

確認問題の解答

問題 15.1

- | | | |
|--------------|--------------------|-------------------|
| (1) 4 | (2) 4, 5, 平均値 | (3) トリム平均 |
| (4) 最大値, 最小値 | (5) 最小値 | (6) レンジ, 分散, 標準偏差 |
| (7) 偏差 | (8) 2 乗, (データの) 個数 | (9) 標準偏差 |
| (10) 分散 | (11) 0, 1 | (12) 平均値, 標準偏差 |
| (13) 3 | (14) 256 | |

問題 15.2

質的変数

性別, 年代, 成績の段階評価, 学生番号, 出身地, 時間帯, 順位, 検定の級, 電話番号, 震度

量的変数

偏差値, 身長, 年齢, 値段, 体重, 時刻, 睡眠時間, 気温, 西暦, 速度

問題 15.3

- | | | |
|----------------|--------------------|------------------|
| (1) AVERAGE 関数 | (2) MEDIAN 関数 | (3) MODE.MULT 関数 |
| (4) MAX 関数 | (5) MIN 関数 | (6) VAR.P 関数 |
| (7) STDEV.P 関数 | (8) STANDARDIZE 関数 | (9) TRIMMEAN 関数 |
| (10) CORREL 関数 | | |

問題 15.4

- | | | |
|------------|--------------|-------------------|
| (1) 折れ線グラフ | (2) 散布図 | (3) ダミー変数 |
| (4) クロス集計 | (5) 相関 | (6) 相関係数 |
| (7) -1, 1 | (8) 1, -1, 0 | (9) 1 |
| (10) 疑似相関 | (11) 最小 2 乗法 | (12) 予測値 |
| (13) 残差 | (14) R-2 乗値 | (15) 0, 1 |
| (16) 1 | (17) 0 | (18) 相関係数, R-2 乗値 |
| (19) 0 | (20) 1 | |

問題 15.5

- | | |
|--------------------|------------------------------|
| (1) 価格, 仕入れ値 | (2) 売上個数, 粗利 (または, 粗利, 売上個数) |
| (3) 売上個数, 価格, 仕入れ値 | |

問題 15.6

- | | | |
|-----------|-----------|--------------|
| (1) 季節変動 | (2) 無作為変動 | (3) 移動平均 |
| (4) 季節変動値 | (5) 季節調整 | (6) 補正, 季節指数 |
| (7) 季節指数 | | |

問題 15.7

- | | | |
|------------------|---------|------------|
| (1) 階級 | (2) 度数 | (3) 相対度数 |
| (4) 累積度数, 累積相対度数 | (5) 1 | (6) 全度数 |
| (7) 1 | (8) 0.5 | (9) ヒストグラム |

問題 15.8

- | | |
|--|-------------|
| (1) 合計, データの個数 | (2) 合計, 平均値 |
| (3) 平均値, データの個数, 合計 (または, データの個数, 平均値, 合計) | |
| (4) 合計, データの個数 | (5) 平均値 |
| (6) 平均値, 合計, データの個数 | (7) データの個数 |

問題 15.9

- | | |
|--|---------|
| (1) 平均値, 中央値, 最頻値 (または, 平均値, 最頻値, 中央値) | (2) 中央値 |
| (3) 最頻値 | (5) 最頻値 |
| (6) 平均値 | (7) 平均値 |
| | (8) 最頻値 |

問題 15.10

	国語	数学	英語
A	65	57	88
B	55	72	68
C	70	75	=I8
D	70	84	60
E	80	66	64
F	65	=H9	86
最大値	=MAX(G2:G7)	=MAX(H2:H6)	=I9+I10
最小値	=MIN(G2:G7)	=H8-H10	=MIN(I2:I3,I5:I7)
レンジ	=G8-G9	27	36

国語の最大値 a は 80, 最小値 b は 55 なので, レンジ c は $80 - 55 = 25$ である.

数学の最大値 e が 84, レンジは 27 なので, 最小値 d は $84 - 27 = 57$ である.

英語の最小値 g は 60, レンジが 36 なので, 最大値 f は $60 + 36 = 96$ である.

	国語	数学	英語
A	65	57	88
B	55	72	68
C	70	75	96
D	70	84	60
E	80	66	64
F	65	57	86
最大値	80	84	96
最小値	55	57	60
レンジ	25	27	36

問題 15.11

各支店について、平均値、中央値、最頻値、最大値、最小値、分散、標準偏差をエクセル関数で求める。
レンジについては、MAX 関数で求めた最大値から MIN 関数で求めた最小値をひいて求める。

	平均値	中央値	最頻値	最大値	最小値	レンジ	分散	標準偏差
A店	848392.5	872685	#N/A	1036890	697410	339480	8402173819	91663.4
B店	571642.5	577485	597780	630990	509220	121770	1623625369	40294.2
C店	317032.5	335790	335790	376380	225090	151290	2093380819	45753.5
D店	780742.5	789660	889290	889290	660510	228780	5511021919	74236.3
E店	542737.5	498150	#N/A	1055340	217710	837630	65249391319	255439.6

- 平均値のほうが中央値より大きい支店は E 店である。
- 最頻値が求められない支店は A 店と E 店である。
- レンジが最も小さい支店は B 店である。
- 分散が最も大きい支店は E 店である。
- 標準偏差が最も小さい支店は B 店であり、その標準偏差の値は約 40294.2 (円) である。B 店について、Excel アドインのデータ分析ツールの「基本統計量」で「統計情報」を求める。

データタブの（分析グループに）[データ分析] ボタンが見当たらないときは、ファイルタブの（「その他...」の）「オプション」から「アドイン」を選択する。下のほうにある「Excel アドイン」の右にある「設定...」をクリックし、「分析ツール」にチェックを入れる。また、チェックを入れても [データ分析] ボタンが出てこない場合は、ファイルを閉じて、もう一度開く。

基本統計量

入力元
入力範囲(I): \$A\$3:\$M\$3

データ方向:
☐ 列(C)
☒ 行(B)

☒ 先頭列をラベルとして使用(L)

出力オプション
☒ 出力先(Q): \$W\$1
☐ 新規ワークシート(P):
☐ 新規ブック(W)
☒ 統計情報(S)
☐ 平均の信頼度の出力(N) 95 %
☐ K 番目に大きな値(A): 1
☐ K 番目に小さな値(M): 1

OK
キャンセル
ヘルプ(H)

B店	
平均	571642.5
標準誤差	12149.16895
中央値 (メジアン)	577485
最頻値 (モード)	597780
標準偏差	42085.95579
分散	1771227675
尖度	-1.191678907
歪度	-0.329353835
範囲	121770
最小	509220
最大	630990
合計	6859710
データの個数	12

これより、B 店について、Excel アドインのデータ分析ツールの「基本統計量」で求められる「標準偏差」は約 42086.0（円）であることがわかる。

- (6) A 店の平均値をセル N2 に求めているとすると、1 月の偏差は「=B2-\$N\$2」（または「=B2-\$N2」）で求められ、これを右にオートフィルすれば 12 月までの偏差を求めることができる（スピルを使って求めるなら「=B2:M2-N2」と入力すればいい）。

A店についての月ごとの売上金額の偏差											
1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
70417.5	188497.5	-73492.5	40897.5	-150983	-114083	-84562.5	18757.5	29827.5	77797.5	-40282.5	37207.5

これより、偏差の絶対値が最も小さいのは 8 月（偏差：18757.5）である。

- (7) 各月の平均値を AVERAGE 関数で求める。

	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
A店	918810	1036890	774900	889290	697410	734310	763830	867150	878220	926190	808110	885600
B店	549810	597780	516600	512910	597780	509220	605160	568260	586710	619920	630990	564570
C店	298890	258300	225090	335790	372690	328410	335790	339480	376380	254610	343170	335790
D店	771210	800730	889290	701100	889290	808110	819180	660510	778590	848700	730620	671580
E店	217710	221400	302580	339480	446490	512910	483390	586710	697410	741690	907740	1055340
平均値	551286	583020	541692	555714	600732	578592	601470	604422	663462	678222	684126	702576

これより、平均値が最も大きい月は 12 月（平均値：702576）であることがわかる。

問題 15.12

学科ごとに AVERAGE 関数で平均値、STDEV.P 関数で標準偏差を求める（理学部数学科の来場者数の平均値をセル B9 に、標準偏差をセル B10 に求めて、右にオートフィルする）。

次に、（セル I3 に）「=STANDARDIZE(」と入力し、理学部数学科の 1 回目のもともとの来場者数（B3）をクリックして指定する。続けて、「,」を入力し、理学部数学科の来場者数の平均値が計算されているセル（B9）をクリックし F4 キーを 2 回押す。さらに、「,」を入力し、理学部数学科の来場者数の標準偏差が計算されているセル（B10）をクリックし F4 キーを 2 回押す。Enter キーを押すと理学部数学科の 1 回目の来場者数を標準化した値が計算される（セル I3 には「=STANDARDIZE(B3,B\$9,B\$10)」と入力される。これは、下にオートフィルするときに「平均」と「標準偏差」の行番号を固定するために、その前に「\$」を付けているのである。一方、あとで右にオートフィルするために、列番号 B の前には「\$」を付けてはいけな

I3															
----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

これより、最も大きい標準化後の値は約 1.451 で、理学部数理学科の 5 回目の来場者数を標準化したものである。また、最も小さい標準化後の値は約-1.983 で、工学部電気・機械工学科の 4 回目の来場者数を標準化したものである（なお、標準化する際にスピルを使うなら「=STANDARDIZE(B3:G7,B9:G9,B10:G10)」と入力すればいい）。

問題 15.13

- (1) 身長と体重についてそれぞれ、AVERAGE 関数で平均値、STDEV.P 関数で標準偏差を求める（身長の平均値をセル B11 に、標準偏差をセル B12 に求めて、右にオートフィルする）。

