

Sample

統計検定®で学ぶ!

# 統計学

SAMPLE

記述  
統計

確率  
・  
確率分布

推測  
統計



代ゼミライセンススクール  
Yozemi License School

## Sample

## ご受講をお考えのみなさまへ

『数学は得意だが、どうも統計だけは…』とか、『数学が苦手だから、まして統計なんて…』といった声はよく耳にします。然るに、『統計は得意だけれど、数学は苦手である』という話は寡聞にして知りません。

しかしながら、数学と統計学の間には単純な包含関係は存在しません。誤解を怖れずに言わせて頂ければ両者は別物であるといえます。そこで、統計を身近に感じ使える道具とするために、本質的理解にこだわった講義を試みようと考えました。

統計検定2級に合格することは、この講座の大きな目標ですが、もちろんそれは結果として獲得されるものの一つにしか過ぎません。ビッグデータの時代を生き抜くためのリベラル・アーツとして、さらには自然科学のみならず、社会科学、人文科学などおよそ科学と称されるもののリング・フランカとして、統計学の修得を目指して行きたいと思えます。

この冊子は、「統計検定®で学ぶ！統計学」HPに掲載中の体験動画用サンプルです。動画と合わせて、講座の雰囲気少しでもお伝え出来ればと思い作成いたしました。実際のテキストは3分野合計で500ページ超、統計検定2級過去問・オリジナル問題を約100題も掲載した大ボリュームとなっています。問題を解くことにこだわった統計学講座、ぜひあなたも受講してみませんか。

「統計検定®で学ぶ！統計学」ホームページはこちら  
※講義の一部を体験動画として公開しています。

【URL】 <http://www.yls.ac/course/statistics/>

【QRコード】



## 記述統計

記述統計とは、収集したデータを集計し、それらを質的ではなく量的に捉えて解釈することで、データの示す性質や分布・傾向などを要約する方法です。特に膨大なデータを扱う場合は、事前に解釈し易い形式にしておかなければ、ただ数字を眺めて時間が過ぎるということになりかねません。

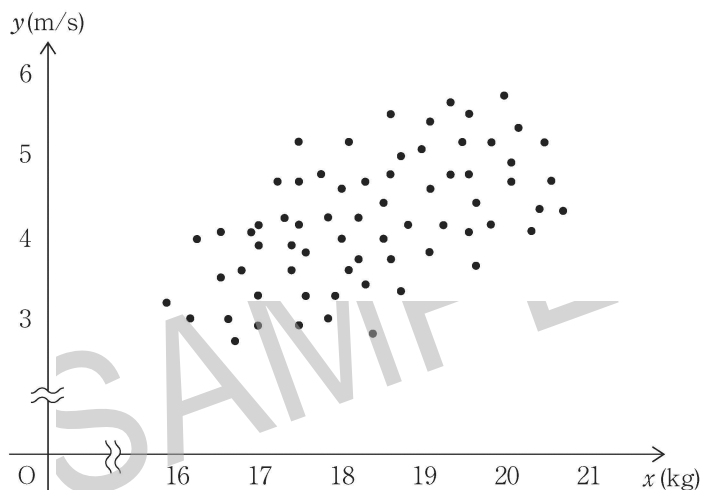
与えられたデータを先ず分析し、次に再総合する還元的(解析的)な手法の巧拙が問われることとなります。手法の基本となるのは、層別・分類に代表される空間的解析と時系列プロットに代表される時間的解析です。

一方、近年の計算機の進化にともない、従来ならば推測統計に頼らざるを得なかった大きな母集団の属性処理が比較的簡単にできるようになってきています。このビッグデータとよばれる新しい解析分野の主役もまた記述統計なのです。

