第7回　　先生のメモ　　15.4362196

今日やること
1. 数行番の第7版
2. 宿題

学生番号　　1234567 ←ここに学生番号を入力する。
氏名　　←ここに氏名を入力する。

データがそろったらさて、次の指標を計算できるようになってください。

平均
中央値
25分位点（しょっぱいてん）
分散
標準偏差
範囲
四分位偏差

度数分布表ができたらさて、次の指標を計算できるようになってください。

平均
分散
標準偏差

では やっていきましょう。「→」でボが書かれているところは、コピペをせずに、自分で入力してください。
↓のような10個のデータで開始です。

40 平均　　平均とは：小学校で習ったとおりに計算しましょう。
30　　30 ←=(10+20+30+40+50)/5
50　　↑の方法は、データが変わっても結果が変わらないので、表計算の利点が活かされていませんね。
10
20　　平均その2：セルの範囲を使いましょう。
　　30 ←=A15+A16+A17+A18+A19/5
　　↑の方法は、データがたくさんあると大変です。数式を入力するのも、数を数えるのも。

　　平均その3：合計を求める関数SUMと、数を数える関数COUNTを使いましょう。
　　30 ←=SUM(A15:A19)/COUNT(A15:A19)
　　↑の方法は、見てないのですが、Excelについて知らなすぎます。

　　平均その4：平均を求める関数AVERAGEを使いましょう
　　30 ←=AVERAGE(A15:A19)
　　平均の話はこれで終わり。

10 中央値　　並べ替えて真ん中にあるデータです。
20　　中央値その1：並べ替えて真ん中にあるデータを確認しましょう。
30
40　　中央値その2：中央値を求める関数MEDIANを使いましょう。
50　　30 ←=MEDIAN(A15:A19)
　　中央値の説明はこれで終わり。

40 四分位点（しぶん　並べ替えて。
30　　最初に来るデータが第1編。
50　　1/4=25%のところが第1四分位点。
10　　2/4=1/2=50%のところが第2四分位点（中央値）。
20　　3/4=75%のところが第3四分位点。
　　最後に来るのが最大編
　　10 ←=MIN(A15:A19)　　最小編その1
　　10 ←=QUARTILE.INC(A15:A19,A14)　　最小編その2
　　20 ←=QUARTILE.INC(A15:A19,A13)　　第1四分位点
　　30 ←=QUARTILE.INC(A15:A19,A12)　　第2四分位点その1
　　30 ←=MEDIAN(A15:A19)　　第2四分位点その2（中央値）
　　40 ←=QUARTILE.INC(A15:A19,A11)　　第3四分位点
　　50 ←=QUARTILE.INC(A15:A19,A10)　　最大編その1
　　50 ←=MAX(A15:A19)　　最大編その2
　　↑すべて以下便です。コピペして入力する場合に便利。
　　四分位点の説明はこれで終わり。

分散　　標準偏差
さて、今まで、たくさんのデータについて平均を出ようとするときに「平均」だけでいいのは小学生までです。
中学生以上は、ばらつき（散らばり）について、何か言えなければなりません。
ながで、次の2人うちこつの会社を比較することしましょう。数値は社員の年収だと思ってください。
会社A　　会社B
640　　300
740　　300
840　　300
940　　300
1040　　3000
840　　840 ←それぞれの平均を計算してください。（A7:F8でコピペ）
平均は同じです。
つまり、平均を出べるだけでは問題の違いはわかりません。

そこで、データがどの程度ばらついて（散らばって）いるかを調べます。
まずは、「平均からのずれ」を調べます。これを偏差といいます。

会社A　　偏差
640　　-200 ←=B2-B8（=A64-A$88）を求めます。
740　　-100 ←=B3-B8をコピペ
840　　0 ←=B4-B8をコピペ
940　　100 ←=B5-B8をコピペ
1040　　200 ←=B6-B8をコピペ
840　　0 ←=B7-B8を求めます。
↑三平均を求めます。

これでわかったことは、偏差の平均は0ということです。
このことはどうしても事実で、あたりまえと思えるようにしてください。
たとえば、試験の平均点が70点だったとしましょう。
全員の点数を、10減らします。
平均点は何点になりますか？　と聞かれたら、60と即答できなければなりません。