次に、（セル E2 に）「=STANDARDIZE(」と入力し、番号 1 の人のもともとの身長（B2）をクリックして指定する。続けて、「,」を入力し、身長の平均値が計算されているセル（B11）をクリックし F4 キーを 2 回押す。さらに、「,」を入力し、身長の標準偏差が計算されているセル（B12）をクリックし F4 キーを 2 回押す。Enter キーを押すと番号 1 の人の身長を標準化した値が計算される（セル E2 には「=STANDARDIZE(B2,B\$11,B\$12)」と入力される）。下にオートフィルして、さらに右にもオートフィルしてデータ全部を標準化する。

E2 : ✕ ✓ fx =STANDARDIZE(B2,B\$11,B\$12)						
	A	B	C	D	E	F
1	番号	身長 (cm)	体重 (kg)		身長 (標準化後)	体重 (標準化後)
2	1	163	54		-1.241081931	-0.850125793
3	2	170	65		-0.171842114	0.218603775
4	3	180	66		1.35564334	0.315761009
5	4	175	75		0.591900613	1.190176111
6	5	161	49		-1.546579022	-1.335911961
7	6	178	80		1.05014625	1.675962278
8	7	175	61		0.591900613	-0.170025159
9	8	167	52		-0.63008775	-1.04444026
10						
11	平均値	171.125	62.75			
12	標準偏差	6.546707188	10.29259443			
13						
14						

これより、標準化後の体重の値のほうが標準化後の身長の値より大きい人の番号は 1, 2, 4, 5, 6 であることがわかる（なお、標準化する際にスピルを使うなら「=STANDARDIZE(B2:C9,B11:C11,B12:C12)」と入力すればいい）。

- (2) 8 つのデータの中央値は 4（= 8/2）番目の大きさのデータと 5（= 8/2 + 1）番目の大きさのデータの平均値になる。

体重について大きさの順になるようにデータの並べ替えをする。C 列の体重が入力されているどこかのセルを選択した状態で、ホームタブの（編集グループにある）「並べ替えとフィルター」をクリックし、「昇順」または「降順」を選択する。

番号	身長 (cm)	体重 (kg)
5	161	49
8	167	52
1	163	54
7	175	61
2	170	65
3	180	66
4	175	75
6	178	80

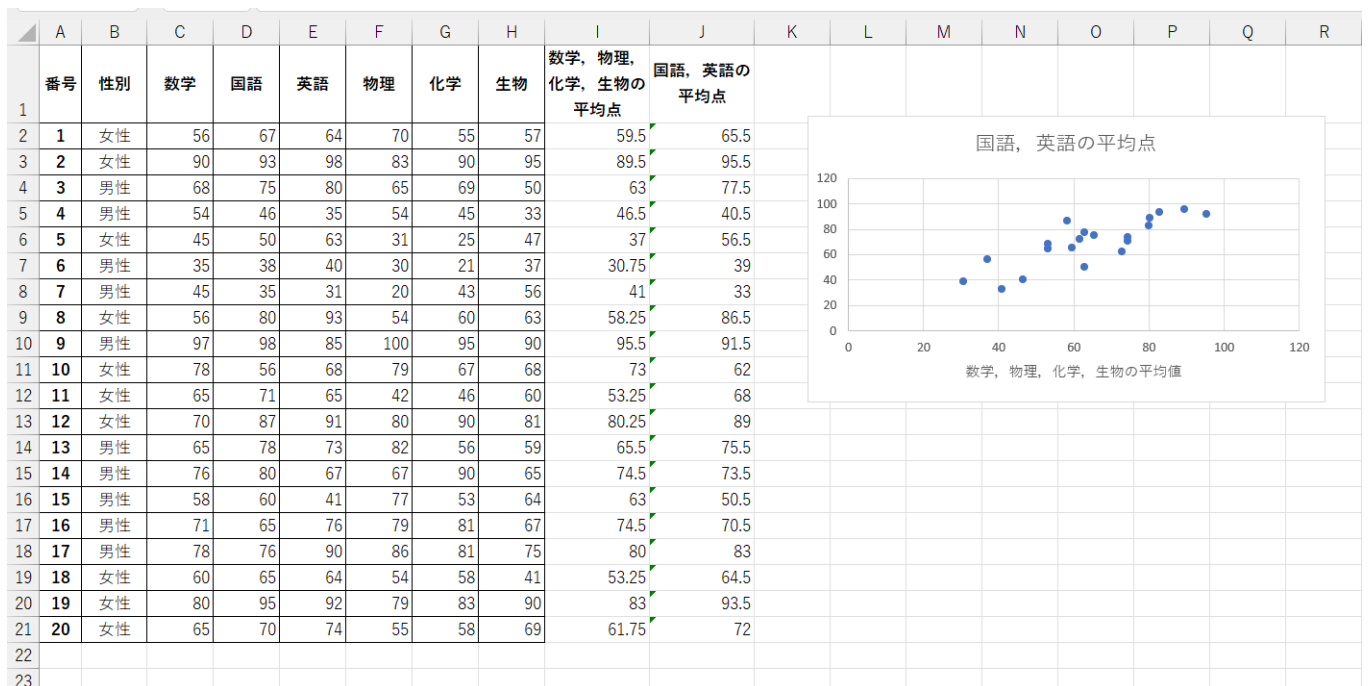
そうすると、4 番目の大きさのデータの番号は 7 であり、5 番目の大きさのデータの番号は 2 であることがわかる。よって、中央値は「番号 7 と番号 2 の平均値」と等しくなる。

問題 15.14

番号 1 についての「数学、物理、化学、生物の平均点」をセル I2 に、「国語、英語の平均点」をセル J2 に、AVERAGE 関数でそれぞれ求める（セル I2 には「=AVERAGE(C2,F2:H2)」，セル J2 には「=AVERAGE(D2:E2)」と入力される）。それらを（21 行目まで）下にオートフィルする。

個々の学生別の「数学、物理、化学、生物の平均点」と「国語、英語の平均点」の 2 列分（I2:J21）を選択して、挿入タブの（グラフグループにある）[散布図 (x, y) またはバブルチャートの挿入] の「散布図」を選ぶ。横軸が「数学、物理、化学、生物の平均点」（I 列）で縦軸が「国語、英語の平均点」（J 列）である（散布図を作成する際、選択範囲の 2 列のうち、左側の列が横軸になり、右側の列が縦軸になる）。

作成したグラフを選択した状態で、（グラフの）デザインタブの（グラフのレイアウトグループにある）[グラフ要素の追加] をクリックし、「軸ラベル」の「第 1 横軸」を選択すると（横）軸ラベルが出てくる。（横）軸ラベルに「数学、物理、化学、生物の平均点」と入力する。



問題 15.15

セル A1 に「価格」、A2 に「仕入れ値」、A3 に「売上個数」、A4 に「粗利」、A5 に「利益」と入力する。セル B1（価格）に「1000」、セル B2（仕入れ値）に「1000」を入力する。セル B3（売上個数）には

$$[-0.88 \times [\text{価格のセル}] + 2222.22], \text{つまり}, [= -0.88 * B1 + 2222.22]$$

を入力する。セル B4（粗利）には

$$[\text{価格のセル}] - [\text{仕入れ値のセル}], \text{つまり}, [= B1 - B2]$$

を入力する。セル B5（利益）には

$$[\text{売上個数のセル}] \times [\text{粗利のセル}], \text{つまり}, [= B3 * B4]$$

を入力する。

価格	1000
仕入れ値	1000
売上個数	1342.22
粗利	0
利益	0

データタブの（分析グループにある）[ソルバー] を選択し、「ソルバーのパラメーター」ダイアログボックスを出す。

[ソルバー] ボタンが見当たらないときは、ファイルタブの（「その他...」の）「オプション」から「アドイン」を選択する。下のほうにある「Excel アドイン」の右にある「設定...」をクリックし、「ソルバーアドイン」にチェックを入れる。また、チェックを入れても [ソルバー] ボタンが出てこない場合は、ファイルを閉じて、もう一度開く。

「目的セルの設定」には利益のセル B5 を指定する。

「目標値」は「最大値」を選択する。

「変数セルの変更」には価格のセル B1 を指定する。

「制約条件の対象」には制約条件を入れる。ここでは、原因となる価格を整数のみで考えたいので、「価格のセル B1 は整数しかとらない」という制約条件を入れる。「制約条件の対象」の右にある「追加」を選択し、「セル参照」に B1 を指定し、真ん中を「int」として OK ボタンを押す。

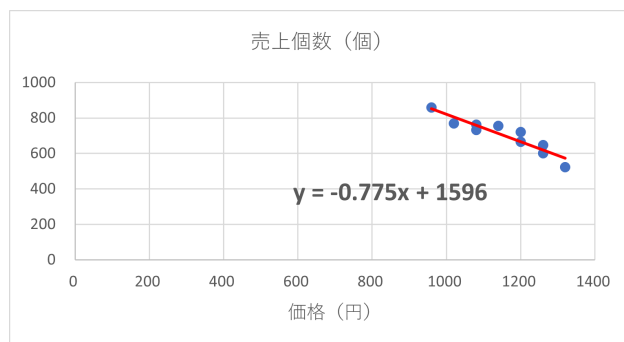
そして、「ソルバーのパラメーター」ダイアログボックスの「解決」を押すと、計算され、「ソルバーの結果」ダイアログボックスが出てくる。OK ボタンを押すと、最大になった予測利益は「約 511805（円）」であり、それを与える販売価格は「1763（円）」であることがわかる。

価格	1763
仕入れ値	1000
売上個数	670.78
粗利	763
利益	511805

問題 15.16

回帰式は下記のようになる。

$$\text{売上個数} = -0.775 \times \text{価格} + 1596$$



データタブの（分析グループにある）[ソルバー] を使って最適化の結果の値を求める。

	A	B	C	D	E	F	G
1		価格 (円)	売上個数 (個)		価格	770	
2	1日目	1200	720		仕入れ値	770	
3	2日目	1080	762		売上個数	=-0.775*F1+1596	
4	3日目	1260	600		粗利	=F1-F2	
5	4日目	1080	732		利益	=F3*F4	
6	5日目	1200	666				
7	6日目	960	858				
8	7日目	1020	768				
9	8日目	1140	756				
10	9日目	1260	648				
11	10日目	1320	522				
12							
13							

