Statistics

- **Statistics** is a branch of mathematics dealing with the collection, analysis, interpretation, presentation, and organization of data. (Wikipedia)

- Statistics converts data to information.

Descriptive vs Inferential Statistics

- **Descriptive Statistics:** Summary of the data (population)
  - Dataset: State.X77 - Average population

- **Inferential Statistics:** Based on random samples from population make inference about the population.
  - Dataset: Chickwt - Effect of feed on chicken weight based on samples.

Qualitative vs Quantitative

Data

Qualitative Data	Quantitative Data
Description	Numbers
Car brand, categories eg Automatic vs Manual, gears	Weight of car, miles per gallon, horsepower

QUALiTYGURUS

Qualitative vs Quantitative

Data

Continuous Data

Discrete Data

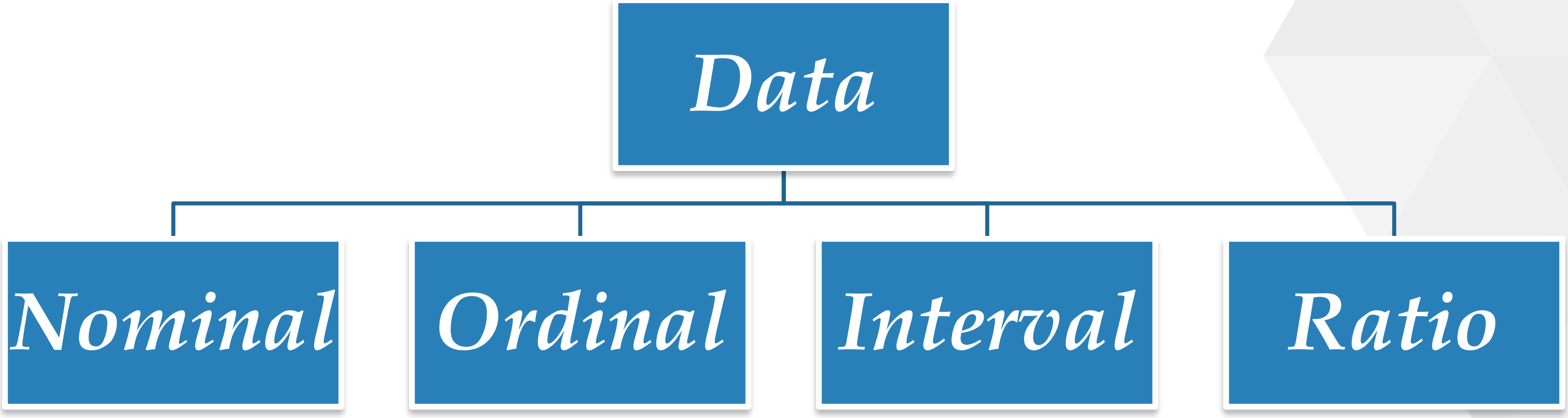
Infinite categories

Finite categories

Weight of car, miles per gallon, horsepower

Number of gears, Automatics vs Manual

Measurement Scales



Measurement Scales

- *Data*
  - *Nominal*
  - *Ordinal*
  - *Interval*
  - *Ratio*

Example:
Color: Blue, Green, Red

Measurement Scales

- **Data**
  - Nominal
  - Ordinal
  - Interval
  - Ratio

Example:
Pass/Fail
Good, Bad, Worst

Measurement Scales

Data

Nominal | *Ordinal* | *Interval* | *Ratio*

Example:
Temperature: Celsius

Measurement Scales

Data

Nominal

Ordinal

Interval

Ratio

Example:
Height, mass, volume

QUALiTYGURUS

Measurement Scales

	Nominal	Ordinal	Interval	Ratio
Ordered	N	Y	Y	Y
Difference	N	N	Y	Y
Absolute Zero	N	N	N	Y
Example	Red, Blue	Good, Bad, Worst	Temperature : Degree C	Length, Weight
Central Tendency Measurement	Mode	Mode, Median	Mode, Median, Mean	Mode, Median, Mean

Qualitative vs Quantitative

Data

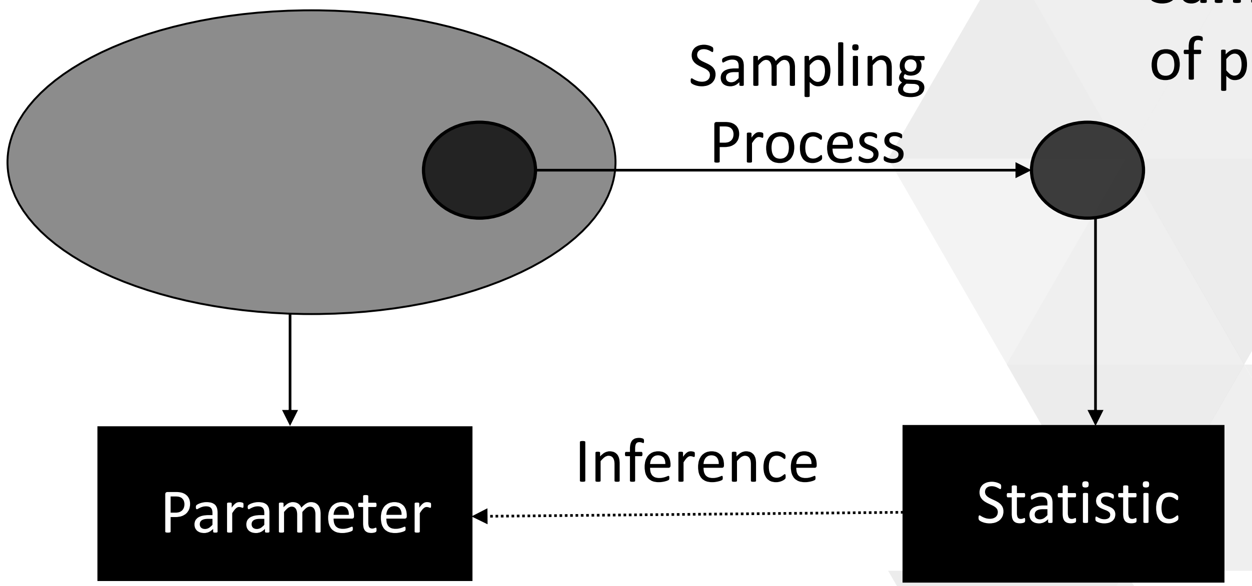
Nominal | *Ordinal* | *Interval* | *Ratio*

Qualitative Data | *Quantitative Data*

Basic Statistical Terms

Population: Complete collection to be studied

Sample: Part of population



Characteristic of a population

Characteristic of a sample

N	number of members	n
μ	mean	$\bar{x}$
σ	standard deviation	s

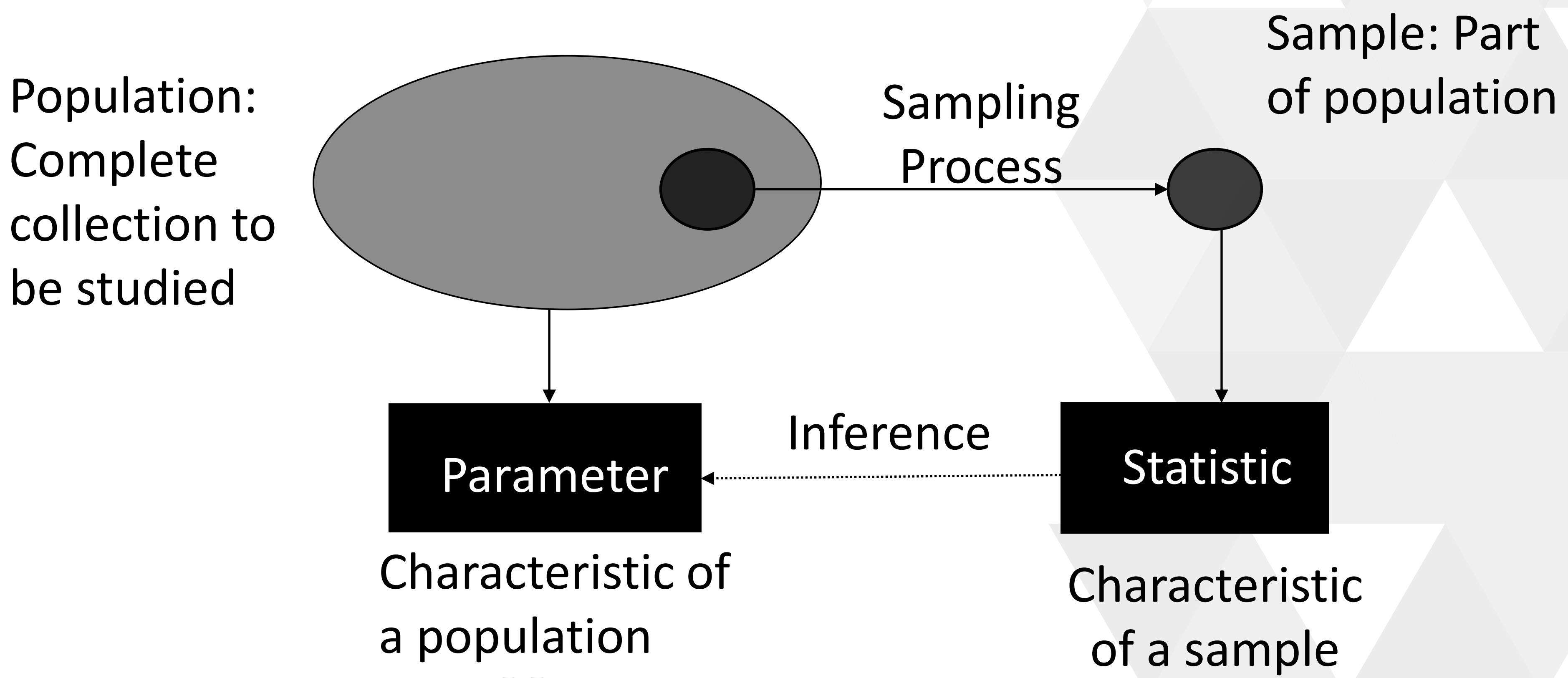
QUALITYGURUS

Notations

	Population Parameters	Sample Statistics
Mean	μ	$\bar{x}$
Standard Deviation	σ	s
Variance	σ^2	s^2
Proportion of population having an attribute	P	p
Proportion of population not having an attribute	Q (=1-P)	q (=1-p)
Correlation coefficient	ρ	r
Number of elements	N	n

QUALITYGURUS

Basic Statistical Terms



N	number of members	n
μ	mean	$\bar{x}$
σ	standard deviation	s

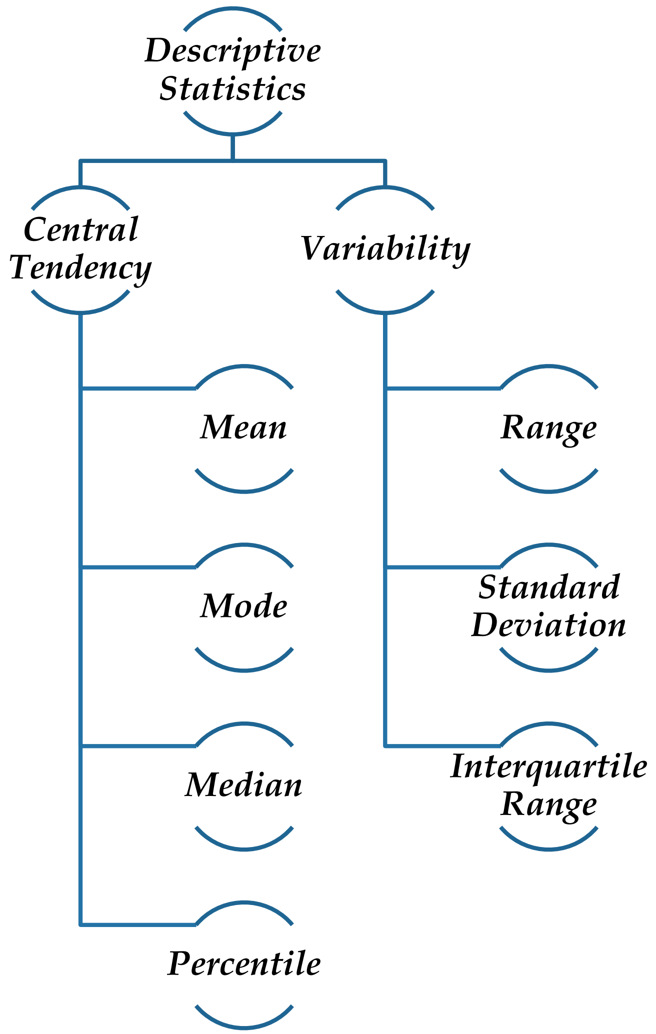
QUALiTYGURUS

Notations

	Population Parameters	Sample Statistics
Mean	μ	$\bar{x}$
Standard Deviation	σ	s
Variance	σ^2	s^2
Proportion of population having an attribute	P	p
Proportion of population not having an attribute	Q (=1-P)	q (=1-p)
Correlation coefficient	ρ	r
Number of elements	N	n

