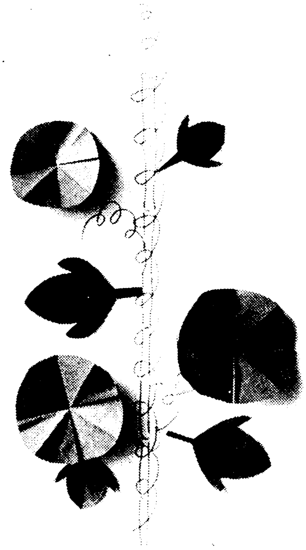
歳時記



「朝顔に　つるべとられて　もらい水」

ものぐさにしているうちに朝顔がはびこってしまい，とうとうつるべにまで巻きついてしまった。掃除するのも面倒だ，どうせ秋になれば枯れてしまうのだから，それまでもらい水をしてすませてしまおう，という魂胆のオカミサンがつくった句。

今月のおもな行事

○8～9日　結城市統計調査員大会（山梨県）
○9～10日　昭和50年事業所基本調査区設定地方事務連絡打合せ会（静岡県）
○15日　工業統計調査結果速報公表
○25～26日　関東ブロック県民所得事務研究会(群馬県)
　那珂郡統計調査員総会・研修会（栃木県）
○31～9月2日　茨城県都市統計事務協議会先進都市視察研修会（広島県・山口県）

さいじき

●シリーズ「統計」

「標準偏差と王のホームラン」

Σ（シグマ）の意味を5月号で見ました。今回もまたシグマの話なのですが，今度は大文字Σでなく小文字のσ（シグマ）です。σは標準偏差をあらわすときに用いられる記号です。もっと面倒臭く言うと，ある集団全体（これを母集団と言います。）の標準偏差（これを母標準偏差と言います）をσであらわし，母集団からひっぱり出した標本（サンプル）の標準偏差（これを標本標準偏差と言います）をローマ字の小文字 s であらわします。

ある集団の姿を何かの特性値を通してあらわす場合，普通はまず算術平均を出してその集団の中心がどの辺にあるのかを見る方法がとられますが，しかし平均だけではなかなか正確な姿をつかむことはむずかしいため，観察値xが平均のまわりにどんなふうに位置しているのか，つまりバラツキ具合を調べる方法が必要です。このバラツキ具合のことを統計では分散度と呼んでいます。分散度をあらわす特性にはレンヂ（範囲）や四分位偏差などがありますが，特性値として優れているためひんぱんに使われるものにかの有名な標準偏差があります。これがσです。これは個々の観察値（変量x_i）とそれらの平均値$\bar{x}$との間の距離（これを統計では偏差と呼びます）について調べる方法をとるのですが，御存知のとおり，算術平均には個々の変量x_iから平均$\bar{x}$をひいた値をそのまま合計すると0になってしまうという性質がありました。0になってしまってはこの先計算になりませんから，$(x_i-\bar{x})$を2乗してマイナスの符号をとり払ってしまう方法をとるわけです。

式であらわすと，

$$\Sigma\,(x_i-\bar{x})^2$$

となります。ここで話のついでですから算術平均が持っているもうひとつの性質を紹介しましょう。$\Sigma\,(x_i-\bar{x})^2$の値というのは同じようにx_iから$\bar{x}$でない任意の数x_oをひいたものを2乗して総和をとった値，つまり$\Sigma\,(x_i-x_o)^2$，よりも絶えず小さい，という性質です。例えば，

2・3・4・5・6

という5つの数字の集まりについて，個々の数字から$\bar{x}$でない数，例えば3をひいて計算してみると，

$$(2-3)^2+(3-3)^2+(4-3)^2+(5-3)^2+(6-3)^2$$
$$=1+0+1+4+9=15$$

となって，算術平均値4をひいて計算したときの値10よりも大きく出ました。3に限らず4以外の数字を入れて計算するとどれも大きくなってしまいます。このように算術平均に対する変量の偏差の2乗和は他のどのような値に対する変量の偏差の2乗和よりも常に小さいのです。算術平均は変量の偏差の最小2乗値であると言えます。ということはつまり，偏差の2乗和を最小にする基準値は算術平均だということです。「最小2乗」というのはどうもどこかで聞いた言葉だと思っておられる方は6月号を読んで下さい。そこに算術平均と最小2乗法についての説明が載っています。算術平均のこの性質を回帰分析に利用されているためにきっと「最小2乗法」というように呼ばれているのでしょう。

