Excel演習で理解する統計学入門

解答例

目次

第 1 章　データの種類

1.1　質的データと量的データ

問題 1-1

質的データ（定期考査時期，リスニングの有無，補習対象，順位，10 段階評価）の書式を「文字列」に，量的データ（点数，偏差値）の書式を「数値」にすると，次のようになります．

	A	B	C	D	E	F	G
1	定期考査時期	リスニングの有無	補習対象	点数	偏差値	順位	10段階評価
2	第1期考査	なし	30点未満	70	57	88	7
3	第2期考査	あり	30点未満	86	63	53	8
4	第3期考査	なし	30点未満	72	60	60	7
5	第4期考査	あり	20点未満	66	69	31	9
6	第5期考査	あり	30点未満	93	71	13	10

1.1 の確認テスト

問 1

(1) 質的データ，量的データ
(2) 名義
(3) 順序
(4) 間隔
(5) 比例

問 2

(1) 売上個数：量的データ（比例尺度）
(2) 年齢：量的データ（比例尺度）
(3) 年代：質的データ（順序尺度）
(4) 時間帯：質的データ（順序尺度）
(5) 時刻：量的データ（間隔尺度）
(6) 震度：質的データ（順序尺度）
(7) ペットの有無：質的データ（名義尺度）
(8) 受験番号：質的データ（名義尺度）
(9) 方角：質的データ（名義尺度）
(10) 体重：量的データ（比例尺度）
(11) テストの問題番号：質的データ（名義尺度）
(12) 年収：量的データ（比例尺度）
(13) 受験の合否：質的データ（名義尺度）
(14) 車の所有台数：量的データ（比例尺度）
(15) テストの点数：量的データ（間隔尺度）

(16) 検定の級：質的データ（順序尺度）
(17) 西暦：量的データ（間隔尺度）
(18) 電話番号：質的データ（名義尺度）
(19) 車のナンバー：質的データ（名義尺度）
(20) 志望順位：質的データ（順序尺度）
(21) 家賃：量的データ（比例尺度）
(22) アンケートの選択肢番号：質的データ（名義尺度）

[**補足**]　以上は解釈の例であり，質的データと量的データの分類は絶対的ではありません．

1.2　データセットの種類

問題 1-2

まずは，ファイル「問題 1-2 元データ」のシート「A 店」の上部にある，列番号「A」から「B」をドラッグして，「コピー」を選択します（この際，「A」という文字から「B」という文字までを右にドラックして，A 列から B 列全体を選択するようにします．また，選択範囲内で右クリックしましょう）．

そして，シート「パネルデータ」のセル A1 を選択し，Enter キーを押します．

つぎに，シート「B 店」の列番号「B」を右クリックし，「コピー」を選択します（この際，「B」という文字を右クリックして，B 列全体を選択するようにしましょう）．

そして，シート「パネルデータ」のセル C1 を選択し，Enter キーを押します．

シート「C 店」以降についても同様のことをくり返すと，次のように 1 セットのパネルデータとしてまとめることができます．

	A	B	C	D	E	F
1		A店での売上金額（円）	B店での売上金額（円）	C店での売上金額（円）	D店での売上金額（円）	E店での売上金額（円）
2	1月	918810	549810	298890	771210	217710
3	2月	1036890	597780	258300	800730	221400
4	3月	774900	516600	225090	889290	302580
5	4月	889290	512910	335790	701100	339480
6	5月	697410	597780	372690	889290	446490
7	6月	734310	509220	328410	808110	512910
8	7月	763830	605160	335790	819180	483390
9	8月	867150	568260	339480	660510	586710
10	9月	878220	586710	376380	778590	697410
11	10月	926190	619920	254610	848700	741690
12	11月	808110	630990	343170	730620	907740
13	12月	885600	564570	335790	671580	1055340

1.2 の確認テスト

問 1

(1) データセット
(2) クロスセクションデータ
(3) 時系列データ
(4) パネルデータ

問 2

「行」は「横方向の並び」のことです．また，「列」は「縦方向の並び」のことです．

問 3

「知りたい対象」は「行」に入り,「その対象についての知りたい情報」は「列」に入る形が想定されています.

問 4

(1) クロスセクションデータ（複数対象，複数変数，同じ時点）
(2) 時系列データ（1 対象，1 変数，複数時点）
(3) パネルデータ（複数対象，複数変数，複数時点）
(4) パネルデータ（複数対象，1 変数，複数時点）
(5) クロスセクションデータ（複数対象，1 変数，同じ時点）
(6) 時系列データ（1 対象，1 変数，複数時点）
(7) パネルデータ（複数対象，1 変数，複数時点）
(8) クロスセクションデータ（1 対象，複数変数，同じ時点）
(9) 時系列データ（1 対象，複数変数，複数時点）

1.3　母集団と標本

問題 1-3

セル L1 に「標本 2」，セル Q1 に「標本 3」と入力します.

E 列の乱数が入力されているセルのひとつ（← どのセルでもいい）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」）を選択します．すると，データがランダムに並び替わります.

セル範囲 A2:D102 を選択して右クリックし，「コピー」を選びます．そして，セル L2 をクリックし，Enter キーを押します．貼り付けられた 100 名分のデータを「標本 2」としましょう.

同じように，150 名分のデータを抽出し，「標本 3」としましょう.

1.3 の確認テスト

問 1

母集団は「調査対象のデータ全体」のことを指す用語です．また，標本は「調査対象のデータ全体から抽出された一部分」を指す用語です.

問 2

標本サイズ（または，標本の大きさ）

問 3

母集団を調べることは全数調査といわれます．また，標本を調べて母集団について推測することは標本調査といわれます.

問 4

データの特徴を把握することを目的とし，母集団にも標本にも適用することができる統計学の分野は記述統計学といわれます．また,「一部分を調べて全体を知る」ための方法を研究する，標本調査にもとづく統計学の分野は推測統計学といわれます.

問 5

(1) 標本調査

(2) 全数調査

(3) 全数調査

(4) 全数調査

(5) 標本調査

(6) 全数調査

(7) 標本調査

第 2 章　データの可視化

2.1　質的データの可視化

問題 2-1

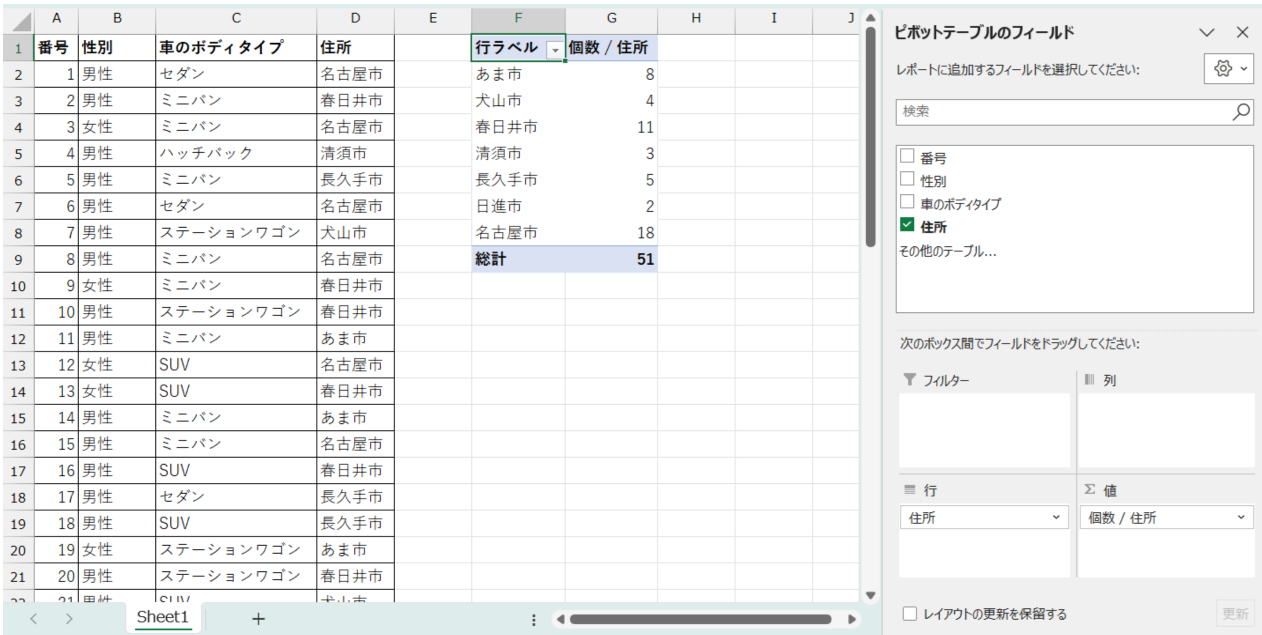
データが入力されている表中のどこかのセル（← どのセルでもいい）を選択した状態で，挿入タブの（テーブルグループにある）［ピボットテーブル］の「テーブルまたは範囲から」を選択します．

「テーブルまたは範囲からのピボットテーブル」ダイアログボックスおいて，「テーブル/範囲」に表全体が選択されていることを確認します．また，ピボットテーブルを配置する場所は，「既存のワークシート」を選択し，「場所」を入力するところにカーソルをおき，セル F1 をクリックして指定します．

「OK」ボタンを押すと「ピボットテーブルのフィールド」が出てきます．「☐ 住所」にチェックを入れると，「行」ボックスに「住所」が入ります．

さらに，チェックを入れた「☑ 住所」（という文字列）を右下へドラッグして「値」ボックスに移動させます．

すると，各市を住所とする人数がそれぞれ表示されます．



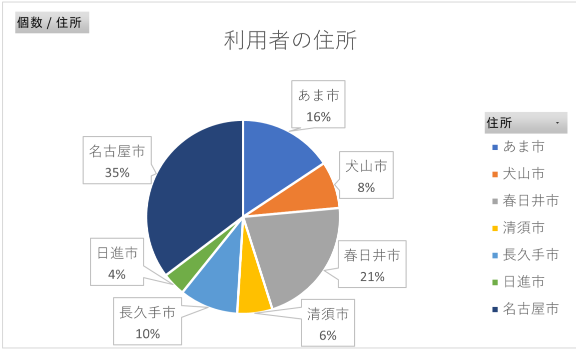
問題 2-2

セル F1 など（← ピボットテーブルが配置されているセル範囲のなかならどれでもいい）を選択した状態で，ピボットテーブル分析タブの（ツールにある）［ピボットグラフ］を選択します．

「グラフの挿入」ダイアログボックスにおいて，（左側のグラフの種類の一覧から）「円」を選び，そのなかから（ふつうの）「円」を選択して「OK」を押します．

出てきた円グラフの上で右クリックして，「データラベルの追加」から「データの吹き出しを追加」を選択します．なお，吹き出しの配置は，ドラッグして調整できます．

グラフタイトルは「利用者の住所」に変更しましょう．



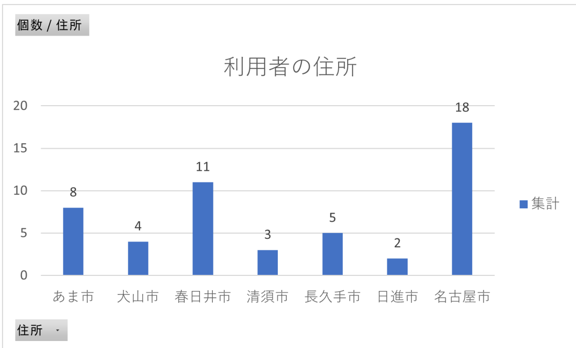
問題 2-3

セル F1 など（← ピボットテーブルが配置されているセル範囲のなかならどれでもいい）を選択した状態で，ピボットテーブル分析タブの（ツールグループにある）［ピボットグラフ］を選択します．

「グラフの挿入」ダイアログボックスにおいて，（左側のグラフの種類の一覧から）「縦棒」を選び，そのなかから「集合縦棒」を選択して「OK」を押します．

出てきた棒グラフの上で右クリックして，「データラベルの追加」から「データラベルの追加」を選択します．

グラフタイトルは「利用者の住所」に変更しましょう．



問題 2-4

データが入力されている表中のどこかのセル（← どのセルでもいい）を選択した状態で，挿入タブの（テーブルグループにある）［ピボットテーブル］の「テーブルまたは範囲から」を選択します．

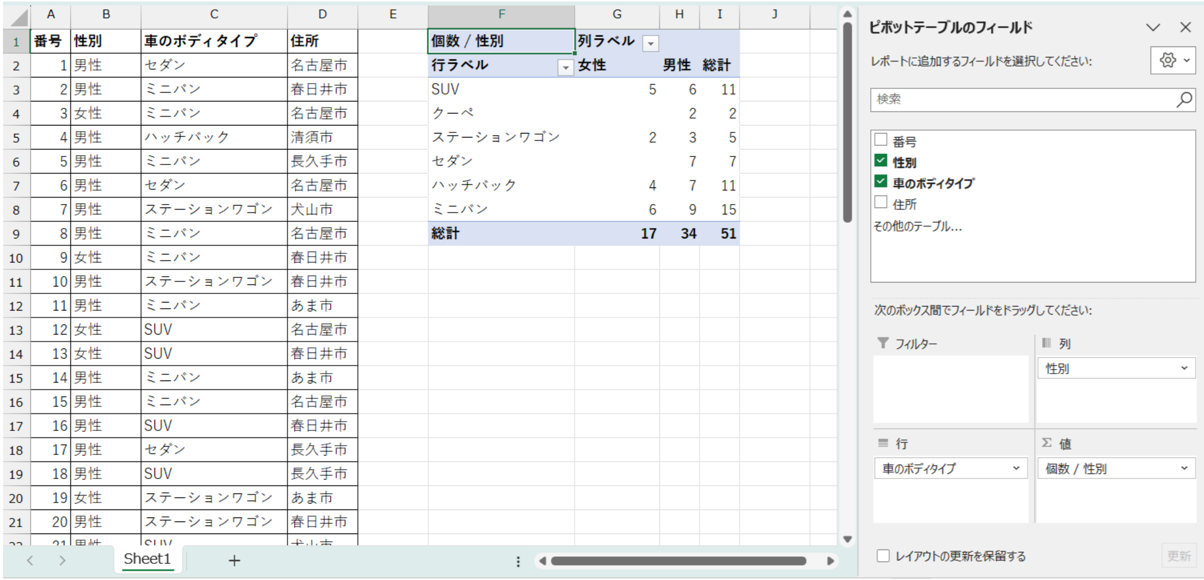
「テーブルまたは範囲からのピボットテーブル」ダイアログボックスにおいて，「テーブル/範囲」に表全体が選択されていることを確認します．また，ピボットテーブルを配置する場所は，「既存のワークシート」を選択し，「場所」を入力するところにカーソルをおき，セル F1 をクリックして指定します．

「OK」ボタンを押すと「ピボットテーブルのフィールド」が出てきます．「☐ 車のボディタイプ」，「☐ 性別」の順にチェックを入れると，「行」ボックスに（上から）「車のボディタイプ」と「性別」が入ります．

そして，チェックを入れた「☑ 性別」（という文字列）を右下へドラッグして「値」ボックスに移動させます．すると，「値」ボックスに「個数/性別」が入ります．

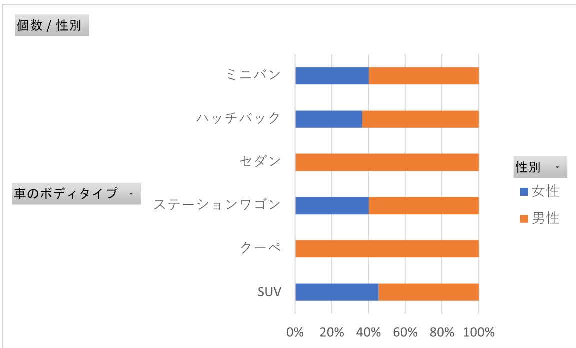
さらに，「行」ボックスに入っている「性別」を右上へドラッグして「列」ボックスに入れます．

すると，車のボディタイプごとに，それを所有している女性の人数と男性の人数がそれぞれ表示されます．



セル F1 など（← ピボットテーブルが配置されているセル範囲のなかならどれでもいい）を選択した状態で，ピボットテーブル分析タブの（ツールグループにある）［ピボットグラフ］を選択します．

「グラフの挿入」ダイアログボックスにおいて，（左側のグラフの種類の一覧から）「横棒」を選び，そのなかから「100% 積み上げ横棒」を選択して「OK」を押します．



2.1 の確認テスト

問 1

(1) 円グラフ

(2) 棒グラフ

(3) 円グラフ

(4) 棒グラフ

問 2

クロス集計

問 3

100% 積み上げ棒グラフ

2.2　量的データの可視化

問題 2-5

データ区間の設定のために，データの最大値と最小値を求めます．

セル E1 に「最大値」，セル E2 に「最小値」と入力します．セル F1 に来場者数の最大値，セル F2 に来場者数の最小値を，以下で説明するようにそれぞれ MAX 関数，MIN 関数を使って求めましょう．

まず，入力モードを「半角英数字」にします．

セル F1 に「=ma」と入力すると，予測変換で関数の候補の一覧が出てくるので，そこから「MAX」をダブルクリックし選択します．すると，「=MAX(」と入力されるので，最大値をとるデータの範囲（セル C2 からセル C50）をドラッグします．Enter キーを押すと，最大値 1893 が計算されます（セルには「=MAX(C2:C50)」と入力されます）．

セル F2 に「=mi」と入力すると，予測変換で関数の候補の一覧が出てくるので，そこから「MIN」をダブルクリックし選択します．すると，「=MIN(」と入力されるので，最大値をとるデータの範囲（セル C2 からセル C50）をドラッグします．Enter キーを押すと，最小値 288 が計算されます（セルには「=MIN(C2:C50)」と入力されます）．

セル H1 に「データ区間」と入力します．最小値が 288，最大値が 1893 なので，たとえば，300 以下，300 より大きく 500 以下，500 より大きく 700 以下，$\cdots$，1700 より大きく 1900 以下というようにデータ区間を設定すれば，どのデータもどれかひとつの区間に入ります．

セル H2 に「300」，セル H3 に「500」，セル H4 に「700」，$\cdots$，セル H10 に「1900」と入力しましょう．

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，分析ツールの「ヒストグラム」を選びます．［データ分析］ボタンが見あたらないときは，ファイルタブの（「その他...」の）「オプション」から「アドイン」を選択します．下のほうにある「Excel アドイン」の右にある「設定...」をクリックして，「分析ツール」にチェックを入れましょう．

「ヒストグラム」ダイアログボックスの「入力範囲」を入れるところにカーソルをおいて，セル C1 からセル C50 をドラッグして指定します．先頭行のセル B1「来場者数」はラベルとして扱うので，「ラベル」にチェックを入れます．

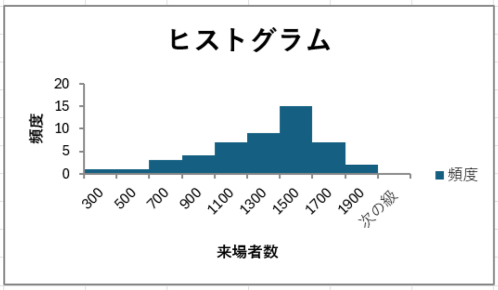
「データ区間」については，セル H1 からセル H10 をドラッグして指定します（セル H1「データ区間」はラベルとして扱います）．

また，「出力先」を選択し，すぐ右のそれを入れるところにカーソルをおいて，セル J1 をクリックして指定します．そして，「グラフ作成」にチェックを入れ，「OK」を押しましょう．

区間（階級）ごとのデータの個数（度数）をあらわす棒グラフ，つまり，ヒストグラムが出てきます．棒と棒の間隔をなくすために，棒の上で右クリック（またはダブルクリック）して「データ系列の書式設定」を出して，「要素の間隔」を「0%」にしましょう．

また，横軸のラベルを「来場者数」に変更しましょう．

E	F	G	H	I	J	K
最大値	1893		データ区間		データ区間	頻度
最小値	288		300		300	1
			500		500	1
			700		700	3
			900		900	4
			1100		1100	7
			1300		1300	9
			1500		1500	15
			1700		1700	7
			1900		1900	2
					次の級	0
					合計	49



問題 2-6

セル J12 に「合計」と入力します.

セル K12 に度数（頻度）の合計を以下で説明するように SUM 関数を使って求めましょう.

セル K12 に「=su」と半角で入力すると，予測変換で関数の候補の一覧が出てくるので，そこから「SUM」をダブルクリックし選択します．すると，「=SUM(」と入力されるので，合計をとるデータの範囲（セル K2 からセル K11）をドラッグします．Enter キーを押すと，合計 49 が計算されます（セルには「=SUM(K2:K11)」と入力されます）.

セル L1 に「相対度数」と入力します.

セル L2 に「=」を入力し，セル K2 をクリックします．続けて，「/」を入力し，度数の合計が計算されたセル（K12）をクリックします．そして，そのまま F4 キー（設定によっては Fn キー +F4 キー）を押すと，「=K2/$K$12」となります．このセル（L2）を 10 行目まで下にドラッグし，他の階級についての相対度数もオートフィルで求めます.

セル M1 に「累積度数」と入力します.

セル M2 に「=」を入力し，セル K2 をクリックします．セルには「=K2」と入力されます.

また，セル M3 には「=M2+K3」と入力します．このセル（M3）を 10 行目まで下にドラッグし，他の階級についての累積度数もオートフィルで求めます.

セル N1 に「累積相対度数」と入力します.

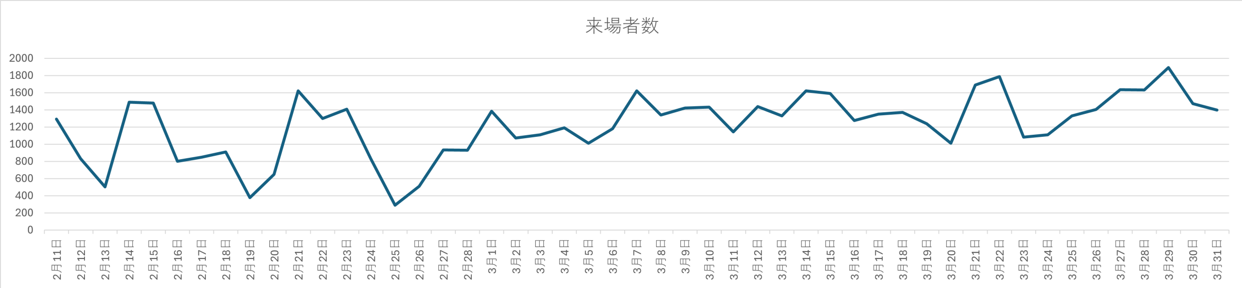
セル範囲 M2:M10 を選択し，そのセル範囲の右下あたりにマウスポインタを合わせ「+」の形にして，この状態のまま右にドラッグします．すると，累積相対度数もオートフィルで求められます.

E	F	G	H	I	J	K	L	M	N	O
最大値	1893		データ区間		データ区間	頻度	相対度数	累積度数	累積相対度数	
最小値	288		300		300	1	0.020408	1	0.020408163	
			500		500	1	0.020408	2	0.040816327	
			700		700	3	0.061224	5	0.102040816	
			900		900	4	0.081633	9	0.183673469	
			1100		1100	7	0.142857	16	0.326530612	
			1300		1300	9	0.183673	25	0.510204082	
			1500		1500	15	0.306122	40	0.816326531	
			1700		1700	7	0.142857	47	0.959183673	
			1900		1900	2	0.040816	49	1	
					次の級	0				
					合計	49				

一番大きい階級の累積度数は合計と等しくなり，一番大きい階級の累積相対度数は 1 になることが確認できます.

問題 2-7

セル範囲 A1:A50 を選択したあと，Ctrl キーを押しながらセル範囲 C1:C50 を選択します．そして，挿入タブの（グラフグループにある）［折れ線/面グラフの挿入］の「2-D 折れ線」の「折れ線」を選びます．



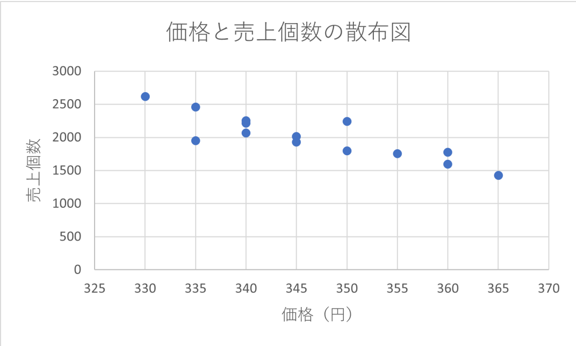
問題 2-8

価格と売上個数のデータの 2 列分（セル範囲 B1:C15）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」を選びます．横軸が価格（B 列）で縦軸が売上個数（C 列）になります．

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」から「第 1 横軸」を選択すると（横）軸ラベルが出てきます．（横）軸ラベルに「価格（円)」と入力します．

同じように，「軸ラベル」から「第 1 縦軸」を選択することにより（縦）軸ラベルを出し，「売上個数」と入力しましょう．

また，グラフタイトルは「価格と売上個数の散布図」に変更しましょう．



2.2 の確認テスト

問 1

(1) 階級
(2) 度数（または，頻度）
(3) 相対度数
(4) 1
(5) 累積度数
(6) 累積相対度数
(7) 合計
(8) 1

(9) ヒストグラム

問 2

(1) 折れ線グラフ

(2) 散布図

問 3

2 つの変数について散布図を作成する際，両者に因果関係があると予想されるならば，原因と思われる変数を x 軸にとり，結果と思われる変数を y 軸にとります．

問 4

（ア）は 3,（イ）は 1/50,（ウ）は 9/50,（エ）は 6/25,（オ）は 3/50,（カ）は 1 となります

	A	B	C
1	階級	度数	相対度数
2	150以上155未満	1	1/50
3	155以上160未満	7	7/50
4	160以上165未満	9	9/50
5	165以上170未満	9	9/50
6	170以上175未満	12	6/25
7	175以上180未満	7	7/50
8	180以上185未満	3	3/50
9	185以上190未満	2	1/25
10	合計	50	1
11			

問 5

	A	B	C	D	E
1	階級	度数	相対度数	累積度数	累積相対度数
2	35以上40未満	1	1/100	1	1/100
3	40以上45未満	14	7/50	15	3/20
4	45以上50未満	20	1/5	35	7/20
5	50以上55未満	15	3/20	50	1/2
6	55以上60未満	30	3/10	80	4/5
7	60以上65未満	16	4/25	96	24/25
8	65以上70未満	4	1/25	100	1
9					

ここで，次のように計算しています．

	A	B	C	D	E
1	階級	度数	相対度数	累積度数	累積相対度数
2	35以上40未満	1	1/100	1	1/100
3	40以上45未満	100×7/50	3/20-1/100	1+14	3/20
4	45以上50未満	100×1/5	1/5	15+20	3/20+1/5
5	50以上55未満	50-35	15/100	50	7/20+3/20
6	55以上60未満	30	30/100	50+30	1/2+3/10
7	60以上65未満	96-80	16/100	96	4/5+4/25
8	65以上70未満	4	4/100	96+4	1
9					

第 3 章　データの中心とデータのばらつき

3.1　代表値

問題 3-1

セル A9 に「国語と英語と社会の平均点」と入力します．列番号「A」（← セル A1 のすぐ上）と列番号「B」（← セル B1 のすぐ上）の間の境界線の上にマウスポインタを合わせてダブルクリックをし，A 列の幅を文字の幅に合った大きさにします．

セル B9 に「=av」と入力して，出てくる関数の候補の一覧から「AVERAGE」をダブルクリックし選択します．すると，「=AVERAGE(」と入力されるので，セル B2 をクリックして選択したあと，Ctrl キーを押しながら，セル B4 をクリックして選択します．そのあとまた，Ctrl キーを押しながら，セル B6 をクリックして選択します．Enter キーを押すと，生徒 A の国語と英語と社会の 3 科目の平均点 73 が計算されます（セルには「=AVERAGE(B2,B4,B6)」と入力されます）．

このセルを右にドラッグし，他の生徒それぞれの国語と英語と社会の平均点もオートフィルで求めます．

	A	B	C	D	E	F	G	H
1	科目	生徒A	生徒B	生徒C	生徒D	生徒E	生徒F	
2	国語	84	89	81	47	90	56	
3	数学	41	78	70	78	70	87	
4	英語	50	60	74	97	80	94	
5	理科	52	80	70	70	53	57	
6	社会	85	79	70	54	91	81	
7	5科目の平均点	62.4	77.2	73	69.2	76.8	75	
8	数学と理科の平均点	46.5	79	70	74	61.5	72	
9	国語と英語と社会の平均点	73	76	75	66	87	77	
10								

問題 3-2

セル A13 に「平均点」，セル A14 に「中央値」，セル A15 に「最頻値」と入力します．

セル B13 に身長の平均値を AVERAGE 関数で求めます（セルには「=AVERAGE(B2:B12)」と入力されます）．

そして，このセルを右にドラッグし，体重と腹囲それぞれの平均値もオートフィルで求めましょう．

入力モードを「半角英数字」にし，セル B14 に「=me」と入力すると，予測変換で関数の候補の一覧が出てくるので，そこから「MEDIAN」をダブルクリックし選択します．すると，「=MEDIAN(」と入力されるので，中央値をとるデータの範囲（セル B2 からセル B12）をドラッグします．Enter キーを押すと，身長の中央値 170 cm が計算されます（セルには「=MEDIAN(B2:B12)」と入力されます）．

