



- One-way analysis of variance (ANOVA)
 - F-test and t-test

1. Mark Anderson and Pat Whitcomb (2000), *DOE Simplified*, Productivity Inc., Chapter 2.

One-Factor Tutorial

1

Welcome to Stat-Ease's introduction to analysis of variance (ANOVA). The objective of this PowerPoint presentation and the associated software tutorial is two-fold. One, to introduce you to the Design-Expert software before you attend a computer-intensive workshop, and two, to review the basic concepts of ANOVA. These concepts are presented in their simplest form, a one-factor comparison. This form of design of experiment (DOE) can compare two levels of a single factor, e.g. two or more vendors. Or perhaps to compare the current process versus a proposed new process. Analysis of variance is based on F-testing and is typically followed up with pairwise t-testing.

Turn the page to get started!

If you have any questions about these materials, please e-mail StatHelp@StatEase.com or call 612-378-9449 and ask for statistical support.



General One-Factor Tutorial (Tutorial Part 1 – The Basics)

Instructions:

1. Turn on your computer.
2. Start Design-Expert® version 7.
3. Work through Part 1 of the “General One-Factor Tutorial.”
4. Don't worry yet about the actual plots or statistics. Focus on learning how to use the software.
5. When you complete the tutorial return to this presentation for some explanations.

First use the computer to complete the General One-Factor Tutorial (Part 1 – The Basics) using Design-Expert software. This tutorial will give you complete instructions to create the design and to analyze the data. Some explanation of the statistics is provided in the tutorial, but more complete explanation is given in this PowerPoint presentation.

Start the Design-Expert tutorial now. After you've completed the tutorial, return here and continue with this slide show.

Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Model	2212.11	2	1106.06	12.57	0.0006
A	2212.11	2	1106.06	12.57	0.0006
Pure Error	1319.50	15	87.97		
Cor Total	3531.61	17			

In this tutorial
Pure Error SS = Residual SS

Model Sum of Squares (SS): $SS_{\text{Model}} = \text{Sum of squared deviations due to treatments.}$

$$= 6(153.67-162.72)^2 + 6(178.33-162.72)^2 + 6(156.17-162.72)^2 = 2212.11$$

Model DF (Degrees of Freedom): The deviation of the treatment means from the overall average must sum to zero. The degrees of freedom for the deviations is, therefore, one less than the number of means.

$$DF_{\text{Model}} = 3 \text{ treatments} - 1 = 2.$$

Model Mean Square (MS): $MS_{\text{Model}} = SS_{\text{Model}}/DF_{\text{Model}} = 2212.11/2 = 1106.06$

Pure Error Sum of Squares: $SS_{\text{Pure Error}} = \text{Sum of squared deviations of response data points from their treatment mean.}$

$$= (160-153.67)^2 + (150-153.67)^2 + \dots + (156-156.17)^2 = 1319.50$$

Pure Error DF: There are $(nt - 1)$ degrees of freedom for the deviations within each treatment.

$$DF_{\text{Pure Error}} = \sum(nt - 1) = (5 + 5 + 5) = 15$$

Pure Error Mean Square: $MS_{\text{Pure Error}} = SS_{\text{Pure Error}}/DF_{\text{Pure Error}} = 1319.50/15 = 87.97$

F Value: Test for comparing treatment variance with error variance.

$$F = MS_{\text{Model}}/MS_{\text{Pure Error}} = 1106.06/87.97 = 12.57$$

Prob > F: Probability of observed F value if the null hypothesis is true. The probability equals the tail area of the F-distribution (with 2 and 15 DF) beyond the observed F-value. Small probability values call for rejection of the null hypothesis.

Cor Total: Totals corrected for the mean. $SS = 3531.61$ and $DF = 17$

Corrected total sum of squared deviations from the grand mean = Between treatment sum of squares + Within treatment sum of squares

$SS_{\text{Corrected total}}$ = SS_{Model} + $SS_{\text{Residuals}}$

$$\sum_{t=1}^k \sum_{i=1}^{n_t} (y_{ti} - \bar{y})^2 = \sum_{t=1}^k n_t (\bar{y}_t - \bar{y})^2 + \sum_{t=1}^k \sum_{i=1}^{n_t} (y_{ti} - \bar{y}_t)^2$$

$$F = \frac{SS_{\text{Model}} / df_{\text{Model}}}{SS_{\text{Residuals}} / df_{\text{Residuals}}} = \frac{MS_{\text{Model}}}{MS_{\text{Residuals}}}$$

where: SS \equiv "Sum of Squares"
MS \equiv "Mean Square"