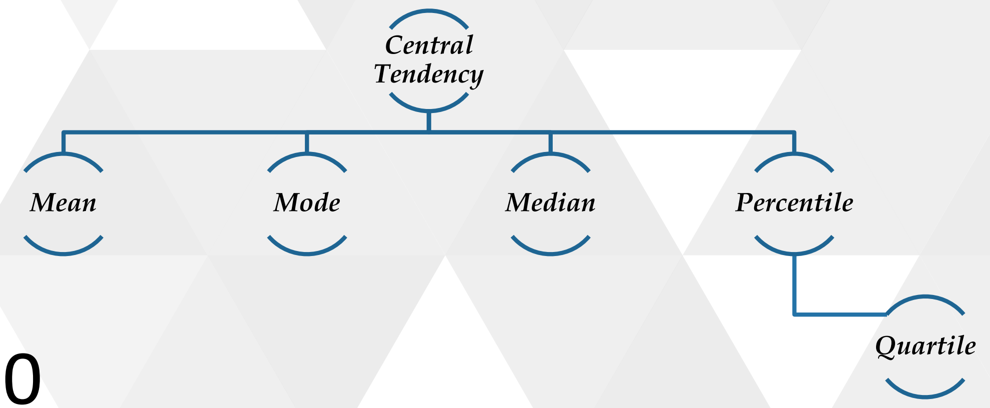
QUALiTYGURUS

Descriptive Statistics



Mean

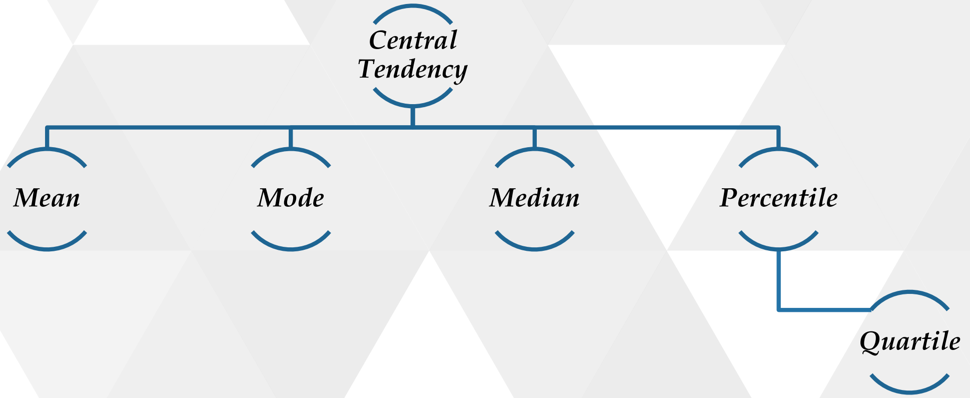
- Also known as Average
- Affected by extreme values
- Example: 10, 11, 14, 9, 6
- Mean = (10+11+14+9+6)/5 = 50/5 = 10



QUALiTYGURUS

Mode

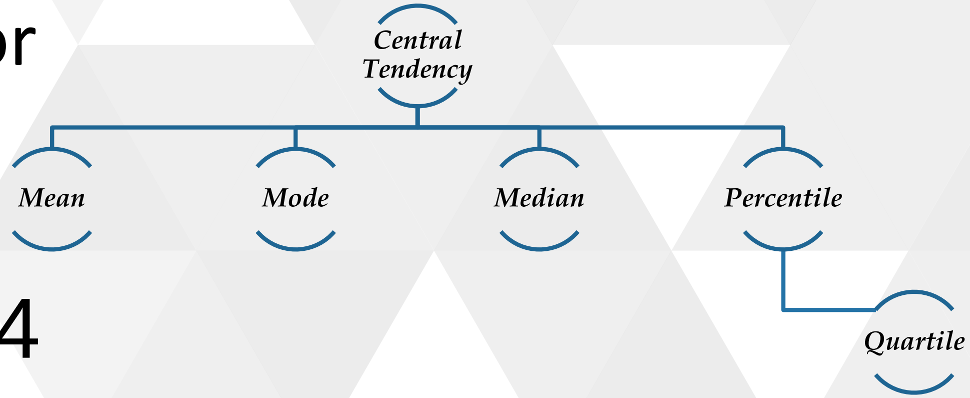
- Most occurring item
- Example: 10, 11, 14, 9, 6, 10
- Mode = 10



Median

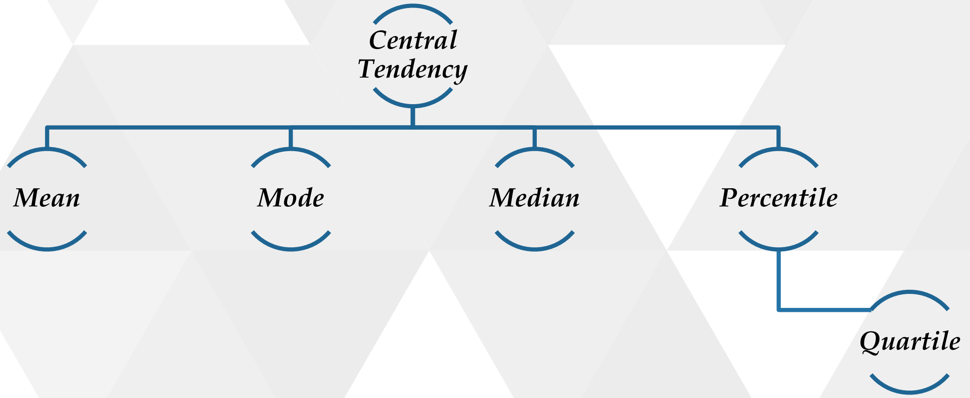
- Middle value when put in ascending or descending order.
- Example: 10, 11, 14, 9, 6
- In ascending order - 6,9,10,11,14
- Median = 10

- Example: 10, 11, 14, 9, 6, 11
- In order - 6,9,10,11, 11,14
- Median = 10.5



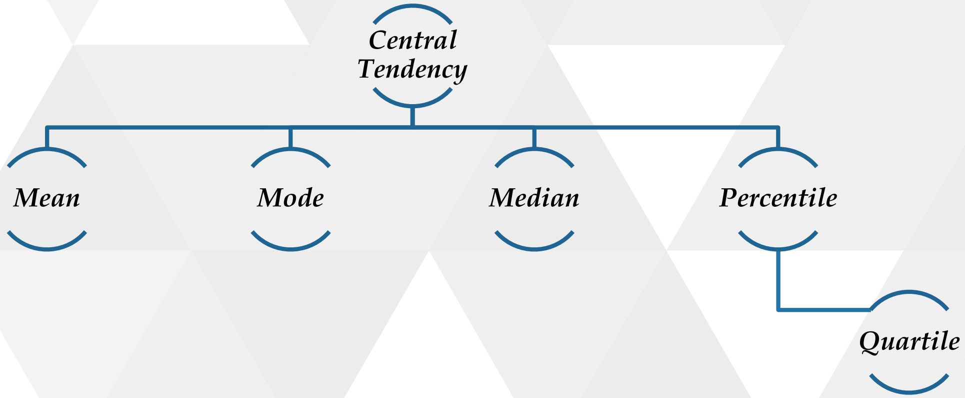
Percentile

- Median divides the data in two equal parts when arranged in ascending or descending order
- Percentile divides data in 99 parts
- Quartile divides data in 4 parts
- Example: 6,9,10,11, 11,14
- Q1=9, Q2=10.5, Q3=11



Percentile/Quartile Steps

- Arrange in ascending or descending order
- Calculate location(i) = P.(n)/100
- P=percentile, n=numbers in data set
- If i is whole number – Percentile is average of (i)th and (i+1)th location
- If i is “not” a whole number – Percentile is located at (i+1)th whole-num.
- Example: 6,9,10,11, 11,14
- Q1=9, Q2=10.5, Q3=11



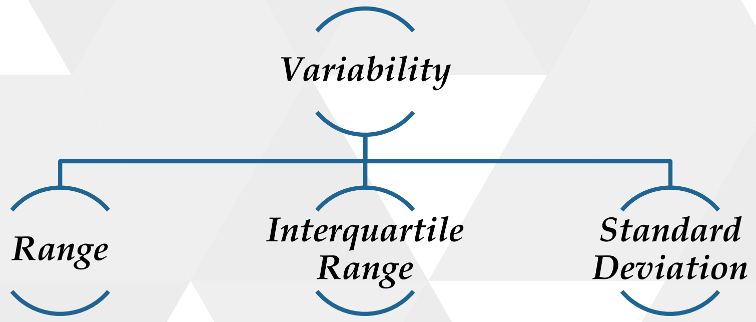
Descriptive Statistics

Variability

- *Range*
- *Interquartile Range*
- *Standard Deviation*

Range

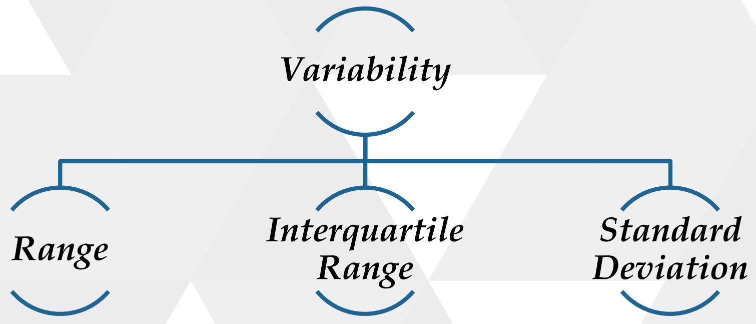
- Difference between lowest and the highest value.
- Example: 6,9,10,11, 11,14
- Range = 14-6 = 8



QUALiTYGURUS

Interquartile Range

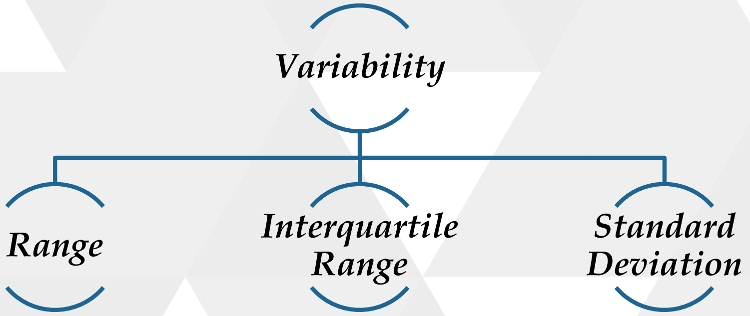
- Range of middle 50% data
- IQR = Q3-Q1
- Example: 6,9,10,11, 11,14
- Q1=9, Q2=10.5, Q3=11
- IQR = 11-9 = 2
- Box-and-Whisker Plot



Standard Deviation

- Variance = average of squared deviation about the arithmetic mean.
- Square root of variance is standard deviation

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



Standard Deviation

x	$x-\bar{x}$	$(x-\bar{x})^2$
100	0	0
101	1	1
99	-1	1
102	2	4
98	-2	4
100	0	0
$\bar{x}=100$	$\sum(x-\bar{x})=0$	$\sum(x-\bar{x})^2=10$

$$S^2 = \frac{\sum(x-\bar{x})^2}{n-1}$$

$$S^2 = 10/5 = 2$$

$$S = \sqrt{2} = 1.414$$

QUALITYGURUS

Skewness

- **Negative Skew or Left Skew**
  - Left tail is longer
- **Positive Skew or Right Skew**
  - Right tail is longer
- **Value of Skewness (Rule of Thumb)**
  - Skew = 0 means perfect symmetric
  - Skew between 0 and +/- 0.5 means approximately symmetric
  - Skew between +/- 0.5 and 1.0 means moderately skewed
  - Skew more than +1 or less than -1 means highly skewed

