Design of Experiments I Spring 2002

Introductory Material

Statistical Concepts and Methods Useful for Design of Experiments

Feel a little uneasy with statistical terms and methods? This section should help.

Apart from her fame as a pioneer in modern nursing and public health, Florence Nightingale was a distinguished member of the American Statistical Association

Hear her actual voice recordings at: **http://www.dnai.com/~borneo/nightingale/nutting.htm**

- **"(Statistics is) the most important science in the whole world, for upon it depends the practical application of every other science and of every art: the one science essential to all political and social administration, all education, for only it gives exact results of our experience....to understand God's thoughts, we must study statistics, for these are the measure of his purpose." --- Florence Nightingale**

Samples Vs. Populations

- **Sample:** A subset of size *n* of observations drawn from a target population.

- **Random Sample:** A sample selected in a random manner -- each member of the population has an equal probability of being selected.

- **Population:** A conceptual or physical set of all possible observations that can be collected (a.k.a.,Target Population).

- **Population Mean:** The theoretical average of a population denoted by the Greek letter μ.

- **Population Standard Deviation:** The theoretical standard deviation of a population denoted by the Greek letter σ.

- **Population Variance:** The theoretical variance of a population denoted by σ^2.

Frequency Distributions

- Data can be displayed using a variety of 'visual' methods. One important tool is the **frequency distribution**.
- In any population, some random variation is always present - no two members of a population are identical.

For example, there is some variation in the following:

- ❖ Height and weight of human beings
- ❖ Number of blue M&M's in randomly selected bags

- For discrete or categorical data, **frequency** (count) is the number of times each value occurs in the data set.
- **For example**: The number of blue M&M's in each bag is a discrete measurement. The 'frequency' of blue M&M's can be counted.

Frequency Distributions

- If the data are continuous measurements, then we must use **grouped frequency**. The range of the data is divided into convenient intervals (bins) and we count the number of data points falling into each interval.

For example: The frequency of '5th grade girls in a particular school that weigh 65-75 lbs' is a 'grouped frequency'.

- The frequency (total count) for each interval can be converted into a **relative frequency, or 'percent'**. Here, we divide each interval frequency by the total number of data points to calculate a 'percent'.

  The relative frequency of 5th grade girls weighing 65-75 lbs is a percent of all of the 5th grade girls.

Frequency Distributions

- On the following slide, **grouped frequencies**, **relative frequencies (percents)** and **cumulative percents** have been calculated for a data set.
- The following steps were followed :
  1. Order the data points from smallest to largest.
  2. Count the number of data points (n).
  3. Determine bin size - square root n is a good guideline.
     Note: bins must be the same size and cannot overlap.
  4. Place a tally mark in the appropriate bin for each data point.
  5. Count the number of tally marks in each bin.
  6. Calculate the grouped frequency, percent and cumulative percent for each bin.

Frequency Distributions

Aluminum Lithium Strength Data

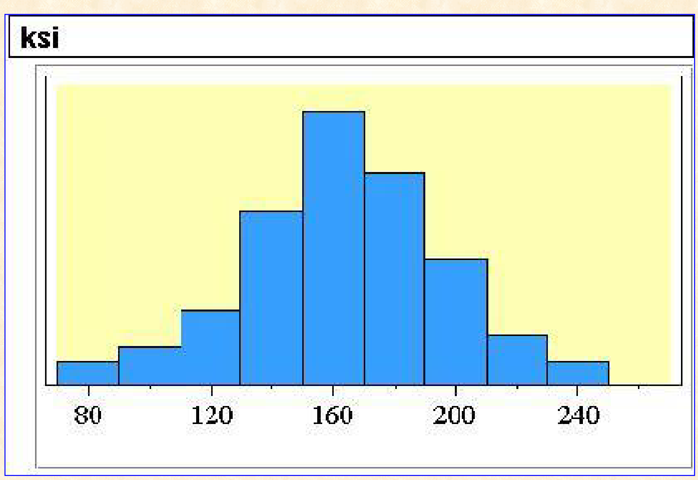
Batch A	Batch B	Batch C	Batch D	Batch E	Batch F	Batch G	Batch H
105	221	183	186	121	181	180	143
97	154	153	174	120	168	167	141
245	228	174	199	181	158	176	110
163	131	154	115	160	208	158	133
207	180	190	193	194	133	156	123
134	178	76	167	184	135	229	146
218	157	101	171	165	172	158	169
199	151	142	163	145	171	148	158
160	175	149	87	160	237	150	135
196	201	200	176	150	170	118	149

Frequency Distribution of Al-Li Data

Class Interval KSI*	Tally Mark	Frequency	Relative Frequency**
$70 \le x < 90$	//	2	0.0250
$90 \le x < 110$	///	3	0.0375
$110 \le x < 130$	///// /	6	0.0750
$130 \le x < 150$	///// ///// ////	14	0.1750
$150 \le x < 170$	///// ///// ///// ///// //	22	0.2750
$170 \le x < 190$	///// ///// ///// //	17	0.2125
$190 \le x < 210$	///// /////	10	0.1250
$210 \le x < 230$	////	4	0.0500
$230 \le x < 250$	//	2	0.0250

*Min value 76, Max value 246, n=80: use range 70-250 with 9 bins of width 20

Histogram of the Aluminum Lithium Data



Frequency Distributions

A **Bar chart** is a graphical display of a frequency distribution for categorical data.

A **Pareto Chart** is a special type of bar chart (more on these charts later).

A **Histogram** is a graphical display of a grouped frequency distribution for continuous data.

There are three important characteristics of histograms:

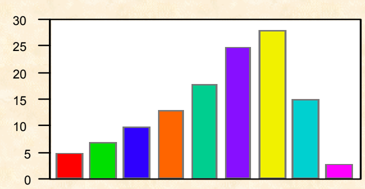
- **Shape**: Is the shape symmetric or 'skewed'?
- **Central tendency**: Is there a mounded shape?
- **Scatter**: How do the data spread about the center?

These characteristics are demonstrated on the following slides:

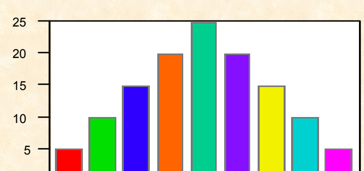
Frequency Distributions

Common Histogram Shapes:

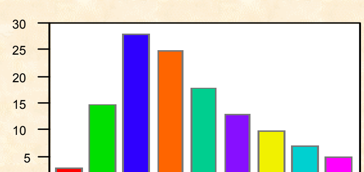
Left Skewed: Data trails off to the left.



Symmetric: Data has approximately the same distribution on either side of the center point.



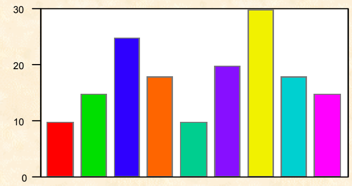
Right Skewed: Data trails off to the right. Time data is almost always right skewed.



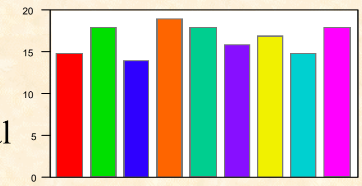
Frequency Distributions

What is the Central Tendency?

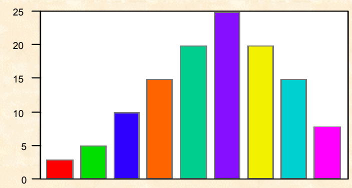
Bi-modal or multi-modal: More than one 'peak'. This may be an indication that there is more than one population in the data.



Uniform: Data is evenly distributed with no peak or central tendency.



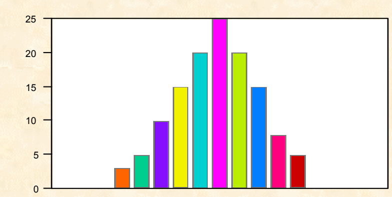
Uni-modal: Data has one central tendency.

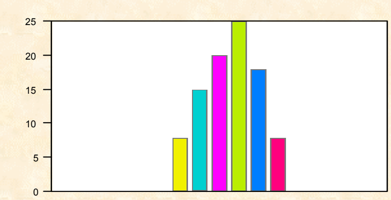


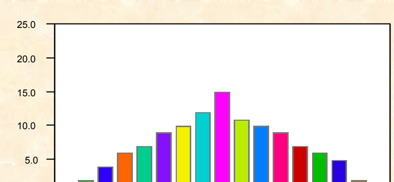
Frequency Distributions

How much 'spread' or variability is in the data?

How much spread is there about the central tendency?





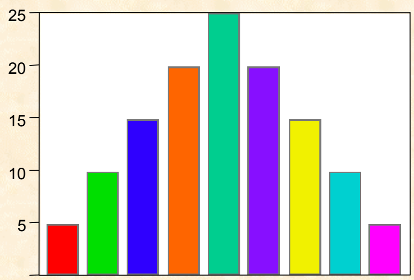


Frequency Distributions

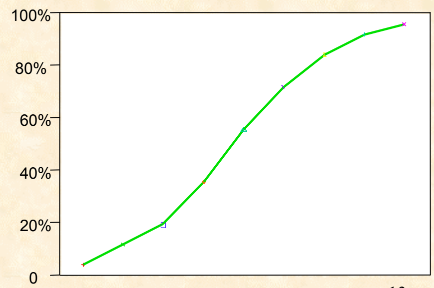
As we saw previously, **Cumulative Frequency Distributions** are another effective method for displaying data.

This type of display is often used when we are interested in the percent of items falling below a particular value.

Frequency Distribution



Cumulative Frequency Distribution



Frequency Distributions

Histograms provide many **benefits**. Histograms:

- Are a graphical representation of a frequency distribution.
- Are good **compact summaries** of a data set.
- Are **easy to interpret** variation in the data set.
- Help to identify **unusual patterns**.

Unfortunately, histograms also have many **limitations** and should be used with caution:

- Individual **data points are lost**.
- Histograms are **unstable with less than 75 data points** and can be very misleading. 100 or more data points is highly recommended.
- **Improper bin sizes**, as we will see on the following slide, can **mask important** data features.

