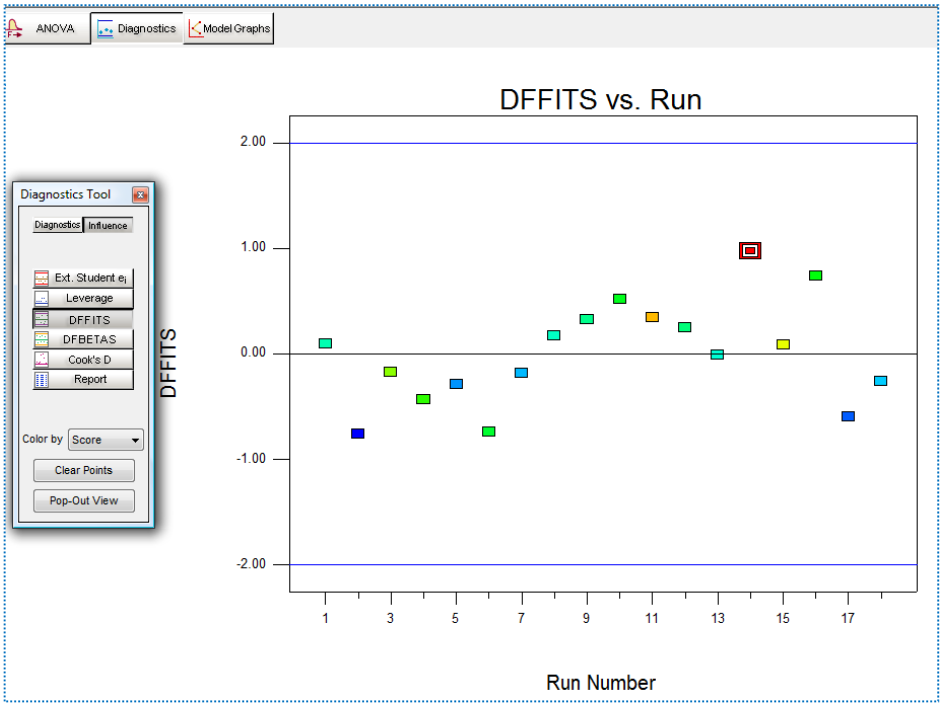
General One-Factor Tutorial
(Part 2 – Advanced Features)

Digging Deeper Into Diagnostics

If your bowling data is active in Design-Expert® software from Part 1 of this tutorial, continue on. If you exited the program, re-start it and use **File**, **Open Design** to open data file (**Bowling.dxp**). Otherwise, set up this data file as instructed in our General One-Factor Tutorial (Part 1 – The Basics). Then, under the **Analysis** branch (you may already be here) click the **Score** node and press the ***Diagnostics*** button.

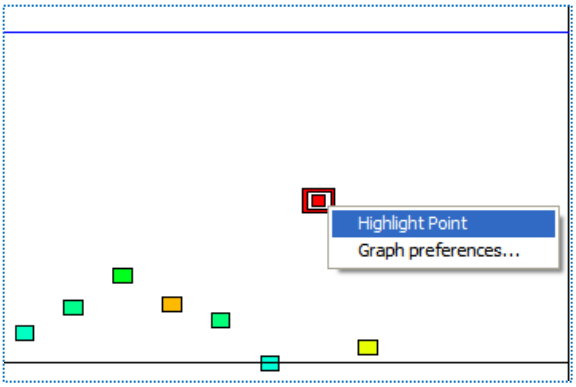
We're going to look at a new graph in the **Diagnostics Tool**. Click the **Influence** option on the Diagnostics Tool palette. Then click on **DFFITS**. This statistic, which stands for difference in fits, measures the change in each predicted value that occurs when that response is deleted. The larger the absolute value of DFFITS, the more it influences the fitted model. (For more details on this statistic and the related deletion diagnostic, DFBETAS, see our program Help or refer to Raymond Myers' *Classical and Modern Regression with Applications, 2nd Edition* (PWS Pub. Co., 1990).)



DFFITS graph – highest point clicked (your graph may differ due to random runs)

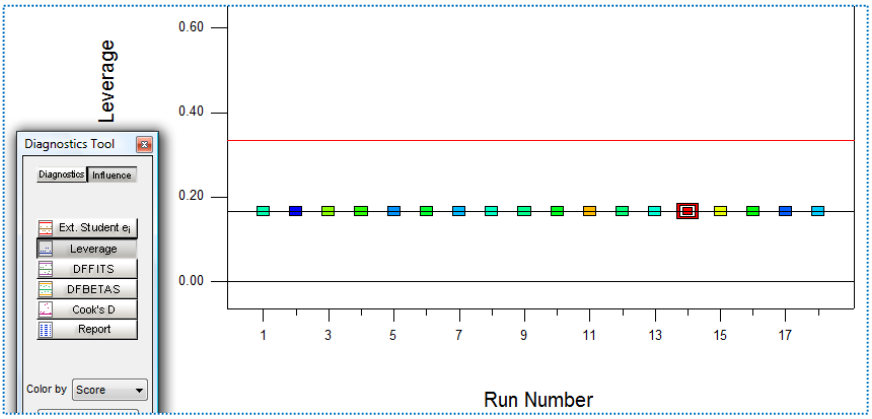
Notice that one point lies above the rest. (The pattern on your graph may differ from what we show here due to randomized run order, but this isn't a concern in this discussion.) The top-most point is Mark's high game, which earlier created controversy, particularly among competitors Pat and Shari. Mark's point falls far