最大になった予測利益は、百の位を四捨五入すると「約 322000 (円)」であることがわかる。

価格	1415
仕入れ値	770
売上個数	499.375
粗利	645
利益	322097

問題 15.17

- (1) 分析ツールの「移動平均」を選び、「入力範囲」は売上高合計のデータ (C2:C25) とする。「先頭行をラベルとして使用」にはチェックをしない（「入力範囲」を C1:C25 とし、項目名（セル C1「売上高合計（百万円）」）を「入力範囲」に含めた場合は、「先頭行をラベルとして使用」にチェックをする）。「区間」は「12」,「出力先」は売上高合計データのとなり (D2) にする。

2021 年 11 月の移動平均値は約 45811 (百万円) であることがわかる。

- (2) 季節変動値は「売上高合計 ÷ 移動平均」で求められる。セル E24 (2021 年 11 月の移動平均値のとなり) に「=C24/D24」と計算式を入れる。

2021 年 11 月の季節変動値は約 1.000 であることがわかる。

E24				=C24/D24		
	A	B	C	D	E	F
1	年	月	売上高合計 (百万円)			
2	2020年	1月	45,442	#N/A		
3		2月	30,474	#N/A		
4		3月	34,218	#N/A		
5		4月	27,852	#N/A		
6		5月	21,588	#N/A		
7		6月	27,534	#N/A		
8		7月	42,743	#N/A		
9		8月	55,122	#N/A		
10		9月	41,649	#N/A		
11		10月	43,366	#N/A		
12		11月	42,573	#N/A		
13		12月	57,733	39,191		
14	2021年	1月	51,877	39,727		
15		2月	36,079	40,195		
16		3月	46,612	41,227		
17		4月	40,371	42,271		
18		5月	34,618	43,356		
19		6月	38,064	44,234		
20		7月	49,456	44,793		
21		8月	61,038	45,286		
22		9月	44,340	45,511		
23		10月	43,744	45,542		
24		11月	45,804	45,811	1.000	
25		12月	59,714	45,976		
26						
27						

問題 15.18

- (1) 月ごとのトリム平均は TRIMMEAN 関数で求める。ここで、TRIMMEAN 関数の「配列」にはトリム平均をとるデータ全部を指定して、「割合」には取り除く両端のデータの個数がデータ全部の個数に占める割合を入力する。この問題の場合、取り除くのは両端の2つのデータであり、それがデータ全部の個数6に占める割合は「2/6」である。つまり、セル B11 を「=TRIMMEAN(B3:B8,2/6)」で求め、これを右にオートフィルすればいい。

1月のトリム平均は1.09であることがわかる。

- (2) 月ごとのトリム平均の合計をセル N11 に SUM 関数「=SUM(B11:M11)」で求める。

そして、補正值(セル B13)を「12÷[トリム平均の合計値](=12/N11)」とし、各月の補正トリム平均を「トリム平均×補正值」で求める。セル B12 に「=B11*\$B\$13」(または「=B11\$B13」)と入力すると、「\$」をつけた列番号 B を固定してオートフィルすることができる(それらの合計値を SUM 関数でセル N12 に求めると 12 になるはずである)。

1月の補正トリム平均は約 1.08 であることがわかる。

	季節変動値												
	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月	
2016年	1.10	1.02	0.97	0.99	0.94	0.77	0.75	0.79	0.90	1.18	1.20	1.37	
2017年	1.00	1.05	0.96	0.99	1.00	0.80	0.82	0.80	0.91	1.19	1.16	1.43	
2018年	1.14	1.08	0.96	0.95	0.98	0.84	0.77	0.81	0.99	1.18	1.19	1.34	
2019年	1.00	1.04	0.95	0.92	0.98	0.85	0.82	0.81	0.95	1.16	1.12	1.28	
2020年	1.18	1.13	0.92	0.91	0.91	0.87	0.80	0.77	0.91	1.13	1.31	1.37	
2021年	1.12	1.09	0.91	0.98	1.01	0.88	0.78	0.78	0.98	1.09	1.28	1.36	
最大値	1.18	1.13	0.97	0.99	1.01	0.88	0.82	0.81	0.99	1.19	1.31	1.43	
最小値	1.00	1.02	0.91	0.91	0.91	0.77	0.75	0.77	0.90	1.09	1.12	1.28	合計値
トリム平均	1.09	1.065	0.9475	0.96	0.975	0.84	0.791307	0.795	0.9375	1.1625	1.2075	1.36	12.13131
補正トリム平均	1.078202	1.053473	0.937244	0.949609	0.964447	0.830908	0.782742	0.786395	0.927353	1.149917	1.19443	1.34528	12
補正值	0.989176												

(3) 12月の補正トリム平均（季節指数）が一番大きいことがわかる。

よって、翌年（2022年）のデータを予測するとき、求めた補正トリム平均（季節指数）のみの影響を受けると仮定すると、最も値が大きくなるのは12月であると予想できる（12月の補正トリム平均（季節指数）は約1.35なので、この月ではふつうは移動平均より大きくなる（約1.35倍になる）ということである）。

問題 15.19

(1) 季節指数は各月の合計が12になるような値なので、2月の季節指数は12から「2月以外の合計」をひいて求める（SUM関数の範囲にE4とM4:P4を指定するときは、E4をクリックして選択したあと、Ctrlキーを押しながらG4:P4をドラッグして選択する。または、E4をクリックして選択したあと、「,」を入力し、そのあと、G4:P4をドラッグして選択する）。0.81となる。

季節指数（補正トリム平均）											
1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
1.17	=12-SUM(E4,G4:P4)	0.95	0.89	0.77	0.82	1.09	1.34	0.99	0.88	0.91	1.38

(2) 季節調整済み売上高合計は、「売上高合計 ÷ 季節指数」で求められる。2009年2月の季節調整済み売上高合計は「2009年2月の売上高合計 ÷ 2月の季節指数」（「=C15/F4」）で求められ、約31,286（百万円）となる。

D15																		
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
1	年	月	売上高合計（百万円）															
2	2008年	1月	35,489		季節指数（補正トリム平均）													
3		2月	25,251		1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月		
4		3月	28,808		1.17	0.81	0.95	0.89	0.77	0.82	1.09	1.34	0.99	0.88	0.91	1.38		
5		4月	26,926															
6		5月	23,349															
7		6月	24,937															
8		7月	33,677															
9		8月	41,077															
10		9月	30,057															
11		10月	27,233															
12		11月	28,145															
13		12月	42,030															
14	2009年	1月	35,811															
15		2月	25,342	31,286														
16		3月	28,401															
17		4月	26,734															
18		5月	22,674															
19		6月	25,944															
20		7月	33,296															
21		8月	41,414															
22	9月	29,399																
23	10月	27,228																
24	11月	28,252																
25	12月	43,985																
26																		
27																		

(3) 5月の季節指数が一番小さいことがわかる。

よって、翌年（2010年）の売上高合計を予測するとき、季節指数のみの影響を受けると仮定すると、最も売上高合計が小さくなるのは5月であると予想できる（5月の季節指数は約0.77なので、この月ではふつうは移動平均より小さくなる（約0.77倍になる）ということである）。

問題 15.20

最後の階級の累積度数が全度数である。「相対度数 = 度数 ÷ 全度数」より、「度数 = 相対度数 × 全度数」であるので、度数を求めることができる。また、最初の階級の度数は累積度数そのものであり、最初の階級の累積相対度数は相対度数そのものである。

下記のように計算式を入れる（一般に、最後の階級の累積相対度数は1なので、「1」と入力してもいい）。

階級	度数	相対度数	累積度数	累積相対度数
0～10	=D2	=B2/D9	3	=C2
10～20	=C3*D9	=E3-E2	=D2+B3	0.1
20～30	=C4*D9	0.18	=D3+B4	=E3+C4
30～40	=D5-D4	=B5/D9	34	=E4+C5
40～50	13	=B6/D9	=D5+B6	=E5+C6
60～70	=C7*D9	=E7-E6	=D6+B7	0.64
80～90	=D8-D7	=B8/D9	93	=E7+C8
90～100	7	=B9/D9	=D8+B9	=E8+C9

下記のような結果が得られる。

階級	度数	相対度数	累積度数	累積相対度数
0～10	3	0.03	3	0.03
10～20	7	0.07	10	0.1
20～30	18	0.18	28	0.28
30～40	6	0.06	34	0.34
40～50	13	0.13	47	0.47
60～70	17	0.17	64	0.64
80～90	29	0.29	93	0.93
90～100	7	0.07	100	1

問題 15.21

- (1) 実家までの帰省時間の最頻値をエクセル関数（MODE.MULT 関数、または、MODE.SNGL 関数）で求める。実家までの帰省時間の最頻値は5（時間）であることがわかる。
- (2) 実家までの帰省時間と帰省回数について、それぞれ平均値をエクセル関数（AVERAGE 関数）で求める。実家までの帰省時間と帰省回数の平均値はそれぞれ約 4.33（時間）と約 16.07（回）である。
- (3) 実家までの帰省時間と帰省回数について、それぞれ標準偏差をエクセル関数（STDEV.P 関数）で求める。実家までの帰省時間と帰省回数の標準偏差はそれぞれ約 2.02（時間）と約 15.00（回）である。
- (4) 実家までの帰省時間と帰省回数について、それぞれトリム平均をエクセル関数（TRIMMEAN 関数）で求める。ここで、TRIMMEAN 関数の「配列」にはトリム平均をとるデータ全部を指定して、「割合」には取り除く両端のデータの個数がデータ全部の個数に占める割合を入力する。この問題の場合、取り除くのは両端の2つのデータであり、それがデータ全部の個数 30 に占める割合は「2/30」である。つまり、実家までの帰省時間の最大値と最小値を取り除いたトリム平均は「=TRIMMEAN(B2:B31,2/30)」で求められ、約 4.32 であることがわかる。これを右にオートフィルすると、帰省回数の最大値と最小値を取り除いたトリム平均が求められ、15.25 であることがわかる（セルには「=TRIMMEAN(C2:C31,2/30)」と入力される）。

