

情報処理演習Ⅱ 2019
(エクセルを使った統計学の演習)

【松本治彦】

2019年3月
宇部フロンティア大学出版会

著者 松本治彦

発行者 宇部フロンティア大学出版会

発効日 2019年3月30日

情報処理演習 II 2019

情報処理演習 II 2019 日程表

- 4月5日のテーマ；平均値、標準偏差、偏差値
- 4月12日のテーマ；度数分布表とヒストグラムの作成その1
- 4月19日のテーマ；度数分布表とヒストグラムの作成その2
- 4月26日のテーマ；階級値の取り方とヒストグラム
- 5月10日のテーマ；今までの復習、質問タイム
- 5月17日のテーマ；パーセンタイル値、箱ヒゲ図その1
- 5月24日のテーマ；パーセンタイル値、箱ヒゲ図その2
- 5月31日のテーマ；相関係数と回帰直線
- 6月7日のテーマ；t分布、区間推定、検定
- 6月14日のテーマ；今までの復習、質問タイム
- 6月21日のテーマ；ピポットテーブル
- 6月28日のテーマ；家計簿1
- 7月5日のテーマ；家計簿2
- 7月12日のテーマ；今までの復習、質問タイム
- 7月19日のテーマ；まとめ（USBの提出）

なお、進捗状況により、テーマを変更する場合があります。
授業の時間は、金曜日の2コマ目です。

なお、授業は10:50より開始します。早めに来て、前回の復習などを行ってください。

成績は、日ごろの授業態度と演習のまとめで提出してもらう成果で評価します。

1. 平均値、標準偏差、偏差値

(データをどのようにみたらよいのか?)

高校で中間試験、期末試験、実力試験、塾の試験、その成績が気になっていたと思います。成績表をみると、個々の成績と全体の平均、さらに偏差値との関係があまり明確には分からないことが多いと思います。そこで、この機会に平均値、標準偏差、偏差値の関係を理解してください。

以下に平均値、標準偏差、分散、偏差値の一般式を表してみました。

平均値とは、データの総和をデータの個数で割った値です。式で示すと、N 個のデータ $x_1, x_2, x_3, \dots, x_N$ の平均値 \bar{x} は、

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

ここで、 Σ はギリシャ文字“シグマ”であり、和を意味しています。すなわち、 $i = 1$ から N までの x_i の和を意味しています。

平均値は、加算と1回の割り算を含むだけですから、統計の尺度としては便利な値です。しかし、細かい情報は、全部捨ててしまっているのです。細かい情報を知るには、平均値の他に、度数分布とか、分散、標準偏差などを用いることが必要です。

次の式が分散 s^2 の一般式です。データのバラツキの度合いをみる尺度の1つです。

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

この式をみると、それぞれのデータから平均値を引いてその二乗した値の総和したものをデータの個数で割った値を意味しています。このようなめんどろな計算をしなくても、それぞれのデータから平均値を引いた値をたして、それをデータの個数で割ればよさそうに思いませんか。しかし、それではマイナスの値とプラスの値が生じて、合計した計算結果がゼロになることもあるのです。それでは、バラツキの度合いをみることができません。

そこで、それぞれのデータについて平均値からの差を求めて、それを二乗することで正の値にして加算する方法をとっているのです。

しかし、分散は元のデータを二乗しているので(点)²となり、もとのデータと単位が合わないのです。バラツキの尺度としては、その平方根をとり、もとのデータと単位をそろえた方が使いやすいことがわかると思います。次式が分散の平方根、すなわち標準偏差なのです。

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

さて、皆さんは偏差値という言葉をよく耳にしますが、偏差値とは、ある集団における、ある個体のある種の値である x_i を次のような式で標準化した値 y_i なのです。

$$\frac{y_i - 50}{10} = \frac{x_i - \bar{x}}{s} \quad \text{すなわち、} \quad y_i = \frac{10}{s}(x_i - \bar{x}) + 50$$

なのです。この式をみると、偏差値 y_i はもとの値 x_i を平均が 50、標準偏差が 10 になるように標準化した値なのです。

以上が統計学のお話です。これをエクセルで計算するのは意外と簡単にできます。平均値を求めるには、【挿入】→【関数】→【関数の分類】→【統計】→【AVERAGE】の順に選択をします。次に数値の範囲を選んで【OK】ボタンをクリックすれば結果が表れます。分散や標準偏差についても、同様に【挿入】→【関数】→【関数の分類】→【統計】→【VAR.S】あるいは【STDEV.S】の順に選択をします。次に数値の範囲を選んで【OK】ボタンをクリックすれば結果が表れます。

さて、偏差値を求めるには上記の一般式のような計算式をエクセル上で作成すればよいのです。平均値、標準偏差をあらかじめ求めておく。次に【=10/（標準偏差）×（ x_i - 平均値）+50】を作れば完成です。

いくつかの例を作って偏差値の正しい見方について考えてみてください。

成績表(得点)

	国語	数学	理科	社会	英語	合計
山田修司	50	35	65	75	55	280
永田和彦	35	65	35	78	89	302
西村雄二	85	95	90	80	91	441
村田力	98	100	75	83	90	446
中村裕子	40	50	50	77	70	287
徳永政治	45	45	45	79	45	259
安部幸三	72	72	70	81	60	355
庄司早苗	60	70	60	85	75	350
高田勇太	65	55	80	87	80	367
平均値	61.1	65.2	63.3	80.6	72.8	343.0
分散(VAR,P)	397.4	424.8	277.8	13.4	245.3	4091.6
標準偏差(STDEV.P)	19.9	20.6	16.7	3.7	15.7	64.0
分散(VAR,S)	447.1	477.9	312.5	15.0	275.9	4603.0
標準偏差(STDEV.S)	21.1	21.9	17.7	3.9	16.6	67.8

成績表(偏差値)

	国語	数学	理科	社会	英語	合計
山田修司	44.4	35.3	51.0	34.8	38.6	40.2
永田和彦	36.9	49.9	33.0	43.0	60.4	43.6
西村雄二	62.0	64.4	66.0	48.5	61.6	65.3
村田力	68.5	66.9	57.0	56.7	61.0	66.1
中村裕子	39.4	42.6	42.0	40.3	48.2	41.2
徳永政治	41.9	40.2	39.0	45.7	32.3	36.9
安部幸三	55.5	53.3	54.0	51.2	41.8	51.9
庄司早苗	49.4	52.3	48.0	62.2	51.4	51.1
高田勇太	52.0	45.0	60.0	67.6	54.6	53.8

2. 度数分布とヒストグラムの作成

表3 身長データ(単位:cm)

155.5 157.5 160.3 172.3 181.6 158.6 175.3 160.5 167.3 170.2
 163.0 163.3 162.8 161.9 161.9 158.8 181.5 171.4 167.3 168.9
 166.3 165.5 164.5 165.2 168.5 157.8 178.9 170.2 163.8 168.3
 167.2 167.8 169.8 168.4 161.8 159.8 180.3 166.5 173.2 177.2
 172.3 169.5 170.3 171.6 161.4 160.3 178.5 172.5 166.9 172.9
 170.8 172.0 165.5 172.5 168.2 165.4 175.2 172.8 166.7 170.3
 173.1 172.6 171.7 176.3 169.9 168.7 167.5 172.4 167.8 169.2
 175.4 166.5 166.8 174.3 167.8 166.6 166.8 173.1 164.8 168.2
 176.6 173.2 174.2 163.9 170.8 168.9 169.4 173.8 166.6 164.3
 162.8 176.5 176.3 173.8 173.5 170.3 169.5 174.5 172.9 170.6
 175.2 164.2 178.6 175.2 174.3 173.5 168.6 167.2 164.2 170.3
 177.8

度数を求めるには、次のカウントイフ文を使う。

COUNTIF(\$B\$3:\$K\$14,"<161.5")-COUNTIF(\$B\$3:\$K\$14,"<158.5")