さて標準偏差σはこの　$\Sigma\,(x_i-\bar{x})^2$を項数nで割って，$\sqrt{\ }$で開いて出します。2乗して計算したので最後に$\sqrt{\ }$で開いて元に戻してやるのです。元に戻さないままだと「σ^2」で，「分散」と呼んで利用します。こうして標準偏差は式であらわすと

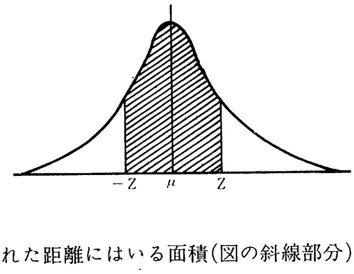
$$\sigma=\sqrt{\frac{\Sigma\,(x_i-\bar{x})^2}{n}}\ \cdots\cdots\cdots\cdots\cdots\cdots(1)$$

とあらわされます。実際の計算では$\bar{x}$の値が小数点以下何位にもなったりして計算が複雑面倒となって，すぐに計算をまちがったりしてしまうため(1)式を変形してもっと計算しやすい形にします。

(1)式は次のようにあらわせます。

$$\sigma=\sqrt{\frac{\Sigma x_i^2}{n}-\bar{x}^2}\ \cdots\cdots\cdots\cdots\cdots\cdots(2)$$

この方が計算しやすいことが多いのです。こうして，算術平均と標準偏差とは切っても切れない縁が有るため，よく2つ一緒に組み合わされて表現されます。それによって平均から標準偏差であらわされる一定の距離の間に変量x_iがどのくらい分布しているのかがわかります。ある観察値が正規分布に従う変量と考えられるときの分布図は下のような形をとります。



いま，この分布図が面積1，平均μ（ミュー，母平均）0，標準偏差σ 1の正規分布をあらわすとき，中央の平均μから＋，－の両方向にZまで離れた距離にはいる面積（図の斜線部分）は下表のようにすでにわかっています。

この表と図とはちょうど「合いかぎ」のようなもので，正規分布というひと言であらわされるさまざまな形のかぎ穴にすべて合う合いかぎだと考えるとわかりやすいでしょう。平均と標準偏差はデータによって当然異なり，正規分

冨　永　重　己

布はこの２つの値によっていろいろな形になるのですが，このμが0，σが１という図表によってどんな形の正規分布もこれにあてはめて考えることが可能なわけです。この図表から，平均から１σ離れたところには観察値の68.27％が含まれることがわかります。同様に２σのところまでに 95.45％，３σのところまでにはほぼ全部の99.73％が含まれます。

Z	斜線部の面積
0	0
0.5	0.3829
1.0	0.6827
1.5	0.8664
2.0	0.9545
2.5	0.9876
3.0	0.9973
∞	1.0000

これを実際のデータを例に見てみることにしましょう。５月号で巨人軍の王選手に出場して頂き過去10年間の本塁打数の合計と平均を出しましたが，あの記録を見つけるのには古新聞をひっぱり出して大変でした。時節柄関心の高いところですのでここに無理やり再登場してもらうことにしましょう。過去10年の結果では分布と呼ぶにはあまりにもデータの数が少ないのですが，まあそこは偉大なる打者，王選手のバットから生み出された本塁打ですから，ここは無理やり，正規分布に従うものと考えてしまった方が話しやすいのです。そこで早速彼の本塁打数について標準偏差を計算してみます。

年　度	本塁打数 x	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	x^2
昭42年度	47	1.4	1.96	2,209
43	49	3.4	11.56	2,401
44	44	−1.6	2.56	1,936
45	47	1.4	1.96	2,209
46	39	−6.6	43.56	1,521
47	48	2.4	5.76	2,304
48	51	5.4	29.16	2,601
49	49	3.4	11.56	2,401
50	33	12.6	158.76	1,089
51	49	3.4	11.56	2,401