番号	実家までの帰省時間（時間）	帰省回数（回）
最頻値	5	
平均値	4.33	16.07
標準偏差	2.02	15.00
トリム平均	4.32	15.25

- (5) 分析ツールの「相関」を選び、「入力範囲」は実家までの帰省時間と帰省回数のデータ（B1:C31）、「データ方向」を「列」とし、「先頭行をラベルとして使用する」にチェックする。

相関

入力元
 入力範囲(I):

データ方向: ☒ 列(C) ☐ 行(B)

☒ 先頭行をラベルとして使用(L)

出力オプション
☒ 出力先(O):

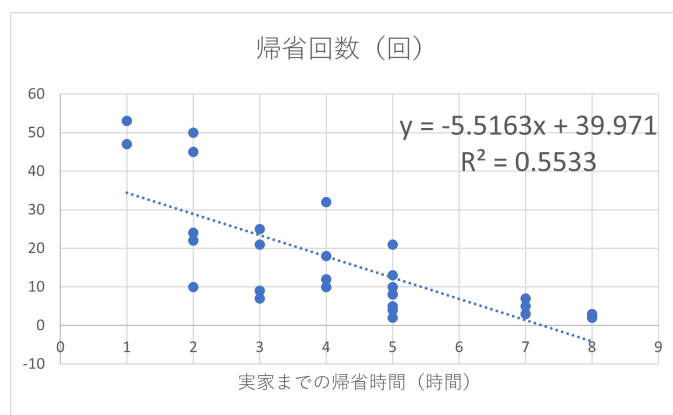
☐ 新規ワークシート(P):

☐ 新規ブック(W)

	実家までの帰省時間（時間）	帰省回数（回）
実家までの帰省時間（時間）	1	
帰省回数（回）	-0.74386243	1

実家までの帰省時間と帰省回数の相関係数は約 -0.744 であることがわかる（CORREL 関数で求めている）。

- (6) 実家までの帰省時間と帰省回数の2列分 (B1:C31) を選択して、挿入タブの「散布図 (x, y) またはバブルチャートの挿入」の「散布図」を選ぶ。そして、散布図のマーカー (点) の上で右クリックして「近似曲線の追加」を選択する。近似曲線の書式設定において、「グラフに数式を表示する」と「グラフに R-2 乗値を表示する」にチェックを入れる。



回帰式は

$$y = -5.5163x + 39.971$$

であり、R-2 乗値は約 0.5533 であることがわかる。

- (7) 分析ツールの「回帰分析」を選び、「入力 Y 範囲」は結果系のデータ (この場合は「帰省回数」、C1:C31)、「入力 X 範囲」は原因系のデータ (この場合は「実家までの帰省時間」、B1:B31)、「ラベル」にチェックする。「残差」にもチェックを入れる。

観測値	予測値: 帰省回数 (回)	残差
1	12.38913043	-10.3891
2	28.93804348	-18.938
3	28.93804348	-6.93804
4	23.42173913	-16.4217
5	34.45434783	12.54565
6	23.42173913	1.578261
7	28.93804348	21.06196
8	12.38913043	-2.38913
9	1.356521739	1.643478
10	12.38913043	-8.38913
11	17.90543478	14.09457
12	-4.159782609	7.159783
13	17.90543478	0.094565
14	12.38913043	-7.38913
15	12.38913043	0.61087
16	28.93804348	-4.93804
17	1.356521739	5.643478
18	12.38913043	-4.38913
19	-4.159782609	6.159783
20	23.42173913	-14.4217
21	12.38913043	8.61087
22	17.90543478	-7.90543
23	17.90543478	-5.90543
24	28.93804348	16.06196
25	1.356521739	3.643478
26	12.38913043	-8.38913
27	12.38913043	-4.38913
28	23.42173913	-2.42174
29	-4.159782609	6.159783
30	34.45434783	18.54565

- すると、残差も求められ、残差の絶対値が最も大きくなる番号は7であることがわかる（その残差は約 21.062 である）。7 番の学生の帰省回数の実測値は 50（回）であり、予測値は約 28.938（回）である。
- (8) 上記 (7) で行った、実家までの帰省時間と帰省回数についての回帰分析の結果、残差の絶対値が最も小さくなる番号は 13 番であることがわかる（その残差は約 0.095 である）。13 番の学生の帰省回数の実測値は 18（回）であり、予測値は約 17.905（回）である。
- (9) 上記 (7) で行った、実家までの帰省時間と帰省回数についての回帰分析の結果、回帰式の傾き（約 -5.516）と切片（約 39.971）が計算されている。

	係数
切片	39.97065217
実家までの帰省時間（時間）	-5.51630435

これより、帰省時間（ x ）と帰省回数（ y ）についての回帰式は

$$y = -5.5163x + 39.971 \quad (\text{傾きと切片は概数})$$

であることがわかる（(6) からでもわかる）。よって、[実家までの帰省時間が 6 時間のときの 1 年間の帰省回数の予測値] は、空いているセルに「= [傾きが入力されたセル番地] *6+ [切片が入力されたセル番地]」と入力すると求められる。約 6.873（回）となることがわかる。

(10) 帰省時間 (x) と帰省回数 (y) についての回帰式は

$$y = -5.5163x + 39.971 \quad (\text{傾きと切片は概数})$$

である。一般に、直線の方程式「 $y = ax + b$ (a, b は定数)」において、 x が 1 増えると y は傾き a の分だけ増える。つまり、実家までの帰省時間が 1 時間増えると、帰省回数は約 -5.5163 (回) だけ増える、つまり、約 5.5163 (回) だけ減ると予測されることになる。

(11) 帰省回数について大きさの順になるようにデータの並べ替えをする。C 列の帰省回数が入力されているどこかのセルを選択した状態で、ホームタブの(編集グループにある)「並べ替えとフィルター」をクリックし、「昇順」または「降順」を選択する。そうすると、「20 回以上 25 回未満」である帰省回数は、

21
21
22
24

の 4 つであることがわかる。よって、帰省回数が「20 回以上 25 回未満」の階級の度数は 4 である。

また、相対度数は度数を全度数でわったものなので、「 $=4/30$ 」で求められ、約 0.13 であることがわかる。

(12) 1 が正解である。

問題の選択肢の 4 つのヒストグラムはどれも階級が 0～5, 5～10, 10～15, 15～20, 25～30, 30～35, 35～40, 40～45, 45～50, 50～55 であることがわかる。よって、(12) により大きさの順に並べ替えられているデータを見て、この階級で度数だけ調べ、それに一致するヒストグラムを選べば正解はわかる。

なお、この階級で度数を調べると下記のようなになる。

階級	度数
0～5	7
5～10	7
10～15	5
15～20	1
20～25	4
25～30	1
30～35	1
35～40	0
40～45	0
45～50	2
50～55	2

問題 15.22

選択肢 4 のヒストグラムを見て各階級における度数を入力し、相対度数、累積度数、累積相対度数について下記のように計算式を入力する。相対度数については最初のセル (C2) を下にオートフィルする。累積度数、累積相対度数については 2 番目のセル (D3, E3) を下にオートフィルする。

	A	B	C	D	E	F
1	階級	度数	相対度数	累積度数	累積相対度数	
2	0～5	7	=B2/\$B\$13	=B2	=C2	
3	5～10	7	=B3/\$B\$13	=D2+B3	=E2+C3	
4	10～15	3	=B4/\$B\$13	=D3+B4	=E3+C4	
5	15～20	1	=B5/\$B\$13	=D4+B5	=E4+C5	
6	20～25	4	=B6/\$B\$13	=D5+B6	=E5+C6	
7	25～30	1	=B7/\$B\$13	=D6+B7	=E6+C7	
8	30～35	3	=B8/\$B\$13	=D7+B8	=E7+C8	
9	35～40	0	=B9/\$B\$13	=D8+B9	=E8+C9	
10	40～45	0	=B10/\$B\$13	=D9+B10	=E9+C10	
11	45～50	2	=B11/\$B\$13	=D10+B11	=E10+C11	
12	50～55	2	=B12/\$B\$13	=D11+B12	=E11+C12	
13	合計	=SUM(B2:B12)				
14						
15						

これより、選択肢 4 のヒストグラムを与える度数分布表は

	A	B	C	D	E	F
1	階級	度数	相対度数	累積度数	累積相対度数	
2	0～5	7	0.233333	7	0.233333333	
3	5～10	7	0.233333	14	0.466666667	
4	10～15	3	0.1	17	0.566666667	
5	15～20	1	0.033333	18	0.6	
6	20～25	4	0.133333	22	0.733333333	
7	25～30	1	0.033333	23	0.766666667	
8	30～35	3	0.1	26	0.866666667	
9	35～40	0	0	26	0.866666667	
10	40～45	0	0	26	0.866666667	
11	45～50	2	0.066667	28	0.933333333	
12	50～55	2	0.066667	30	1	
13	合計	30				
14						
15						

となる。15 以上 20 未満の階級で累積相対度数が 0.6 となっているので、この階級までのデータ（つまり 20 未満のデータ）が全体のちょうど 6 割であることがわかる。

問題 15.23

- (1) 表中のどこかのセルを選択した状態で、挿入タブの「ピボットテーブル」(の「テーブルまたは範囲から」)を選択する。「テーブル/範囲」に表全体が選択されていることを確認する。OK ボタンを押すと「ピボットテーブルのフィールド」が出てくる。