ただし最初は、COUNTIF(\$B\$3:\$K\$14,"<158.5")で

表 4 身長の度数分布表

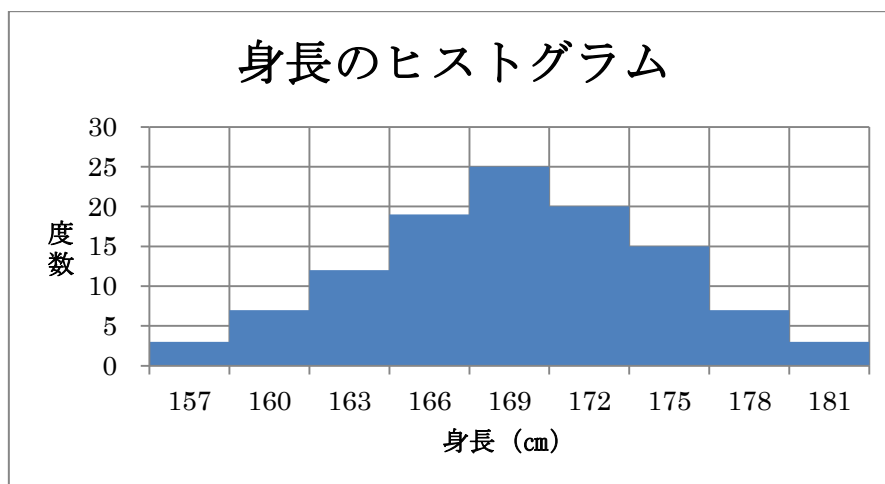
身長階級	度数	相対度数	累積度数	累積相対度数	階級値
155.5～158.5	3	0.027	3	0.027	157
158.5～161.5	7	0.063	10	0.090	160
161.5～164.5	12	0.108	22	0.198	163
164.5～167.5	19	0.171	41	0.369	166
167.5～170.5	25	0.225	66	0.595	169
170.5～173.5	20	0.180	86	0.775	172
173.5～176.5	15	0.135	101	0.910	175
176.5～179.5	7	0.063	108	0.973	178
179.5～182.5	3	0.027	111	1.000	181
計	111	1.000	111	1.000	

ヒストグラムの作成ポイント；

度数の列を選択→挿入→棒グラフ

棒の間隔をゼロにするには、棒グラフを作成した後に、グラフの真ん中をクリック→データ系列の書式設定→要素の間隔を0%にする。

横軸を階級値に変更するには、右クリックをして→データの選択→横（項目）軸ラベル→編集→階級値を選択→OK



3. 階級値の取り方とヒストグラム

ここでの目的は、男女でヘモグロビン濃度の分布が異なることがわかるヒストグラムを描くことです。

ヘモグロビン濃度 (mg/dl)

番号	男子	女子
1	15.5	10.0
2	15.0	15.0
3	14.0	14.5
4	14.5	12.5
5	13.5	14.0
6	10.0	17.5
7	16.0	8.5
8	16.5	10.0
9	17.0	11.0
10	15.0	14.5

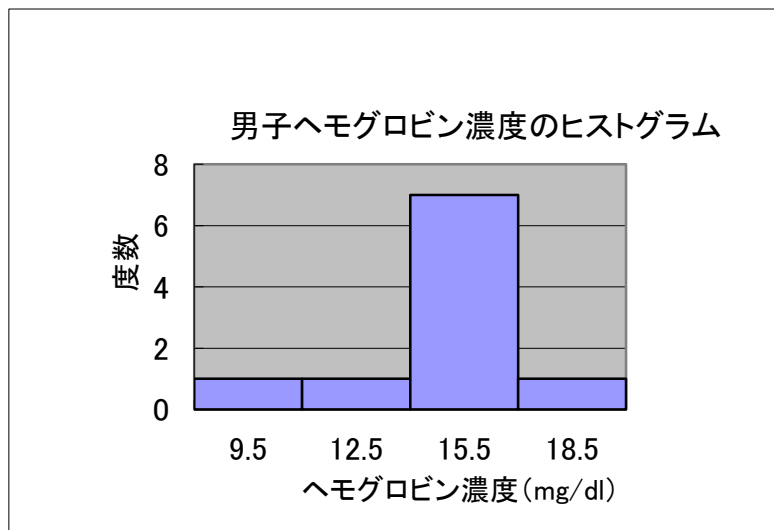
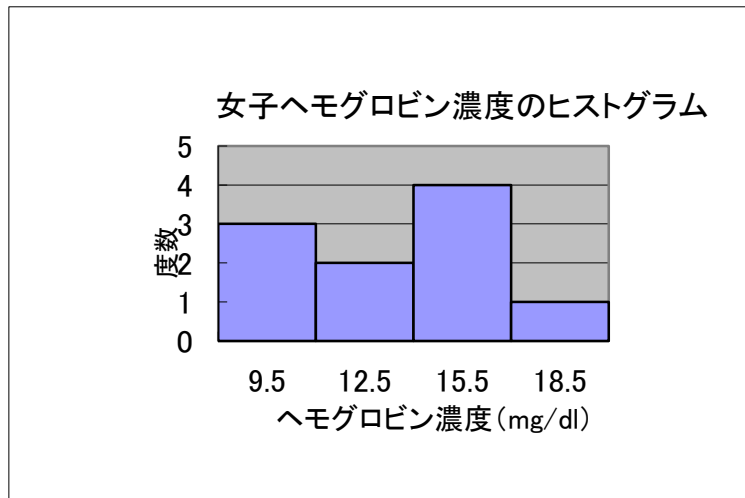
次の度数分布表は階級値の取り方の一例です。階級値の取り方を変えて、度数分布表を作成してみてください。

男子の度数分布表

階級	度数	相対度数	累積度数	累積相対度数	階級値
8~11	1	0.1	1	0.1	9.5
11~14	1	0.1	2	0.2	12.5
14~17	7	0.7	9	0.9	15.5
17~20	1	0.1	10	1.0	18.5

女子の度数分布表

階級	度数	相対度数	累積度数	累積相対度数	階級値
8~11	3	0.3	3	0.3	9.5
11~14	2	0.2	5	0.5	12.5
14~17	4	0.4	9	0.9	15.5
17~20	1	0.1	10	1.0	18.5



4. マスターテーブル

1年の統計学で取り上げた食中毒を疑う場合に作成したマスターテーブルです。簡単な表なので各自で作成してみてください。

表 マスターテーブル

	酢の物を食べた(+)	酢の物を食べなかった(-)	合計
症状あり(+)	52(94.5%) 【88.1%】	3(5.5%) 【8.3%】	55 (100%)
症状なし(-)	7(17.5%) 【11.9%】	33(82.5%) 【91.7%】	40 (100%)
合計	59【100%】	36【100%】	95

()内は行の変数の相対度数、【】内は列の変数の相対度数。

症状(+)
55人のうち、52人が酢の物を食べていることから、症状と酢の物の関係は濃厚です。X²乗検定によって統計的有意差をもとめることができます。

事例:ある料亭で懐石料理を食べた人のうち、吐き気・嘔吐・下痢・腹痛を訴える患者が続出した。状況はマスターテーブルの通りである。

5. 中央値、四分位数、パーセンタイル値

データを数値の大小順に並べた時、ちょうど真ん中の順位になるデータの値を使っても分布を代表させることができます。この代表値を中央値 (Me、メディアン、median) と呼ぶ。中央値は、データ数 n が基数の時は $\{(n+1) \div 2\}$ 番目のデータ値、偶数の時は $(n \div 2)$ 番目と $(n \div 2 + 1)$ 番目のデータを足して 2 で割った値となる。

中央値はデータ全体をその値の上下に 2 分するものですが、さらに、それぞれを 2 分、つまりデータを 4 等分する値を求めることができます。これらを四分位数 (quartile) と呼び、下から第 1、第 2、第 3 四分位数 (Q1、Q2、Q3) と呼びます。もちろん、第 2 四分位数は中央値 Me と一致します。

また、データ数が十分に多い場合には、四分位数をさらに拡張して、全データを 100 等分にする値を求めるともできます。これを百分位数 (パーセンタイル値、percentile) と呼び、第 1、第 2、第 3 四分位数は、それぞれ 25、50、75 パーセンタイル値となります。