below a relatively conservative high benchmark of plus-or-minus two for the DFFITS. So, taking all other diagnostics into consideration, we don't advise that this particular run be investigated further. Nevertheless, for purposes of learning how to use new Design-Expert software features, right-click Mark's top point with your mouse and select **Highlight Point** as shown below.



Highlighting a point

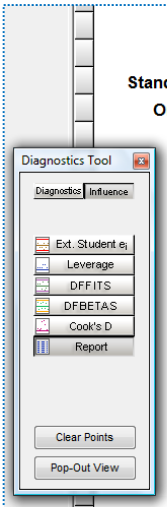
Myers demonstrates mathematically that the DFFITS statistic is really the externally studentized residual multiplied by high leverage points. Click the **Leverage** button and you'll see that all runs exhibit equal leverage here because an equal number of runs were made at each treatment level (all three bowlers rolled six games each).



Leverages

Therefore, this DFFITS exhibits a pattern identical to that shown on the externally studentized residual graph, which you studied in the preceding tutorial. (If you don't recall this, click "Ext. Student e_i" in your Diagnostics Tool.) The reason we're reviewing this is to set the stage for what you'll do later in this tutorial – unbalance the leverages to make this session more significant for diagnostic purposes.

Here's one final Design-Expert software feature for you before we leave the **Diagnostics Tool**: Click the **Report** button to get a table of statistics case-by-case in standard order for the entire experiment. For those of you who prefer numbers over pictures (statisticians for sure!), this should satisfy your appetite. Notice that Mark's high 195 game is highlighted in blue text as shown below.



Diagnostics Case Statistics

Standard Order	Actual Value	Predicted Value	Residual	Leverage	Internally Studentized Residual	Externally Studentized Residual	Influence on Fitted Value DFFITS	Cook's Distance	Run Order
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

Report with case statistics used for preceding diagnostics graphs

Remember, you can right-click any value in reports of this nature within Design-Expert software to view context-sensitive Help with statistical details.

Modifying the Design Layout

Design-Expert offers great flexibility when modifying data in its design layout. We'll see in this bowling scenario how our software allows you to modify an existing design with added blocks and factor levels.

The outcome of the bowling match appears to be definitive, especially from Mark's perspective. But Pat and Shari demand one more chance to prove themselves worthy of the team. They still think Mark's high 195 game was a fluke, even though this isn't supported by the diagnostic analysis. Mark objects and a dispute ensues.

Attempting compromise, the team captain decides to toss out the highest and lowest games for each of the three bowlers and replace them with two new scores each. But Ben, a newly hired programmer and avid bowler, arrives at the alley and is allowed to participate in this second block of runs. (Yes, this makes little sense, but it will add some interest to this tour of Design-Expert's flexibility for design and analysis of experiments – no matter how convoluted they become in actuality.)

It quickly becomes apparent that this new kid does things differently. He's a lefty with a huge hook that's hard to control. To aggravate this variability, Ben does something very different from other bowlers – he does not put his thumb in the ball's hole made for that purpose. When Ben's odd approach works, the pins go flying. But as likely as not, that ball slides off into the left gutter or careens over the edge on the right.

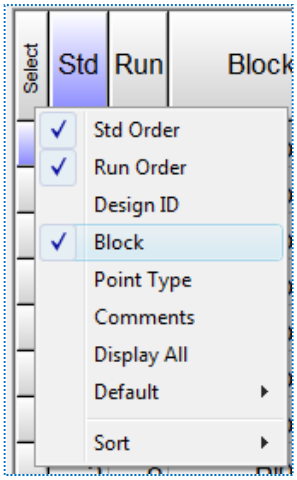
The results for Ben and the three original bowling team candidates are below.