「性別」,「数学」にチェックを入れる。すると,「行」ボックスに「性別」,「値」ボックスに「合計/数学」が入る。ピボットテーブルの行(横)に性別が出てきて,男女ごとの「数学」の成績の合計が出てくる。

「値」ボックス内の「合計/数学」を「列」ボックスに移動させる。そして,(「ピボットテーブルのフィールド」の上部にある)ボックスにチェックの入った「性別」(という文字列)をドラッグして「値」ボックスに移動させる。すると,「値」ボックスに「個数/性別」が入る。

個数 / 性別	列ラベル						
行ラベル		1	2	3	4	5	総計
女性		2	4	5	2	2	15
男性		1	1	4	5	4	15
総計		3	5	9	7	6	30

すると、男女別の数学の成績ごとの人数が表示される。数学の成績が「3」である女性の人数は 5（人）であることがわかる。

(2) (1) の結果より、数学の成績が「4」である学生的人数は 7（人）であることがわかる。

(3) 「ピボットテーブルのフィールド」の「性別」, 「英語」にチェックを入れる（これ以外にチェックが入っていたら外す）（「ピボットテーブルのフィールド」が出ていない場合には、ピボットテーブルのどこかのセルが選択されている状態で、（ピボットテーブル）分析タブの（表示グループにある）[フィールドリスト] を選択する）。

「値」ボックス内の「合計／英語」を「列」ボックス内に移動させる。また、（「ピボットテーブルのフィールド」の上部にある）ボックスにチェックの入った「性別」（という文字列）をドラッグして「値」ボックスに移動させる。すると、「値」ボックスに「個数/性別」が入る。

個数 / 性別	列ラベル						
行ラベル		1	2	3	4	5	総計
女性		1	1	7	3	3	15
男性		2	4	4	4	1	15
総計		3	5	11	7	4	30

英語の成績が「3」である女性の人数は 7（人）であることがわかる。

(4) 「ピボットテーブルのフィールド」の「化学」にチェックを入れる（これ以外にチェックが入っていたら外す）。

「値」ボックス内が「個数／化学」になるように、値フィールドの設定の「選択したフィールドのデータ」を「個数」にする。

また、ボックスにチェックの入った「化学」（という文字列）をドラッグして「列」ボックスに移動させる（この結果より、化学の成績が「3」である学生の割合をわり算で求めてもいい）。

	列ラベル						
		1	2	3	4	5	総計
個数 / 化学		3	6	11	7	3	30

そして、「値」ボックス内の「個数／化学」をクリックし、「値フィールドの設定」を選び、「計算の種類」タブを選択する。「計算の種類」を「行集計に対する比率」に変更する。

	列ラベル						
		1	2	3	4	5	総計
個数 / 化学		10.00%	20.00%	36.67%	23.33%	10.00%	100.00%

化学の成績が「3」である学生の割合は約 36.67%であることがわかる。

問題 15.24

- (1) 表中のどこかのセルを選択した状態で、挿入タブの「ピボットテーブル」(の「テーブルまたは範囲から」)を選択する。「テーブル/範囲」に表全体が選択されていることを確認する。OK ボタンを押すと「ピボットテーブルのフィールド」が出てくる。

「性別」,「数学」にチェックを入れる。すると,「行」ボックスに「性別」,「値」ボックスに「合計/数学」が入る。

「値」ボックス内の「合計/数学」を「列」ボックスに移動させる。そして,(「ピボットテーブルのフィールド」の上部にある)ボックスにチェックの入った「性別」(という文字列)をドラッグして「値」ボックスに移動させる。すると,「値」ボックスに「個数/性別」が入る。

個数 / 性別	列ラベル					
行ラベル	1	2	3	4	5	総計
女性	2	4	5	2	2	15
男性	1	1	4	5	4	15
総計	3	5	9	7	6	30

すると,男女別の数学の成績ごとの人数が表示される。(ここまで問題 15. 22 (1) と同じ)

ピボットテーブルの数学の成績の項目(最上行の 1, 2, 3, 4, 5) が書かれたセルで右クリックし,「グループ化」を選択する。「グループ化」ダイアログボックスが出てくるので,先頭の値を「4」,末尾の値を「5」,単位を「2」にする。

グループ化 ? X

自動

☐ 先頭の値(S): 4
 ☐ 末尾の値(E): 5
 単位(B): 2

OK キャンセル

個数 / 性別	列ラベル		
行ラベル	<4	4-5	総計
女性	11	4	15
男性	6	9	15
総計	17	13	30

「値」ボックス内の「個数/数学」をクリックし,「値フィールドの設定」を選び,「計算の種類」タブを選択する。「計算の種類」を「列集計に対する比率」に変更する。

個数 / 性別	列ラベル			
行ラベル	<4	4-5	総計	
女性	64.71%	30.77%	50.00%	
男性	35.29%	69.23%	50.00%	
総計	100.00%	100.00%	100.00%	

すると,数学の成績が「4」または「5」である学生の男女構成比は女性 約 30.77%:男性 約 69.23%であることがわかる。

別解

問題 15.23(1) の結果から、数学の成績が「4」または「5」である学生は $(7 + 6)$ 人で、そのうち女性は $(2 + 2)$ 人、男性は $(5 + 4)$ 人であることがわかる。よって、数学の成績が「4」または「5」である学生の男女構成比を求めるために、女性については「 $=(2+2)/(7+6)$ 」、男性については「 $=(5+4)/(7+6)$ 」と入力する。これらをパーセントであらわすために 100 をかけるか、または、ホームタブの「パーセントスタイル」ボタン (%) をクリックすると、女性 約 30.77%：男性 約 69.23%であることがわかる。

- (2) 「ピボットテーブルのフィールド」の「性別」、「生物」にチェックが入っている状態にする（これら以外にチェックが入っていたら外す）。

「値」ボックス内の「合計／生物」を「列」ボックス内に移動させる。また、ボックスにチェックの入った「性別」（という文字列）をドラッグして「値」ボックスに移動させる。すると、「値」ボックスに「個数/性別」が入る。

ピボットテーブルの生物の成績の項目（最上行の 1, 2, 3, 4, 5）が書かれたセルで右クリックし、「グループ化」を選択する。「グループ化」ダイアログボックスが出てくるので、先頭の値を「3」、末尾の値を「4」、単位を「2」にする。

そして、「値」ボックス内の「個数／生物」をクリックし、「値フィールドの設定」を選び、「計算の種類」タブを選択する。「計算の種類」を「列集計に対する比率」に変更する。

個数 / 性別	列ラベル			
行ラベル	<3	3-4	>5	総計
女性	20.00%	58.82%	50.00%	50.00%
男性	80.00%	41.18%	50.00%	50.00%
総計	100.00%	100.00%	100.00%	100.00%

生物の成績が「3」または「4」である学生の男女構成比は女性 約 58.82%、男性 約 41.18%であることがわかる。

別解

上記の解答で、グループ化する前の状態から、

個数 / 性別	列ラベル					
行ラベル	1	2	3	4	5	総計
女性		1	6	4	4	15
男性		2	2	3	4	15
総計	2	3	9	8	8	30

生物の成績が「3」または「4」である学生は $(9 + 8)$ 人で、そのうち女性は $(6 + 4)$ 人、男性は $(3 + 4)$ 人であることがわかる。よって、女性の割合は $(6 + 4)/(9 + 8) = 0.58824 \dots$ で、男性の割合は $(3 + 4)/(9 + 8) = 0.41176 \dots$ である。それぞれ 100 をかけて、これらをパーセントであらわすと、女性 約 58.82%、男性 約 41.18%となる。

問題 15.25

データの範囲内のどこかのセルを選んでから、ホームタブの（編集グループにある）「並べ替えとフィルター」の「昇順」または「降順」を選択する。重さが 1367g 未満または 1634g 以上のものは出荷しないので、

出荷しないかぼちゃはこの中で3個であることがわかる。

かぼちゃの重さ (g)
1219
1341
1367
1421
1432
1459
1477
1493
1523
1543
1548
1548
1551
1572
1580
1587
1598
1611
1622
1629
1631
1633
1634

問題 15.26

- (1) 表中のどこかのセルを選択した状態で、挿入タブの「ピボットテーブル」(の「テーブルまたは範囲から」)を選択する。「テーブル/範囲」に表全体が選択されていることを確認する。OK ボタンを押すと「ピボットテーブルのフィールド」が出てくる。

「性別」,「大学の成績平均」にチェックを入れる。すると、「行」ボックスに「性別」,「値」ボックスに「合計/大学の成績平均」が入る。ピボットテーブルの行(横)に性別の項目が出てきて、それぞれの項目(男女)ごとの大学の成績平均の合計が出てくる。これらの値は「平均値」ではなく「合計」なので、「ピボットテーブルのフィールド」の「値」ボックス内の「合計/大学の成績平均」をクリックし、値フィールドの設定を選び、「選択したフィールドのデータ」を「平均」に変更する。

行ラベル	平均 / 大学の成績平均
女	2.669230769
男	2.835294118
総計	2.763333333

これより、女性についての「大学の成績平均」の平均値は約2.67であり、男性についての「大学の成績平均」の平均値は約2.84であることがわかる(また、全員についての「大学の成績平均」の平均値は約2.76であることもわかる)。

- (2) 「ピボットテーブルのフィールド」の「学部」,「学年」,「勉強時間(分)」にチェックを入れる(これら以外にチェックが入っていたら外す)。「ピボットテーブルのフィールド」が出ていない場合には、ピボットテーブルのどこかのセルが選択されている状態で、(ピボットテーブル) 分析タブの(表示グループにある)「フィールドリスト」を選択する)。

「行」ボックスに入った「学部」はそのまま、「値」ボックス内の「合計／学年」をドラッグして「列」ボックスへ移動させる。「値」ボックス内の「合計／勉強時間（分）」をクリックし、値フィールドの設定を選び、「選択したフィールドのデータ」を「平均」に変更する。すると、学部、学年別に勉強時間の平均値が表示される（これ以外の配置でもいい）。