ある学年の物理学の成績

35 78 95 55 25 78 88 45 46 30 40 65
 44 70 40 60 80 90 35 44 55 61 65 70
 30 38 78 72 56 44 68 78 53 30 96 78
 86 81 89 72 75 74 36 62 56 47 51 61
 72 36 81 94 25 10 40 30 20 10 50 70
 89 56 23 41 55 70 80 90 12 15 20 60
 80 70 60 50 40 30 20 10 56 46 79 36
 92 81 75 34 26 38 97 50 60 19 84 74
 92 87 73 18 26 38 76 85 65 61 49 56
 59 57 32 45 40 60 18 79 70 60 54 30

パーセンタイル値を求めるには、例えば PERCENTILE.INC(B5:K12,0.975)

最大	97
最小	10
平均	55.8
範囲	87

上限	94.0
Q3	75.0
Me	56.0
Q1	38.0
下限	12.0

6. 箱ヒゲ図

30代、40代、50代、60代の中性脂肪データのパーセンタイル値を示す。

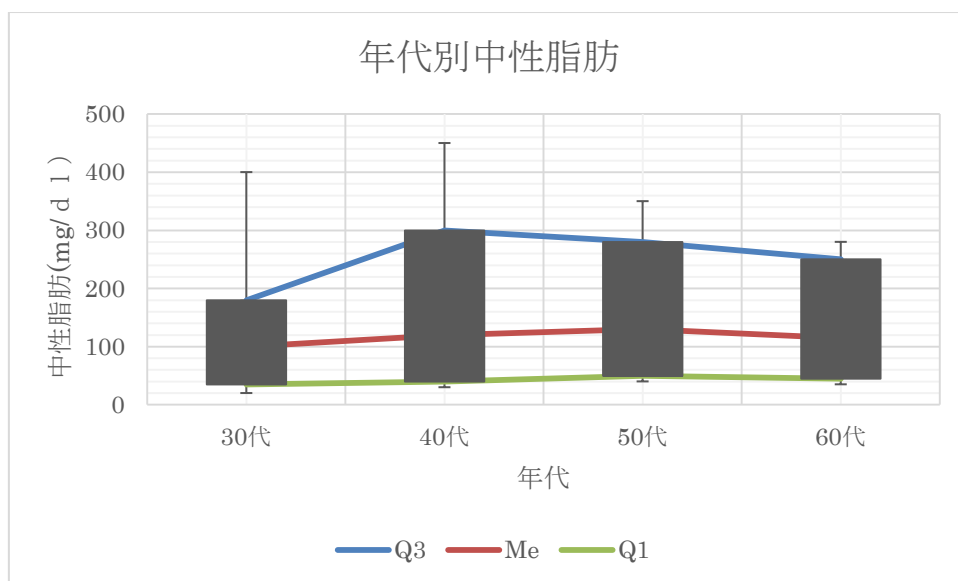
(単位;mg/dl)

	30代	40代	50代	60代
上限	400	450	350	280
Q3	180	300	280	250
Me	100	120	130	115
Q1	35	40	50	45
下限	20	30	40	35

40代、50代、60代の中性脂肪から、次のようなパーセンタイルが求められた。このデータから、箱ヒゲ図を作成してください。

	30代	40代	50代	60代
上限-Q3	220	150	70	30
Q3	180	300	280	250
Me	100	120	130	115
Q1	35	40	50	45
Q1-下限	15	10	10	10

	30代	40代	50代	60代
Q3	180	300	280	250
Me	100	120	130	115
Q1	35	40	50	45



上記した箱ヒゲ図の作り方の手順

まず、下記の表を作ってください。

30代、40代、50代、60代の中性脂肪データのパーセント
イル値を示す。

(単位;mg/dl)

	30代	40代	50代	60代
上限	400	450	350	280
Q3	180	300	280	250
Me	100	120	130	115
Q1	35	40	50	45
下限	20	30	40	35

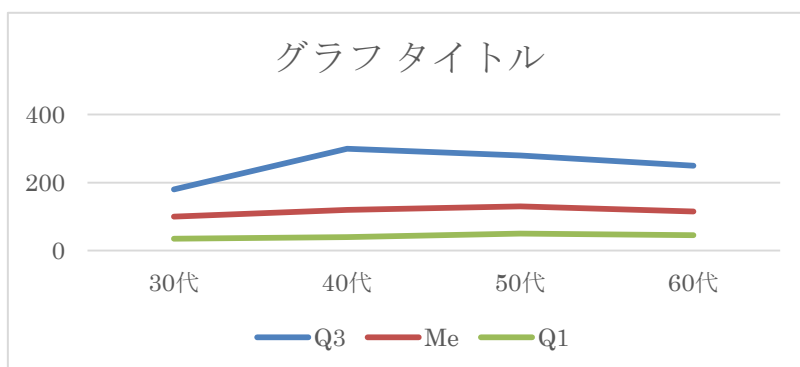
この表に、(上限-Q3)と、(Q1-下限)を加えて、以下の表を作る。

	30代	40代	50代	60代
上限-Q3	220	150	70	30
Q3	180	300	280	250
Me	100	120	130	115
Q1	35	40	50	45
Q1-下限	15	10	10	10

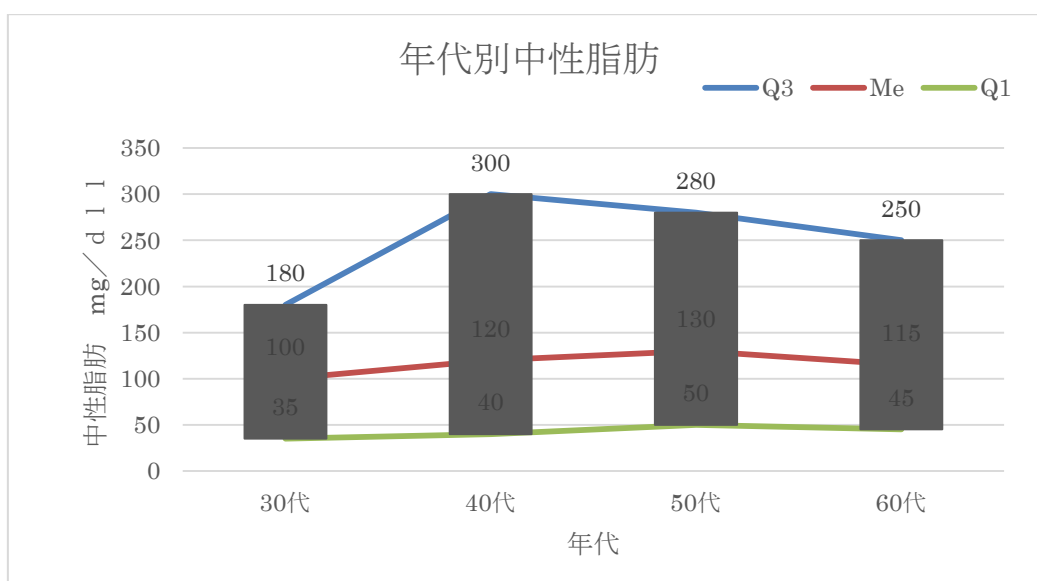
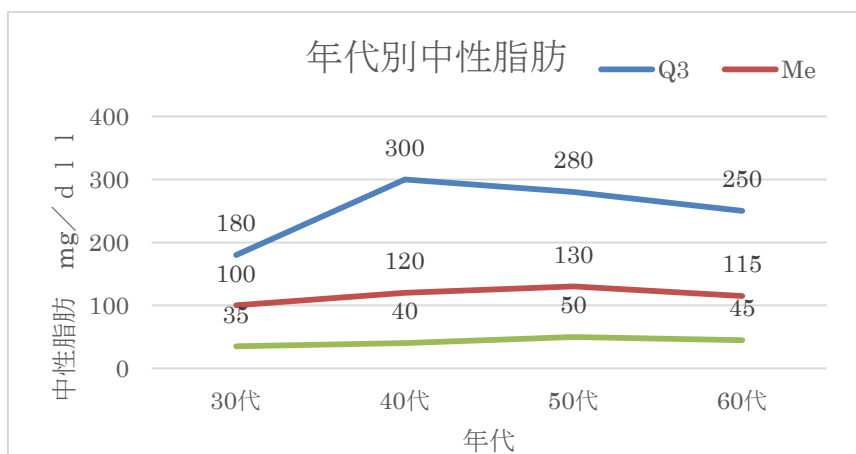
折れ線図を作るために、以下の表を作る。