そして，このセルを右にドラッグし，体重と腹囲それぞれの中央値もオートフィルで求めましょう．

セル B15 に「=mo」と入力すると，予測変換で関数の候補の一覧が出てくるので，そこから「MODE.MULT」をダブルクリックし選択します．すると，「=MODE.MULT(」と入力されるので，最頻値をとるデータの範囲（セル B2 からセル B12）をドラッグします．Enter キーを押すと，身長の最頻値 175 cm が計算されます（セルには「=MODE.MULT(B2:B12)」と入力されます）．

そして，このセルを右にドラッグし，体重と腹囲それぞれの最頻値もオートフィルで求めましょう（セル C15 には「#N/A」と表示されます．「#N/A」は該当なしという意味で，この場合，最頻値が存在しないことを示しています）．

	A	B	C	D	E
1	番号	身長（cm）	体重（kg）	腹囲（cm）	
2	1	163	54	70	
3	2	170	65	75	
4	3	180	66	81	
5	4	175	75	86	
6	5	161	49	59	
7	6	178	80	85	
8	7	175	61	77	
9	8	167	95	105	
10	9	182	101	90	
11	10	152	43	58	
12	11	155	50	59	
13	平均値	168.909091	67.1818182	76.8181818	
14	中央値	170	65	77	
15	最頻値	175	#N/A	59	
16					

問題 3-3

セル範囲 B1:B13 に各データを入力します．

A15 に「平均値」，セル A16 に「中央値」，セル A17 に「最頻値」と入力します．

セル B15 に平均値を AVERAGE 関数で求めます（セルには「=AVERAGE(B1:B13)」と入力されます）．

セル B16 に中央値を MEDIAN 関数で求めます（セルには「=MEDIAN(B1:B13)」と入力されます）．

セル B17 に最頻値を MODE.MULT 関数で求めます（セルには「=MODE.MULT(B1:B13)」と入力されます）．

	A	B
1		33
2		33
3		110
4		52
5		55
6		33
7		50
8		180
9		33
10		33
11		83
12		33
13		52
14		
15	平均値	60
16	中央値	50
17	最頻値	33

平均値は 60 万円，中央値は 50 万円，最頻値は 33 万円と計算されました．

平均値が一番大きくなりました．これは，値が大きいデータのせいで合計が大きくなるからです．

一方，中央値「50 万円」は 7 番目に大きい値「そのもの」であり，また，最頻値「33 万円」はもっともよく現れるデータ「そのもの」であることが確認できました．

3.1 の確認テスト

問 1

(1) 相加平均値

(2) 7

(3) 50，51，平均値

問 2

(1) 最頻値

(2) 中央値

(3) 最頻値

(4) 平均値

(5) 平均値，中央値，最頻値（または，平均値，最頻値，中央値）

問 3

$$\text{平均値} = \frac{\text{合計}}{\text{個数}} \quad \text{より} \quad \text{個数} = \frac{\text{合計}}{\text{平均値}}$$

なので，求めるデータの個数は

$$\frac{385}{55} = 7$$

より，7 であることがわかります．

問 4

学生 11 名のテストの平均点は 70 点なので，点数の合計は

$$70 \times 11 = 770$$

より，770 となります．つぎに，一番大きい点数を除いた平均点は 67 点なので，一番大きい点数を除いた合計は

$$67 \times 10 = 670$$

より，670 となります．よって，一番大きい点数は

$$770 - 670 = 100$$

より，100 であることがわかります．また，一番小さい点数を除いた平均点は 73.4 点なので，一番小さい点数を除いた合計は

$$73.4 \times 10 = 734$$

より，734 となります．よって，一番小さい点数は

$$770 - 734 = 36$$

より，36 であることがわかります．

問 5

$$\text{平均値} = \frac{\text{合計}}{\text{個数}} \quad \text{より} \quad \text{合計} = \text{平均値} \times \text{個数}$$

なので，求めるデータの合計は

$$4 \times 10 = 40$$

より，40 になります．よって，残りのデータは

$$40 - (9 + 2 + 3 + 0 + 4 + 8 + 8 + 6 + 0) = 0$$

より，0 であることがわかります．この 10 個のデータを小さい順に並べ替えると，

$$0,\ 0,\ 0,\ 2,\ 3,\ 4,\ 6,\ 8,\ 8,\ 9$$

となるので，中央値は

$$\frac{3 + 4}{2} = 3.5$$

であり，最頻値は 0 であることがわかります．

3.2　分散，標準偏差

問題 3-4

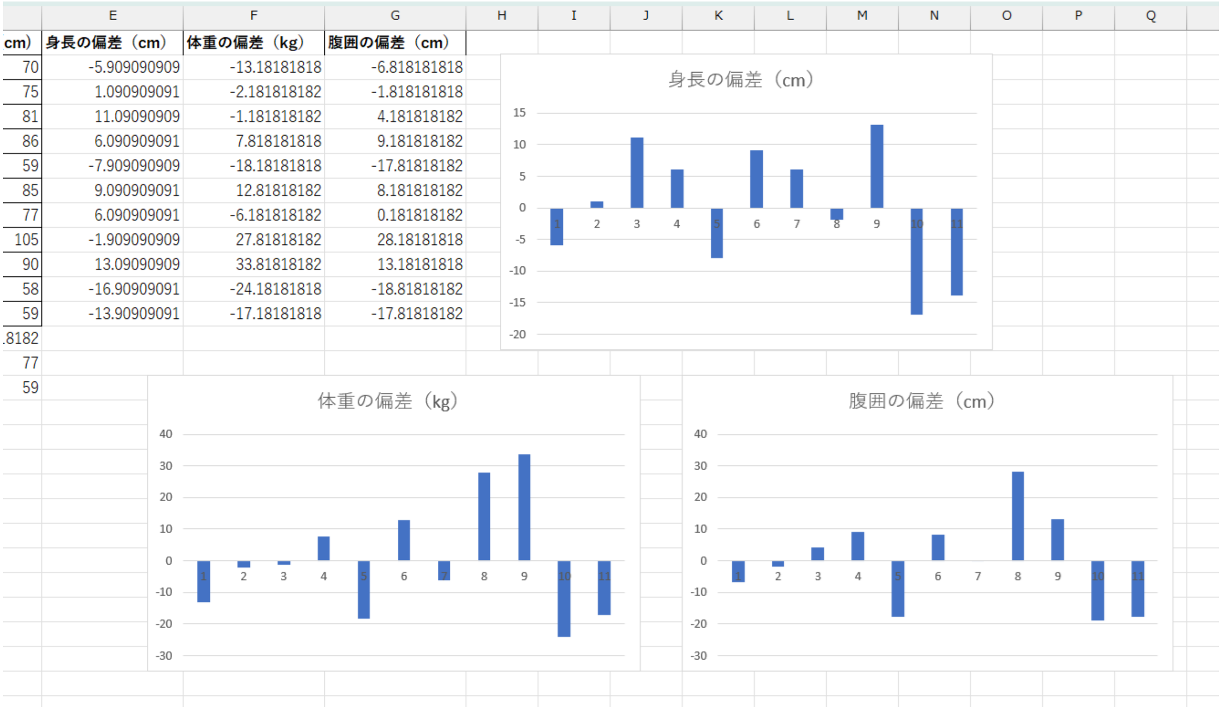
セル E1 に「身長の偏差（cm）」，セル F1 に「体重の偏差（kg）」，セル G1 に「腹囲の偏差（cm）」と入力します．

セル E2 に，番号 1 の身長の偏差を求めます．そのため，セル E2 に「=」を入力し，セル B2 をクリックします．続けて，「-」を入力し，平均値が計算されたセル（B13）をクリックします．そして，そのまま F4 キー（設定によっては Fn キー +F4 キー）を 2 回押すと，「=B2-B$13」となります．「13」の前に「$」記号が付き，オートフィルする際に「13（行目）」が固定されます．

このセル（E2）を 12 行目まで下にドラッグして，番号 2 から番号 11 の身長の偏差もオートフィルで求めます．さらにそのまま，選択されているセル範囲 E2:E12 の右下あたりにマウスポインタを合わせ「+」の形にして，この状態のまま右にドラッグします．すると，体重の偏差，腹囲の偏差もオートフィルで求められます．

セル範囲 E1:E12 を選択して，挿入タブの（グラフグループにある）[縦棒/横棒グラフの挿入］の「2-D 縦棒」の「集合縦棒」を選びます．

セル範囲 F1:F12，また，セル範囲 G1:G12 についても，同じように棒グラフを作成しましょう．



cm)	身長の偏差（cm）	体重の偏差（kg）	腹囲の偏差（cm）
70	-5.909090909	-13.18181818	-6.818181818
75	1.090909091	-2.181818182	-1.818181818
81	11.09090909	-1.181818182	4.181818182
86	6.090909091	7.818181818	9.181818182
59	-7.909090909	-18.18181818	-17.81818182
85	9.090909091	12.81818182	8.181818182
77	6.090909091	-6.181818182	0.181818182
105	-1.909090909	27.81818182	28.18181818
90	13.09090909	33.81818182	13.18181818
58	-16.90909091	-24.18181818	-18.81818182
59	-13.90909091	-17.18181818	-17.81818182
.8182			
77			
59			

問題 3-5

セル A16 に「分散」，セル A17 に「標準偏差」と入力します．

入力モードを「半角英数字」にし，セル B16 に「=v」と入力すると，予測変換で関数の候補の一覧が出てくるので，そこから「VAR.P」をダブルクリックし選択します．すると，「=VAR.P(」と入力されるので，分散をとるデータの範囲（セル B2 からセル B12）をドラッグします．Enter キーを押すと，身長の分散が計算されます（セルには「=VAR.P(B2:B12)」と入力されます）．そして，このセルを右にドラッグし，体重の分散，腹囲それぞれの分散もオートフィルで求めましょう．

つぎに，セル B17 に「=st」と入力すると，予測変換で関数の候補の一覧が出てくるので，そこから「STDEV.P」をダブルクリックし選択します．すると，「=STDEV.P(」と入力されるので，標準偏差をとるデータの範囲（セル B2 からセル B12）をドラッグします．Enter キーを押すと，身長の標準偏差が計算されます（セルには「=STDEV.P(B2:B12)」と入力されます）．そして，このセルを右にドラッグし，体重，腹囲それぞれの標準偏差もオートフィルで求めましょう．

	A	B	C	D	
1	番号	身長（cm）	体重（kg）	腹囲（cm）	身長の偏
2	1	163	54	70	-5.
3	2	170	65	75	1.
4	3	180	66	81	11
5	4	175	75	86	6.
6	5	161	49	59	-7.
7	6	178	80	85	9.
8	7	175	61	77	6.
9	8	167	95	105	-1.
10	9	182	101	90	13
11	10	152	43	58	-16
12	11	155	50	59	-13
13	平均値	168.909091	67.1818182	76.8181818	
14	中央値	170	65	77	
15	最頻値	175	#N/A	59	
16	分散	93.9008264	324.694215	197.785124	
17	標準偏差	9.69024388	18.0192734	14.0636099	

問題 3-6

セル A18 に「変動係数」と入力します．

セル B18 に「=B17/B13」と入力すると，身長の変動係数が計算されます．そして，これを右にドラッグし，体重の変動係数，腹囲の変動係数もオートフィルで求めましょう．

	A	B	C	D	
1	番号	身長（cm）	体重（kg）	腹囲（cm）	身長の偏
2	1	163	54	70	-5.
3	2	170	65	75	1.
4	3	180	66	81	11
5	4	175	75	86	6.
6	5	161	49	59	-7.
7	6	178	80	85	9.
8	7	175	61	77	6.
9	8	167	95	105	-1.
10	9	182	101	90	13
11	10	152	43	58	-16
12	11	155	50	59	-13
13	平均値	168.909091	67.1818182	76.8181818	
14	中央値	170	65	77	
15	最頻値	175	#N/A	59	
16	分散	93.9008264	324.694215	197.785124	
17	標準偏差	9.69024388	18.0192734	14.0636099	
18	変動係数	0.05736958	0.26821652	0.18307658	

3.2 の確認テスト

問 1

(1) 偏差

(2) 2 乗

(3) 標準偏差

(4) 分散

(5) 変動係数

(6) 10

(7) 10000

(8) 平均値

問 2

平均値は，

$$\frac{1+0+3+1+6+1}{6} = \frac{12}{6} = 2$$

と計算され，2 となります．偏差は各データから平均値を引いたものなので，各データの偏差はそれぞれ

$$1-2,\ 0-2,\ 3-2,\ 1-2,\ 6-2,\ 1-2$$

つまり，

$$-1,\ -2,\ 1,\ -1,\ 4,\ -1$$

であることがわかります．これより分散は

$$\frac{(-1)^2+(-2)^2+1^2+(-1)^2+4^2+(-1)^2}{6} = \frac{24}{6} = 4$$

と計算され，4 となり，標準偏差は

$$\sqrt{4} = 2$$

より，2 となります．よって，変動係数は

$$\frac{2}{2} = 1$$

と計算され，1 となります．

問 3

偏差は各データから平均値を引いたものなので，各データは偏差に平均値を加えたものということになります．よって，求める各データは

$$-3+2,\ -1+2,\ 0+2,\ 4+2,\ -5+2,\ 2+2,\ 3+2,\ 3+2,\ -4+2,\ 1+2$$

つまり，

$$-1,\ 1,\ 2,\ 6,\ -3,\ 4,\ 5,\ 5,\ -2,\ 3$$

であることがわかります．これより分散は

$$\frac{(-3)^2+(-1)^2+0^2+4^2+(-5)^2+2^2+3^2+3^2+(-4)^2+1^2}{10} = \frac{90}{10} = 9$$

と計算され，9 となり，標準偏差は

$$\sqrt{9} = 3$$

より，3 となります．よって，変動係数は

$$\frac{3}{2} = 1.5$$

と計算され，1.5 となります．

問 4

問 2 の各データを 10 倍すると，

$$-1\times 10,\ 1\times 10,\ 2\times 10,\ 6\times 10,\ -3\times 10,\ 4\times 10,\ 5\times 10,\ 5\times 10,\ -2\times 10,\ 3\times 10$$

つまり，

$$-10,\ 10,\ 20,\ 60,\ -30,\ 40,\ 50,\ 50,\ -20,\ 30$$

となります．よって，平均値は

$$\frac{-10+10+20+60+(-30)+40+50+50+(-20)+30}{10} = \frac{200}{10} = 20$$

と計算され，20 となります．各データの偏差はそれぞれ

$$-10-20,\ 10-20,\ 20-20,\ 60-20,\ -30-20,\ 40-20,\ 50-20,\ 50-20,\ -20-20,\ 30-20$$

つまり，

$$-30,\ -10,\ 0,\ 40,\ -50,\ 20,\ 30,\ 30,\ -40,\ 10$$

となります．これより分散は

$$\frac{(-30)^2+(-10)^2+0^2+40^2+(-50)^2+20^2+30^2+30^2+(-40)^2+10^2}{10}=\frac{9000}{10}=900$$

と計算され，900 になり，標準偏差は

$$\sqrt{900}=30$$

より，30 となります．よって，変動係数は

$$\frac{30}{20}=1.5$$

と計算され，1.5 となります．

よって，各データを 10 倍したとき，平均値は 10 倍，標準偏差も 10 倍，変動係数は 1 倍（つまり変化なし）になることが確認できました．

第 4 章　2 項分布

4.1　ベルヌーイ試行

問題 4-1

セル A2 に「=RANDBETWEEN(0,1」と入力し，Enter キーを押します．ふたたびセル A2 を選択し，このセルの右下あたりにマウスポインタを合わせ「+」の形にして，この状態のまま 61 行目まで下にドラッグし，オートフィルします．

すると，A 列に「0」または「1」がランダムに 60 個発生します．

つぎに，条件をみたすセルの個数を返す COUNTIF 関数を使って，「1」が出た回数，つまり，A 列にある「1」と表示されているセルの個数を求めましょう．そのため，セル C2 に「=COUNTIF(」と入力し，列番号「A」(← セル A1 のすぐ上) をクリックし，A 列全体を指定します．続けて，「,1」と入力し，Enter キーを押します．セルには「=COUNTIF(A:A,1)」と入力され，A 列にある「1」と表示されているセルの個数が返されます．

また，セル C4 には「=60-C2」と入力し，「0」が出た回数を求めます．

そして，セル D2 には「=C2/60」と入力し，「1」が出た割合

$$\frac{「1」が出た回数}{コインを投げた回数}$$

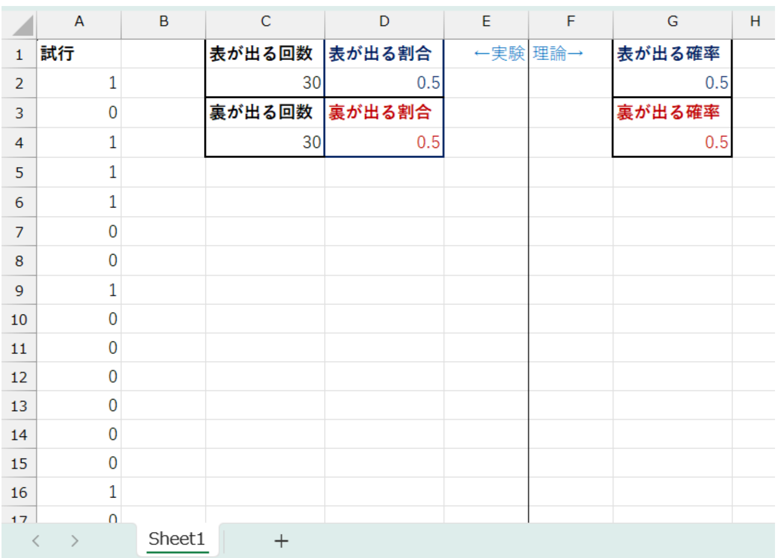
を計算します．

また，セル D4 には「=1-D2」(または，「=C4/60」) と入力し，「0」が出る割合を求めます．

セル G2 に「=1/2」と入力し，表が出る確率を求めます．

セル D2 の「1」が出る割合が，この確率 1/2 (= 0.5) に近いことが確認できます．

また，セル G4 に「=1-G2」と入力し，裏が出る確率 1/2 (= 0.5) を求めます．



	A	B	C	D	E	F	G	H
1	試行		表が出る回数	表が出る割合	←実験	理論→	表が出る確率	
2	1		30	0.5			0.5	
3	0		裏が出る回数	裏が出る割合			裏が出る確率	
4	1		30	0.5			0.5	
5	1							
6	1							
7	0							
8	0							
9	1							
10	0							
11	0							
12	0							
13	0							
14	0							
15	0							
16	1							
17	0							

Sheet1

はじめの 30 回は表が出て残りの 30 回は裏が出る確率は

「1 回目に「1」が出る確率」×「2 回目に「1」が出る確率」× ··· ×「30 回目に「1」が出る確率」
　×「31 回目に「0」が出る確率」× ··· ×「60 回目に「0」が出る確率」

$$= \left(\frac{1}{2}\right)^{30} \times \left(\frac{1}{2}\right)^{30}$$

と計算して求められます（**独立事象の乗法定理**）．セル I2 には「=G2^30*G4^30」と入力し，この確率を求めましょう．

表が出る回数が 30 であるときの表と裏の並び方のパターンは，上記の

はじめの 30 回は表が出て残りの 30 回は裏が出る

以外にもあります．全部で何通りあるのでしょうか．これは，60 回のなかから「1」が出る回を 30 個選ぶときの選び方の総数であり，

相異なる 60 個のものから 30 個を重複なく選んだものの総数（組み合わせの総数）

となります．つまり，これは $\boldsymbol{{}_{60}\mathrm{C}_{30}}$ とあらわされます．これら ${}_{60}\mathrm{C}_{30}$ 通りのパターンそれぞれについて，それが起こる確率は上記で求めた

はじめの 30 回は表が出て残りの 30 回は裏が出る確率

と等しくなります．

セル I5 に「=COMBIN(60,30)」と入力し，${}_{60}\mathrm{C}_{30}$ の値を求めましょう．

以上より，「1」が出る回数が 30 である確率は，

はじめの 30 回は表が出て残りの 30 回は裏が出る確率

を ${}_{60}\mathrm{C}_{30}$ 個足し合わせることによって求められます（**加法定理（2 つの事象が互いに排反である場合）**）．

セル I8 に「=I5*I2」と入力し，コインを 60 回投げたときの表が出る回数が 30 である確率を求めましょう．

F	G	H	I
理論一	表が出る確率		はじめの30回は表が出て残りの30回は裏が出る確率
	0.5		8.67362E-19
	裏が出る確率		
	0.5		表が出る回数が30であるときの表と裏の並び方のパターンの総数
			1.18265E+17
			表が出る回数が30である確率
			0.102578173

4.1 の確認テスト

問 1

(1) 試行
(2) 事象，全事象，根元事象
(3) ベルヌーイ試行
(4) 余事象

問 2

「1 の目が出る」という事象 A は,

$$B = \{1\}$$

というようにあらわすことができ,「素数の目が出る」という事象 P は,

$$P = \{2,\ 3,\ 5\}$$

というようにあらわすことができます.

問 3

根元事象は 6 個あり，それぞれ次のようにあらわすことができます.

$$\{1\},\quad \{2\},\quad \{3\},\quad \{4\},\quad \{5\},\quad \{6\}$$

問 4

出る目についてのすべての場合は $\{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$ なので 6 通りあります.

2 の目が出る場合は $\{2\}$，3 の目が出る場合は $\{3\}$，5 の目が出る場合は $\{5\}$ なので，それぞれ 1 通りあります．よって，サイコロを振ったときに 2 の目が出る確率，3 の目が出る確率，5 の目が出る確率は，どれも 1/6 であることがわかります.

また，奇数の目が出る場合は $\{2,\ 3,\ 5\}$ なので 3 通りあります．よって，奇数の目が出る確率は（3/6 =）1/2 であることがわかります.

問 5

すべての場合は $\{(表, 表),\ (表, 裏),\ (裏, 表),\ (裏, 裏)\}$ なので 4 通りあり，どちらも裏が出る場合は $\{(裏, 裏)\}$ なので 1 通りあります．よって，どちらも裏が出る確率は 1/4 であることがわかります.

また，少なくとも 1 回は表が出る場合は $\{(表, 表),\ (表, 裏),\ (裏, 表)\}$ なので 3 通りあります．よって，少なくとも 1 回は表が出る確率は 3/4 であることがわかります.

[別解]　コインを 2 回投げたときに「少なくとも 1 回は表が出ること」は「どちらも裏が出ること」の余事象なので,

$$「少なくとも 1 回は表が出る確率」= 1 -「どちらも裏が出る確率」= 1 - \frac{1}{4} = \frac{3}{4}$$

と計算して求めることもできます.

問 6

「1 回目に裏が出るという事象」と「2 回目に裏が出るという事象」と「3 回目に裏が出るという事象」と「4 回目に裏が出るという事象」と「5 回目に裏が出るという事象」は独立なので，コインを 5 回投げたときにどれも裏が出る確率は

$$\begin{aligned}&「1 回目に裏が出る確率」\times「2 回目に裏が出る確率」\times「3 回目に裏が出る確率」\\&\quad\times「4 回目に裏が出る確率」\times「5 回目に裏が出る確率」\\&\quad= \frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2} = \frac{1}{32}\end{aligned}$$

と計算して求めることができ，1/32 であることがわかります（独立事象の乗法定理).

また，コインを 5 回投げたときに「少なくとも 1 回は表が出ること」は「どれも裏が出ること」の余事象なので,

$$「少なくとも 1 回は表が出る確率」= 1 −「どれも裏が出る確率」= 1 - \frac{1}{32} = \frac{31}{32}$$

と計算して求めることができます．

問 7

コインを 5 回投げたときに，はじめの 1 回は表が出て残りの 4 回は裏が出る確率は

「1 回目に表が出る確率」×「2 回目に裏が出る確率」×「3 回目に裏が出る確率」
×「4 回目に裏が出る確率」×「5 回目に裏が出る確率」

$$= \left(\frac{1}{2}\right)^1 \times \left(\frac{1}{2}\right)^4 = \frac{1}{32}$$

と計算して求められます（独立事象の乗法定理）．

コインを 5 回投げたときに，表が出る回数が 1 であるときの表と裏の並び方のパターンは，上記の

はじめの 1 回は表が出て残りの 4 回は裏が出る

以外にもあります．全部で何通りあるのでしょうか．これは，5 回のなかから表が出る回を 1 個選ぶときの選び方の総数であり，

相異なる **5** 個のものから **1** 個を重複なく選んだものの総数（組み合わせの総数）

となります．つまり，これは $\mathbf{{}_5C_1}$ とあらわされます．これら ${}_5C_1$ 通りのパターンそれぞれについて，それが起こる確率は上記で求めた

はじめの 1 回は表が出て残りの 4 回は裏が出る確率

と等しくなります．

よって，表が出る回数が 1 である確率は，

はじめの 1 回は表が出て残りの 4 回は裏が出る確率

を ${}_5C_1$ 個足し合わせることによって求められます（加法定理（2 つの事象が互いに排反である場合））．つまり，

$${}_5C_1 \times \left(\frac{1}{2}\right)^1 \times \left(\frac{1}{2}\right)^4$$

を計算することにより求められます．これは，

$$5 \times \frac{1}{32} = \frac{5}{32}$$

と計算されるので，表が出る回数が 1 である確率は 5/32 であることがわかります．

同様に，表が出る回数が 2 である確率は，

$${}_5C_2 \times \left(\frac{1}{2}\right)^2 \times \left(\frac{1}{2}\right)^3 = \frac{5 \times 4}{2 \times 1} \times \frac{1}{32} = \frac{10}{32} = \frac{5}{16}$$

となり，表が出る回数が 3 である確率は，

$${}_5C_3 \times \left(\frac{1}{2}\right)^3 \times \left(\frac{1}{2}\right)^2 = \frac{5 \times 4 \times 3}{3 \times 2 \times 1} \times \frac{1}{32} = \frac{10}{32} = \frac{5}{16}$$

となり，表が出る回数が 4 である確率は，

$$
{}_5\mathrm{C}_4 \times \left(\frac{1}{2}\right)^4 \times \left(\frac{1}{2}\right)^1 = \frac{5 \times 4 \times 3 \times 2}{4 \times 3 \times 2 \times 1} \times \frac{1}{32} = \frac{5}{32}
$$

となります.

4.2　2 項分布

問題 4-2

「コインを 60 回投げる」という試行を 100 回おこないます.

セル H4 に「=RANDBETWEEN(0,1」と入力し，Enter キーを押します．このセル（H4）を 63 行目まで下にドラッグしてオートフィルします．さらにそのまま，選択されているセル範囲 H4:H63 の右下あたりにマウスポインタを合わせ「+」の形にして，この状態のまま右にドラッグし，試行 100 までオートフィルします．すると，H 列から右に 100 列分の各列に，「0」または「1」がランダムに 60 個ずつ発生します（つまり，合計 6000 個発生します）.

つぎに，「コインを 60 回投げる」という 100 回おこなわれた試行について，表が出た回数，つまり，「1」が出た回数をそれぞれ求めます.

まず，COUNTIF 関数を使って，H 列にある「1」と表示されているセルの個数をセル H2 に求めましょう．そのため，セル H2 に「=COUNTIF(」と入力し，セル範囲 H4:H63 を選択して指定します．続けて，「,1」と入力し，Enter キーを押します．セルには「=COUNTIF(H4:H63,1)」と入力され，セル範囲 H4:H63 にある「1」と表示されているセルの個数が返されます.

このセル（H2）を右にドラッグし，試行 100 までオートフィルしましょう.

そして，COUNTIF 関数を使って，「1」が出た回数が 0 である試行の個数（度数）をセル B3 に求めます.

そのため，セル B3 に「=COUNTIF(」と入力し，セル範囲 H2:DC2（H2 から右に 100 セル分）を選択して指定し，そのまま F4 キー（設定によっては Fn キー +F4 キー）を押します．続けて，「,」と入力したあと，セル A3 をクリックして，Enter キーを押します．セルには「=COUNTIF($H$2:$DC$2,A3)」と入力され，セル範囲 H2:DC2 にある「1」と表示されているセルの個数が返されます.