Kurtosis

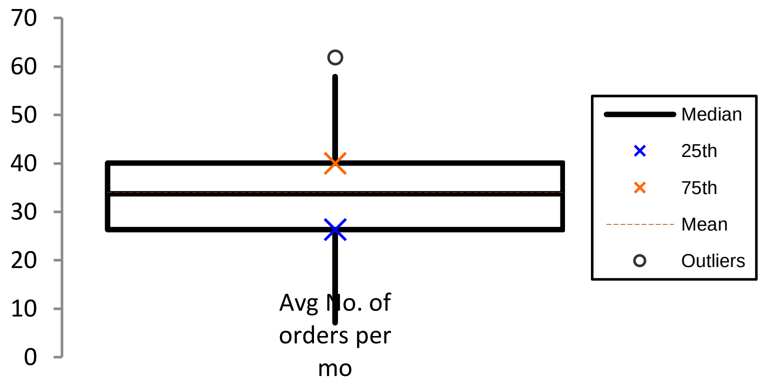
- Normal distribution has Kurtosis = 0
- Kurtosis < 0 means peak is short and broad, tails are shorter
- Kurtosis > 0 means peak is higher and thinner, tails are longer

Graphical Methods

- Box-and-whisker plots
- Scatter diagrams
- Histograms

Box and Whisker Plots

- Also known as Box Plot
- Shows the median
- Shows Q1, Q3 and IQR



QUALITYGURUS

Scatter Diagram

- One of seven basic quality tools
- To see relationship between two variables
- Relationship should make practical sense
- Temperature(X) vs Ice cream sale (Y)
- Some times relationship between two variables is because of a third variable. (ice cream sale vs heat stroke cases)
- Correlation/Regression is covered in the Analyze Phase

Histogram

- Graphical representation of the distribution of numerical data
- Values are assigned “bins” and frequency for each bin is plotted.

Histograms

Scatter Diagram

- One of seven basic quality tools
- To see relationship between two variables
- Relationship should make practical sense
- Temperature(X) vs Ice cream sale (Y)
- Some times relationship between two variables is because of a third variable. (ice cream sale vs heat stroke cases)
- Correlation/Regression is covered in the Analyze Phase

Probability

- Classic Model

$$\frac{\text{Number of outcomes in which the event occurs}}{\text{Total Number of possible outcomes of an experiment}}$$

Probability

❖ Relative Frequency of Occurrence

$$\frac{\text{Number of times an event occurred}}{\text{Total number of opportunities for an event to occur}}$$

Probability

- Experiment/Trial: Some thing done with an expectation of result.

- Event or Outcome: Result of experiment

- Sample Space: A sample space of an experiment is the set of all possible results of that random experiment.

  {1, 2, 3, 4, 5, 6}

Probability

- Union: Probability that events A <u>or</u> B occur: $P(A \cup B)$

- Intersection: Probability that events A <u>and</u> B occur: $P(A \cap B)$

Probability

- Mutually Exclusive Events: When two events cannot occur at the same time

- Independent Events: The occurrence of Event A does not change the probability of Event B

- Complementary Events: The probability that Event A will NOT occur is denoted by P(A').

Probability

- Rule of Addition

The probability that Event A **<u>or</u>** Event B occurs

=

Probability that Event A occurs

+

Probability that Event B occurs

-

Probability that both Events A and B occur

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

QUALITYGURUS

Probability

❖ Rule of Multiplication:

The probability that Events A **and** B both occur

=

Probability that Event A occurs

x

Probability that Event B occurs, given that A has occurred

$$P(A \cap B) = P(A)\,P(B|A)$$

Probability

❖ Independent Events

Probability

- Dependent Events

Factorial

- **Factorial** of a non-negative integer n, denoted by n!, is the product of all positive integers less than or equal to n

Permutation/ Combination

- Permutation: A set of objects in which position (or order) is important.
  - e.g. Lock combination: 3376

- Combination: A set of objects in which position (or order) is NOT important.
  - e.g. Selecting 2 students out of 5

Central Limit Theorem

Central Limit Theorem

- For almost all populations, the sampling distribution of the mean can be approximated closely by a normal distribution, provided the sample size is sufficiently large.

Central Limit Theorem

Central Limit Theorem

Normal Probability Distribution

- Symmetrically distributed
- Long Tails / Bell Shaped
- Mean/ Mode and Median are same

Normal Probability Distribution

- Two factors define the shape of the curve:
  - Mean
  - Standard Deviation

Normal Probability Distribution

- About 68% of the area under the curve falls within **<u>1 standard deviation</u>** of the mean.
- About 95% of the area under the curve falls within **<u>2 standard deviations</u>** of the mean.
- About 99.7% of the area under the curve falls within **<u>3 standard deviations</u>** of the mean.

Normal Probability Distribution

- The total area under the normal curve = 1.
- The probability of any particular value is 0.
- The probability that X is greater than or less than a value = area under the normal curve in that direction

Normal Probability Distribution

- The value of the random variable Y is:

$$Y = \{ 1/[\sigma * \mathrm{sqrt}(2\pi)] \} * e^{-(x - \mu)2/2\sigma 2}$$

- where X is a normal random variable,
- μ = mean,
- σ = standard deviation,
- π is approximately 3.14159,
- e is approximately 2.71828.

QUALITYGURUS

Normal Probability Distribution

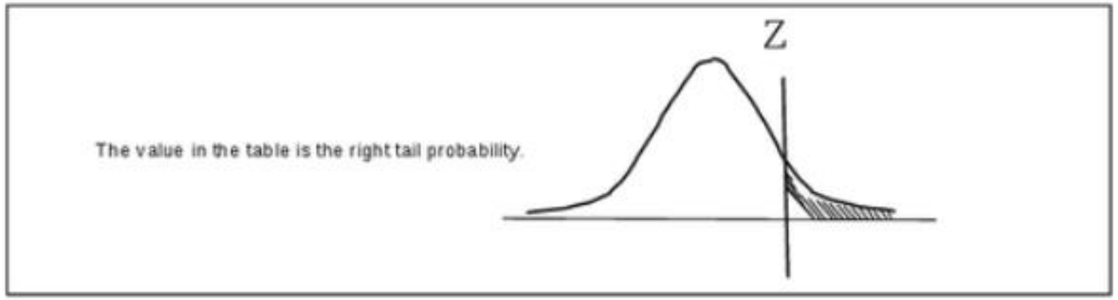
- Z Value / Standard Score

- How many standard deviations an element is from the mean.
- $z = (X - \mu) / \sigma$

- z is the z-score,
- X is the value of the element,
- μ is the population mean,
- σ is the standard deviation.

QUALITYGURUS

Z Table



Hundredth place for Z-value →

Z-Value	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.49601	0.49202	0.48803	0.48405	0.48006	0.47608	0.47210	0.46812	0.46414
0.1	0.46017	0.45620	0.45224	0.44828	0.44433	0.44038	0.43644	0.43251	0.42858	0.42465
0.2	0.42074	0.41683	0.41294	0.40905	0.40517	0.40129	0.39743	0.39358	0.38974	0.38591
0.3	0.38209	0.37828	0.37448	0.37070	0.36693	0.36317	0.35942	0.35569	0.35197	0.34827
0.4	0.34458	0.34090	0.33724	0.33360	0.32997	0.32636	0.32276	0.31918	0.31561	0.31207
0.5	0.30854	0.30503	0.30153	0.29806	0.29460	0.29116	0.28774	0.28434	0.28096	0.27760
0.6	0.27425	0.27093	0.26763	0.26435	0.26109	0.25785	0.25463	0.25143	0.24825	0.24510
0.7	0.24196	0.23885	0.23576	0.23270	0.22965	0.22663	0.22363	0.22065	0.21770	0.21476
0.8	0.21186	0.20897	0.20611	0.20327	0.20045	0.19766	0.19489	0.19215	0.18943	0.18673
0.9	0.18406	0.18141	0.17879	0.17619	0.17361	0.17106	0.16853	0.16602	0.16354	0.16109
1.0	0.15866	0.15625	0.15386	0.15151	0.14917	0.14686	0.14457	0.14231	0.14007	0.13786
1.1	0.13567	0.13350	0.13136	0.12924	0.12714	0.12507	0.12302	0.12100	0.11900	0.11702
1.2	0.11507	0.11314	0.11123	0.10935	0.10749	0.10565	0.10383	0.10204	0.10027	0.09853
1.3	0.09680	0.09510	0.09342	0.09176	0.09012	0.08851	0.08691	0.08534	0.08379	0.08226
1.4	0.08076	0.07927	0.07780	0.07636	0.07493	0.07353	0.07215	0.07078	0.06944	0.06811
1.5	0.06681	0.06552	0.06426	0.06301	0.06178	0.06057	0.05938	0.05821	0.05705	0.05592
1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551

Continuous Probability Distributions

- **Normal probability distribution**

- Student's t distribution
- Chi-square distribution
- F distribution

Continuous vs Discrete Variable

- If a variable can take on any value between two specified values, it is called a **continuous variable**; otherwise, it is called a **discrete variable**.

Discrete Probability Distributions

- **Binomial Probability Distribution**

- Bernoulli Distribution
- Hypergeometric Probability Distribution
- Geometric Distribution
- Negative Geometric Distribution

- **Poisson Probability Distribution**

Binomial Probability Distribution

- **A binomial experiment** has the following properties:
  - The experiment consists of n repeated trials.
  - Each trial can result in just two possible outcomes. We call one of these outcomes a success and the other, a failure.
  - The probability of success, denoted by p, is the same on every trial.
  - The trials are independent; that is, the outcome on one trial does not affect the outcome on other trials.

QUALITYGURUS

Binomial Probability Distribution

- **A binomial experiment** has the following properties:
  - The experiment consists of *n* repeated trials.
  - Each trial can result in just two possible outcomes. We call one of these outcomes a success and the other, a failure.
  - The probability of success, denoted by *p*, is the same on every trial.
  - The trials are independent; that is, the outcome on one trial does not affect the outcome on other trials.

QUALiTYGURUS

Binomial Probability Distribution

$$P(x) = {}_nC_x \cdot p^x \cdot (1 - p)^{n - x}$$

- ***x***: The number of successes that result from the binomial experiment.
- ***n***: The number of trials in the binomial experiment.
- ***p***: The probability of success on an individual trial.
- ***q***: The probability of failure on an individual trial. (This is equal to 1 - *p*.)
- ***n!***: The factorial of n (also known as n factorial).
- **P(x)** : Binomial probability - the probability that an *n*-trial binomial experiment results in exactly *x* successes, when the probability of success on an individual trial is *p*.
- **$_nC_x$**: The number of combinations of *n* things, taken *x* at a time.

QUALiTYGURUS

Binomial Probability Distribution

- The **binomial probability** refers to the probability that a binomial experiment results in exactly *x* successes.
- Suppose a binomial experiment consists of *n* trials and results in *x* successes. If the probability of success on an individual trial is *p*, then the binomial probability is:
- $P(x) = {}_nC_x \cdot p^x \cdot (1 - p)^{n - x}$
  or
  $P(x) = \{ n! / [x! (n - x)!] \} \cdot p^x \cdot (1 - p)^{n - x}$

QUALiTYGURUS

Binomial Probability Distribution

- The mean of the distribution (μ_x) is

  n . p

- The variance (σ^2_x) is

  n . p . (1 - p)

- The standard deviation (σ_x) is

  sqrt[*n . p . (1 - p)*]

n: The number of trials in the binomial experiment.

p: The probability of success on an individual trial.

QUALiTYGURUS

Five Conditions - Binomial

- 1. There is a fixed number, n , of identical trials.
- 2. For each trial, there are only two possible outcomes (success/failure).
- 3. The probability of success, p, remains the same for each trial.
- 4. The trials are independent of each other.
- 5. x = the number of successes observed for the n trials.

$$P(x) = {}_nC_x \cdot p^x \cdot (1 - p)^{n - x}$$

QUALITYGURUS

Bernoulli Distribution

- Distribution of successes on a single trial.
  - What is the probability of getting head in tossing of a coin once?