	30代	40代	50代	60代
Q3	180	300	280	250
Me	100	120	130	115
Q1	35	40	50	45

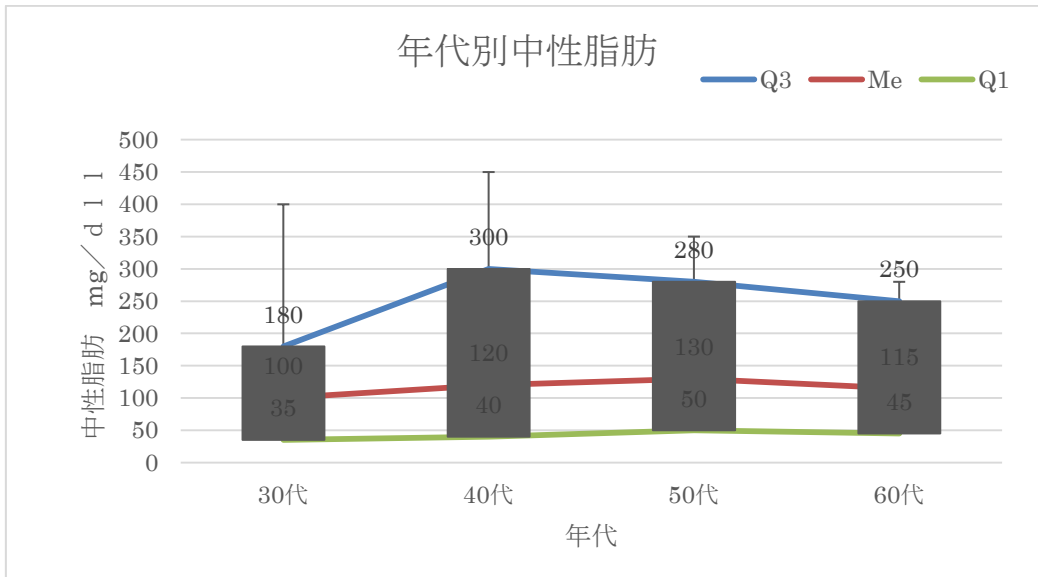
上記の表をすべて選択して、「挿入」→「折れ線グラフ」→「一定順序」を選択。



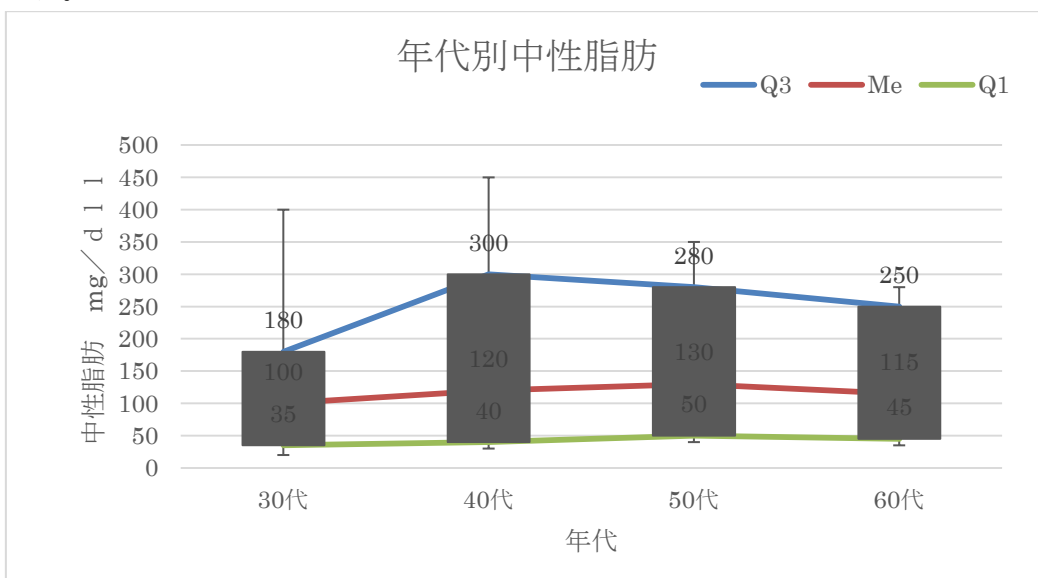
つぎに、グラフタイトル、軸の調整などを行ったうえで、「Me」を選択して、「+マークをクリック」→「ローソク」を選択。



次に、Q3 を選択して、「+マークをクリック」→「誤差範囲」→「その他のオプション」→「正方向」→「ユーザー設定」→「他の指定」を選択。「正範囲」に「上限-Q3」の値を選択。



次に、Q1 を選択して、「+マークをクリック」→「誤差範囲」→「その他のオプション」→「負方向」→「ユーザー設定」→「他の指定」を選択。「負範囲」に「Q1-下限」の値を選択。



演習問題

次のデータは、糖負荷試験後の血糖値の変化を18人について調べた結果です。このデータから、各時点での2.5、25、50、75、97.5パーセンタイル値を求めてください。さらに、その結果に基づいて、箱ヒゲ図を描いてください。

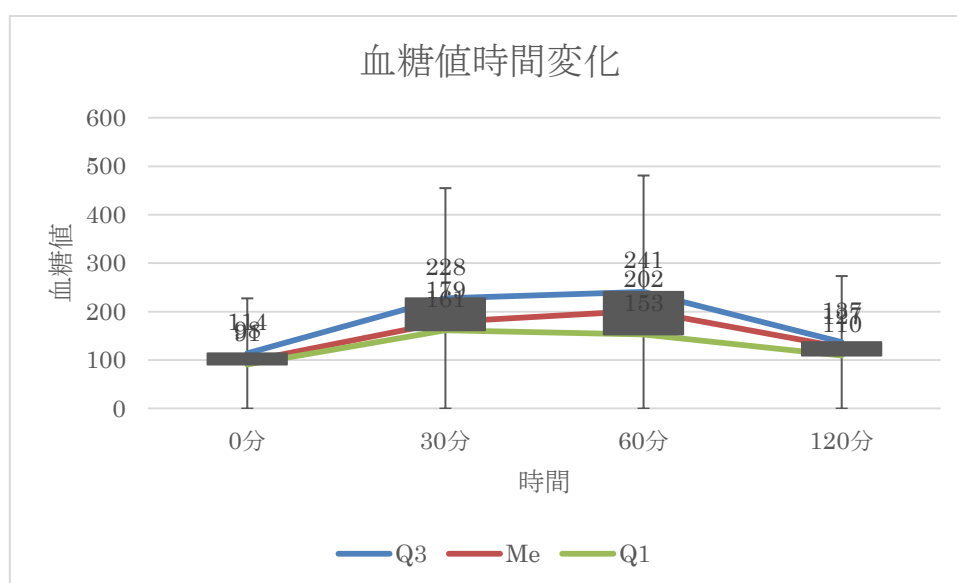
ID	0分	30分	60分	120分
個体1	89	157	148	102
個体2	90	174	181	123
個体3	110	232	284	162
個体4	119	232	278	298
個体5	99	176	168	114
個体6	94	175	221	137
個体7	88	127	139	115
個体8	96	226	217	132
個体9	96	102	115	91
個体10	96	123	144	97
個体11	82	142	182	108
個体12	86	177	137	88
個体13	115	237	299	157
個体14	117	196	208	136
個体15	137	254	311	174
個体16	99	180	236	132
個体17	126	198	195	126
個体18	101	228	242	127

パーセンタイル値	0分	30分	60分	120分
2.5	83.7	110.9	124.4	89.3
25	91.0	161.3	153.0	109.5
50	97.5	178.5	201.5	126.5
75	113.8	227.5	240.5	136.8
97.5	132.3	246.8	305.9	245.3