$n=10,\ \Sigma x=456,\ \Sigma(x_i-\bar{x})=0$

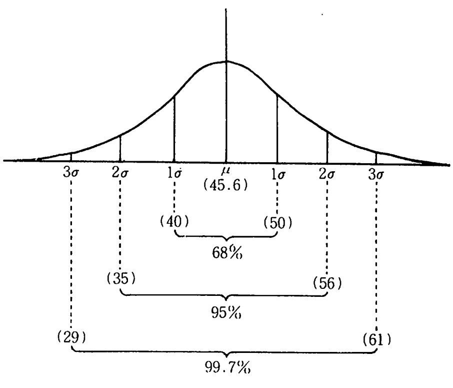
$\Sigma(x_i-\bar{x})^2=278.4,\ \Sigma x^2=21{,}072$

$\bar{x}=\Sigma x/n=456/10=45.6,\ \bar{x}^2=2{,}079.36$

先の①式または②式を使って計算すると標準偏差は 5.3になります。

こうして過去10年間の王選手の本塁打の平均は45.6本，標準偏差は 5.3本とわかりました。これをさきほどの正規分布の面積の表と照らしあわせてみてみることにしましょう。σが5.3ですから平均から１σ離れた範囲というのは，王の本塁打の場合45.6±5.3ですから40.3〜50.9本の間になり，この間にデータの68％が含まれるということです。つまり40.3本から50.9本の本塁打を打った年は，過去10年のうちの68％，約７割ですから，７回はあるだろうと予想されます。逆に言えば，40.3本から50.9本の間からはみ出してしまう本塁打数が記録された年が３回有ると思われます。どれ，表を見てみると……46年度の39本と48年度の51本と50年度の33本……やっぱり３回はみ出していました。では次にＺの値が２のときは，表の面積は 95.45％，45.6本±(2×5.3)ですからこれは35.0〜56.2です。35本から56本の本塁打を打ったのは，データの約95％です。王が100年打ったとしたらそのうちの95回ということになりますが，ここでは10年間の記録ですから，まあ10回のうち全部か９回のどちらかというところでしょう。逆に見て，35〜56本からはみ出てしまう本数を記録した年度は１回有るかないか，です。表を見てみると……その１回がありました。50年の33本です。50年は33本しか打てませんでした。確か50年は巨人軍が最下位の年でした。王が打てなかったのがその原因のひとつになったのか，それともチームがふるわないので王のバットも湿ってしまったのか，その辺はわかりませんが，Ｚが３になるともう99.7％ですから10年の記録では29本から61本以外の本塁打数が記録された年はありません。

これを図にすると下のようになります。一般にσが大きくなるほどバラツキ方も大きいことをあらわします。10年の結果をうまく利用できたのも王選手だからこそです。王さまさまです。ことし，王選手はアーロンの記録を破り，まだまだ活躍するでしょうが，図から見て，30本打てなくなったときが王が王でなくなるとき（引退）かもしれません。



標準偏差から少し確率の話になってきたところで今回はこれまで。

（県統計課　消費統計係）

県民所得の簡易推計とは

利用者の声

統計いばらきの1月号（1977年）に，「昭和50年度県民所得簡易推計結果」が掲載されていたが，簡易推計とはどんなものなのか。

担当者の声

毎年発表している県民所得統計は，県経済の実態やその動向を把握するための最も重要な指標の1つであり，県行政施策をはじめ，国民経済の地域的分析や地域開発施策等の基礎資料として各方面で利用され，今後さらにその充実が望まれているところです。

しかし，県民所得統計は，その推計方法が各種統計情報の組み合せ，加工によっているために，各方面にわたり膨大な資料を必要としています。ところが，これらの基礎資料のうちある部分については，官公庁，民間での発表時期に制約があるため，県民所得統計の公表は対象年度に比較してかなり遅れざるを得ない状況にあります。茨城県の場合，対象年度から1年を経過した時点で公表されているのが現状ですので，経済指標としての有用性に比べて，時期的な面で利用価値が減殺されています。

そこで，最近は早期推計による公表が強く要請されてきており，いくつかの都道府県では実施に移されています。その方法としては，おおよそ次の3つに大別して考えることができます。

(1) 積上げ式による方法（積上げ方式）

(2) 統計式による方法（統計式）

(3) (1)と(2)を併用する方法（併用方式）

現在実施している各県をみてみますと，(1)，(2)の方法が多く，一部の県で(3)の折衷方法を用いているようです。

本県においては，従来，積上げ方法による速報を公表していましたが，推計方法，内容とも確報と変わるものでなく，時期的な面でもより早期の公表が今後とも難しいと思われますので，これにかわる早期推計の可能性を検討してきました。