QUALITYGURUS

Hypergeometric Distribution

- There is a fixed number, n , of identical trials.
- For each trial, there are only two possible outcomes (success/failure).
- ~~The probability of success, p, remains the same for each trial.~~
- ~~The trials are independent of each other.~~
- **<u>Finite and known population without replacement.</u>**
- **<u>Number of successes in population are known</u>**
- x = the number of successes observed for the n trials.

$$P(x) = {}_nC_x \cdot p^x \cdot (1 - p)^{n - x}$$

Hypergeometric Distribution

- **N**: size of population
- **A**: number of successes in population
- ***x***: The number of successes that result from the experiment.
- ***n***: The number of trials without replacement.
- ~~***p***: The probability of success on an individual trial.~~
- ~~***q***: The probability of failure on an individual trial. (This is equal to 1 - *p*.)~~
- **P(x)** : The probability that an *n*-trial experiment results in exactly *x* successes
- $_n C_x$: The number of combinations of *n* things, taken *x* at a time.

$$P(x) = {}_A C_x \cdot {}_{N-A} C_{n-x} / {}_N C_n$$

QUALITYGURUS

Hypergeometric Distribution

$P(x) = {}_{A}C_{x} \cdot {}_{N\text{-}A}C_{x} / {}_{N}C_{n}$

- Out of 10 people (6M, 4F), 3 people are selected without replacement. What is the probability that two of them are females?
- $P(2) = {}_{4}C_{2} \cdot {}_{10\text{-}4}C_{2} \,/\, {}_{10}C_{3}$
- $= {}_{4}C_{2} \cdot {}_{6}C_{2} \,/\, {}_{10}C_{3}$

- When sample size is less than 5% population then can use Binomial.

QUALiTYGURUS

Geometric Distribution

- Number of trials needed to get the first success.
  - What is the probability that if the coin is tossed repeatedly the first head appears on 5^{th} trial?

QUALITYGURUS

Negative Binomial Distribution

- Generalization of the Geometric distribution
- Number of trials needed to get the ~~first~~ number of successes.
  - What is the probability that if the coin is tossed repeatedly the ~~first~~ third time head appears on 5^{th} trial?
- In Binomial distribution trials are fixed, in Negative Binomial number of successes are fixed.

QUALITYGURUS

Poisson Distribution

- A **Poisson experiment** has the following properties:
- The experiment results in outcomes that can be classified as successes or failures.
- The average number of successes (μ) that occurs in a specified region is known.
- Outcomes are random. Occurrence of one outcome does not influence the chance of another outcome of interest.
- The outcomes of interest are rare relative to the possible outcomes.

Poisson Distribution

- *e*: A constant equal to approximately 2.71828. (Actually, *e* is the base of the natural logarithm system)
- μ: The mean number of successes that occur in a specified region.
- *x*: The actual number of successes that occur in a specified region.
- P(*x*; μ): The **Poisson probability** that exactly *x* successes occur in a Poisson experiment, when the mean number of successes is μ.

$$P(x; \mu) = (e^{-\mu})\,(\mu^{x}) / x!$$

QUALITYGURUS

Poisson Distribution

- **Poisson Formula.** Suppose we conduct a Poisson experiment, in which the average number of successes within a given region is μ. Then, the Poisson probability is:

- $P(x; \mu) = (e^{-\mu}) (\mu^{x}) / x!$

- where *x* is the actual number of successes that result from the experiment, and *e* is approximately equal to 2.71828.

QUALITYGURUS

Poisson Distribution

- The Poisson distribution has the following properties:
- The mean of the distribution is equal to μ .
- The variance is also equal to μ .

Poisson Distribution

- On a booking counter on the average 3.6 people come every 10 minute on weekends. What is the probability of getting 7 people in 10 minutes?
- $\mu = 3.6$, $x=7$
- $P(x; \mu) = (e^{-\mu})\,(\mu^{x}) / x! = (e^{-3.6})\,(3.6^{7}) / 7!$
- $=0.02732 \times 7836.41 / 5040 = 0.0424$

QUALITYGURUS

Errors of Statistical Tests

		True State of Nature	
		H_0 Is true	H_a Is true
Conclusion	Support H_0 / Reject H_a	Correct Conclusion	Type II Error
	Support H_a / Reject H_0	Type I Error	Correct Conclusion (Power)

QUALITYGURUS

Errors of Statistical Tests

	Type I error (alpha)	Type II error (beta)
Name	Producer's risk/ Significance level	Consumer's risk
1 minus error is called	Confidence level	Power of the test
Example of Fire Alarm	False fire alarm leading to inconvenience	Missed fire leading to disaster
Effects on process	Unnecessary cost increase due to frequent changes	Defects may be produced
Control method	Usually fixed at a pre-determined level, 1%, 5% or 10%	Usually controlled to < 10% by appropriate sample size
Simple definition	Innocent declared as guilty	Guilty declared as innocent

Significance Level

Level of Confidence / Confidence Interval:

C = 0.90, 0.95, 0.99 (90%, 95%, 99%)

Level of Significance:

$\alpha = 1 - C$ (0.10, 0.05, 0.01)

QUALITYGURUS

Power

- Power = 1 – β (or 1 - type II error)
- Type II Error: Failing to reject null hypothesis when null hypothesis is false.
- Power: Likelihood of rejecting null hypothesis when null hypothesis is false.

- Or: Power is the ability of a test to correctly reject the null hypothesis.

Alpha vs Beta

- Researcher can not commit both Type I and II error. Only one can be committed.
- As the value of α increases (say 0.01 to 0.05) β goes down and the Power of test increases.
- To reduce both Type I and II errors increase sample size.

Hypothesis Testing

1. State the Alternate Hypothesis.
2. State the Null Hypothesis.
3. Select a probability of error level (alpha level). Generally 0.05
4. Select and compute the test statistic (e.g t or z score)
5. Critical test statistic
6. Interpret the results.

Hypothesis Testing

- Lower Tail Tests
  - H_0: $\mu \geq 150cc$
  - H_a: $\mu < 150cc$
- Upper Tail Tests
  - H_0: $\mu \leq 150cc$
  - H_a: $\mu > 150cc$

Hypothesis Testing

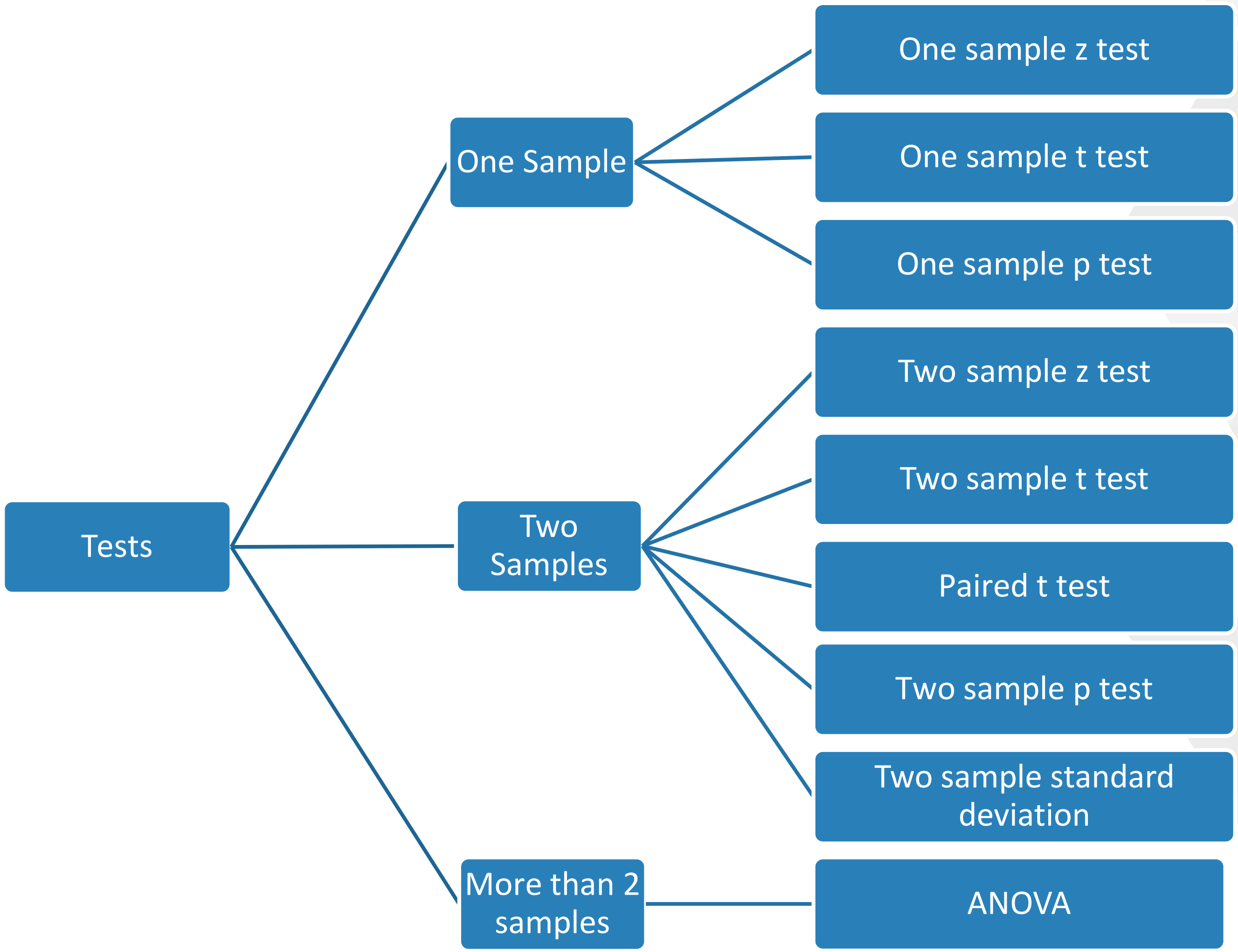
- Two Tail Tests
  - $H_0: \mu = 150cc$
  - $H_a: \mu \neq 150cc$

QUALiTYGURUS

Calculate Test Statistic

- Single sample
  - $z = (x - \mu) / \sigma$
- Mean of Multiple samples
  - $z = (\bar{x} - \mu) / (\sigma / \sqrt{n})$

Tests for Mean, Variance & Proportion



One Sample z Test

- Calculated value
- $z = [\bar{x} - \mu] / [\sigma / \mathrm{sqrt}(n)]$
- Example: Perfume bottle producing 150cc with sd of 2 cc, 100 bottles are randomly picked and the average volume was found to be 152cc. Has mean volume changed? (95% confidence)
- $z_{calculated} = (152-150)/[2 / \mathrm{sqrt}(100)] = 2/0.2 = 10$
- $z_{critical} = ?$

One Sample z Test

$z_{critical} = 1.96$

z	0	1	2	3	4	5	6	7	8	9
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
3.7	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
3.8	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
3.9	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

One Sample z Test

- Calculated value
- $z = [\bar{x} - \mu] / [\sigma / \text{sqrt}(n)]$
- Example: Perfume bottle producing 150cc with sd of 2 cc, 100 bottles are randomly picked and the average volume was found to be 152cc. Has mean volume changed? (95% confidence)
- $z_{calculated} = (152-150)/[2 / \text{sqrt}(100)] = 2/0.2 = 10$
- $z_{critical} = 1.96$ > Reject Ho

QUALiTYGURUS

p Value

- p value is the lowest value of alpha for which the null hypothesis can be rejected. (Probability that the null hypothesis is correct)
- If $p = 0.01$ you can reject the null hypothesis at $\alpha = 0.05$
- p is low the null must go / p is high the null fly.

