

年 組 番 氏名

SS 理数探究 I 統計処理①(代表値、ヒストグラム)

統計学とは「数値の集団から、その集団についてのどのような情報を引き出せるか」を研究する数学です。

気候や降水量、運動の記録、身長や体重などのように、ある集団を構成する人や物の特性を数量的に表すものを変量といい、調査や実験などで得られた変量の観測値や測定値の集まりを**データ**という。

統計的な調査には、対象全体からデータを集めて調べる**全数調査**があるが、これは多くの費用と労力を必要とするので簡単には行えません。これに対して、調査の対象全体からその一部を抜き出して調べる**標本調査**があります。標本調査の場合、調査の対象全体を**母集団**(ぼしゅうだん)といい、それに属する個々の対象を**個体**、個体の総数を**母集団の大きさ**とといいます。また、母集団から抜き出された個体の集合を**標本**といい、母集団から標本を抜き出すことを**抽出**とといいます。

統計学には2本の柱、記述統計と推測統計があります。

- (1) 記述統計・・・データの平均値や散らばりの大きさを調べたり、図やグラフを描いてデータの性質を説明するデータの処理の仕方。たとえば、データの性質を自分で納得し、その内容を第三者に公表したい場合など。
- (2) 推測統計・・・既知の標本から未知の母集団を推測するという方針でのデータの処理の仕方。たとえば、同じ学年の生徒の中から10人くらいの代表を選んでその体重を測り、そのデータによって同学年の生徒全体の体重の現状を推し測るなど。

○母集団の中の割合を推定する。

問：いま、黒と白、合計1000個の豆が入った袋から100個を取り出したとき、黒が20個含まれていたとします。このとき袋の中の黒豆の数はいくつでしょうか。

解答：

○母集団の大きさを推定する。(再捕獲法)

問：全体数がわからない白豆の入った袋から、100個を取り出して黒く塗ります。その黒く塗った豆を再び袋に戻して、十分に混ぜます。そして再度100個の豆を取り出したとき、黒い豆がそのうち20個入っていたとします。袋の中の全部の白豆の数はいくつでしょうか。

解答：

統計処理①

○薬の効果を調べる。

問：生徒 830 人の学校で、インフルエンザにかかった人が 43 人もありました。それらの生徒を調べると、約半数の 22 人が予防注射を受けていて、残りの 21 人は受けていないことがわかりました。予防注射を受けた人のほうが、少しですが受けなかった人より多いという結果が出ました。しかし、無効とはいえないようです。その理由はなぜですか。

解答：

○データの代表値 最頻値（モード）、中央値（メジアン）、平均値（ミーン）

次の 8 個のデータについて考えましょう。

5 4 3 6 4 8 5 5

これらのデータの代表値を考えてみましょう。

- ・データにおいて、もっとも個数の多い値を、**最頻値またはモード**という。

上の例では、最頻値は（ ）です。

- ・データを小さい順に並べたとき、中央の位置にくる値を**中央値またはメジアン**という。

データの大きさが偶数のとき、中央に 2 つの値が並ぶが、その場合は 2 つの値の平均値を中央値とします。

上の例では、小さい順に並べ替えると

3 4 4 5 5 5 6 8

よって中央値は（ ）です。

- ・データの総和をデータの個数で割った値を**平均値**といいます。

上の例では、平均値は（ ）です。

問：A 班 5 人、B 班 5 人の試験の成績の結果が発表されました。A 班の 5 人の成績の平均点は、B 班の平均点より少し下でした。それなのに先生の説明では、A 班のほとんどの人は、B 班の誰よりも成績が上だということです。それはなぜでしょうか。