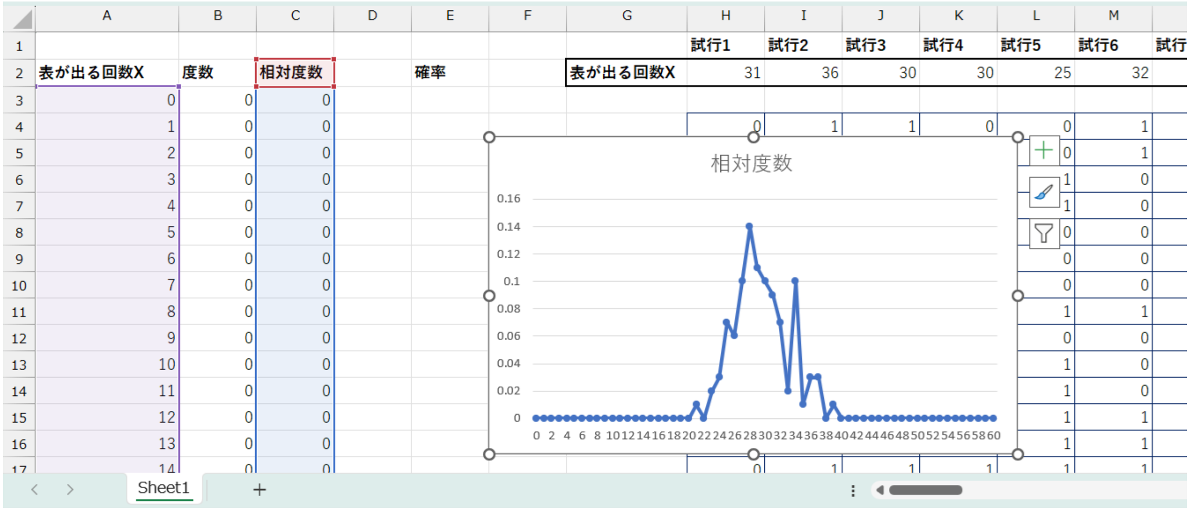
このセル（B3）を下にドラッグし，「1」が出た回数が 60 である試行の個数（度数）までオートフィルで求めましょう.

セル C3 に「=B3/100」と入力し，「1」が出た回数が 0 である試行の個数（度数）を全度数 100 で割った値（相対度数）を求めます.

このセル（C3）を下にドラッグし，「1」が出た回数が 60 である試行の個数（度数）を全度数 100 で割った値（相対度数）までオートフィルで求めましょう.

相対度数のマーカー付き折れ線グラフを作成します.

そのため，セル範囲 C2:C63 を選択し，挿入タブの（グラフグループにある）[折れ線/面グラフの挿入] の「2-D 折れ線」の「マーカー付き折れ線」を選びます．作成されたグラフが選択されたまま，グラフのデザインタブの（データグループにある）[データの選択] をクリックします.「データソースの選択」ダイアログボックスが出てくるので，横（項目）軸ラベルの「編集」をクリックします.「軸ラベル」ダイアログボックスが出てくるので，A 列の該当箇所（A3:A63）をドラッグして表示させて「OK」を押します．さらに，「データソースの選択」ダイアログボックスでも「OK」を押すと，グラフの横軸の値が A 列の値に変わります.



問題 4-3

はじめの k 回は「1」が出て残りの $(60-k)$ 回は「0」が出る確率は

「1 回目に「1」が出る確率」×「2 回目に「1」が出る確率」× ⋯ ×「k 回目に「1」が出る確率」
×「$(k+1)$ 回目に「0」が出る確率」× ⋯ ×「60 回目に「0」が出る確率」

$$= \left(\frac{1}{2}\right)^{k} \times \left(\frac{1}{2}\right)^{(60-k)}$$

と計算して求められます（独立事象の乗法定理）．ここで，k は 0 から 60 までの自然数です．

「1」が出る回数が k であるときの「1」と「0」の並び方のパターンは，上記の

はじめの k 回は「1」が出て残りの $(60-k)$ 回は「0」が出る

以外にもあります．全部で何通りあるのでしょうか．これは，60 回のなかから「1」が出る回を k 個選ぶときの選び方の総数であり，

相異なる **60** 個のものから $\boldsymbol{k}$ 個を重複なく選んだものの総数（組み合わせの総数）

となります．つまり，これは ${}_{60}\mathrm{C}_k$ とあらわされます．これら ${}_{60}\mathrm{C}_k$ 通りのパターンそれぞれについて，それが起こる確率は上記で求めた

はじめの n 回は「1」が出て残りの $(60-k)$ 回は「0」が出る確率

と等しくなります．

よって，「1」が出る回数が k である確率は，

はじめの k 回は「1」が出て残りの $(60-k)$ 回は「0」が出る確率

を ${}_{60}\mathrm{C}_k$ 個足し合わせることによって求められます（加法定理（2 つの事象が互いに排反である場合））．

セル E3 に「=COMBIN(60,A3)*(1/2)^A3*(1/2)^(60-A3)」と入力し，「1」が出る回数が 0 である確率を求めましょう．

そして，このセル（E3）を下へドラッグし，「1」が出る回数が 60 である確率までオートフィルで求めましょう．

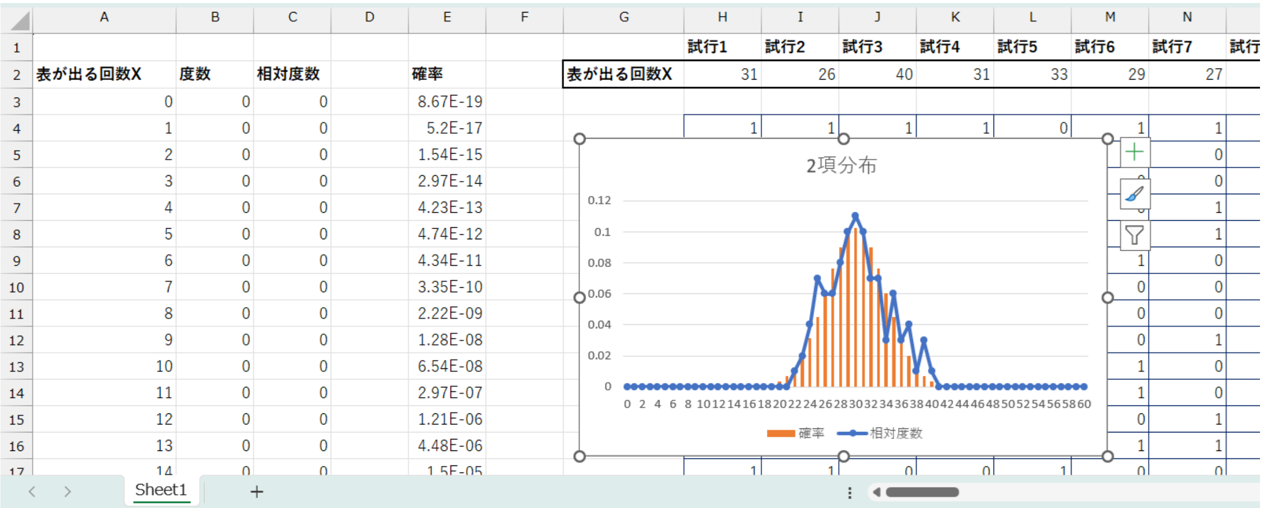
「相対度数のマーカー付き折れ線グラフ」と「確率分布の棒グラフ」を，同一グラフエリアにそれぞれ作成します．

そのため，セル範囲 C2:C63 を選択し，Ctrl キーを押しながら，セル範囲 E2:E63 を選択したあと，挿入タブの（グラフグループにある）［折れ線/面グラフの挿入］の「2-D 折れ線」の「マーカー付き折れ線」を選びます．作成されたグラフが選択されたまま，グラフのデザインタブの（データグループにある）［データの選択］をクリックします．「データソースの選択」ダイアログボックスが出てくるので，横（項目）軸ラベルの「編集」をクリックします．「軸ラベル」ダイアログボックスが出てくるので，A 列の該当箇所（A3:A63）をドラッグして表示させて「OK」を押します．さらに，「データソースの選択」ダイアログボックスでも「OK」を押すと，グラフの横軸の値が A 列の値に変わります．

つぎに，作成したグラフをクリックして選択し，その状態で，グラフのデザインタブの（種類グループにある）［グラフの種類の変更］を選びます．

すると，「グラフの種類の変更」ダイアログボックスが出てくるので，「確率」のグラフの種類を「集合縦棒」に変更しましょう（注意:「グラフの種類の変更」ダイアログボックス左側は「組み合わせ」が選ばれています）.

グラフタイトルは「2 項分布」にしましょう．



2 つのグラフの形がお互い似ていることが確かめられます．

問題 4-4

セル E3 に「=BINOM.DIST(A3,60,1/2,FALSE)」と入力し直し，試行回数が 60，成功率 1/2 である 2 項分布 $B(60, 1/2)$ における成功数が 0 である確率を求めましょう．

そして，このセル（E3）を下へドラッグし，成功数が 60 である確率までオートフィルで求めましょう．

まったく値が変わらないことがわかります．

問題 4-5

セル C65 に「=SUMPRODUCT(A3:A63,C3:C63)」と入力し，「A 列（表が出る回数 X）」と「C 列（相対度数）」の積和を求めましょう．これで，

$$\text{セル A3} \times \text{セル C3} + \text{セル A4} \times \text{セル C4} + \cdots + \text{セル A63} \times \text{セル C63}$$

つまり，表が出た回数 X の平均値が計算されます．

なお，ためしに，AVERAGE 関数を使い（空いているセルに）「=AVERAGE(H2:DC2)」と入力して平均値を求めてみると，セル C65 で求めた値と一致することが確かめられます

セル E65 に「=SUMPRODUCT(A3:A63,E3:E63)」と入力し，「A 列の値（表が出る回数 X）」と「E 列の値（確率）」の積和を求めましょう．これで，

$$\text{セル A3} \times \text{セル E3} + \text{セル A4} \times \text{セル E4} + \cdots + \text{セル A63} \times \text{セル E63}$$

つまり，2 項分布 $B(60, 1/2)$ の期待値 30 が計算されます．

60	57	0	0		2.97E-14		1	0	1	0	0
61	58	0	0		1.54E-15		0	0	0	1	0
62	59	0	0		5.2E-17		1	1	1	1	1
63	60	0	0		8.67E-19		1	1	0	0	0
64											
65	**平均値（期待値）**		30.43		30						
66	**分散**										
67											

問題 4-6

セル D3 に「=(A3-$C$65)^2」と入力し，このセルを 63 行目まで下にドラッグすることにより，「A 列の値（表が出る回数 X）とセル C65 に求めた平均値の差の 2 乗」をオートフィルで求めましょう．ここで，$記号は，式中の「C65」を入力直後に F4 キー（設定によっては Fn キー +F4 キー）を押すことによって入力できます．

そして，セル C66 に「=SUMPRODUCT(D3:D63,C3:C63)」と入力し，「A 列の値（表が出る回数 X）とセル C65 に求めた平均値の差の 2 乗」と「C 列の値（相対度数）」の積和を求めましょう．これで，

$$\text{セル D3} \times \text{セル C3} + \text{セル D4} \times \text{セル C4} + \cdots + \text{セル D63} \times \text{セル C63}$$

つまり，表が出た回数 X の分散が計算されます．

なお，ためしに，VAR.P 関数を使い（空いているセルに）「=VAR.P(H2:DC2)」と入力して分散を求めてみると，セル C66 で求めた値と一致することが確かめられます．

セル F3 に「=(A3-$E$65)^2」と入力し，このセルを 63 行目まで下にドラッグすることにより，「A 列の値（表が出る回数 X）とセル E65 に求めた期待値の差の 2 乗」をオートフィルで求めましょう．

そして，セル E66 に「=SUMPRODUCT(F3:F63,E3:E63)」と入力し，「A 列の値（「1」が出る回数 X）とセル E65 に求めた期待値の差の 2 乗」と「E 列の値（確率）」の積和を求めましょう．これで，

$$\text{セル F3} \times \text{セル E3} + \text{セル F4} \times \text{セル E4} + \cdots + \text{セル F63} \times \text{セル E63}$$

つまり，2 項分布 $B(60, 1/2)$ の分散 15 が計算されます．

60	57	0	0	754.0516	2.97E-14	729	0	0	1	1	1
61	58	0	0	809.9716	1.54E-15	784	1	1	1	1	0
62	59	0	0	867.8916	5.2E-17	841	0	1	0	0	0
63	60	0	0	927.8116	8.67E-19	900	0	1	0	0	1
64											
65	**平均値（期待値）**		29.54		30						
66	**分散**		14.4084		15						
67											

問題 4-7

セル E65 に「=60*(1/2)」と入力しなおし，$B(60, 1/2)$ の期待値を計算すると，問題 4-5 で求めた結果 30 と一致することがわかります．

また，セル E66 に「=60*(1/2)*(1-1/2)」と入力しなおし，$B(60, 1/2)$ の分散を計算すると，問題 4-6 で求めた結果 15 と一致することがわかります．

4.2 の確認テスト

問 1

(1) 離散型確率変数，連続型確率変数

(2)（離散型）確率分布，1，確率質量関数

(3) 2 項分布

(4) n，p

(5) 期待値

(6) 分散，標準偏差

問 2

(1) X の確率分布は

x	1	2	3	4	5	6
$P(X=x)$	1/6	1/6	1/6	1/6	1/6	1/6

それぞれの確率を全部足すと，

$$\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{6}{6}=1$$

となり，1 になることが確かめられます．

(2)「X がとる値」と「その値をとる確率」の積すべての和は

$$1\times\frac{1}{6}+2\times\frac{1}{6}+3\times\frac{1}{6}+4\times\frac{1}{6}+5\times\frac{1}{6}+6\times\frac{1}{6}=\frac{21}{6}=3.5$$

となるので，X の期待値 $E(X)$ は 3.5 となります．

(3)「X がとる値と期待値の差の 2 乗」と「その値をとる確率」の積すべての和は

$$(1-3.5)^2\times\frac{1}{6}+(2-3.5)^2\times\frac{1}{6}+(3-3.5)^2\times\frac{1}{6}+(4-3.5)^2\times\frac{1}{6}+(5-3.5)^2\times\frac{1}{6}+(6-3.5)^2\times\frac{1}{6}=\frac{35}{12}$$

となるので，X の分散 $V(X)$ は 35/12 となります．

問 3

(1) X の確率分布は

x	0	1
$P(X=x)$	1/2	1/2

それぞれの確率を全部足すと，

$$\frac{1}{2}+\frac{1}{2}=1$$

となり，1 になることが確かめられます．

(2)「X がとる値」と「その値をとる確率」の積すべての和は

$$0\times\frac{1}{2}+1\times\frac{1}{2}=\frac{1}{2}=0.5$$

となるので，X の期待値 $E(X)$ は 0.5 となります．

(3)「X がとる値と期待値の差の 2 乗」と「その値をとる確率」の積すべての和は

$$\left(0-\frac{1}{2}\right)^2 \times \frac{1}{2} + \left(1-\frac{1}{2}\right)^2 \times \frac{1}{2} = \frac{1}{4} = 0.25$$

となるので，X の分散 $V(X)$ は 0.25 となります．

問 4

(1) X の確率分布は

x	0	1	2	3	4	5
$P(X=x)$	$1/32$	$5/32$	$5/16$	$5/16$	$5/32$	$1/32$

それぞれの確率を全部足すと，

$$\frac{1}{32}+\frac{5}{32}+\frac{5}{16}+\frac{5}{16}+\frac{5}{32}+\frac{1}{32}=\frac{32}{32}=1$$

となり，1 になることが確かめられます．

(2) $n=5$，$p=1/2$

(3)「X がとる値」と「その値をとる確率」の積すべての和は

$$0\times\frac{1}{32}+1\times\frac{5}{32}+2\times\frac{5}{16}+3\times\frac{5}{16}+4\times\frac{5}{32}+5\times\frac{1}{32}=\frac{80}{32}=2.5$$

となるので，X の期待値 $E(X)$ は 2.5 となります．

また，試行回数 n は 5，成功率 p は $1/2$ なので，公式 $E(X)=np$ で求めると，

$$E(X)=5\times\frac{1}{2}=2.5$$

となり，両者は一致することが確かめられます．

(4)「X がとる値と期待値の差の 2 乗」と「その値をとる確率」の積すべての和は

$$(0-2.5)^2\times\frac{1}{32}+(1-2.5)^2\times\frac{5}{32}+(2-2.5)^2\times\frac{5}{16}+(3-2.5)^2\times\frac{5}{16}+(4-2.5)^2\times\frac{5}{32}$$
$$+(5-2.5)^2\times\frac{1}{32}=1.25$$

となるので，X の分散 $V(X)$ は 1.25 となります．

また，試行回数 n は 5，成功率 p は $1/2$ なので，公式 $V(X)=np(1-p)$ で求めると，

$$V(X)=5\times\frac{1}{2}\times\left(1-\frac{1}{2}\right)=1.25$$

となり，両者は一致することが確かめられます．

問 5

試行回数 n は 100，成功率 p は $1/2$ なので，

$$E(X)=100\times\frac{1}{2}=50,\quad V(X)=100\times\frac{1}{2}\times\left(1-\frac{1}{2}\right)=25$$

問 6

試行回数 n は 500，成功率 p は $1/5$ なので，

$$E(X) = 500 \times \frac{1}{5} = 100, \quad V(X) = 500 \times \frac{1}{5} \times \left(1 - \frac{1}{5}\right) = 80$$

問 7

「くじを引いたときあたるかあたらないか」は，起こる結果が 2 つしかないので，ベルヌーイ試行です．あたる確率が一定なので，くじ引きを 40 回おこなって，あたる回数 X とすると，X は 2 項分布 $B(40, 1/20)$ に従います．

試行回数 n は 40，成功率 p は 1/20 なので，

$$E(X) = 40 \times \frac{1}{20} = 2, \quad V(X) = 40 \times \frac{1}{20} \times \left(1 - \frac{1}{20}\right) = 1.9$$

第 5 章　正規分布

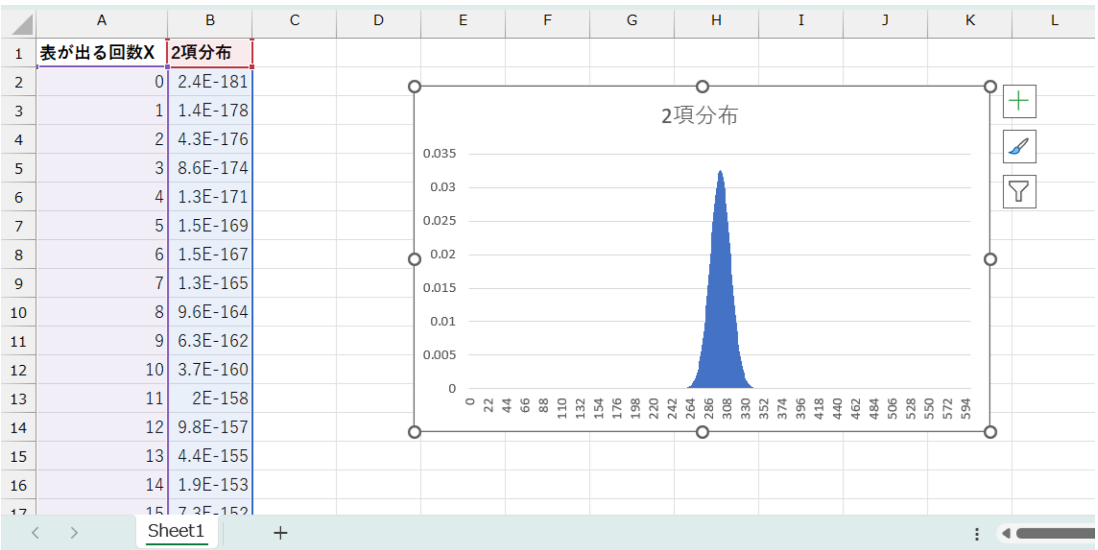
5.1　2 項分布から正規分布へ

問題 5-1

セル B2 に「=BINOM.DIST(A2,600,1/2,FALSE)」と入力し，試行回数が 600，成功率 1/2 である 2 項分布 $B(600, 1/2)$ における成功数が 0 である確率を求めます．

そして，このセル（B2）を下へドラッグし，成功数が 600 である確率までオートフィルで求めましょう．

セル範囲 B1:B602 を選択したあと，挿入タブの（グラフグループにある）［縦棒/横棒グラフの挿入］の「2-D 縦棒」の「集合縦棒」を選びます．作成されたグラフが選択されたまま，グラフのデザインタブの（データグループにある）［データの選択］をクリックします．「データソースの選択」ダイアログボックスが出てくるので，横（項目）軸ラベルの「編集」をクリックします．「軸ラベル」ダイアログボックスが出てくるので，A 列の該当箇所（A2:A602）をドラッグして表示させて「OK」を押します．さらに，「データソースの選択」ダイアログボックスでも「OK」を押すと，グラフの横軸の値が A 列の値に変わります．



問題 5-2

セル B2 に「=BINOM.DIST(A2,60,1/2,FALSE)」と入力し，試行回数が 60，成功率 1/2 である 2 項分布 $B(60, 1/2)$ における成功数が 0 である確率を求めましょう．

そして，このセル（B2）を下へドラッグし，成功数が 60 である確率までオートフィルで求めましょう．

セル B64 に「=60*(1/2)」と入力し，$B(60, 1/2)$ の期待値 30 を求めましょう．

また，セル B65 に「=60*(1/2)*(1-1/2)」と入力し，$B(60, 1/2)$ の分散 15 を求めましょう．

つぎに，正規分布の確率密度関数によって，確率密度を求めてみましょう．

セル C2 に「=(1/SQRT(2*PI()*$B$65))*EXP(-((A2-$B$64)^2)/(2*$B$65))」と入力し，期待値がセル B64 の値 30，分散がセル B65 の値 15 である正規分布の確率密度関数の $x = 0$ における値を求めます．

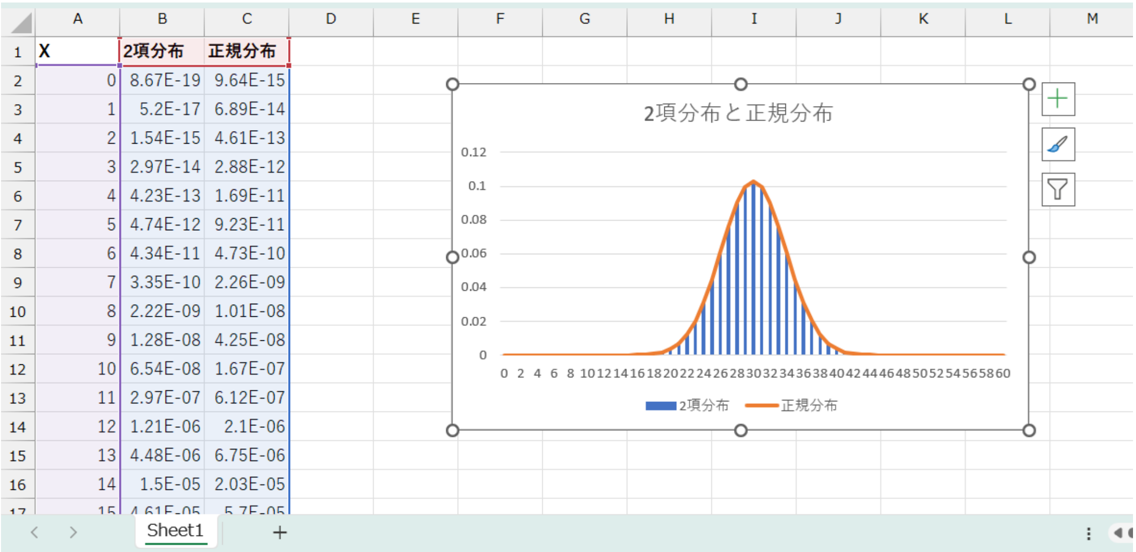
そして，このセル（C2）を下へドラッグし，$x = 60$ における値までオートフィルで求めましょう．

セル範囲 B1:C62 を選択したあと，挿入タブの（グラフグループにある）［縦棒/横棒グラフの挿入］の「2-D 縦棒」の「集合縦棒」を選びます．作成されたグラフが選択されたまま，グラフのデザインタブの（データグループにある）［データの選択］をクリックします．「データソースの選択」ダイアログボックスが出てくるので，横（項目）軸ラベルの「編集」をクリックします．「軸ラベル」ダイアログボックスが出てくるので，A 列の該当箇所（A2:A62）をドラッグして表示させて「OK」を押します．さらに，「データソースの選択」ダイアログボックスでも「OK」を押すと，グラフの横軸の値が A 列の値に変わります．

つぎに，作成したグラフをクリックして選択し，その状態で，グラフのデザインタブの（種類グループにある）［グラフの種類の変更］を選びます．

すると，「グラフの種類の変更」ダイアログボックスが出てくるので，「正規分布」のグラフの種類を「折れ線」に変更しましょう．

また，グラフタイトルは「2 項分布と正規分布」にしましょう．



X	2項分布	正規分布
0	8.67E-19	9.64E-15
1	5.2E-17	6.89E-14
2	1.54E-15	4.61E-13
3	2.97E-14	2.88E-12
4	4.23E-13	1.69E-11
5	4.74E-12	9.23E-11
6	4.34E-11	4.73E-10
7	3.35E-10	2.26E-09
8	2.22E-09	1.01E-08
9	1.28E-08	4.25E-08
10	6.54E-08	1.67E-07
11	2.97E-07	6.12E-07
12	1.21E-06	2.1E-06
13	4.48E-06	6.75E-06
14	1.5E-05	2.03E-05
15	4.61E-05	5.7E-05

2 項分布 $B(60, 1/2)$ のグラフが正規分布 $N(30, 15)$ のグラフで近似されていることが確かめられます．

問題 5-3

セル C2 に「=NORM.DIST(A2,$B$64,SQRT($B$65),FALSE)」と入力しなおし，期待値がセル B64 の値 30，分散がセル B65 の値 15 である正規分布の確率密度関数の $x = 0$ における値を求めましょう．

そして，このセル（C2）を下へドラッグし，$x = 60$ における値までオートフィルで求めましょう．

まったく値が変わらないことがわかります．

5.1 の確認テスト

問 1

(1) 確率密度関数，1

(2) 確率密度

(3) μ，σ^2，σ

(4) 36，9^2（または，81），9

(5) np，$np(1-p)$，np，$np(1-p)$

(6) 10，9，3，10，3^2（または，9）

問 2

正規分布は連続型確率分布，2 項分布は離散型確率分布です．

5.2　累積分布関数

問題 5-4

セル B2 に「=NORM.DIST(A2,60,5,FALSE)」と入力し，期待値が 60，標準偏差が 5 である正規分布の確率密度関数の $x = 0$ における値を求めます．

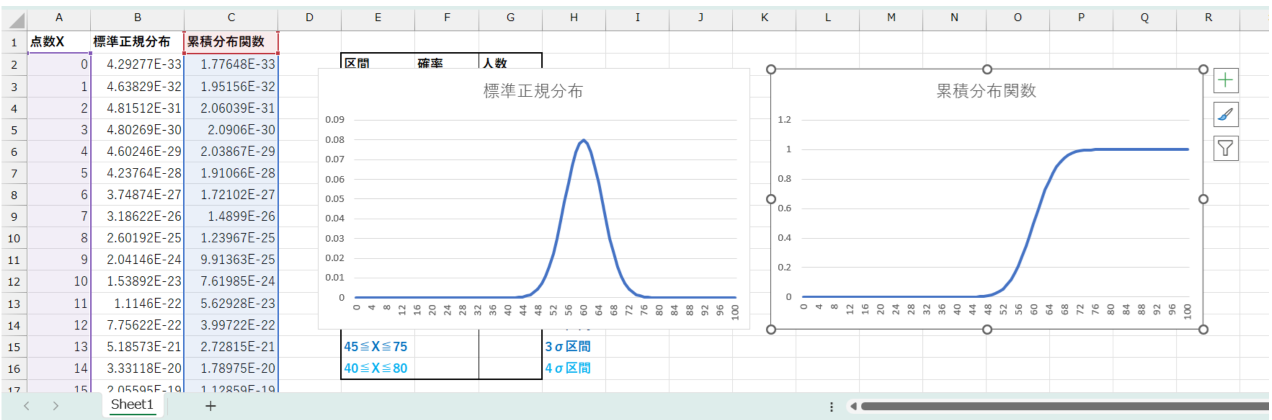
そして，このセル（B2）を下へドラッグし，$x = 100$ における値までオートフィルで求めましょう．

セル範囲 B1:B102 を選択したあと，挿入タブの（グラフグループにある）［折れ線/面グラフの挿入］の「2-D 折れ線」の「折れ線」を選びます．作成されたグラフが選択されたまま，グラフのデザインタブの（データグループにある）［データの選択］をクリックします．「データソースの選択」ダイアログボックスが出てくるので，横（項目）軸ラベルの「編集」をクリックします．「軸ラベル」ダイアログボックスが出てくるので，A 列の該当箇所（A2:A102）をドラッグして表示させて「OK」を押します．さらに，「データソースの選択」ダイアログボックスでも「OK」を押すと，グラフの横軸の値が A 列の値に変わります．

セル C2 に「=NORM.DIST(A2,60,5,TRUE)」と入力し，期待値が 60，標準偏差が 5 である正規分布の累積分布関数の $x = 0$ における値を求めます．

そして，このセル（C2）を下へドラッグし，$x = 100$ における値までオートフィルで求めましょう．

セル範囲 C1:C102 を選択したあと，挿入タブの（グラフグループにある）［折れ線/面グラフの挿入］の「2-D 折れ線」の「折れ線」を選びます．作成されたグラフが選択されたまま，グラフのデザインタブの（データグループにある）［データの選択］をクリックします．「データソースの選択」ダイアログボックスが出てくるので，横（項目）軸ラベルの「編集」をクリックします．「軸ラベル」ダイアログボックスが出てくるので，A 列の該当箇所（A2:A102）をドラッグして表示させて「OK」を押します．さらに，「データソースの選択」ダイアログボックスでも「OK」を押すと，グラフの横軸の値が A 列の値に変わります．