One Sample t Test

- Calculated value
- $t = [\bar{x} - \mu] / [s / \mathrm{sqrt}(n)]$
- Example: Perfume bottle producing 150cc, 4 bottles are randomly picked and the average volume was found to be 151cc and sd of sample was 2 cc. Has mean volume changed? (95% confidence)
- $t_{cal} = (151-150)/[2 / \mathrm{sqrt}(4)] = 1/1 = 1$
- $t_{critical} = ?$

QUALITYGURUS

One Sample t Test

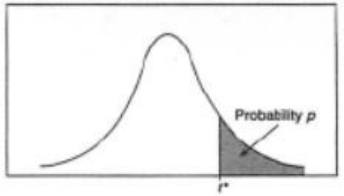


TABLE B *t* Distribution Critical Values

df	TAIL PROBABILITY P											
	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646

$t_{critical}$ = 3.182

QUALITYGURUS

One Sample t Test

- Calculated value
- $t = [\bar{x} - \mu] / [s / \mathrm{sqrt}(n)]$
- Example: Perfume bottle producing 150cc, 4 bottles are randomly picked and the average volume was found to be 151cc and sd of sample was 2 cc. Has mean volume changed? (95% confidence)
- $t_{cal} = (151-150)/[2 / \mathrm{sqrt}(4)] = 1/1 = 1$
- $t_{critical} = 3.182 >$ Fail to reject Ho

QUALITYGURUS

One Sample p Test

- H_0: $p = p_0$
- Calculated value

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

- Example: Smoking rate in a town in past was 21%, 100 samples were picked and found 14 smokers. Has smoking habit changed?

One Sample p Test

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

- Example: Smoking rate in a town in past was 21%, 100 samples were picked and found 14 smokers. Has smoking habit <u>changed</u> at 95% confidence? (two tail)
- $p_0 = 0.21$, $p=0.14$
- $np_0 = 0.21x100 = 21$ and $n(1-p_0)= 0.79x100 = 79$
- >5 means sample size is sufficient.
- z = (0.14-0.21)/sqt (0.21x0.79/100)
- z = -0.07/0.0407 = -1.719
- $z_{critical} = 1.96$

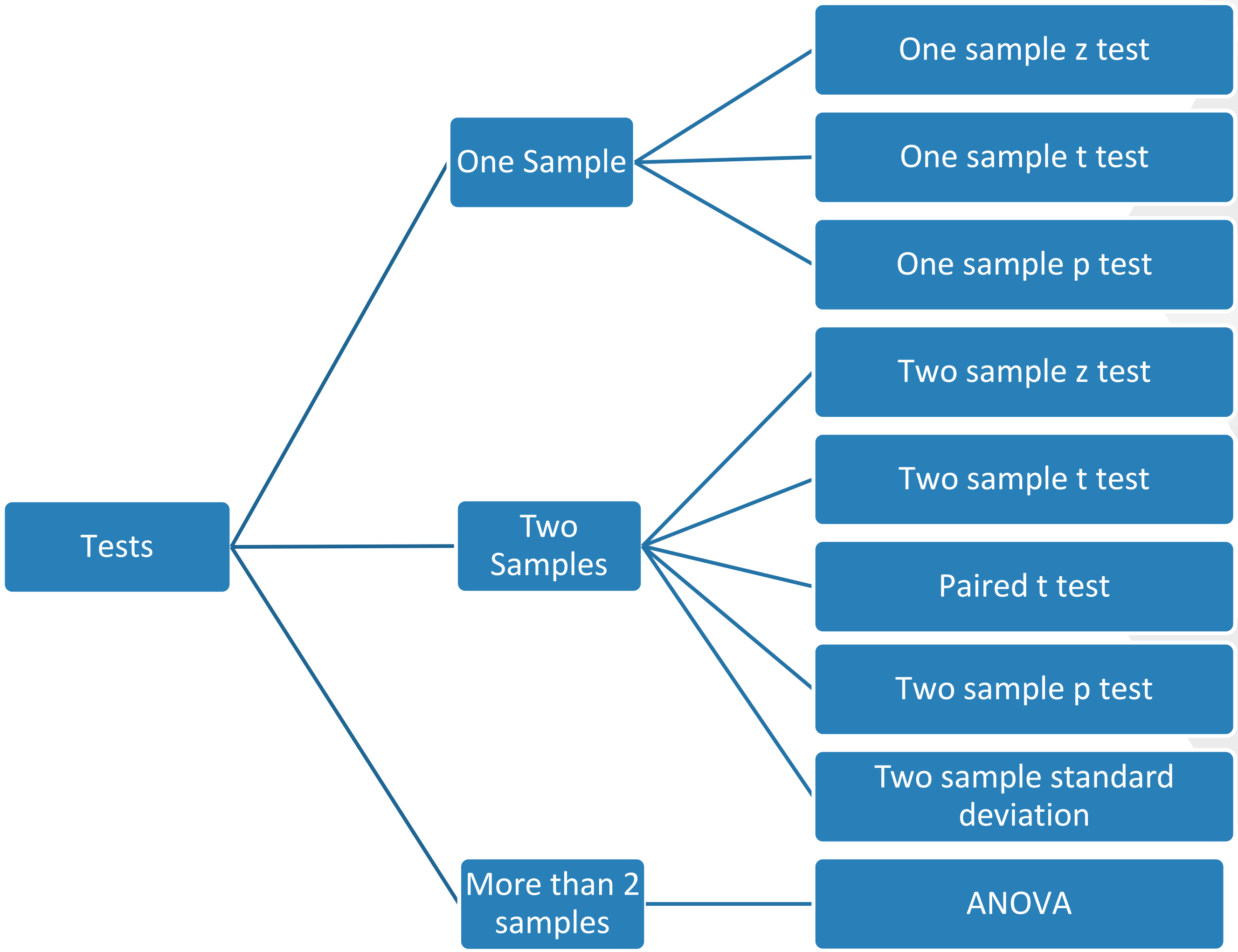
QUALITYGURUS

One Sample p Test

- Example: Smoking rate in a town in past was 21%, 100 samples were picked and found 14 smokers. Has smoking habit <u>reduced</u> at 95% confidence? (one tail)
- H_0: p < p_0
- p_0 = 0.21, p=0.14
- z = (0.14-0.21)/sqt (0.21x0.79/100)
- z = -0.07/0.0407 = -1.719
- z $_{critical}$ = 1.645

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Tests for Mean, Variance & Proportion



QUALiTYGURUS

Two Sample z Test

- **Null hypothesis**: H_0: $\mu_1 = \mu_2$
- or H_0: $\mu_1 - \mu_2 = 0$
- **Alternative hypothesis**: H_a : $\mu_1 \neq \mu_2$

$$z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Two Sample z Test

- Example: From two machines 100 samples each were drawn.
  - Machine 1: Mean = 151.2 / sd = 2.1
  - Machine 2: Mean = 151.9 / sd = 2.2
  - Is there difference in these two machines. Check at 95% confidence level.

$$z = \frac{(\overline{x_1} - \overline{x_2})}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Two Sample z Test

- Example: From two machines 100 samples each were drawn.
  - Machine 1: Mean = 151.2 / sd = 2.1
  - Machine 2: Mean = 151.9 / sd = 2.2
  - Is there difference in these two machines. Check at 95% confidence level.
- Z_{cal} = -0.7 / 0.304 = -2.30
- $Z_{critical}$ = 1.96
- Reject Null.
- There is a difference.

$$z = \frac{(\overline{x}_1 - \overline{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

QUALITYGURUS

Two Sample z Test

- Example: From two machines 100 samples each were drawn.
  - Machine 1: Mean = 151.9 / sd = 2.1
  - Machine 2: Mean = 151.2 / sd = 2.2
  - Is there **difference of more than 0.2 cc** in these two machines. Check at 95% confidence level.

$$z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- $H_0: \mu_1 - \mu_2 = 0.2$

QUALITYGURUS

Two Sample z Test

- Example: From two machines 100 samples each were drawn.
  - Machine 1: Mean = 151.9 / sd = 2.1
  - Machine 2: Mean = 151.2 / sd = 2.2
  - Is there **difference of more than 0.2 cc** in these two machines. Check at 95% confidence level.
- Z_{cal} = 0.5/0.304 = 2.30
- $Z_{critical}$ = 1.64
- Reject Null Hypothesis.

$$z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

QUALITYGURUS

Two Sample t Test

- If two set of data are independent or dependent.
  - If the values in one sample reveal no information about those of the other sample, then the samples are independent.
    - Example: Blood pressure of male/female
  - If the values in one sample affect the values in the other sample, then the samples are dependent.
    - Example: Blood pressure before and after a specific medicine

QUALITYGURUS

Two Sample t Test

- If two set of data are independent or dependent.
  - If the values in one sample reveal no information about those of the other sample, then the samples are independent.
    - Example: Blood pressure of male/female

Two sample t test

  - If the values in one sample affect the values in the other sample, then the samples are dependent.
    - Example: Blood pressure before and after a specific medicine

Paired t test

Two Sample t Test

- Is variance for two samples equal?

  - If yes: Pooled variance calculate Sp for finding out t

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Paired t Test

- Where you have two samples in which observations in one sample can be paired with observations in the other sample.
- Or
- If the values in one sample affect the values in the other sample, (the samples are dependent.)
  - Example: Blood pressure before and after a specific medicine

Paired t Test

- Find the difference between two set of readings as d1, d2 …. dn.
- Find the mean and standard deviation of these differences.

$$t = \frac{\bar{d}}{\sqrt{s^2/n}}$$

Paired t Test

- Example: Before and after medicine BP was measured. Is there a difference at 95% confidence level?

$$t = \frac{\bar{d}}{\sqrt{s^2/n}}$$

Patient	Before	After
1	120	122
2	122	120
3	143	141
4	100	109
5	109	109

Paired t Test

- Example: Before and after medicine BP was measured. Is there a difference at 95% confidence level?

$$t = \frac{\bar{d}}{\sqrt{s^2/n}}$$

Patient	Before	After	difference
1	120	122	2
2	122	120	-2
3	143	141	-2
4	100	109	9
5	109	109	0

- d-bar = 1.4 , s = 4.56 , n=5
- $t_{cal.}$ = 1.4/1.99 = 0.70

Paired t Test

$$t = \frac{\bar{d}}{\sqrt{s^2/n}}$$

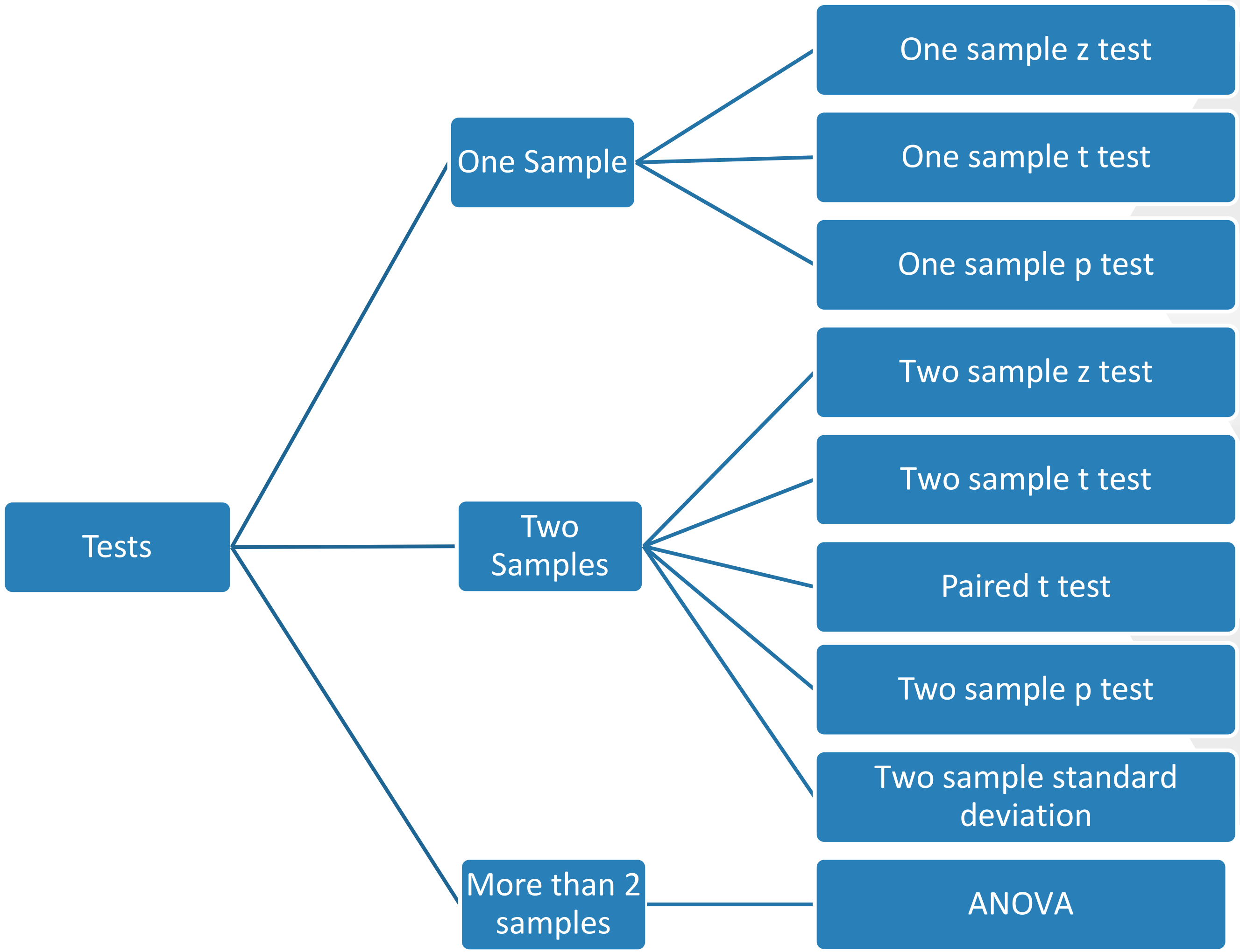
- Example: Before and after medicine BP was measured. Is there a difference at 95% confidence level?

