

2年目の「データの分析」

有朋高校単位制課程 大谷 健介

1 度数分布表の復習「紙テープ 10 cm切り」

何森 仁 先生(神奈川大学)が、25 年前から講義で実践されている統計の学習です。生徒の興味・関心を引きつけて、なおかつヒストグラム・度数分布表が復習できて、代表値のことも理解が深まります。

この実践を私も夏休み明けに早速やってみました。ふだん、まじめにやっている生徒も、あまり前向きでない生徒も全員取り組んでしまうという優れものです。ぜひ、先生方も取り入れてみてください…紙テープが意外と安くなくて…2つで105円です。

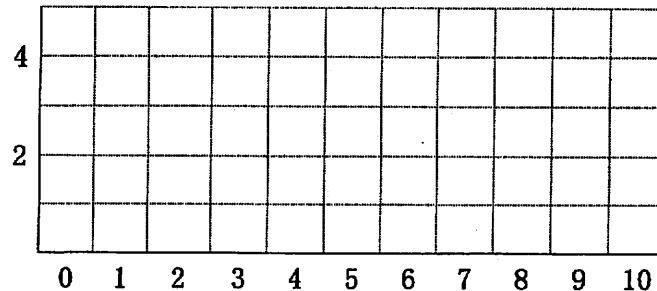
ちなみに、この実践は実教出版の高校数学 I (大判教科書)の課題学習 5 として、掲載されていることを最近知りました。参考までに資料につけておきます。

ところで昨年は、1つ1つのデータを正の字で数えて度数分布表をつくり、それをヒストグラムに反映させていましたが、紙テープ切りをしてみて、データを直接グラフに書き込んでかぞえ、あとから度数分布表を完成させる方が楽なことに気がつきました。中央値についても、わざわざ値の小さなデータから順に書くことは不要でした。私のつくったプリントの練習問題です。

問題 次のデータはある講座における生徒 20 人の小テストの点数である。次の問いに答えよ。

7, 6, 4, 7, 10, 5, 1, 9, 10, 8, 3, 4, 6, 8, 10, 9, 4, 6, 7, 6

(1)ヒストグラムを作成せよ。



(2)度数分布表を完成させよ。

点数(点)	0	1	2	3	4	5	6	7	8	9	10	合計
度数(人)												

(3)このデータの中央値、最頻値、平均値をそれぞれ求めよ。

2 教材の適切な精選と配列

つぎに箱ひげ図の話です。四分位数や箱ひげ図にはいろいろな問題点が見られることや、担当した講座が、受験等に関わりがないことから、今年度は箱ひげ図と四分位数をきわめて薄く取り扱い、分散や標準偏差を手厚く指導するという、「教材の精選」を行いました。

