

サンプルサイズの決め方は、分析の目的や方法によって異なりますが、今回の講座で見た母平均の区間推定の場合であれば、信頼区間 $100(1-\alpha)\%$ で、誤差を E 以内で押さえたい場合、サンプルサイズ N は、

$$N = \left(\frac{z\left(\frac{\alpha}{2}\right)}{E} \sigma \right)^2$$

で決めます。 $z(\alpha/2)$ は標準正規分布の累積確率が $\alpha/2$ となる点の値を表し、 σ は母集団の標準偏差(またはその近似値)です。例えば、ある製品のサイズの標準偏差が 1mm であることが分かっている、 95% 信頼区間($\alpha=0.05$)で、誤差を 0.1mm 以内に収めたいときには、

$$N = \left(\frac{z\left(\frac{0.05}{2}\right)}{0.1} 1 \right)^2$$

となります。 $z(0.05/2)$ の値は Excel なら「=NORM.S.INV(1-0.05/2)」で求められます($\cong 1.96$)。したがって、

$$N = \left(\frac{1.96}{0.1} \right)^2 = 10.96^2 \cong 120$$

となります。理由については長くなるので省略しますが、『統計解析のはなし』(石村貞夫著・東京図書)の P.183 あたりに書かれています。また、母平均の検定の場合であれば、『統計的方法のしくみ』(永田靖著、日科技連)の P.93 あたりに詳しいです。なお、サンプルサイズの決め方全般については、『サンプルサイズの決め方』(永田靖著、朝倉書店)に詳しいようです(値段か高くて、私は読んでいないのですが、目次を見るとひとつとおり網羅してあるようです)。

時系列分析についてはあまり詳しくないのですが、何らかの周期的な変化が仮定されるのであれば、その周期を使うのが普通だと思います。例えば、曜日によって売上にパターンが見られるのであれば、1日ごとにデータを何週間か取り、曜日によるパターンを検出、予測するとか、降水量など月ごとの変化を見るのであれば、月ごとにデータを何年分か取り、パターンを検出、予測するといった具合です(Excel であれば、FORECAST.ETS.STAT 関数で季節変動(周期的な変化のパターン)を加味した予測ができ、FORECAST.ETS.SEASONALITY 関数で繰り返しパターンの周期の検出ができます)。