Patient	Before	After	difference
1	120	122	2
2	122	120	-2
3	143	141	-2
4	100	109	9
5	109	109	0

- $t_{cal.} = 1.4/1.99 = 0.70$
- $t_{0.025,\,4} = 2.766$
- Fail to reject null hypothesis

QUALITYGURUS

Tests for Mean, Variance & Proportion



$$Z = \frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1 - p_o)}{n}}}$$

Two Sample p Test

- **Null hypothesis**: H_0: $p_1 = p_2$
- or H_0: $p_1 - p_2 = 0$
- **Alternative hypothesis**: H_a : $p_1 \neq p_2$

- **Normal approximation – Pooled**

$$\frac{(\bar{p}_1 - \bar{p}_2)}{\sqrt{\bar{p}(1 - \bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

- **Normal approximation – Un-pooled**

$$Z = \frac{(p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

QUALITYGURUS

Tests for Variance

- F-test
  - for testing equality of *two* variances from different populations
  - for testing equality of several means with technique of ANOVA.
- Chi-square test
  - For testing the population variance against a specified value
  - testing goodness of fit of some probability distribution

Two Sample Variance – F Test

- F-test
  - for testing equality of *two* variances from different populations

  - H_0: $\sigma^2_1 = \sigma^2_2$
  - F $_{\text{calculated}}$

$$F_{\text{calculated}} = \frac{S_1^2}{S_2^2}$$

  - Keep higher value at the top for right tail test.
  - Remember: Variance is square of standard deviation

Two Sample Variance – F Test

- $F_{critical}$
  - Use table with appropriate degrees of freedom
  - For two tail test use the table for $\alpha/2$

F - Distribution (α = 0.05 in the Right Tail)

Numerator Degrees of Freedom

df_2 \ df_1	1	2	3	4	
1	161.45	199.50	215.71	224.58	230
2	18.513	19.000	19.164	19.247	19
3	10.128	9.5521	9.2766	9.1172	9
4	7.7086	9.9443	6.5914	6.3882	6
5	6.6079	5.7861	5.4095	5.1922	5
6	5.9874	5.1433	4.7571	4.5337	4
7	5.5914	4.7374	4.3468	4.1203	3
8	5.3177	4.4590	4.0662	3.8379	3
9	5.1174	4.2565	3.8625	3.6331	3
10	4.9646	4.1028	3.7083	3.4780	3
11	4.8443	3.9823	3.5874	3.3567	3
12	4.7472	3.8853	3.4903	3.2592	3
13	4.6672	3.8056	3.4105	3.1791	3
14	4.6001	3.7389	3.3439	3.1122	2
15	4.5431	3.6823	3.2874	3.0556	2
16	4.4940	3.6337	3.2389	3.0069	2
17	4.4513	3.5915	3.1968	2.9647	2
18	4.4139	3.5546	3.1599	2.9277	2
19	4.3807	3.5219	3.1274	2.8951	2
20	4.3512	3.4928	3.0984	2.8661	2
21	4.3248	3.4668	3.0725	2.8401	2
22	4.3009	3.4434	3.0491	2.8167	2
23	4.2793	3.4221	3.0280	2.7955	2
24	4.2597	3.4028	3.0088	2.7763	2
25	4.2417	3.3852	2.9912	2.7587	2
26	4.2252	3.3690	2.9752	2.7426	2
27	4.2100	3.3541	2.9604	2.7278	2
28	4.1960	3.3404	2.9467	2.7141	2
29	4.1830	3.3277	2.9340	2.7014	2
30	4.1709	3.3158	2.9223	2.6896	2
40	4.0847	3.2317	2.8387	2.6060	2
60	4.0012	3.1504	2.7581	2.5252	2
120	3.9201	3.0718	2.6802	2.4472	2
∞	3.8415	2.9957	2.6049	2.3719	2

Denominator Degrees of Freedom

F - Distribution (α = 0.01 in the Right Tail)

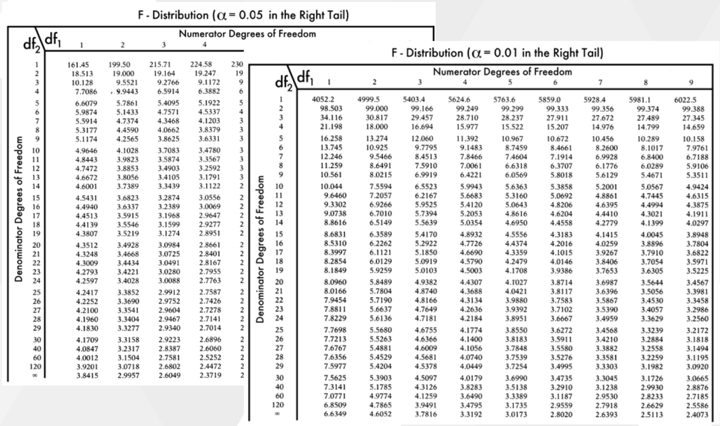
Numerator Degrees of Freedom

df_2 \ df_1	1	2	3	4	5	6	7	8	9
1	4052.2	4999.5	5403.4	5624.6	5763.6	5859.0	5928.4	5981.1	6022.5
2	98.503	99.000	99.166	99.249	99.299	99.333	99.356	99.374	99.388
3	34.116	30.817	29.457	28.710	28.237	27.911	27.672	27.489	27.345
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158
6	13.745	10.925	9.7795	9.1483	8.7459	8.4661	8.2600	8.1017	7.9761
7	12.246	9.5466	8.4513	7.8466	7.4604	7.1914	6.9928	6.8400	6.7188
8	11.259	8.6491	7.5910	7.0061	6.6318	6.3707	6.1776	6.0289	5.9106
9	10.561	8.0215	6.9919	6.4221	6.0569	5.8018	5.6129	5.4671	5.3511
10	10.044	7.5594	6.5523	5.9943	5.6363	5.3858	5.2001	5.0567	4.9424
11	9.6460	7.2057	6.2167	5.6683	5.3160	5.0692	4.8861	4.7445	4.6315
12	9.3302	6.9266	5.9525	5.4120	5.0643	4.8206	4.6395	4.4994	4.3875
13	9.0738	6.7010	5.7394	5.2053	4.8616	4.6204	4.4410	4.3021	4.1911
14	8.8616	6.5149	5.5639	5.0354	4.6950	4.4558	4.2779	4.1399	4.0297
15	8.6831	6.3589	5.4170	4.8932	4.5556	4.3183	4.1415	4.0045	3.8948
16	8.5310	6.2262	5.2922	4.7726	4.4374	4.2016	4.0259	3.8896	3.7804
17	8.3997	6.1121	5.1850	4.6690	4.3359	4.1015	3.9267	3.7910	3.6822
18	8.2854	6.0129	5.0919	4.5790	4.2479	4.0146	3.8406	3.7054	3.5971
19	8.1849	5.9259	5.0103	4.5003	4.1708	3.9386	3.7653	3.6305	3.5225
20	8.0960	5.8489	4.9382	4.4307	4.1027	3.8714	3.6987	3.5644	3.4567
21	8.0166	5.7804	4.8740	4.3688	4.0421	3.8117	3.6396	3.5056	3.3981
22	7.9454	5.7190	4.8166	4.3134	3.9880	3.7583	3.5867	3.4530	3.3458
23	7.8811	5.6637	4.7649	4.2636	3.9392	3.7102	3.5390	3.4057	3.2986
24	7.8229	5.6136	4.7181	4.2184	3.8951	3.6667	3.4959	3.3629	3.2560
25	7.7698	5.5680	4.6755	4.1774	3.8550	3.6272	3.4568	3.3239	3.2172
26	7.7213	5.5263	4.6366	4.1400	3.8183	3.5911	3.4210	3.2884	3.1818
27	7.6767	5.4881	4.6009	4.1056	3.7848	3.5580	3.3882	3.2558	3.1494
28	7.6356	5.4529	4.5681	4.0740	3.7539	3.5276	3.3581	3.2259	3.1195
29	7.5977	5.4204	4.5378	4.0449	3.7254	3.4995	3.3303	3.1982	3.0920
30	7.5625	5.3903	4.5097	4.0179	3.6990	3.4735	3.3045	3.1726	3.0665
40	7.3141	5.1785	4.3126	3.8283	3.5138	3.2910	3.1238	2.9930	2.8876
60	7.0771	4.9774	4.1259	3.6490	3.3389	3.1187	2.9530	2.8233	2.7185
120	6.8509	4.7865	3.9491	3.4795	3.1735	2.9559	2.7918	2.6629	2.5586
∞	6.6349	4.6052	3.7816	3.3192	3.0173	2.8020	2.6393	2.5113	2.4073

Denominator Degrees of Freedom

Two Sample Variance – F Test

- Example: We took 8 samples from machine A and the standard deviation was 1.1. For machine B we took 5 samples and the variance was 11. Is there a difference in variance at 90% confidence level?
- n1 = 8, $s_1 = 1.1$, $s^2{}_1 = 1.21$, df = 7 (denominator)
- n2 = 5, $s^2{}_2 = 11$, df = 4 (numerator)
- $F_{\text{calculated}} = 11/1.21 = 9.09$ (higher value at top)



QUALiTYGURUS

Two Sample Variance – F Test

F - Distribution (α = 0.01 in the Right Tail)

Denominator Degrees of Freedom df_2 \ df_1	1
1	4052.2
2	98.5
3	34.1
4	21.1
5	16.2
6	13.7
7	12.2
8	11.2
9	10.5
10	10.0
11	9.6
12	9.3
13	9.0
14	8.8
15	8.6
16	8.5
17	8.3
18	8.2
19	8.1
20	8.0
21	8.0
22	7.9
23	7.8
24	7.8
25	7.7
26	7.7
27	7.6
28	7.6
29	7.5
30	7.5
40	7.3
60	7.0
120	6.8
∞	6.6

F - Distribution (α = 0.05 in the Right Tail)

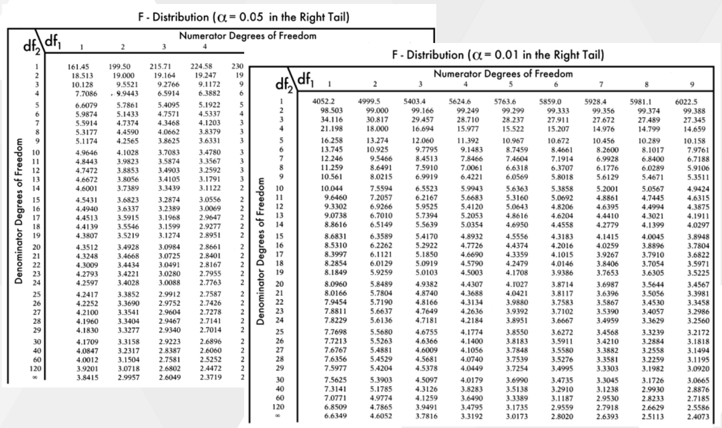
Denominator Degrees of Freedom df_2 \ Numerator Degrees of Freedom df_1	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385
3	10.128	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123
4	7.7086	9.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	6.9988
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588
∞	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799

$$F_{calculated} = \frac{S_1^2}{S_2^2}$$

$$F_{critical} = 4.1203$$

Two Sample Variance – F Test

- Example: We took 8 samples from machine A and the standard deviation was 1.1. For machine B we took 5 samples and the variance was 11. Is there a difference in variance at 90% confidence level?
- $n1 = 8, s_1 = 1.1, s^2_1 = 1.21, df = 7$ (denominator)
- $n2 = 5, s^2_2 = 11, df = 4$ (numerator)
- $F_{calculated} = 11/1.21 = 9.09$ (higher value at top)
- Reject H_0



$$F_{critical} = 4.1203$$

QUALITYGURUS

Tests for Variance

- F-test
  - for testing equality of *two* variances from different populations
  - for testing equality of several means with technique of ANOVA.
- Chi-square test
  - For testing the population variance against a specified value
  - testing goodness of fit of some probability distribution

One Sample Chi Square

- For testing the population variance against a specified value σ

$$\chi^2 = \frac{(n - 1)S^2}{\sigma^2}$$

One Sample Chi Square

- Example: A sample of 25 bottles was selected. The variance of these 25 bottles as 5 cc. Has it **increased** from established 4 cc? 95% confidence level.

