第３章　調査データの処理

１．数値データの取扱い

合計、平均について思い出して下さい。60 点、50 点、70 点、40 点、80 点の合計（Sum）は…

S＝60+50＋70＋40＋80＝300　　より、　S は 300 です。　平均は $\bar{x}$＝300÷5＝60 点

合計　$S = x_1 + x_2 + x_3 + \cdots\cdots + x_n = \sum_{i=1}^{n} x_i$　　　x_i　と書きます。

右の式は、i＝1 から n までの x_i の合計を表します。（シグマ）

平均の一般式は、合計（S）を N で割ります。その式は、下のようになります。

平均(mean)　$\bar{x} = \frac{1}{\mathrm{N}}(x_1 + x_2 + x_3 + \cdots\cdots + x_n) = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{n}} x_i$

ここで、平均　$\bar{x} = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{n}} x_i$　これは、今後よく使います。

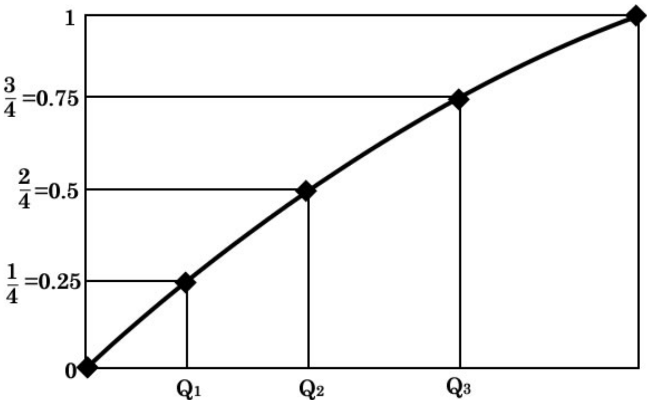
（１）四分偏差（Q　Quartile deviation）

データ（測定値）がどのように分布しているか知るために、累積度数曲線を作り、

その $\frac{1}{4}$ 、 $\frac{2}{4}$ 、 $\frac{3}{4}$ の値で示す方法がよく用いられます。データを大きさの順に並べたとき、

$\frac{1}{4}$ の値を Q_1 、 $\frac{2}{4}$ の値を Q_2 、 $\frac{3}{4}$ の値を Q_3 として表します。

Q_2は、大きさの順に並べたとき中央の値ですので、メディアン（Me）になります。
（500 個のデータがあれば、大きさの順に並べて 250 番目のデータの値です。）



Me : Mediam（メディアン）　Q_2
数値を大きさの順に並べたとき、中央になる値

Q_1 、Q_3 ：　四分位数（Quartile）
　Q_1は、大きい順に並べたとき、全体の 1/4 番目になる値
　Q_3は、大きい順に並べたとき、全体の 3/4 番目になる値

問題　全体で 500 個のデータがあります。

これを四分位数の　$\sim Q_1$、$Q_1 \sim Q_2$、$Q_2 \sim Q_3$、$Q_3 \sim$　には、それぞれおおよそ何個のデータが入るでしょうか。

