

Re: Need help understanding Homogeneity of Variance please

Re: Need help understanding Homogeneity of Variance please

Source: <http://sci.tech-archive.net/Archive/sci.stat.edu/2006-11/msg00001.html>

- *From:* Bruce Weaver <bweaver@xxxxxxxxxxxxx>
 - *Date:* Tue, 31 Oct 2006 09:38:51 -0500
-

Reef Fish wrote:

Richard Ulrich wrote:

On 30 Oct 2006 06:50:28 -0800, "Reef Fish"

RU:

I would prefer to say, "the method *may* be all wrong," and I think that RF expresses that more relaxed idea in his closing comments, where some violations are more serious than others
[snip, some detail]

RF:

BUt those are TWO DIFFERENT sets of statements.

In the above, it means If the ASSUMPTION(s) are NOT valid, then the statistical results based on the method WILL be all wrong. There is no "may be" about it. If you have two binary variables X and Y and you test its correlation with the test statistic T for the Pearson correlation coefficient (which would be phi for the two binary variables), the result WILL be wrong because the assumption is violated 100%, without question.

In the situation below, it's about the VALIDATION of the assumption. If Normality is required of a variable, and it is not known 100% to be nonnormal, then there is leeway in deciding what is a serious violation and what is not, because in that case (unlike the case it does not require any thinking to know that the (0,1) variable is NOT normal) the DATA can never prove with 100% certainty whether it came from a Normal population or not.

There is a BIG difference in the above two situations.

Re: Need help understanding Homogeneity of Variance please

But as Box put it, "in nature there never was a normal distribution". So if we're talking about real data, we *know* with 100% certainty that it's not normal, and the real question is whether it is *useful* to assume that it is. Here's the Box quote in some context.

"In applying mathematics to subjects such as physics or statistics we make tentative assumptions about the real world which we know are false but which we believe may be useful nonetheless. The physicist knows that particles have mass and yet certain results, approximating what really happens, may be derived from the assumption that they do not. Equally, the statistician knows, for example, that in nature there never was a normal distribution, there never was a straight line, yet with normal and linear assumptions, known to be false, he can often derive results which match, to a useful approximation, those found in the real world."

Box GEP. Science and Statistics. JASA, Vol. 71, No. 356 (Dec., 1976), 791–799.

—

Bruce Weaver

bweaver@xxxxxxxxxxxx

www.angelfire.com/wv/bwhomedir

.