何れにせよ記述統計は統計学の基礎となる分野ですから、ここで用いられる諸々の用語や手法を確実なものにしておくことは肝要です。

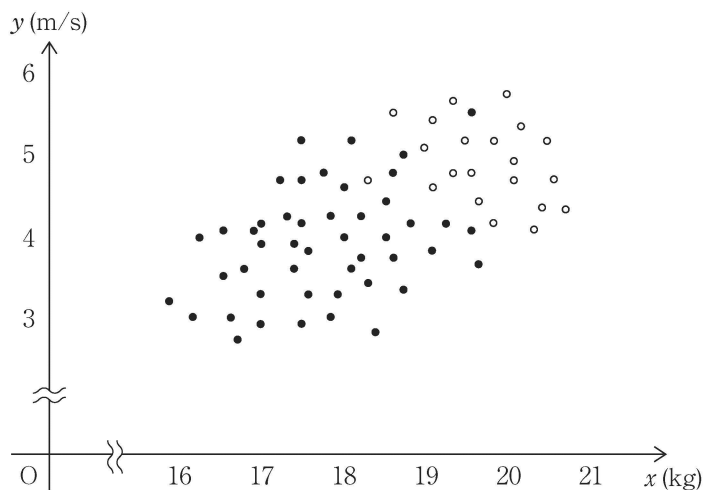
## Sample 1.1 偏相関係数 (partial correlation coefficient)

ある幼稚園の園児を対象に体重  $x$  (kg) と走力 (短距離を走る速さ)  $y$  (m/s) について散布図をつくったところ次図のようになりました。



Sample 図1.1

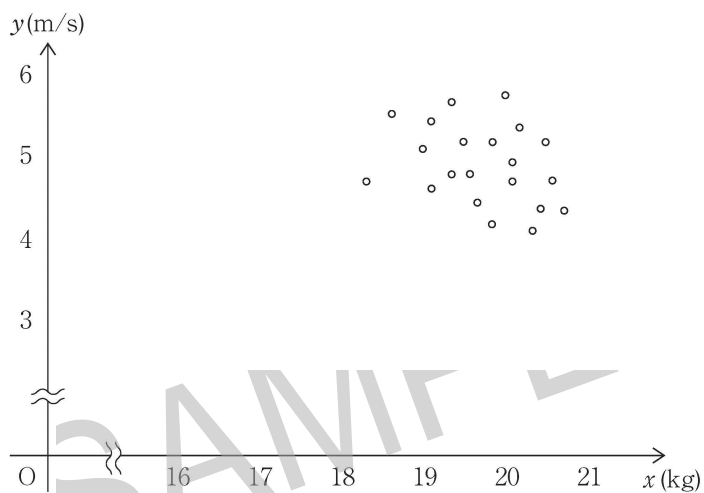
この散布図だけ眺めれば、体重と走力には中位の(やや強い)正の相関があると判断できます。則ち、太った子ほど足が速いということになります。これは日常の観察からすると不自然な結果です。そこで、体重、走力以外の第三の変量として年齢を考慮することになります。先の散布図において、年長児の変量だけを○点で再表示すると、次図のようになります。



Sample 図1.2

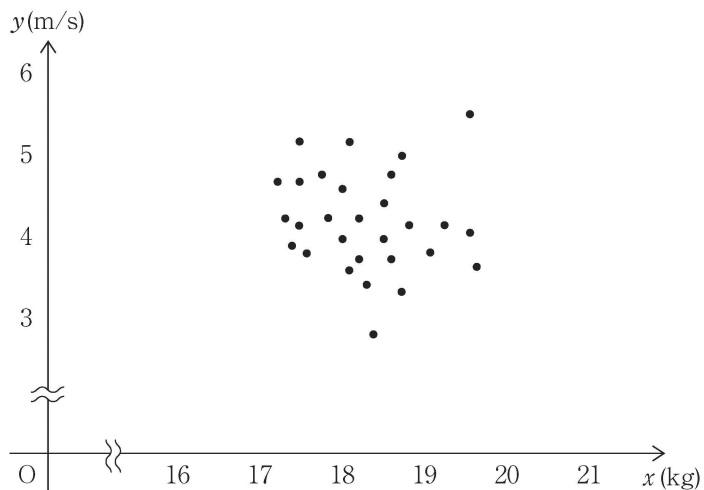


改めて、年長児だけを対象にした散布図を示すと次図のようになり、体重と走力には中位の(やや強い)負の相関があると判断できます。



Sample 図1.3

年中児だけを対象にした散布図は次図のようになり、体重と走力には低位ではありますが、やはり負の相関があると判断できます。



Sample 図1.4

この例に観るように、年齢という交絡因子の影響を考慮しないとき、体重と走力の間に見かけ上の相関、則ち擬相関(擬似相関)が出現します。そこで、交絡因子の影響を排除した相関を考えることが重要になります。