The Empirical Rule

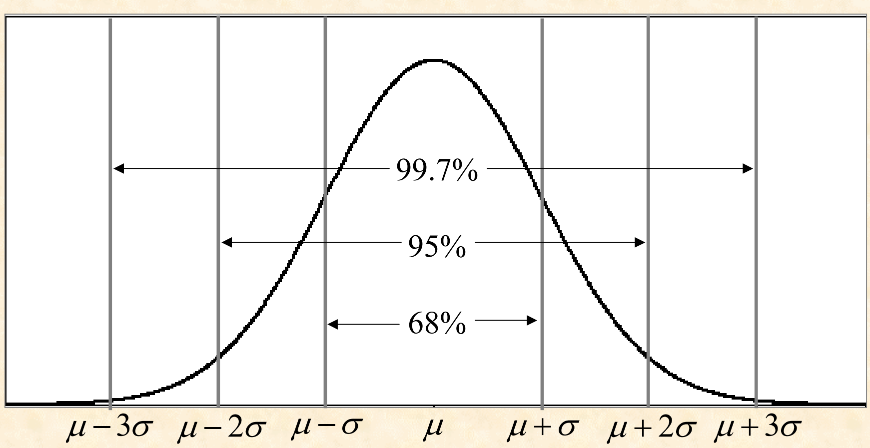
In manufacturing we are often interested in predicting the proportion of products or units that will fall within a specified range (or specification limit).

According to the **Empirical Rule**: If a frequency distribution is approximately symmetric and mounded in shape, then we will observe the following:

- Approximately **68%** of all values will fall within **±1*S*** (standard deviation) of the mean.
- Approximately **95%** will fall within **±2*S*** of the mean.
- Nearly **100%** will fall within **±3*S*** of the mean.

As we will see, this rule is used a great deal in industrial statistics to calculate quantities such as spec limits and process limits.

The Empirical Rule



If a random variable is Normally distributed, then the Empirical Rule works exactly.

The Empirical Rule

- The sample mean is 162.66 and the sample standard deviation is 33.77.
- Applying the Empirical Rule to this data:
  - ❖ We expect 68% of the data to be in the interval [129, 196].
  - ❖ The observed frequency is 57 values or 71%.
  - ❖ We expect 95% of the data within the interval [95, 230].
  - ❖ The observed frequency is 75 values or 94%.
- The Empirical Rule typically works well if the frequency distribution is not too skewed (with a long tail) or too flat (no central mound).

Measures of Location and Spread

While graphical displays can show us trends or patterns in our data, we are often interested in two important characteristics:

- **Measures of the 'location'** of the center of the population or sample, also referred to as the 'central tendency'.

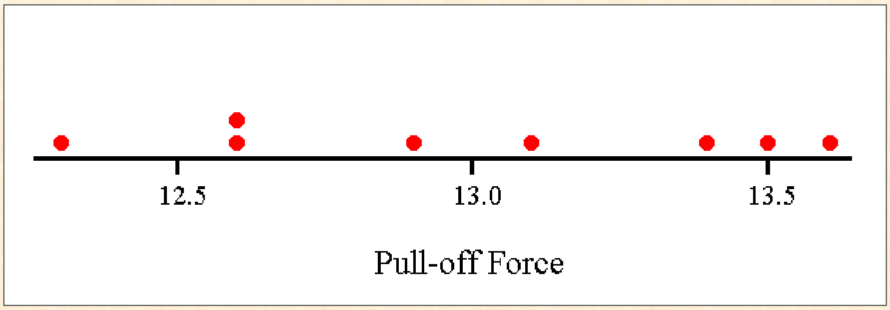
  Three measures of central tendency are mean, median, and mode.

- **Measures of 'scatter' or variability** of the population or sample. These are measures of the spread of the values around the central tendency.

  Three measures of spread are variance, standard deviation, and range.

Measures of Location and Spread

- For the pull-off force data presented earlier, what can we conclude about the distribution of values?



- What is a typical pull-off force?
- How do the measurements vary about the center?

Measures of Location and Spread

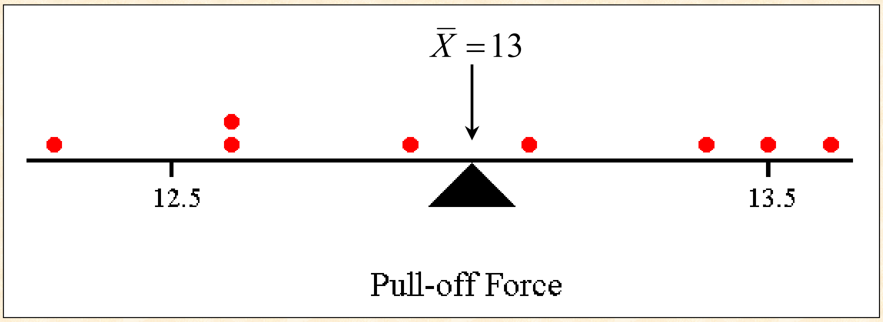
- The **Sample Mean** or average is the most important measure of 'the middle" of the data.
- The sample mean, referred to as 'X-bar', is the average of all observations from a sample:

$$\bar{X} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

- The mean is the center of gravity or balancing point of a data set.
- The sample mean is an estimate of the population mean, which is the average of all observations from a population.

Measures of Location and Spread

An example of a sample mean for the pull-off force data.



- Notice that the mean is denoted by a fulcrum to emphasize that it is the balancing point of the distribution of values.

Measures of Location and Spread

- The **Sample Median** is the 50th percentile of the sample data. Half of the data values lie below the median and half lie above the median.

  The median is the middle value.

  Notion: We usually denote the median by placing a ~ (tilde) over the variable, or we use the subscript 0.50.

  Example: The sample median is denoted $\tilde{X}$ or $X_{0.50}$.

- The sample median and sample mean are approximately the same if the distribution is symmetric. If the distribution is highly skewed, then the mean and median are quite different.

- Can you explain why?

Measures of Location and Spread

- The two most important measures of the variability or spread of sample data are **Variance** and **Standard Deviation**.
- **Sample Variance**, denoted by S^2:
  - ❖ S^2 is an estimate of the population variance σ^2.
  - ❖ S^2 = average squared distance the data points lie from their sample mean.
- **Sample Standard Deviation**, denoted by S:
  - ❖ S is an estimate of the population standard deviation σ.
  - ❖ S = average distance the data points lie from their sample mean.

Measures of Location and Spread

- The sample variance is calculated as the average squared distance from the sample mean.
- In other words, the variance is computed by calculating distance from each data point to the sample mean:

$$\boxed{S^2 = \frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}{n-1}}$$

- The sample standard deviation, the average distance from the sample mean, is calculated from the sample variance:

$$\boxed{S = \sqrt{S^2}}$$

Measures of Location and Spread

- Calculation of variance and standard deviation for the pull-off force data.

i	Pull-off (x_i)	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
1	12.6	-0.4	0.16
2	12.9	-0.1	0.01
3	13.4	0.4	0.16
4	12.3	-0.7	0.49
5	13.6	0.6	0.36
6	13.5	0.5	0.25
7	12.6	-0.4	0.16
8	13.1	0.1	0.01
	104.0	0.0	1.60

$$S^2 = \frac{1.60}{7} = 0.2286(ft - lbs^2), \quad S = \sqrt{0.2286} = 0.4781(ft - lbs)$$

Measures of Location and Spread

- The denominator for the calculation of sample variance, $S^{2,}$ is *n-1*. This is known as the '**Degrees of Freedom**'.
- Degrees of freedom is an important concept in statistics which will be referred to often.
- The degrees of freedom are the number of '**freely varying observations**' in a sample data set. **The larger the sample size, the larger the degrees of freedom**. Larger degrees of freedom allows us to make stronger inferences about the population characteristic.

(This concept is extremely important, particularly when we are discussing Designed Experiments).

The Normal Distribution

The most important distribution is known as the **Normal** or **Gaussian Distribution**. This distribution, known as the *Bell Shaped Curve*, is the basis for most of the statistical techniques we will cover throughout the course.

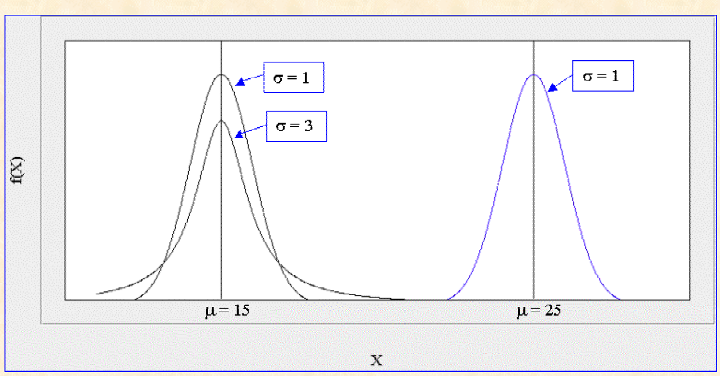
For all continuous distributions, calculus is required to calculate the probabilities - the area under the curve.

To determine probabilities for normally distributed populations, a ‘normal probability table’ has been created (as we will see in a few slides).

There are infinitely many possible normal probability curves, each depending on the mean and the variance. On the next slide, we see examples of normal curves with different means and variances.

The Normal Distribution

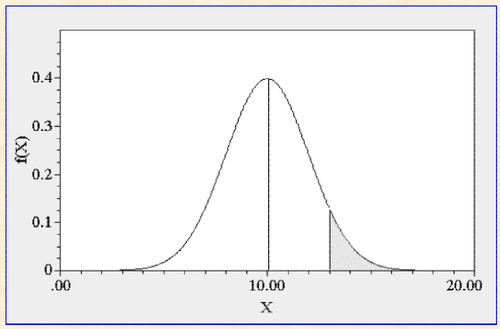
There are infinitely many possible normal distributions defined by values of the population parameters μ and σ^2.



For any normal random variable X with mean μ and variance σ^2, we denote the distribution as **$X \sim N(\mu; \sigma^2)$**.

The Normal Distribution

Example: The distribution for pull-off force for bonded wires (in the semi-conductor industry) is normally distributed, and has a mean of 10 and a variance of 4. We denote the distribution as X~N(10;4):



The shaded area shows P(X>13) for a normal random variable with $\mu = 10$ and $\sigma = 2$.

