A_2\bar{R}$

	S Chart	X-bar Chart
LCL	$LCL = B_3\bar{S}$ Or 0 if LCL < 0	$LCL = \bar{\bar{X}} - A_3\bar{S}$
CL	$\bar{S} = \frac{\text{sum of subgroup sigmas}}{\# \text{ of subgroups}}$	$\bar{\bar{X}} = \frac{\text{sum of subgroup averages}}{\# \text{ of subgroups}}$
UCL	$UCL = B_4\bar{S}$	$UCL = \bar{\bar{X}} + A_3\bar{S}$

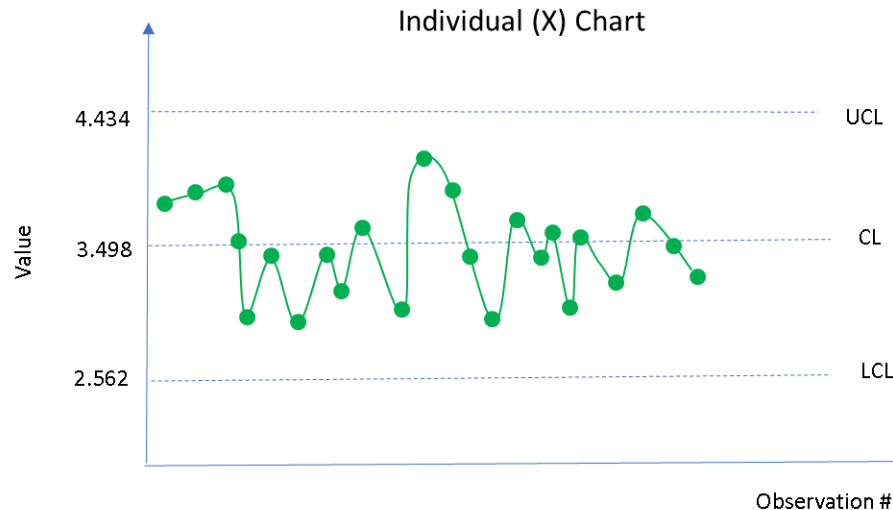
	X Chart
LCL	$LCL = \bar{X} - 2.66 * \bar{R}$
CL	$\bar{X} = \frac{\text{sum of measurements}}{\# \text{ of measurements}}$
UCL	$UCL = \bar{X} + 2.66 * \bar{R}$

	np Chart	p Chart
LCL	$LCL = n\bar{p} - 3\sqrt{n\bar{p}(1-\bar{p})}$ Or 0 if LCL < 0	$LCL = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$ Or 0 if LCL < 0
CL	$n\bar{p} = \frac{\text{sum of subgroup defective counts}}{\# \text{ of subgroups}}$	$\bar{p} = \frac{\text{sum of subgroup defective counts}}{\# \text{ of subgroup sizes}}$
UCL	$UCL = n\bar{p} + 3\sqrt{n\bar{p}(1-\bar{p})}$ Or n if UCL > n	$UCL = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$ Or 1 if UCL > 1

	c Chart	u Chart
LCL	$LCL = \bar{c} - 3\sqrt{\bar{c}}$ Or 0 if LCL < 0	$LCL = \bar{u} - 3\sqrt{\frac{\bar{u}}{n}}$ Or 0 if LCL < 0
CL	$\bar{c} = \frac{\text{sum of subgroup occurrences}}{\# \text{ of subgroups}}$	$\bar{u} = \frac{\text{sum of subgroup occurrences}}{\# \text{ of units in all subgroups}}$
UCL	$UCL = \bar{c} + 3\sqrt{\bar{c}}$	$UCL = \bar{u} + 3\sqrt{\frac{\bar{u}}{n}}$

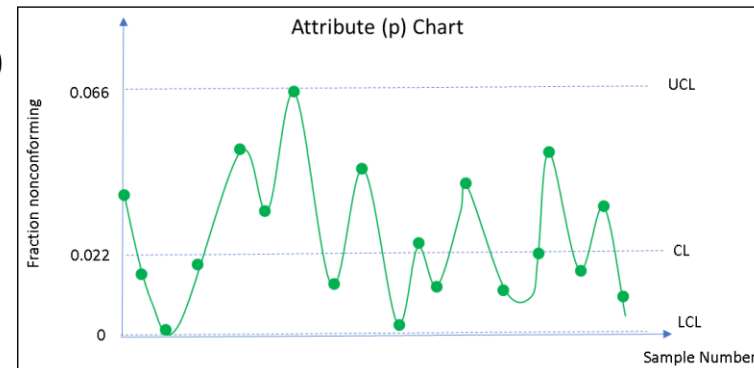
SPC: Example 1

- Consider a set of observations measuring the % of aluminum in a chemical process, with $\bar{x} = 3.498$ and $\bar{R} = 0.35$. Is this process under control?
- Solution:
 - For $n = 2 \Rightarrow 3/d_2 = 2.66 \Rightarrow LCL = \bar{X} - 2.66 * \bar{R} = 3.498 - 2.66*0.352 = 2.562$
 - $UCL = \bar{X} + 2.66 * \bar{R} = 3.498 + 2.66*0.352 = 4.434$
 - The resulting X-chart shows the process is under control



SPC: Example 2

- The operators of automated sorting machines in a post office must read the ZIP code on a letter and divert the letter to the proper carrier route. Over one month's time, 25 samples of 1200 letters were chosen, and the number of errors was recorded. The fraction non-conforming was calculated by dividing the number of errors by 100. From the results, the average fraction non-conforming was determined to be $\bar{p} = 0.022$. Is this process under control?
- Solution:
 - The standard deviation:
 - $s_{\bar{p}} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = \sqrt{\frac{0.022(1-0.022)}{100}} = 0.01467$
 - $UCL = 0.022 + 3*0.01467 = 0.066$
 - $LCL = 0.022 - 3*0.01467 = -0.022 < 0 \Rightarrow LCL = 0$
 - The process appears to be in control



Thank You.

**I hope you enjoyed this overview of
basic statistics.**