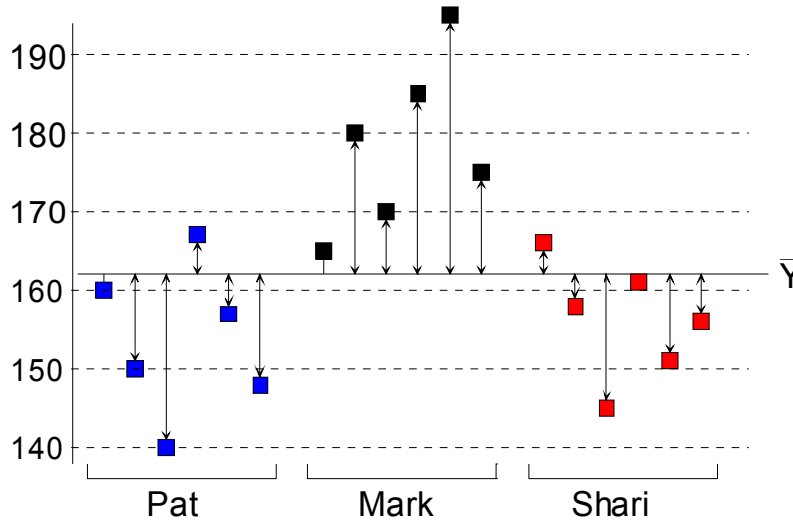
Let's breakdown the ANOVA a bit further. The starting point is the sum of squares: It becomes convenient to introduce a mathematical partitioning of sums of squares that are analogous to variances. We find that the total sum of squared deviations from the grand mean can be broken into two parts: that caused by the treatment (factor levels) and the unexplained remainder which we call residuals or error.

Note on slide:

k = # treatments

n = # replicates

$$SS_{\text{corrected total}} = \sum_{t=1}^3 \sum_{i=1}^{6,6,6} (y_{ti} - \bar{y})^2 = 3531.61$$

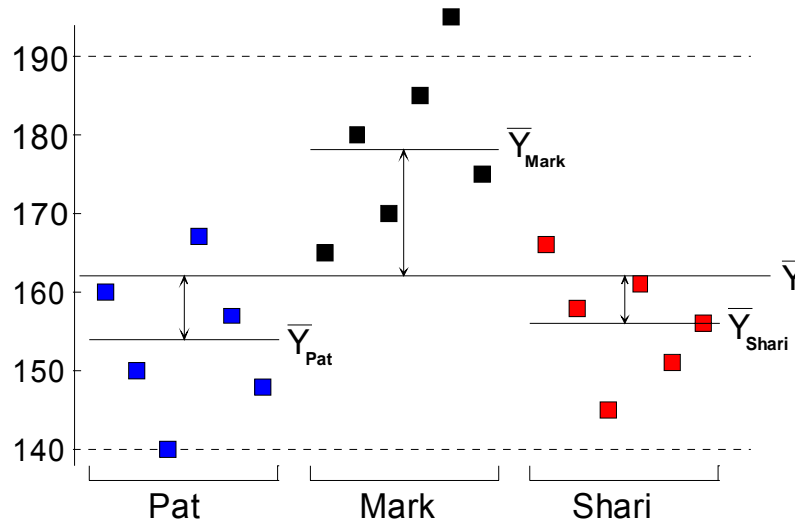


One-Factor Tutorial

5

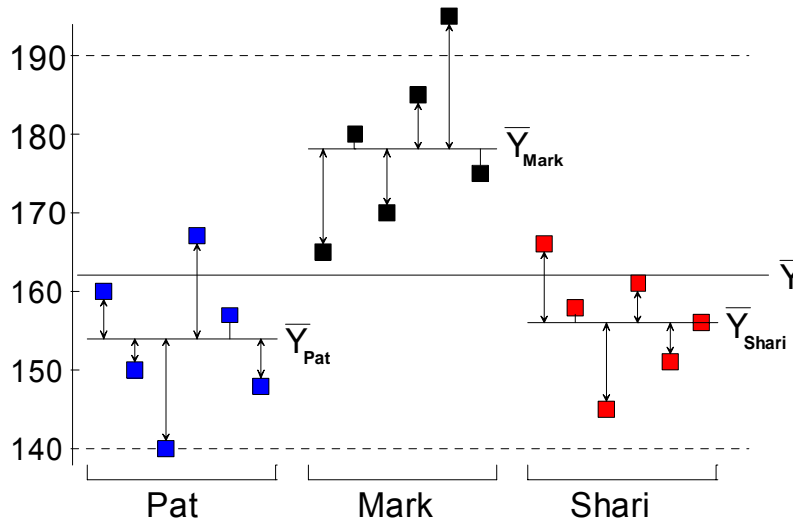
This is the total sum of squares in the data. It is calculated by taking the squared differences between each bowling score and the overall average score, then summing those squared differences. Next, this total sum of squares is broken down into the treatment (or Model) SS and the Residual (or Error) SS. These are shown on the following pages.

$$SS_{\text{model}} = \sum_{t=1}^3 n_t (\bar{y}_t - \bar{y})^2 = 2212.11$$



This is a graph of the between treatment SS. It is the squared difference between the treatment means and the overall average, weighted by the number of samples in that group, and summed together.

$$SS_{\text{residuals}} = \sum_{t=1}^3 \sum_{i=1}^{6,6,6} (y_{ti} - \bar{y}_t)^2 = 1319.50$$



One-Factor Tutorial

7

This is the within treatment SS. Think of this as the differences due to normal process variation. It is calculated by taking the squared difference between the individual observations and the treatment mean, and the summed across all treatments. Notice that this calculation is independent of the overall average – if any of the means shift up or down, it won't effect the calculation.

