Dropping The Ball

August 01, 2005

"If you don't want to discover true associations, ignore what is going on among modern [statistical] techniques," said Rand Wilcox, University of Southern California, playfully addressing the overflow audience.

Wilcox's invited address entitled, "More Reasons Why Discoveries Are Lost by Ignoring Modern Statistical Methods: Some Recent Advances When Comparing Groups and Studying Associations" was part of the popular, new methodology programming at the APS 17th Annual Convention in Los Angeles.

Wilcox expressed deep dissatisfaction with standard statistical techniques, and offered some modern alternatives to routinely used ANOVA, or analysis of variance, and regression, where the results are fitted to an equation for analysis. According to Wilcox, before you decide not to reject the null hypothesis, you better wait. "The possibility is that there is something going on [in your data], but you are just not using the right tool to detect it," he said.

"ANOVA performs well when distributions do not differ. Least squares regression and Pearson's correlation work well when variables are independent," Wilcox said. In fact, "if the underlying assumptions are true, these methods have considerable power." But, "suppose the assumptions are not true."

According to Wilcox, there was already enough evidence 40 years ago indicating that, in the case of ANOVA, even a very small departure from normality could destroy power. When distributions differ "ANOVA falls apart," Wilcox argued. "When we look at regression, things get even worse. Regression basically inherits all the problems [of] ANOVA, plus new problems are introduced."

Testing the Assumptions

Researchers often find themselves analyzing data without knowing that the findings have been controlled and handled the correct way. But which process is most effective in confirming the results and checking for errors?

"Diagnostics is a very reasonable suggestion," Wilcox said, however, "consistent throughout all papers, this strategy is not working well." The standard methods "don't have enough power to detect problems."

"Now, assume an outlier," Wilcox said, referring to a rogue subject that bears results inconsistent with the majority. "Outliers can be disastrous," he added. "There are about half a dozen ways of detecting outliers.

Undetected outliers can considerably distort correlations. "It actually happened to somebody in our department doing clinical study," Wilcox said. A student found a high correlation in his data on

schizophrenia. "He was very excited. He solved schizophrenia! Then he found an outlier. When he removed it, the correlation disappeared," Wilcox said. "The good news is that we have the technology to address these problems."

Statistical Tools

Most important is a more robust theory that enables us to deal with non-normality. Wilcox pointed out that problems associated with the Central Limit Theorem, a theorem used to examine converging results, using either percentile or bootstrap-T, a method used for evaluating generalities of the data, can be overcome with this more robust theory. "As the sample size gets larger and larger, they converge in the correct answer more quickly," Wilcox explained.

Wilcox suggested that familiarity with the properties of modern methods increases our chances of getting the most from data. For example, there are conditions where a percentile bootstrap method beats the bootstrap-T, and there are conditions where the reverse is true.

"We [also] really need some better ways of doing an effect size," Wilcox argued. He suggested using Kernel Density Estimator, a nonparametric technique in which data are averaged to create a smooth approximation. This method "has been around for a long time, but within the last year major advances have taken place," Wilcox said.

"Modern technology is wonderful. It is just sitting there waiting for people to take advantage of it," Wilcox said. He stressed the importance of keeping up with the rapid progression of statistical methods. "Advances continue to appear in the statistics literature, but if we choose to ignore these techniques, discoveries are going to be lost," he said.