$\sim Q_1$

$Q_1 \sim Q_2$

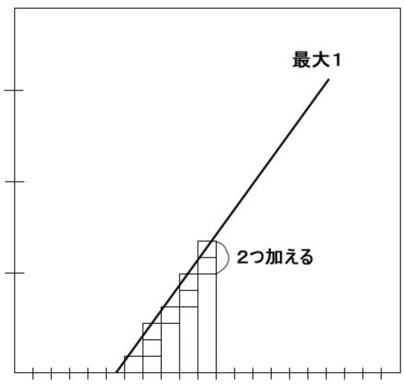
$Q_2 \sim Q_3$

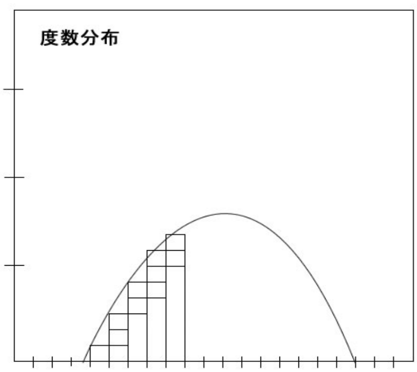
$Q_3 \sim$

問題　全体で 600 個のデータのうち、$Q_1 \sim Q_3$ には、おおよそ何個のデータが入るでしょうか。

問題　Q_2（Me）と平均とは、同じものですか。それとも違うものですか。
それぞれの特性を踏まえて説明して下さい。

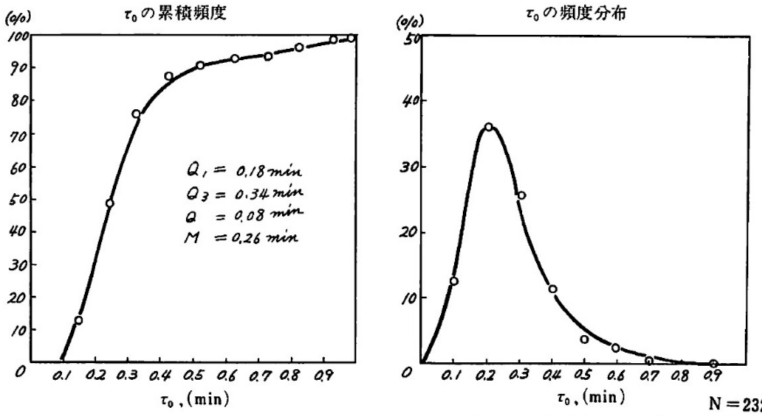
（注）





発問の反応時間を調べたところ、次のようになりました。蓄積度数分布は、何を表わすでしょうか。次のデータから考えて下さい。

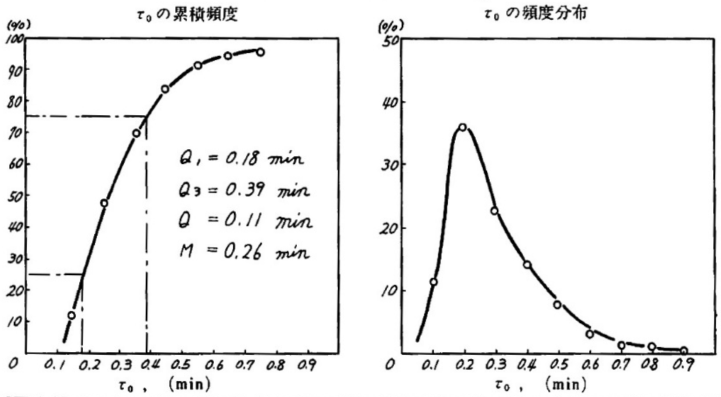
I－6　小学校・探査　　小　学　校・探　査　の　τ_0



N＝232

τ_0の区分 (min)	0～0.04	0.05～0.14	0.15～0.24	0.25～0.34	0.35～0.44	0.45～0.54	0.55～0.64	0.65～0.74	0.75～0.84	0.85～0.94	0.95 1.04	1.05～	計
出現頻度 (%)	0	12.5	36.2	26.7	11.6	3.4	2.6	1.3	3.0	0.9	0.9	0.8	100

I－2　高校・探査　　・　探査の　τ_0



τ_0の区分 (min)	0～0.04	0.05～0.14	0.15～0.24	0.25～0.34	0.35～0.44	0.45～0.54	0.55～0.64	0.65～0.74	0.75～0.84
出現頻度 (%)	0	11.4	35.9	22.3	13.9	7.3	2.6	1.8	1.8
τ_0の区分 (min)	0.85～0.94	0.95～1.04	1.05～1.14	1.15～1.24	1.25～1.34	1.35～1.44	1.44～		計
出現頻度 (%)	0.4	0.7	0.7	0	0.4	0.4	0.4		100