その目的で用意されるのが**偏相関係数**(partial correlation coefficient)で、  
体重( $x$ )、走力( $y$ )、年齢( $z$ )の各々のデータに対し、

体重と走力の相関係数： $r_{xy}$

体重と年齢の相関係数： $r_{xz}$

走力と年齢の相関係数： $r_{yz}$

とすると、 $z$ の影響を除いた $x$ 、 $y$ の偏相関係数 $r_{xy\cdot z}$ は、

Sample (1-1) 
$$r_{xy\cdot z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{1-r_{xz}^2}\sqrt{1-r_{yz}^2}}$$

と定義されます。

ここで、 $r_{xy}=0.55$ 、 $r_{xz}=0.89$ 、 $r_{yz}=0.76$ である場合、 $r_{xy\cdot z}$ は、

$$r_{xy\cdot z} = \frac{0.55 - 0.89 \times 0.76}{\sqrt{1 - 0.89^2} \sqrt{1 - 0.76^2}} = \frac{-0.126}{0.456 \times 0.649} = -0.425 \dots$$

と求められます。

## Sample 1.2 回帰直線 (regression line)

2つの変量が冬の最低気温 $x$ と防寒具の売り上げ $y$ のような対等な関係ではないものとして、喫煙量 $x$ と慢性閉塞性肺疾患の発症率 $y$ の関係、最高気温 $x$ と電力消費量 $y$ の関係など様々な例を考えることができます。このような場合、 $x$ と $y$ の関連性の強弱より、 $x$ から $y$ への**対応**(correspondence)が重視されます。則ち、ある $x$ を入力すると、いかなる $y$ が出力されるのかという関数(写像)を求めることが主たる目的になります。そして、この関数を用いて、両者の対応関係を説明したり、将来の予測を図ります。

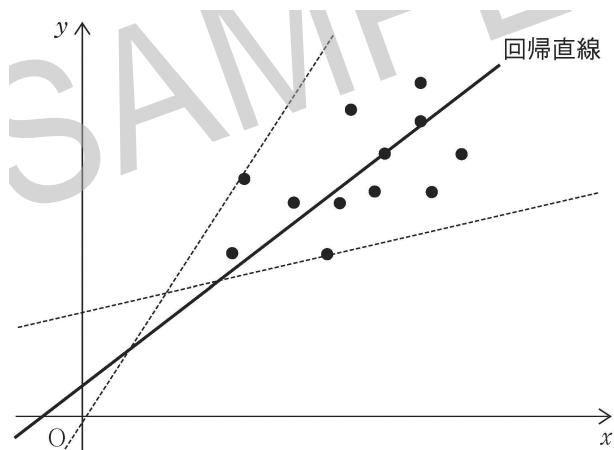
入力側(出発側)の変数 $x$ を**説明変数**(explanatory variable)とか**独立変数**(independent variable)などおよび、出力側(到着側)の変数 $y$ を**応答変数**(response variable)とか**従属変数**(dependent variable)などおよびます。散布図を描く際は、横軸に説明変数 $x$ を縦軸に応答変数 $y$ をとります。

ところで、このような  $x$  から  $y$  への対応の中でもっとも単純なものは線形な関係で、それは1次関数

$$y = \alpha + \beta x$$

で表現されます。

$xy$  平面上では直線で表わされ、 $\alpha$  が  $y$  切片、 $\beta$  が直線の傾きを示します。相関係数が  $\pm 1$  でない限り、散布図において全ての点が同一直線上にのるわけではありませんから、与えられた散布図のデータにもっともよく適応した直線を探ることになります。この直線を**回帰直線**とよび、そのときの  $\alpha$ 、 $\beta$  を**回帰係数**(regression coefficient)とよびます。



Sample 図1.5

# SAPIX YOZEMI GROUP

授業期間の前後を問わず、次の行為は禁じられています。

1. 販売を目的とした複製，複写，転載，加工
2. 他者への譲渡，転売，貸与

※統計検定® は一般財団法人統計質保証推進協会の登録商標です。  
本書籍の内容について、一般財団法人統計質保証推進協会は関与していません。