Block	Game	Pat	Mark	Shari	Ben
1	1	160	~~165~~	~~166~~	NA
1	2	150	180	158	NA
1	3	~~140~~	170	~~145~~	NA
1	4	~~167~~	185	161	NA
1	5	157	~~195~~	151	NA
1	6	148	175	156	NA
2	1	162	175	163	200
2	2	153	180	166	130

Bowling scores with high and low games replaced by two new games (plus a new guy)

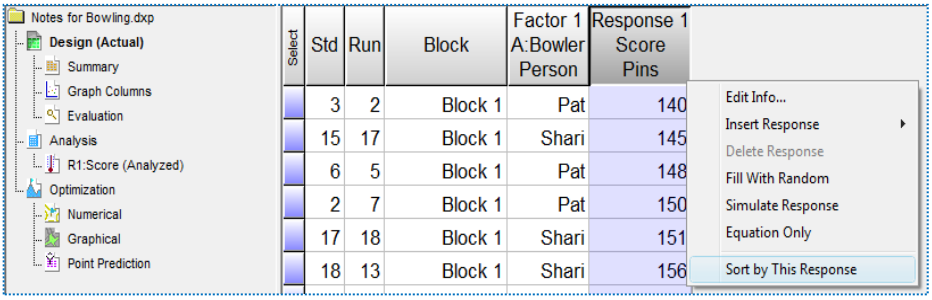
To enter this new data (and ignore some of the old), click the **Design** node near the upper left of your screen. You should now see the bowling data from the first tutorial. Mark's high 195 game remains highlighted in blue text (assuming you clicked on it as instructed on page 2 of this tutorial while performing the diagnostics).

Right click the **Select** column header and click **Block.** This design attribute is now needed to accommodate the new bowler's (Ben's) incoming score data.



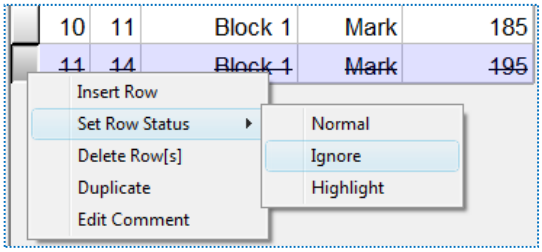
Selecting block to display it as a column in the design layout

Right click the **Response** column header and choose **Sort by This Response**.



Sorting Runs by Response

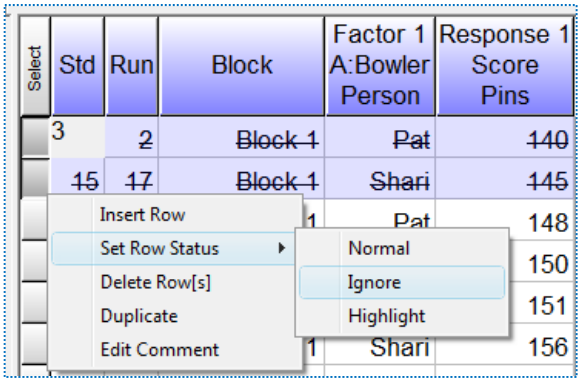
Mark's best game now drops to the very bottom. Let's single him out first to placate Pat and Shari. Right-click the square button in the Select column's last row (Mark's 195 score). Click **Set Row Status**, then **Ignore** as shown below.



Ignoring Mark's high game

By the way, it's OK to change your mind when modifying your design layout: You can 'un-ignore' a row by clicking Set Row Status, Normal.

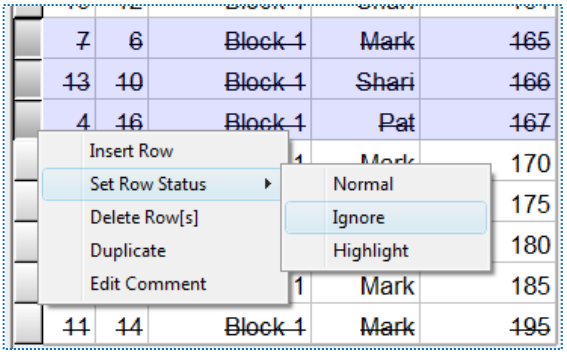
Now let's really get Pat's and Shari's hopes high by excluding their low games from consideration. Click the square button in the Select column's top row (Pat's low 140 game) and, while pressing down the **Shift** key, also click the button in the Select column's second row (Shari's low 145 game). Release the **Shift** key. Keep your mouse within the Select column's first or second row, right-click and choose **Set Row Status**, **Ignore** for these two low games, as shown below.



Ignoring the low games for Pat and Shari

Now move down a few rows and click the square button in the Select column's row showing Mark's low 165 game.

Notice the two rows below Mark's low 165 game – the high games for Shari (166) and Pat (167). It's now time for Shari and Pat to pay the price for complaining. While first pressing and holding down the **Shift** key, click the following two square buttons in the Select column's row: Shari's high 166 game and Pat's high 167 game. Release the **Shift** key. Three rows should now be highlighted in light blue as shown below. Keep your mouse within the Select column's highlighted three rows, right-click and choose **Set Row Status**, **Ignore**.



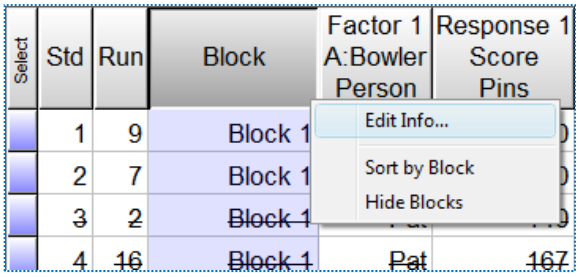
Ignoring Mark's low game and the high games for Shari and Pat

Now let's restore the original layout order. Right-click the **Std** column header, then choose **Sort by Standard Order**. Compare your screen with what we show below. If there are differences, fix them now to match this screenshot. However, remember that the run number is random, so you don't need to fix that.