問題　Q_1、Q_2、Q_3を求めて下さい。（分=min）

小学校

Q_1	Q_2	Q_3
分	分	分

使い易いように秒にします

Q_1	Q_2	Q_3
10　秒	14　秒	20　秒

高等学校

Q_1	Q_2	Q_3
分	分	分

使い易いように秒にします

Q_1	Q_2	Q_3
秒	秒	秒

問題　発問後、学習者が最初に手を挙げた時間は、発問から8秒後でした。教師としては、どのように受け止めればよいでしょうか。

問題　発問後、25秒後で最初の学習者が手をあげました。教師はどのように受け止めればよいでしょうか。

問題　発問後に、学習者にかんがえさせるためにはどのような配慮が必要でしょうか。

問題　Q_1、Q_2、Q_3、を分布状態を示す参考として使ってみたい事例を考えて下さい。

問題　平均と Q_2（Me）の違いを説明して下さい。

問題　次の表は学習活動での正答率（Q_1、Q_2、Q_3 での正答率）です。ここから、どのようなことが言えますか。各自で考察して下さい。

四分位 / 正答	Q_1	Q_2	Q_3
発問の応答	５６％	７７％	９２％
問題の課題解決	４６％	６３％	８３％
グループ討論	５０％	６９％	８７％
全体討論	５３％	７３％	８７％
分節の通過率	６５％	８５％	９５％

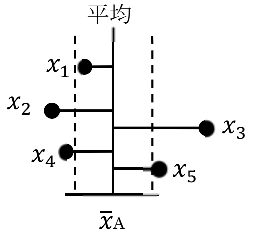
（２）平均偏差（Average deviation）

下の図で、(A)（B）の 2 つの分布について、平均値からどれほど離れているかを知るひとつの方法を考えてみましょう。

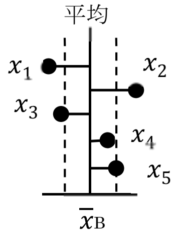
図でみれば、(A) の方が分布の広がりが大きいですね。

平均からどの程度離れているか知るため、各数値について平均からの差を求め、それぞれの差を合計します。そうすると、合計した値は “0” になってしまいます。なぜでしょう。各数値は図で示すように平均から+、－になります。

(A)



(B)



式でも考えてみて下さい。

$$\frac{1}{\mathrm{N}}\sum_{i=1}^{n}(x_i-\bar{x}) \quad = \quad \frac{1}{\mathrm{N}}\sum_{i=1}^{n}x_i \;-\; \frac{1}{\mathrm{N}} \times \mathrm{N}\bar{x} \;=\; \bar{x} \;-\; \bar{x}=0$$

これでは答えが 0 になるので困ります。

ひとつの方法として、下のように差の絶対値を取ってもよいのですが、絶対値の計算は困難です。

$$\frac{1}{\mathrm{N}}\sum_{i=1}^{n}|x_i-\bar{x}|$$

そこで数学でよく使う方法ですが、差を自乗

$(x_i－\bar{x})^2$ すれば、

$$\frac{1}{\mathrm{N}}\sum_{i=1}^{n}(\ x_i-\bar{x}\ )^2$$

全て正の数になり、自乗した値の合計が計算できます。

この式は、自乗の和ですので、点数と同じように取り扱うには、平方根にするとよいですね。

（数学でよく使いましたね。）

$$\sqrt{\frac{1}{\mathrm{N}}\sum_{i=1}^{n}(\ x_i-\bar{x}\ )^2}$$

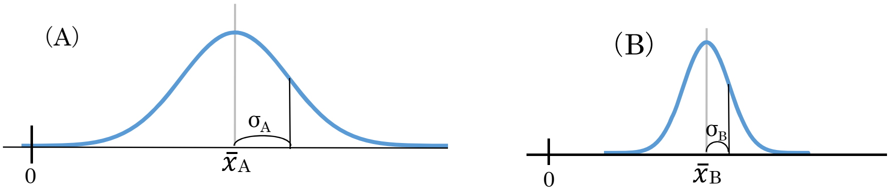
（３）標準偏差（Standard deviation）

このように求めた平均からの分布の状況の表し方を標準偏差と言います。

標準偏差　　$\sigma = \sqrt{\frac{1}{\mathrm{N}}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