The F statistic that is derived from the mean squares is converted into its corresponding p-value. In this case, the p-value tells the probability that the differences between the means of the bowler's scores can be accounted for by random process variation.

As the p-value decreases, it becomes less likely the effect is due to chance, and more likely that there was a real cause. In this case, with a p-value of 0.0006, there is only a 0.06% chance that random noise explains the differences between the average bowlers scores. Therefore, it is highly likely that the difference detected is a true difference in Pat, Mark and Shari's bowling skills.

The p-value is a statistic that is consistent among all types of hypothesis testing so it is helpful to try to understand it's meaning. This was reviewed in the tutorial and will be covered much more extensively in the Stat-Ease workshop. For now, let's just lay out some general guidelines:

- If $p < 0.05$, then the model (or term) is statistically significant.
- If $0.05 < p < 0.10$, then the model might be significant, you will have to decide based on subject matter knowledge.
- If $p > 0.10$, then the model is not significant.

Std. Dev	9.38	R-Squared	0.6264
Mean	162.72	Adj R-Squared	0.5766
C.V.	5.76	Pred R-Squared	0.4620
PRESS	1900.08	Adeq Precision	6.442

The second part of the ANOVA provides additional summary statistics. The value of each of these will be discussed during the workshop. For now, simply learn their names and basic definitions.

Std Dev: Square root of the Pure (experimental) Error. (Sometimes referred to as Root MSE)
 $= \text{SqRt}(87.97) = 9.38$

Mean: Overall mean of the response = $[160 + 150 + 140 + \dots + 156] / 18 = 162.72$

C.V.: Coefficient of variation, the standard deviation as a percentage of the mean.
 $= (\text{Std Dev}/\text{Mean})(100) = (9.38/162.72)(100) = 5.76\%$

Predicted Residual Sum of Squares (PRESS): A measure of how this particular model fits each point in the design. The coefficients are calculated without the first point. This model is then used to estimate the first point and calculate the residual for point one. This is done for each data point and the squared residuals are summed.

R-Squared: The multiple correlation coefficient. A measure of the amount of variation about the mean explained by the model.
 $= 1 - [\text{SS}_{\text{Pure Error}} / (\text{SS}_{\text{Model}} + \text{SS}_{\text{Pure Error}})] = 1 - [1319.50 / (2212.11 + 1319.50)] = 0.6264$

Adj R-Squared: Adjusted R-Squared. Measure of the amount of variation about the mean explained by the model adjusted for the number of parameters in the model.
 $= 1 - (\text{SS}_{\text{Pure Error}} / \text{DF}_{\text{Pure Error}}) / [(\text{SS}_{\text{Model}} + \text{SS}_{\text{Pure Error}}) / (\text{DF}_{\text{Model}} + \text{DF}_{\text{Pure Error}})]$
 $= 1 - (1319.50/15) / [(2212.11 + 1319.50) / (2 + 15)] = 0.5766$

Pred R-Squared: Predicted R-Squared. A measure of the predictive capability of the model.
 $= 1 - (\text{PRESS} / \text{SS}_{\text{Total}}) = 1 - (1900.08 / 3531.61) = 0.4620$

Adeq Precision: Compares the range of predicted values at design points to the average prediction error. Ratios greater than four indicate adequate model discrimination.

Treatment Means (Adjusted, If Necessary)

	Estimated Mean	Standard Error
1-Pat	153.67	3.83
2-Mark	178.33	3.83
3-Shari	156.17	3.83

Estimated Mean: The average response for each treatment.

Standard Error: The standard deviation of the average is equal to the standard deviation of the individuals divided by the square root of the number of individuals in the average. $SE = 9.38/\sqrt{6} = 3.83$

This is a table of the treatment means and the standard error associated with that mean. As long as all the treatment sample sizes are the same, the standard errors will be the same. If the sample sizes differed, the SE's would vary accordingly.

Note that the standard errors associated with the means are based on the pooling of within treatment variances.

Conclusion from ANOVA:

There are differences between the bowlers that cannot be explained by random variation.

Next Step:

Run pairwise t-tests to determine which bowlers are different from the others.

The ANOVA does not tell us which bowler is best, it simply indicates that at least one of the bowlers is statistically different from the others. This difference could be either positive or negative. In a one-factor design, the ANOVA is followed by pairwise t-tests to gain more understanding of the specific results.

