Re: assumption of Classification

Source: http://sci.tech−archive.net/Archive/sci.stat.edu/2005−04/msg00119.html

- *From*: Richard Ulrich <Rich.Ulrich@xxxxxxxxxxx>
- *Date*: Wed, 27 Apr 2005 12:17:06 −0400

On 26 Apr 2005 22:21:23 −0700, "Data Matter" <fungile@xxxxxxxxx>
wrote:

> He's asking if these procedures make distributional assumptions.
> Classification trees do not. Most clustering algorithms (k−means,
> single link, average link, etc.) do not. However, there is a class of
> clustering algorithms which assumes that each cluster is multivariate
> normal and then proceed to find the means and covariances of these
> clusters.

A classification tree that tries to break at every value
will not care whether the distance between 1 and 10 is
the same as the distance between 10 and 100 (or not).
(It is going to have a lot of opportunity to over−capitalize
on chance, so the N needs to be large.)

A classification tree that uses the mean will have some of
the same difficulty that "link" clustering does, if it wrongly
assumes that equal measures of intervals are equivalent.

>
> Nonetheless, normality is not the only assumption to be checked. Every
> method has its own list of assumptions and you should make sure that
> your data agree with the method you choose.
>

It's always good to check.

For methods of ordinary least squares, normality is not
as important as having decently behaved residuals − mainly,
absence of outliers, absence of pattern. And that behavior
matters for the *tests*, not for carrying out the fit.
[...]

−−
Rich Ulrich, wpilib@xxxxxxxx
http://www.pitt.edu/~wpilib/index.html
.

- ***Follow−Ups***:
  - ♦ ***Re: assumption of Classification***
    - ◊ *From:* Reef Fish

- ***References***:
  - ♦ ***assumption of Classification***
    - ◊ *From:* apple0811
  - ♦ ***Re: assumption of Classification***
    - ◊ *From:* Richard Ulrich
  - ♦ ***Re: assumption of Classification***
    - ◊ *From:* Data Matter

- Prev by Date: ***Re: assumption of Classification***
- Next by Date: ***Re: MatLab randn and Simulation Step Numbers***
- Previous by thread: ***Re: assumption of Classification***
- Next by thread: ***Re: assumption of Classification***
- Index(es):
  - ♦ ***Date***
  - ♦ ***Thread***