問題 5-5

セル F3 に「=NORM.DIST(60,60,5,TRUE)」と入力し，点数が 60 以下となる確率 $P(X \leq 60)$ を求めます．または，「=C62」と入力しても求められます．

セル F4 に「=1-NORM.DIST(60,60,5,TRUE)」と入力し，点数が 60 以上となる確率 $P(X \geq 60)$ を求めます．または，「=1-C62」と入力しても求められます．

ここで，「全事象が起こる確率は 1 である」ことから，「1 −「 点数が 60 以下となる確率」」を計算して求めています．

セル F5 に「=NORM.DIST(55,60,5,TRUE)」，セル F6 に「=1-NORM.DIST(65,60,5,TRUE)」，
セル F7 に「=NORM.DIST(50,60,5,TRUE)」，セル F8 に「=1-NORM.DIST(70,60,5,TRUE)」，
セル F9 に「=NORM.DIST(45,60,5,TRUE)」，セル F10 に「=1-NORM.DIST(75,60,5,TRUE)」，
セル F11 に「=NORM.DIST(40,60,5,TRUE)」，セル F12 に「=1-NORM.DIST(80,60,5,TRUE)」
とそれぞれ入力し，各確率を求めましょう．

	A	B	C	D	E	F	G	H
1	点数X	確率密度関数	累積分布関数					
2	0	4.29277E-33	1.77648E-33		区間	確率	人数	
3	1	4.63829E-32	1.95156E-32		X≦60	0.5		
4	2	4.81512E-31	2.06039E-31		X≧60	0.5		
5	3	4.80269E-30	2.0906E-30		X≦55	0.158655		
6	4	4.60246E-29	2.03867E-29		X≧65	0.158655		
7	5	4.23764E-28	1.91066E-28		X≦50	0.02275		
8	6	3.74874E-27	1.72102E-27		X≧70	0.02275		
9	7	3.18622E-26	1.4899E-26		X≦45	0.00135		
10	8	2.60192E-25	1.23967E-25		X≧75	0.00135		
11	9	2.04146E-24	9.91363E-25		X≦40	3.17E-05		
12	10	1.53892E-23	7.61985E-24		X≧80	3.17E-05		
13	11	1.1146E-22	5.62928E-23		55≦X≦65			1σ区間
14	12	7.75622E-22	3.99722E-22		50≦X≦70			2σ区間
15	13	5.18573E-21	2.72815E-21		45≦X≦75			3σ区間
16	14	3.33118E-20	1.78975E-20		40≦X≦80			4σ区間
17	15	2.05595E-19	1.12859E-19					

Sheet1

問題 5-6

セル F13 に「=NORM.DIST(65,60,5,TRUE)-NORM.DIST(55,60,5,TRUE)」と入力し，点数が 55 以上 65 以下となる確率 $P(55 \leq X \leq 65)$ を求めます．または，「=C67-C57」と入力しても求められます．

ここで，「「点数が 65 以下となる確率」−「点数が 55 以下となる確率」」を計算して求めています．

セル F14 に「=NORM.DIST(70,60,5,TRUE)-NORM.DIST(50,60,5,TRUE)」，
セル F15 に「=NORM.DIST(75,60,5,TRUE)-NORM.DIST(45,60,5,TRUE)」，
セル F16 に「=NORM.DIST(80,60,5,TRUE)-NORM.DIST(40,60,5,TRUE)」
とそれぞれ入力し，各確率を求めましょう．

	A	B	C	D	E	F	G	H
1	点数X	確率密度関数	累積分布関数					
2	0	4.29277E-33	1.77648E-33		区間	確率	人数	
3	1	4.63829E-32	1.95156E-32		X≦60	0.5		
4	2	4.81512E-31	2.06039E-31		X≧60	0.5		
5	3	4.80269E-30	2.0906E-30		X≦55	0.158655		
6	4	4.60246E-29	2.03867E-29		X≧65	0.158655		
7	5	4.23764E-28	1.91066E-28		X≦50	0.02275		
8	6	3.74874E-27	1.72102E-27		X≧70	0.02275		
9	7	3.18622E-26	1.4899E-26		X≦45	0.00135		
10	8	2.60192E-25	1.23967E-25		X≧75	0.00135		
11	9	2.04146E-24	9.91363E-25		X≦40	3.17E-05		
12	10	1.53892E-23	7.61985E-24		X≧80	3.17E-05		
13	11	1.1146E-22	5.62928E-23		55≦X≦65	0.682689		1σ区間
14	12	7.75622E-22	3.99722E-22		50≦X≦70	0.9545		2σ区間
15	13	5.18573E-21	2.72815E-21		45≦X≦75	0.9973		3σ区間
16	14	3.33118E-20	1.78975E-20		40≦X≦80	0.999937		4σ区間
17	15	2.05595E-19	1.12859E-19					

Sheet1

問題 5-7

セル G3 に「=F3*500」と入力し，点数が 60 以下となる人数を求めます．

ここで，「「点数が 60 以下となる確率」 × 「全人数」」を計算して求めています．

このセル（G3）を下にドラッグし，点数が 40 以上 80 以下となる人数までオートフィルで求めましょう．

	A	B	C	D	E	F	G	H
1	点数X	確率密度関数	累積分布関数					
2	0	4.29277E-33	1.77648E-33		区間	確率	人数	
3	1	4.63829E-32	1.95156E-32		X≦60	0.5	250	
4	2	4.81512E-31	2.06039E-31		X≧60	0.5	250	
5	3	4.80269E-30	2.0906E-30		X≦55	0.158655	79.32763	
6	4	4.60246E-29	2.03867E-29		X≧65	0.158655	79.32763	
7	5	4.23764E-28	1.91066E-28		X≦50	0.02275	11.37507	
8	6	3.74874E-27	1.72102E-27		X≧70	0.02275	11.37507	
9	7	3.18622E-26	1.4899E-26		X≦45	0.00135	0.674949	
10	8	2.60192E-25	1.23967E-25		X≧75	0.00135	0.674949	
11	9	2.04146E-24	9.91363E-25		X≦40	3.17E-05	0.015836	
12	10	1.53892E-23	7.61985E-24		X≧80	3.17E-05	0.015836	
13	11	1.1146E-22	5.62928E-23		55≦X≦65	0.682689	341.3447	1σ区間
14	12	7.75622E-22	3.99722E-22		50≦X≦70	0.9545	477.2499	2σ区間
15	13	5.18573E-21	2.72815E-21		45≦X≦75	0.9973	498.6501	3σ区間
16	14	3.33118E-20	1.78975E-20		40≦X≦80	0.999937	499.9683	4σ区間
17	15	2.05595E-19	1.12859E-19					

5.2 の確認テスト

問 1

(1) 累積分布関数

(2) 1σ 区間，　2σ 区間，　3σ 区間，　4σ 区間

(3) 0.5

問 2

4.2 の確認テスト　問 2 より，X の確率分布は次のようになることがわかっています．

x	1	2	3	4	5	6
$P(X=x)$	1/6	1/6	1/6	1/6	1/6	1/6

よって，累積分布関数の値は次のようになります．

x	1	2	3	4	5	6
$P(X \leq x)$	1/6	1/3	1/2	2/3	5/6	1

問 3

4.2 の確認テスト　問 3 より，X の確率分布は次のようになることがわかっています．

x	0	1
$P(X=x)$	1/2	1/2

よって，累積分布関数の値は次のようになります．

x	0	1
$P(X \leq x)$	1/2	1

問 4

4.2 の確認テスト　問 4 より，X の確率分布は次のようになることがわかっています．

x	0	1	2	3	4	5
$P(X = x)$	1/32	5/32	5/16	5/16	5/32	1/32

よって，累積分布関数の値は次のようになります．

x	0	1	2	3	4	5
$P(X \leq x)$	1/32	3/16	1/2	13/16	31/32	1

問 5

1σ 区間は $100 - 5 \leq X \leq 100 + 5$，つまり，$95 \leq X \leq 105$,

2σ 区間は $100 - 2 \times 5 \leq X \leq 100 + 2 \times 5$，つまり，$90 \leq X \leq 110$,

3σ 区間は $100 - 3 \times 5 \leq X \leq 100 + 3 \times 5$，つまり，$85 \leq X \leq 115$,

4σ 区間は $100 - 4 \times 5 \leq X \leq 100 + 4 \times 5$，つまり，$80 \leq X \leq 120$,

問 6

1σ 区間は $-211 - 16 \leq X \leq -211 + 16$，つまり，$-227 \leq X \leq -195$,

2σ 区間は $-211 - 2 \times 16 \leq X \leq -211 + 2 \times 16$，つまり，$-243 \leq X \leq -179$,

3σ 区間は $-211 - 3 \times 16 \leq X \leq -211 + 3 \times 16$，つまり，$-259 \leq X \leq -163$,

4σ 区間は $-211 - 4 \times 16 \leq X \leq -211 + 4 \times 16$ つまり，$-275 \leq X \leq -147$

問 7

(1) 正規分布は期待値を中心として左右対称なので，$P(X \geq 60) = a$

(2)「全事象が起こる確率は 1 である」ことから，$P(X \leq 60) = 1 - P(X \geq 60) = 1 - a$

(3) $P(40 \leq X \leq 60) = P(X \leq 60) - P(X \leq 40) = (1 - a) - a = 1 - 2a$

第 6 章　標準正規分布

6.1　正規分布のグラフ

問題 6-1

セル B2 に「=NORM.DIST(A2,-2,1,FALSE)」と入力し，期待値 −2，標準偏差 1 である正規分布の確率密度関数の $x = -5$ における値を求めましょう．

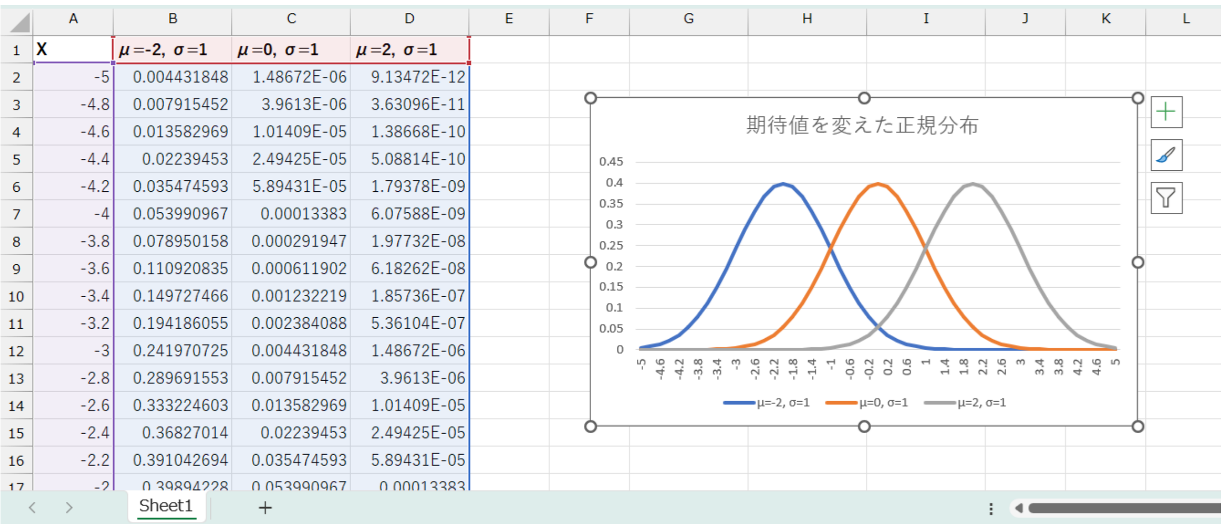
セル C2 に「=NORM.DIST(A2,0,1,FALSE)」と入力し，期待値 0，標準偏差 1 である正規分布の確率密度関数の $x = -5$ における値を求めましょう．

セル D2 に「=NORM.DIST(A2,2,1,FALSE)」と入力し，期待値 2，標準偏差 1 である正規分布の確率密度関数の $x = -5$ における値を求めましょう．

そして，セル範囲 B2:D2 を選択し，その右下あたりにマウスポインタを合わせ「+」の形にして，この状態のままダブルクリックします．すると，$x = 5$ における値までオートフィルで求めることができます．

セル範囲 B1:D52 を選択したあと，挿入タブの（グラフグループにある）［折れ線/面グラフの挿入］の「2-D 折れ線」の「折れ線」を選びます．作成されたグラフが選択されたまま，グラフのデザインタブの（データグループにある）［データの選択］をクリックします．「データソースの選択」ダイアログボックスが出てくるので，横（項目）軸ラベルの「編集」をクリックします．「軸ラベル」ダイアログボックスが出てくるので，A 列の該当箇所（A2:A52）をドラッグして表示させて「OK」を押します．さらに，「データソースの選択」ダイアログボックスでも「OK」を押すと，グラフの横軸の値が A 列の値に変わります．

グラフタイトルは「期待値を変えた正規分布」にしましょう．



	A	B	C	D
1	X	μ=-2, σ=1	μ=0, σ=1	μ=2, σ=1
2	-5	0.004431848	1.48672E-06	9.13472E-12
3	-4.8	0.007915452	3.9613E-06	3.63096E-11
4	-4.6	0.013582969	1.01409E-05	1.38668E-10
5	-4.4	0.02239453	2.49425E-05	5.08814E-10
6	-4.2	0.035474593	5.89431E-05	1.79378E-09
7	-4	0.053990967	0.00013383	6.07588E-09
8	-3.8	0.078950158	0.000291947	1.97732E-08
9	-3.6	0.110920835	0.000611902	6.18262E-08
10	-3.4	0.149727466	0.001232219	1.85736E-07
11	-3.2	0.194186055	0.002384088	5.36104E-07
12	-3	0.241970725	0.004431848	1.48672E-06
13	-2.8	0.289691553	0.007915452	3.9613E-06
14	-2.6	0.333224603	0.013582969	1.01409E-05
15	-2.4	0.36827014	0.02239453	2.49425E-05
16	-2.2	0.391042694	0.035474593	5.89431E-05
17	-2	0.39894228	0.053990967	0.00013383

期待値を変えても分散はそのままなので，山の形はかわりません．そして，期待値が 2 大きくなると，山が 2 だけ右に移動することが確認できます．

問題 6-2

セル B2 に「=NORM.DIST(A2,0,1/3,FALSE)」と入力し，期待値 0，標準偏差 1/3 である正規分布の確率密度関数の $x = -5$ における値を求めましょう．

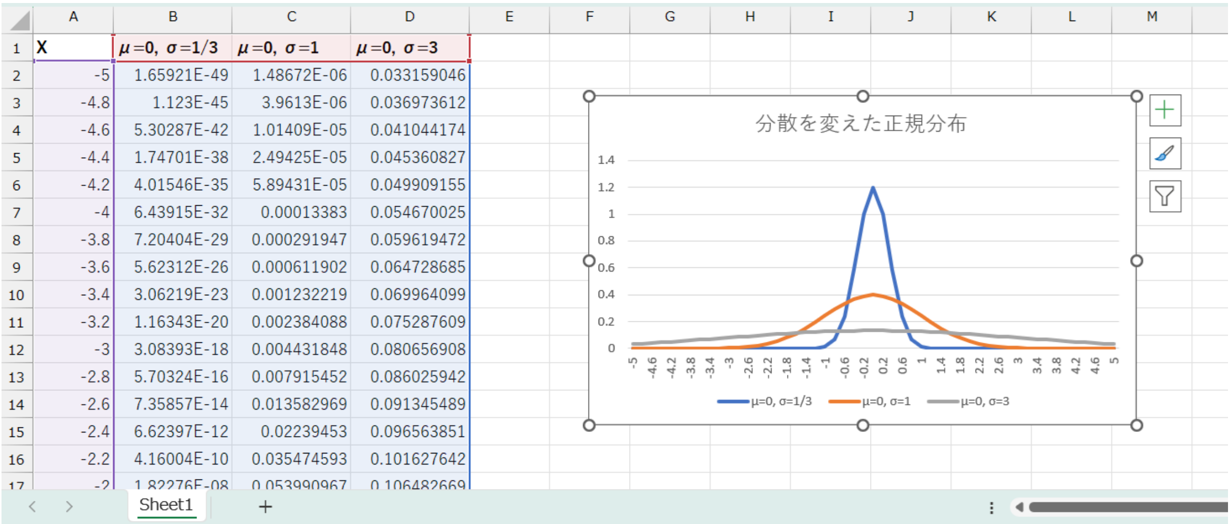
セル C2 に「=NORM.DIST(A2,0,1,FALSE)」と入力し，期待値 0，標準偏差 1 である正規分布の確率密度関数の $x=-5$ における値を求めましょう．

セル D2 に「=NORM.DIST(A2,0,3,FALSE)」と入力し，期待値 0，標準偏差 3 である正規分布の確率密度関数の $x=-5$ における値を求めましょう．

そして，セル範囲 B2:D2 を選択し，その右下あたりにマウスポインタを合わせ「+」の形にして，この状態のままダブルクリックします．すると，$x=5$ における値までオートフィルで求めることができます．

セル範囲 B1:D52 を選択したあと，挿入タブの（グラフグループにある）[折れ線/面グラフの挿入］の「2-D 折れ線」の「折れ線」を選びます．作成されたグラフが選択されたまま，グラフのデザインタブの（データグループにある）[データの選択］をクリックします．「データソースの選択」ダイアログボックスが出てくるので，横（項目）軸ラベルの「編集」をクリックします．「軸ラベル」ダイアログボックスが出てくるので，A 列の該当箇所（A2:A52）をドラッグして表示させて「OK」を押します．さらに，「データソースの選択」ダイアログボックスでも「OK」を押すと，グラフの横軸の値が A 列の値に変わります．

グラフタイトルは「分散を変えた正規分布」にしましょう．



	A	B	C	D
1	X	μ=0, σ=1/3	μ=0, σ=1	μ=0, σ=3
2	-5	1.65921E-49	1.48672E-06	0.033159046
3	-4.8	1.123E-45	3.9613E-06	0.036973612
4	-4.6	5.30287E-42	1.01409E-05	0.041044174
5	-4.4	1.74701E-38	2.49425E-05	0.045360827
6	-4.2	4.01546E-35	5.89431E-05	0.049909155
7	-4	6.43915E-32	0.00013383	0.054670025
8	-3.8	7.20404E-29	0.000291947	0.059619472
9	-3.6	5.62312E-26	0.000611902	0.064728685
10	-3.4	3.06219E-23	0.001232219	0.069964099
11	-3.2	1.16343E-20	0.002384088	0.075287609
12	-3	3.08393E-18	0.004431848	0.080656908
13	-2.8	5.70324E-16	0.007915452	0.086025942
14	-2.6	7.35857E-14	0.013582969	0.091345489
15	-2.4	6.62397E-12	0.02239453	0.096563851
16	-2.2	4.16004E-10	0.035474593	0.101627642
17	-2	1.82276E-08	0.053990967	0.106482669

分散を変えても期待値はそのままなので，山の位置は変わりません．そして，分散が小さくなると山が高く，すそがうすくなり，分散が大きくなると山が低く，すそが厚くなることが確認できます．

6.1 の確認テスト

問 1

(1) 3，1^2（または，1）

(2) 16，12^2（または，144）

問 2

正規分布のグラフにおいて，分散が小さくなると山が高くなります．また，分散が大きくなると山が低くなります．

問 3

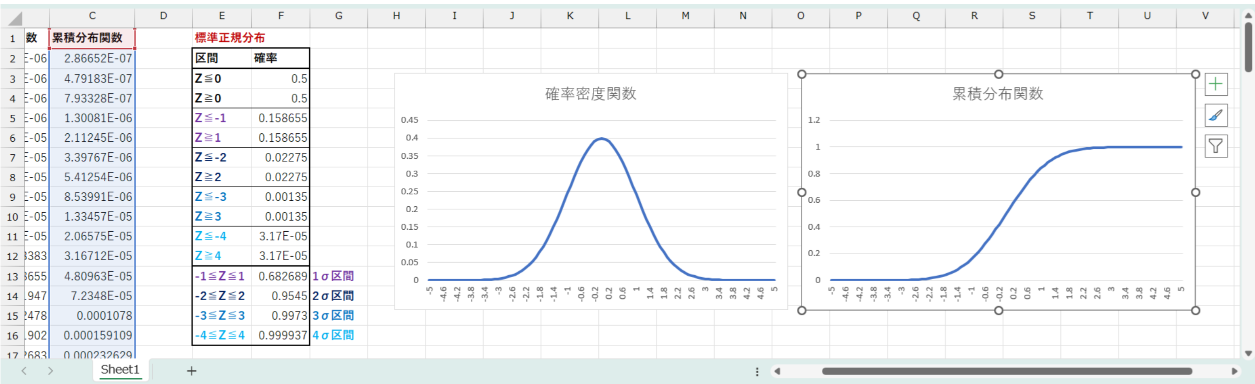
正規分布のグラフにおいて，分散が小さくなると山のすそはうすくなります．また，分散が大きくなると山のすそは厚くなります．

6.2　標準正規分布

問題 6-3

セル範囲 B1:B102 を選択したあと，挿入タブの（グラフグループにある）［折れ線/面グラフの挿入］の「2-D 折れ線」の「折れ線」を選びます．作成されたグラフが選択されたまま，グラフのデザインタブの（データグループにある）［データの選択］をクリックします．「データソースの選択」ダイアログボックスが出てくるので，横（項目）軸ラベルの「編集」をクリックします．「軸ラベル」ダイアログボックスが出てくるので，A 列の該当箇所（A2:A102）をドラッグして表示させて「OK」を押します．さらに，「データソースの選択」ダイアログボックスでも「OK」を押すと，グラフの横軸の値が A 列の値に変わります．

セル範囲 C1:C102 を選択したあと，挿入タブの（グラフグループにある）［折れ線/面グラフの挿入］の「2-D 折れ線」の「折れ線」を選びます．作成されたグラフが選択されたまま，グラフのデザインタブの（データグループにある）［データの選択］をクリックします．「データソースの選択」ダイアログボックスが出てくるので，横（項目）軸ラベルの「編集」をクリックします．「軸ラベル」ダイアログボックスが出てくるので，A 列の該当箇所（A2:A102）をドラッグして表示させて「OK」を押します．さらに，「データソースの選択」ダイアログボックスでも「OK」を押すと，グラフの横軸の値が A 列の値に変わります．



問題 6-4

セル E13 に「=I6」，セル G13 に「=I7」と入力し，「下位 5%と上位 5%を除いた 90%以内に入るのはどの値以上どの値以下か」を求めます．

セル E15 に「=I8」，セル G15 に「=I9」と入力し，「下位 2.5%と上位 2.5%を除いた 95%以内に入るのはどの値以上どの値以下か」を求めます．

セル E17 に「=I10」，セル G17 に「=I11」と入力し，「下位 0.5%と上位 0.5%を除いた 99%以内に入るのはどの値以上どの値以下か」を求めます．

	A	B	C	D	E	F	G	H	I	J	K	L
1	Z	確率密度関数	累積分布関数		標準正規分布							
2	-5	1.48672E-06	2.86652E-07		下位50%以下に入るのはどの値以下か				0		P(Z≦0)	0.5
3	-4.8	3.9613E-06	7.93328E-07		上位50%以上に入るのはどの値以上か				0		P(Z≧0)	0.5
4	-4.6	1.01409E-05	2.11245E-06		下位15.86553%以下に入るのはどの値以下か				-1		P(Z≦-1)	0.1586553
5	-4.4	2.49425E-05	5.41254E-06		上位15.86553%以上に入るのはどの値以上か				1		P(Z≧1)	0.1586553
6	-4.2	5.89431E-05	1.33457E-05		下位5%以下に入るのはどの値以下か				-1.64485			
7	-4	0.00013383	3.16712E-05		上位5%以上に入るのはどの値以上か				1.644854			
8	-3.8	0.000291947	7.2348E-05		下位2.5%以下に入るのはどの値以下か				-1.95996			
9	-3.6	0.000611902	0.000159109		上位2.5%以上に入るのはどの値以上か				1.959964			
10	-3.4	0.001232219	0.000336929		下位0.5%以下に入るのはどの値以下か				-2.57583			
11	-3.2	0.002384088	0.000687138		上位0.5%以上に入るのはどの値以上か				2.575829			
12	-3	0.004431848	0.001349898		下位5%と上位5%を除いた90%以内に入るのはどの値以上どの値以下か							
13	-2.8	0.007915452	0.00255513		-1.64485	以上	1.6448536	以下				
14	-2.6	0.013582969	0.004661188		下位2.5%と上位2.5%を除いた95%以内に入るのはどの値以上どの値以下か							
15	-2.4	0.02239453	0.008197536		-1.95996	以上	1.959964	以下				
16	-2.2	0.035474593	0.013903448		下位0.5%と上位0.5%を除いた99%以内に入るのはどの値以上どの値以下か							
17	-2	0.053990967	0.022750132		-2.57583	以上	2.5758293	以下				
18	-1.8	0.078950158	0.035930319									

Sheet1

問題 6-5

セル D2 に「=(A2-60)/5」と入力し，点数 0 の z 値を求めます．そして，このセル（B2）を下へドラッグし，点数 100 の z 値までオートフィルで求めましょう．

求めた Z は標準正規分布 $N(0,1^2)$ に従います．

セル E2 に「=NORM.S.DIST(D2,TRUE)」と入力し，標準正規分布の累積分布関数の $z = -12$ における値を求めましょう．そして，このセル（E2）を下へドラッグし，$z = 8$ における値までオートフィルで求めましょう．

B 列の「X の累積分布関数の値」と，E 列で求めた「Z の累積分布関数の値」が一致することが確かめられます．

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	点数X	累積分布関数（X）		Z	累積分布関数（Z）		区間（X）	確率			区間（Z）	確率	
2	0	1.77648E-33		-12	1.77648E-33		X≦60	0.5			Z≦0	0.5	
3	1	1.95156E-32		-11.8	1.95156E-32		X≧60	0.5			Z≧0	0.5	
4	2	2.06039E-31		-11.6	2.06039E-31		X≦55	0.158655			Z≦-1	0.158655	
5	3	2.0906E-30		-11.4	2.0906E-30		X≧65	0.158655			Z≧1	0.158655	
6	4	2.03867E-29		-11.2	2.03867E-29		X≦50	0.02275			Z≦-2	0.02275	
7	5	1.91066E-28		-11	1.91066E-28		X≧70	0.02275			Z≧2	0.02275	
8	6	1.72102E-27		-10.8	1.72102E-27		X≦45	0.00135			Z≦-3	0.00135	
9	7	1.4899E-26		-10.6	1.4899E-26		X≧75	0.00135			Z≧3	0.00135	
10	8	1.23967E-25		-10.4	1.23967E-25		X≦40	3.17E-05			Z≦-4	3.17E-05	
11	9	9.91363E-25		-10.2	9.91363E-25		X≧80	3.17E-05			Z≧4	3.17E-05	
12	10	7.61985E-24		-10	7.61985E-24		55≦X≦65	0.682689	1σ区間		-1≦Z≦1	0.682689	1σ区間
13	11	5.62928E-23		-9.8	5.62928E-23		50≦X≦70	0.9545	2σ区間		-2≦Z≦2	0.9545	2σ区間
14	12	3.99722E-22		-9.6	3.99722E-22		45≦X≦75	0.9973	3σ区間		-3≦Z≦3	0.9973	3σ区間
15	13	2.72815E-21		-9.4	2.72815E-21		40≦X≦80	0.999937	4σ区間		-4≦Z≦4	0.999937	4σ区間
16	14	1.78975E-20		-9.2	1.78975E-20								
17	15	1.12859E-19		-9	1.12859E-19								

Sheet1

問題 6-6

セル E2 に「=5*F2+60」と入力し，z 値「0」の標準化前の点数を求めます．そして，このセル（E2）を 11 行目まで下へドラッグし，オートフィルしましょう．

	A	B	C	D	E	F	G	H	I	J	K	L
1	点数X	Z			点数X	Z		区間（X）	確率		区間（Z）	確率
2	0	-12		下位50%以下に入るのは何点以下か	60	0		X≦60	0.5		Z≦0	0.5
3	1	-11.8		上位50%以上に入るのは何点以上か	60	0		X≧60	0.5		Z≧0	0.5
4	2	-11.6		下位15.86553%以下に入るのは何点以下か	55	-1		X≦55	0.158655		Z≦-1	0.158655
5	3	-11.4		上位15.86553%以上に入るのは何点以上か	65	1		X≧65	0.158655		Z≧1	0.158655
6	4	-11.2		下位5%以下に入るのは何点以下か	51.77573	-1.64485						
7	5	-11		上位5%以上に入るのは何点以上か	68.22427	1.644854						
8	6	-10.8		下位2.5%以下に入るのはどの値以下か	50.20018	-1.95996						
9	7	-10.6		上位2.5%以上に入るのはどの値以上か	69.79982	1.959964						
10	8	-10.4		下位0.5%以下に入るのは何点以下か	47.12085	-2.57583						
11	9	-10.2		上位0.5%以上に入るのは何点以上か	72.87915	2.575829						
12	10	-10										
13	11	-9.8										
14	12	-9.6										
15	13	-9.4										
16	14	-9.2										
17	15	-9										