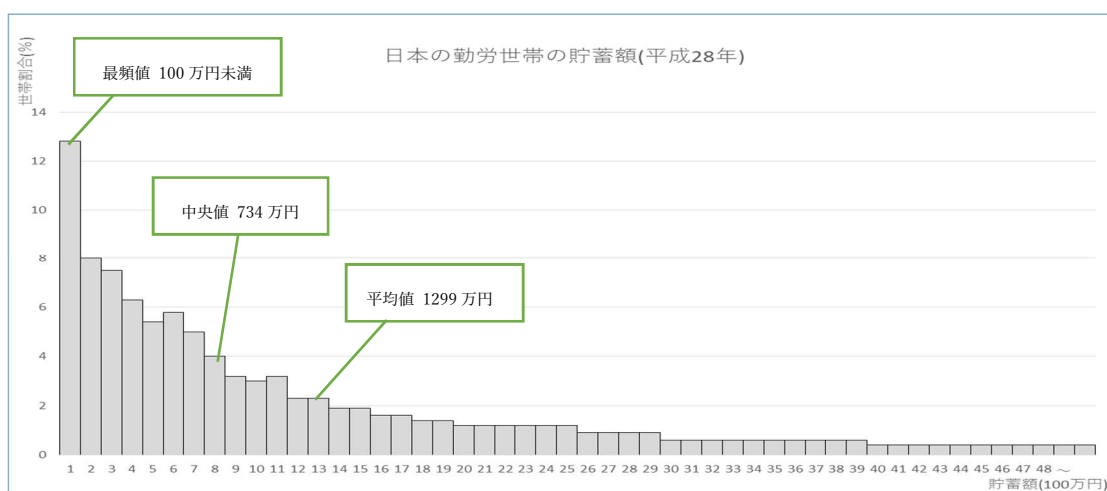
解答：

統計処理①

○平均値が多数を代表するとは限らない

「勤労者世帯(2人以上の世帯)の平均貯蓄は約1300万円」と聞くと多いと思いますか。それは「平均値近辺の人が最も多い」と無意識に判断したためでしょう。

下の貯蓄の分布を見ると、貯蓄が100万円以下の世帯の割合が最も多く12.8%を占めます。そして、貯蓄額が多い世帯ほど、世帯の割合が小さくなっていきます。平均値は1299万円ですが、平均値を下回る世帯が3分の2以上を占めています。中央値は734万円で、最頻値は100万円未満です。



貯蓄が極めて多い、ごく一部の少数の世帯が、平均値を上げています。

このように平均値には、極端なデータ(外れ値)の影響を受けやすいという欠点があります。平均値をデータの代表値とするにはデータの分布を調べる必要があります。

○ヒストグラム(度数分布図)

度数分布表を分かりやすく表現するグラフとして**ヒストグラム(度数分布図)**があります。

ヒストグラムとは、ある特定のデータを区間ごとに区切り、各区間の個数や数値のばらつきを棒グラフに似た形の図で表現するグラフです。

ヒストグラムでは、横軸の階級に連続した順番があります。あくまでも、一つの連続したデータの度数の分布を伝えるツールがヒストグラムです。

棒グラフでは横軸の順番は、作成者が自由に並べ替えられます。各グラフの長方形ごとに独立しており、他の長方形とは連続した関係性はありません。

グラフの表現方法としても、ヒストグラムは各階級の長方形同士をくっつける事によって、連続した意味合いがあることを表現しています。一方の棒グラフは、各データを表す長方形同士の間隔が空いていることが特徴です。

統計処理①

○平均値と中央値の比較

ある2つの会社 A、B の平均給与と中央値を調べたものです。ただし、1,000 円未満は切り捨てました。

平均給与は A社 55 万円 B社 50 万円 です。

中央値は A社 30 万円 B社 40 万円 です。

問1 平均値だけを考えた場合、あなたはどの会社に入社希望しますか。

解答：

問2 中央値だけを考えた場合、あなたはどの会社に入社希望しますか。

解答：

問3 2社の平均値と中央値を比べると、2社とも平均値より中央値が低くなっています。平均値より高い人と平均値より低い人ではどちらが多いでしょうか。また、平均値より中央値が低いのはなぜだと思えますか。

解答：

問4 平均値はA社がB社より高いのですが、中央値はA社がB社より低くなっています。平均値より低い給与の人の割合が高いのはどちらだと考えられますか。

また、最頻値はどちらが低いと予想されますか。

解答：

問5 給与の代表値として平均値、最頻値、中央値のうちどれを選ぶと良いでしょうか。

解答：