平均 / 勉強時間（分）	列ラベル				
行ラベル		1	2	3	4 総計
経営学部		75	18.75	10	60
経済学部	28.33333333		20	20	65
法学部		45	30	60	135
総計		46.25	23.125	27.14285714	82.85714286

これより、勉強時間の平均値が最も大きいのは平均値 135（時間）の法学部の 4 年生であることがわかる。

- (3) 「ピボットテーブルのフィールド」の「学部」と「履修科目数」にチェックを入れる（これら以外にチェックが入っていたら外す）。

「値」ボックス内の「合計／履修科目数」をドラッグして「列」ボックスへ移動させる。さらに、ボックスにチェックの入った「学部」（という文字列）をドラッグして「値」ボックスに移動させる。

個数 / 学部		列ラベル									
行ラベル		4	5	6	7	8	9	10	11	12	総計
経営学部				1	2	4		2	3		12
経済学部			1	1			4	1		1	8
法学部		2				2	3	2		1	10
総計		2	1	2	2	6	7	5	3	2	30

これより、経営学部の履修科目数の最頻値は 8、経済学部の履修科目数の最頻値は 9、法学部の履修科目数の最頻値は 9 であることがわかる。

- (4) 「ピボットテーブルのフィールド」の「学部」と「履修科目数」にチェックを入れる。

「行」ボックスには「学部」が入り、「列」ボックスは空で、「値」ボックスには「平均／履修科目数」が入っている状態にする（もし「値」ボックス内にあるのが「合計／履修科目数」であればそれをクリックし、「値フィールドの設定」を選び、「選択したフィールドのデータ」を「平均」に変更する）。

行ラベル	平均 / 履修科目数
経営学部	8.75
経済学部	8.625
法学部	8.3
総計	8.56666667

これより、経営学部の履修科目数の平均値は 8.75、経済学部の履修科目数の平均値は 8.625、法学部の履修科目数の平均値は 8.3 であることがわかる。

- (5) 「ピボットテーブルのフィールド」の「性別」と「履修科目数」にチェックを入れる（これら以外にチェックが入っていたら外す）。

「値」ボックス内の「合計／履修科目数」（または「平均／履修科目数」）をドラッグして「列」ボックスへ移動させる。さらに、ボックスにチェックの入った「性別」（という文字列）をドラッグして「値」ボックスに移動させる。

次に、ピボットテーブルの履修科目数の項目（最上行の 4, 5, 6, 7, 8, 9, 10, 11, 12）が書かれたセルで右クリックし、「グループ化」を選択する。「グループ化」ダイアログボックスが出てくるので、先頭の値を「10」、末尾の値を「12」、単位を「3」にする。

さらに、「値」ボックス内の「個数/性別」をクリックし、「値フィールドの設定」を選び、「計算の種類」タブを選択する。「計算の種類」を「列集計に対する比率」に変更する。

個数 / 性別 列ラベル ▼			
行ラベル ▼	<10	10-12	総計
女	35.00%	60.00%	43.33%
男	65.00%	40.00%	56.67%
総計	100.00%	100.00%	100.00%

すると、履修科目数が 10 未満である学生の男女構成比は女性 35%：男性 65%であり、履修科目数が 10 以上である学生の男女構成比は女性 60%：男性 40%であることがわかる。

- (6) (5) の結果の状態において、「値」ボックス内の「個数/性別」をクリックし、「値フィールドの設定」を選び、「計算の種類」タブを選択する。「計算の種類」を「行集計に対する比率」に変更する。

個数 / 性別 列ラベル ▼			
行ラベル ▼	<10	10-12	総計
女	53.85%	46.15%	100.00%
男	76.47%	23.53%	100.00%
総計	66.67%	33.33%	100.00%

すると、男性のうち、履修科目数が 10 以上である人の割合は約 23.53%であることがわかる。

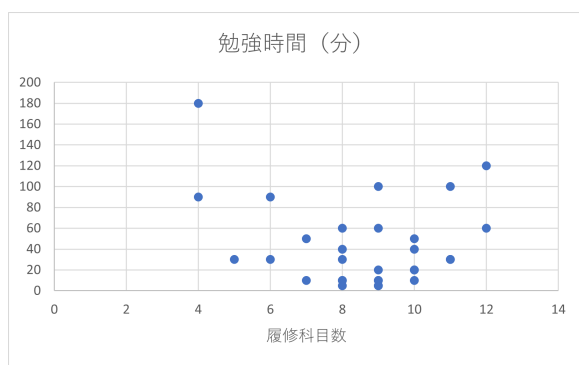
- (7) データタブの（分析グループにある）「データ分析」を選択する。分析ツールの「相関」を選び、「入力範囲」は「履修科目数」「通学時間（分）」「勉強時間（分）」「大学の成績平均」、および、「高校の評点平均」のデータ（E1:I31）、「データ方向」を「列」とし、「先頭行をラベルとして使用する」にチェックすればいい。小数第 3 位までの表示にすると下記ようになる。

	履修科目数	通学時間（分）	勉強時間（分）	大学の成績平均	高校の評定平均
履修科目数	1				
通学時間（分）	0.086	1			
勉強時間（分）	-0.211	-0.082	1		
大学の成績平均	0.050	-0.269	0.537	1	
高校の評定平均	0.013	-0.183	0.765	0.714	1

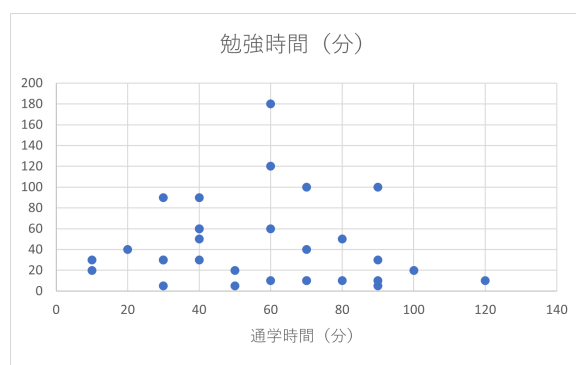
「履修科目数」と「勉強時間（分）」の散布図を作成するために、これらのデータ 2 列分（E1:E31 と G1:G31）を選択して、挿入タブの「散布図 (x, y) またはバブルチャートの挿入」の「散布図」を選ぶ（E1:E31 と G1:G31 を選択するときは、E1:E31 を選択したあと、Ctrl キーを押しながら G1:G31 を選択すればいい）。

「履修科目数」（E 列）が横軸、「勉強時間（分）」（G 列）が縦軸となる（散布図を作成する際、選択範囲の 2 列のうち、左側の列が横軸になり、右側の列が縦軸になる）。作成したグラフを選択した状態で、（グラフの）デザインタブの「グラフ要素の追加」をクリックし、「軸ラベル」の「第 1 横軸」を選択すると（横）軸ラベルが出てくる。（横）軸ラベルに「履修科目数」と入力する。

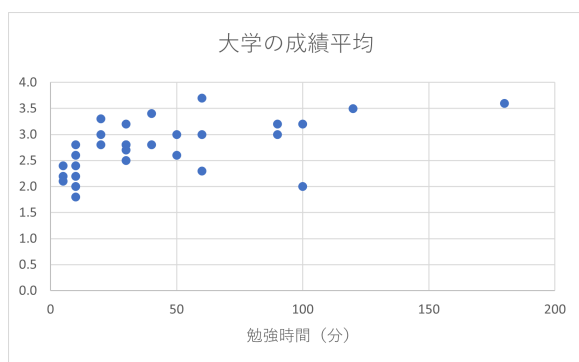
同様に、それぞれの散布図を作成し、相関係数と比較すると、下記ようになる。



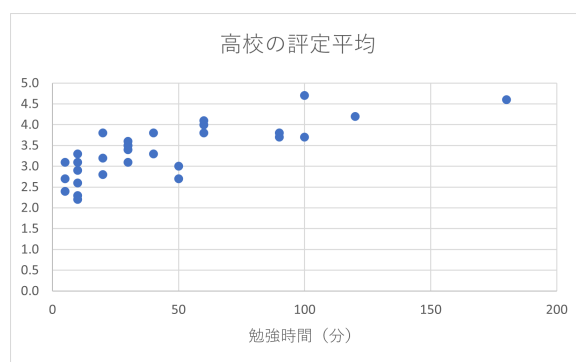
相関係数 0.086



相関係数 -0.082



相関係数 0.537



相関係数 0.765

(8) 分析ツールの「回帰分析」を選び、「入力 Y 範囲」は結果系のデータ（この場合は「大学の成績平均」、H1:H31）、「入力 X 範囲」は原因系のデータ（この場合は「勉強時間 (分)」, G1:G31）、「ラベル」にチェックする。

概要								
回帰統計								
重相関 R	0.537236							
重決定 R2	0.288623							
補正 R2	0.263217							
標準誤差	0.433295							
観測数	30							
分散分析表								
	自由度	変動	分散	観測された分散比	有意 F			
回帰	1	2.132827553	2.132827553	11.36028137	0.002203769			
残差	28	5.256839114	0.187744254					
合計	29	7.389666667						
	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	2.47445	0.116637049	21.21496124	0.000000000	2.235530316	2.713370645	2.235530316	2.713370645
勉強時間 (分)	0.006541	0.001940585	3.37050165	0.002203769	0.002565636	0.010515851	0.002565636	0.010515851

よって、回帰式の傾きは「勉強時間 (分)」の約 0.0065 であり、回帰式の切片は「切片」の約 2.4745 である。つまり、勉強時間 (x) と大学の成績平均 (y) についての回帰式は

$$y = 0.0065x - 2.4745 \quad (\text{傾きと切片は概数})$$

となることがわかる。一般に、直線の方程式「 $y = ax + b$ (a, b は定数)」において、 x が 1 増えると y は傾き a の分だけ増える。よって、勉強時間が 1 分増えたとき、大学の成績平均は約 0.0065 だけ増加すると予測される。つまり、勉強時間が 100 分増えたときは、大学の成績平均は約 0.65 だけ増加すると予測される。