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

- $X^2 = 24 \times 5 / 4 = 30$

- What is critical value of Chi Square for 24 degrees of freedom?

One Sample Chi Square

- Example: A sample of 25 bottles was selected. The variance of these 25 bottles as 5 cc. Has it **increased** from established 4 cc? 95% confidence level.

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

- $X^2 = 24x5\ / 4 = 30$

- What is critical value of Chi Square for 24 degrees of freedom?

One Sample Chi Square

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x²								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

One Sample Chi Square

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

- Example: A sample of 25 bottles was selected. The variance of these 25 bottles as 5 cc. Has it **increased** from established 4 cc? 95% confidence level.
- $X^2 = 24 \text{x} 5 / 4 = 30$
- Critical value of Chi Square for 24 degrees of freedom = 36.42
- Fail to reject H_0

QUALITYGURUS

ANOVA

- F-test
  - for testing equality of *two* variances from different populations
  - for testing equality of several means with technique of ANOVA.
- Chi-square test
  - For testing the population variance against a specified value
  - testing goodness of fit of some probability distribution
  - testing for independence of two attributes

ANOVA

- F-test
  - for testing equality of *two* variances from different populations
  - H_0: $\sigma^2_1 = \sigma^2_2$
  - $F_{\text{calculated}}$

$$F_{\text{calculated}} = \frac{S_1^2}{S_2^2}$$

  - Keep higher value at the top for right tail test.
  - Remember: Variance is square of standard deviation

ANOVA

- Why ANOVA?
  - We used t test to compare the means of two populations.
  - What if we need to compare more than two populations? With ANOVA e can find out if one or more populations have different mean or comes from a different population.
  - We could have conducted multiple t Test.
  - How many t Test we need to conduct if have to compare 4 samples? … 6

ANOVA

- Why ANOVA?
  - How many t Test we need to conduct if have to compare 4 samples? … 6

4				x
3			x	3 vs 4
2		x	2 vs 3	2 vs 4
1	x	1 vs 2	1 vs 3	1 vs 4
	1	2	3	4

ANOVA

- Why ANOVA?
  - How many t Test we need to conduct if have to compare 4 samples? … 6
  - Each test is done with alpha = 0.05 or 95% confidence.
  - 6 tests will result in confidence level of 0.95x0.95x0.95x0.95x0.95x0.95 = 0.735

ANOVA

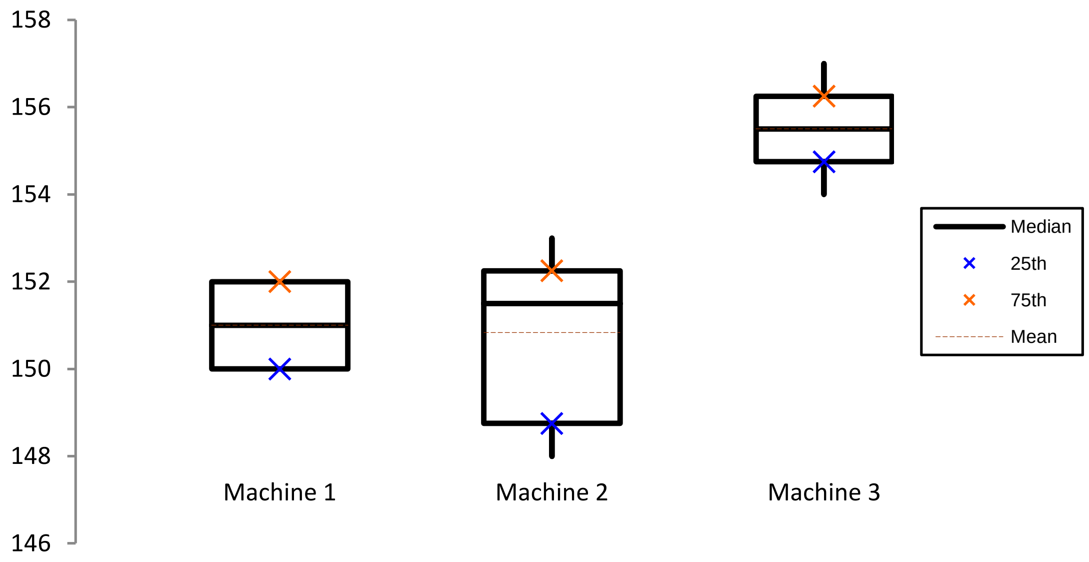
- Comparing three machines:

Machine 1	Machine 2	Machine 3
150	153	156
151	152	154
152	148	155
152	151	156
151	149	157
150	152	155
$\bar{x}1 = 151$	**$\bar{x}2 = 150.83$**	**$\bar{x}3 = 155.50$**

QUALiTYGURUS

ANOVA

- Comparing three machines:



Machine 1	Machine 2	Machine 3
150	153	156
151	152	154
152	148	155
152	151	156
151	149	157
150	152	155
$\bar{x}1$ = 151.00	**$\bar{x}2$ = 150.83**	**$\bar{x}3$ = 155.50**

ANOVA

- Comparing three machines:

Machine 1	Machine 2	Machine 3
150	153	156
151	152	154
152	148	155
152	151	156
151	149	157
150	152	155
$\bar{x}1$ = 151.00	**$\bar{x}2$ = 150.83**	**$\bar{x}3$ = 155.50**

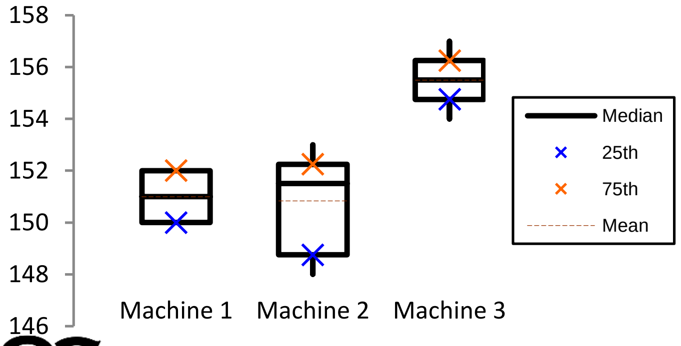
Machine 4	Machine 5	Machine 6
130	163	166
155	152	154
160	143	155
158	141	151
152	149	152
145	157	155
$\bar{x}1$ = 151.00	**$\bar{x}2$ = 150.83**	**$\bar{x}3$ = 155.50**

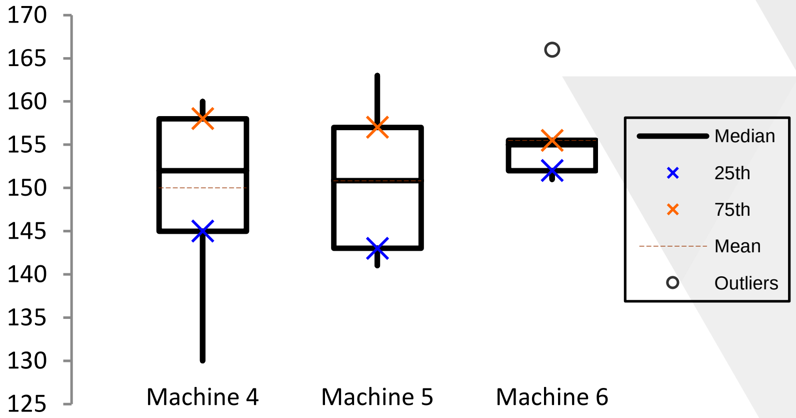
ANOVA

- Comparing three machines:

Machine 1	Machine 2	Machine 3
150	153	156
151	152	154
152	148	155
152	151	156
151	149	157
150	152	155
x̄1 = 151.00	**x̄2 = 150.83**	**x̄3 = 155.50**

Machine 4	Machine 5	Machine 6
130	163	166
155	152	154
160	143	155
158	141	151
152	149	152
145	157	155
x̄1 = 151.00	**x̄2 = 150.83**	**x̄3 = 155.50**





QUALiTYGURUS

ANOVA

- ANOVA is Analysis of Variance
- Variance

$$s^2 = \frac{\sum(x_i - \bar{X})^2}{n-1}$$

- Numerator of this formula is Sum of Squares
- Total of Sum of Squares (SST) =

$$SS_{\text{between/or treatment}} + SS_{\text{within/or error}}$$

ANOVA

- $SST = SS_{\text{between(or treatment)}} + SS_{\text{within(or error)}}$

- Ratio:

$SS_{\text{between(or treatment)}} / SS_{\text{within(or error)}}$

$F = MS_{\text{between(or treatment)}} / MS_{\text{within(or error)}}$

ANOVA

- $SST = SS_{\text{between(or treatment)}} + SS_{\text{within(or error)}}$

Machine 1	Machine 2	Machine 3
150	153	156
151	152	154
152	148	155
152	151	156
151	149	157
150	152	155
$\bar{x}1$ = 151.00	**$\bar{x}2$ = 150.83**	**$\bar{x}3$ = 155.50**

ANOVA

Machine 1	Machine 2	Machine 3
150	153	156
151	152	154
152	148	155
152	151	156
151	149	157
150	152	155
$\bar{x}1$ = 151.00	$\bar{x}2$ = 150.83	$\bar{x}3$ = 155.50

- SST = SS $_{\text{between(or treatment)}}$ +SS $_{\text{within(or error)}}$

Machine 1	x1 - $\bar{x}1$	Sqr(x1 - $\bar{x}1$)	Machine 2	x2 - $\bar{x}2$	Sqr(x2 - $\bar{x}2$)	Machine 3	x3 - $\bar{x}3$	Sqr(x3 - $\bar{x}3$)	
150.00	-1.00	1.00	153.00	2.17	4.69	156.00	0.50	0.25	
151.00	0.00	0.00	152.00	1.17	1.36	154.00	-1.50	2.25	
152.00	1.00	1.00	148.00	-2.83	8.03	155.00	-0.50	0.25	
152.00	1.00	1.00	151.00	0.17	0.03	156.00	0.50	0.25	
151.00	0.00	0.00	149.00	-1.83	3.36	157.00	1.50	2.25	
150.00	-1.00	1.00	152.00	1.17	1.36	155.00	-0.50	0.25	
151.00			150.83			155.50			152.44
		4.00			18.83			5.50	

- SS $_{\text{within}}$ = 4.00+18.83+5.50 = 28.33

QUALiTYGURUS

ANOVA

Machine 1	Machine 2	Machine 3
150	153	156
151	152	154
152	148	155
152	151	156
151	149	157
150	152	155
$\bar{x}1$ = 151.00	$\bar{x}2$ = 150.83	$\bar{x}3$ = 155.50