(1) 箱ひげ図の良さ

昨年度は、2つのデータの比較を考えて、こんな練習問題に取り組んでいました。

問題 次のデータは、A、Bの2つの地点における、日中の歩行者の交通量を10日間調べたものである。

A地点	52	62	80	65	40	70	77	58	49	55
B地点	62	75	90	77	51	80	88	69	57	65

(1) 2地点における5数要約の値を求めよ。

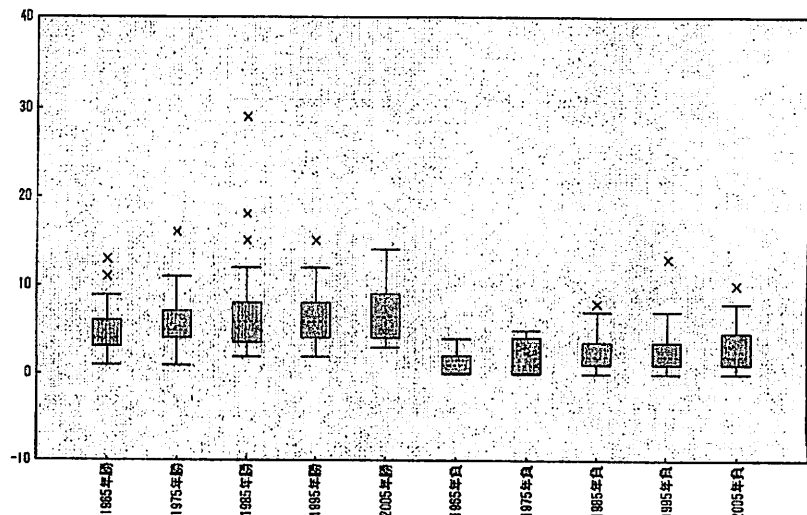
	最小値	第1四分位数	中央値	第3四分位数	最大値
A地点					
B地点					

(2) A地点、B地点のデータの箱ひげ図を書け。

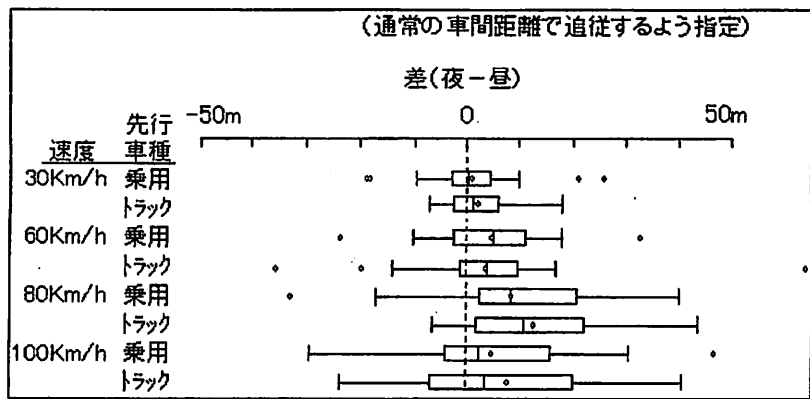
統計の目標のようなことがわかってくると、この学習内容にはあまり意味がないように感じてきました。箱ひげ図は「多次元的な表現ができる」ところにその良さがあると考えられますので、そういう意味ではたかだか2つのデータの比較に使う必要はなく、それならば、累積度数曲線やヒストグラムを使った方が優秀です。

下の表は筑波大学の先生が夏の甲子園のデータについて、2005年に記したサイトから引用してきたものです。この年の大会は打撃戦が多かった印象があったことから10年ごとの出場校の得点を箱ひげ図にして本当に打撃戦が多かったかどうかを調べたものです。度数分布表で示そうとすると、10回分のデータを横書きしてたくさん書かなくてははいませんが、箱ひげ図を利用することによって、簡素に経年変化を調べることができ、この図の良さを感じ

ることができます。なお、×印は「はずれ値」として処理されていますが、おそらく箱の大きさ(長さ)の1.5倍を超えた値(得点)を「はずれ値」として処理しており、いたる年でははずれ値が複数発生していることが見られますが、このはずれ値の扱いについても注意が必要です。



次の箱ひげ図は、自動車安全運転センターのサイトからの引用です。被験者に通常の車間距離で追従運転してもらったときの、昼夜の車間距離の差を、速度や車種を変えて実験した結果です。これも、度数分布表で示すとたてにたくさんのグラフを描くことになりますので、簡素に車種・速度の変化を見ることができます。



(2) 「教材の精選」の理由

① 四分位数には現在、世界的に10以上の定義があつて、これが標準的というものがない中で、学習指導要領ではそのうちの1つを取り上げて定義しているため、教科書の内容と(Excel をふくめた)種々の統計ソフトの考え方がマッチしていない現実がある…箱ひげ図はこの世に出てきてから10数年しか経っていない(三角比の歴史はいかほどか…)

* あるウェブサイトによると、

例 10個のデータ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 について

Excel や R では … 第1四分位数 3.25 中央値 5.5 第3四分位数 7.75

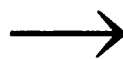
JMP や SPSS では … 第1四分位数 2.75 中央値 5.5 第3四分位数 8.25

と計算されるらしい

② 教科書の130ページに「四分位範囲が大きいほど…」とあり、133ページには「分散、標準偏差が大きいほど…」と女の子が言っている指導の難しさ

(統計の専門書や論文には、四分位数と分散を同列に扱い、ともにデータの散らばり具合を表す量としている例はないそうです)

四分位範囲が大きいほど、データの散らばり具合が大きいといえます。



分散、標準偏差の値が大きいほど、データの散らばり具合が大きいといえます。



③ 標準レベルの「数学 I」で使用している実教出版の例題から

例題 下の表は、ある高校のA組とB組の男子9人について、ボール投げの記録(m)を大きさの順に並べたものである。

生徒番号	①	②	③	④	⑤	⑥	⑦	⑧	⑨
A組	26	27	29	31	31	33	34	35	35
B組	23	25	26	27	29	31	31	32	33