One-Factor Tutorial

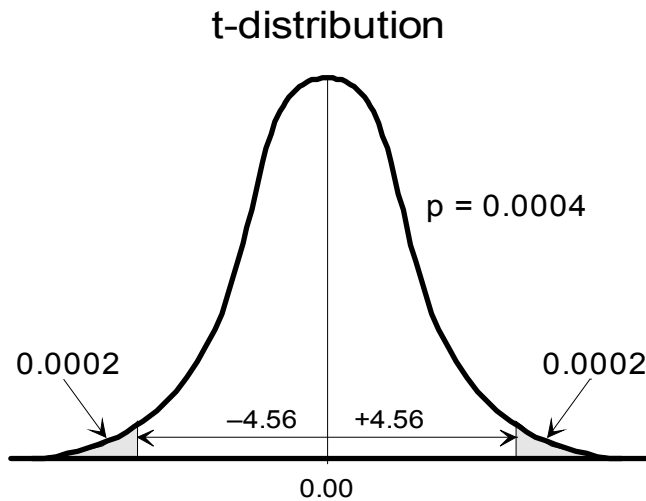
Treatment Comparisons via t-Tests

	Estimated Mean		Standard Error		
1-Pat	153.67		3.83		
2-Mark	178.33		3.83		
3-Shari	156.17		3.83		
Treatment	Mean Difference	DF	Standard Error	t for H ₀ Coeff=0	Prob > t
1 vs 2	-24.67	1	5.41	-4.56	0.0004
1 vs 3	-2.50	1	5.41	-0.46	0.6509
2 vs 3	22.17	1	5.41	4.09	0.0010

These t-tests are valid only when the null hypothesis is rejected during ANOVA!

This section is called post ANOVA because you need the protection given by the preceding analysis of variance before making specific treatment comparisons. In this case we did get a significant overall F test in the ANOVA. Therefore it's OK to see how each bowler did relative to the others. All possible pairs are shown with the appropriate t-test and associated probability.

Note that the difference between Pat and Mark is statistically significant, as is the difference between Mark and Shari. However, the difference between Pat and Shari is not statistically significant.



Can say Pat and Mark are different: $p < 0.05$
(1 vs 2: $t = -4.56$, $p = 0.0004$)

Likewise:

Cannot say Pat and Shari are different: $p > 0.10$
(1 vs 3: $t = -0.46$, $p = 0.6509$)

Can say Shari and Mark are different: $p < 0.05$
(2 vs 3: $t = 4.09$, $p = 0.0010$)

Here's a picture of the t-distribution. It's very much like the normal curve but with heavier tails. The p-value comes from the proportion of the area under the curve that falls outside plus and minus t.

The t-value for the comparison of Pat and Shari isn't nearly as large. In fact, it's insignificant ($p > 0.10$). But the t-value for Mark vs Shari is almost as high as that for Mark vs Pat, and therefore this difference is significant at the 0.05 level.



One-Factor Tutorial

Diagnostics Case Statistics

Std Ord	Actual Value	Predicted Value	Residual	Leverage	Internally Studentized Residual	Externally Studentized Residual	Influence on Fitted Value DFFITS	Cook's Distance	Run Ord
1	160.00	153.67	6.33	0.167	0.740	0.728	0.326	0.036	9
2	150.00	153.67	-3.67	0.167	-0.428	-0.416	-0.186	0.012	7
3	140.00	153.67	-13.67	0.167	-1.596	-1.693	-0.757	0.170	2
4	167.00	153.67	13.33	0.167	1.557	1.643	0.735	0.162	16
5	157.00	153.67	3.33	0.167	0.389	0.378	0.169	0.010	8
6	148.00	153.67	-5.67	0.167	-0.662	-0.649	-0.290	0.029	5
7	165.00	178.33	-13.33	0.167	-1.557	-1.643	-0.735	0.162	6
8	180.00	178.33	1.67	0.167	0.195	0.188	0.084	0.003	15
9	170.00	178.33	-8.33	0.167	-0.973	-0.971	-0.434	0.063	4
10	185.00	178.33	6.67	0.167	0.779	0.768	0.343	0.040	11
11	195.00	178.33	16.67	0.167	1.947	2.175	0.973	0.253	14
12	175.00	178.33	-3.33	0.167	-0.389	-0.378	-0.169	0.010	3
13	166.00	156.17	9.83	0.167	1.149	1.162	0.520	0.088	10
14	158.00	156.17	1.83	0.167	0.214	0.207	0.093	0.003	1
15	145.00	156.17	-11.17	0.167	-1.304	-1.338	-0.598	0.113	17
16	161.00	156.17	4.83	0.167	0.565	0.551	0.247	0.021	12
17	151.00	156.17	-5.17	0.167	-0.603	-0.590	-0.264	0.024	18
18	156.00	156.17	-0.17	0.167	-0.019	-0.019	-0.008	0.000	13
One-Factor Tutorial									14

Residual analysis is a key component to understanding the data. This is used to validate the ANOVA, and to detect any problems in the data. Although the easiest review of residuals is done graphically, Design-Expert also produces a table of case statistics that can be seen from the Diagnostics button, either from the View menu, or on the “Influential” portion of the tool palette.