The Normal Distribution

A normal random variable with $\mu = 0$ and $\sigma^2 = 1$ is called a **standard normal random** variable and is denoted by '**Z**'.

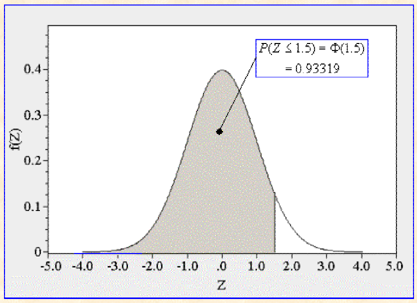
We denote the distribution for the **standard normal** variable by **Z~N(0;1)**. This is a reference distribution that all normal distributions can be compared to.

The normal probability table, shown on the next slide, is based on the cumulative standard normal distribution - the cumulative area under the standard normal curve from right to left.

Note that the table starts at Z=0.00 and goes to Z=3.09 (negative values of Z are not listed). This is because the normal distribution is symmetric around the mean - both sides of the distribution have the same probabilities.

The Normal Distribution

This table can also be found in the Appendix (**Normal Curve Probabilities**).



Z	0.00	0.01	0.02	...
0.0	0.50000	0.50399	0.50398	...
.	.	.	.	...
1.5	0.93320	0.93448	0.93574	...

The figure above depicts a typical probability calculation for *Z* using the normal probability table on the previous slide.

From the table, we find that the probability that $Z \leq 1.5$, written $P(Z \leq 1.5)$, is .93.

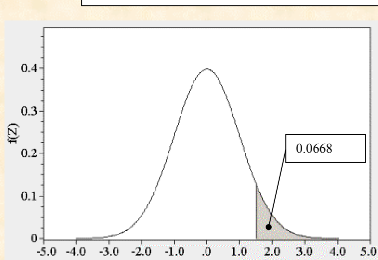
The Normal Distribution

Example: Probability calculations for *Z*:

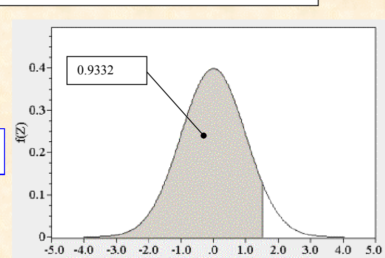
The table gives the cumulative probability (area under the curve) from the left.

To find a probability (area) to the right, subtract from 1, since the total area under the curve = 1.

$$\boxed{P(Z > 1.5) = 1 - P(Z \leq 1.5) = 1 - 0.9332 = 0.0668}$$



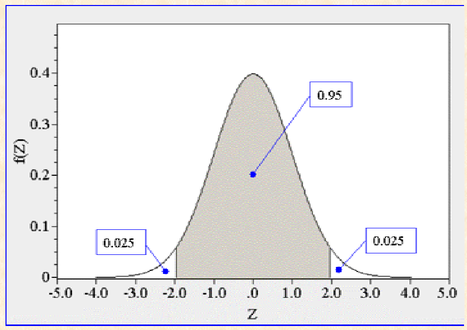
= 1 -



The Normal Distribution

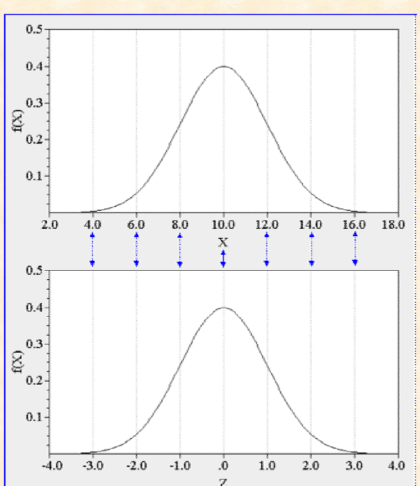
Suppose we wish to find percentiles of *Z* such that P(*z*<*Z*<*z*)=0.95. Because the normal distribution is symmetric, each tail outside of this interval contains an area (probability) of 0.025.

From the Standard Normal table, P(Z < 1.96) =0.975. Therefore, P(Z < -1.96) = 0.025 and our desired interval is P(-1.96 < Z < 1.96) = 0.95.



The Normal Distribution

In the figure below we compare the normal random variable *X* to its standard normal counterpart *Z*:



$$X : \mu = 10, \sigma = 2$$

$$Z : \mu = 0, \sigma = 1$$

$$Z = \frac{X - \mu}{\sigma}$$

$$Z = \frac{13 - 10}{2} = 1.5$$

The Normal Distribution

We may be interested in whether our process is capable of producing parts that are within the specs that have been established by either an internal or an external customer.

Example: For a machining operation, the most critical measurement of the operation is the outer diameter of the parts produced.

A specification has been established by the external customer. The target is .25 mm +/- .0015 mm (the **lower spec is .2485 mm and the upper spec is .2515 mm**).

The outer diameter, based on process data, has the distribution $X \sim N(0.2508; 0.0005^2)$.

What percentage of parts will be within specifications?

The Normal Distribution

We start by **turning X into Z values**.

$$Z = \frac{X - \mu}{\sigma}$$

$$Z = \frac{.2485 - .2508}{.0005} = -4.6$$

$$Z = \frac{.2515 - .2508}{.0005} = 1.4$$

Then, we **use the standard normal table to calculate the probabilities of Z**.

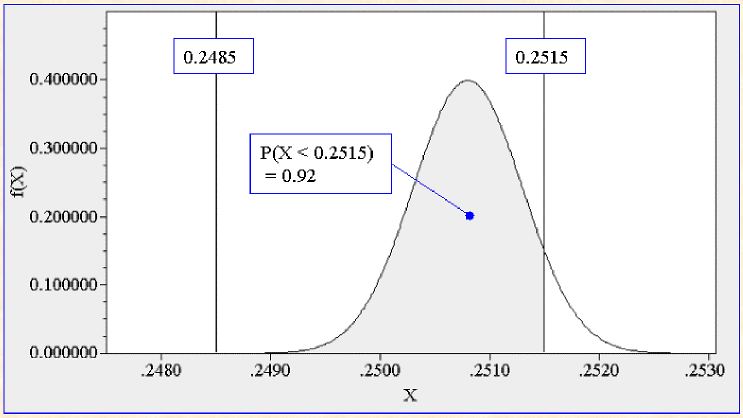
$P(Z<-4.6) = 0.00$, and $P(Z<1.4) = 0.92$

Thus, $P(-4.6 < Z < 1.4) = 0.92 - 0.00 = 0.92$.

The Normal Distribution

Therefore, P(.2485 ≤ X ≤ .2515) = 0.92.

Since **the probabilities of Z apply to our random variable X**, approximately 92% will be within specification.



Normal Quantile Plot

How do we know if our data are normally distributed? In practice we are often not certain if a sample data set came from a normally distributed random variable.

For the normal distribution: $X_p = Z_p\sigma + \mu$

We can calculate percentile values *p* from the sample data set.

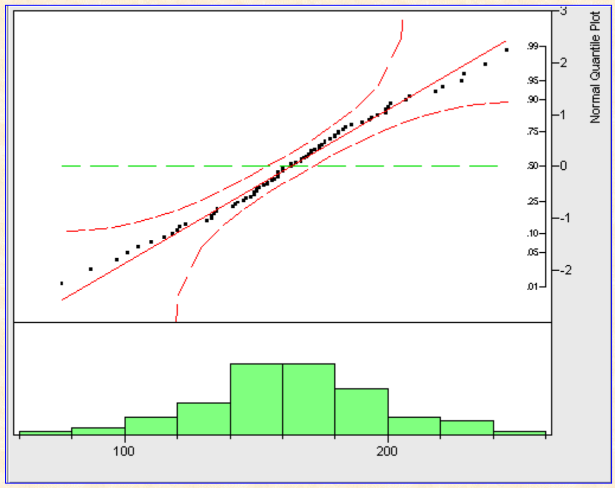
We can then plot these percentiles of X versus the corresponding percentiles of Z.

If the plot is approximately a straight line, then we assume that the data come from a normal distribution.

This plot is referred to as a **Normal Quantile Plot**.

Normal Quantile Plot

We will continue with the Aluminum Lithium Data. Do the KSI data seem to come from a normal distribution?



The Rule of Averages

A very powerful rule used throughout industrial statistics is **The Rule of Averages** (or the **Central Limit Theorem**):

Suppose we have a random variable X with a population mean μ and a population standard deviation σ, then sample averages of size n drawn from the population of X are normally distributed with population mean = μ and population standard deviation = $\sigma/\sqrt{n}$.

The term **"standard error of the mean",** $S/\sqrt{n}$, is the standard deviation of the sample mean, where n is the sample size.

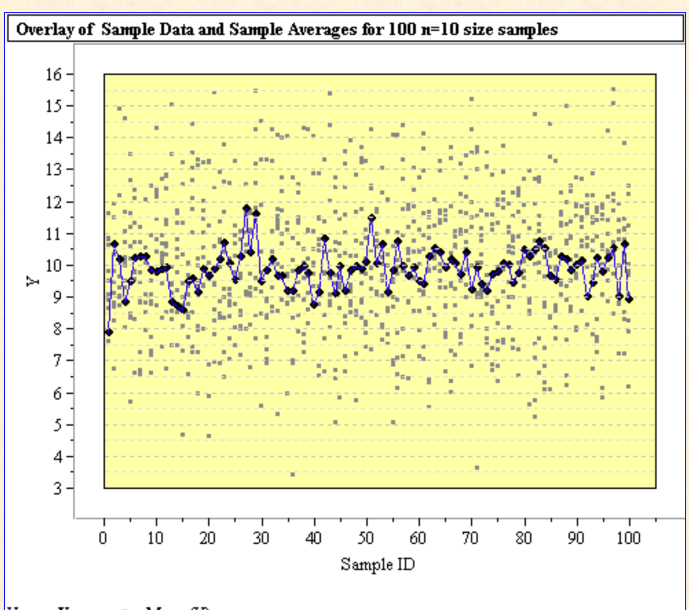
The Rule of Averages

Why is this rule important? It is one of the fundamental reasons we use sample data:

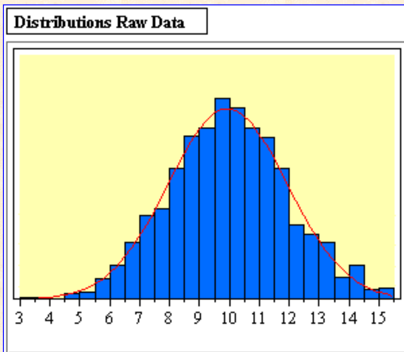
- The **standard deviation of the sample mean is always less** than the standard deviation of the raw data (by a factor of the square root of the sample size).
- **We get more precise estimates** of population behavior by taking samples. Larger samples give more precise estimates than smaller samples.
- **Averaging is a very effective noise filter**.
- **Sample means are normally distributed**.