Sheet1

問題 6-7

セル E2 に「=NORM.INV(0.5,60,5)」と入力しなおし，「下位 50%以下に入るのはどの値以下か」を求めます．

なお，問題 5-5 で，「=NORM.DIST(60,60,5,TRUE)」と入力したら，0.5 が返されました．これにより，正規分布 $N(60, 5^2)$ の累積分布関数の $x = 60$ における値が 0.5 であることが確認できました．すなわち，逆関数については，累積分布関数の値 0.5 に対応する値は 60 になるので，「=NORM.INV(0.5,60,5)」は 60 になるはずです．

セル E3 に「=NORM.INV(1-0.5,60,5)」と入力しなおし，「上位 50%以上に入るのはどの値以上か」を求めます．

セル E4 に「=NORM.INV(0.1586553,60,5)」と入力しなおし，「下位 15.86553%以下に入るのはどの値以下か」を求めます．

セル E5 に「=NORM.INV(1-0.1586553,60,5)」と入力しなおし，「上位 15.86553%以上に入るのはどの値以上か」を求めます．

セル E6 に「=NORM.INV(0.05,60,5)」と入力しなおし，「下位 5%以下に入るのはどの値以下か」を求めます．

セル E7 に「=NORM.INV(1-0.05,60,5)」と入力しなおし，「上位 5%以上に入るのはどの値以上か」を求めます．

セル E8 に「=NORM.INV(0.025,60,5)」と入力しなおし，「下位 2.5%以下に入るのはどの値以下か」を求めます．

セル E9 に「=NORM.INV(1-0.025,60,5)」と入力しなおし，「上位 2.5%以上に入るのはどの値以上か」を求めます．

セル E10 に「=NORM.INV(0.005,60,5)」と入力しなおし，「下位 0.5%以下に入るのはどの値以下か」を求めます．

セル E11 に「=NORM.INV(1-0.005,60,5)」と入力しなおし，「上位 0.5%以上に入るのはどの値以上か」を求めます．

以上において，まったく値が変わらないことがわかります．

6.2 の確認テスト

問 1

(1) 0，1．1^2（または，1）

(2) 上側確率，下側確率，両側確率

(3) 50

(4) 上側 c%点，下側 c%点，両側 c%点

(5) 標準化，z 値

(6) 0，1

(7) 0，1^2（または，1）

(8) 10，6

問 2

X から Z へ標準化したとき，$Z = (X - \mu)/\sigma$ とあらわされ，また，$X = \sigma Z + \mu$ とあらわされます．

問 3

1σ 区間は $-1 \leq Z \leq 1$，　2σ 区間は $-2 \leq Z \leq 2$，
3σ 区間は $-3 \leq Z \leq 3$，　4σ 区間は $-4 \leq Z \leq 4$

問 4

-1 についての下側確率は 16%，1 についての両側確率は 32%

問 5

下側 5%点は -1.64，両側 10%点は 1.64

問 6

下側 2.5%点は -1.96，両側 5%点は 1.96

問 7

(1) 点数 30 の z 値は，$(30 - 45)/15 = -1$ と計算され，-1 となります．また，点数 51 の z 値は，$(51 - 45)/15 = 0.4$ と計算され，0.4 となります．

(2) z 値が 1 であるときの標準化前の点数は，$15 \times 1 + 45 = 60$ と計算され，60 となります．また，z 値が -2 であるときの標準化前の点数は，$15 \times (-2) + 45 = 15$ と計算され，15 となります．

問 8

(1) 身長 178 cm の z 値は，$(178 - 170)/8 = 1$ と計算され，1 となります．また，身長 166 cm の z 値は，$(166 - 170)/8 = -0.5$ と計算され，-0.5 となります．

(2) z 値が 0.5 であるときの標準化前の身長は，$8 \times 0.5 + 170 = 174$ と計算され，174 cm となります．また，z 値が -0.25 であるときの標準化前の身長は，$8 \times (-0.25) + 170 = 168$ と計算され，168 cm となります．

(3) 標準正規分布に従う確率変数 Z について，上位 50%以上に入るのは 0 以上です．
z 値が 0 であるときの標準化前の身長は，$8 \times 0 + 170 = 170$ と計算され，170 cm となります．
よって，上位 50%以上に入るのは，身長 170 cm 以上ということになります．

(4) 標準正規分布に従う確率変数 Z について，上位 15.9%以上に入るのは 1 以上であることがわかっています．
z 値が 1 であるときの標準化前の身長は，$8 \times 1 + 170 = 178$ と計算され，178 cm となります．
よって，上位 15.9%以上に入るのは，身長 178 cm 以上ということになります．

(5) 標準正規分布に従う確率変数 Z について，上位 15.9%以上に入るのは 1 以上であることがわかっていて，標準正規分布は期待値 0 を中心として左右対称なので，下位 15.9%以下に入るのは -1 以下となります．

z 値が -1 であるときの標準化前の身長は，$8 \times (-1) + 170 = 162$ と計算され，162 cm となります．

よって，下位 15.9%以下に入るのは，身長 162 cm 以下ということになります．

第 7 章　点推定

7.1　母数と推定量

問題 7-1

次のように入力します．

G	H	I	J	K	L	M	N	O	P	Q	R	S	T
標本平均			50.64		標本平均					標本平均			
標本分散			310.6304		標本分散					標本分散			
標本標準偏差			17.62471		標本標準偏差					標本標準偏差			
不偏分散					不偏分散					不偏分散			
不偏分散による標準偏差					不偏分散による標準偏差					不偏分散による標準偏差			
標準誤差					標準誤差					標準誤差			
標本1					標本2					標本3			
学生番号	学生所属	年次	点数		学生番号	学生所属	年次	点数		学生番号	学生所属	年次	点数
27-8063	法学部	4	44		28-7044	経営学部	3	40		28-6063	経済学部	3	46
29-7002	経営学部	2	20		30-9034	人間生活訁	1	18		28-9555	人間生活訁	3	50
28-6009	経済学部	3	90		29-6117	経済学部	2	58		29-6008	経済学部	2	46
28-7110	経営学部	3	70		30-8026	法学部	1	58		28-6091	経済学部	3	50
27-7102	経営学部	4	44		27-6096	経済学部	4	32		27-6034	経済学部	4	66
29-7069	経営学部	2	50		29-6096	経済学部	2	58		30-7043	経営学部	1	44
30-7062	経営学部	1	90		28-8078	法学部	3	50		29-6106	経済学部	2	80
30-7122	経営学部	1	40		27-6108	経済学部	4	48		28-7062	経営学部	3	50

セル O1 に標本 2 の点数の平均値，つまり，標本平均を AVERAGE 関数で求めます（セルには「=AVERAGE(O10:O109)」と入力されます）．

セル O2 に標本 2 の点数の分散，つまり，標本分散を VAR.P 関数で求めます（セルには「=VAR.P(O10:O109)」と入力されます）．

セル O3 に標本 2 の点数の標準偏差，つまり，標本標準偏差を STDEV.P 関数で求めます（セルには「=STDEV.P(O10:O109)」と入力されます）．

また，セル T1 に標本 3 の点数の平均値，つまり，標本平均を AVERAGE 関数で求めます（セルには「=AVERAGE(T10:T159)」と入力されます）．

セル T2 に標本 3 の点数の分散，つまり，標本分散を VAR.P 関数で求めます（セルには「=VAR.P(T10:T159)」と入力されます）．

セル T3 に標本 3 の点数の標準偏差，つまり，標本標準偏差を STDEV.P 関数で求めます（セルには「=STDEV.P(T10:T159)」と入力されます）．

G	H	I	J
標本平均			50.64
標本分散			310.6304
標本標準偏差			17.62471
不偏分散			
不偏分散による標準偏差			
標準誤差			
標本1			
学生番号	学生所属	年次	点数
27-8063	法学部	4	44
29-7002	経営学部	2	20
28-6009	経済学部	3	90
28-7110	経営学部	3	70
27-7102	経営学部	4	44
29-7069	経営学部	2	50
30-7062	経営学部	1	90
30-7122	経営学部	1	40

L	M	N	O
標本平均			50.88
標本分散			355.5456
標本標準偏差			18.85592
不偏分散			
不偏分散による標準偏差			
標準誤差			
標本2			
学生番号	学生所属	年次	点数
28-7044	経営学部	3	40
30-9034	人間生活部	1	18
29-6117	経済学部	2	58
30-8026	法学部	1	58
27-6096	経済学部	4	32
29-6096	経済学部	2	58
28-8078	法学部	3	50
27-6108	経済学部	4	48

Q	R	S	T
標本平均			50.52
標本分散			377.3563
標本標準偏差			19.42566
不偏分散			
不偏分散による標準偏差			
標準誤差			
標本3			
学生番号	学生所属	年次	点数
28-6063	経済学部	3	46
28-9555	人間生活部	3	50
29-6008	経済学部	2	46
28-6091	経済学部	3	50
27-6034	経済学部	4	66
30-7043	経営学部	1	44
29-6106	経済学部	2	80
28-7062	経営学部	3	50

上記の計算結果は一例です．

問題 7-2

セル O4 に標本 2 の点数の不偏分散を VAR.S 関数で求めます（セルには「=VAR.S(O10:O109)」と入力されます）．

セル O5 に標本 2 の点数の不偏分散による標準偏差を STDEV.S 関数で求めます（セルには「=STDEV.S(O10:O109)」と入力されます）．

また，セル T4 に標本 3 の点数の不偏分散を VAR.S 関数で求めます（セルには「=VAR.S(T10:T159)」と入力されます）．

セル T5 に標本 3 の点数の不偏分散による標準偏差を STDEV.S 関数で求めます（セルには「=STDEV.S(T10:T159)」と入力されます）．

G	H	I	J
標本平均			50.64
標本分散			310.6304
標本標準偏差			17.62471
不偏分散			316.9698
不偏分散による標準偏差			17.80365
標準誤差			
標本1			
学生番号	学生所属	年次	点数
27-8063	法学部	4	44
29-7002	経営学部	2	20
28-6009	経済学部	3	90
28-7110	経営学部	3	70
27-7102	経営学部	4	44
29-7069	経営学部	2	50
30-7062	経営学部	1	90
30-7122	経営学部	1	40

L	M	N	O
標本平均			50.88
標本分散			355.5456
標本標準偏差			18.85592
不偏分散			359.137
不偏分散による標準偏差			18.95091
標準誤差			
標本2			
学生番号	学生所属	年次	点数
28-7044	経営学部	3	40
30-9034	人間生活部	1	18
29-6117	経済学部	2	58
30-8026	法学部	1	58
27-6096	経済学部	4	32
29-6096	経済学部	2	58
28-8078	法学部	3	50
27-6108	経済学部	4	48

Q	R	S	T
標本平均			50.52
標本分散			377.3563
標本標準偏差			19.42566
不偏分散			379.8889
不偏分散による標準偏差			19.49074
標準誤差			
標本3			
学生番号	学生所属	年次	点数
28-6063	経済学部	3	46
28-9555	人間生活部	3	50
29-6008	経済学部	2	46
28-6091	経済学部	3	50
27-6034	経済学部	4	66
30-7043	経営学部	1	44
29-6106	経済学部	2	80
28-7062	経営学部	3	50

上記の計算結果は一例です．

問題 7-3

セル O6 に「=O5/SQRT(100)」と入力すると，標本 2 の点数の標準誤差が求められます．

また，セル T6 に「=T5/SQRT(150)」と入力すると，標本 3 の点数の標準誤差が求められます．

G	H	I	J	K	L	M	N	O	P	Q	R	S	T
標本平均			50.64		標本平均			50.88		標本平均			50.52
標本分散			310.6304		標本分散			355.5456		標本分散			377.3563
標本標準偏差			17.62471		標本標準偏差			18.85592		標本標準偏差			19.42566
不偏分散			316.9698		不偏分散			359.137		不偏分散			379.8889
不偏分散による標準偏差			17.80365		不偏分散による標準偏差			18.95091		不偏分散による標準偏差			19.49074
標準誤差			2.517816		標準誤差			1.895091		標準誤差			1.591412
標本1					標本2					標本3			
学生番号	学生所属	年次	点数		学生番号	学生所属	年次	点数		学生番号	学生所属	年次	点数
27-8063	法学部	4	44		28-7044	経営学部	3	40		28-6063	経済学部	3	46
29-7002	経営学部	2	20		30-9034	人間生活部	1	18		28-9555	人間生活部	3	50
28-6009	経済学部	3	90		29-6117	経済学部	2	58		29-6008	経済学部	2	46
28-7110	経営学部	3	70		30-8026	法学部	1	58		28-6091	経済学部	3	50
27-7102	経営学部	4	44		27-6096	経済学部	4	32		27-6034	経済学部	4	66
29-7069	経営学部	2	50		29-6096	経済学部	2	58		30-7043	経営学部	1	44
30-7062	経営学部	1	90		28-8078	法学部	3	50		29-6106	経済学部	2	80
30-7122	経営学部	1	40		27-6108	経済学部	4	48		28-7062	経営学部	3	50

上記の計算結果は一例です.

7.1 の確認テスト

問 1

(1) 母平均，母分散，母標準偏差

(2) 母数

(3) 統計的推測

(4) 推定，検定（または，仮説検定）

(5) 統計量，標本平均，標本分散，標本標準偏差

(6) 不偏分散

(7) 不偏分散

(8) 推定量，一致推定量，不偏推定量

(9) 母平均，母分散

(10) 母平均，母分散

(11) 標準誤差

問 2

標本サイズが大きいほど，標準誤差は小さくなります.

また，標本のばらつきが大きいほど，標準誤差は大きくなります.

問 3

母分散の不偏推定量であり，母集団のばらつきを標本から推定する場合に使われるのは，不偏分散です.

問 4

標本標準偏差は，標本分散の正の平方根をとったものなので，$\sqrt{81} = 9$ より，9 であることがわかります.

変動係数は，標本標準偏差を平均値で割ったものなので，

$$\frac{9}{2} = 4.5$$

より，4.5 となります.

また,

$$\text{標準誤差} = \sqrt{\frac{\text{不偏分散}}{\text{標本サイズ}}} = \sqrt{\frac{90}{10}} = \sqrt{9} = 3$$

より，標準誤差は 3 となります．

問 5

標本標準偏差を平均値で割ったものが変動係数なので，標本標準偏差は，平均値と変動係数をかけたものになります．よって，標本標準偏差は,

$$4 \times 0.5 = 2$$

より，2 であることがわかります．

標本分散は，標本標準偏差を 2 乗したものなので，$2^2 = 4$ より，4 となります．

不偏分散は，標本分散に「(データの個数)/(データの個数 −1)」をかけたものなので,

$$4 \times \frac{5}{5-1} = 5$$

より，5 となります．

また,

$$\text{標準誤差} = \sqrt{\frac{\text{不偏分散}}{\text{標本サイズ}}} = \sqrt{\frac{5}{5}} = 1$$

より，標準誤差は 1 となります．

問 6

$$\text{標準誤差} = \sqrt{\frac{\text{不偏分散}}{\text{標本サイズ}}}$$

より，$\sqrt{\text{不偏分散}}$ は，標準誤差と $\sqrt{\text{標本サイズ}}$ をかけたものとなります．よって，$\sqrt{\text{不偏分散}}$ は,

$$2 \times \sqrt{10} = \sqrt{40}$$

より，$\sqrt{40}$ となります．

これより，不偏分散は $\sqrt{40}^2 = 40$ より，40 となります．

標本分散は，不偏分散に「(データの個数 −1)/(データの個数)」をかけたものなので,

$$40 \times \frac{10-1}{10} = 36$$

より，36 となります．

よって，標本標準偏差は，$\sqrt{36} = 6$ より，6 であることがわかります．

また，標本標準偏差を平均値で割ったものが変動係数なので，平均値は標本標準偏差を変動係数で割って求めることができます．よって，平均値は,

$$\frac{6}{1.2} = 5$$

より，5 となります．

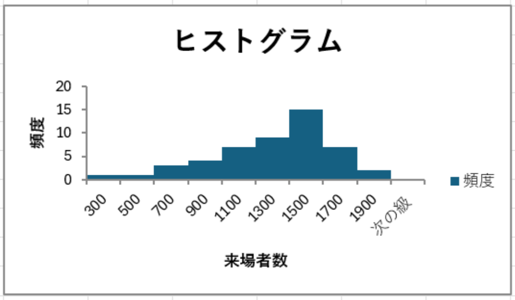
7.2　分布の形状をあらわす基本統計量

問題 7-4

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，分析ツールの「基本統計量」を選びます．［データ分析］ボタンが見あたらないときは，ファイルタブの（「その他...」の）「オプション」から「アドイン」を選択します．下のほうにある「Excel アドイン」の右にある「設定...」をクリックして，「分析ツール」にチェックを入れましょう．

「基本統計量」ダイアログボックスの「入力範囲」を入れるところにカーソルをおいて，セル C1 からセル C50 をドラッグして指定します．データ方向は「列」のままです．先頭行のセル C1「来場者数」はラベルとして扱うので，「先頭行をラベルとして使用」にチェックを入れます．

また，「出力先」を選択し，すぐ右のそれを入れるところにカーソルをおいて，セル T1 をクリックして指定します．そして，「統計情報」にチェックを入れて OK を押しましょう．



来場者数	
平均	1215.224
標準誤差	52.45363
中央値 （メジアン）	1298
最頻値 （モード）	1620
標準偏差	367.1754
分散	134817.8
尖度	0.066743
歪度	-0.65182
範囲	1605
最小	288
最大	1893
合計	59546
データの個数	49

すると，上記のような，平均，標準誤差，中央値，最頻値，標準偏差，分散，尖度，歪度，範囲，最小，最大，合計，データの個数のそれぞれの値の一覧が出てきます．

7.2 の確認テスト

問 1

(1) 尖度，0
(2) 歪度，0
(3) 最大値，最小値
(4) 範囲，分散，標準偏差，標準誤差（以上は順序がちがってもいい）
(5) 変動係数，尖度，歪度（以上は順序がちがってもいい）

問 2

尖度が大きいほど，「ピークがとがっている」分布になります．

問 3

歪度が正のとき，「ピークが左にかたよっている」分布になります．

第 8 章　中心極限定理

8.1　大数の法則

問題 8-1

　この問題では，コインを投げたときに表が出たら 1，裏が出たら 0 を割り当てる対応 X の確率分布をとるような母集団を想定しています（4.2 の確認テスト　問 3 より，X の期待値は 0.5，分散は 0.25）．母平均は 0.5，母分散は 0.25 ということになります．

x	0	1
$P(X=x)$	1/2	1/2

　このような母集団から大きさ 4 の標本を 100 回無作為抽出することを，「RANDBETWEEN 関数を使って 0 または 1 のどちらかを 4 行 100 列分ランダムに取り出す」ことで代用します．それぞれの標本について，標本平均を求めて，グラフを作成することにより母平均 0.5 と比較してみましょう．

「コインを 4 回投げる」という試行を 100 回おこないます．

　セル B2 に「=RANDBETWEEN(0,1)」と入力し，このセル（B2）を 5 行目まで下にドラッグしてオートフィルします．さらにそのまま，選択されているセル範囲 B2:B5 の右下あたりにマウスポインタを合わせ「+」の形にして，この状態のまま CW 列目まで右にドラッグし，オートフィルします．すると，0 または 1 のどちらかがランダムに 4 行 100 列分，つまり，400 個発生します．

　これで，母集団から大きさ 4 の標本を 100 回無作為抽出したことになります．

　つぎに，「コインを 4 回投げる」という 100 回おこなわれた各試行について，割り当てられた数値の平均値（標本平均）をそれぞれ求めます．

　セル B8 に「=AVERAGE(B2:B5)」と入力し，このセル（B8）を CW 列目まで右にドラッグしてオートフィルします．

　これで，それぞれの標本について，標本平均を求めたことになります．

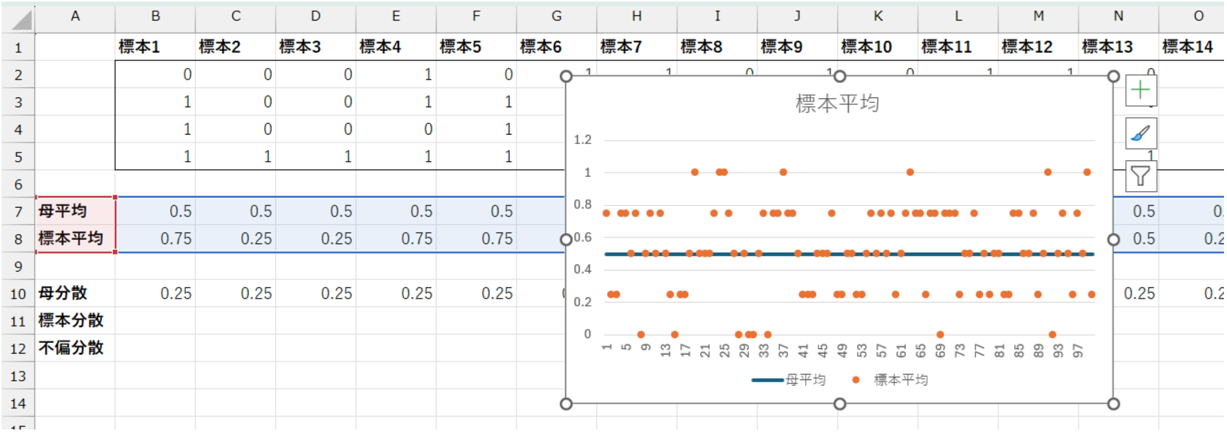
「標本平均の散布図」と「母平均の折れ線グラフ」を，同一グラフエリアにそれぞれ作成します．

　そのため，セル範囲 A7:CW8 を選択し，挿入タブの（グラフグループにある）［折れ線/面グラフの挿入］の「2-D 折れ線」の「折れ線」を選びます．

　つぎに，作成したグラフをクリックして選択し，その状態で，グラフのデザインタブの（種類グループにある）［グラフの種類の変更］を選びます．

　すると，「グラフの種類の変更」ダイアログボックスが出てくるので，「標本平均」のグラフの種類を「散布図」に変更しましょう．

　また，グラフタイトルは「標本平均」にしましょう．



標本平均は母平均の付近に集まっていることが確かめられます．

問題 8-2

「コインを 4 回投げる」という 100 回おこなわれた各試行について，割り当てられた数値の標本分散をそれぞれ求めます．

セル B11 に「=VAR.P(B2:B5)」と入力し，このセル（B11）を CW 列目まで右にドラッグしてオートフィルします．

これで，それぞれの標本について，標本分散を求めたことになります．

「コインを 4 回投げる」という 100 回おこなわれた各試行について，割り当てられた数値の不偏分散をそれぞれ求めます．

セル B12 に「=VAR.S(B2:B5)」と入力し，このセル（B12）を CW 列目まで右にドラッグしてオートフィルします．

これで，それぞれの標本について，不偏分散を求めたことになります．

「標本分散の散布図」と「母分散の折れ線グラフ」を，同一グラフエリアにそれぞれ作成します．

そのため，セル範囲 A10:CW11 を選択し，挿入タブの（グラフグループにある）［折れ線/面グラフの挿入］の「2-D 折れ線」の「折れ線」を選びます．

つぎに，作成したグラフをクリックして選択し，その状態で，グラフのデザインタブの（種類グループにある）［グラフの種類の変更］を選びます．

すると，「グラフの種類の変更」ダイアログボックスが出てくるので，「標本分散」のグラフの種類を「散布図」に変更しましょう．

また，グラフタイトルは「標本分散」にしましょう．

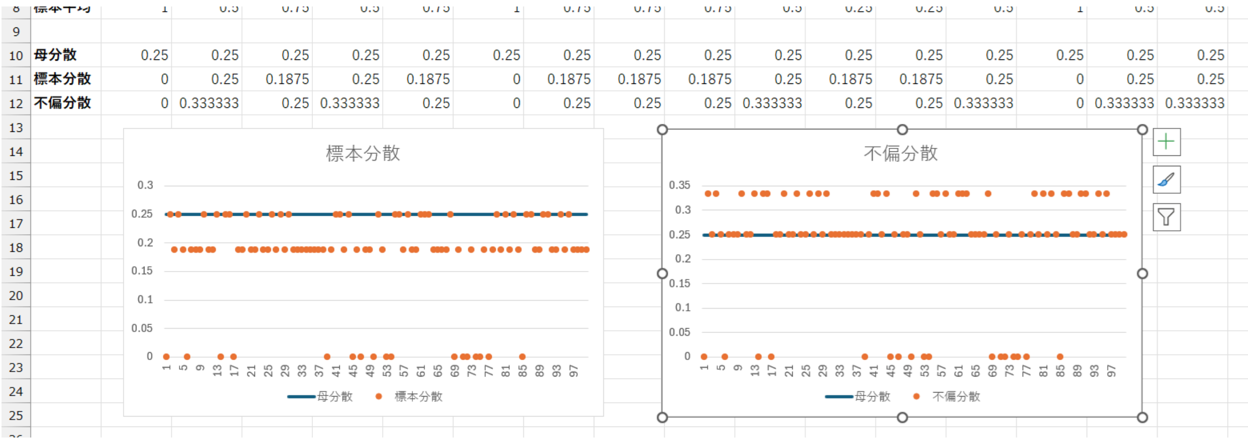
次に，「不偏分散の散布図」と「母分散の折れ線グラフ」を，同一グラフエリアにそれぞれ作成します．

そのため，セル範囲 A10:CW10 を選択し，Ctrl キーを押しながら，セル範囲 A12:CW12 を選択したあと，挿入タブの（グラフグループにある）［折れ線/面グラフの挿入］の「2-D 折れ線」の「折れ線」を選びます．

つぎに，作成したグラフをクリックして選択し，その状態で，グラフのデザインタブの（種類グループにある）［グラフの種類の変更］を選びます．

すると，「グラフの種類の変更」ダイアログボックスが出てくるので，「不偏分散」のグラフの種類を「散布図」に変更しましょう．

また，グラフタイトルは「不偏分散」にしましょう．



不偏分散は母分散の付近に集まっていることが確かめられます．一方，標本分散は，母分散と比較すると全体的に小さくなっていることがわかります．

問題 8-3

「コインを 40 回投げる」という試行を 100 回おこないます．

セル B2 に「=RANDBETWEEN(0,1)」と入力し，このセル（B2）を 41 行目まで下にドラッグしてオートフィルします．さらにそのまま，選択されているセル範囲 B2:B41 の右下あたりにマウスポインタを合わせ「+」の形にして，この状態のまま CW 列目まで右にドラッグし，オートフィルします．すると，0 または 1 のどちらかがランダムに 40 行 100 列分，つまり，4000 個発生します．

これで，母集団から大きさ 40 の標本を 100 回無作為抽出したことになります．

つぎに，「コインを 40 回投げる」という 100 回おこなわれた各試行について，出た目の平均値（標本平均）をそれぞれ求めます．

セル B44 に「=AVERAGE(B2:B41)」と入力し，このセル（B44）を CW 列目まで右にドラッグしてオートフィルします．

これで，それぞれの標本について，標本平均を求めたことになります．

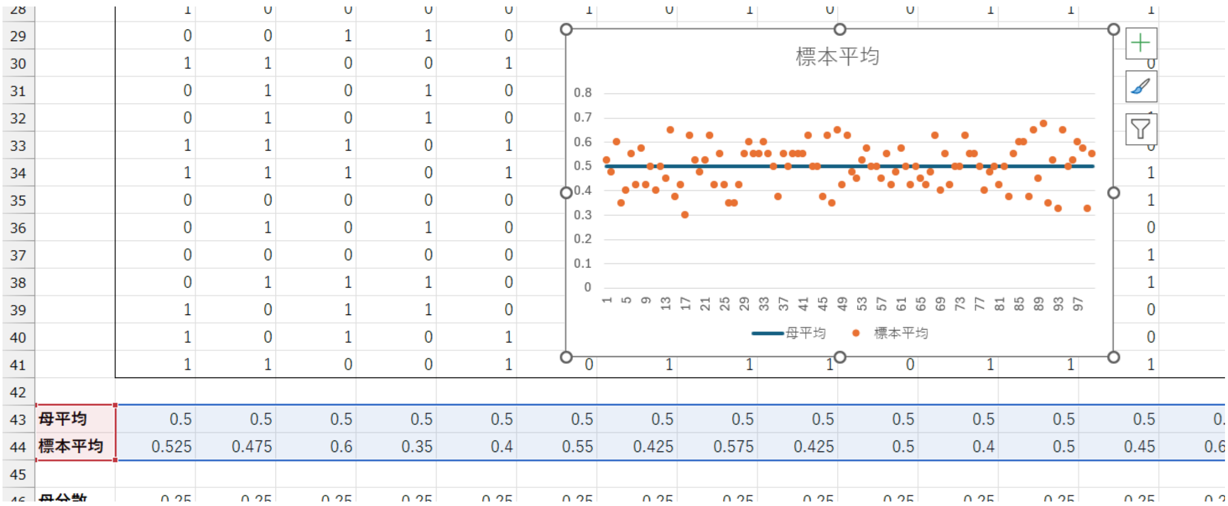
「標本平均の散布図」と「母平均の折れ線グラフ」を，同一グラフエリアにそれぞれ作成します．