上限-Q3	18.575	19.275	65.400	108.550
-------	--------	--------	--------	---------

	75	113.8	227.5	240.5	136.8
	50	97.5	178.5	201.5	126.5
	25	91.0	161.3	153.0	109.5
Q1-下限		7.3	50.3	28.7	20.2

	0分	30分	60分	120分
Q3	113.8	227.5	240.5	136.8
Me	97.5	178.5	201.5	126.5
Q1	91.0	161.3	153.0	109.5



7. 相関係数と回帰直線

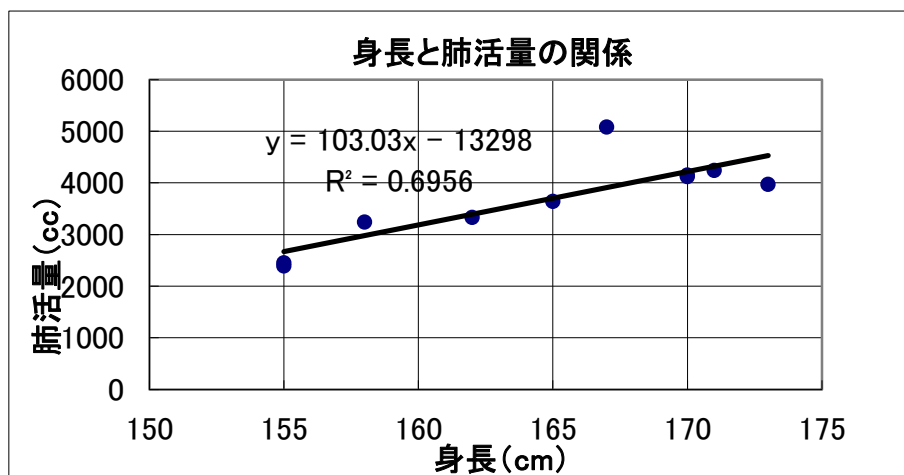
身長から肺活量を予測した表とグラフです。

No.	身長 (cm)	肺活量 (cc)	予測した 肺活量
1	171	4240	4320.4
2	165	3640	3702.2
3	167	5080	3908.3
4	170	4120	4217.4
5	158	3240	2981.0
6	162	3330	3393.1

7	173	3970	4526.5
8	170	4150	4217.4
9	155	2450	2671.9
10	155	2390	2671.9
平均	164.6	3661.0	
標準偏差	6.8	834.2	
身長から	180		5247.7
肺活量を	190		6278.0

予測

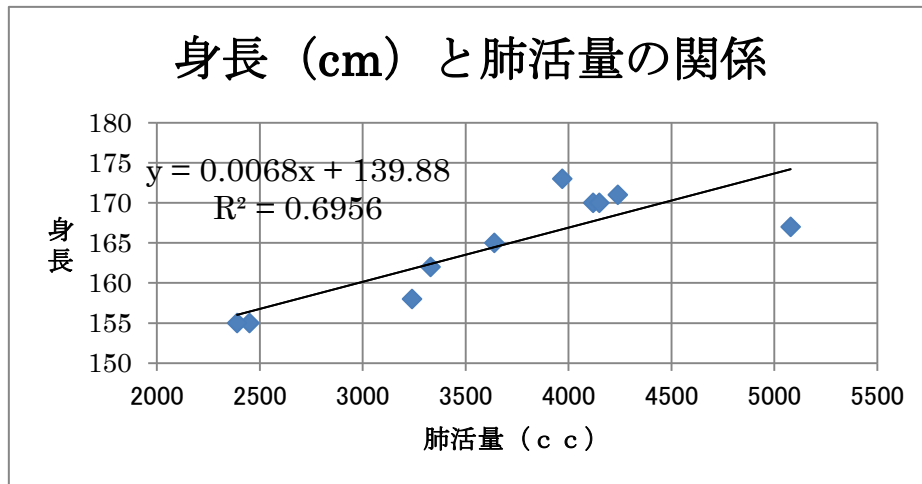
FORECAST (x、既知のy、既知のx)を使う。ここで、xは予測する肺活量の値に対する身長を、数値で示す。既知のyは、既知の従属変数(ここでは肺活量)の値が入力されているセル範囲または配列を指定する。既知のxは、既知の独立変数(ここでは身長)の値が入力されているセル範囲または配列を指定する。



下の表とグラフは、逆に、肺活量から身長を予測した表とグラフです。

No.	肺活量(cc)	身長(cm)	予測した身長
1	4240	171	168.5
2	3640	165	164.5
3	5080	167	174.2
4	4120	170	167.7
5	3240	158	161.8
6	3330	162	162.4
7	3970	173	166.7

	8	4150	170	167.9
	9	2450	155	156.4
	10	2390	155	156.0
肺活量から 身長を予測		6000		180.4
		7000		187.1



ここでは、先ほどとは、逆に肺活量が独立変数、身長が従属変数として、肺活量から身長を予測しています。

相関係数は、CORREL (配列 1、配列 2) を使います。

8. 区間推定

例題 1

区間推定

小学 6 年生 123 人を対象に HDL-コレステロール値を調べた。その結果、平均値は 58.9mg/dl、分散は 220.8 であった。これらの値から HDL-コレステロールの母平均値の 95%信頼区間を求めよ。

新バージョン

Excel では TINV2T
t値は、 1.9796
上限値は 61.55231
下限値は 56.24769

したがって、HDL-コレステロールの母平均値の 95%信頼区間は次のようになります。

$$56.25 \text{ mg/dl} < \mu < 61.55 \text{ mg/dl}$$

例題 2

生徒 500 人の中で 10 人のテスト結果が次のようであった。

1	60
2	70
3	80
4	80
5	80
6	90
7	100
8	80
9	70
10	90
平均	80
u	10.95
s	11.55

nで割った値

n-1 で割った値

標準偏差が未知のとき

t 分布の逆関数

は 2.3

下限値は 71.7

上限値は 88.3

したがって、生徒 500 人のテストの平均値の 95%信頼区間は次のようになります。

$$71.7 < \text{平均値} < 88.3$$

9. 検定

対応のあるt検定の方法

患者	術前の体温	術後の体温	差	差の偏差	差の偏差平方
イ	36.7	35.5	-1.2	0.4	0.16
ロ	36.5	35.6	-0.9	0.7	0.49
ハ	36.5	35.6	-0.9	0.7	0.49
ニ	36.4	35.6	-0.8	0.8	0.64
ホ	36.6	35.1	-1.5	0.1	0.01

差の平均 -1.1 偏差平方和 1.79

データ数 5 標準誤差 0.30

t -3.54

対応のない場合のt検定の方法

男性と女性の身長に差があるかどうか			
	個数	平均値	標準偏差
男性	25	152.9	5.19
女性	34	139.7	5.547
t 分布表から求めた値	3.47		
t 値	9.13		

$\alpha = 0.001$ で、t分布表から自由度=57、両側検定により、棄却p値(限界危険率)は 3.47、 $t = 9.13 > t_{57}(0.001) = 3.47$ であるから、男女の身長の平均が等しいという仮説は、有意水準 $\alpha = 0.001$ で棄却される。すなわち、男性の身長の平均は女性よりも高い。

t 分表から t を求めるには、TINV2T (確率、自由度) で求める。
計算から求める t は、

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

です。

10. ピボットテーブル

ピボットテーブルの作成

データ内のセルを選択、「挿入」、「テーブル、ピボットテーブル」をクリック、「ピボットテーブルの作成」ダイアログボックスを確認、自動的に選択されたセルの範囲を確認、セル範囲の異なる場合は、ここをクリックして選択し直す、作成場所を確認、OK をクリック

データ作成には、次のことに注意する。先頭行に列ラベル（項目名）があること、データの範囲内に空白行、または空白列がないこと、小計や合計の行、列がないこと、ある場合は削除しておくこと。

ピボットテーブルをデータと同じワークシート内に作成するには、「ピボットテーブルの作成」ダイアログボックスで既存のワークシートを選択、場所にピボットテーブルを作成するセルを選択する。