As we will see in the following examples, the Rule of Averages applies to both normal and non-normal distributions.

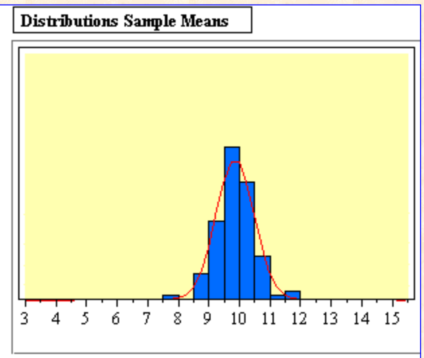
Overlay Plot of Raw Values of X and Mean Values of X, n=10



• The **Rule of Averages** at Work: Raw values X~N(10;4) compared to their sample means for sample size n=10

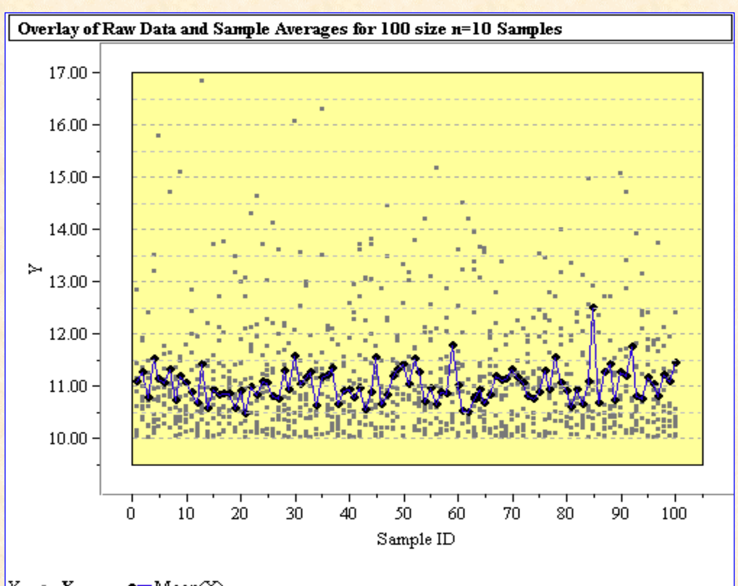


Mean	9.9627856
Std Dev	1.9867533
Std Err Mean	0.0628267
upper 95% Mean	10.086075
lower 95% Mean	9.8394962
N	1000

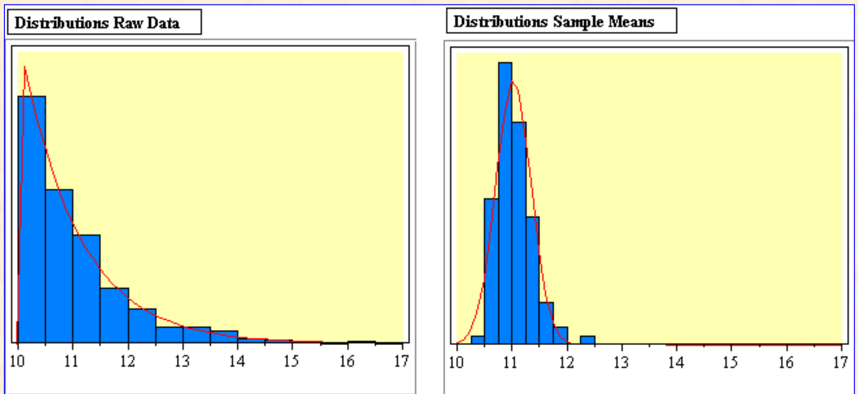


Mean	9.8798881
Std Dev	0.6190726
Std Err Mean	0.0195768
upper 95% Mean	9.9183051
lower 95% Mean	9.8414711
N	1000

Overlay Plot of Raw Values of X and Mean Values of X for n=10



The Rule of Averages at Work: Raw exponential values X compared to their sample means for sample size n=10



Mean	11.034718
Std Dev	1.0031181
Std Err Mean	0.0317214
upper 95% Mean	11.096968
lower 95% Mean	10.972469
N	1000

Mean	11.035997
Std Dev	0.3213689
Std Err Mean	0.0101626
upper 95% Mean	11.05594
lower 95% Mean	11.016054
N	1000

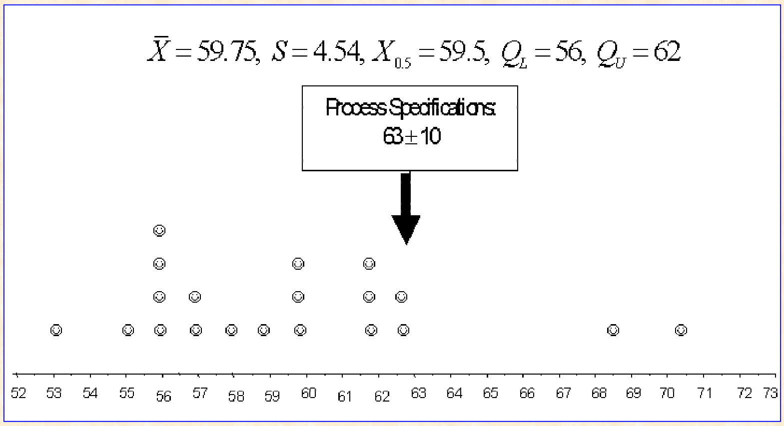
Confidence Intervals

• Point estimates of parameters, like the sample mean, do not tell us anything about the precision of the estimate. As an example, a single estimate of the population mean μ may fall very far from the true underlying value.

• Recall that point estimates are sample statistics that have a probability distribution. We make use of these distributions in order to determine how close our individual sample estimates, of a population parameter, may be to the true underlying parameter value.

• Two popular statistical methods to quantify the precision in a point estimate are called: (1) Confidence Intervals, and (2) Hypothesis Tests.

Case Study: Color Rendition Index (CRI) is an important characteristic of fluorescent lamps (100 is sunlight). The dot plot depicts 20 CRI values for a lot of manufactured Compact Fluorescent Lamps (CFLs). Is the process on target?



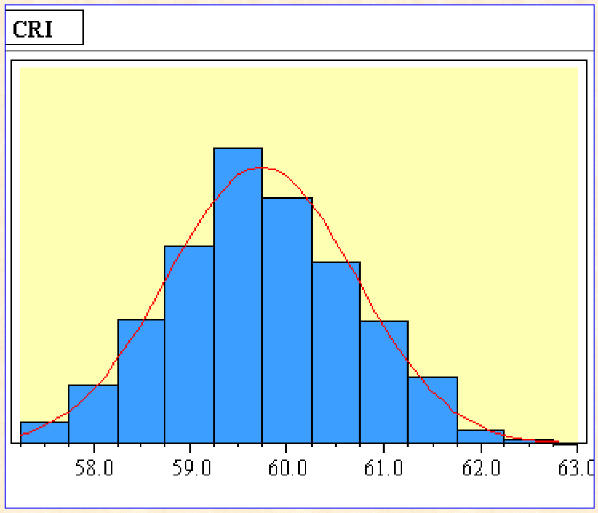
- If the lamp manufacturing process is on target (μ=63), then is it likely to observe a sample average as low as **59.75**? Is the process on target? To answer this question we must look at the probability distribution for the sample average and determine if the observed sample value is probable; i.e., have we observed an error in our process?

$$S_{\bar{X}} = \frac{S}{\sqrt{n}}$$

- Use the sample standard deviation S to obtain the standard error for the sample mean (S.E.=1.02).
- Use the *Empirical Rule* to determine if 59.75 is a probable sample average given the process is on target.
- 95% of sample averages should be in the interval $63 \pm 2(1.02) = [60.96, 65.04]$, 99% in $63 \pm 3(1.02) = [59.94, 66.06]$.
- Neither interval contains the sample average 59.75.

Histogram for 1,000 simulated sample mean CRI values

CRI~N(59.75;1.02^2). Not one sample mean exceeds 63.



Finding a Confidence Interval for "μ" With Small n

- We wish to calculate an interval of possible values for the true unknown population mean μ.

- We must be able to quantify our level of confidence that the interval contains the true unknown mean.

- Statistical **confidence intervals** provide a way to compute such an interval of possible values for μ.

- Student's t distribution is the basis for statistical **confidence intervals.**

Student's t Distribution

• If $\boxed{X \sim N(\mu;\sigma^2)}$ then $\boxed{Z = \frac{X-\mu}{\sigma}}$ follows the "standard normal distribution", N(0;1).

• σ is usually unknown and is estimated by the sample standard deviation S.

• σ must be a known value in order for Z to follow a standard normal distribution.

• When S is used to estimate σ, we have Student's t distribution

$\boxed{t_{n-1} = \frac{X-\mu}{S}}$ or $\boxed{t_{n-1} = \frac{\bar{X}-\mu}{S/\sqrt{n}}}$ with μ_t=0.0 and σ_t>1.0.

• The added variability (σ > 1.0) of Student's t comes from estimating σ with S.

