Data analysis: a very powerful tool for monitoring performance

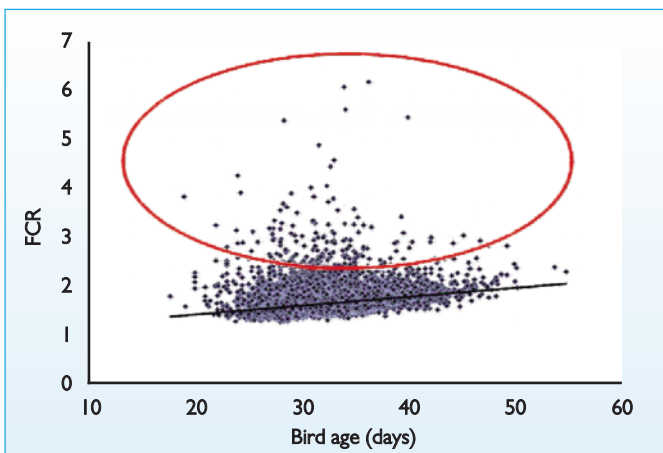*by Dr Nick French, global head of technical transfer, Aviagen Ltd.*

Today's poultry industry collects vast quantities of data covering all aspects of the production process: from breeder flock records to meat yield in the processing plant; from feed formulations in the feed mill to hatchery performance data.

We use the data to monitor progress and make sure we meet our business targets, but data can be used to do more. Data can be used to help investigate problems and to look for areas to improve performance.

Modern computing power and the development of computer programs for working with data, such as Excel and Access, has made analysis much easier than in the past. Most users of computers will have the tools to do quite complicated data analysis installed on their machines, but often have not been trained how to do more than simple data spread sheets.

This is one of the reasons Aviagen have made data handling a focus of their Europe, Middle East and Africa Production Management Schools. Whilst it is not hands on with the birds, good data handling can have a real impact on the bottom line of the business.

With data being such an important aspect of all parts of the business today it is more important than ever for poultry producers to have an understanding of the tools available to them and what can be done.

The purpose of this article is to show how more complicated data analysis can be performed on any computer using Excel as an example. Excel has been chosen here as the most widely used system but other data handling packages can also be used equally well.

Good analysis requires:
● Good data: we must be able to trust that the numbers we are analysing are correct.
● That the data is stored in a way that it can be analysed, in a database format.
● The analysis is carried out by someone who understands what the data means and looks for relationships that are meaningful.

Trusting the numbers

Clearly we can only do sensible data analysis if the data we are analysing is correct. If the numbers are not entered correctly then we are less likely to come to the correct conclusion from the data analysis. In computing terminology, this is known as 'garbage in – garbage out'.

Fig. 1. Plotting results can quickly highlight potential data errors – the very high FCR numbers circled should be checked to make sure they are correct.

Fig. 2. An example of data stored in an Excel sheet in a report format. Each tab in the workbook holds the results for an individual hatch.

Fig. 3. Example of the data from Fig. 2 converted into database format. All the results for all the hatches are now on one sheet.

When analysing data we should not assume that the numbers are correct, we need to check.

Even the best systems have data errors.
● Check that your data makes sense, and that includes chicken sense. The person entering or checking the data should understand what the numbers mean and what the numbers look like.
● Beware of averages as they can hide poor data – look at the range around the average to make sure they are also sensible.
● Graph your data to look for data points that stand out from the trend (Fig. 1) – are these points correct or data errors?
● Beware of numbers that are too good to be true and show no variation – biological data nearly always shows some variation.

Correcting errors in data should be done with care. If it is clear what the correct number should be then the data can be corrected. If it is not clear, it is normally best to leave the value as blank rather than enter a value that we think is correct.

Storing data for analysis

A common problem when asked to carry out a data analysis is to find that all the data are stored on Excel sheets in a format designed to report results (Fig. 2). Each sheet contains the results from one flock, one week, one day etc.

While this is useful for inspecting data it is very difficult to do an analysis of all the data across all the sheets to look for trends and relationships. What is required for data analysis is for the data to be stored in a database.

Databases are used for storing data (in some cases very large quantities of data) in a structured format. For data analysis, ideally we want all the data we wish to analyse on one sheet with all the information for each data record in a single row on the sheet.

As an example, the report format data shown in Fig. 2 has been converted into database format in Fig. 3.
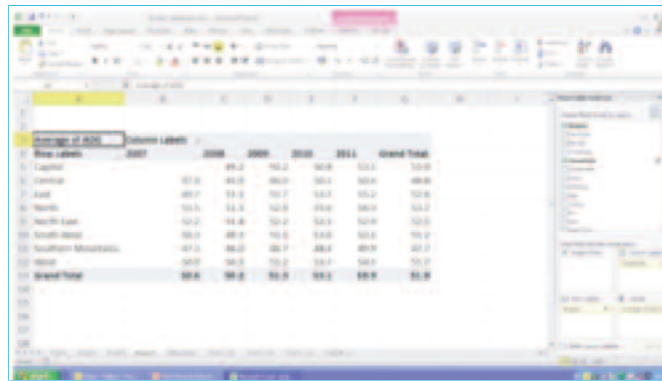
Note that now all the data for each individual sheet for each hatch has been transferred into a single sheet. The test date shown at the top of the report in Fig. 2 has been moved into a column in the database and now listed for each record for that hatch, changing when another record sheet is copied into the database sheet. Other information is stored in the database such as the incubator number.