そのため，セル範囲 A43:CW44 を選択し，挿入タブの（グラフグループにある）［折れ線/面グラフの挿入］の「2-D 折れ線」の「折れ線」を選びます．

つぎに，作成したグラフをクリックして選択し，その状態で，グラフのデザインタブの（種類グループにある）［グラフの種類の変更］を選びます．

すると，「グラフの種類の変更」ダイアログボックスが出てくるので，「標本平均」のグラフの種類を「散布図」に変更しましょう．

また，グラフタイトルは「標本平均」にしましょう．



標本平均は母平均の付近に集まっていることが確かめられます．また，問題 8-1 における標本平均のグラフと比べると，さらに母平均の付近に集まっていることがわかります．

問題 8-4

「コインを 40 回投げる」という 100 回おこなわれた各試行について，出た目の標本分散をそれぞれ求めます．

セル B47 に「=VAR.P(B2:B41)」と入力し，このセル（B47）を CW 列目まで右にドラッグしてオートフィルします．

これで，それぞれの標本について，標本分散を求めたことになります．

「コインを 40 回投げる」という 100 回おこなわれた各試行について，出た目の不偏分散をそれぞれ求めます．

セル B48 に「=VAR.S(B2:B41)」と入力し，このセル（B48）を CW 列目まで右にドラッグしてオートフィルします．

これで，それぞれの標本について，不偏分散を求めたことになります．

「標本分散の散布図」と「母分散の折れ線グラフ」を，同一グラフエリアにそれぞれ作成します．

そのため，セル範囲 A46:CW47 を選択し，挿入タブの（グラフグループにある）［折れ線/面グラフの挿入］の「2-D 折れ線」の「折れ線」を選びます．

つぎに，作成したグラフをクリックして選択し，その状態で，グラフのデザインタブの（種類グループにある）［グラフの種類の変更］を選びます．

すると，「グラフの種類の変更」ダイアログボックスが出てくるので，「標本分散」のグラフの種類を「散布図」に変更しましょう．

また，グラフタイトルは「標本分散」にしましょう．

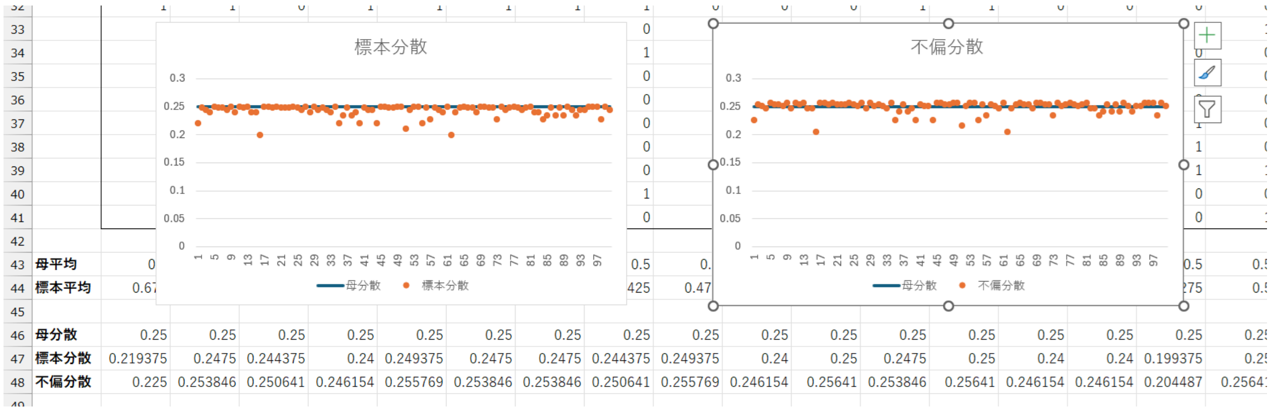
「不偏分散の散布図」と「母分散の折れ線グラフ」を，同一グラフエリアにそれぞれ作成します．

そのため，セル範囲 A46:CW46 を選択し，Ctrl キーを押しながら，セル範囲 A48:CW48 を選択したあと，挿入タブの（グラフグループにある）［折れ線/面グラフの挿入］の「2-D 折れ線」の「折れ線」を選びます．

つぎに，作成したグラフをクリックして選択し，その状態で，グラフのデザインタブの（種類グループにある）［グラフの種類の変更］を選びます．

すると，「グラフの種類の変更」ダイアログボックスが出てくるので，「不偏分散」のグラフの種類を「散布図」に変更しましょう．

また，グラフタイトルは「不偏分散」にしましょう．



不偏分散．標本分散ともに母分散の付近に集まっていることが確かめられます（よく見ると，標本分散のほうが全体的に小さくなっていますが，標本サイズが比較的大きいため目立ちません）．また，問題 8-2 における標本分散のグラフ，不偏分散のグラフと比べると，さらに母分散の付近に集まっていることがわかります．

問題 8-5

セル A1 に「試行」，セル B1 に「表が出た割合」と入力します．

セル A2 に「=RANDBETWEEN(0,1)」と入力し，このセル（A2）を 1001 行目以上まで下にドラッグしてオートフィルします．

すると，A 列に 0 または 1 の乱数が 1000 個以上発生します．

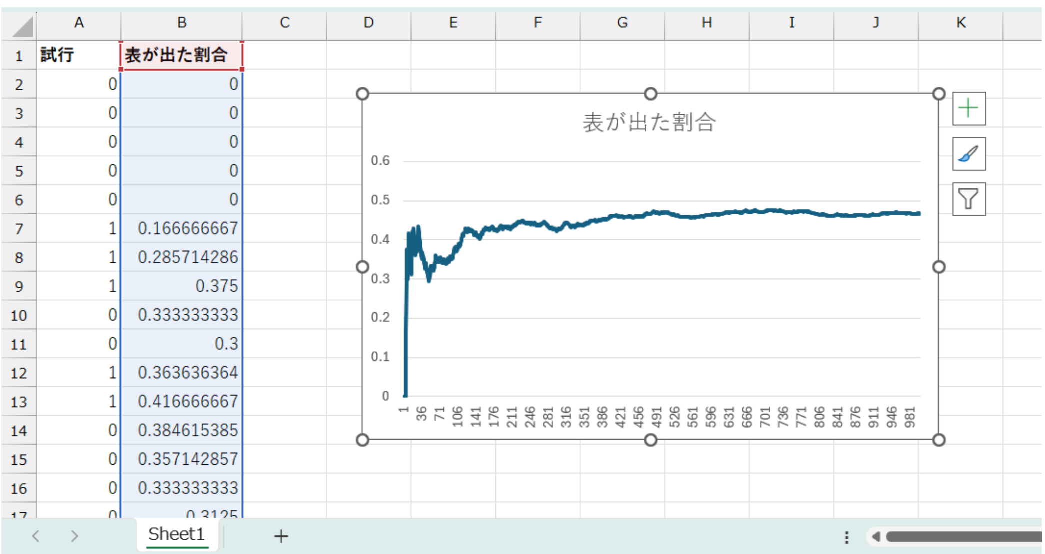
つぎに，条件をみたすセルの個数を返す COUNTIF 関数と数値が入力されているセルの個数を返す COUNT 関数を使って，各行までの A 列にある「1」と表示されているセルの割合を求めましょう．

そのため，セル B2 に「=COUNTIF($A$2:A2,1)/COUNT($A$2:A2)」と入力します．ここで，$記号は，F4 キー（設定によっては Fn キー +F4 キー）を押すことによって入力できます．

ふたたびセル B2 を選択し，そのセルの右下あたりにマウスポインタを合わせると，マウスポインタが「+」の形になります．この状態のままダブルクリックすると，オートフィルされ，セル B3 以降についても各行までの A 列にある「1」と表示されているセルの割合が返されます．

各回までの表が出た割合の折れ線グラフを作成します．

そのため，B 列を選択し，挿入タブの（グラフグループにある）［折れ線/面グラフの挿入］の「2-D 折れ線」の「折れ線」を選びます．



グラフを見ると，コインを投げる回数が増えるにつれて，表が出た割合が「表が出る確率 1/2 (= 0.5)」に収束していくのがわかります.

8.1 の確認テスト

問 1

(1) 母平均，大数の法則

(2) 1/6，3.5

(3) 1/2，1/2

問 2

ある母集団から大きさ n の標本の無作為抽出して母平均を推測する場合には，標本サイズ n が大きいほどより正確に母平均を推測できます.

8.2　中心極限定理

問題 8-6

「サイコロを 10 回振る」という試行を 1000 回おこないます.

セル B2 に「=RANDBETWEEN(1,6)」と入力し，このセル（B2）を K 列まで右にドラッグしてオートフィルします．さらにそのまま，選択されているセル範囲 B2:K2 の右下あたりにマウスポインタを合わせ「+」の形にして，この状態のままダブルクリックすることにより，標本 1000 までオートフィルします．すると，2 行目から 1001 行目に，1 から 6 までの整数の乱数が各行 10 個ずつ発生します.

つぎに，「サイコロを 10 回振る」という 1000 回おこなわれた試行について，出た目の平均値（標本平均）をそれぞれ求めます.

セル L2 に「=AVERAGE(B2:K2)」と入力し，このセル（L2）を 1001 行目までオートフィルします.

そして，標本平均の分布を求めます.

セル O2 に，「=COUNTIF($L$2:$L$1001,N2)/1000」と入力し，標本平均が「1」である標本の割合，つまり，相対度数を求めます．ここで，$記号は，式中の「L2:L1001」を入力直後に F4 キー（設定によっては Fn キー +F4 キー）を押すことによって入力できます．このセル（O2）を，52 行目までオートフィルし，標本平均が「6」である標本の相対度数まで求めましょう.

相対度数の棒グラフを作成します.

そのため，セル範囲 O2:O52 を選択し，挿入タブの（グラフグループにある）[縦棒/横棒グラフの挿入] の「2-D 縦棒」の「集合縦棒」を選びます．作成されたグラフが選択されたまま，グラフのデザインタブの（データグループにある）[データの選択] をクリックします.「データソースの選択」ダイアログボックスが出てくるので，横（項目）軸ラベルの「編集」をクリックします.「軸ラベル」ダイアログボックスが出てくるので，N 列の該当箇所（N2:N52）をドラッグして表示させて「OK」を押します．さらに，「データソースの選択」ダイアログボックスでも「OK」を押すと，グラフの横軸の値が N 列の値に変わります.

グラフタイトルは「標本平均の分布」にしましょう.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	標本平均		標本平均	相対度数
標本1	2	5	3	4	3	1	1	3	3	2	2.7		1	0
標本2	6	6	2	1	6	1	4	2	3	3	3.4		1.1	0
標本3	4	5	3	1	6	4	2	4	3	3	3.5		1.2	0
標本4	3	1	6	1	4	3	4	4	5	1	3.2		1.3	0
標本5	6	1	1	5	3	1	3	5	4	2	3.1		1.4	0
標本6	4	5	5	3	2	4	6	5	4	1	3.9		1.5	0
標本7	3	1	5	1	4	1	5	5	6	6	3.7		1.6	0
標本8	3	5	5	5	1	2	6	2	4	3	3.6		1.7	0
標本9	5	1	1	6	4	5	4	5	5	4	4		1.8	0.001
標本10	4	2	1	4	6	6	4	6	1	1	3.5		1.9	0.001
標本11	4	6	5	3	5	1	3	3	4	2	3.6		2	0
標本12	2	6	4	5	6	2	2	5	1	5	3.8		2.1	0.003
標本13	2	2	4	4	1	4	5	4	5	5	3.6		2.2	0.001
標本14	6	2	6	4	1	4	4	2	4	6	3.9		2.3	0.008
標本15	1	2	4	5	2	5	2	5	6	6	3.8		2.4	0.007
標本16	5	3	3	1	3	5	1	4	6	3	3.4		2.5	0.011
標本17	2	2	1	3	2	4	1	5	1	6	2.7		2.6	0.017

正規分布にさらにもっと似ていることが確かめられます．

問題 8-7

セル P1 に「累積相対度数」，セル Q1 に「正規分布の累積分布関数」と入力します．

標本平均の累積相対度数を求めます．

そのため，まず，セル P2 に「=O2」と入力します．また，セル P3 に「=P2+O3」と入力し，このセル（P3）を 52 行目まで下にドラッグしてオートフィルします．

正規分布 $N(3.5, 35/(12 \times 10))$ の累積分布関数の値を求めます．

そのため，セル Q2 に「=NORM.DIST(N2,3.5,SQRT(35/(12*10)),TRUE)」と入力し，このセル（Q2）を 52 行目まで下にドラッグしてオートフィルします．

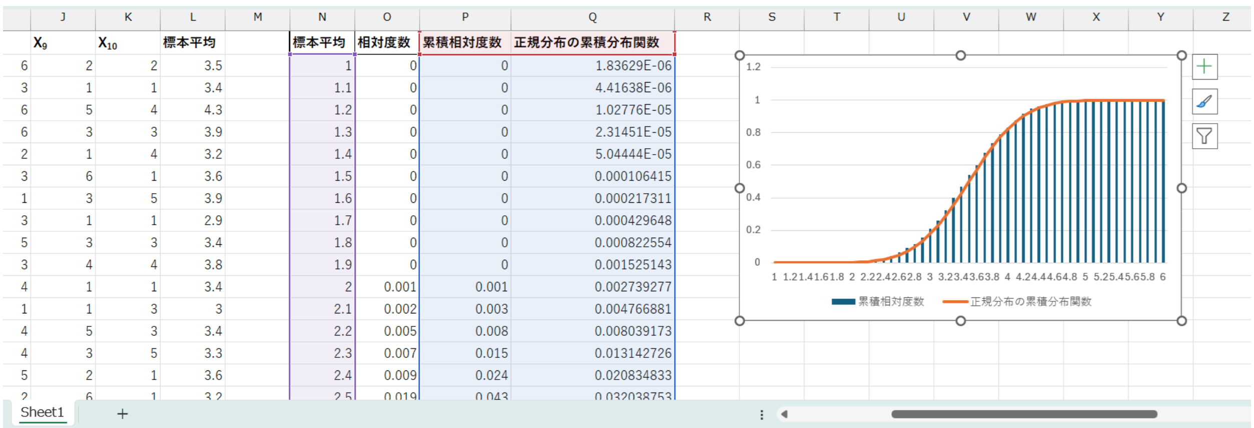
「標本平均の累積相対度数の棒グラフ」と「正規分布の累積分布関数の折れ線グラフ」を，同一グラフエリアにそれぞれ作成します．

そのため，セル範囲 P1:Q52 を選択し，挿入タブの（グラフグループにある）［縦棒/横棒グラフの挿入］の「2-D 縦棒」の「集合縦棒」を選びます．作成されたグラフが選択されたまま，グラフのデザインタブの（データグループにある）［データの選択］をクリックします．「データソースの選択」ダイアログボックスが出てくるので，横（項目）軸ラベルの「編集」をクリックします．「軸ラベル」ダイアログボックスが出てくるので，N 列の該当箇所（N2:N52）をドラッグして表示させて「OK」を押します．さらに，「データソースの選択」ダイアログボックスでも「OK」を押すと，グラフの横軸の値が N 列の値に変わります．

つぎに，作成したグラフをクリックして選択し，その状態で，グラフのデザインタブの（種類グループにある）［グラフの種類の変更］を選びます．

すると，「グラフの種類の変更」ダイアログボックスが出てくるので，「正規分布の累積分布関数」のグラフの種類を「折れ線」に変更しましょう．

また，グラフタイトルは削除しましょう．



	X_9	X_{10}	標本平均		標本平均	相対度数	累積相対度数	正規分布の累積分布関数
6	2	2	3.5		1	0	0	1.83629E-06
3	1	1	3.4		1.1	0	0	4.41638E-06
6	5	4	4.3		1.2	0	0	1.02776E-05
6	3	3	3.9		1.3	0	0	2.31451E-05
2	1	4	3.2		1.4	0	0	5.04444E-05
3	6	1	3.6		1.5	0	0	0.000106415
1	3	5	3.9		1.6	0	0	0.000217311
3	1	1	2.9		1.7	0	0	0.000429648
5	3	3	3.4		1.8	0	0	0.000822554
3	4	4	3.8		1.9	0	0	0.001525143
4	1	1	3.4		2	0.001	0.001	0.002739277
1	1	3	3		2.1	0.002	0.003	0.004766881
4	5	3	3.4		2.2	0.005	0.008	0.008039173
4	3	5	3.3		2.3	0.007	0.015	0.013142726
5	2	1	3.6		2.4	0.009	0.024	0.020834833

「大きさ 10 の標本を 1000 回無作為抽出したときの標本平均の累積相対度数のグラフ」が,「正規分布 $N(3.5, 35/(12\times 10))$ の累積分布関数のグラフ」に非常に似ていることが確かめられます.

8.2 の確認テスト

問 1

(1) μ, σ^2/n, $n\mu$, $n\sigma^2$, 中心極限定理
(2) μ, σ^2/n

問 2

ある母集団から,大きさ n の標本の無作為抽出をくり返すとき,標本サイズ n が大きいほど,その標本平均のばらつきは小さくなります.

問 3

サイコロを振ったときに出る目を X とするとき,X の期待値 $E(X)$ は 3.5 となり,X の分散 $V(X)$ は 35/12 となります(4.2 の確認テスト　問 2 参照).

よって,中心極限定理より,求める正規分布の期待値は 3.5 となり,分散は,$(35/12)/35 = 1/12$ より,1/12 となります.

問 4

コインを投げたときに表が出たら 1,裏が出たら 0 を割り当てる対応を X とするとき,X の期待値 $E(X)$ は 0.5 となり,X の分散 $V(X)$ は 0.25 となります(4.2 の確認テスト　問 3 参照).

よって,中心極限定理より,求める正規分布の期待値は 0.5 となり,分散は,$0.25/500 = 1/2000$ より,1/2000 となります.

問 5

コインを 5 回投げたときに表が出る回数を X とするとき,X の期待値 $E(X)$ は 2.5 となり,X の分散 $V(X)$ は 1.25 となります(4.2 の確認テスト　問 4 参照).

よって,中心極限定理より,求める正規分布の期待値は 2.5 となり,分散は,$1.25/125 = 1/100$ より,1/100 となります.

問 6

(1) 中心極限定理より,求める正規分布の期待値は 170 cm となり,標準偏差は,$\sqrt{5^2/100} = \sqrt{1/4} = 1/2 = 0.5$ より,0.5 cm となります.
(2) $Z = (\overline{X} - 170)/0.5$ とあらわされます.
(3) 求める正規分布は標準正規分布なので,期待値は 0,分散は 1 となります.

ここで,期待値 μ,標準偏差 σ の確率変数 X から確率変数 $Z = (X - \mu)/\sigma$ への変換を標準化といいます.X が正規分布 $N(\mu, \sigma^2)$ に従うとき,Z は標準正規分布 $N(0, 1^2)$ に従います.

第 9 章　母平均の区間推定（母分散既知）

9.1　母平均の 95%信頼区間

問題 9-1

セル D1 に「=AVERAGE(A4:A23)」と入力し，標本平均を求めます．168.0 であることがわかります．

セル F3 に「=NORM.S.INV(1-0.025)」と入力し，標準正規分布における上側 2.5%点を求めます．約 1.96 であることがわかります．

セル G5 に「=D1+F3*SQRT(64/20)」と入力し，母平均 μ の 95%信頼区間の上限を求めます．約 171.5 であることがわかります．

また，セル E5 に「=D1-F3*SQRT(64/20)」と入力し，母平均 μ の 95%信頼区間の下限を求めます．約 164.5 であることがわかります．

よって，母平均 μ の 95%信頼区間はおおよそ

$$164.5 \leq \mu \leq 171.5$$

となります．

	A	B	C	D	E	F	G
1	母分散 σ^2=64		標本平均	168.0			
2							
3	標本		標準正規分布における上側2.5%点			1.959963985	
4	169.3						
5	162.8		母平均 μ の95%信頼区間		164.4939	≤ 母平均 μ ≤	171.5061
6	161.4						
7	165.7						
8	167.0						
9	170.8						
10	181.8						
11	163.8						
12	169.7						
13	170.0						
14	166.0						
15	159.5						
16	170.9						

Sheet1

問題 9-2

セル B2 に「=NORM.INV(RAND(),100,50)」と入力し，このセル（B2）を 5 行目まで下にドラッグしてオートフィルします．さらにそのまま，選択されているセル範囲 B2:B5 の右下あたりにマウスポインタを合わせ「+」の形にして，この状態のまま CW 列目まで右にドラッグし，オートフィルします．すると，期待値 100，標準偏差 50（分散 2500）である正規乱数が 4 行 100 列分，つまり，400 個発生します．

これで，母集団から大きさ 4 の標本を 100 回無作為抽出したことになります．

セル B6 に「=AVERAGE(B2:B5)」と入力し，標本平均を求め，このセル（B6）を CW 列目まで右にドラッグし，オートフィルします．

セル E8 に「=NORM.S.INV(1-0.025)」と入力し，標準正規分布における上側 2.5%点を求めます．約 1.96 であることがわかります．

セル B11 に「=B6+$E$8*SQRT(2500/4)」と入力し，母平均 μ の 95%信頼区間の上限を求め，このセル（B11）を CW 列目まで右にドラッグし，オートフィルします．

また，セル B12 に「=B6-$E$8*SQRT(2500/4)」と入力し，母平均 μ の 95%信頼区間の下限を求め，このセル（B12）を CW 列目まで右にドラッグし，オートフィルします．

母平均の 95%信頼区間の上限と下限についての折れ線グラフを作成します．

そのため，セル範囲 A11:CW12 を選択し，挿入タブの（グラフグループにある）［折れ線/面グラフの挿入］の「2-D 折れ線」の「折れ線」を選びます．

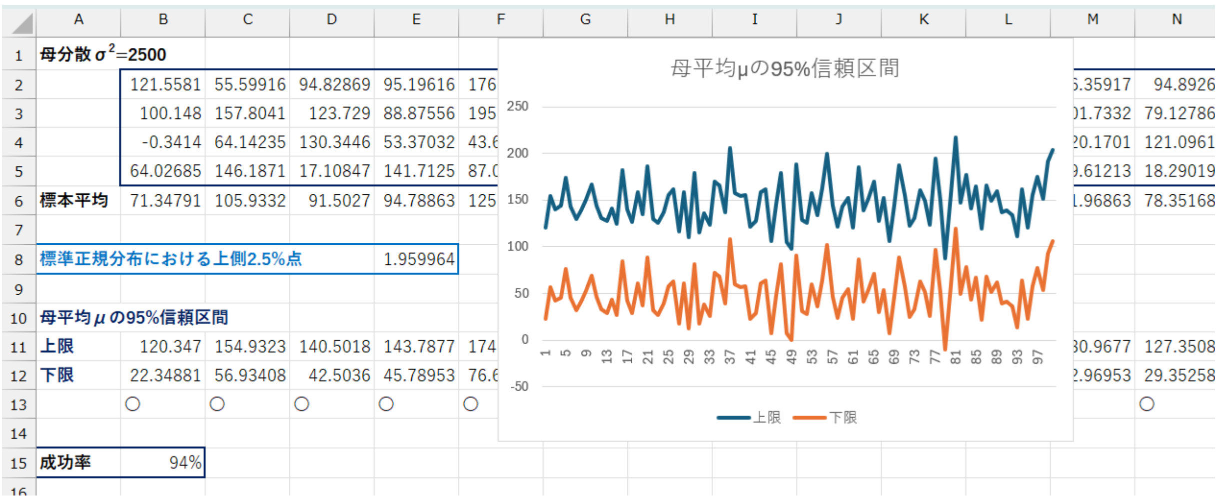
出てきた折れ線グラフの縦 (値) 軸について，表示される最小値を 80.0，最大値を 120.0 に変更します．

そのため，折れ線グラフの縦軸の上で右クリック（またはダブルクリック）し，「軸の書式設定」を出し，(軸のオプションの)「最小値」を −50 にして，「最大値」を 250 に変更します．

また，グラフタイトルは「母平均μの 95%信頼区間」にしましょう．

セル B13 に「=IF(OR(B11<100,B12>100),”×”,” ○”)」と入力し，95%信頼区間のなかに母平均 100 が含まれるなら「○」，含まれないなら「×」が表示されるようにします．そして，このセル（B13）を CW 列目まで右にドラッグし，オートフィルします．

セル B15 に「=COUNTIF(B13:CW13,” ○”)/100」と入力し，「○」の割合を求めます．ホームタブの（数値グループにある）［パーセント スタイル］を選択することによって，%スタイルにしましょう．



問題 9-3

セル B2 に「=NORM.INV(RAND(),100,50)」と入力し，このセル（B2）を 201 行目まで下にドラッグしてオートフィルします．さらにそのまま，選択されているセル範囲 B2:B201 の右下あたりにマウスポインタを合わせ「+」の形にして，この状態のまま CW 列目まで右にドラッグし，オートフィルします．すると，期待値 100，標準偏差 50（分散 2500）である正規乱数が 200 行 100 列分，つまり，20000 個発生します．

これで，母集団から大きさ 200 の標本を 100 回無作為抽出したことになります．

セル B202 に「=AVERAGE(B2:B201)」と入力し，標本平均を求め，このセル（B202）を CW 列目まで右にドラッグし，オートフィルします．

セル E204 に「=NORM.S.INV(1-0.025)」と入力し，標準正規分布における上側 2.5%点を求めます．約 1.96 であることがわかります．

セル B207 に「=B202+$E$204*SQRT(2500/200)」と入力し，母平均 μ の 95%信頼区間の上限を求め，このセル（B207）を CW 列目まで右にドラッグし，オートフィルします．

また，セル B208 に「=B202-$E$204*SQRT(2500/200)」と入力し，母平均 μ の 95%信頼区間の下限を求め，このセル（B208）を CW 列目まで右にドラッグし，オートフィルします．

母平均の 95%信頼区間の上限と下限についての折れ線グラフを作成します．

そのため，セル範囲 A207:CW208 を選択し，挿入タブの（グラフグループにある）［折れ線/面グラフの挿入］の「2-D 折れ線」の「折れ線」を選びます．

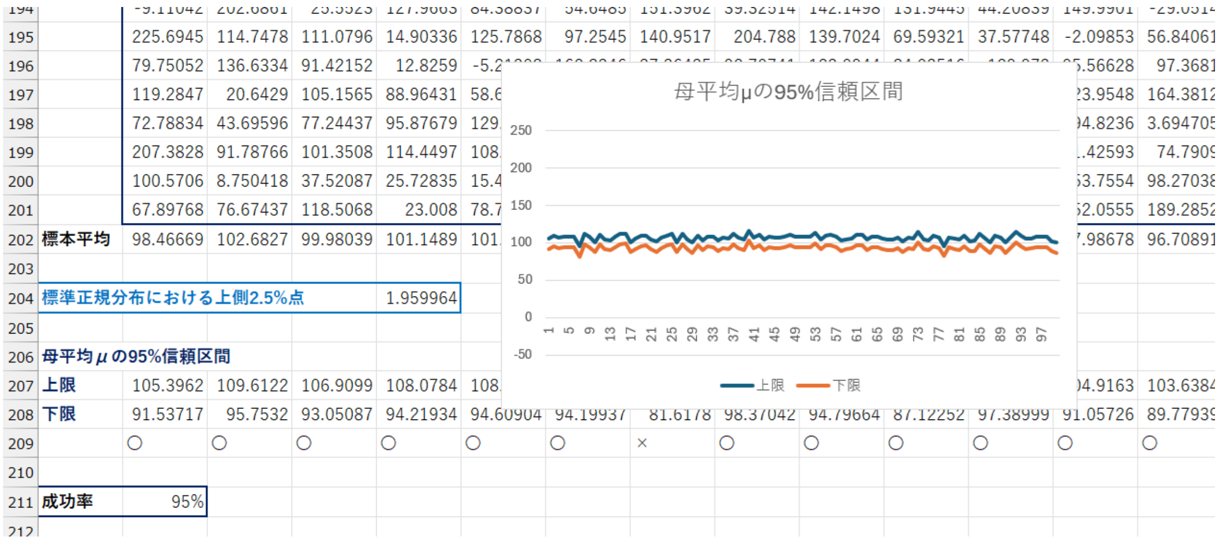
出てきた折れ線グラフの縦 (値) 軸について，表示される最小値を −50，最大値を 250 に変更します．

そのため，折れ線グラフの縦軸の上で右クリック（またはダブルクリック）し，「軸の書式設定」を出し，(軸のオプションの)「最小値」を −50 にして，「最大値」を 250 に変更します．

また，グラフタイトルは「母平均 μ の 95%信頼区間」にしましょう．

セル B209 に「=IF(OR(B207<100,B208>100),”×”,” ○”)」と入力し，95%信頼区間のなかに母平均 100 が含まれるなら「○」，含まれないなら「×」が表示されるようにします．そして，このセル（B209）を CW 列目まで右にドラッグし，オートフィルします．

セル B211 に「=COUNTIF(B209:CW209,” ○”)/100」と入力し，「○」の割合を求めます．ホームタブの（数値グループにある）［パーセント スタイル］を選択することによって，%スタイルにしましょう．