四分位範囲は A組 $34.5 - 28 = 6.5$ B組 $31.5 - 25.5 = 6$

よって、四分位範囲からデータの散らばりの度合は、B組よりA組の方が大きい

↓

しかし、分散 A組 計算省略=9.95 標準偏差 A組 3.15

B組 =10.69 B組 3.27

よって、分散、標準偏差からデータの散らばりの度合は、A組よりB組の方が大きい

(3) 統計の何を学ぶか

2年目の指導を終えて、まだまだこの分野は勉強不足なのですが、少しわかってきたことを書きます。

- ① 箱ひげ図の「見方」を知ることは必要だと思うが、「書き方」がわかることはたいした問題ではない。
- ② 基本となる代表値はその意味も求めることもともに大切であろう。
- ③ 少ないデータでも「分散」や「標準偏差」の計算方法やそのおおざっぱな意味を知ることは統計を知る上では重要であろう。
→ 「標準偏差を手計算で求める」ことをモチベーションにして、開平算をやるのもおもしろいかも…
- ④ 本来的には非常にたくさんのデータをコンピュータソフトで取り扱って、そのデータの特徴を分析したり、相関を調べたりできることが統計を学ぶ意義であろう。→「数学」ではなく「情報」に近いもの…
- ⑤ データを使って本当にその本質に迫っていくためには、多角的な多くのデータの関連によって見えてくるものであり、本来「ホームランの多いチームが強い」程度の話では無いはずである。
- ⑥ 割り切って「受験数学」として指導する！

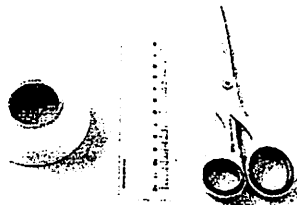
3 雑感

と、統計の学習について少し問題点を洗ってみるレポートになりました。本課程では、「数学 I 基礎X」という講座の中で、“方程式と不等式”と、この“データの分析”の2章だけ扱います(他の章は「I 基礎Y」の講座で…)。中学校まで数学が極端に苦手だった生徒は、数式を嫌がります。分数はもっと嫌がります。考える前から思考障害が働いてどんな簡単な質問をしても「わかりません」となります。ところが、この単元はそういった数学から少し離れるため、「これはできるかも…」という生徒が増えます。学習意欲も上がってきます。そういう視点で捉えると、実はこの単元は一役買っているような感覚を持っていることも事実です。生徒のそのような意識と指導者側の「何を教えたいか」がもっとマッチしてくると良いのですが…。

課題学習 5 紙テープを切ってデータを分析してみよう

紙テープをくり返し切り取り、切り取った長さを測ってみよう。5章で学んだ内容を利用して、テープの長さのデータを分析してみよう。

長さが10cmの紙テープを机の上において、それを見ながら、1巻の紙テープから目測で10cmの長さになるように切り取る作業をする。
得られたデータを、整理してみよう。



手順を確認しよう

- ① 定規を用いて、10cmの紙テープを1つ切り取る
- ② 1の紙テープを机の上において、定規を用いずに、目測で同じ長さになるように紙テープを切り取る作業をくり返す
- ③ ②で切り取った紙テープの長さをすべて定規で測り、そのデータを記録する

実は、紙テープを切り取る作業を50回くり返したところ、次のようなデータを得た。

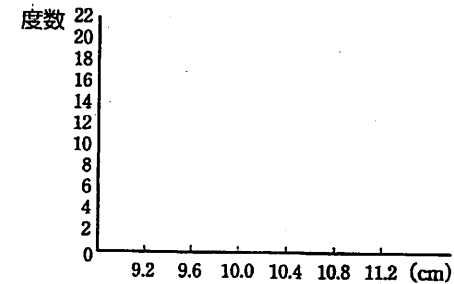
9.3	10.5	10.0	9.6	10.2	9.7	9.9	9.8	(cm)
10.1	9.9	10.7	9.9	10.0	9.9	9.7	10.8	
10.0	10.2	10.9	10.0	9.9	10.4	9.6	10.0	
11.0	10.1	10.1	10.4	10.2	10.1	10.2	9.8	
9.7	10.0	10.2	10.5	10.3	10.0	10.4	10.4	
10.4	10.1	10.4	10.2	9.5	10.7	9.9	9.9	
10.0	10.1							

計算してみよう

① 実さんの実験で得られたデータを、右の度数分布表に整理しなさい。

階級 (cm)	度数	正の字
9.2 ^{以上} ~ 9.6 ^{未満}		
9.6 ~ 10.0		
10.0 ~ 10.4		
10.4 ~ 10.8		
10.8 ~ 11.2		
計		