エクセル以外のデータからピボットテーブルを作成するには、「ピボットテーブルの作成」ダイアログボックスで「外部データソースを使用」を選択、「接続の選択」ボタンをクリック、ダイアログボックスが表示、「参照」をクリック、「データファイルの選択」ダイアログボックスでデータベースファイルを選択する。

行ラベル、列ラベル、値エリアにフィールド（列見出し；項目名）をドラック

値エリアにドラックしたフィールドの集計は、数値データの場合は合計、文字データの場合はデータの個数を数える計算をする。フィールドを間違えた場合は、そのフィールドを削除し、操作をやり直す。

フィールドリストウィンドウは、フィールドセクションとエリアセクションの配置を変更するときに使う。「フィールドリスト」の右上のボタンをクリックして、レイアウトを選択する。

レポートフィルターエリアにフィールドを追加すると、設定したフィールドのアイテムごとにピボットテーブルの切り替えが可能になる。

ピボットテーブルを作成すると、列や行の総計が表示される。列や行の集計は「ピボッ

「テーブルツール」「デザイン」「レイアウト」グループ、「総計」をクリック、「列のみ集計を行う」をクリックする。

小計行の表示を変更するには、「レイアウト」グループ、「小計」をクリック、「すべての小計をグループの先頭に表示する」をクリックする。

空白行を挿入して、表を見やすくするには、「レイアウト」グループ、「空白行」をクリック、「アイテムの後ろに空行を入れる」を選択する。

元データを変更した場合、ピボットテーブルの更新を行うことが必要になる。更新には2つの方法がある。「ピボットテーブルツール」、「オプションタブ」「データ」、「更新」をクリックすると、「更新」と「すべて更新」が表示される。「更新」をクリックすると、選択しているピボットテーブルのみが更新。「すべて更新」をクリックすると、元データで作成したすべてのピボットテーブルが更新される。

元データに新しいデータを追加した場合、ピボットテーブルのデータ範囲を変更する。「ピボットテーブルツール」「オプション」「データ」グループにある「データソースの変更」ボタンをクリック、「ピボットテーブルのデータソースの変更」ダイアログボックスで新しいデータ範囲を指定する。

データ表示の切り替え

ピボットテーブルに追加した各フィールドは、必要に応じてアイテムの表示と非表示を切り替えできる。必要なアイテムだけを表示してデータを分析できる。

これには、「スライサー」「レポートフィルターフィールド」「行ラベルフィールド」や「列ラベルフィールド」を利用する。

スライサーでフィールドの項目を絞り込むには、表の中をクリックし、「ピボットテーブルツール」「オプション」をクリック、「並べ替えとフィルター」グループの「スライサー」をクリック、「設定したいフィールド名をクリックしてチェックマークをつける。「OK」をクリックする。指定したフィールドのスライサーが表示、表示したいアイテムのボタンをクリックする。複数のアイテムを選択するには、スライサー内の最初のアイテムをクリックし、2つ目以降のアイテムは「コントロール」キーを押しながらクリックする。選択したアイテムを解除するには、スライサーの右上にある「フィルターのクリア」ボタンをクリックする。スライサーを削除するには、スライサーを選択して「デリート」キーを押す。スライサーを複数表示するには、「スライサーの挿入」ダイアログボックスで、複数のフィールド名にチェックマークをつける。

レポートフィルターフィールドを使うとドロップダウンリストでアイテムを絞り込むことができ、アイテムごとに集計したピボットテーブルを表示できる。「レポートフィルター」のフィルターボタンをクリック、項目名をクリック、OK ボタンをクリックする。項目を複数選択するには「フィルターボタン」をクリック、「複数のアイテムを選択」をクリックしてチェックマークをつける。アイテムごとのピボットテーブルを新しいワークシートに作

成するには、「レポートフィルター」は「すべて」を選んでおく。「ピポットテーブルツール」「オプション」をクリック、「レポートフィルターページの表示」をクリック、フィールドの選択、OK をクリックする。

行ラベルや列ラベルフィールドにあるフィールドのアイテムは、いろいろな条件で絞り込むことができる。絞り込みの条件には、特定のアイテムや値を指定できる。行ラベルフィールドのフィルターボタンをクリック、すべて選択をクリックしてチェックマークをつけ、OK をクリックする。上位項目だけを表示するには、行ラベルフィールドのフィルターボタンをクリック、値フィルターにマウスポインターを合わせ、例えば「トップテン」をクリックする。条件を設定、OK ボタンをクリックする。詳細データを表示するには、詳細データを表示したい項目をクリック、「ピポットテーブルツール」「オプション」タブをクリック、「アクティブなフィールド」グループ、「フィールド全体の展開」をクリック、「詳細データを表示するフィールド名を選択」、OK をクリックする。

レイアウトの変更

ピポットテーブルに追加したフィールドは、簡単に追加や移動、削除ができる。フィールドを移動するには、例えば「列ラベル」エリアにある項目を「行ラベル」エリアにドラック、「行ラベル」エリアにある項目を「列ラベル」エリアにドラックする。フィールドのレベルを変更するには、例えば「列ラベル」エリアの項目のレベルをドラックして入れ替える。

フィールドのグループ化

フィールドのアイテムは、いろいろなグループにまとめることができる。例えばフィールドのデータが日付データの場合、月単位や年単位にまとめた集計結果を表示できる。日付が入力されているフィールドをクリック、「ピポットテーブルツール」「オプション」タブをクリック、「グループ」グループ、「グループフィールド」をクリック、「月」をクリック、OK ボタンをクリックする。数値のデータをグループ化するには、数値データが入力されているフィールドをクリック、「ピポットテーブルツール」「オプション」タブをクリック、「グループ」グループ、「グループフィールド」をクリック、「先頭の値」に「0」と入力、「末尾の値」に25000 と入力、「単位」に「5000」と入力、OK とクリックする。

文字データが配置されているフィールドでは、アイテムを指定してグループ化することができる。例えば「支店」のデータの中から「宇都宮」と「東京本社」を1つのグループにまとめることが可能。グループ名は任意の名前を付けられる。

グループ化するデータを「コントロール」キーを押しながら選択、「ピポットテーブルツール」「オプション」タブをクリックする。「グループ」グループ、「グループの選択」をクリックする。

集計方法の変更

レイアウトセクションの「値」エリアに数値データのフィールドを追加すると、自動的に計算方法が決められる。この計算方法は後から変更できる。また、フィールドを追加すれば複数の集計結果を表示することが可能。

「値」エリアには複数のフィールドを配置できる。同じフィールドのデータで異なる計算をしたいときは、まず2つ目のフィールドを追加、計算方法は後で変更する。例えば「購入金額」を「値」エリアにドラッグ、集計用のフィールドが追加、必要に応じて追加したフィールドの集計方法を変更する。ここでは、値フィールドの集計方法を変更する方法を考える。

値フィールドに配置した2つ目の「購入金額」フィールドの計算方法を合計から戸数に変更してみる。集計方法を変更したいフィールドをクリック、「ピボットテーブルツール」、「オプション」タブをクリック、「アクティブなフィールド」グループ、「フィールドの設定」をクリック、「集計方法」タブをクリック、「データの個数」を選択、OKをクリックする。

計算の種類を変更するには、「計算の種類を変更したい」フィールドをクリック、「ピボットテーブルツール」、「オプション」タブをクリック、「計算の種類」タブをクリック、「総計に対する比率」を選択、OKをクリックする。

元の数値に戻すには、「ピボットテーブルツール」、「オプション」 「計算方法」グループにある「計算の種類」ボタンをクリックし、一覧から「計算なし」を選択する。

ピボットテーブルの書式

ピボットテーブルを見やすくするには、数値データの表示形式やピボットテーブル全体のスタイルを変更する。

値フィールドの表示形式を変更するには、表示形式を変更したいフィールドをクリック、「ピボットテーブルツール」、「オプション」タブをクリック、「アクティブなフィールド」グループ、「フィールドの設定」をクリック、「表示形式」をクリック、「数値」をクリック、「桁区切り (,) を使用する」をクリックしてチェックマークをつける。OK をクリックする。