しかしどうやるか、会社Aを会社Bも、偏差の平均を比べても意味がないはずです（同次元なので）。
偏差が正になるのと正になるので、打ち消しあってしまうからです。
そこで、偏差の2乗を調べます。
2乗すれば、すべて正になるので、打ち消しあわなくなるわけです。

会社A　　偏差　　偏差の2乗
640　　-200　　40000 ←=B2-B8^2（=A10^2）を求めます。
740　　-100　　10000 ←=C2-C$16をコピペ
840　　0　　0 ←=C2-C$16をコピペ
940　　100　　10000 ←=C2-C$16をコピペ
1040　　200　　40000 ←=C2-C$16をコピペ
840　　0　　20000 ←=C7-C$16を求めます。
↑三平均を↑の偏差とその 141.4214 ←=平均の平方根（=SQRT(C10)）を求めます。

上の三の偏差を分散といいます。
上の三の偏差を標準偏差といいます。
分散と標準偏差はいずれも、データがどの程度ばらついているかの指標です。
今、データの単位は「円」。分散は「円」の2乗の平均なので、その単位は「円の2乗」です。
標準偏差は分散の平方根なので、その単位は「円」です。
ですから、感覚的には、標準偏差のほう使いやすいです。
会社Aの社員の年収のばらつきは、約141万円ようだけです。

会社Aの社員の年収の分散と標準偏差を求めてください。

会社A　　偏差　　偏差の2乗
300　　-540　　291600
300　　-540　　291600
300　　-540　　291600
300　　-540　　291600
2000　　2160　　4665600
840　　0　　1166400
　　10倍

以上をまとめると、会社Aと会社Bの社員の年収は、平均だけ見れば同じですが、ばらつきは会社Bのほうがかなり大きいということになります。

さて、これまで見てきたような、分散の定義（偏差の2乗の平均）どおりの計算は大事なのですが、いつもこんなことはやっていられません。
分散と標準偏差は、関数一発で求められます。

会社A　　会社B
640　　300
740　　300
840　　300
940　　300
1040　　3000
840　　840 ←それぞれの平均を計算してください。（A14:F14にコピペ）
20000　　1166400 ←それぞれの分散を（=VAR.P関数）で計算してください。（A14:F14にコピペ）
141.4214　　1080 ←それぞれの標準偏差を（=STDEV.P関数）で計算してください。（A14:F14にコピペ）
　　分散と標準偏差が、前に定義どおりに計算した結果と同じであることを確認してください。

範囲
四分位偏差
データの散らばりの指標は分散と標準偏差が決定版ですが、他の指標も紹介しておきましょう。
範囲：「最大値-最小値」です。
四分位偏差：「第3四分位点の値-第1四分位点の値」です。

会社A　　会社B
640　　300
740　　300
840　　300
940　　300
1040　　3000
640　　300 ←それぞれの最小値をMINで求めてください。
1040　　3000 ←それぞれの最大値をMAXで求めてください。
400　　2700 ←それぞれの範囲を求めてください。
740　　300 ←それぞれの第1四分位点の値を求めてください。
940　　300 ←それぞれの第3四分位点の値を求めてください。
200　　0 ←それぞれの四分位偏差を求めてください。

ここからちょっと、度数分布表からデータの平均、分散、標準偏差を求める方法を学びます。
まずは、練習用の度数分布表を作りましょう。
↓は問題練習（10人分）です。
0　　5　　15　　20　　35　　51　　65　　77　　80　　90
43.8 ←平均
994.56 ←分散
31.52681 ←標準偏差