Select	Std	Run	Block	Factor 1 A:Bowler Person	Response 1 Score Pins
	1	9	Block 1	Pat	160
	2	7	Block 1	Pat	150
	~~3~~	~~2~~	~~Block 1~~	~~Pat~~	~~140~~
	~~4~~	~~16~~	~~Block 1~~	~~Pat~~	~~167~~
	5	8	Block 1	Pat	157
	6	5	Block 1	Pat	148
	~~7~~	~~6~~	~~Block 1~~	~~Mark~~	~~165~~
	8	15	Block 1	Mark	180
	9	4	Block 1	Mark	170
	10	11	Block 1	Mark	185
	~~11~~	~~14~~	~~Block 1~~	~~Mark~~	~~195~~
	12	3	Block 1	Mark	175
	~~13~~	~~10~~	~~Block 1~~	~~Shari~~	~~166~~
	14	1	Block 1	Shari	158
	~~15~~	~~17~~	~~Block 1~~	~~Shari~~	~~145~~
	16	12	Block 1	Shari	161
	17	18	Block 1	Shari	151
	18	13	Block 1	Shari	156

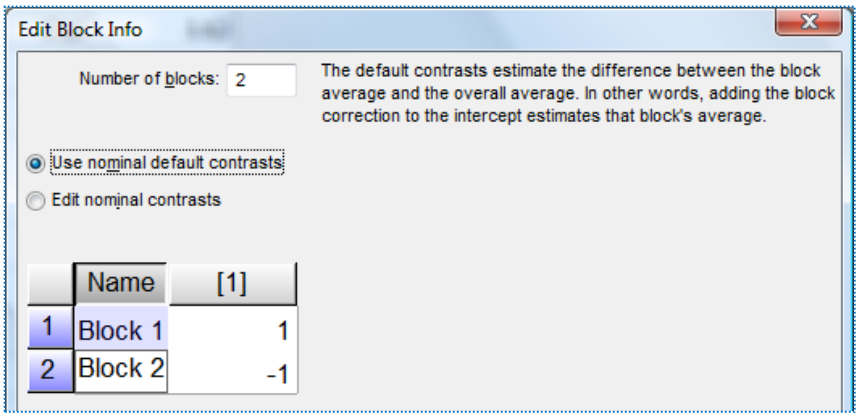
Back to standard order after low and high games ignored for each bowler

Now create a new block (needed for the second round of bowling) by right-clicking the **Block** column header and choosing **Edit Info** as shown below.



Creating a new block

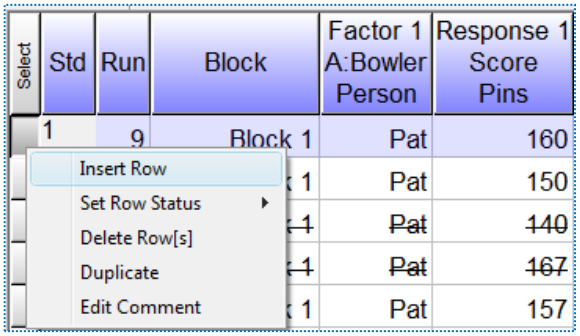
You'll see a form allowing you to assign names to the block(s). Don't bother doing this now. As shown below, change **Number of Blocks** at the top to **2**. Press the **Tab** key to see the change take effect. (If the name field truncates, click and move the right border of the column header to re-size it.)



Adding a second block of runs

Click **OK**. It seems that nothing changed, but actually the program now knows that you will be conducting another block of runs.

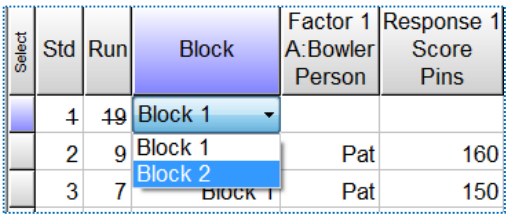
Now you are ready to begin adding and/or duplicating rows. This can be accomplished in different ways, depending on your ingenuity. We'll follow routes revealing as many of the editing features as possible, although they may not demonstrate the most elegant approaches. As shown below, right click the Select column's square button at the left of the first row (Pat's 160 game) to bring up the editing menu. Click the first selection, **Insert Row**.