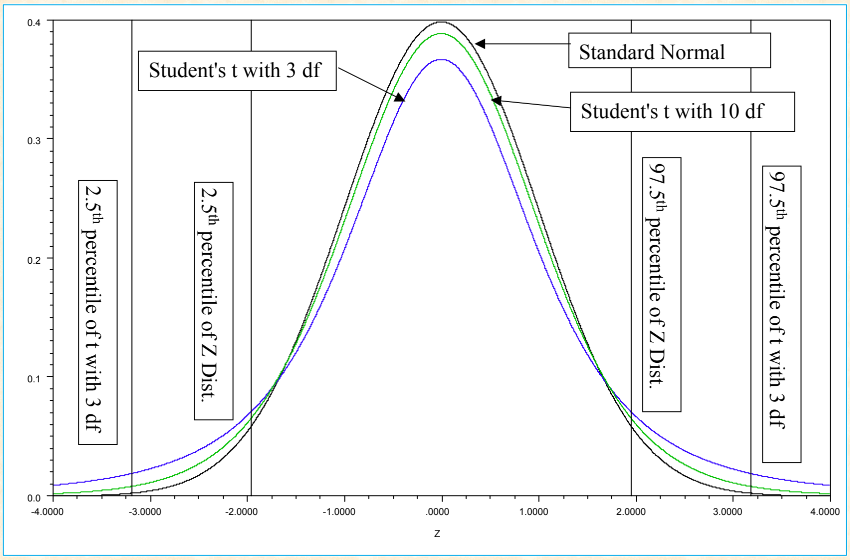
• The exact shape of Student's t distribution depends upon the (*n-1*) degrees of freedom used to estimate S.

• Student's t has thicker tails compared to the standard normal for small *degrees of freedom*. As the sample size *n* increases, Student's t approaches the shape of the standard normal.

• A table of percentile values for each distribution is required, since the distribution shape changes with the *degrees of freedom n-1*.

• The appendix contains a table of percentiles for Student's t distribution -- the rows are indexed by increasing degrees of freedom; i.e., each row represents a different distribution curve. The columns are indexed by percentile values for each distribution.

Standard Normal Distribution Vs. Student's t Distributions



Calculating a Confidence Interval for μ Using Student's t Distribution

1. Specify a desired level of confidence for the interval -- usually denoted by the Greek letter α (alpha) and typical values are 0.10, 0.05, or 0.01.

2. For the selected α, compute the *Percent Confidence* for interval. The *Percent Confidence* = (1- α)*100.

3. Compute the required percentile from Student's t distribution.
   * The percentile = (1- α/2)
   Example: for a 95% C.I., α=0.05, the percentile is 0.975.

(For the *degrees of freedom n-1* and the specified percentile, use the Student's t table to find the appropriate "Student's t value").

Example: α=0.05, n=11, the desired percentile, from Student's t Table is $t_{0.975,10}$=2.228.

5. Using and sample standard deviation S, calculate the desired confidence interval for μ using the formula:

$$\left[\bar{X} - t_{1-\alpha/2,n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{1-\alpha/2,n-1}\frac{S}{\sqrt{n}}\right]$$

- We are $(1-\alpha)\%$ confident the interval contains the unknown value for the population mean μ.

- Confidence intervals are easy to calculate but take a little practice at first.

Partial Student's t Table

	CUMULATIVE PROBABILITY					
DF	0.75	0.90	0.95	0.975	0.99	0.995
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	.765	1.638	2.353	3.182	4.541	5.841
4	.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	.718	1.440	1.943	2.447	3.143	3.707
7	.711	1.415	1.895	2.365	2.998	3.499
8	.706	1.397	1.860	2.306	2.896	3.355
9	.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	.697	1.363	1.796	2.201	2.718	3.106
12	.695	1.356	1.782	2.179	2.681	3.055
13	.694	1.350	1.771	2.160	2.650	3.012
14	.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	.690	1.337	1.746	2.120	2.583	2.921
17	.689	1.333	1.740	2.110	2.567	2.898
18	.688	1.330	1.734	2.101	2.552	2.878
19	.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845

Example: Calculate a 95% Confidence Interval for the CRI data. For n=20 sample measurements we have:

$$\bar{X} = 59.75,\ S = 4.54,\ \ t_{19,0.975} = 2.093$$

Recall our formula for a C.I.: $\bar{X} \pm t_{n-1,1-\alpha/2} \dfrac{S}{\sqrt{n}}$

Therefore our 95% C.I. is

$$59.75 \pm 2.093 \frac{4.54}{\sqrt{20}} = 59.75 \pm 2.093(1.02) =$$
$$57.63 \le \mu \le 61.88$$

Notice that the interval does not contain 63 which indicates that the target value it is not a plausible value for the process mean.

What Does a Confidence Interval Tell Us?

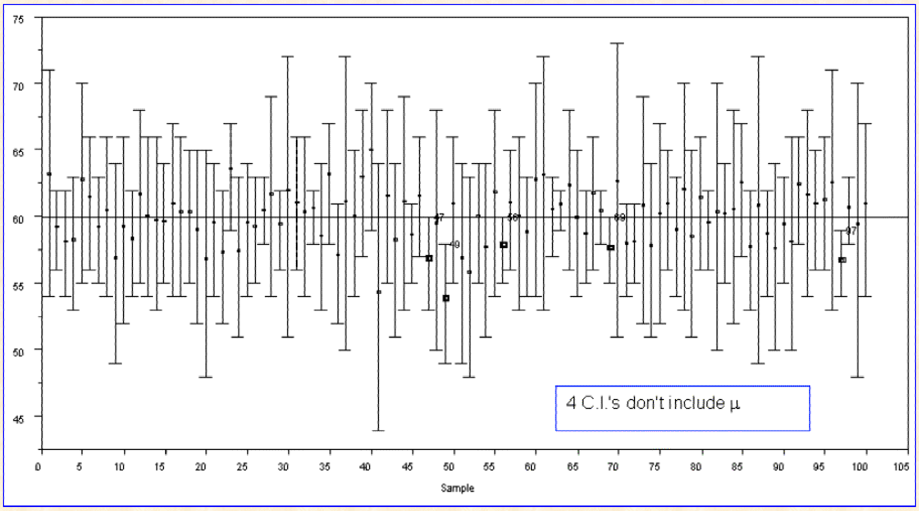
• Although fairly easy to calculate, confidence intervals have a subtle interpretation and are often misinterpreted -- even by statisticians.

• A (1-α)% confidence interval does not imply a probability that any one interval actually includes the true mean μ.

• The confidence interval level actually states that, "if I collected 100 sample sets and constructed a C.I. for each set, then (1-α)% of those C.I.'s will bracket the true unknown mean μ -- no claims are made for anyone of the individual computed C.I.'s.

* As an example, a 95% C.I. implies that if we collected 100 sample sets and computed 100 C.I.'s, on average 95 out of the 100 intervals will contain the true unknown mean μ.

100 Simulated 95% Confidence Intervals for, n=5, Using Student's t, 4 df



Calculating Confidence Intervals in JMP

• Many find the arithmetic of confidence intervals rather obscure, however modern software like JMP provide confidence interval estimates for the data analyst.

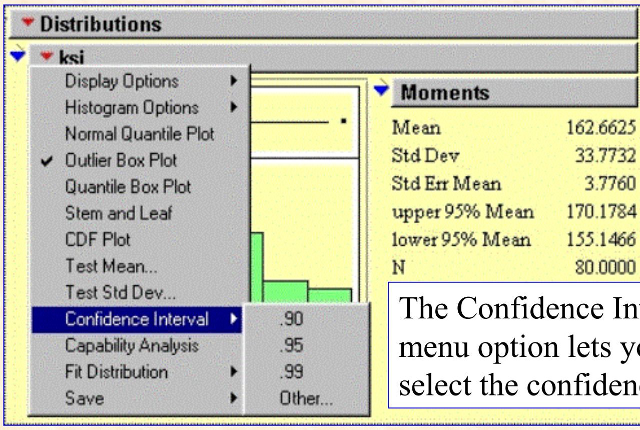
• It is important, however, that the data analyst properly understand the interpretation of the confidence intervals that are computed by the software.

• JMP provides confidence interval estimates for a population mean μ in a number of platforms where such intervals make sense for the data analysis.

• We will only cover the more common places where a data analyst might wish to compute a confidence interval for a single population mean for some number of different population means.

• The most common platform to use, when confidence intervals are desired for a population mean, is **Distribution of Y.**

• In **Distribution of Y** one can specify the confidence level to use in estimating a confidence interval, but JMP provides a 95% confidence interval by default.



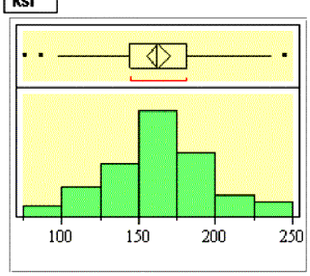
A 95% interval is provided by default

The Confidence Interval menu option lets you select the confidence level

Example: Use the dataset Al_Li.JMP to compute a 90%, 99%, and 95% confidence interval for the unknown population mean KSI value. Use the Distribution of Y platform.

Distributions

ksi



Moments

Mean	162.6625
Std Dev	33.7732
Std Err Mean	3.7760
upper 95% Mean	170.1784
lower 95% Mean	155.1466
N	80.0000

95% CI

Confidence Intervals

Parameter	Estimate	Lower CI	Upper CI	1-Alpha
Mean	162.6625	156.3779	168.9471	0.900
Std Dev	33.77324	29.90657	38.90867	

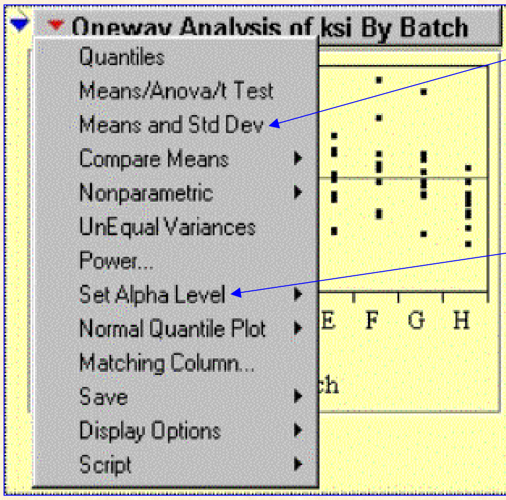
90% CI

Confidence Intervals

Parameter	Estimate	Lower CI	Upper CI	1-Alpha
Mean	162.6625	152.6958	172.6292	0.990
Std Dev	33.77324	27.97804	42.29352	

99% CI

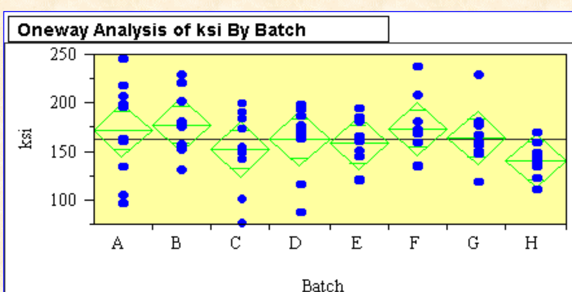
• One can also compute confidence intervals for a number of different population means using Fit Y by X.