See “Case Statistics – Definitions” on the following two slides, but these will also be covered in the workshop. Plots of the case statistics are used to validate our model.

Actual Value: The value observed at that design point.

Predicted Value: The value predicted at that design point using the current polynomial.

Residual: Difference between the actual and predicted values for each point in the design.

Leverage: Leverage of a point varies from 0 to 1 and indicates how much an individual design point influences the model's predicted values. A leverage of 1 means the predicted value at that particular case will exactly equal the observed value of the experiment, i.e., the residual will be 0. The sum of leverage values across all cases equals the number of coefficients (including the constant) fit by the model. The maximum leverage an experiment can have is $1/k$, where k is the number of times the experiment is replicated

Internally Studentized Residual: The residual divided by the estimated standard deviation of that residual. The number of standard deviations (s) separating the actual from predicted values.

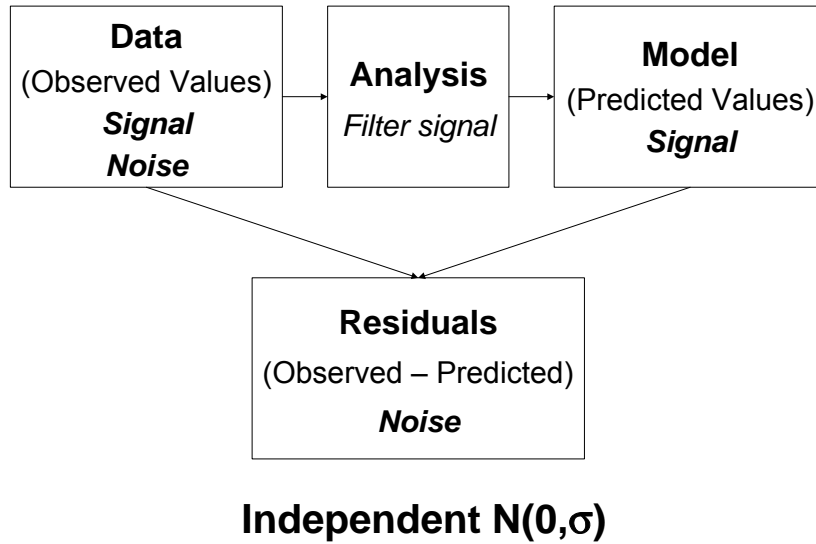
The case statistics will be explained in the workshop; don't sweat the details.

Externally Studentized Residual (outlier t value, R-Student): Calculated by leaving the run in question out of the analysis and estimating the response from the remaining runs. The t value is the number of standard deviations difference between this predicted value and the actual response. This tests whether the run in question follows the model with coefficients estimated from the rest of the runs, that is, whether this run is consistent with the rest of the data for this model. Runs with large t values should be investigated.

DFFITs: Measures the influence the i th observation has on the predicted value. It is the studentized difference between the predicted value with observation i and the predicted value without observation i . DFFITS is the externally studentized residual magnified by high leverage points and shrunk by low leverage points:

Cook's Distance: A measure of how much the regression would change if the case is omitted from the analysis. Relatively large values are associated with cases with high leverage and large studentized residuals. Cases with large D_i values relative to the other cases should be investigated. Large values can be caused by mistakes in recording, an incorrect model, or a design point far from the rest of the design points.

The case statistics will be explained in the workshop; don't sweat the details.



Examine the residuals to look for patterns that indicate something other than noise is present. If the residual is pure noise (it contains no signal) then the analysis is complete.

Additive treatment effects

One Factor: Treatment means adequately represent response behavior.

- Model F-test
- Lack-of-Fit
- Box-Cox plot

Independence of errors

Knowing the residual from one experiment gives no information about the residual from the next.

S Residuals
versus
Run Order

Studentized residuals $N(0, \sigma^2)$:

- Normally distributed
- Mean of zero
- Constant variance, $\sigma^2=1$

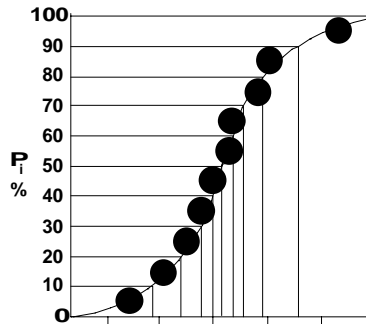
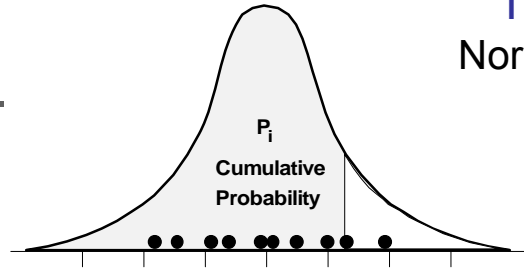
Normal Plot of
S Residuals

S Residuals
versus
Predicted

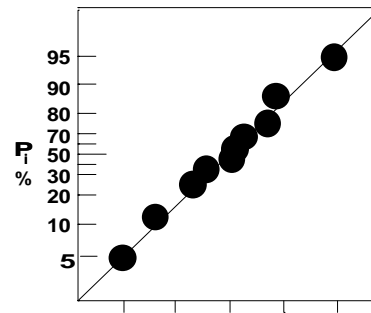
Check assumptions by plotting internally studentized residuals (S Residuals)!