② 上の度数分布表を、ヒストグラムで表しなさい。



③ 実さんの実験で得られたデータについて、平均値と標準偏差を求めなさい。

課題

長さが10cmの紙テープを机の上において、それを見ながら、1巻の紙テープから目測で10cmの長さになるように切り取る作業を100回行い、得られたデータを度数分布表に整理して、ヒストグラムで表しなさい。

また、得られたデータについて、平均値と標準偏差を求めなさい。

考えてみよう

実験結果からわかったことをまとめ、クラスで話し合ってみよう。

班 学番 _____ 氏名 _____

テープの色は



これが10cmです



じっくり、しっかり見て、10cmの長さを頭の中に！！

10cm
測る

5cm

階級

cm

5

10

15

20

25

30

35

40

45

計

1) 5.25 ~ 5.75

2) 5.75 ~ 6.25

3) 6.25 ~ 6.75

4) 6.75 ~ 7.25

5) 7.25 ~ 7.75

6) 7.75 ~ 8.25

7) 8.25 ~ 8.75

8) 8.75 ~ 9.25

9) 9.25 ~ 9.75

10cm

10) 9.75 ~ 10.25

11) 10.25 ~ 10.75

12) 10.75 ~ 11.25

13) 11.25 ~ 11.75

14) 11.75 ~ 12.25

15) 12.25 ~ 12.75

16) 12.75 ~ 13.25

17) 13.25 ~ 13.75

18) 13.75 ~ 14.25

19) 14.25 ~ 14.75

15cm

20) 14.75 ~ 15.25

21) 15.25 ~ 15.75

22) 15.75 ~ 16.25

23) 16.25 ~ 16.75

24) 16.75 ~ 17.25

25) 17.25 ~ 17.75

26) 17.75 ~ 18.25

27) 18.25 ~ 18.75

28) 18.75 ~ 19.25

29) 19.25 ~ 19.75

20cm

30) 19.75 ~ 20.25

31) 20.25 ~ 20.75

32) 20.75 ~ 21.25

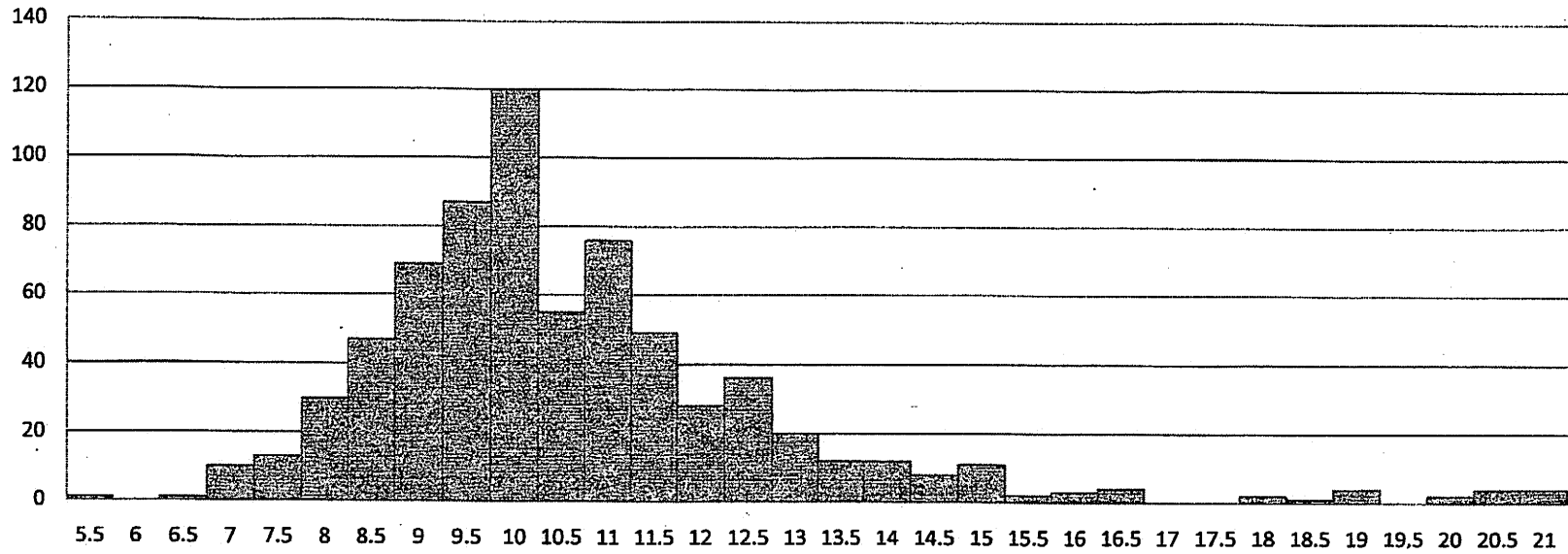
33) 21.25 ~ 21.75

34) 21.75 ~ 22.25

「あなたの10cm切りの分析」集計

No.	階級(cm)		階級値 (cm)	1F	2E	1F	1C	1C	1E	1D	1E	1C	1E	2G	1G	1E	1G	2G	1F	3D	1F	3B	3C	2D	2B	度数 (回)	累積 度数	階級値×度数	偏差	(偏差)×度数	
	以上	未満		K	H	N	M	K	M	C	M	G	M	M	T	H	A	T	S	O	K	S	H	W	Y						
1)	5.25	~	5.75																	1						1	1	5.5	-5.16	26.7	
2)	5.75	~	6.25																							1	1	0		0	
3)	6.25	~	6.75																		1						1	2	6.5	-4.16	17.3
4)	6.75	~	7.25												1					4	5					10	12	70	-3.66	134	
5)	7.25	~	7.75										1						2	9		1			13	25	97.5	-3.16	130		
6)	7.75	~	8.25										4	5	2		1		3	6	8	1			30	55	240	-2.66	213		
7)	8.25	~	8.75						3		3		17	7	1	1			3	6	5	1			47	102	399.5	-2.16	220		
8)	8.75	~	9.25						8		8	2	4		1	8	9	3		6	1	7	11	1	69	171	621	-1.66	191		