Inserting a new row

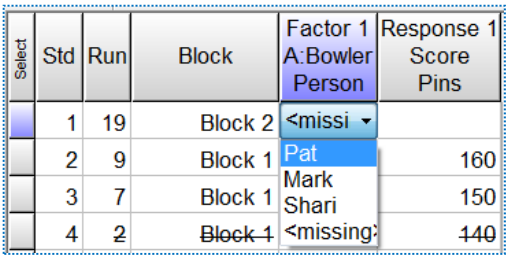
You now see a new row containing blanks for the bowler and the score. (Don't worry if it's being ignored – crossed out, that is – for the moment.) Click the first

row's block cell directly below the block field header, then click the list arrow. Select **Block 2** as shown below.



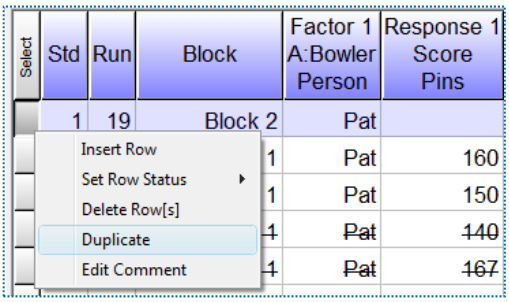
Changing the block number

Click the blank field for bowler and press the list arrow (▼). Select **Pat**. (We're using categorical factors here, but if this were a numerical field, you'd enter a value.)



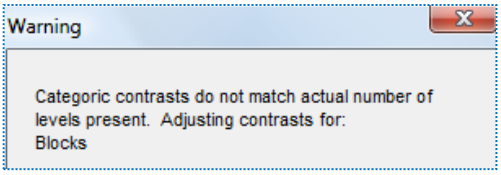
Entering a categorical value for factor

Again, right-click the Select column's square button at the left of the first row to bring up the editing menu as shown below. Click **Duplicate**.



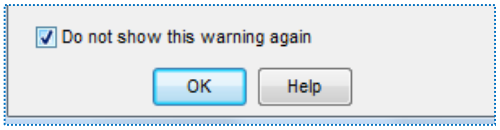
Duplicating a row

Design-Expert may pop up a warning like the one shown below.



Warning about categoric contrasts

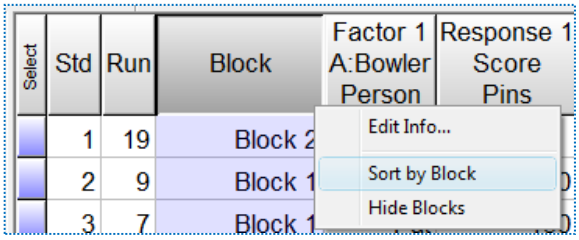
The program is recognizing a potential problem here and is alerting you that only one bowler is in the second block. You need not worry at this stage because you will be adding others. Click the check option **Do not show this warning again.** This will save you aggravation later. Don't worry – you will not be unprotected indefinitely. This warning will be re-enabled the next time you start the program.



Turning off a warning (it will come back the next time you run the program)

Press **OK** to proceed.

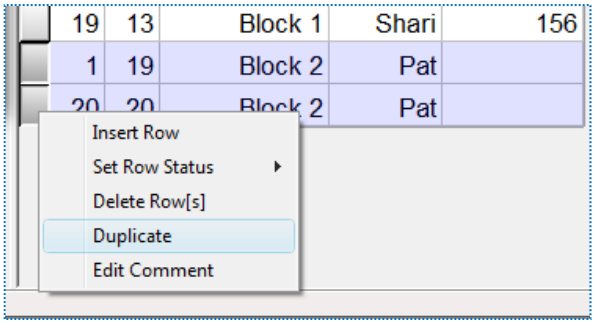
Right-click the **Block** column header and choose **Sort by Block**.



Sorting by block

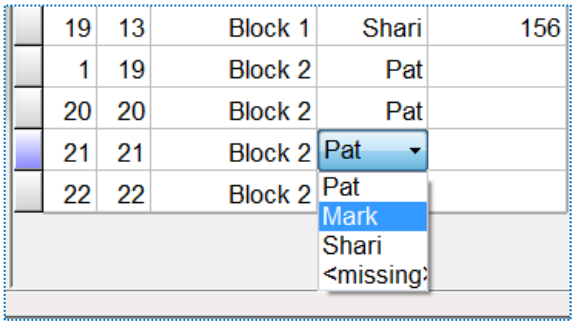
Two new rows are now seen at the bottom of your design layout. We need two new rows apiece for Shari and Mark. Let's simply duplicate Pat's two new rows and update the names. Do this by first clicking the Select column's square button at the left of Pat's first new row, so it is highlighted. Then while holding down the Shift key, click the Select column's square button at the left of Pat's second new row. Both rows should now be highlighted. (This is a bit tricky, but it saves time.)

Now right-click any Select column's square button at the left of the highlighted block and select **Duplicate**. (If the warning screen pops up again, click OK.)