(9) (8) での回帰分析の結果より、R-2 乗値は約 0.289 である。

(10) (8) での回帰分析の結果、勉強時間 (x) と大学の成績平均 (y) についての回帰式は下記である。

$$y = 0.0065x - 2.4745 \quad (\text{傾きと切片は概数})$$

	係数
切片	2.47445
勉強時間 (分)	0.006541

よって、[勉強時間が 150 分のときの大学の成績平均の予測値] は、空いているセルに「= [傾きが入力されたセル番地] *150+ [切片が入力されたセル番地]」と入力すると求められる。約 3.46 と計算される。

(11) 男女別に分かれるようにデータの並べ替えをする。そのため、B 列の性別が入力されているどこかのセルを選択した状態で、ホームタブの (編集グループにある) [並べ替えとフィルター] をクリックし、「昇順」または「降順」を選択する。男女別に、それぞれの相関係数を CORREL 関数で求めればよい。

男性のデータが上部にあるならば、男性についての相関係数は「=CORREL(G2:G18,H2:H18)」で求められ、女性についての相関係数は「=CORREL(G19:G31,H19:H31)」で求められる。男性については約 0.680 となり、女性については約 0.264 であることがわかる。

これより、男性のほうが「勉強時間 (分)」と「大学の成績平均」との正の相関が強いと考えられる。

別解

男女別に分かれるようにデータの並べ替えをしたあと、分析ツールを使って男女それぞれについての相関係数を別々に求めてもいい。つまり、分析ツールの「相関」を選び、男性のデータが上部にあるならば、「入力範囲」を男性の「勉強時間 (分)」と「大学の成績平均」のデータ (G1:H18)、「データ方向」を「列」とし、「先頭行をラベルとして使用する」にチェックして男性についての相関係数をまず求める。

	勉強時間 (分)	大学の成績平均
勉強時間 (分)	1	
大学の成績平均	0.680340525	1

次に、「入力範囲」を女性の「勉強時間 (分)」と「大学の成績平均」のデータ (G19:H31)、「データ方向」を「列」とし、「先頭行をラベルとして使用する」のチェックを外して女性についての相関係数を求めればよい。

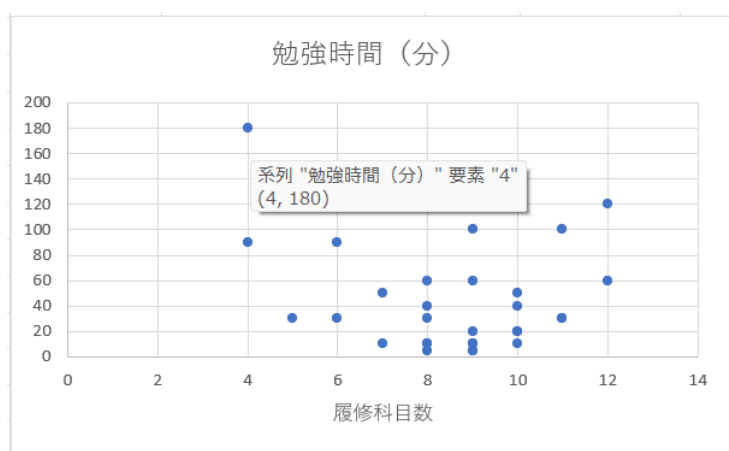
	列 1	列 2
列 1	1	
列 2	0.264493507	1

(12) 分析ツールの「相関」を選び、「入力範囲」は「学年」、「履修科目数」、「通学時間 (分)」、「勉強時間 (分)」、「大学の成績平均」、「高校の評点平均」のデータ (D1:I31)、「データ方向」を「列」とし、「先頭行をラベルとして使用する」にチェックすればよい。

	学年	履修科目数	通学時間 (分)	勉強時間 (分)	大学の成績平均	高校の評定平均
学年	1					
履修科目数	-0.730084518	1				
通学時間 (分)	-0.159616222	0.086455239	1			
勉強時間 (分)	0.297341005	-0.21065878	-0.081684489	1		
大学の成績平均	-0.163887639	0.049654045	-0.269031008	0.537236421	1	
高校の評定平均	0.017915493	0.013482308	-0.18335276	0.765125162	0.714424127	1

「学年」との間の相関係数が最も -1 に近いのは「履修科目数」(相関係数は約 -0.73) であることがわかる。よって、学年が上がるほど減少する傾向が最も強いのは、「履修科目数」であると考えられる。

- (13) 「履修科目数」と「勉強時間 (分)」の散布図において、1 点だけ上に飛び出しているのが見える。その点にマウスポインタを近づけ確認すると、履修科目数 4、勉強時間 180 (分) のデータであり、それは 30 番のデータであることがわかる。よって、外れ値ではないかと思われるデータの番号は 30 である。



- (14) 「履修科目数」について大きさの順になるようにデータの並べ替えをする。E 列の帰省回数が入力されているどこかのセルを選択した状態で、ホームタブの(編集グループにある)[並べ替えとフィルター]をクリックし、「昇順」または「降順」を選択する。そうすると、指定された各階級に属するデータの個数、つまり、度数を調べやすくなる。

度数を入力したら、相対度数、累積度数、累積相対度数について下記のように計算式を入力する。相対度数については最初のセル (C2) を下にオートフィルする。累積度数、累積相対度数については 2 番目のセル (D3, E3) を下にオートフィルする。

	A	B	C	D	E	F
1	階級	度数	相対度数	累積度数	累積相対度数	
2	4~6	3	=B2/\$B\$7	=B2	=C2	
3	6~8	4	=B3/\$B\$7	=D2+B3	=E2+C3	
4	8~10	13	=B4/\$B\$7	=D3+B4	=E3+C4	
5	10~12	8	=B5/\$B\$7	=D4+B5	=E4+C5	
6	12~14	2	=B6/\$B\$7	=D5+B6	=E5+C6	
7	合計	=SUM(B2:B6)				
8						
9						

これより、求める度数分布表は次のようになる。

	A	B	C	D	E	F
1	階級	度数	相対度数	累積度数	累積相対度数	
2	4～6	3	0.1	3	0.1	
3	6～8	4	0.13333	7	0.233333333	
4	8～10	13	0.43333	20	0.666666667	
5	10～12	8	0.26667	28	0.933333333	
6	12～14	2	0.06667	30	1	
7	合計	30				
8						
9						

(15) レンジは、MAX 関数で求めた最大値から MIN 関数で求めた最小値をひいて求める。

「大学の成績平均」と「高校の評定平均」のレンジはそれぞれ 1.9 と 2.5 であることがわかる。

(16) 「大学の成績平均」と「高校の評定平均」のそれぞれについて、AVERAGE 関数で平均値、STDEV.P 関数で標準偏差をそれぞれ求める（「大学の成績平均」の平均値をセル H37 に、標準偏差をセル H38 にそれぞれ求めて、右にオートフィルする）。

次に、（セル K2 に）「=STANDARDIZE(」と入力し、1 番の学生のもとの「大学の成績平均」（H2）をクリックして指定する。続けて、「,」を入力し、「大学の成績平均」の平均値が計算されているセル（H37）をクリックし F4 キーを 2 回押す。さらに、「,」を入力し、「大学の成績平均」の標準偏差が計算されているセル（H38）をクリックし F4 キーを 2 回押す。Enter キーを押すと 1 番の学生の「大学の成績平均」を標準化した値が計算される（セル K2 には「=STANDARDIZE(H2,H\$37,H\$38)」と入力される）。下にオートフィルして、さらに右にもオートフィルしてデータ全部を標準化する。

K2		fx		=STANDARDIZE(H2,H\$37,H\$38)									
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	番号	性別	学部	学年	履修科目数	通学時間（分）	勉強時間（分）	大学の成績平均	高校の評定平均		大学の成績平均（標準化後）	高校の評定平均（標準化後）	
2	1	男	経営学部	3	8	70	10	2.0	2.9		-1.5	-0.7	
3	2	女	経済学部	1	10	100	20	3.0	2.8		0.5	-0.9	
4	3	女	経営学部	4	8	70	40	2.8	3.3		0.1	-0.1	
5	4	男	法学部	3	9	40	60	2.3	3.8		-0.9	0.7	
6	5	女	法学部	3	8	40	60	3.0	4.0		0.5	1.0	
7	6	女	経営学部	2	11	40	30	2.8	3.5		0.1	0.2	
8	7	女	法学部	2	8	90	30	2.7	3.5		-0.1	0.2	
9	8	男	経済学部	4	5	30	30	2.5	3.1		-0.5	-0.4	
10	9	男	経済学部	2	9	10	20	2.8	3.2		0.1	-0.3	
11	10	男	経営学部	2	8	30	5	2.2	2.7		-1.1	-1.1	
12	11	女	経営学部	2	11	10	30	2.8	3.4		0.1	0.1	
13	12	男	法学部	1	9	120	10	2.6	3.3		-0.3	-0.1	
14	13	男	経営学部	3	7	70	10	2.8	3.1		0.1	-0.4	
15	14	男	経済学部	1	9	50	5	2.4	2.4		-0.7	-1.5	
16	15	女	経済学部	4	9	90	100	2.0	3.7		-1.5	0.5	
17	16	女	法学部	1	9	90	5	2.1	3.1		-1.3	-0.4	
18	17	女	法学部	4	4	30	90	3.0	3.8		0.5	0.7	
19	18	女	経営学部	1	11	70	100	3.2	4.7		0.9	2.1	
20	19	男	経営学部	4	6	40	90	3.2	3.7		0.9	0.5	
21	20	男	経営学部	4	7	80	50	2.6	2.7		-0.3	-1.1	
22	21	男	法学部	1	12	60	120	3.5	4.2		1.5	1.3	
23	22	男	経済学部	3	6	30	30	3.2	3.6		0.9	0.4	
24	23	男	経営学部	1	10	40	50	3.0	3.0		0.5	-0.6	
25	24	男	経済学部	1	12	60	60	3.7	4.1		1.9	1.2	
26	25	女	経済学部	3	9	80	10	2.2	2.3		-1.1	-1.7	
27	26	女	法学部	2	10	50	20	3.3	3.8		1.1	0.7	
28	27	男	法学部	2	10	20	40	3.4	3.8		1.3	0.7	
29	28	女	経営学部	3	10	60	10	1.8	2.2		-1.9	-1.9	
30	29	男	経営学部	2	8	90	10	2.4	2.6		-0.7	-1.2	
31	30	男	法学部	4	4	60	180	3.6	4.6		1.7	2.0	
32													
33							最大値	3.7	4.7				
34							最小値	1.8	2.2				
35							レンジ	1.9	2.5				
36													
37							平均値	2.8	3.4				
38							標準偏差	0.496308596	0.627950812				
39													
40													