度数分布表を完成させてください。

下限値	階級値	上限値	①度数	②階級値×度数	③階級値−①で求めた平均	④階級値の2乗	⑤階級値の2乗×度数
0	5	10	2	10	-38.8	1474.56	2949.12
11	15.5	20	2	31	-27.8	778.41	1556.82
21	25.5	30	0	0	-17.8	320.41	0
31	35.5	40	1	35.5	-7.8	62.41	62.41
41	45.5	50	0	0	2.1	4.41	0
51	55.5	60	1	55.5	12.1	146.41	146.41
61	65.5	70	1	65.5	22.1	488.41	488.41
71	75.5	80	2	151	32.1	1030.41	2060.82
81	85.5	90	1	85.5	42.1	1772.41	1772.41
91	95.5	100	0	0	52.1	2714.41	0
			10	434			9026.4
			↑合計オ↑合計ア				↑合計ウ

平均：43.4　↑合計オ↑人数ケ平均
ここでいったん解説を頑張。

分散：903.64　↑合計ウ↑人数ケ分散
標準偏差：30.06061　↑分散の平方根＝標準偏差

⑤の平均が得られるのはなぜでしょうか。
平均を求めるためには、合計が必要です。
試験の結果の合計は、0+5+・・・・ですが、度数分布表には点数は載っていないので、この計算はできません。
しかし、0は2回で対2人いるとことはわかりの。とりあえずその2人は5点（階級値）ということにして、2*5+・・・とするのです。
そんないかげんなことを、していいのか？思うかもしれません。
たしかに、試験の結果ってたまやるのはちょっと問題ですが、たとえば160, 160, 170点の平均身長を求めようとする場合、
本当は160,1160, 160,8170, 170,2770と行かもしれないとしても、160+160+170=160*2+170と計算するのがずっと。
このように、おおよその値を使って計算するというのはよくあることなのです。
もちろんおおよその値で計算するので、もとの値で計算した結果と同じにはなりません。

次の計算は、83*0+87*1+88*0+91*1=・・・というように、2行の要素を掛算に掛けてから足しています。
こういう計算はよくあるので、専用の関数が用意されています。
434 ←=SUMPRODUCT(B2:B6,D2:D6)
これはいわゆる「ベクトルの内積」と同じ計算です。
この計算はこれもよく使うので、この関数は覚えておくといいです。

総合問題（できあがりの結果は同じです。見直を使っているので、同じにはなりません。）
試験結果1003人分を乱数で作ります。
30　　37　　46　　45　　62　　80　　72　　11　　64　　63　　74　　43　　88　　35　　49　　58　　49　　60　　41　　60　　73　　36　　51　　91　　26　　74　　34　　59　　44　　41　　74　　74　　58　　52　　41　　71　　52　　50　　72　　55　　35　　55　　71　　57　　44　　40　　75　　40　　66　　49　　56　　64　　63　　63
乱数を使うのは試験が固定されるから。
操作を固定するために、2次行目を選択してください。
↓は最初の値を代行してください。
29　　61　　66　　27　　25　　36　　68　　32　　46　　36　　17　　49　　35　　30　　86　　63　　48　　44　　80　　47　　52　　39　　39　　39　　47　　55　　35　　45　　28　　67　　72　　49　　36　　59　　58　　39　　39　　59　　47　　36　　58　　41　　12　　68　　72　　26　　26　　41　　26　　62　　54　　59　　58　　39

まずは、ふつうに平均、分散、標準偏差を求める。ヒント：データの範囲はA215:AL1229です。
48.996 ←平均
229.822 ←分散
15.36021 ←標準偏差

次に、度数分布表を作り、度数分布表だけから平均、分散、標準偏差を求めます。（先の作業の反復練習まとめのようです。）

下限値	階級値	上限値	度数	階級値×度数	（階級値−平均）^2×度数
0	5	10	10	50	19382.00625
11	15.5	20	19	294.5	21354.58688
21	25.5	30	74	1887	40953.49625
31	35.5	40	190	6745	34755.86875
41	45.5	50	241	10965.5	2994.075625
51	55.5	60	243	13486.5	10187.92688
61	65.5	70	147	9628.5	38859.56688
71	75.5	80	58	4379	40613.68625
81	85.5	90	13	1111.5	17791.53213
91	95.5	100	5	477.5	10769.62813
			1000	49025	238276.875
			↑合計	↑合計	↑合計
					49.025
					238.276875
				↑合計/人数＝平均	↑合計/人数＝分散
					15.43621968
					↑分散の平方根＝標準偏差