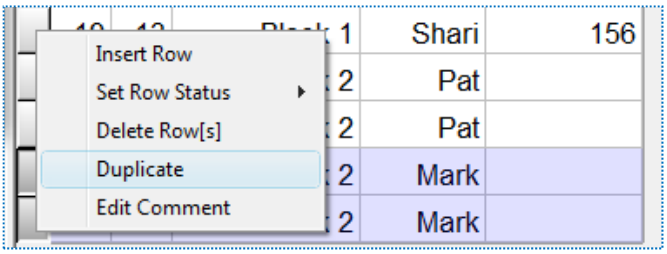
Duplicating a block of rows

In the first duplicated row, click the field for **Bowler** and select **Mark**.



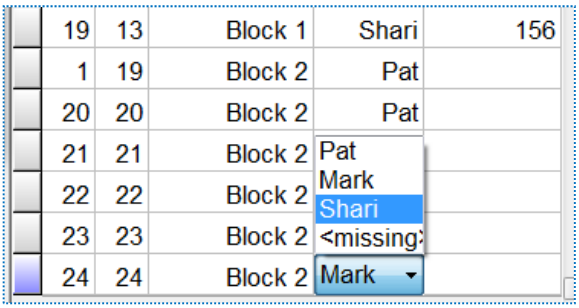
Changing name of bowler

Do the same for the last row. You now should have two new rows for both Pat and Mark. Click the Select column's square button at the left of Mark's first new row, so it is highlighted. Then while holding down the **Shift** key, click the Select column's square button at the left of Mark's second new row. Both rows should now be highlighted. As before, right-click any Select column's square button at the left of the highlighted block and select **Duplicate**.



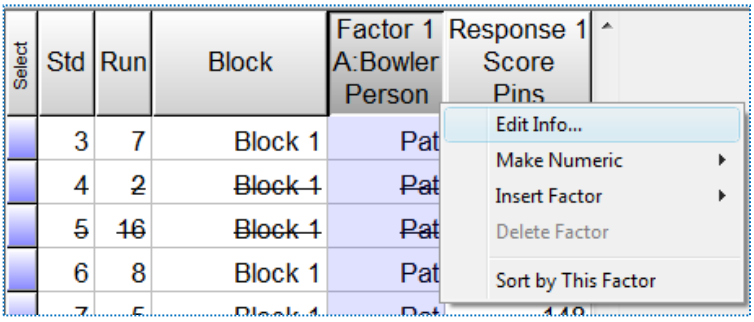
Duplicating two more rows

In the first duplicated row, click the field for **Bowler** and select **Shari**. Do the same for the last row.



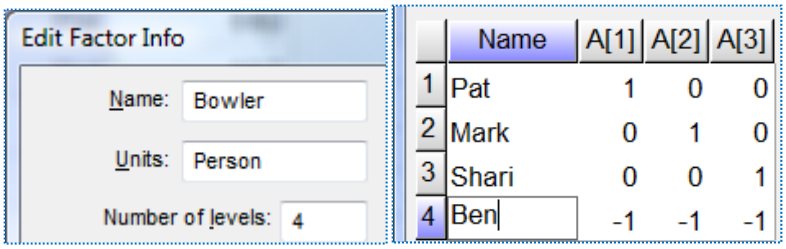
Completing lineup for block 2 – the second round of bowling

But what about the new kid – Ben? We need to identify him as a new competitor in this bowling contest. Do this by right-clicking the header for **Bowler** and selecting **Edit Info**.



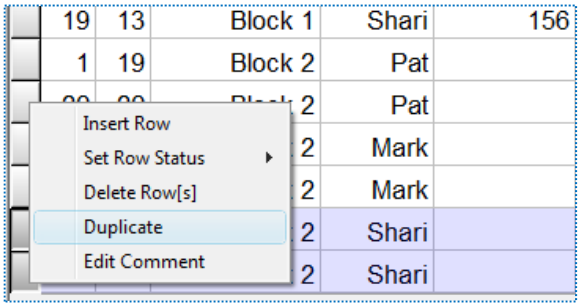
Getting ready to add a new level for the factor

Change **Number of Levels** to **4** (see below left). Press **Tab** once. Click the field intersecting at **Name** column and row **4** (below right). Type the name **Ben**.



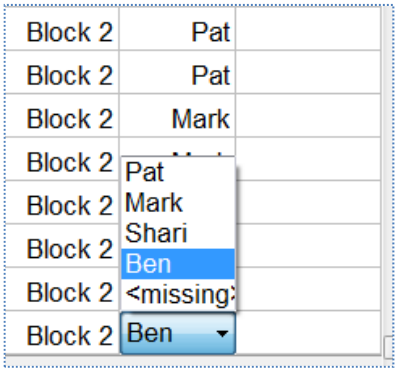
Entering the new bowler

Press **OK**. Now duplicate two more rows by clicking the Select column's square button at the left of the first of Shari's two new games. While holding down the **Shift** key, click the Select column's square button at the left of the last run. Finally, right-click any Select column's square button at the left of the highlighted block and select **Duplicate.**



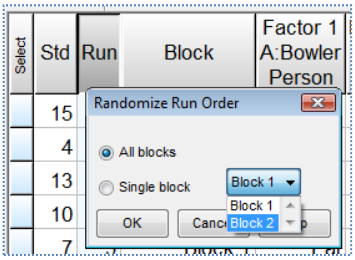
Duplicating two more rows so the new bowler can be included

In both of these new duplicated rows, click the fields for **Bowler** and select **Ben**.