195		225.6945	114.7478	111.0796	14.90336	125.7868	97.2545	140.9517	204.788	139.7024	69.59321	37.57748	-2.09853	56.84061
196		79.75052	136.6334	91.42152	12.8259	-5.2							.56628	97.3681
197		119.2847	20.6429	105.1565	88.96431	58.6							3.9548	164.3812
198		72.78834	43.69596	77.24437	95.87679	129.							4.8236	3.694705
199		207.3828	91.78766	101.3508	114.4497	108.							.42593	74.7909
200		100.5706	8.750418	37.52087	25.72835	15.4							3.7554	98.27038
201		67.89768	76.67437	118.5068	23.008	78.7							2.0555	189.2852
202	標本平均	98.46669	102.6827	99.98039	101.1489	101.							.98678	96.70891
203														
204	標準正規分布における上側2.5%点				1.959964									
205														
206	母平均 μ の95%信頼区間													
207	上限	105.3962	109.6122	106.9099	108.0784	108.							4.9163	103.6384
208	下限	91.53717	95.7532	93.05087	94.21934	94.60904	94.19937	81.6178	98.37042	94.79664	87.12252	97.38999	91.05726	89.7793
209		○	○	○	○	○	○	×	○	○	○	○	○	○
210														
211	成功率	95%												

ここで作成した「大きさ 200 の標本から算出される母平均の 95%信頼区間の上限と下限についての折れ線グラフ」は，問題 9-2 で作成した「大きさ 4 の標本から算出される母平均の 95%信頼区間の上限と下限についての折れ線グラフ」と比べると，区間の幅が小さく，ばらつきも小さいことが確認できます．

9.1 の確認テスト

問 1

(1) 区間推定，信頼区間

(2) 95，信頼係数

(3) 2.5，2.5

問 2

$$\overline{X} - 1.96 \times \sqrt{\frac{\sigma^2}{n}} \leq \mu \leq \overline{X} + 1.96 \times \sqrt{\frac{\sigma^2}{n}}$$

に，$\overline{X}=10$，$\sigma^2=2^2=4$，$n=16$ を代入すると，

$$10-1.96\times\sqrt{\frac{4}{16}}\leq\mu\leq 10+1.96\times\sqrt{\frac{4}{16}}$$

つまり，

$$10-1.96\times\frac{1}{2}\leq\mu\leq 10+1.96\times\frac{1}{2}$$

となるので，求める母平均 μ の 95%信頼区間は，これを計算した次になります．

$$9.02\leq\mu\leq 10.98$$

問 3

(1) $$\overline{X}-1.96\times\sqrt{\frac{\sigma^2}{n}}\leq\mu\leq\overline{X}+1.96\times\sqrt{\frac{\sigma^2}{n}}$$

に，$\overline{X}=100$，$\sigma^2=500$，$n=5$ を代入すると，

$$100-1.96\times\sqrt{\frac{500}{5}}\leq\mu\leq 100+1.96\times\sqrt{\frac{500}{5}}$$

つまり，

$$100-1.96\times 10\leq\mu\leq 100+1.96\times 10$$

となるので，求める母平均 μ の 95%信頼区間は，これを計算した次になります．

$$80.4\leq\mu\leq 119.6$$

(2) $$\overline{X}-1.96\times\sqrt{\frac{\sigma^2}{n}}\leq\mu\leq\overline{X}+1.96\times\sqrt{\frac{\sigma^2}{n}}$$

に，$\overline{X}=100$，$\sigma^2=500$，$n=500$ を代入すると，

$$100-1.96\times\sqrt{\frac{500}{500}}\leq\mu\leq 100+1.96\times\sqrt{\frac{500}{500}}$$

つまり，

$$100-1.96\times 1\leq\mu\leq 100+1.96\times 1$$

となるので，求める母平均 μ の 95%信頼区間は，これを計算した次になります．

$$98.04\leq\mu\leq 101.96$$

問 4

標本から母分散を使って算出される母平均の 95%信頼区間の幅は，標本サイズが大きいほど小さくなります．

9.2 母平均のさまざまな信頼区間（母分散既知）

問題 9-4

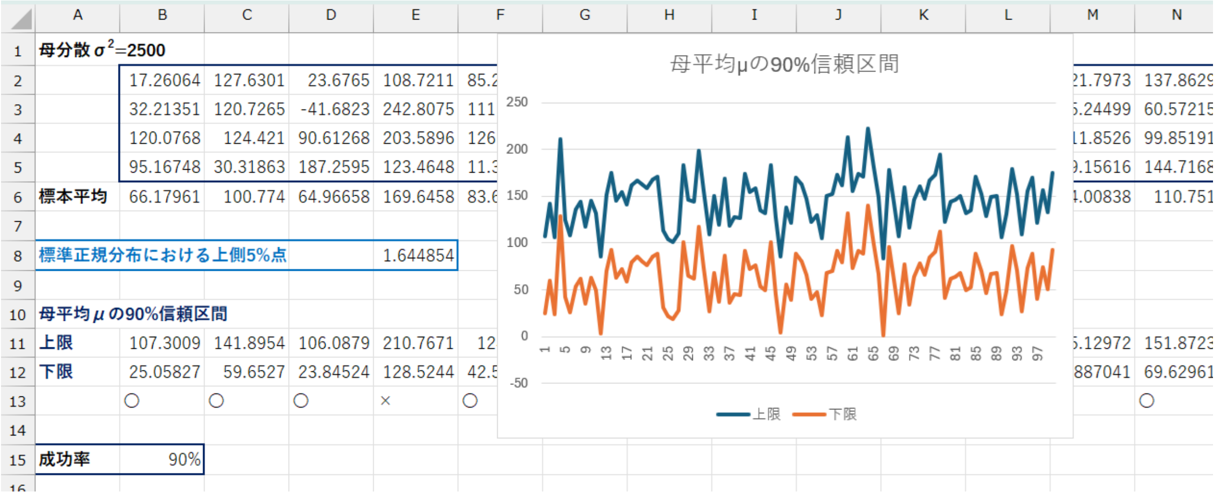
セル A8 を「標準正規分布における上側 5%点」に入力しなおします．

セル E8 を「=NORM.S.INV(1-0.05)」に入力しなおし，標準正規分布における上側 5%点を求めます．約

1.64 であることがわかります．

セル A10 を「母平均 μ の 90%信頼区間」に入力しなおします．

また，グラフタイトルは「母平均 μ の 90%信頼区間」に変更しましょう．



	A	B	C	D	E	F	M	N
1	母分散 σ^2=2500							
2		17.26064	127.6301	23.6765	108.7211	85.2	21.7973	137.8629
3		32.21351	120.7265	-41.6823	242.8075	111	.24499	60.57215
4		120.0768	124.421	90.61268	203.5896	126	11.8526	99.85191
5		95.16748	30.31863	187.2595	123.4648	11.3	.15616	144.7168
6	標本平均	66.17961	100.774	64.96658	169.6458	83.6	.00838	110.751
7								
8	標準正規分布における上側5%点				1.644854			
9								
10	母平均 μ の90%信頼区間							
11	上限	107.3009	141.8954	106.0879	210.7671	12	.12972	151.8723
12	下限	25.05827	59.6527	23.84524	128.5244	42.5	887041	69.62961
13		○	○	○	×	○		○
14								
15	成功率	90%						

ここで作成した「母平均の 90%信頼区間の上限と下限についての折れ線グラフ」は，問題 9-2 で作成した「母平均の 95%信頼区間の上限と下限についての折れ線グラフ」と比べると，区間の幅が小さいことが確認できます．

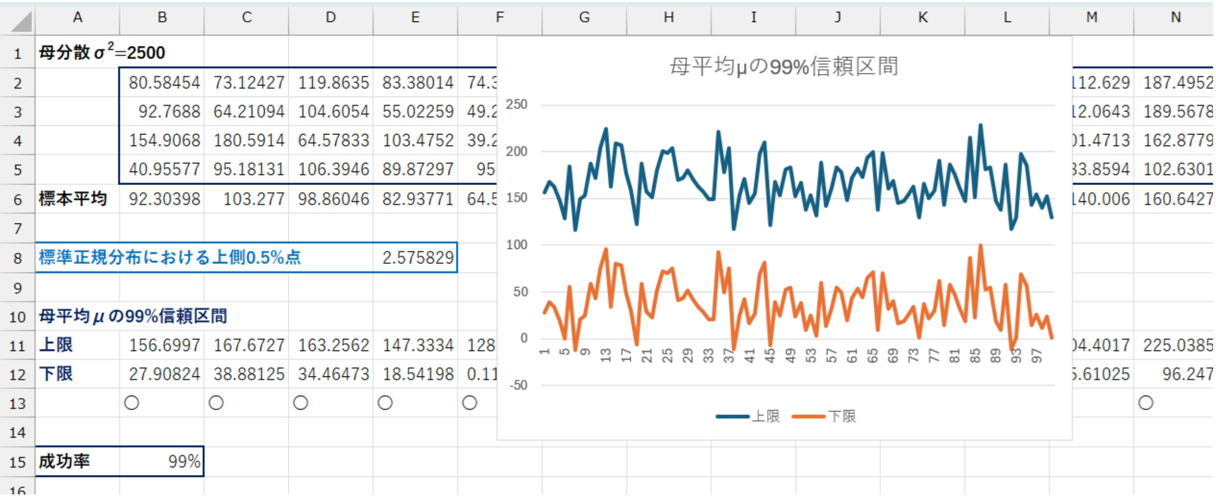
問題 9-5

セル A8 を「標準正規分布における上側 0.5%点」に入力しなおします．

セル E8 を「=NORM.S.INV(1-0.005)」に入力しなおし，標準正規分布における上側 0.5%点を求めます．約 2.58 であることがわかります．

セル A10 を「母平均 μ の 99%信頼区間」に入力しなおします．

また，グラフタイトルは「母平均 μ の 99%信頼区間」に変更しましょう．



	A	B	C	D	E	F	M	N
1	母分散 σ^2=2500							
2		80.58454	73.12427	119.8635	83.38014	74.3	112.629	187.4952
3		92.7688	64.21094	104.6054	55.02259	49.2	12.0643	189.5678
4		154.9068	180.5914	64.57833	103.4752	39.2	01.4713	162.8779
5		40.95577	95.18131	106.3946	89.87297	95	33.8594	102.6301
6	標本平均	92.30398	103.277	98.86046	82.93771	64.5	140.006	160.6427
7								
8	標準正規分布における上側0.5%点				2.575829			
9								
10	母平均 μ の99%信頼区間							
11	上限	156.6997	167.6727	163.2562	147.3334	128	04.4017	225.0385
12	下限	27.90824	38.88125	34.46473	18.54198	0.11	.61025	96.247
13		○	○	○	○	○		○
14								
15	成功率	99%						

ここで作成した「母平均の 99%信頼区間の上限と下限についての折れ線グラフ」は，問題 9-2 で作成した「母平均の 95%信頼区間の上限と下限についての折れ線グラフ」と比べると，区間の幅が大きいことが確認できます．

9.2 の確認テスト

問 1

(1) 5，5

(2) 0.5，0.5

問 2

$$\overline{X} - 1.64 \times \sqrt{\frac{\sigma^2}{n}} \leq \mu \leq \overline{X} + 1.64 \times \sqrt{\frac{\sigma^2}{n}}$$

に，$\overline{X} = 10$，$\sigma^2 = 2^2 = 4$，$n = 16$ を代入すると，

$$10 - 1.64 \times \sqrt{\frac{4}{16}} \leq \mu \leq 10 + 1.64 \times \sqrt{\frac{4}{16}}$$

つまり，

$$10 - 1.64 \times \frac{1}{2} \leq \mu \leq 10 + 1.64 \times \frac{1}{2}$$

となるので，求める母平均 μ の 90%信頼区間は，これを計算した次になります．

$$9.18 \leq \mu \leq 10.82$$

また，99%信頼区間は，2 つ上の式において 1.64 を 2.58 におきかえた

$$10 - 2.58 \times \frac{1}{2} \leq \mu \leq 10 + 2.58 \times \frac{1}{2}$$

を計算した次になります．

$$8.71 \leq \mu \leq 11.29$$

問 3

(1) $$\overline{X} - 1.64 \times \sqrt{\frac{\sigma^2}{n}} \leq \mu \leq \overline{X} + 1.64 \times \sqrt{\frac{\sigma^2}{n}}$$

に，$\overline{X} = 100$，$\sigma^2 = 500$，$n = 5$ を代入すると，

$$100 - 1.64 \times \sqrt{\frac{500}{5}} \leq \mu \leq 100 + 1.64 \times \sqrt{\frac{500}{5}}$$

つまり，

$$100 - 1.64 \times 10 \leq \mu \leq 100 + 1.64 \times 10$$

となるので，求める母平均 μ の 90%信頼区間は，これを計算した次になります．

$$83.6 \leq \mu \leq 116.4$$

また，99%信頼区間は，2 つ上の式において 1.64 を 2.58 におきかえた

$$100 - 2.58 \times 10 \leq \mu \leq 100 + 2.58 \times 10$$

を計算した次になります．

$$74.2 \leq \mu \leq 125.8$$

(2) $\overline{X} - 1.64 \times \sqrt{\dfrac{\sigma^2}{n}} \leq \mu \leq \overline{X} + 1.64 \times \sqrt{\dfrac{\sigma^2}{n}}$

に，$\overline{X} = 100$，$\sigma^2 = 500$，$n = 500$ を代入すると，

$$100 - 1.64 \times \sqrt{\frac{500}{500}} \leq \mu \leq 100 + 1.64 \times \sqrt{\frac{500}{500}}$$

つまり，

$$100 - 1.64 \times 1 \leq \mu \leq 100 + 1.64 \times 1$$

となるので，求める母平均 μ の 99%信頼区間は，これを計算した次になります.

$$98.36 \leq \mu \leq 101.64$$

また，99%信頼区間は，2 つ上の式において 1.64 を 2.58 におきかえた

$$100 - 2.58 \times 1 \leq \mu \leq 100 + 2.58 \times 1$$

を計算した次になります.

$$97.42 \leq \mu \leq 102.58$$

問 4

標本から母分散を使って算出される母平均の信頼区間の幅は，信頼係数が大きいほど大きくなります.

第 10 章　母平均の区間推定（母分散未知）

10.1　t 分布

問題 10-1

セル B2 に「=T.DIST(A2,10,FALSE)」と入力し，自由度 10 の t 分布の確率密度関数の $x = -5$ における値を求めましょう．

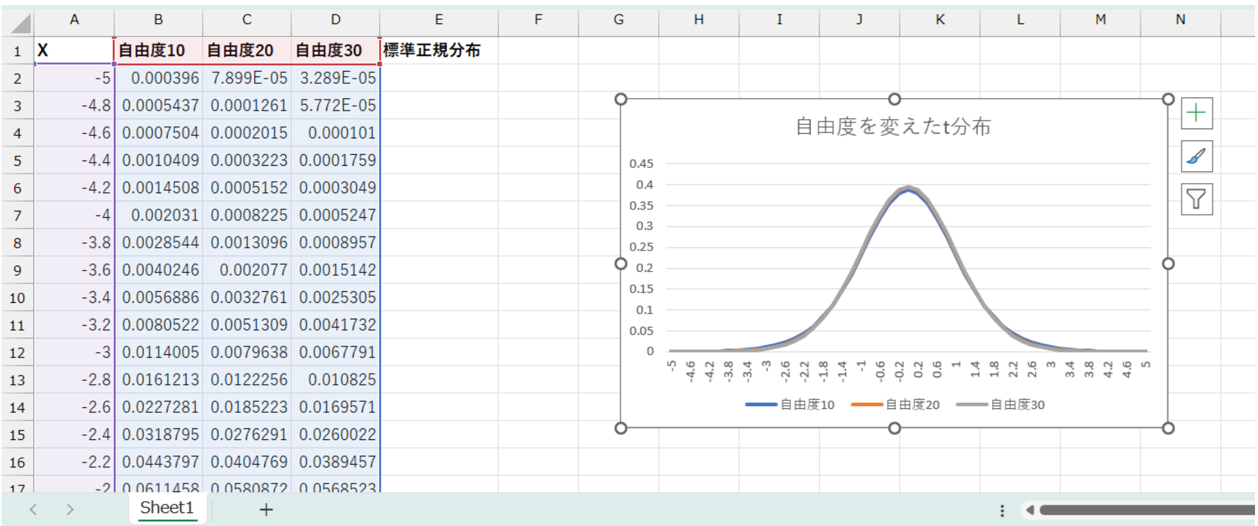
セル C2 に「=T.DIST(A2,20,FALSE)」と入力し，自由度 20 の t 分布の確率密度関数の $x = -5$ における値を求めましょう．

セル D2 に「=T.DIST(A2,30,FALSE)」と入力し，自由度 30 の t 分布の確率密度関数の $x = -5$ における値を求めましょう．

そして，セル範囲 B2:D2 を選択し，その右下あたりにマウスポインタを合わせ「+」の形にして，この状態のままダブルクリックします．すると，$x = 5$ における値までオートフィルで求めることができます．

セル範囲 B1:D52 を選択したあと，挿入タブの（グラフグループにある）［折れ線/面グラフの挿入］の「2-D 折れ線」の「折れ線」を選びます．作成されたグラフが選択されたまま，グラフのデザインタブの（データグループにある）［データの選択］をクリックします．「データソースの選択」ダイアログボックスが出てくるので，横（項目）軸ラベルの「編集」をクリックします．「軸ラベル」ダイアログボックスが出てくるので，A 列の該当箇所（A2:A52）をドラッグして表示させて「OK」を押します．さらに，「データソースの選択」ダイアログボックスでも「OK」を押すと，グラフの横軸の値が A 列の値に変わります．

グラフタイトルは「自由度を変えた t 分布」にしましょう．



X	自由度10	自由度20	自由度30	標準正規分布
-5	0.000396	7.899E-05	3.289E-05	
-4.8	0.0005437	0.0001261	5.772E-05	
-4.6	0.0007504	0.0002015	0.000101	
-4.4	0.0010409	0.0003223	0.0001759	
-4.2	0.0014508	0.0005152	0.0003049	
-4	0.002031	0.0008225	0.0005247	
-3.8	0.0028544	0.0013096	0.0008957	
-3.6	0.0040246	0.002077	0.0015142	
-3.4	0.0056886	0.0032761	0.0025305	
-3.2	0.0080522	0.0051309	0.0041732	
-3	0.0114005	0.0079638	0.0067791	
-2.8	0.0161213	0.0122256	0.010825	
-2.6	0.0227281	0.0185223	0.0169571	
-2.4	0.0318795	0.0276291	0.0260022	
-2.2	0.0443797	0.0404769	0.0389457	
-2	0.0611458	0.0580872	0.0568523	

これらのグラフはほとんどちがいがわかりませんが，よく見ると，自由度が大きくなるにつれて山が高くなることが確認できます．

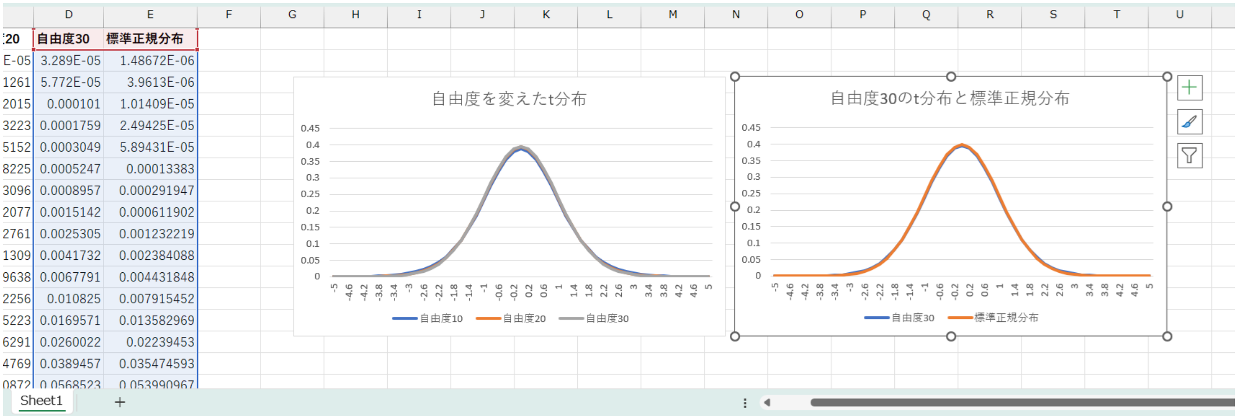
問題 10-2

セル E2 に「=NORM.S.DIST(A2,FALSE)」と入力し，標準正規分布の確率密度関数の $x = -5$ における値を求めましょう．

そして，そのセル（E2）の右下あたりをダブルクリックすることにより，$x = 5$ における値までオートフィルで求めましょう．

セル範囲 D1:E52 を選択したあと，挿入タブの（グラフグループにある）［折れ線/面グラフの挿入］の「2-D 折れ線」の「折れ線」を選びます．作成されたグラフが選択されたまま，グラフのデザインタブの（データグループにある）［データの選択］をクリックします．「データソースの選択」ダイアログボックスが出てくるので，横（項目）軸ラベルの「編集」をクリックします．「軸ラベル」ダイアログボックスが出てくるので，A 列の該当箇所（A2:A52）をドラッグして表示させて「OK」を押します．さらに，「データソースの選択」ダイアログボックスでも「OK」を押すと，グラフの横軸の値が A 列の値に変わります．

グラフタイトルは「自由度 30 の t 分布と標準正規分布」にしましょう．



ふたつのグラフは重なって見え，ほとんどちがいがわからないことが確認できます．

問題 10-3

セル B2 に「=T.DIST(A2,5,FALSE)」と入力し，自由度 5 の t 分布の確率密度関数の $x = -5$ における値を求めましょう．

セル C2 に「=T.DIST(A2,5,TRUE)」と入力し，自由度 5 の t 分布の累積分布関数の $x = -5$ における値を求めましょう．

そして，セル範囲 B2:C2 を選択し，その右下あたりにマウスポインタを合わせ「+」の形にして，この状態のままダブルクリックします．すると，$x = 5$ における値までオートフィルで求めることができます．

セル I2 に「=T.INV(0.05,5)」と入力し，「下位 5%以下に入るのはどの値以下か」を求めます．

セル I3 に「=T.INV(1-0.05,5)」と入力し，「上位 5%以上に入るのはどの値以上か」を求めます．

セル I4 に「=T.INV(0.025,5)」と入力し，「下位 2.5%以下に入るのはどの値以下か」を求めます．

セル I5 に「=T.INV(1-0.025,5)」と入力し，「上位 2.5%以上に入るのはどの値以上か」を求めます．

セル I6 に「=T.INV(0.005,5)」と入力し，「下位 0.5%以下に入るのはどの値以下か」を求めます．

セル I7 に「=T.INV(1-0.005,5)」と入力し，「上位 0.5%以上に入るのはどの値以上か」を求めます．

	A	B	C	D	E	F	G	H	I
1	X	確率密度関数	累積分布関数		自由度5のt分布				
2	-5	0.001757438	0.002052358		下位5%以下に入るのはどの値以下か				-2.01505
3	-4.8	0.002152335	0.002441826		上位5%以上に入るのはどの値以上か				2.015048
4	-4.6	0.002650517	0.002920161		下位2.5%以下に入るのはどの値以下か				-2.57058
5	-4.4	0.003282555	0.003510935		上位2.5%以上に入るのはどの値以上か				2.570582
6	-4.2	0.004088976	0.004244779		下位0.5%以下に入るのはどの値以下か				-4.03214
7	-4	0.005123727	0.005161708		上位0.5%以上に入るのはどの値以上か				4.032143
8	-3.8	0.006458848	0.006314241						
9	-3.6	0.008190773	0.007771628						
10	-3.4	0.010448715	0.009625522						
11	-3.2	0.013405684	0.011997588						
12	-3	0.017292579	0.015049624						
13	-2.8	0.022415519	0.018996812						
14	-2.6	0.029175742	0.024124727						
15	-2.4	0.038089657	0.030810396						
16	-2.2	0.049803352	0.039546949						
17	-2	0.06509031	0.050969739						

Sheet1

10.1 の確認テスト

問 1

(1) 不偏分散

(2) t 分布

(3) 標準正規分布

問 2

自由度 5 の t 分布のグラフの山のほうが高い．

問 3

標準正規分布のグラフの山のほうが高い．

問 4

自由度 5 の t 分布のグラフのすそのほうがうすい．

問 5

標準正規分布のグラフのすそのほうがうすい．

10.2 母平均の信頼区間（母分散未知）

問題 10-4

セル D1 に「=AVERAGE(A4:A9)」と入力し，標本平均を求めます．約 169.7 であることがわかります．

また，セル D2 に「=VAR.S(A4:A9)」と入力し，不偏分散を求めます．約 53.6 であることがわかります．

セル F4 に「=T.INV(1-0.025,5)」と入力し，自由度 5 の t 分布における上側 2.5%点を求めます．約 2.57 であることがわかります．

セル G6 に「=D1+F4*SQRT(D2/6)」と入力し，母平均 μ の 95%信頼区間の上限を求めます．約 177.4 であることがわかります．

また，セル E6 に「=D1-F4*SQRT(D2/6)」と入力し，母平均 μ の 95%信頼区間の下限を求めます．約 162.0 であることがわかります．

よって，母平均 μ の 95%信頼区間はおおよそ

$$162.0 \leq \mu \leq 177.4$$

となります．

	A	B	C	D	E	F	G
1	母分散 σ^2=?		標本平均	169.7167			
2			不偏分散	53.63767			
3	標本						
4	170.8		自由度5のt分布における上側2.5%点			2.570581836	
5	159.5						
6	166.3		母平均 μ の95%信頼区間		162.0308	≤ 母平均 μ ≤	177.4025
7	170.0						
8	169.7						
9	182.0						
10							

問題 10-5

セル D1 に「=AVERAGE(A4:A33)」と入力し，標本平均を求めます．約 169.7 であることがわかります．

また，セル D2 に「=VAR.S(A4:A33)」と入力し，不偏分散を求めます．約 53.6 であることがわかります．

セル F4 に「=T.INV(1-0.025,29)」と入力し，自由度 29 の t 分布における上側 2.5%点を求めます．約 2.05 であることがわかります．

セル G6 に「=D1+F4*SQRT(D2/30)」と入力し，母平均 μ の 95%信頼区間の上限を求めます．約 172.4 であることがわかります．

また，セル E6 に「=D1-F4*SQRT(D2/30)」と入力し，母平均 μ の 95%信頼区間の下限を求めます．約 167.0 であることがわかります．

よって，母平均 μ の 95%信頼区間はおおよそ

$$167.0 \leq \mu \leq 172.4$$

となります．

	A	B	C	D	E	F	G
1	母分散 σ^2=?		標本平均	169.6867			
2			不偏分散	53.63775			
3	標本						
4	170.2		自由度29のt分布における上側2.5%点			2.045229642	
5	162.0						
6	170.7		母平均 μ の95%信頼区間		166.9519	≤ 母平均 μ ≤	172.4214
7	166.7						
8	167.0						
9	171.3						
10	180.7						
11	158.9						
12	169.7						
13	169.9						
14	166.0						
15	165.5						
16	170.9						

Sheet1

ここで求めた「大きさ 30 の標本から算出した母平均 μ の 95%信頼区間」：

$$167.0 \leq \mu \leq 172.4$$

は，問題 10-4 で求めた「大きさ 6 の標本から算出した母平均 μ の 95%信頼区間」：

$$162.0 \leq \mu \leq 177.4$$

に比べて，幅が小さいことが確認できます（標本平均も不偏分散も互いにほぼ等しいです）．

問題 10-6

セル C4 を「自由度 5 の t 分布における上側 5%点」に入力しなおします．

セル F4 を「=T.INV(1-0.05,5)」に入力しなおし，自由度 5 の t 分布における上側 5%点を求めます．約 2.02 であることがわかります．

セル C6 を「母平均 μ の 90%信頼区間」に入力しなおします．

母平均 μ の 90%信頼区間はおおよそ

$$163.7 \leq \mu \leq 175.7$$

となります．

	A	B	C	D	E	F	G
1	母分散 σ^2=?		標本平均	169.7167			
2			不偏分散	53.63767			
3	標本						
4	170.8		自由度5のt分布における上側5%点			2.015048373	
5	159.5						
6	166.3		母平均μの90%信頼区間		163.6918	≤ 母平均μ ≤	175.7415
7	170.0						
8	169.7						
9	182.0						
10							
11							

ここで求めた「母平均 μ の 90%信頼区間」：

$$163.7 \leq \mu \leq 175.7$$

は，問題 10-4 で求めた「母平均 μ の 95%信頼区間」：

$$162.0 \leq \mu \leq 177.4$$

に比べて，幅が小さいことが確認できます．

問題 10-7

セル C4 を「自由度 5 の t 分布における上側 0.5%点」に入力しなおします．

セル F4 を「=T.INV(1-0.005,5)」に入力しなおし，自由度 5 の t 分布における上側 0.5%点を求めます．約 4.03 であることがわかります．

セル C6 を「母平均 μ の 99%信頼区間」に入力しなおします．

母平均 μ の 99%信頼区間はおおよそ

$$157.7 \leq \mu \leq 181.8$$

となります．

	A	B	C	D	E	F	G
1	母分散σ^2=?		標本平均	169.7167			
2			不偏分散	53.63767			
3	標本						
4	170.8		自由度5のt分布における上側0.5%点			4.032142984	
5	159.5						
6	166.3		母平均μの99%信頼区間		157.6609	≤母平均μ≤	181.7724
7	170.0						
8	169.7						
9	182.0						
10							