Diagnostic plots of the case statistics are used to check these assumptions.

Test for Normality Normal Probability Plot



Ordinary Graph Paper



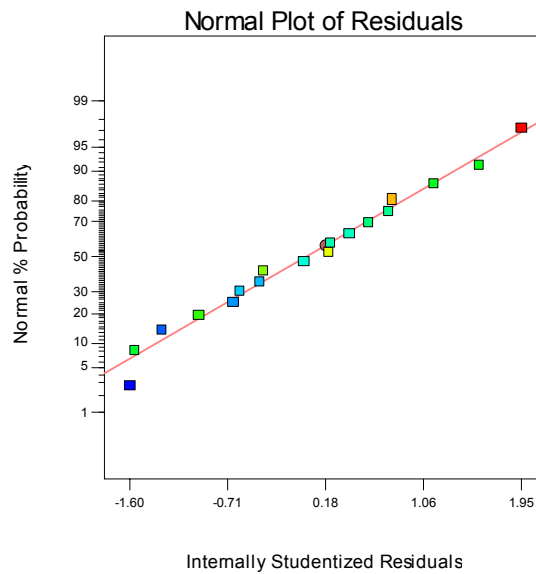
Normal Probability Paper

A major assumption is that the residuals are normally distributed which can be easily checked via normal probability paper. For example, assume that you've taken a random sample of 10 observations from a normal distribution (dots on top figure). The shaded area under the curve represents the cumulative probability (P) which translates to the chance that you will get a number less than (or equal to) the value on the response (Y) scale.

Now let's look at a plot of cumulative probability (lower left). In this example, the first point (1 out of 10) represent 10% of all the data. This point is plotted in the middle of the range, so P equals 5%. The second point represents an additional 10% of the available data so we put that at P of 15% and so on (3rd point at 25, 4th at 35, etc.). Notice how the "S" shape that results gets straightened out on the special paper at the right, which features a P axis that is adjusted for a normal distribution. If the points line up on the normal probability paper, then it is safe to assume that the residuals are normally distributed.

Design-Expert® Software
Score

Color points by value of
Score:
■ 140
■ 168
■ 195



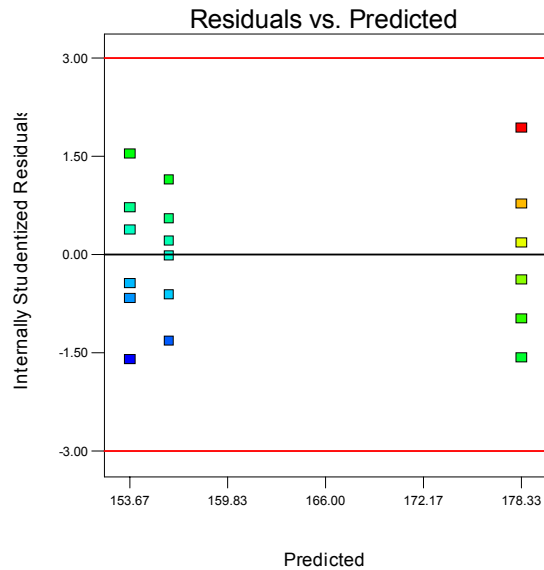
One-Factor Tutorial

20

The normal plot of residuals is used to confirm the normality assumption. If all is okay, residuals follow an approximately straight line, i.e., are normal. Some scatter is expected, look for definite patterns, e.g., "S" shape. The data plotted here exhibits normal behavior - - that's good

Design-Expert® Software
Score

Color points by value of
Score:
■ 140
■ 168
■ 195



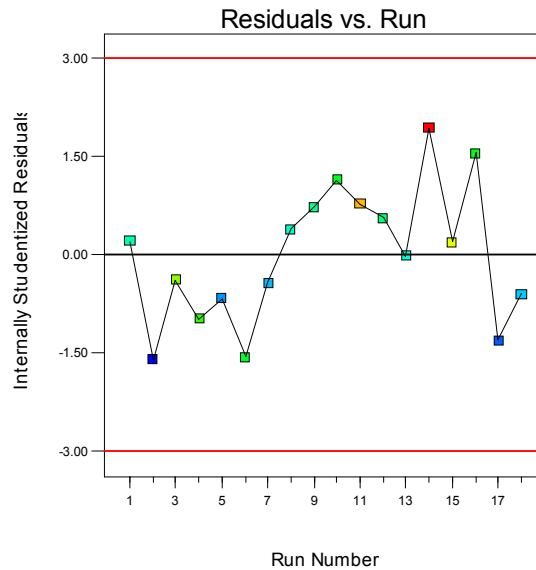
One-Factor Tutorial

21

The graph of studentized residuals versus predicted values is used to confirm the constant variance assumption. The size of the residuals should be independent of the size of the predicted values. Watch for residuals increasing in size as the predicted values increase, e.g., a megaphone shape. The data plotted here exhibits constant variance - - that's good.