Ben now on the list as a bowler

An important aside: Always randomize your run orders for actual experiments. For purposes of this tutorial, this will just be a bother, so do this only if you wish to try it out, but it's very easy to do – simply right-click the Run column-header and do this for Block 2 as shown.



How to randomize the run order in the second block

Now enter the new data as shown below.

	1	19	Block 2	Pat	162
	20	20	Block 2	Pat	153
	21	21	Block 2	Mark	175
	22	22	Block 2	Mark	180
	23	23	Block 2	Shari	163
	24	24	Block 2	Shari	166
	25	25	Block 2	Ben	200
	26	26	Block 2	Ben	130

Data entered for second block of games

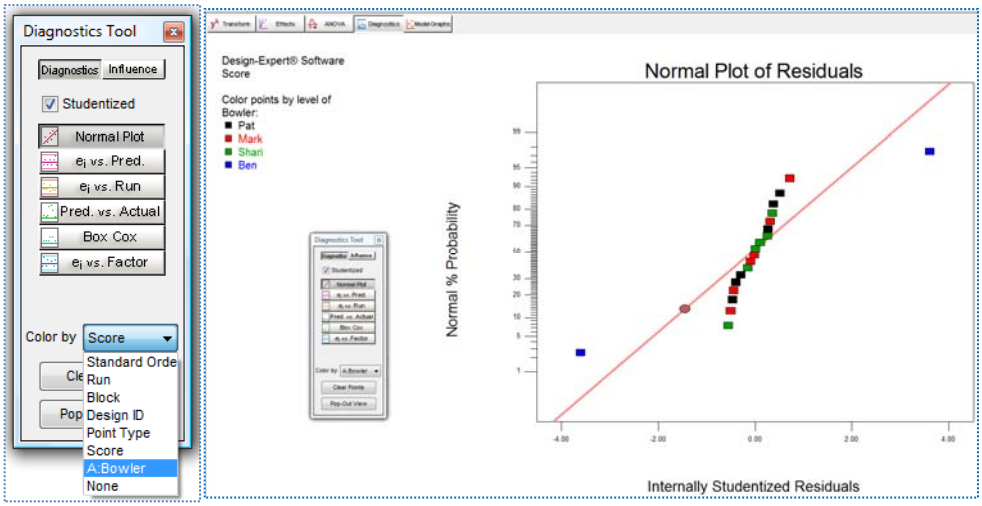
Go ahead now and re-analyze your data by clicking the **Score** node under **Analysis**. Move through ***Transform*** and click on the ***Effects*** button. Here, you need to make sure "A-Bowler" is included in the model. To include this term, make sure it is identified by an "M". If not, **double click on "A-Bowler"** in the table to toggle it from "e" to **"M"**.

Selection: Manual Order: Main effect

Term	df	Sum of Squares	Mean Square	F Value	Prob > F
Intercept					
M A-Bowler	3	1722.82	574.27	3.05	0.0611
e Lack Of Fit	2	42.72	21.36		
e Pure Error	13	2782.25	214.02		
Residuals	15	2824.97	188.33		

Effects button results

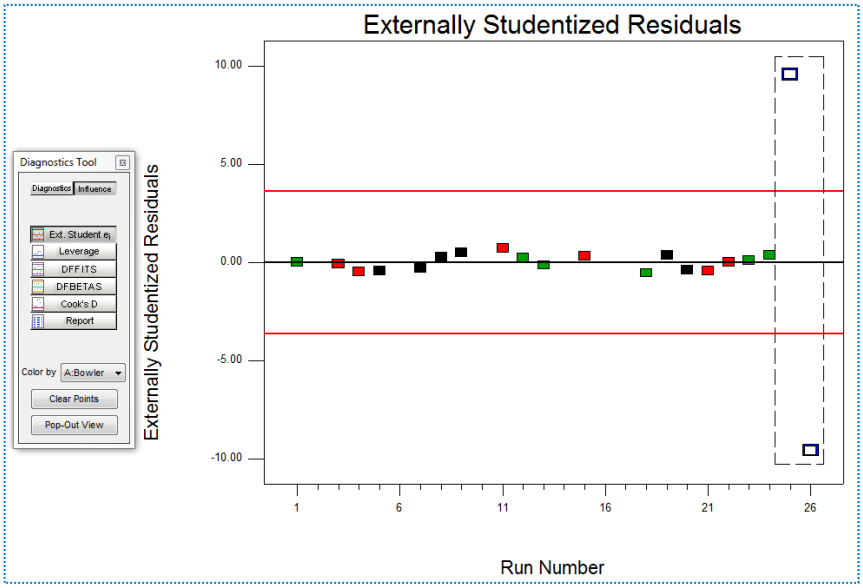
Proceed to ***ANOVA*** and then to ***Diagnostics***. On the floating **Diagnostics Tool,** change (click) from **Influence** to **Diagnostics**. Do not worry if this model gets labeled "insignificant." As you will see, something is abnormal about this data. Do you notice that the residuals now line up very poorly, especially at the extreme points as shown below? On the floating **Diagnostics Tool** change **Color by** to **A: Bowler**.