この σ は、シグマです。σ は、平均からの分布の広がりの具合を示しています。

例えば上の（A）の σ_A と、（B）の σ_B の広がりの大小を示すとこのようになります。



（４）分散（variance）V

標準偏差の自乗σ^2は、分散 V　と言います。これは、σを求める前の式でしたね。

$$\mathrm{V} = \sigma^2 = \frac{1}{\mathrm{N}}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

逆に、この V の平方根が σ です。

$$\sigma = \sqrt{\frac{1}{\mathrm{N}}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

（５）z=値、T 得点

平均から、σの何倍離れているかで表します。

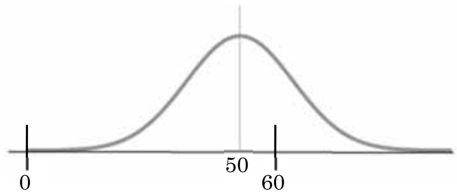
$$z = \frac{x_i - \bar{x}}{\sigma}$$

これを z 値と言います。

もし、　$x_i - \bar{x} = \sigma$

これだと、この値は 1 となりますが、点数に近い表現にするために便宜上 z を 10 倍にします。平均値を 50 に移行すると下のようになります。そうすると、

$\frac{(x_i - \bar{x})}{\sigma} \times 10 + 50$

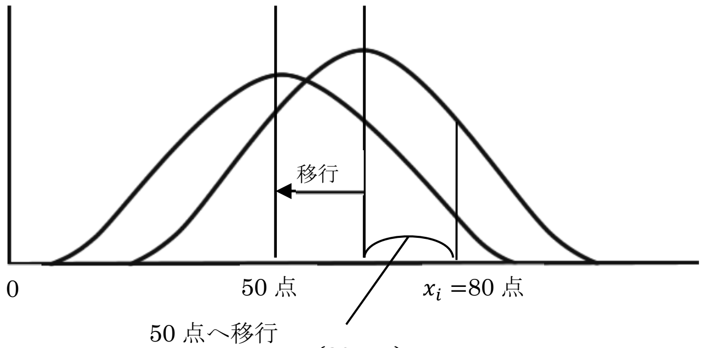


$\frac{x_i - \bar{x}}{\sigma} = 1$　のとき、$\mathrm{T} = 60$ 点

T : 得点　　＊T が正規分布（ガウス分布）と仮定

$$\mathrm{T} = \frac{10(x_i - \bar{x})}{\sigma} + 50$$

たとえば、$\sigma = 15$、平均$\bar{x} = 65$、x_iが 80 とすると　$\mathrm{T} = \frac{80 - 65}{15} \times 10 + 50 =$　となります。



$\frac{(80 - \bar{x})}{\sigma}$　$\times 10$　　σ の巾を 10 とします

これだけは覚えておいて下さい！

T － 得点

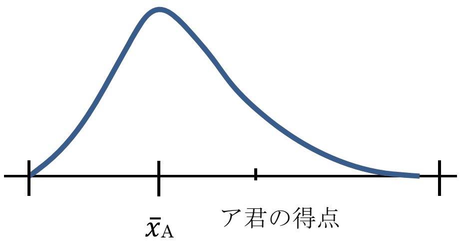
$$\mathrm{T} = \frac{x_i - \bar{x}}{\sigma} \times 10 + 50$$

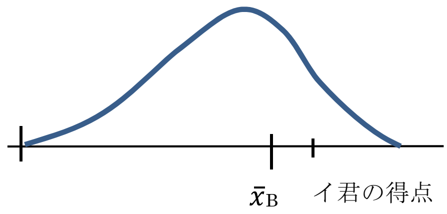
$$\mathrm{T} = \frac{（得点－平均値）}{標準偏差} \times 10 + 50$$

試験をして、科目等により平均や広がりに差があるときに使います。

（とくに、入試などで受験科目によって平均点に差が出たとき、どのようにしますか。選択科目で、生物の平均点が 70 点で科学の平均点が 40 点だとした時、どうすればよいでしょう。）