Design-Expert® Software
Score

Color points by value of
Score:
■ 140
■ 168
■ 195



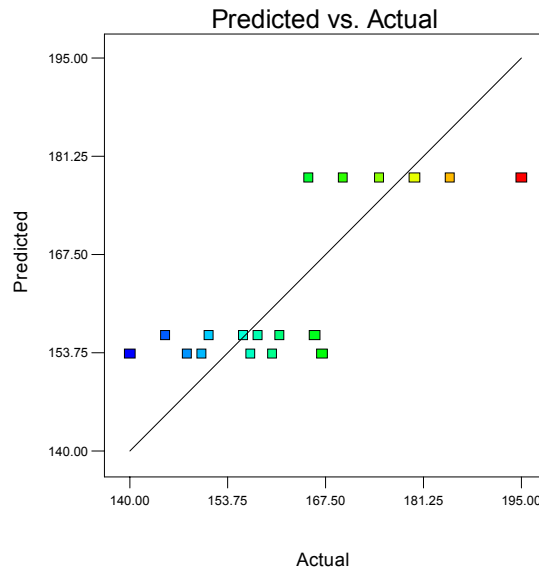
One-Factor Tutorial

22

If all is okay, there will be random scatter on the residuals versus run plot. Look for trends, which may be due to a time-related lurking variable. Also look for distinct patterns indicating autocorrelation. Randomization provides insurance against autocorrelation and trends. This data exhibits no time trends - - that's good.

Design-Expert® Software
Score

Color points by value of
Score:
■ 140
■ 168
■ 195

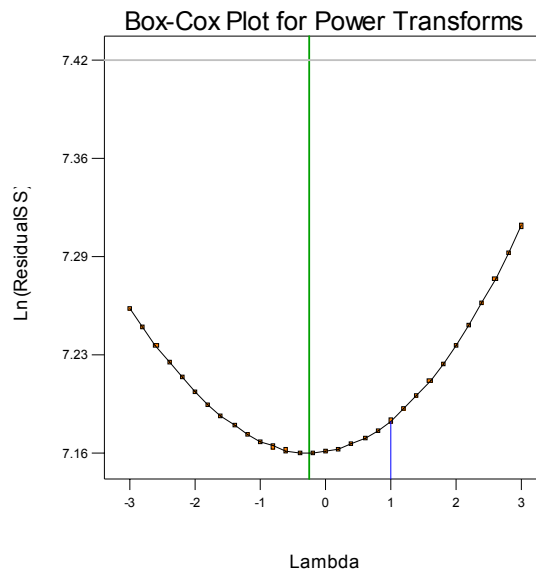


The predicted versus actual plot shows how the model predicts over the range of data. Plot should exhibit random scatter about the 45 degree line. Clusters of points above or below the line indicate problems of over or under predicting. The line is going through the middle of the data over the whole range of the data - - that's good. The scatter shows the bowling scores cannot be predicted very precisely.

Design-Expert® Software
Score

Lambda
Current = 1
Best = -0.24
Low C.I. = -4.59
High C.I. = 4.12

Recommend transform:
None
(Lambda = 1)



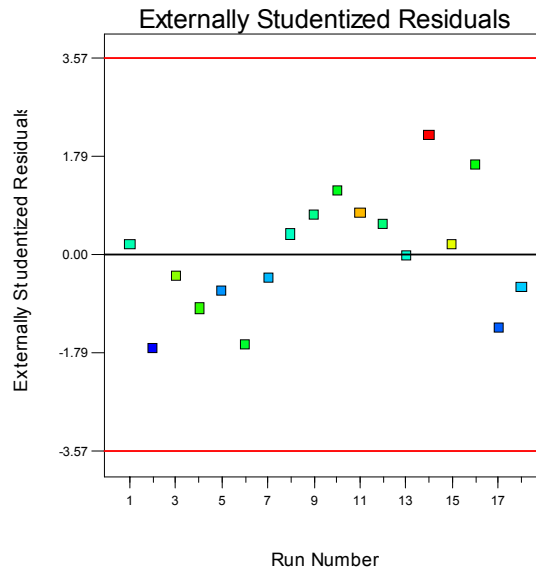
One-Factor Tutorial

24

The Box-Cox plot tells you whether a transformation of the data may help – transformations will be covered during the workshop. Without getting into the details, notice that it says “none” for recommended transform. That’s all you need to know for now.