Diagnostics for bowling results – part two: Normal plot with poorly aligned residuals

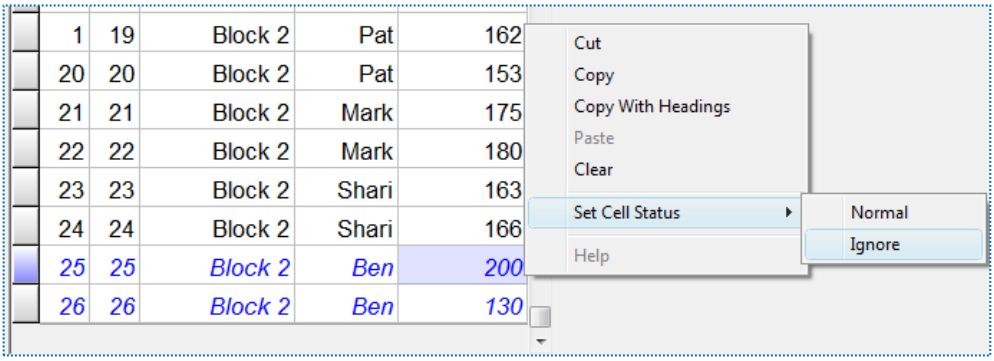
Now you see that the results from Ben do not fit with the others (his games are the two outliers – low and high), which should be no surprise considering his odd style of bowling. Click the **Influence** button to bring up the externally studentized residuals – a good tool for detecting outliers.

Drag your mouse over Ben's residuals at the far right. Both points should now be highlighted. We must ignore or delete them. (Sorry Ben, odd behavior by programmers is considered normal at Stat-Ease, but not when it comes to bowling!)



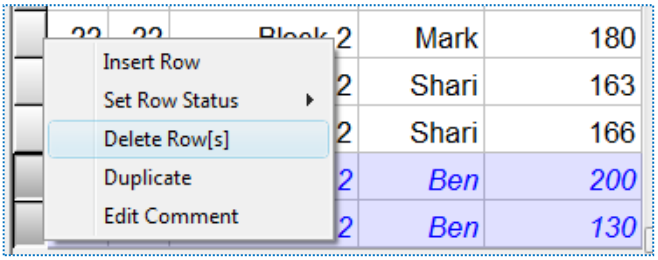
Ben's games selected on 'outlier' plot (externally studentized residuals)

Click the **Design** node (upper left) to get back to the home base of the design layout. Notice that Ben's games are conveniently highlighted in blue text so they can easily be deleted. There are many ways to do this in Design-Expert software. It provides no advantage in this case, which features only one response measure, but you can ignore a specific result by right-clicking that cell and setting Set Cell Status to Ignore as shown below.



Ignoring a single cell – an option that's not recommended for this case

In this case it is preferable to ignore his entire runs (not just individual cells) as described earlier in this tutorial. Better yet, simply delete them altogether. No offense to Ben, but given that he only bowled two games and his unorthodox style creates such abnormal variability, it is best now to click the Select column's square button at the left of his first score of 200 (making him feel really bad ☹), shift-click the button below it for the second game of 130 (not so sorry to see this gone!), then without moving your mouse, right-click and select **Delete Row(s)**.



Deleting Ben's games

Click **Yes** on the warning that pops up about deleting rows (a safety precaution) and **OK** for the heads-up the program gives you about eliminating an entire categoric contrast (no more Ben). Then go ahead and re-analyze the results.

It turns out that the added games cause no change in the overall conclusions as to who's the better bowler. Mark remains on top. It would now be appropriate to recover the low and high games for each bowler from block 1. Because this data was not deleted, only ignored, getting it back is simply a matter of right-clicking to the left of each of the six suspect rows and changing Set Row Status to Normal. (Or, if you're adept at manipulating lines of text or data with your mouse, do all rows at once using a click and shift-click.) Give this a try! Then re-analyze one last time.

By working through this exercise, you now see how easy it is to manipulate Design-Expert's design layout.

PS. Still feeling bad about deleting Ben's scores? Don't worry – he gets to bowl with Pat and Shari in a lesser league. After bowling for an entire year (roughly 100 games), it will become clear whether Ben's crazy way of bowling will pay off by achieving a good average overall. After all, his 2 game average of 165 wasn't so bad, just inconsistent (high variability). With more data, his true ability will become more apparent.