The indicated menu option computes a population mean confidence interval for each level of the Nominal factor

The indicated menu option sets the confidence level for the confidence intervals. 95% is the default

Example: Using the dataset Al_Li.JMP calculate a 99% confidence interval for each of the population Batch means.



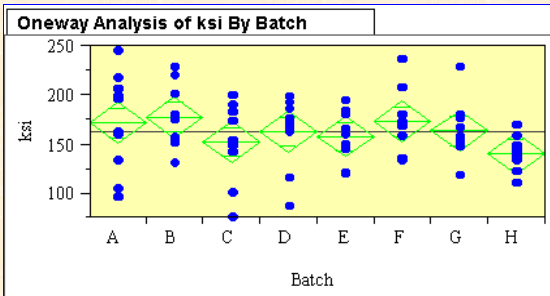
The vertical tips of the means diamonds represent 99% confidence intervals

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 99%	Upper 99%
A	10	172.400	10.517	144.57	200.23
B	10	177.600	10.517	149.77	205.43
C	10	152.200	10.517	124.37	180.03
D	10	163.100	10.517	135.27	190.93
E	10	158.000	10.517	130.17	185.83
F	10	173.300	10.517	145.47	201.13
G	10	164.000	10.517	136.17	191.83
H	10	140.700	10.517	112.87	168.53

Std Error uses a pooled estimate of error variance

Example: Using the dataset Al_Li.JMP calculate a 95% confidence interval for each of the population Batch means.



The vertical tips of the means diamonds now represent 95% confidence intervals

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
A	10	172.400	10.517	151.43	193.37
B	10	177.600	10.517	156.63	198.57
C	10	152.200	10.517	131.23	173.17
D	10	163.100	10.517	142.13	184.07
E	10	158.000	10.517	137.03	178.97
F	10	173.300	10.517	152.33	194.27
G	10	164.000	10.517	143.03	184.97
H	10	140.700	10.517	119.73	161.67

Std Error uses a pooled estimate of error variance

Hypothesis Testing: Quantifying Risk

General Idea behind Hypothesis Testing

• Recall that a confidence interval is constructed from sample data and provides a set of plausible values for the true (and unknown) mean of the population from which the sample is drawn.

• In ***hypothesis testing***, we suggest plausible values for the mean and then use data to see if these values make sense.

• More precisely, we construct hypotheses about the unknown mean, then collect data and, using statistical analysis, quantify risks of incorrect decisions relative to the hypotheses.

- We will discuss hypothesis testing in the context of testing for the mean of a single normal population. We shall see later that the **F-Ratio** (F test) hypothesis test can be used to determine if several population means are equal.

- In fact, hypothesis testing has very broad applications. It is a fairly involved statistical topic and we will only cover the basics.

- An hypothesis test requires a statement of our conjecture as to the true value(s) of a population parameter, such as μ, we call this value the **null hypothesized value(s) denoted as H_o**

- In the case where we reject our null hypothesis as false, we require a statement of an **alternative hypothesis H_A**, which is automatically accepted if H_o is rejected

- We will begin with an example of a hypothesis test.

Example: Average Lumen Level for CFL's

- We are interested in determining if the average lumen value for a lot of compact fluorescent lights is close to the specification target of 3500 lumens.

- A technician has measured the intensities for each of 7 suitably aged lamps from the specified lot:

3200, 3400, 4000, 3700, 2500, 3400, and 3700.

- We need to determine if the lot average is significantly distant from the target of 3500 lumens. We will employ a **one-sample t test** to determine if we are **significantly** off target.

• For this data,

$$\bar{X} = 3414.3, \quad S = 481.1$$

• The question we are attempting to address is whether or not the distance between 3414.3 and 3500 is sufficient to indicate that the true unknown mean for the lot may be different from the target value of 3500. Recall that we talk about statistical distance in terms of standard deviation or standard error.

• Our null hypothesis is:

$$H_o : \mu = 3500$$

• Our alternative hypothesis is:

$$H_A : \mu \neq 3500$$

• Note that values of $\bar{X}$ that are either much larger or smaller than 3500 would cause us to doubt H_0.

Underlying Theory

• Recall that $t = \frac{\bar{X} - \mu}{S/\sqrt{n}}$ has a **t distribution** on n-1 df, where **S** is the estimate of **σ** and n-1 is the df for **S**.

• So, under the assumption stated in H_0 that $\mu = 3500$:

$$\frac{\bar{X} - 3500}{S/\sqrt{n}}$$

follows a **t distribution** on 6 df.

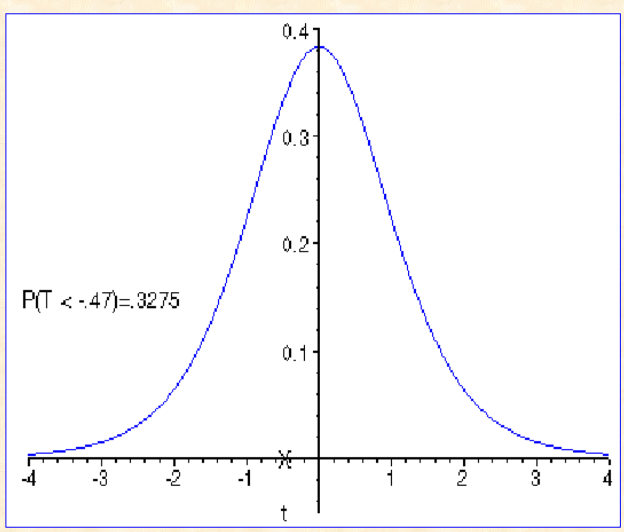
• This quantity, evaluated for our sample, is called the **test statistic**:

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{3414.1 - 3500}{481.1/7} = -0.4714$$

• Since large or small values of $\bar{X}$ correspond to large or small values of the statistic *t*, **we reject H_0 for values of *t* that are out in the tails of the distribution.**

• Notice that **we have just evaluated the distance between the target and the sample average in terms of standard errors** (since we are dealing with averages). We have found that the sample average is about 0.47 standard errors less than the target of 3500. Is this a significant distance?

• Since our t statistic behaves according to **Student's t distribution** with *n-1* degrees of freedom, we can use the t distribution to determine whether or not this is an unusually large t value. The density for a t distribution on 6 df looks as follows; the t value of -0.47 is marked with an "X".



• This value is not sufficiently extreme to cause us to doubt H_0.

P-Values

• Throughout the course we have referred to the **p-values** as a means by which we can judge the significance of main effects and interaction terms in our models. We now explain the concept behind **p-values.**

• In fact, the probability of obtaining a value of t at least as extreme as -.47, if H_0 is true, is the area under the t distribution from negative infinity to -.47 and from .47 to positive infinity. This area is .665 (.3275 in each tail). This value, the probability that, if H_0 is true, we obtain a value of the test statistic as extreme as what was in fact observed, is called the **p-value** for the test.

• **P-values are** routinely reported on computer output. (Prob > F in our one-way ANOVA printout, for example, is a **p-value** for the test that the factor level means are all identical.)

• Sufficiently extreme values to reject H_0 in the lumens example would be in the tails of the distribution. These would be values of *t* for which the **p-values**, or Prob > |t|, is small, say below .10.

• A p-value of .10 in a t-test for a single mean would indicate that we would only expect to see such an extreme value about 10% of the time, if the mean were as hypothesized.

• **The smaller the p-value, the more significant the result!**

Significance Levels

• We can also think of setting our p-value in advance, and defining which values of the test statistic will lead to rejection of H_0. Here we would reject if, for some small value of α,

$$\boxed{|t| > t_{n-1,1-\alpha/2}}$$

• If we reject H_0 using this **rejection region**, then we would reject H_0 only a proportion a of the time when H_0 is true.

• The value α is called the **significance level** of the test.

• Suppose we had set out to conduct a test of H_0 at the 0.05 significance level. Then we could look up $t_{6, .025}$ in the t tables. This gives $t_{6, 0.025} = 2.447$. We would then have noted that the absolute value of -.47 does not exceed 2.447, and so we would not have rejected of H_0 at a 0.05 significance level.

• The test we have conducted in this example is called a **one-sample t-test**.

• A one sample t test can be performed in JMP using the **Distribution of Y** platform.

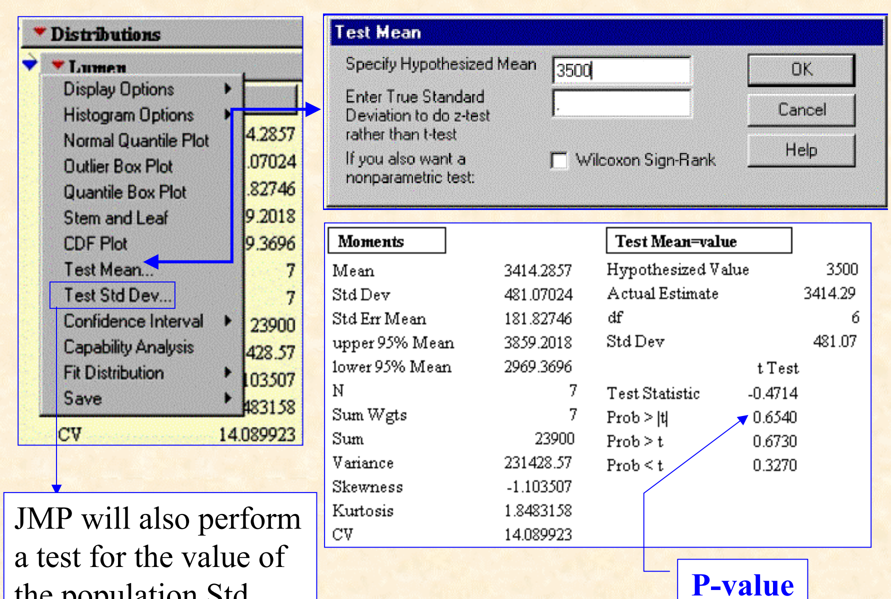
• Enter Distribution of Y with the variable containing your data entered as Y

• Under the red button on the Display header you will notice a a menu option, **"Test Mean Value".**

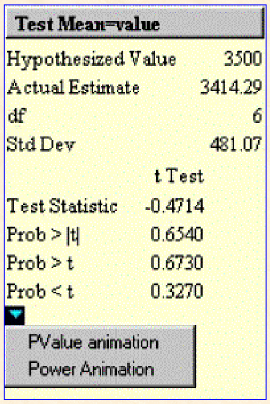
• A dialogue window appears and asks for your hypothesized mean value, that is your hypothesized value for the true mean.

