データの対数変換後の分散

岩瀬晃盛:NPO 環境統計統合機構

2016 年 1 月 4 日

メートル法（MKS 単位系）で計測した身長の確率変量を $X_{メートル}$ とし、尺貫法で計測した同じ身長の確率変量を $X_{尺}$ とする（確率変量と確率変数とを使い分けている）。

$X_{メートル}$ の単位はメートルであり、$X_{尺}$ の単位は尺である。従って、

$$0.30303[メートル/尺]X_{尺} = X_{メートル}$$

なる関係がある。例えば

$$X_{尺} = 5.0[尺] \qquad X_{MKS} = 1.51515[メートル]$$

など。

ここで両辺を数にするために

$$0.30303[メートル/尺]\mu_{尺}\frac{X_{尺}}{\mu_{尺}} = \mu_{メートル}\frac{X_{メートル}}{\mu_{メートル}}$$

とした後に

$$0.30303[メートル/尺]\frac{\mu_{尺}}{\mu_{メートル}}\frac{X_{尺}}{\mu_{尺}} = \frac{X_{メートル}}{\mu_{メートル}}$$

と変形すれば、上式は数に関する等式となっていることを確認されたい。

ここで上式右辺の対数、$\log\left(\frac{X_{メートル}}{\mu_{メートル}}\right)$ の分散を計算すると

$$Var\left[\log\left(\frac{X_{メートル}}{\mu_{メートル}}\right)\right] = Var\left[\log\left(0.30303[メートル/尺]\frac{\mu_{尺}}{\mu_{メートル}}\frac{X_{尺}}{\mu_{尺}}\right)\right]$$

$$= Var\left[\log\left(0.30303[メートル/尺]\frac{\mu_{尺}}{\mu_{メートル}}\right) + \log\left(\frac{X_{尺}}{\mu_{尺}}\right)\right]$$

ここで

$$0.30303[\text{メートル}/\text{尺}]\frac{\mu_{\text{尺}}}{\mu_{\text{メートル}}}$$

は定数であるから

$$= Var\left[\log\left(\frac{X_{\text{尺}}}{\mu_{\text{尺}}}\right)\right]$$

となる。つまり

$$Var\left[\log\left(\frac{X_{\text{メートル}}}{\mu_{\text{メートル}}}\right)\right] = Var\left[\log\left(\frac{X_{\text{尺}}}{\mu_{\text{尺}}}\right)\right]$$

メートル法で表現したデータを同じ単位を持つ任意の量で割り算をして数にした対数の分散は、尺貫法で表現したデータを同じ単位を持つ任意の量で割り算をして数にした対数の分散に等しいことが示されたことになる。

上記の議論は母集団での議論であるけれども、標本に関する議論に於いても同様であり、ある集団の構成員の身長をメートル単位で記述したデータと、同じ集団の構成員の身長を尺貫法で記述したデータがあったときに、どちらのデータも数字だけにして対数変換した後の標本分散を計算すれば結果として一致するということを意味している。（この具体的な展開が必要であると判断されれば後に提示する）

ざっくり言えば、分散の計算に限っては気兼ねなく量の数値だけを取り出して対数変換後にその計算をしても支障は生じないし、その（数である）値には意味がある。ここで言う意味とは、対数正規分布 $LN(\mu, c^2)$ での母パラメータ（数であるから母数と表現できる）c^2 と関連付けられることである。

このデータの単位を無視して数だけに注目して対数変換をした後に、標本分散を計算することの ”結果として ”の正当性が示された。

本来は量の対数変換は許されないのであるが、形式的に機械的にその量の数の部分だけを取り出して議論されることが見受けられる。このような議論は誤ったものなのであるけれども、これらの議論の中で結果的には正当な操作と同じ結果をもたらす場合として、データの対数変化後の分散があることを示したものである。

一般論としては、確率変量 X について

$$Var\left[\log\frac{X}{\mu_0}\right]=Var\left[\log\left(\frac{\mu_1}{\mu_0}\frac{\dot{X}}{\mu_1}\right)\right]=Var\left[\log\frac{\mu_1}{\mu_0}+\log\frac{X}{\mu_1}\right]=Var\left[\log\frac{X}{\mu_1}\right]$$

つまり正値確率変量 X をこれと同じ単位を持っているどんな量で割り算することによって数へ変換しても、数へ変換後の分散は一定で変化しない。つまり数へ変換する量に分散は依存しないことを意味している。尚、数へ変換後の分散は当然のことながら数である。

平均に関しては勿論このようなことはいえない。また、くどいけれども、量を対数変換することは定義され得ないことに注意のこと。