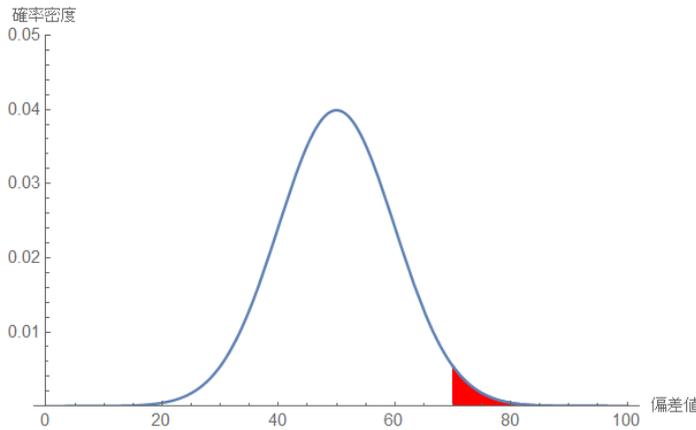


一家に遊女も寝たり 6σ <仮説検定 p 値>

2020年10月上旬 白田由香利

ありえない事を比較することを考えてみる。目白太郎君が全国模試で、数学の偏差値70点を取ることと、目白太郎君のお母さんがもっているX社の株が800円から明日、858.8円に価格が上昇することは、どちらが有りえなさ度が高いだろうか？



まず偏差値であるが、その分布は  $N(50, 100)$  と決まっている。観測値を正規分布として表して、平均値が50点、標準偏差が10点となるようにデータ調整したものが偏差値である。よって70点は  $+2\sigma$  のところである。70点以上の点を取る確率は、右図の赤い面積で表されるが、この確率は約2.275%である。

p-値とは、与えられた値よりも大きい値が出る確率を言う。この場合、70点のp値は2.275%である。逆に点数が平均の50点よりも低い30点だったとしよう。30点は  $-2\sigma$  である。30点のp値はそれ以下の点の確率を表すので、同じく2.275%となる。

とある全国模試で、偏差値20点を取って、親から「お前の下に何人いると思っているのだ」としかられたら、以下のように計算しよう。

受験者数は1万人、20点は  $-3\sigma$  なので、p値は0.00135。  $10000 \times 0.00135 = 13.5$

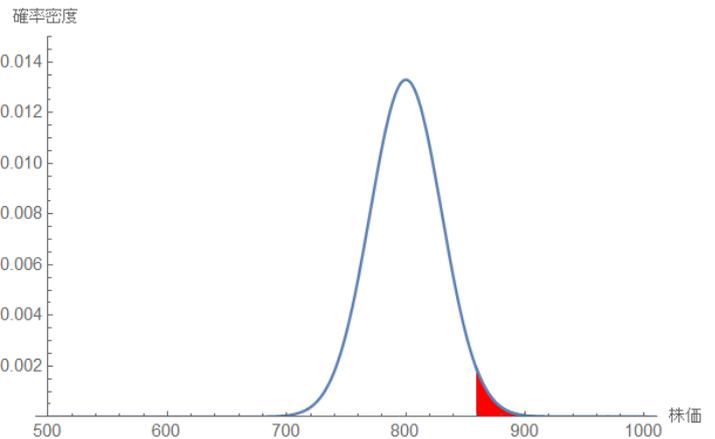
「私を含めて13.5人です」

もちろん、親はこの計算を期待しているのではない。

次に株の値上がり問題を考える。この場合、X社の株値が最近どのような確率分布に従っているかという情報が必要である。仮に、  $N(800, 900)$  としよう(右図参照)。  $\sigma = 30$  である。

$$(858.8 \text{ 円} - 800 \text{ 円}) \div 30 \text{ 円} = 1.96$$

であるので、858.8円は  $1.96\sigma$  に相当する。よって、そのp値は2.5%である。正規分布の95%信頼区間が  $-1.96\sigma$  から  $+1.96\sigma$  の区間であったことを思い出そう。



P値が小さいほうが、有り得なさ度が高い。

よって、p値が2.275%であった、偏差値70点を取る、のほうが有り得なさ度が高い。

もちろん、偏差値の正規分布  $N(50, 100)$  はあくまで偏差値の分布であって、目白太郎君と、高田馬場恵子さんでは、点数の確率分布は異なる。また、株価が正規分布に従うのは、平穏な動きのときだけであり、株価の大幅な下落及び上昇の際には、べき分布に従っている、ということが分かってきた(青山2008, マンデルブロ2008)。

P 値とは、仮説検定において、その値以上の極端なことが起こる確率を表している。片側検定で有意水準を 5% に設定して、帰無仮説を検証する場合、p 値が 2% であれば、有意水準 5% よりも小さく、それは非常に有り得ないことと解釈でき、帰無仮説は棄却される。気持ちとしては、「このような有り得ない、そして、稀なる、珍なることが起きたのであるからして、帰無仮説が間違っていたのだろう」である。

例を示す。1 ポンドステーキを注文したのに、どうも小さい。帰無仮説と対立仮説を立てる。

帰無仮説：重さの母平均=1 ポンド

対立仮説：重さの母平均<1 ポンド

5 回注文して、標本平均を計算したら、その p 値が 2% であった。よって帰無仮説棄却。この店の 1 ポンドステーキの重さの母平均は 1 ポンド未満である、となる。

有意水準 1% の場合はどうなるであろうか？ p 値が 2% であれば、有意水準 1% よりも大きいので、非常に有り得ないことではない、となり、帰無仮説は棄却されない。積極的に支持できる状況でもないの、結論は「重さの母平均=1 ポンドを否定することはできなかった」くらいにとどめる。

仮説検定では、統計値 XXX は、YYY という確率分布に従う、という前提がある。例えば、統計量 t は自由度 29 の t 分布に従う、というようにである。経営で統計を使う多くの人は、証明の理解にこだわらず、「そういう統計の理論があって、先人が証明してくれたのだから、有難く使おう」位の気持ちで使えばよいと思う。証明ができなくても気にしなくてよい。母集団の確率分布の形状も、特定されるパラメトリックなものだけで、正規分布、t 分布、 $\chi$  二乗分布、F 分布が使えればまずは十分である。注意点は、使えるときの条件があることである。例えば、確率変数 X は正規分布に従うこと、あるいは、分散分析で各グループの確率変数の分散は等しいこと、などがある。この前提条件は、その都度、説明していく。

「クリスマスイブの夜、家の前に、前からほしいと思っていたロシアンブルーの猫が捨てられていたの、ラッキーだわ」自分の周りで有り得ないことがおこったら、どの位の p 値になるか、ちょっと考えてみましょう。

参考文献：

\* 青山秀明 他、「経済物理学」、共立出版、2008。

\* ベノワ・B・マンデルブロ、リチャード・L・ハドソン、(高安 秀樹、雨宮 絵理 訳)、「禁断の市場フラクタルでみるリスクとリターン」、東洋経済新報社、2008。

終わり

引用元：一家に遊女も寝たり萩と月 芭蕉