表の形式を変更するには、「ピボットテーブルツール」、「デザイン」タブをクリック、「レイアウト」グループ、「レポートのレイアウト」をクリック、「アウトライン形式で憑依」をクリックする。

スタイルを変更するには、「ピボットテーブルツール」、「デザイン」タブをクリック、「ピボットテーブルスタイル」グループ、「その他」をクリック、スタイルを選択する。

ピボットグラフ

ピボットグラフは、ピボットテーブルをグラフ化したもの。ピボットテーブルとピボットグラフは連動、ピボットテーブルを変更するとそれに合わせてピボットグラフも変更さ

れる。また、グラフのフィールドは、テーブルと同じようにドラック操作で変更可能。

テーブルからグラフを作成するには、表の中のセルをクリック、「ピポットテーブルツール」、「オプション」タブをクリック、「ツール」グループ、「ピポットグラフ」をクリック、グラフの種類を選択、OK をクリックする。

グラフをグラフシートに移動するには、「ピポットグラフ」を選択、「ピポットテーブルツール」、「デザイン」タブの「場所」グループにある「グラフの移動」ボタンをクリック、「グラフの移動」ダイアログボックス、「新しいシート」を選択、OK をクリックする。

入力したデータから直接グラフを作成するには、ピポットテーブルとピポットグラフが同時に作成できる。表の中をクリックしておく。「挿入」タブ、「テーブル」グループ、「ピポットテーブル」をクリック、「ピポットグラフ」をクリック、作成場所を選択、OK をクリックする。

グラフにフィールドを追加するには、あらかじめ、グラフをシートに移動する。ピポットグラフを選択、例えば「分類」を「軸フィールド（項目）」エリアにドラックする。

グラフのフィールドを別のエリアに移動するには、例えば「分類」を「レポートフィルター」エリアにドラックする。フィールドを削除するには、削除したいフィールドをレイアウトセクションの外にドラックする。フィールドのレベルを変更するには、フィールドの上下を入れ替える。

11. 家計簿 2019

1. 家計簿作成に必要なエクセルの機能について

1.1 複数シートの選択（オートフィル）；

シートを複数選択して 1 度に表を作成したり、罫線を引くことができます。選択した複数のシートのまとまりを「作業グループ」と呼んでいます。連続したシートを選択するときは、最初のシートをクリックし、Shift キーを押しながら最後のシートまでクリックします。離れたシートを選択するときは、最初のシートをクリックし、Ctrl キーを押しながら選択するシートをクリックして行きます。なお、作業グループを解除するには、選択しているシート以外の見出しをクリックします（または、作業グループ内のどれかのシート見出しを右クリックして、作業グループモードを解除します）。

1.2 数式を入力する；

セルに入力したデータを使って計算をするには、数式を入力します。数式は計算するセルの番地を使って作成します。セルの番地で計算することで、数式が参照しているセルの値が変わっても、常に最新の計算結果を求めることができます。

1.3 数式をコピーする；

セル番地を使って入力した数式は、同じような計算をする他の行や列にコピーするこ

とができます。数式をコピーすると大きな表の計算項目もすぐに完成します。相対的なセル参照式のコピーと、絶対的なセル参照式のコピーの 2 通りがあります。絶対番地は列番号と行番号の前に「\$」記号を付けて指定します。

1.4 関数を入力する；

エクセルでは合計や平均、またはセルの個数を数えるなどの多くの関数が用意されています。関数を使うと単純な四則演算だけでなく、ローンの返済額や減価償却額などを計算することができます。

1.5 シートを串刺しで合計する；

月や部署などで分けて作成した複数のシートを、串刺しにして合計を求めることができます。串刺しで合計を求めるには、シートのどの表も同じような構成でなければなりません。串刺しの合計を求めるシートとセルを選択します。ツールバーで「オート SUM」ボタンをクリックして、SUM 関数を仮入力します。串刺し合計する先頭のシートとセルを指定します。次にシフトキーを押しながら串刺し合計する最後のシートを指定します。**Enter** キーを押して、セル SUM 関数を入力します。

1.6 IF 関数について；

この関数は家計簿の作成に必要な論理関数です。

1.7 セルのデータをコピーする；

セルのデータだけや書式だけなど指定して貼り付けるには、ショートカットメニューの「形式を選択して貼り付け」を使います。

2. 論理関数

論理関数は条件が TRUE (真) か FALSE (偽) か確認したり、複数の条件を検査するときに使います。特に、IF 関数は、関数の中でも多用されるものの 1 つです。

- ・ AND 関数；書式 → =AND (論理式 1, [論理式 2, 論理式 3, ...])

引数のすべての論理式が真のときに「TRUE」を返し、1 つでも偽があるときは「FALSE」を返します。

- ・ OR 関数；書式 → =OR (論理式 1, [論理式 2, 論理式 3, ...])

引数のいずれかの論理式が真のとき「TRUE」を返し、すべてが偽であるときは「FALSE」を返します。

- ・ NOT 関数；書式 → =NOT (論理式)

引数が真であるとき「FALSE」を返し、偽であるとき「TRUE」を返します。

- ・ IF 関数；書式 → =IF (論理式, 真の場合, 偽の場合)

論理式が真であるとき「真の場合」を返し、偽であるとき「偽の場合」を返します。

使用例：=IF (50<=100, “正しい”, “誤り”) → 正しい

=IF (AND (50<=100, 10>=5), “正しい”, “誤り”) → 正しい

=IF (AND (50<=100, 10<=5), “正しい”, “誤り”) → 誤り

=IF (OR (50<=100, 10<=5), “正しい”, “誤り”) → 正しい

・SUMIF 関数；書式 → =SUMIF (範囲, 検索条件[, 合計範囲])

指定された検索条件に一致するセルの値を合計します。

範囲：検索条件の対象となる数値を含む範囲を指定します。

検索条件：検索条件を数値、式、文字列で指定します。

合計範囲：実際に計算の対象となる数値の範囲を指定します。

範囲の中で検索条件に一致する値だけが合計されます。合計範囲を省略すると、範囲の中で条件を満たすセルが合計されます。

3. 家計簿の作成

まず、シートを12枚ほど作成して、各シート名を1月分、2月分～12月分までのシート名に変更します（シート見出しをダブルクリックして）。

1月分～12月分までの12枚のシートを選択して作業グループを作り、入力を開始します。

3.1 家計簿の外枠；

- 1) A2→ “家計簿” ; タイトル、文字サイズ、フォント飾りなどを付ける
- 2) D2→ “1月分” ; 何月分の家計簿であるかを入力する（後で修正が必要）
- 3) E2→ “前月繰り越し額”；セル E2 と F2 を結合する、セルの下に罫線を設定
- 4) G2→ ; 前月繰り越し額を入力する（2月以降については後で説明します）
- 5) A4→ “No.” ; この列は通し番号とする
- 6) A5→ “1”
- 7) A6→ “2”
- 8) A7～A60→ ; A5、A6 を選択してオートフィルで連続番号を入力する
- 9) B4→ “日付”
- 1 0) B5～B60→ ; この部分には、家計簿の入出金記録の日付を入力する
- 1 1) C4→ “分類”
- 1 2) C5～C60→ ; この部分には、入出金の分類を入力する。入力にはリストを使用する
- 1 3) D4→ “明細”
- 1 4) D5～D60→ ; この部分には、入出金の明細を入力する
- 1 5) E4→ “入金”
- 1 6) E5～E60→ ; この部分には、入金額を入力する
- 1 7) F4→ “出金”
- 1 8) F5～F60→ ; この部分には、出金額を入力する
- 1 9) G4→ “残高”
- 2 0) G5→ “=IF(AND(E5=0,F5=0), G2,G2+E5-F5)”
- 2 1) G6→ “=IF(AND(E6=0,F6=0), G5,G5+E6-F6)”

- 2 2) G7～G60→ ; G6 をオートフィルして計算式を入力する
- 2 3) E61→ “当月残高” ;セル E61 と F61 を結合する。セルの下に罫線を設定する
- 2 4) G61→ “=G2+SUM(E5:E60)-SUM(F5:F60)”