9)	9.25	~	9.75			5			18	2	18	5	1		5	6	9	1	2	2	1	5	6		87	258	826.5	-1.16	118		
10)	9.75	~	10.25	2		8	5		11	2	11	12	1	2	8	8	12	2	8	3	1	4	6	7	7	120	378	1200	-0.66	52.9	
11)	10.25	~	10.75	2		5	2		2		2	5	2	1	3	2	1	2	5			1	4	7	9	55	433	577.5	-0.16	1.48	
12)	10.75	~	11.25	11	2	6	9		2	4	2	4		5	1			9	5	1			1	4	10	76	509	836	0.34	8.59	
13)	11.25	~	11.75	3	5	3	7		5		2		4					4	7				1	5	3	49	558	563.5	0.84	34.3	
14)	11.75	~	12.25	7	2	2	6		2				1		2			2		1				3	28	586	336	1.34	50		
15)	12.25	~	12.75	3	7	1		2		8			5					5	3				1		36	622	450	1.84	121		
16)	12.75	~	13.25	2	3		1	2		4			6					1						1	20	642	260	2.34	109		
17)	13.25	~	13.75		5			2					5												12	654	162	2.84	96.5		
18)	13.75	~	14.25		3			1	2				6												12	666	168	3.34	134		
19)	14.25	~	14.75		4				1				3												8	674	116	3.84	118		
20)	14.75	~	15.25		2			2		3			4												11	685	165	4.34	207		
21)	15.25	~	15.75						1				1												2	687	31	4.84	46.8		
22)	15.75	~	16.25						1				2												3	690	48	5.34	85.4		
23)	16.25	~	16.75					2		1			1												4	694	66	5.84	136		
24)	16.75	~	17.25																							694	0	6.34	0		
25)	17.25	~	17.75																							694	0	6.84	0		
26)	17.75	~	18.25					1		1															2	696	36	7.34	108		
27)	18.25	~	18.75						1																1	697	18.5	7.84	61.4		
28)	18.75	~	19.25					4																	4	701	76	8.34	278		
29)	19.25	~	19.75																							701	0	8.84	0		
30)	19.75	~	20.25					2																	2	703	40	9.34	174		
31)	20.25	~	20.75					4																	4	707	82	9.84	387		
32)	20.75	~	21.25					4																	4	711	84	10.34	427		
合計				30	33	30	30	27	44	37	44	30	30	46	30	30	32	30	30	26	30	30	33	28	31	711		7582		3687	

10cm切り集計



中央値 = 10 cm

平均値 = 10.66 cm

最頻値 = 10 cm

分散 = 5.18517727

標準偏差 = 2.27709843