• Click OK and a text report of your test appears below.

• JMP provides the p-value for the test rather and lets the user make the decision to reject or fail to reject the null hypothesis.



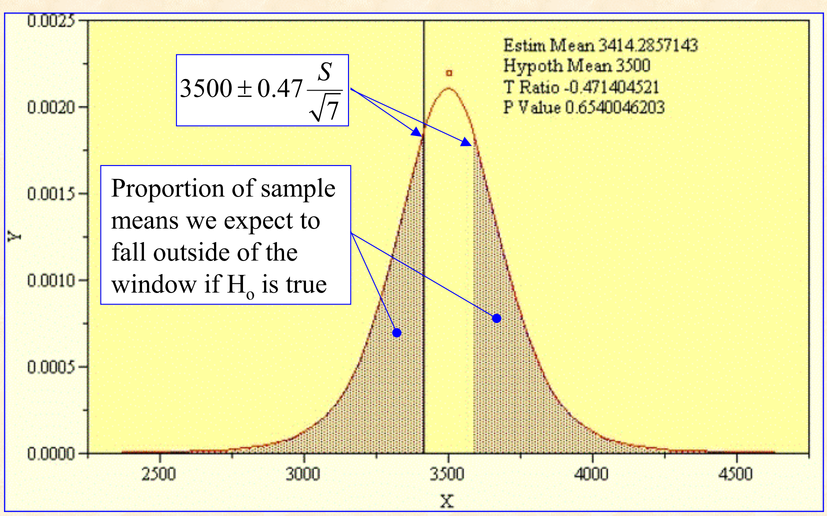
JMP will also perform a test for the value of the population Std. Dev. σ



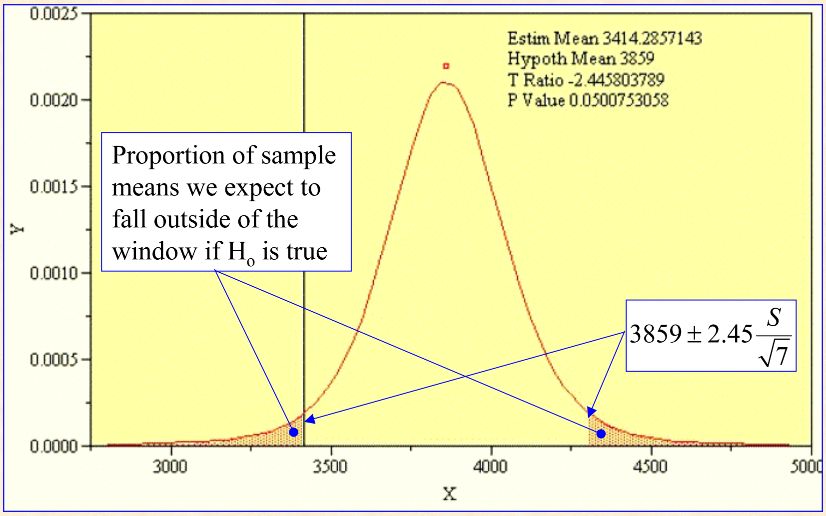
Test Mean=value	
Hypothesized Value	3500
Actual Estimate	3414.29
df	6
Std Dev	481.07
	t Test
Test Statistic	-0.4714
Prob > |t|	0.6540
Prob > t	0.6730
Prob < t	0.3270

PValue animation
Power Animation

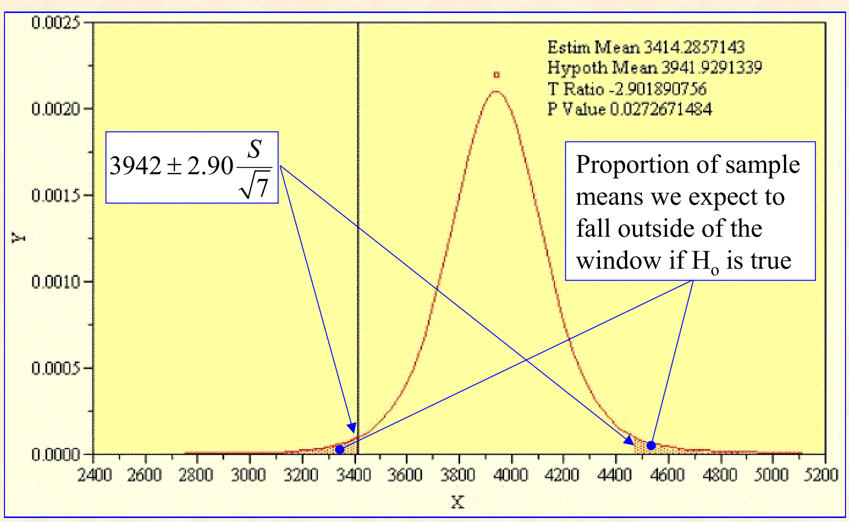
When the analysis appears for the One Sample t-test, a red button shows up at the bottom of the results display. If you depress the red button you are given two menu options for graphical displays. For the present we select **PValue animation**, which allows to graphically interpret the meaning of a p-value.



The p-value represents the proportion of mean values for X, that would fall outside of the window specified by the T ratio if μ=3500 (H_o). **We want this proportion to be < 0.05.**



If our null hypothesized mean had been 3859, then we would be on the edge of the rejection region because the **p-value** is approximately equal to 0.05, our significance level.



Suppose we had specified H_o:μ=3942, then our p-value would be 0.027 < 0.05. We expect less than 3% of sample means to fall outside of the observed window, if the null is true. **We strongly reject H_o.**

Statistical Decision Making

• Hypothesis testing requires a binary decision as to whether the data supports our hypothesis (called the null hypothesis) or not.
• Binary, accept or reject, decisions have four possible results:

• The **True Positive** -- the null hypothesis is truly false and is correctly rejected as false.

• The **False Positive** -- the null hypothesis is true, but is erroneously rejected as false (called a **Type I error**).

• The **True Negative** -- the null hypothesis is true and is correctly accepted as true.

• The **False Negative** -- the null hypothesis is truly false but is erroneously accepted as true (called a **Type II error**).

• A test that results in lots of **false positives** is virtually useless since a rejection is most likely an error.

• A test which results in lots of **false negatives** is dangerous since the customer now receives poor quality product, which is missed by the quality tests or checks performed by the producer.

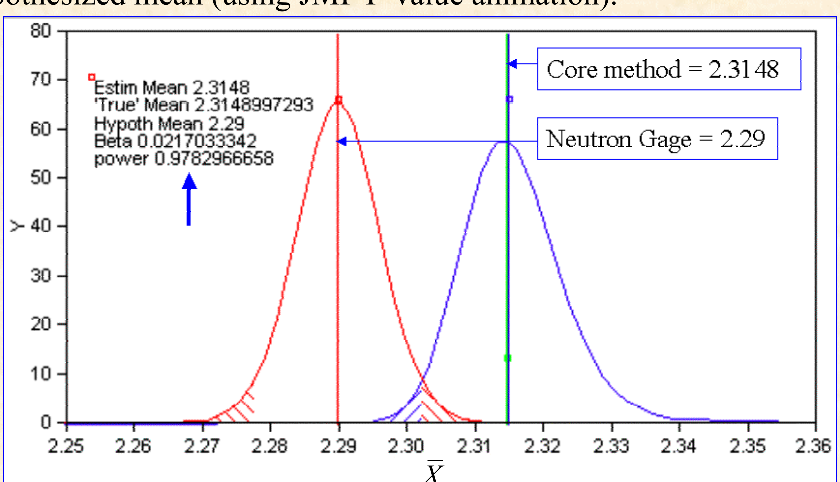
• Any analytic method (including hypothesis testing) will result in all four of the outcomes given enough opportunity. No test or analytic method is fool proof.

•Hypothesis testing attempts to keep the errors to a reasonable minimum. The risk of a false positive is controlled by the **significance level**.

• False negatives are controlled by designing tests with sufficient **power** to detect that the null hypothesis is false. We shall see how this is done later on.

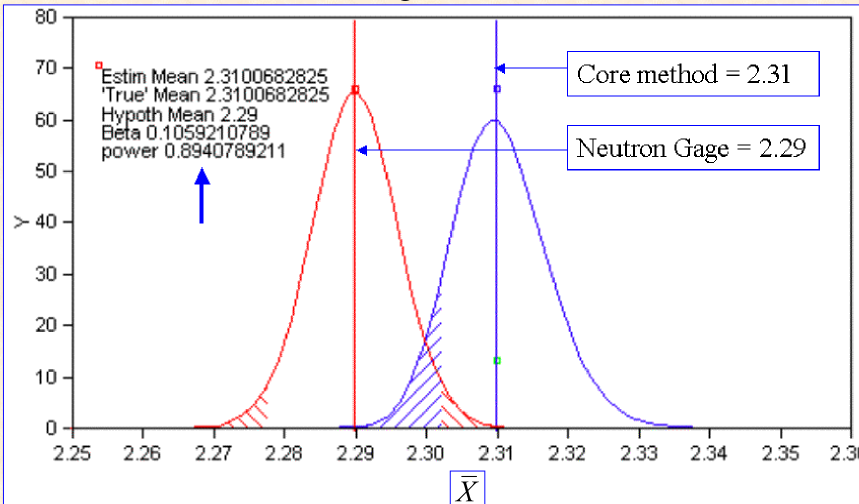
Case Study: An important characteristic of hot melt asphalt pavement is the specific gravity/density of the mix that is applied to the surface. Two methods exist to measure density/specific gravity. One method is a wet chemical method that requires taking a sample of cores from the pavement and determining the density/specific gravity in the laboratory. The other method is nondestructive and is based upon the scattering of a neutron radiation source applied to the pavement. For 30 cores from a paving project the neutron gage first measured the density/specific gravity in situ, then the cores were taken to the laboratory and the density/specific gravity measured chemically for each of the cores. The DOT wishes to use the neutron gages because they are quick and less expensive. Paving contractors believe that a substantial bias exists between the two methods and are worried that too much pavement will be declared "out of spec." For the 30 core areas, the neutron gage measured an average specific gravity of 2.29.