- $SST = SS_{\text{between(or treatment)}} + SS_{\text{within(or error)}}$

Machine 1	x1 - $\bar{x}1$	Sqr(x1 - $\bar{x}1$)	Machine 2	x2 - $\bar{x}2$	Sqr(x2 - $\bar{x}2$)	Machine 3	x3 - $\bar{x}3$	Sqr(x3 - $\bar{x}3$)	
150.00	-1.00	1.00	153.00	2.17	4.69	156.00	0.50	0.25	
151.00	0.00	0.00	152.00	1.17	1.36	154.00	-1.50	2.25	
152.00	1.00	1.00	148.00	-2.83	8.03	155.00	-0.50	0.25	
152.00	1.00	1.00	151.00	0.17	0.03	156.00	0.50	0.25	
151.00	0.00	0.00	149.00	-1.83	3.36	157.00	1.50	2.25	
150.00	-1.00	1.00	152.00	1.17	1.36	155.00	-0.50	0.25	
151.00			150.83			155.50			152.44
		4.00			18.83			5.50	
	-1.44	2.07		-1.61	2.58		3.06	9.36	

- $SS_{\text{between}} = (2.07+2.58+9.36)x6 = 84.06$

ANOVA

Machine 1	Machine 2	Machine 3
150	153	156
151	152	154
152	148	155
152	151	156
151	149	157
150	152	155
$\bar{x}1$ = 151.00	**$\bar{x}2$ = 150.83**	**$\bar{x}3$ = 155.50**

- $SST = SS_{\text{between(or treatment)}} + SS_{\text{within(or error)}}$
- $SST = 84.06 + 28.33 = 112.39$
- **<u>Degrees of freedom</u>**
- Total df = $df_{\text{treatment}} + df_{\text{error}}$
- $(N-1) = (C-1) + (N-C)$
- $df_{\text{treatment}} = 3-1=2$, $df_{\text{error}} = 18-3=15$
- $df_{\text{total}} = 17$

ANOVA

Machine 1	Machine 2	Machine 3
150	153	156
151	152	154
152	148	155
152	151	156
151	149	157
150	152	155
$\bar{x}1$ = 151.00	$\bar{x}2$ = 150.83	$\bar{x}3$ = 155.50

- **<u>Mean Sum of Square = SS / df</u>**
- $MS_{between} = SS_{between(or\ treatment)} / df_{treatment}$
- $MS_{between} = 84.06 / 2 = 42.03$
- $MS_{within} = SS_{within(or\ error)} / df_{within}$
- $MS_{within} = 28.33 / 15 = 1.89$
- $F = Ms_{between} / Ms_{within} = 42.03/1.89 = 22.24$

QUALITYGURUS

ANOVA

- $F = MS_{between} / MS_{within} = 42.03/1.89 = 22.24$

- Compare this with $F_{critical}$
- $F(2, 15, 0.95) = 3.68$

- Reject Null Hypothesis

- DEMONSTRATE MS Excel

F - Distribution ($\alpha = 0.05$ in the Right Tail)

df_2 \ df_1 (Numerator Degrees of Freedom)	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385
3	10.128	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123
4	7.7086	9.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	6.9988
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588
∞	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799

Denominator Degrees of Freedom

QUALiTYGURUS

Goodness of Fit Test (Chi Square)

- To test if the sample is coming from a population with specific distribution.
- Other goodness-of-fit tests are
  - Anderson-Darling
  - Kolmogorov-Smirnov
- Chi Square Goodness of Fit can be used for any time of data: Continuous or Discrete.

Goodness of Fit Test (Chi Square)

- H_0: The data follow a specified distribution.
- Ha: The data do not follow the specified distribution.
- Calculated Statistic: $\chi^2 = \sum_{i=1}^{k} \frac{(O-E)^2}{E}$
- Critical Statistic: Chi square for k-1 degrees of freedom for specific alpha.

Goodness of Fit Test (Chi Square)

- A coin is flipped 100 times. Number of heads are noted. Is this coin biased?

$$\chi^2 = \sum_{i=1}^{k} \frac{(O-E)^2}{E}$$

Expected	Observed
50	51
50	52
50	56
50	82
50	65

QUALiTYGURUS

Goodness of Fit Test (Chi Square)

- A coin is flipped 100 times. Number of heads are noted. Is this coin biased?

$$\chi^2 = \sum_{i=1}^{k} \frac{(O-E)^2}{E}$$

Expected	Observed	O-E	$(O-E)^2$	$(O-E)^2/E$
50	51	1	1	0.02
50	52	2	4	0.08
50	56	6	36	0.72
50	82	32	1024	20.48
50	65	15	225	4.5
				$X^2 = 25.8$

QUALITYGURUS

Goodness of Fit Test (Chi Square)

- A coin is flipped 100 times. Number of heads are noted. Is this coin biased?

$$\chi^2 = \sum_{i=1}^{k} \frac{(O-E)^2}{E}$$

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x ²								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38

$$X^2_{cal} = 25.8$$

$$X^2_{(4,0.95)} = 9.49$$

Goodness of Fit Test (Chi Square)

$$\chi^2 = \sum_{i=1}^{k} \frac{(O-E)^2}{E}$$

- A coin is flipped 100 times. Number of heads are noted. Is this coin biased?

- $X^2_{cal} = 25.8$
- $X^2_{(4,0.95)} = 9.49$

- Reject Null Hypothesis
  - Coin is biased

Contingency Tables

- To find relationship between two discrete variables.

	Smoker	Non Smoker	
Male	60	40	100
Female	35	40	75
	95	80	175

	Operator 1	Operator 2	Operator 3	
Shift 1	22	26	23	71
Shift 2	28	62	26	116
Shift 3	72	22	66	160
	122	112	115	347

Contingency Tables

- Null hypothesis is that there is no relationship between the row and column variables.
- Alternate hypothesis is that there is a relationship. Alternate hypothesis does not tell what type of relationship exists.

	Operator 1	Operator 2	Operator 3	
Shift 1	22	26	23	71
Shift 2	28	62	26	116
Shift 3	72	22	66	160
	122	112	115	347

Contingency Tables

❖ Calculate Chi square statistic.

$$\chi^2 = \sum_{i=1}^{k} \frac{(O-E)^2}{E}$$

	Operator 1	Operator 2	Operator 3	
Shift 1	22	26	23	71
Shift 2	28	62	26	116
Shift 3	72	22	66	160
	122	112	115	347

Contingency Tables

❖ Calculate Chi square statistic.

$$\chi^2 = \sum_{i=1}^{k} \frac{(O-E)^2}{E}$$

OBSERVED	Operator 1	Operator 2	Operator 3	
Shift 1	22	26	23	71
Shift 2	28	62	26	116
Shift 3	72	22	66	160
	122	112	115	347

EXPECTED	Operator 1	Operator 2	Operator 3	
Shift 1	122x71/347	112x71/347	115x71/347	71
Shift 2	122x116/347	112x116/347	115x116/347	116
Shift 3	122x160/347	112x160/347	115x160/347	160
	122	112	115	347

QUALiTYGURUS

Contingency Tables

- Calculate Chi square statistic.

$$\chi^2 = \sum_{i=1}^{k} \frac{(O-E)^2}{E}$$

EXPECTED	Operator 1	Operator 2	Operator 3	
Shift 1	122x71/347	112x71/347	115x71/347	71
Shift 2	122x116/347	112x116/347	115x116/347	116
Shift 3	122x160/347	112x160/347	115x160/347	160
	122	112	115	347

EXPECTED	Operator 1	Operator 2	Operator 3	
Shift 1	24.96	22.91	23.53	71
Shift 2	40.78	37.44	38.44	116
Shift 3	56.25	51.64	53.02	160
	122	112	115	347

Contingency Tables

- Calculate Chi square statistic.

$$\chi^2 = \sum_{i=1}^{k} \frac{(O-E)^2}{E}$$

OBSERVED	Operator 1	Operator 2	Operator 3	
Shift 1	22	26	23	71
Shift 2	28	62	26	116
Shift 3	72	22	66	160
	122	112	115	347

EXPECTED	Operator 1	Operator 2	Operator 3	
Shift 1	24.96	22.91	23.53	71
Shift 2	40.78	37.44	38.44	116
Shift 3	56.25	51.64	53.02	160
	122	112	115	347

$(O-E)^2/E$	Operator 1	Operator 2	Operator 3	
Shift 1	$(22-24.96)^2/24.96 = 0.35$	0.42	0.01	71
Shift 2	$(28-40.78)^2/40.78 = 4.00$	16.11	4.03	116
Shift 3	$(72-56.25)^2/56.25 = 4.41$	17.01	3.18	160
	122	112	115	347

$$X^2 = 49.52$$

Contingency Tables

- Calculate Chi square statistic = 49.52
- Degrees of freedom = (r-1)(c-1) = 4
- Chi square critical = 9.49
- Reject null hypothesis
- There is a relationship between the shift and the operator.

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38

Correlation

- $Y = f(X)$,
  - where Y is Dependent variable or the result (output)
  - X is Independent variable, input or the controllable variable

- For example in the study of marks obtained by students in a subject (Y) vs hours of study (X)

Correlation

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n(\sum x^2) - (\sum x)^2]\,[n(\sum y^2) - (\sum y)^2]}}$$

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Hours Studied (X)	Test Score % (Y)
20	40
24	55
46	69
62	83
22	27
37	44
45	61
27	33
65	71
23	37

QUALITYGURUS

Correlation

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n(\sum x^2) - (\sum x)^2]\,[n(\sum y^2) - (\sum y)^2]}}$$

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

	Hours Studied (X)	Test Score % (Y)	XY	X2	Y2
	20	40	800	400	1600
	24	55	1320	576	3025
	46	69	3174	2116	4761
	62	83	5146	3844	6889
	22	27	594	484	729
	37	44	1628	1369	1936
	45	61	2745	2025	3721
	27	33	891	729	1089
	65	71	4615	4225	5041
	23	37	851	529	1369
SUM	**371**	**520**	**21764**	**16297**	**30160**

Correlation Coefficient

- **Correlation**
- Measures the strength of linear relationship between Y and X
- Pearson Correlation Coefficient, r (r varies between -1 and +1)
  - Perfect positive relationship: r = 1
  - No relationship: r = 0
  - Perfect negative relationship: r = -1

Correlation Coefficient

Correlation vs Causation

- **Correlation does not imply causation**
  - a correlation between two variables does not imply that one causes the other

QUALITYGURUS

Coefficient of Determination

- Coefficient of Determination, r^2
- Proportion of the variance in the dependent variable that is predictable from the independent variable
- (varies from 0.0 to 1.0 or zero to 100%)
  - None of the variation in Y is explained by X, $r^2 = 0.0$
  - All of the variation in Y is explained by X, $r^2 = 1.0$

  - $r = 0.88$, $r^2 = 0.77$

QUALITYGURUS

Regression Analysis

- Quantifies the relationship between Y and X ($Y = a + bX$)

Regression Analysis

❖ Quantifies the relationship between Y and X (Y = a + bX)

	Hours Studied (X)	Test Score % (Y)	XY	X2	Y2
	20	40	800	400	1600
	24	55	1320	576	3025
	46	69	3174	2116	4761
	62	83	5146	3844	6889
	22	27	594	484	729
	37	44	1628	1369	1936
	45	61	2745	2025	3721
	27	33	891	729	1089
	65	71	4615	4225	5041
	23	37	851	529	1369
SUM	371	520	21764	16297	30160

$$Y = a + bX$$

$$b = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2} \qquad a = \frac{\sum Y - b\sum X}{N}$$

QUALITYGURUS

Regression Analysis

❖ Quantifies the relationship between Y and X (Y = 15.79 + 0.97.X)

$$Y = a + bX$$

$$b = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2} \qquad a = \frac{\sum Y - b\sum X}{N}$$

	Hours Studied (X)	Test Score % (Y)	XY	X2	Y2
	20	40	800	400	1600
	24	55	1320	576	3025
	46	69	3174	2116	4761
	62	83	5146	3844	6889
	22	27	594	484	729
	37	44	1628	1369	1936
	45	61	2745	2025	3721
	27	33	891	729	1089
	65	71	4615	4225	5041
	23	37	851	529	1369
SUM	371	520	21764	16297	30160

Regression Analysis

- For a student studying 50 hrs what is the expected test score %?

$$Y = a + bX$$

$$b = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2} \qquad a = \frac{\sum Y - b\sum X}{N}$$