

Re: Goodness of fit measures for a distribution

Source: <http://sci.tech-archive.net/Archive/sci.stat.math/2005-04/msg00222.html>

- *From:* "Reef Fish" <Large_Nassau_Grouper@xxxxxxxxx>
 - *Date:* 12 Apr 2005 16:23:23 -0700
-

Aleks Jakulin wrote:

> Reef Fish wrote:

>> Chi-square is based on histograms -- it's worthless.

>>

>> Kolmogorov uses on ONE POINT in the difference between the empirical

>> and theoretical cdfs, the point of maximum departure.

>>

>> Your EYEBALLS can do an infinitely better job than that, looking at
>> the plot of the entire cdfs.

>

> In principle yes.

In principle AND in practice.

Here is a very simple PRINCIPLE. If something is INAPPROPRIATE for 1 (ONE) <such as a histogram>, it's inappropriate for the other 999 of a thousand also.

> But what do you do when you have a few dozen or
> hundred variables in a complex data mining task?

You should mine more carefully and delicately than running a steam-roller over all of them when more delicate tools are required.

When you have a few thousand variables, the first task is to selectively consider ONLY a few dozen, if that many, that seem most appropriate, for substantive reasons.

> Then you do want a one-number summary: you have no time to manually
> inspect a few hundred thousand QQ plots.

It's easy to do a one number summary. Just generate one RANDOM NUMBER, and say "that's what the 'puter gave me!" And that number is probably as meaningful <or meaningless> as your single number from GIGO (Garbage In, Garbage Out).

Re: Goodness of fit measures for a distribution

What you have argued is in fact the WORST that has happened to the application of statistics -- when computer programs and packages are readily available for any Tom, Dick, and Harry to throw data into the bin to get some meaningless and useless number(s) out.

Progress takes 10 steps backwards.

-- Bob.

> mag. Aleks Jakulin
> <http://kt.ijs.si/aleks/>
> Department of Knowledge Technologies,
> Jozef Stefan Institute, Ljubljana, Slovenia.

- *Follow-Ups:*

- ◆ **Re: Goodness of fit measures for a distribution**
◇ *From:* Aleks Jakulin

- *References:*

- ◆ **Goodness of fit measures for a distribution**
◇ *From:* Unknown
- ◆ **Re: Goodness of fit measures for a distribution**
◇ *From:* Reef Fish
- ◆ **Re: Goodness of fit measures for a distribution**
◇ *From:* Aleks Jakulin

- Prev by Date: **Re: P-value from chi-square value: source code**
- Next by Date: **Re: Reduction of variance due to grouping**
- Previous by thread: **Re: Goodness of fit measures for a distribution**
- Next by thread: **Re: Goodness of fit measures for a distribution**
- Index(es):
 - ◆ **Date**
 - ◆ **Thread**