The power for a one sample t test using the neutron gage mean as the hypothesized mean (using JMP P-value animation).



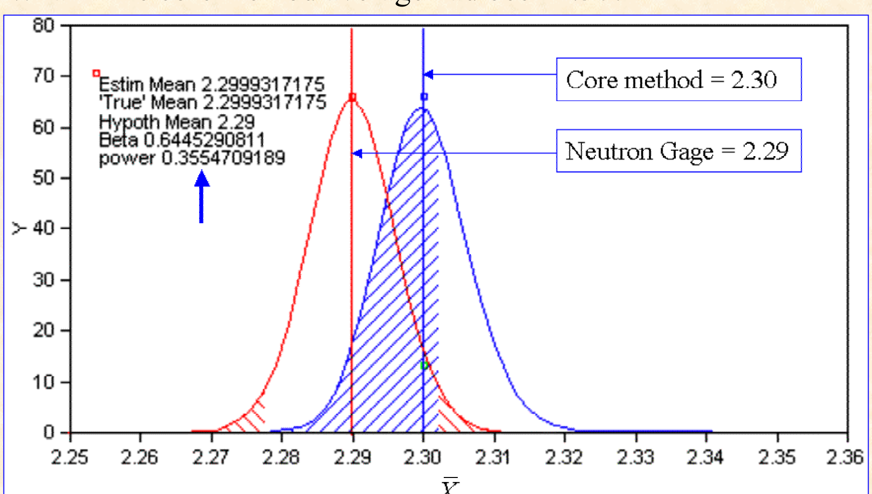
N	Std Err Mean	Alpha
30	0.00604	0.05

What if the core method average had been 2.31?



N	Std Err Mean	Alpha
30	0.00604	0.05

What if the core method average had been 2.30?



N	Std Err Mean	Alpha
30	0.00604	0.05

Improving the Power of a Hypothesis Test

- Power represents the ability of a hypothesis test to detect true differences between null hypothesized values and alternative hypothesized values.

- Power is analogous to the resolution of a microscope. The higher the power, the smaller the features that can be distinguished.

- The sample size *n* is commonly adjusted in order to achieve a hypothesis test with a desired level of power.

- Recall that the standard deviation of an average is estimated by $\boxed{\sigma/\sqrt{n}}$ By increasing the sample size, we can increase the power of our test since the statistical distance between the null and alternative values is increased by a factor of root *n*.

- Example. Suppose we wish to detect a mean shift of 200 lumens in a lamp manufacturing process and σ = 200. If samples of size $n = 10$ are collected, then the standard error of the sample mean is

$$\boxed{200/\sqrt{10} = 63.25}$$

- So a 200 lumens difference is 3.2 standard errors.

- If $n = 100$ then the standard error is 20 lumens, and the 200 lumen difference is 10 standard errors. We could easily detect a shift of this magnitude.

- **Power is the probability of detecting a true difference between the null and alternative values of the population mean, when the alternative value is the true mean.** We are never absolutely certainty of detecting a true difference.

• For a one-sample t test, here is the formula one uses to calculate the sample size *n* required to achieve a desired power. Let

α = P(false positive) = significance level of the test
β = P(false negative), so power = 1 - β
Δ = a specified difference we wish to detect
σ^2 = population variance

• The required sample size is

$$\boxed{n = \left(\frac{Z_{1-\alpha/2} + Z_{1-\beta}}{\Delta} \right)^2 \sigma^2}$$

where $Z_{1-\alpha/2}$ and $Z_{1-\beta}$ are percentiles of the standard normal distribution.

Example: Suppose α = 0.05, β = 0.05, power=0.95, Δ = 200, and σ = 200. Then

$$\boxed{n = \left(\frac{Z_{0.975} + Z_{0.95}}{\Delta} \right)^2 \sigma^2 = \left(\frac{1.96 + 1.645}{200} \right)^2 200^2 = 13}$$

• The power of a test can also be adjusted by changing the significance level. The smaller the significance level, then the higher the power of the test for a given sample size and specified difference.

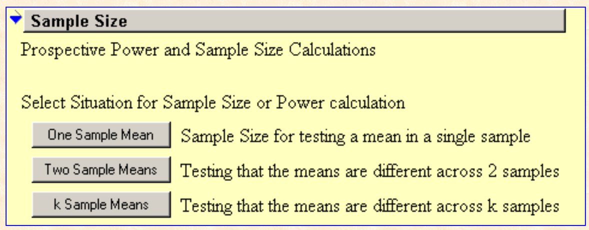
• The JMP software has an excellent facility that allows users to generate **Power Curves** from which sample size determinations can be made for any particular experimental application.

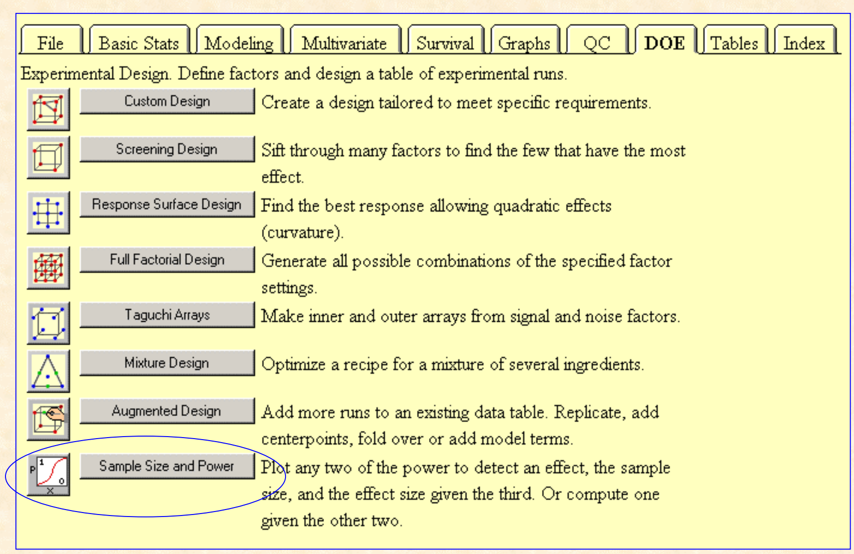
• The Power Curve routine is in the **DOE Platform**

• Select the DOE platform by clicking on the DOE tab in the JMP Starter screen.

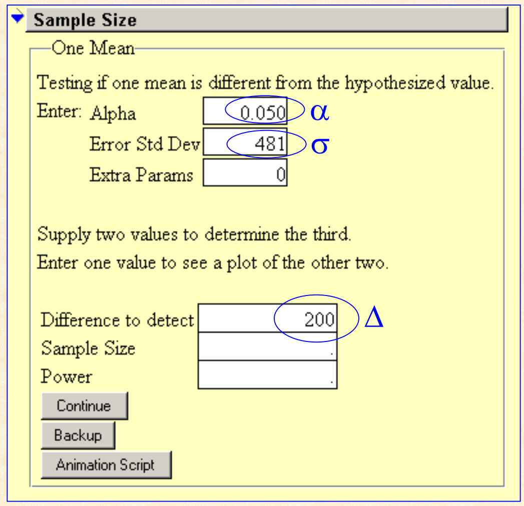
• Select the Sample Size and Power menu option from the DOE platform (see picture on next slide).

• There are 3 options for Power calculations depending upon the application.

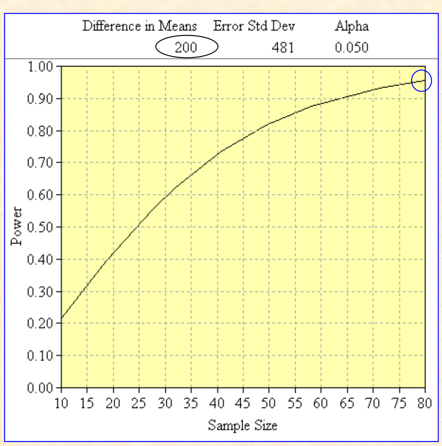


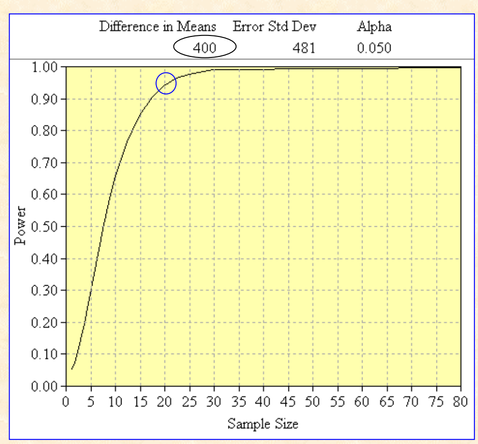


Example: Fitting Power Curves in JMP using the Lumens data (One Sample Mean case). To generate power curves simply fill in the desired difference to detect. You must generate separate curves if a range of differences is to be evaluated.



Power curves for two different delta (difference) values





Sample size of 80 required in order to have a Power = 0.95 to detect 200 lumen shift.

Sample size of 20 required in order to have a Power = 0.95 to detect 400 lumen shift.

Confidence Intervals Versus Hypothesis Tests

• Although it is not obvious, confidence intervals and hypothesis tests are mathematically equivalent. Each provides a slightly different perspective. However, for a given situation, the inferences drawn from a hypothesis test are equivalent to those inferences based upon a confidence interval.

• A confidence interval provides a set of possible values for the true unknown population parameter. One gains a sense of the precision in the inferences made about the possible value of the population parameter.

• A hypothesis test provides a p-value with which to assess the observed significance or risk associated with making inferences about the possible value of the population parameter.

• Statistical software often provides both confidence intervals and hypothesis tests for population parameters.