積上げ式による早期推計の場合，いかなる推計方法であれ，どうしても必要とする資料が多くなり，時期的な面で限界がでてきます。それに比べて，統計式による方法は，少ない情報を用いて，より早い時期に推計値を得ることができ，推計作業の進め方によっては，かなり高い精度の予測値を得ることも可能です。

本県でも，統計式による方法で昭和48年度以降試算してきましたが，昭和50年度分から，「県民所得簡易推計結果」として公表しています。統計式等による早期推計結果は，早期の利用に供する見込額であるという点で，本来の県民所得統計を時期的な面で補完するもので，地域間の比較等の資料として使用するためには各県共通の標準方式の開発を待たなければなりません。国においても，標準方式を検討中ですが，確立したモデルが得られず，モデルの理論構成が不充分でした。なかでも各種特性値の形式的判断はともかく，相互関連等の実質的判断になると，十分にそれらの数値を利用しきれない所もあります。しかしながら，各方面からの早期公表の要望に答えるために，昭和50年度から公表に踏み切ったものです。

公表の時期は，早いほど利用者の要請に答えることになる訳ですが，早ければそれだけ予測精度が低まると一般的には考えられますので，その兼合いが問題となっています。ただ早期推計に対する要望は，早期の経済見通しのためとか，経済動向の把握のためという場合がほとんどであるため，確報に比べれば，精度よりもむしろ速報性に比重をおくことが必要といえましょう。

―参考―

統計式による推計には，時系列回帰分析の手法が考えら

れる。

時系列回帰分析とは，同時に変化する２つ以上の変量の時系列があり，その間に何らかの関係が考えられる場合，その一方を被説明変数（従属変数），他を説明変数（独立変数）とし，説明変数の変化による被説明変数の変化の関係を，最もよくあてはまる１つの式（回帰方程式）で表わそうとするものであり，また，その式から将来の予測値を推計しようとする統計分析手法である。回帰方程式の描く線（回帰線）のあてはめ方法として最も多く使われるのが最小２乗法で，あてはめる直線（または曲線）と実績値（観測値）との残差の２乗和が最小になるように直線（または曲線）の位置を定めるものである（最小２乗法については，６月号参照）。

説明変数の選択基準

(1) 早期に入手可能なデータであること
(2) 被説明変数に対し，一般的に相関が高いと思われること
(3) 時系列的に整備された資料であること

方程式の作成および選定

回帰方程式には，一元一次式（または二元一次式）を想定し，昭和35年度～昭和49年度（または昭和40年度～昭和49年度）の時系列データを用いて，最小２乗法により，各項目について数本の方程式を作成した。

例：

Ｙ＝ａ＋bx，Ｙ＝a＋bx＋cx

（昭和49年度までの時系列データによって，ａ・ｂ・ｃを計算し，50年度の説明変数の値をＸに代入することによって，50年度の各所得項目Ｙが推計される。）

ａ・ｂ・ｃは，回帰線の位置と方向をきめる回帰母数（パラメータ）である。

しかし，この方程式だけでは，推計に十分使用しうるかどうか判断できないので，各方程式ごとに次のような特性値を算出し，最良と思われる方程式を各項目につき１本選定し，推計に用いた。

(1) 推定値の標準誤差
(2) 重相関係数
(3) 回帰係数の標準誤差およびｔ一値
(4) ダーヴィン・ワトソン比
(5) 推計式による各年度の推定値と実績値との乖離率

推定値の標準誤差の値が小さく，重相関係数の値が大きいもの（１に近いもの）を基準にし，また，ｔ一値の小さいもの，ダーヴィン・ワトソン比が2から大きく乖離するものは，できるだけ選定からはずした。

また，回帰方程式は，直接最小２乗法と残差（実績値ー推定値）の系列相関による誤差を取り除くため，一般化最小２乗法による回帰方程式をあわせて作成した。

資料や内容の問い合せは

水戸市三の丸１－５－38　〒310
茨城県企画部統計課県勢統計係
ＴＥＬ0292－21－8111　内線426
までお寄せください。

このページに，どんな質問でもお寄せ下さい。担当の係からお答します。ハガキ，ＴＥＬ，なんでも結構です。

宛先
水戸市三の丸１－５－38
茨城県企画部統計課企画調整係
「統計いばらき担当」まで
ＴＥＬ　0292－21－8111
内線４２０