上記のようにして項目名やナンバー、計算式などを入力して、家計簿を作成します。

3.2 費目名リストの作成 ;

次に、分類をリストから入力できるように、費目名のリストを作成してみます。

- 1) I4→ 費目
- 2) I5→ 給与
- 3) I6→ 賞与
- 4) I7→ 雑収入
- 5) I8→ 食料費
- 6) I9→ 飲料費
- 7) I10→ 外食費
- 8) I11→ 住居費
- 9) I12→ 娯楽・教養費
- 1 0) I13→ 衣料費
- 1 1) I14→ 交際費
- 1 2) I15→ 医療費
- 1 3) I16→ 交通・通信費
- 1 4) I17→ 水道・光熱費
- 1 5) I18→ 保険料
- 1 6) I19→ 教育費
- 1 7) I20→ 貯金
- 1 8) I21→ 生活雑貨費
- 1 9) I22→ 車両費
- 2 0) I23→ 雑費

罫線など引いて表を完成させてください。

3.3 入力規則

次に、費目名のリスト範囲、セル I5～I23 を選択します。【名前ボックス】に“費目名”と入力します。これで、セル I5～I23 に「費目名」という名前が設定されます。これが終了したら作業グループを解除してください。

分類のセル C5～C60 をリストから選択できるようにするために、入力規則を設定します。まず、セル C5～C60 を選択します。次に、【データ】メニューから【入力規則】を選びます。【設定】タブで【入力値の種類】から【リスト】を選びます。【元の値】に【=費目名】と入力し、【ドロップダウンリストから選択する】をチェックして、【OK】をクリックしま

す。そうすると、分類を費目名のリストから入力できるようになります（=が小文字であることを確認すること！）。

次に、1月分のシートのC5～C60をコピーします。そして、2月分～12月分までのシートを選択（作業グループ）して、2月分のシートのC5にアクティブセルを移動して、貼り付けを行います。ここで、作業グループを解除して、各月で分類が費目名リストから入力できることを確認します。

3.4 費目の合計；

次に、作成した家計簿に、分類別の出費を表す表とグラフを追加します。まず、全シートを選択して同時にこれらの作業を実施します。

- 1) J4→“費目合計”
- 2) J5→“=SUMIF(\$C\$5:\$C\$60,I5,\$E\$5:\$E\$60)”
- 3) J6,J7→J5をコピーして貼り付けます（あるいは、J5をオートフィル）。
- 4) J8→“=SUMIF(\$C\$5:\$C\$60,I8,\$F\$5:\$F\$60)”
- 5) J9～J23→；セルJ8をコピーして貼り付けます（あるいは、J8をオートフィル）。

ここでグループを解除。

次に、前月繰り越し額の入力について説明します。1月分のシートでは、G2に昨年の残額を入力します。2月分のシートでは、1月分のシートのG61をコピーして2月分のシートのG2に貼り付けます。このとき、貼り付けについては、注意が必要です。[編集]から[形式を選択して貼り付け]を選択し、さらに[リンク貼り付け]を選択して実行します。以降は同じようにして12月分までの繰り越し額を入力します。

3.5 年間集計；

最後に年間集計表の作成をします。まず、年間集計表を作る新しいシートを作成します。新しく挿入する右側のシートを選択して、【挿入】－【ワークシート】を選択します。新しくできたシート見出しを年間集計とします。

次に、1月分のシートのI4～I23までをコピーして年間集計のシートのI4に貼り付けます。次に、J5にアクティブセルを移動して、ツールバーで「オートSUM」ボタンをクリックして、SUM関数を仮入力します。次に、串刺し合計する先頭のシートである1月分のシートのJ5を選択し、シフトキーを押したまま、串刺し合計する最後のシートである12月分のシートのJ5を選択して、エンターキーを押します。これで年間合計のシートで給与に関する合計が入力されました。次に、オートフィル機能を使ってJ5の内容をJ23まで貼り付けてください。これで、一応、年間集計のシートは完成です。

4. グラフ作成

グラフ作成の範囲を選択します。ここでは、支出の割合をグラフ化するのでI8～J23を選択します。【グラフウィザード】ボタンをクリックし、グラフウィザードを起動します。【グラフの種類】から、【円】を選びます。このグラフの形は、好みで選んでください。選

扱が終了したら、【次へ】をクリックします。グラフウィザードの【3/4】で【データラベル】タブをクリックし、【データラベル】から、【ラベルとパーセンテージを表示する】を選択します。次に、【凡例】タブをクリックして【凡例を表示する】のチェックをはずして【次へ】をクリックします。グラフの場所を【オブジェクト】にして、【完了】をクリックします。これで、グラフが完了します。グラフの大きさや場所などは表に合わせて見やすいように調整してください。

最後に複数シート選択を解除してください。

【復習】 区間推定と検定

母集団と標本；

集団の中である知りたい値を母数 (parameter) という。母数を知ろうとする対象を母集団 (population) と呼び、そこから抽出する一部を標本 (sample) と呼ぶ。統計学では、その標本から母集団の母数を推定する。

正規分布；

身長や知能指数などの分布は正規分布に従う。この分布は平均値を中心とした左右対称の釣鐘状の分布です。この分布によると、「平均値±1.96×標準偏差」の範囲にすべてのデータの 95%が存在するなどの性質が分かっています。この性質を利用して推定や検定を行います。

t分布；

標本数が少ない時には、正規分布の代わりに t 分布を使って推定とか検定を行います。

区間推定；

「真の母平均は値 A と値 B の間 (区間) に “おそらく” 含まれているだろう」というような考え方で母平均の推定値をあらわすことができる。

このような考え方で、標本から母集団の母数を推定することを区間推定という。母集団の正規分布が仮定できる場合、“おそらく “の程度を信頼度 (信頼係数) といい、一般に利用されるのは 0.95 (95%) である。また、設定した信頼度 (95%) で母数を含む区間のことを (95%) 信頼区間という。

t分布を使った区間推定；

データ数が n 、標本平均が $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$ 、標本標準偏差が $s = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$

のとき、次のように母平均 μ の取り得る範囲を推定することができます。 μ の 95% の信頼区間は

$$\bar{x} - t_0 \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_0 \frac{s}{\sqrt{n}}$$

ここで、 t_0 は t 分布表より求めることができます。自由度とは $n - 1$ です。

検定 (平均値の差の検定) :

2つの学校における体重の差やヘモグロビン濃度 (貧血検査) の差、あるいは 75g 経口ブドウ糖負荷後 2 時間値の 2つの地域における差など、2つの集団における平均値の差の検定を必要とする場合がしばしばある。

「差がある」という仮説を検定する場合、差の程度が不明なため、そのままでは検定できない。そこで考え出されたのが、その逆の「差がない」という仮説を検定して、それに何らかの矛盾が見つければ、もとの「差がある」という仮説を採用する。逆に、明らかな矛盾がないときにはその判定を保留する。このような論法で検定を行います。

ここで、「差がない」という仮説は本来「無」に帰すべきものとして「帰無仮説」(null hypothesis) と呼び、 H_0 と略す。また、もとの「差がある」という仮説は、「対立仮説」(alternative hypothesis) と呼び、 H_1 と略す。

仮説の設定としては :

例えば、2つのグループ (A, B) 間で「差がある」という仮説について検定する場合、まず「2つのグループ間には差がない ; $A=B$ 」とう仮説 (帰無仮説) を設定し、元々証明したい「差がある ; $A \neq B$ 」という仮説 (対立仮説) は、一旦伏せておく。

実際の検定には、帰無仮説を誤って捨てる確率 α (第1種の過誤の確率で危険率とも言う) を示し、そのときの検定統計量から確率を出して、

その値が α よりも小さいときに

↓

「有意水準 α で統計的に有意である」といい、帰無仮説を棄却する。

検定統計量より得られる実際の確率を P 値 (P-value) と呼び、

P 値が α 以下となるような統計量の範囲を

棄却域という。

一般に有意水準 α には 5% (0.05)、1% (0.01) が用いられる。
 t 分表から t_0 を求めるには、TINV (確率、自由度) で求める。
計算から求める t は、

$$\frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

です。