たとえば、テスト A とテスト B の平均、分布が次のようになったとき





平均：$\bar{x}_{\mathrm{A}}$ ＝40 点　$\sigma_{\mathrm{A}} = 12$　　　　平均：$\bar{x}_{\mathrm{B}}$ ＝70 点　$\sigma_{\mathrm{B}} = 18$

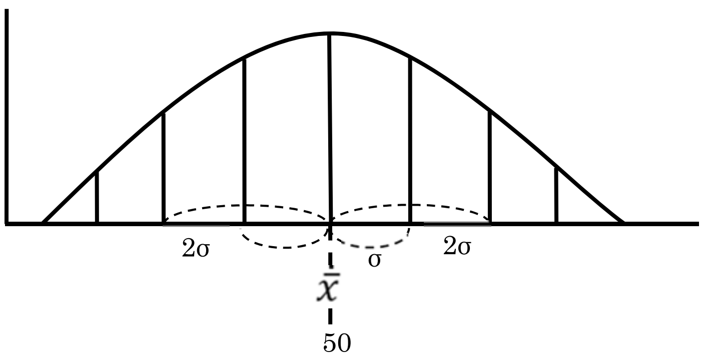
ア君の得点　$\bar{x}_{Ai}$ ＝55 点（素点）　　　　イ君の得点　$\bar{x}_{Bi}$ ＝75 点（素点）

もし、試験の選択科目としたとき、両者の点数を同じ素点で成績を比較してもよいでしょうか。

これを補正するのが、T 得点の処理です。

（注）T 得点の処理は、パソコン等で表計算を用いると簡単に処理することができますが、その意味を理解しておくことが大切です。

参考　T 得点の分布	
20 以下　0.1%	80 以上　0.1%
30 以下　2.3%	70 以上　2.3%
40 以下　16%	60 以上　16%
45 以下　31%	55 以上　31%
50 以下　50%	50 以上　50%



問題　T 得点は何に使われていますか。調べて下さい。

問題　社会科の地理、日本史、世界史の入学試験で平均が大きく違った時、どのような方法で比較しますか。

次の計算をして下さい。

	xi	$xi-\bar{x}$	$(xi-\bar{x})^2$	$\frac{xi-\bar{x}}{\sigma}$	T	
1	76					
2	50					
3	42					
4	37					
5	92					
6	66					
7	84					
8	56					
9	47					
10	68					
11	79					
12	49					
13	63					
14	83					
15	58					
16	22					
17	69					
18	41					
19	33					
20	91					
21	75					
22	53					
23	72					
24	81					
25	59					
26	67					
27	31					
28	64					
29	64					
30	61					
31	21					
32	71					
33	48					
34	45					
35	62					
36	35					
37	78					
38	51					
39	57					
40	85					
合計						

$\sum_{i=1}^{n} xi$

$T=\frac{xi-\bar{x}}{\sigma}\times 10+50$

x_iの蓄積度数分布と四分位数を求めて下さい。

x_iの分布と四分位数を求めて下さい

Q_1＝

Q_3＝

Me＝

Q＝

次の計算をして下さい。

	xi	$xi - \bar{x}$	$(xi - \bar{x})^2$	$\frac{xi - \bar{x}}{\sigma}$	T	
1	61					
2	35					
3	43					
4	46					
5	51					
6	32					
7	28					
8	49					
9	25					
10	30					
11	39					
12	21					
13	53					
14	48					
15	46					
16	56					
17	52					
18	44					
19	66					
20	20					
21	29					
22	37					
23	54					
24	42					
25	38					
26	37					
27	43					
28	29					
29	33					
30	37					
31	55					
32	23					
33	18					
34	35					
35	41					
36	47					
37	51					
38	46					
39	35					
40	29					
合計						

$\sum_{i=1}^{n} xi$

x_iの蓄積度数分布と四分位数を求めて下さい

x_iの分布と四分位数を求めて下さい。

Q_1＝

Q_3＝

Me＝

Q＝