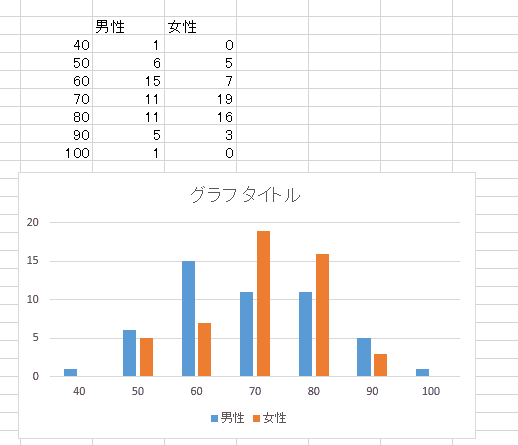
【第7部】様々な分布

性別によるテストの成績の分布から「男女の間にテストの成績の差がある」といえるでしょうか。差がある、もしくは差がないとするならば数学的に検証（検定といいます）する必要があります。

検定するうえでは「分布」を理解する必要があります。特に正規分布の概念を理解し、正規分布から派生するその他の分布と、分散や平均などの「数値の意味」を理解する必要があります。

　理解したうえで分析すると、性別による成績の分布には「差がない」という結論になります。どうしてそのような結論になるか、まずは分布の理解からはじめましょう。



# 二項分布

二項分布とは「n回のベルヌーイ試行（表と裏など、2つの種類しかない試行）の中で何回成功するか」という成功回数の分布です。「サイコロを10回振った時の目の合計が30になる確率」や「コインを10回投げたときに表が5回出る確率」などを考えるときに使います。高校の確率の授業などでも取り上げられる、基本的な分布の１つです。

試行回数をn 、成功する回数をx 、成功確率をpとした場合、n回中x回成功する確率f(x)は

となります。の部分は二項係数として　　と表記されることもあります。 例えばコインを投げて表が出る確率確率を50％としたとき、2回投げて1回しか表が出ない確率は、

2\*25%=50%

となります。

## メレとパスカル

シュバリエ・ド・メレは友人のパスカルに相談しました。「1つのサイコロを4回投げて6が1回でも出れば勝ちという賭けでは勝てた。しかし2つのサイコロを24回投げて6のゾロ目が出れば勝ちという賭けになってから勝てなくなった」というものです。

パスカルは勝率を計算しました。１つのサイコロを4回投げて一度も6が出ない確率は＝0.482なので、最初の賭けの勝率は1－0.482＝0.518＝51.2％となります。２つのサイコロを24回投げて1度も6のゾロ目が出ない確率は＝0.509なので2番目の賭けの勝率は1-0.509=0.491=49.1％となります。確かに最初の賭けは勝率5割を超えますが、2番目の賭けは勝率5割を切ります。メレの勝てなくなったという証言は正しいといえます。

## ギャンブラーの法則

では2番目の賭けはどうすれば有利なるのでしょう。事象の起きる確率をPとしたときに、それをn回したときに勝率が5割を超えれば勝てるとすると、nを1-(1-p)n>50%となるようにすれば良いことになります。確率が50％になるnを求めると

両辺の対数をとると

両辺をで割って01𝑛𝑛𝑗𝑖 =

となります。pにとをそれぞれ入れると、前者や約3.8、後者は約24.6になります。つまりメレは前者であれば4回以上、後者であれば25回以上行うようなゲーム設計にすれば賭けに有利になるといえます。

## 二項分布の練習問題

【問題1】

コインを10回投げたとき、表が3回出る割合を求めなさい。（コインは表と裏が出る確率はそれぞれ50％とする）

【問題2】

１～６までの目があるサイコロを10回投げたとき、合計が35になる確率を求めなさい。（サイコロの目が出る確率はそれぞれとする。）

# ポアッソン分布

ポアソン分布とは、単位当たりにn回の事象が起こる割合の分布です1日あたりの交通事故の数、大量生産された部品において不良品が発生する割合などは、ポアソン分布になります。ポアソン分布は二項分布において成功確率が非常に低い場合の分布です。

## 二項分布からポアソン分布を作る

ポアソン分布の公式は確率λと回数xに対して、eは自然対数の底とすると

とあらわされます。この公式は二項分布からも導き出せます。

1. ポアソン分布の成功回数をλ=npとします。として二項分布に代入します。
2. 一部を入れ替えます
3. nの極限（無限に大きくする）をとります。の極限は1になります。 の極限はに、 は0になります。

1時間に1回しか起こらない出来事が、1時間に2回起こる確率は

eは自然対数の底（2.718281828）なので、e-1は0.367879、これを1/2にすればいいので、0.183939721、つまり1時間に1回しか起こらない出来事が、1時間に2回起こる確率は約18.4％です。

## 待ち時間を計算してみよう（指数分布）

では非常に希な事象が起こる時間間隔を計算してみましょう。時間tが経過した時に事象が起きる確率f(t)と確率密度f(t)dtを求めます。

単位時間当たりで事象が起こる確率はλなので、時間0からtまでに起こる確率はλtとなります。

まずは事象が時間0からtの間にx=0回起こる確率を求めます。ポアソン分布を使って計算すると、

次に時間0~xのあとの微細な時間t~dtの間に事象が起こる確率を求めます。これはλにdtを掛けたλdtです。よって事象が起こる時間間隔xの確率密度関数f(t)Δtは時間0~xまでに起こらなかった確率に時間x~Δxまでに起こった確率を掛ければいいので

両辺からdtを割ると

となります。この分布は指数分布と呼ばれています。

## ポアソン分布の練習問題

【問題1】

学食のカレーには肉が平均2つ入っているとします。カレーに肉が１つしか入っていない確率と、3つ入っている確率をそれぞれ求めなさい

【解答】

ポアソン分布はλ＝2　x=3の時の確率を求める

　答え：18.0%

【問題2】

　1000年間に1回しか出現しないアイドルが10年間に2回出現する確率を求めなさい。

【ヒント】λに0.01、xに2を入れて計算します。答えは0.0050%です。