Design-Expert® Software
Score

Color points by value of
Score:
■ 140
■ 168
■ 195



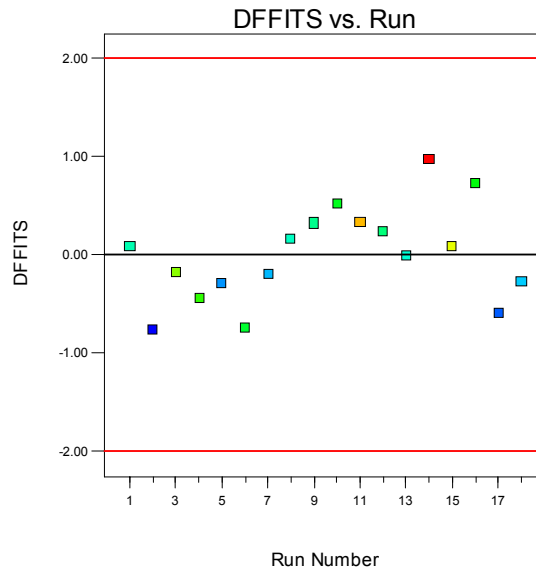
One-Factor Tutorial

25

The externally studentized residuals versus run to identifies model and data problems by highlighting outliers. Look for values outside the red limits. These 95% confidence limits are generally between 3 and 4, but vary depending on the number of runs. A high value indicates a particular run does not agree well with the rest of the data, when compared using the current model.

Design-Expert® Software
Score

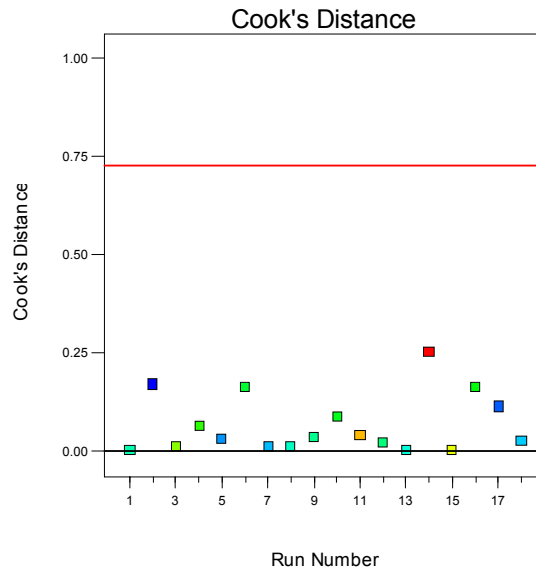
Color points by value of
Score:
■ 140
■ 168
■ 195



The DFFITS versus run is also used to highlight outliers that identify model and data problems. Look for values outside the ± 2 limits. A high value indicates a particular run does not agree well with the rest of the data, when compared using the current model.

Design-Expert® Software
Score

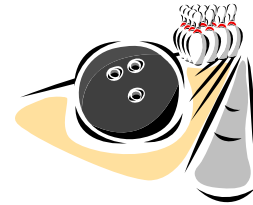
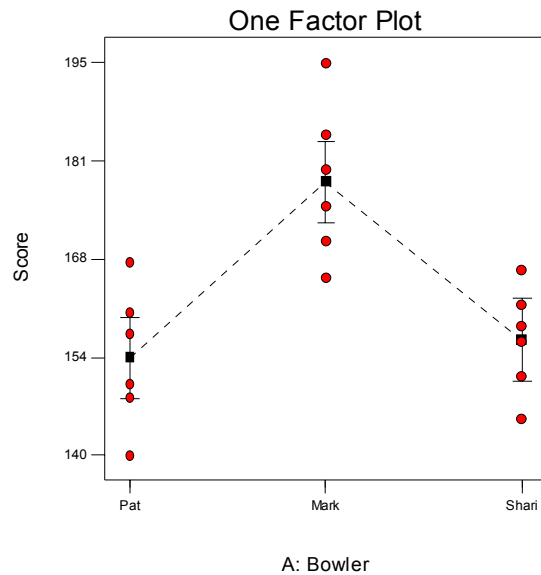
Color points by value of
Score:
■ 140
■ 168
■ 195



One-Factor Tutorial

27

Cook's distance versus run is also used to highlight outliers that identify model and data problems. Look for values outside the red limits. A high value indicates a particular run is very influential and the model will change if that run is not included.



We know from ANOVA that something happened in this experiment. The effect plot shows clear shifts which we could use in a mathematical model to predict the response.

Notice the bars. These show the least significant differences (LSD bars) at the 95 percent confidence level. If bars overlap, then the associated treatments are not significantly different. In this case the LSD's for Pat and Shari overlap, but Mark's stands out. Details on LSD calculation will be covered in the workshop.

Congratulations!

You have completed the One-Factor Tutorial.

The objective of this tutorial was to provide a refresher on the ANOVA process and an exposure to Design-Expert software. In-depth discussion on these topics will be provided during the workshop.

We look forward to meeting you!

If you have any questions about these materials, please e-mail them to StatHelp@StatEase.com or call 612-378-9449 and ask for statistical support.

If you want more of the basics read chapters 2 and 3 of *DOE Simplified*, by Mark Anderson and Pat Whitcomb (2000), Productivity Inc.