これより、「大学の成績平均」と「高校の評定平均」をそれぞれ標準化したときに、少なくともどちらか一方の標準化後の値が 1.5 を超える学生の番号は 18, 24, 30 であることがわかる（なお、標準化する際にスピルを使うなら、「=STANDARDIZE(H2:I31,H37:I37,H38:I38)」のように入力すればいい）。

問題 15.27

データタブの（分析グループにある）[データ分析] を選択して、分析ツールを使って相関係数を求める。

	総排気量 (cc)	全長 (mm)	全幅 (mm)	全高 (mm)	ホイールベース (mm)	最低地上高 (mm)	車両重量 (kg)	モード燃費 (km/l)
総排気量 (cc)	1							
全長 (mm)	-0.382666547	1						
全幅 (mm)	0.095255196	0.43323162	1					
全高 (mm)	-0.350411501	0.539311729	0.731406073	1				
ホイールベース (mm)	0.052735078	0.001291044	-0.688858493	-0.692364315	1			
最低地上高 (mm)	-0.084924761	0.192392349	0.802503137	0.836244431	-0.957791795	1		
車両重量 (kg)	0.019734093	0.626757179	0.851224213	0.757745659	-0.500495902	0.663604135	1	
モード燃費 (km/l)	-0.651483393	0.19853259	0.042235417	0.510129185	-0.380791017	0.384028156	0.175418069	1

- 選択肢の中で最も相関係数が大きいのは「車両重量 (kg) と全幅 (mm)」である（相関係数は約 0.851）。つまり、答えは (c) である。
- 選択肢の中で最も相関係数が 0 に近いのは「ホイールベース (mm) と全長 (mm)」である（相関係数は約 0.001）。つまり、答えは (d) である。
- 選択肢の中で、モード燃費 (km/l) との相関係数が負になるのは「総排気量 (cc)」のみである（相関係数は約 -0.651）。つまり、答えは (a) である。
- 散布図を作成する際、選択範囲の 2 列のうち、左側の列が横軸になり、右側の列が縦軸になる。なので、
 - 縦軸：モード燃費 (km/l)，横軸：全高 (mm)
 - 縦軸：モード燃費 (km/l)，横軸：車両重量 (kg)
 - 縦軸：モード燃費 (km/l)，横軸：総排気量 (cc)
 の散布図については、列の配置を変えずにそのまま範囲選択すれば作成できる。それぞれ作成すると、(f) の「縦軸：モード燃費 (km/l)，横軸：総排気量 (cc)」の散布図が問題の散布図であることがわかる。つまり、答えは (f) である。

補足

たとえば、(a) の「縦軸：全高 (mm)，横軸：モード燃費 (km/l)」の散布図を作成するには、F 列の全高 (mm) のデータを J 列のモード燃費 (km/l) の右側の列に移動させてから、その 2 列のデータを範囲選択すればいい。

- 表中のどこかのセルを選択した状態で、挿入タブの [ピボットテーブル]（の「テーブルまたは範囲から」）を選択する。「テーブル/範囲」に表全体が選択されていることを確認する。OK ボタンを押すと「ピボットテーブルのフィールド」が出てくる。
 「駆動形式」に「総排気量 (cc)」にチェックを入れる。すると、「行」ボックスに「駆動形式」，「値」ボックスに「合計/総排気量 (cc)」が入る。
 ピボットテーブルの行（横）に駆動形式の項目が出てきて、それぞれの項目ごとの「総排気量 (cc)」のそれぞれの合計が出てくる。
 求めたいのは平均値なので、「値」ボックス内の「合計/総排気量 (cc)」をクリックし、値フィールドの設定を選び、「選択したフィールドのデータ」を「平均」にする。

行ラベル	平均 / 総排気量 (cc)
4WD	1928.111111
FF	1797
FR	2387
総計	1954.333333

これより、総排気量 (cc) の平均値が最も大きいのは FR (平均値は 2387cc) であることがわかる。

つまり、答えは (c) である。

- (6) 「ピボットテーブルのフィールド」の「駆動形式」、「全高 (mm)」にチェックを入れる。すると、「行」ボックスに「駆動形式」、「値」ボックスに「合計/全高 (mm)」が入る。求めたいのは平均値なので、「値」ボックス内の「合計/全高 (mm)」をクリックし、値フィールドの設定を選び、「選択したフィールドのデータ」を「平均」にする。

行ラベル	平均 / 全高 (mm)
4WD	1545.833333
FF	1467.5
FR	1310
総計	1493.5

これより、全高 (mm) の平均値が最も大きいのは 4WD (平均値は約 1545.8mm) であることがわかる。

つまり、答えは (a) である。

- (7) 駆動形式別に分かれるようにデータの並べ替えをする。そのため、B 列の駆動形式が入力されているどこかのセルを選択した状態で、ホームタブの（編集グループにある）[並べ替えとフィルター] をクリックし、「昇順」または「降順」を選択する。

駆動形式別にデータが上下に分かれるので、[4WD]、[4WD 以外] それぞれについての「ホイールベース (mm) と車両重量 (kg) の相関係数」を CORREL 関数で求めればよい。[4WD] については約 -0.329 となり、[4WD 以外] については約 0.826 となるので、「ホイールベース (mm) と車両重量 (kg) の相関係数」が正であるのは [4WD 以外] であることがわかる。よって、ホイールベース (mm) が大きくなるほど車両重量 (kg) が大きくなる傾向にあるのは [4WD 以外] である。

つまり、答えは (b) である。

問題 15.28

- (1) 分析ツールの「回帰分析」を選び、「入力 Y 範囲」は結果系のデータ（この場合は「モード燃費 (km/l)」, J1:J31), 「入力 X 範囲」は原因系のデータ（この場合は「総排気量 (cc)」, C1:C31), 「ラベル」にチェックする。「残差」にもチェックを入れておく。

	係数
切片	18.58383
総排気量 (cc)	-0.00269

これより、回帰式の傾きは約 -0.00269 、切片は約 18.584 であることがわかる。つまり、総排気量 (cc) (x) とモード燃費 (km/l) (y) についての回帰式は

$$y = -0.00269x + 18.584 \quad (\text{傾きと切片は概数})$$

である。よって、[総排気量 2000cc のときのモード燃費 (km/l) の予測値] は、空いているセルに「=[傾きが入力されたセル番地]*2000+[切片が入力されたセル番地]」と入力すると求められる。約 13.2 (km/l) と計算される。

- (2) 回帰分析の結果、残差が最も大きくなるのは観測値 9、つまり、9 番の乗用車であり、その残差は約 2.0 (km/l) である。
- (3) 回帰分析の結果、残差の絶対値が最も小さくなるのは観測値 23、つまり、23 番の乗用車である。よって、モード燃費 (km/l) の実測値と回帰式による予測値が最も近いのは 23 番の乗用車であり、その予測値は約 13.7 (km/l) である。

観測値	予測値: モード燃費 (km/l)	残差
1	12.15464842	-0.25465
2	12.15464842	-0.45465
3	12.15464842	-0.15465
4	12.15464842	-0.35465
5	12.15464842	-1.35465
6	13.21046773	-0.81047
7	13.21046773	-0.21047
8	13.21046773	0.789532
9	13.21046773	1.989532
10	14.2770607	-0.77706
11	14.2770607	-0.27706
12	14.2770607	-0.17706
13	13.21046773	-0.81047
14	13.21046773	-0.21047
15	13.21046773	0.189532
16	13.21046773	0.789532
17	14.2770607	-0.77706
18	14.2770607	-0.27706
19	14.2770607	-0.17706
20	12.15464842	-1.15465
21	13.74915104	-0.14915
22	13.74915104	-0.14915
23	13.74915104	-0.04915
24	13.21046773	0.789532
25	13.21046773	0.789532
26	13.21046773	0.789532
27	13.74915104	-0.14915
28	13.21046773	1.789532
29	13.21046773	1.789532
30	14.2770607	-0.97706

- (4) 回帰分析の結果、総排気量 (cc) (x) とモード燃費 (km/l) (y) についての回帰式は

$$y = -0.00269x + 18.584 \quad (\text{傾きと切片は概数})$$

である。一般に、直線の方程式「 $y = ax + b$ (a, b は定数)」において、 x が 1 増えると y は傾き a の分だけ増える。よって、総排気量が 1 cc 上がると、モード燃費は約 -0.00269 km/l だけ増えると予測できることになる。つまり、総排気量が 100 cc 上がると、モード燃費は約 0.269 km/l だけ減ると予測できるということがわかる。

- (5) 総排気量 (cc) とモード燃費 (km/l) についての回帰分析における R^2 乗値は約 0.424 であることがわかる。

回帰統計	
重相関 R	0.651483393
重決定 R2	0.424430611
補正 R2	0.403874561
標準誤差	0.86510384
観測数	30