Specialised programs such as Access are available for large databases. These systems are particularly useful when several databases need to be connected together, such as a flock record database with a hatchery database. However, a database can be easily set up in Excel which can still hold a very large number of records.

Working with a database
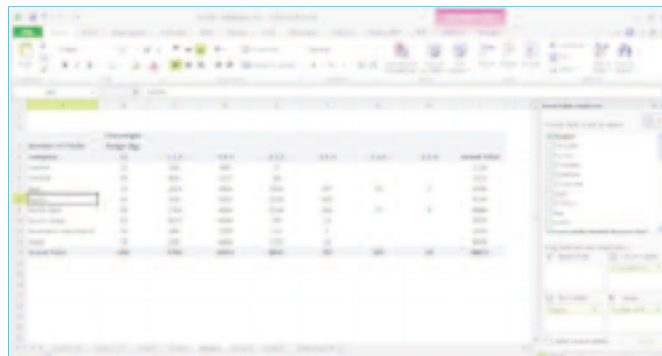
Once we have our data in a database what can we do with it? One of the most useful tools in Excel for analysing data is Pivot Tables. Pivot Tables allow you to organise, report, analyse and graph your data quickly and the time taken to learn how to use Pivot Tables is quickly repaid in time saved later.

Different versions of Excel will have different forms of Pivot Tables: in Excel 2003 it can be found in the Data Menu, in Excel 2010 on the Insert Ribbon. However, the different versions of Pivot Table all work by summarising all the data in a large

Fig. 4. Example of a Pivot Table used to summarise average broiler liveweight from more than 48,000 flocks in eight different company complexes.

Fig. 5. Using a Pivot Table to count the number of flocks within liveweight ranges in a broiler operation dataset.

database into a table. The table can summarise the raw data to show totals, averages, minimums, maximums, number of records or standard deviations depending on what is selected, in a few mouse clicks.

The table can be set up to show the results of the data summaries by any categories (such as flock name, hatchery, feed mill etc.) in the database to produce a report in the format required.

An example of a Pivot Table is shown in Fig. 4, showing a summary of broiler performance data for different company regions.

A very useful tool within Pivot Tables is the ability to take a column of numerical data and use the values as either a column or row heading which is then grouped into ranges. For example it is possible to report body weight as ranges and then show a count of the number of flocks in each weight range (Fig. 5).

A particularly useful feature of this tool is the ability to work with dates so they can be grouped into months, quarters and years (Fig. 6).

There are many features to Pivot Tables which can be used to make working with data much easier for the user.

Producing standard reports from a database is made easy as once the table has been created it can be automatically updated when new data are added into the database. Graphs can be attached to the Pivot Table which would also be automatically updated when new data are

added. The purpose of this article is not to provide a guide to using Pivot Tables in Excel but to give the reader some idea of the possibilities of data analysis using this powerful tool.

Going further

For most users, applying Pivot Tables and graphs will allow them to do the type of data analysis that they require for their business. It will allow for simple comparisons, look at changes over time and investigate relationships between different factors.
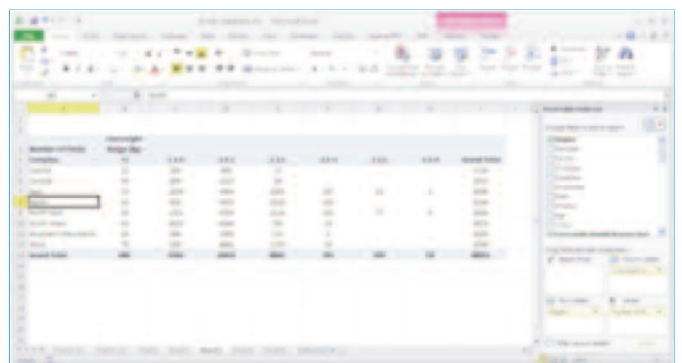
However, it is possible to go further and to look at the data using statistical techniques that will allow an analysis of multiple factors at the same time, to understand which fac-

tors have the biggest effect on performance.

The statistical approach has several advantages:
● Firstly, it shows the factors that have a significant impact on performance; by significant we mean that the differences found are unlikely to be caused by random variation in performance.
● Secondly, a statistical approach allows any one factor that can affect performance to be analysed once all the other factors in the analysis have been taken into account. For example we know that many factors can affect body weight: season, nutrition, feeder space, stocking density, farm management, ventilation etc. and we can measure the effect of each of these factors once the variation in performance caused by the other factors have been accounted for. It is called measuring the pure effect of each factor and it can give a different result to just measuring the average body weight for each factor.
● The statistical analysis can give a measure of the relative importance of each factor on performance. In the body weight example, it could show that nutrition has a larger effect than stocking density, which in turn has a larger effect than feeder space. The analysis highlights where improvements can be made to have the largest effect on performance.

The main difficulty with the statistical approach is that it needs to be carried out with a proper understanding of statistical methodology. There is a wide range of statistical techniques available and the correct selection of the right method is essential to getting meaningful results. Some training in statistics is needed to do a proper analysis.

Final thoughts

Data analysis can be a very powerful tool for monitoring performance and looking for where improvements can be made.

It must be done in conjunction with a good understanding of both the chicken and the economic factors affecting the business: the conclusions drawn must make both chicken and economic sense. ■

Fig. 6. Using the grouping tool to summarise date data into periods.