ここで求めた「母平均 μ の 99%信頼区間」：

$$157.7 \leq \mu \leq 181.8$$

は，問題 10-4 で求めた「母平均 μ の 95%信頼区間」：

$$162.0 \leq \mu \leq 177.4$$

に比べて，幅が大きいことが確認できます．

10.2 の確認テスト

問 1

(1) $(n-1)$，2.5，$(n-1)$，2.5
(2) $(n-1)$，5，$(n-1)$，5
(3) $(n-1)$，0.5，$(n-1)$，0.5

問 2

$$\overline{X} - 2.35 \times \sqrt{\frac{s^2}{n}} \leq \mu \leq \overline{X} + 2.35 \times \sqrt{\frac{s^2}{n}}$$

に，$\overline{X} = 10$，$s^2 = 2^2 = 16$，$n = 4$ を代入すると，

$$10 - 2.35 \times \sqrt{\frac{16}{4}} \leq \mu \leq 10 + 2.35 \times \sqrt{\frac{16}{4}}$$

つまり，

$$10 - 2.35 \times 2 \leq \mu \leq 10 + 2.35 \times 2$$

となるので，求める母平均 μ の 90%信頼区間は，これを計算した次になります．

$$5.3 \leq \mu \leq 14.7$$

また，95%信頼区間は，2 つ上の式において 2.35 を 3.18 におきかえた

$$10 - 3.18 \times 2 \leq \mu \leq 10 + 3.18 \times 2$$

を計算した次になります．

$$3.64 \leq \mu \leq 16.36$$

99%信頼区間は，同じ式において 2.35 を 5.84 におきかえた

$$10 - 5.84 \times 2 \leq \mu \leq 10 + 5.84 \times 2$$

を計算した次になります

$$-1.68 \leq \mu \leq 21.68$$

問 3

(1) $\overline{X} - 2.13 \times \sqrt{\dfrac{s^2}{n}} \leq \mu \leq \overline{X} + 2.13 \times \sqrt{\dfrac{s^2}{n}}$

に，$\overline{X} = 100$，$\sigma^2 = 500$，$n = 5$ を代入すると，

$$100 - 2.13 \times \sqrt{\frac{500}{5}} \leq \mu \leq 100 + 2.13 \times \sqrt{\frac{500}{5}}$$

つまり，

$$100 - 2.13 \times 10 \leq \mu \leq 100 + 2.13 \times 10$$

となるので，求める母平均 μ の 90%信頼区間は，これを計算した次になります．

$$78.7 \leq \mu \leq 121.3$$

また，95%信頼区間は，2 つ上の式において 2.13 を 2.78 におきかえた

$$100 - 2.78 \times 10 \leq \mu \leq 100 + 2.78 \times 10$$

を計算した次になります．

$$72.2 \leq \mu \leq 127.8$$

99%信頼区間は，同じ式において 2.13 を 4.60 におきかえた

$$100 - 4.60 \times 2 \leq \mu \leq 100 + 4.60 \times 2$$

を計算した次になります

$$54 \leq \mu \leq 146$$

(2) $\overline{X} - 1.65 \times \sqrt{\dfrac{s^2}{n}} \leq \mu \leq \overline{X} + 1.65 \times \sqrt{\dfrac{s^2}{n}}$

に，$\overline{X} = 100$，$\sigma^2 = 500$，$n = 500$ を代入すると，

$$100 - 1.65 \times \sqrt{\frac{500}{500}} \leq \mu \leq 100 + 1.65 \times \sqrt{\frac{500}{500}}$$

つまり，

$$100 - 1.65 \times 1 \leq \mu \leq 100 + 1.65 \times 1$$

となるので，求める母平均 μ の 90%信頼区間は，これを計算した次になります．

$$98.35 \leq \mu \leq 101.65$$

また，95%信頼区間は，2 つ上の式において 1.65 を 1.96 におきかえた

$$100 - 1.96 \times 1 \leq \mu \leq 100 + 1.96 \times 1$$

を計算した次になります．

$$98.04 \leq \mu \leq 101.96$$

そして，99%信頼区間は，同じ式において 1.65 を 2.59 におきかえた

$$100 - 2.59 \times 1 \leq \mu \leq 100 + 2.59 \times 1$$

を計算した次になります．

$$97.41 \leq \mu \leq 102.59$$

問 4

標本から不偏分散を使って算出される母平均の 95%信頼区間の幅は，不偏分散が同じであるとすると，標本サイズが大きいほど小さくなります．

問 5

標本から不偏分散を使って算出される母平均の信頼区間の幅は，信頼係数が大きいほど大きくなります．

問 6

標本平均は

$$\frac{85+75+80+75+45}{5} = \frac{360}{5} = 72$$

と計算され，72 であることがわかります．

よって，不偏分散は，

$$\frac{(85-72)^2+(75-72)^2+(80-72)^2+(75-72)^2+(45-72)^2}{4} = \frac{169+9+64+9+729}{4} = \frac{980}{4} = 245$$

と計算され，245 であることがわかります．

ここで，

$$\overline{X} - 2.78 \times \sqrt{\frac{s^2}{n}} \leq \mu \leq \overline{X} + 2.78 \times \sqrt{\frac{s^2}{n}}$$

に，$\overline{X} = 72$，$\sigma^2 = 245$，$n = 5$ を代入すると，

$$72 - 2.78 \times \sqrt{\frac{245}{5}} \leq \mu \leq 72 + 2.78 \times \sqrt{\frac{245}{5}}$$

つまり，

$$72 - 2.78 \times 7 \leq \mu \leq 72 + 2.78 \times 7$$

となるので，求める母平均 μ の 95%信頼区間は，これを計算した次になります．

$$52.54 \leq \mu \leq 91.46$$

第 11 章　母比率の区間推定

11.1　2 項分布の正規近似

問題 11-1

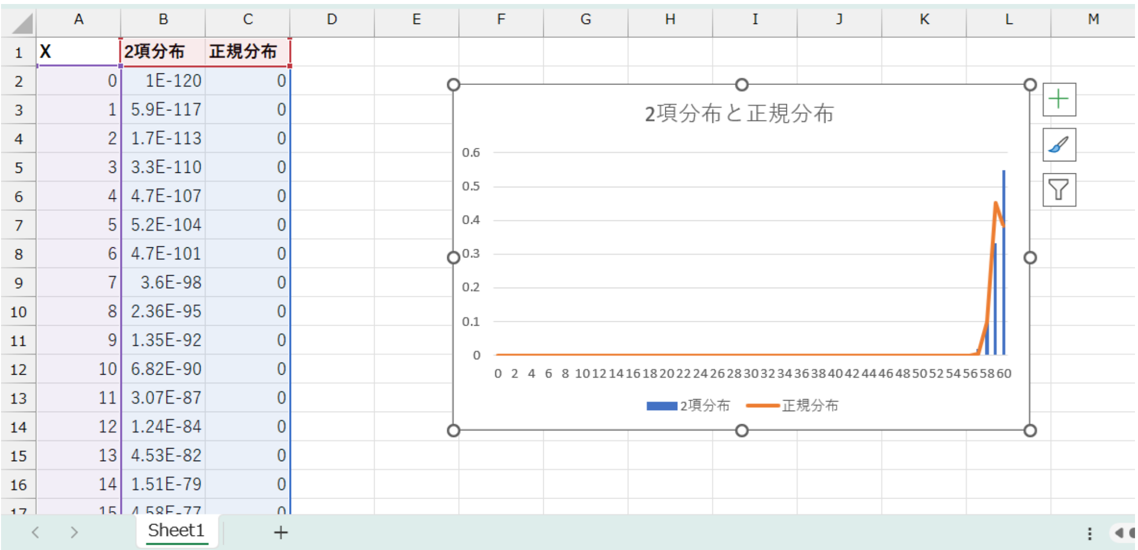
セル B2 を「=BINOM.DIST(A2,60,99/100,FALSE)」に入力しなおし，試行回数が 60，成功率 99/100 である 2 項分布 $B(60, 99/100)$ における成功数が 0 である確率を求めましょう．

そして，このセル（B2）の右下あたりをダブルクリックすることにより，成功数が 60 である確率までオートフィルで求めましょう．

セル B64 を「=60*(99/100)」に入力しなおし，$B(60, 99/100)$ の期待値 297/5 (= 59.4) を求めましょう．

また，セル B65 を「=60*(99/100)*(1-99/100)」に入力しなおし，$B(60, 99/100)$ の分散 297/500 (= 0.594) を求めましょう．

期待値と分散を変更したことにより，C 列が正規分布 $N(297/5, 297/500)$ に修正されます．



X	2項分布	正規分布
0	1E-120	0
1	5.9E-117	0
2	1.7E-113	0
3	3.3E-110	0
4	4.7E-107	0
5	5.2E-104	0
6	4.7E-101	0
7	3.6E-98	0
8	2.36E-95	0
9	1.35E-92	0
10	6.82E-90	0
11	3.07E-87	0
12	1.24E-84	0
13	4.53E-82	0
14	1.51E-79	0

グラフが変化し，2 項分布のグラフと正規分布のグラフのズレがめだつようになりました．つまり，2 項分布における成功率を 1 に近くすると（0.5 から遠くすると），正規近似の精度が下がりました．

問題 11-2

A 列（セル範囲 A2:A62）の X の値を，540, 541, $\cdots$, 600 に変更します．そのため，セル A2 を「540」，セル A3 を「541」に入力しなおし，そのセル範囲 A2:A3 を選択します．そして，その右下あたりにマウスポインタを合わせ「+」の形にして，この状態のままダブルクリックします．

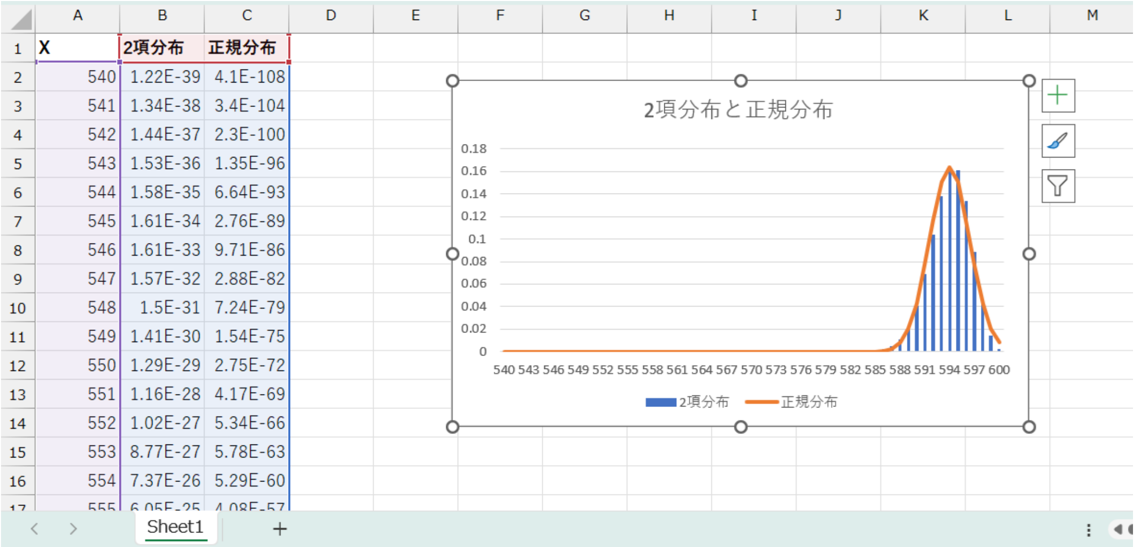
セル B2 を「=BINOM.DIST(A2,600,99/100,FALSE)」に入力しなおし，試行回数が 600，成功率 99/100 である 2 項分布 $B(600, 99/100)$ における成功数が 540 である確率を求めましょう．

そして，このセル（B2）の右下あたりをダブルクリックすることにより，成功数が 600 である確率までオートフィルで求めましょう（成功数は 539 以下も考えられますが省略します）．

セル B64 を「=600*(99/100)」に入力しなおし，$B(600, 99/100)$ の期待値 594 を求めましょう．

また，セル B65 を「=600*(99/100)*(1-99/100)」に入力しなおし，$B(600, 99/100)$ の分散 297/50 (= 5.94) を求めましょう．

期待値と分散を変更したことにより，C 列が正規分布 $N(594, 297/50)$ に修正されます．



グラフが変化し，問題 5-3 と同じように，2 項分布と正規分布のズレがめだたなくなりました．つまり，2 項分布における成功回数を大きくすると，正規近似の精度が上がりました．

11.1 の確認テスト

問 1

(1) np，$np(1-p)$

(2) 正規分布

問 2

2 項分布を正規分布で近似する際，試行回数が同じとすると，成功率が 0.5 から遠いほど近似の精度は下がります．

問 3

2 項分布を正規分布で近似する際，成功率が同じとすると，試行回数が大きいほど近似の精度は上がります．

11.2　母比率の信頼区間

問題 11-3

セル B1 に標本サイズである「1000」を入力します．

また，セル B2 に標本比率である「40%」を入力します．

セル D4 に「=NORM.S.INV(1-0.025)」と入力し，標準正規分布における上側 2.5%点を求めます．約 1.96 であることがわかります．

セル E6 に「=B2+D4*SQRT(B2*(1-B2)/B1)」と入力し，母比率 p の 95%信頼区間の上限を求めます．ホームタブの（数値グループにある）［パーセント スタイル］をクリックして，パーセントスタイルにしましょう．約 43.0% であることがわかります．

また，セル D6 に「=B2-D4*SQRT(B2*(1-B2)/B1)」と入力し，母比率 p の 95%信頼区間の下限を求めます．これもパーセントスタイルにしましょう．約 37.0% であることがわかります．

これらの数値について下図では，ホームタブの（数値グループにある）[小数点以下の表示桁数を減らす] ボタンをクリックすることにより，小数第 1 位までの表示にしています．

よって，母比率 p の 95%信頼区間はおおよそ

$$37.0\% \le p \le 43.0\%$$

となります．

	A	B	C	D	E
1	標本サイズn	1000			
2	標本比率	40%			
3					
4	標準正規分布における上側2.5%点			1.959963985	
5					
6	母比率pの95%信頼区間		37.0%	≤ 母比率p ≤	43.0%
7					
8					
9					

問題 11-4

セル B1 を標本サイズである「10000」に入力しなおします．

母比率 p の 95%信頼区間はおおよそ

$$39.0\% \le p \le 41.0\%$$

となります．

	A	B	C	D	E
1	標本サイズn	10000			
2	標本比率	40%			
3					
4	標準正規分布における上側2.5%点			1.959963985	
5					
6	母比率pの95%信頼区間		39.0%	≤ 母比率p ≤	41.0%
7					
8					
9					

ここで求めた「大きさ 10000 の標本から算出される母比率 p の 95%信頼区間」：

$$39.0\% \le p \le 41.0\%$$

は，問題 11-3 で求めた「大きさ 1000 の標本から算出される母比率 p の 95%信頼区間」：

$$37.0\% \le p \le 43.0\%$$

に比べて，幅が小さいことが確認できます．

問題 11-5

セル B2 を標本比率である「90%」に入力しなおします．

母比率 p の 95%信頼区間はおおよそ

$$88.1\% \leq p \leq 91.9\%$$

となります．

	A	B	C	D	E
1	標本サイズn	1000			
2	標本比率	90%			
3					
4	標準正規分布における上側2.5%点			1.959963985	
5					
6	母比率pの95%信頼区間		88.1%	≤ 母比率p ≤	91.9%
7					
8					
9					

ここで求めた「標本比率 90%の標本から算出される母比率 p の 95%信頼区間」：

$$88.1\% \leq p \leq 91.9\%$$

は，問題 11-3 で求めた「標本比率 40%の標本から算出される母比率 p の 95%信頼区間」：

$$37.0\% \leq p \leq 43.0\%$$

に比べて，幅が小さいことが確認できます．

問題 11-6

セル D1 に標本サイズである「180」を入力します．

また，セル D2 に「=COUNTIF(A2:A181,1)/D1」と入力し，標本比率を求めます．ホームタブの（数値グループにある）［パーセント スタイル］をクリックして，パーセントスタイルにしましょう．約 18.3% であることがわかります．

セル F4 に「=NORM.S.INV(1-0.025)」と入力し，標準正規分布における上側 2.5%点を求めます．約 1.96 であることがわかります．

セル G6 に「=D2+F4*SQRT(D2*(1-D2)/D1)」と入力し，母比率 p の 95%信頼区間の上限を求めます．これもパーセントスタイルにしましょう．約 24.0% であることがわかります．

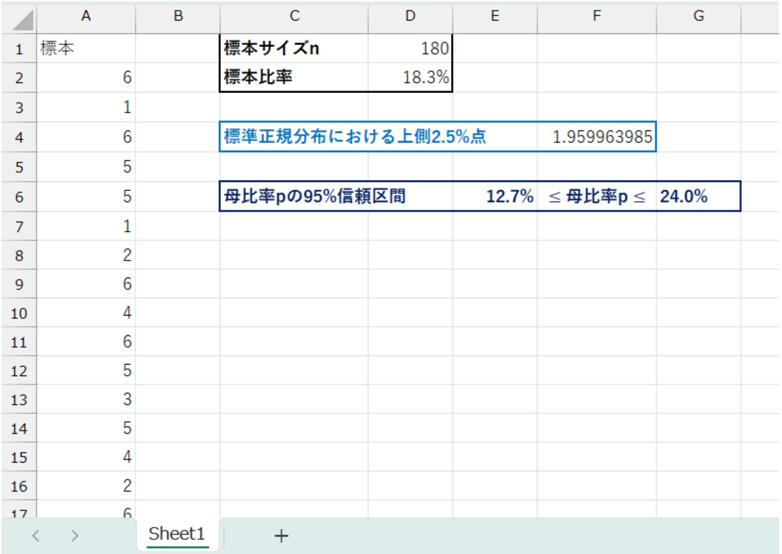
また，セル E6 に「=D2-F4*SQRT(D2*(1-D2)/D1)」と入力し，母比率 p の 95%信頼区間の下限を求めます．これもパーセントスタイルにしましょう．約 12.7% であることがわかります．

これらの数値について下図では，ホームタブの（数値グループにある）［小数点以下の表示桁数を減らす］ボタンをクリックすることにより，小数第 1 位までの表示にしています．

よって，母比率 p の 95%信頼区間はおおよそ

$$12.7\% \leq p \leq 24.0\%$$

となります．



	A	B	C	D	E	F	G
1	標本		標本サイズn	180			
2	6		標本比率	18.3%			
3	1						
4	6		標準正規分布における上側2.5%点			1.959963985	
5	5						
6	5		母比率pの95%信頼区間		12.7%	≤ 母比率p ≤	24.0%
7	1						
8	2						
9	6						
10	4						
11	6						
12	5						
13	3						
14	5						
15	4						
16	2						
17	6						

Sheet1

問題 11-7

セル C4 を「標準正規分布における上側 5%点」に入力しなおします．

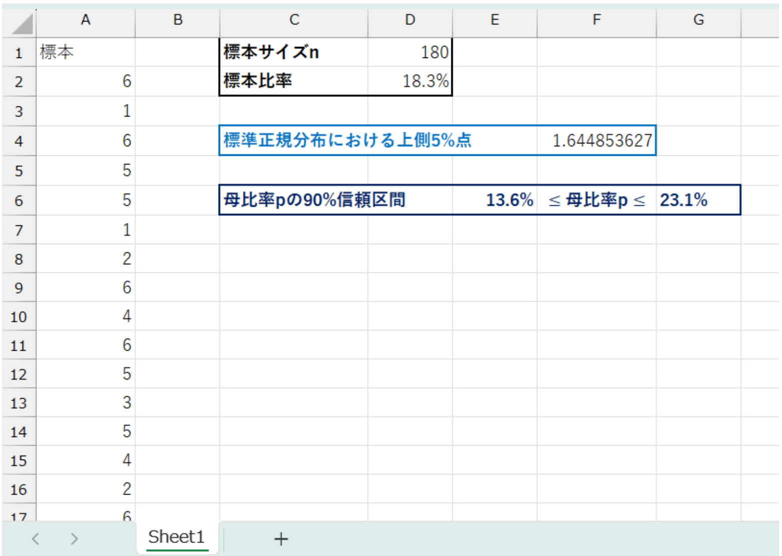
セル F4 を「=NORM.S.INV(1-0.05)」に入力しなおし，標準正規分布における上側 5%点を求めます．約 1.64 であることがわかります．

セル C6 を「母比率 p の 90%信頼区間」に入力しなおします．

母比率 p の 90%信頼区間はおおよそ

$$13.6\% \leq p \leq 23.1\%$$

となります．



	A	B	C	D	E	F	G
1	標本		標本サイズn	180			
2	6		標本比率	18.3%			
3	1						
4	6		標準正規分布における上側5%点			1.644853627	
5	5						
6	5		母比率pの90%信頼区間		13.6%	≤ 母比率p ≤	23.1%
7	1						
8	2						
9	6						
10	4						
11	6						
12	5						
13	3						
14	5						
15	4						
16	2						
17	6						

Sheet1

ここで求めた「母比率 p の 90%信頼区間」：

$$13.6\% \leq p \leq 23.1\%$$

は，問題 11-6 で求めた「母比率 p の 95%信頼区間」：

$$12.7\% \leq p \leq 24.0\%$$

に比べて，幅が小さいことが確認できます．

問題 11-8

セル C4 を「標準正規分布における上側 0.5%点」に入力しなおします．

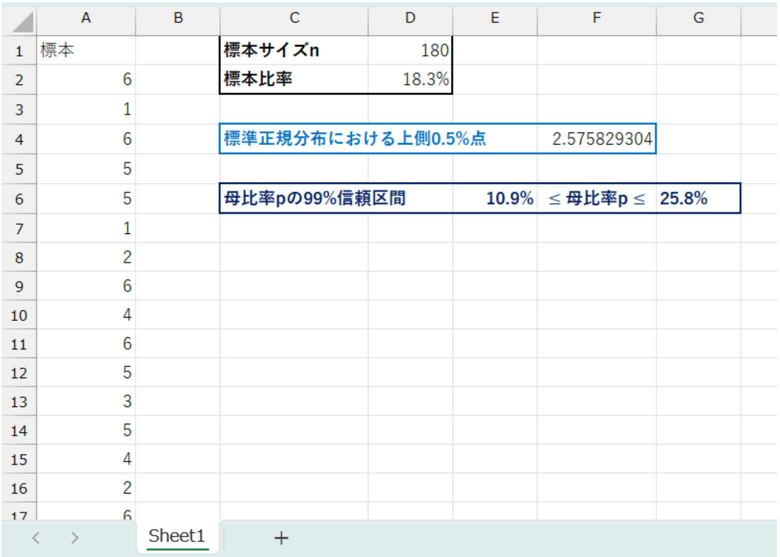
セル F4 を「=NORM.S.INV(1-0.005)」に入力しなおし，標準正規分布における上側 0.5%点を求めます．約 2.58 であることがわかります．

セル C6 を「母比率 p の 99%信頼区間」に入力しなおします．

母比率 p の 99%信頼区間はおおよそ

$$10.9\% \le p \le 25.8\%$$

となります．



	A	B	C	D	E	F	G
1	標本		標本サイズn	180			
2	6		標本比率	18.3%			
3	1						
4	6		標準正規分布における上側0.5%点			2.575829304	
5	5						
6	5		母比率pの99%信頼区間		10.9%	≤ 母比率p ≤	25.8%
7	1						
8	2						
9	6						
10	4						
11	6						
12	5						
13	3						
14	5						
15	4						
16	2						
17	6						

Sheet1

ここで求めた「母比率 p の 99%信頼区間」：

$$10.9\% \le p \le 25.8\%$$

は，問題 11-6 で求めた「母比率 p の 95%信頼区間」：

$$12.7\% \le p \le 24.0\%$$

に比べて，幅が大きいことが確認できます．

11.2 の確認テスト

問 1

(1) 母比率

(2) n，p，標本比率

(3) p，$p(1-p)/n$

(4) 2.5，2.5

(5) 5，5

(6) 0.5，0.5

問 2

(1) $\widehat{p}-1.64\times\sqrt{\dfrac{\widehat{p}\,(1-\widehat{p})}{n}}\leq p\leq\widehat{p}+1.64\times\sqrt{\dfrac{\widehat{p}\,(1-\widehat{p})}{n}}$

に，$\widehat{p}=0.5=\dfrac{1}{2}$，$n=400$ を代入すると，

$$\frac{1}{2}-1.64\times\sqrt{\frac{1/2\,(1-1/2)}{400}}\leq p\leq\frac{1}{2}+1.64\times\sqrt{\frac{1/2\,(1-1/2)}{400}}$$

つまり，

$$\frac{1}{2}-1.64\times\frac{1}{40}\leq p\leq\frac{1}{2}+1.64\times\frac{1}{40}$$

となるので，求める母比率 p の 90%信頼区間は，これを計算した次になります．

$$0.459\leq p\leq 0.541$$

また，95%信頼区間は，2 つ上の式において 1.64 を 1.96 におきかえた

$$\frac{1}{2}-1.96\times\frac{1}{40}\leq p\leq\frac{1}{2}+1.96\times\frac{1}{40}$$

を計算した次になります．

$$0.451\leq p\leq 0.549$$

99%信頼区間は，同じ式において 1.64 を 2.58 におきかえた

$$\frac{1}{2}-2.58\times\frac{1}{40}\leq p\leq\frac{1}{2}+2.58\times\frac{1}{40}$$

を計算した次になります

$$0.4355\leq p\leq 0.5645$$

(2) $\widehat{p}-1.64\times\sqrt{\dfrac{\widehat{p}\,(1-\widehat{p})}{n}}\leq p\leq\widehat{p}+1.64\times\sqrt{\dfrac{\widehat{p}\,(1-\widehat{p})}{n}}$

に，$\widehat{p}=0.1=\dfrac{1}{10}$，$n=400$ を代入すると，

$$\frac{1}{10}-1.64\times\sqrt{\frac{1/10\,(1-1/10)}{400}}\leq p\leq\frac{1}{10}+1.64\times\sqrt{\frac{1/10\,(1-1/10)}{400}}$$

つまり，

$$\frac{1}{10}-1.64\times\frac{3}{200}\leq p\leq\frac{1}{10}+1.64\times\frac{3}{200}$$

となるので，求める母比率 p の 90%信頼区間は，これを計算した次になります．

$$0.0754\leq p\leq 0.1246$$

また，95%信頼区間は，2 つ上の式において 1.64 を 1.96 におきかえた

$$\frac{1}{10}-1.96\times\frac{3}{200}\leq p\leq\frac{1}{10}+1.96\times\frac{3}{200}$$

を計算した次になります．

$$0.0706\leq p\leq 0.1294$$

99%信頼区間は，同じ式において 1.64 を 2.58 におきかえた

$$\frac{1}{10} - 2.58 \times \frac{3}{200} \leq p \leq \frac{1}{10} + 2.58 \times \frac{3}{200}$$

を計算した次になります

$$0.0613 \leq p \leq 0.1387$$

問 3

$$\widehat{p} - 1.96 \times \sqrt{\frac{\widehat{p}\,(1-\widehat{p})}{n}} \leq p \leq \widehat{p} + 1.96 \times \sqrt{\frac{\widehat{p}\,(1-\widehat{p})}{n}}$$

に，$\widehat{p} = \dfrac{20000}{40000} = \dfrac{1}{2}$，$n = 40000$ を代入すると，

$$\frac{1}{2} - 1.96 \times \sqrt{\frac{1/2\,(1-1/2)}{40000}} \leq p \leq \frac{1}{2} + 1.64 \times \sqrt{\frac{1/2\,(1-1/2)}{40000}}$$

つまり，

$$\frac{1}{2} - 1.96 \times \frac{1}{400} \leq p \leq \frac{1}{2} + 1.96 \times \frac{1}{400}$$

となるので，求める母比率 p の 95%信頼区間は，これを計算した次になります．

$$0.4951 \leq p \leq 0.5049$$

問 4

標本比率が同じであるとすると，標本サイズが大きいほど，その標本から算出される母比率の信頼区間の幅は小さくなります．

問 5

標本サイズが同じであるとすると，標本比率が 0.5 に近いほど，その標本から算出される母比率の信頼区間の幅は大きくなります．

問 6

ある標本から算出される母比率の信頼区間の幅は，信頼係数が大きいほど大きくなります．

第 12 章　母比率の検定

12.1　仮説検定の考え方

問題 12-1

セル I1 に「=1/6」と入力します．

帰無仮説は，「このサイコロは正しいサイコロである」です．これは，「このサイコロを投げたとき 1 の出る確率（母比率）は 1/6 である」ということになります．

また，セル I2 に「=I1」と入力します．

対立仮説は，「このサイコロは正しいサイコロでない」です．これは，「このサイコロを投げたとき 1 の出る確率（母比率）は 1/6 でない」ということになります．

コインを投げた回数，つまり，標本サイズ n である「180」を，セル D4 に入力します．

セル G6 に「=D4」と入力し，セル I6 に「=I1」と入力します．

「帰無仮説を仮定すると，このサイコロを 180 回振ったときに 1 の出る回数 X は，試行回数 180，成功率 1/6 $(= 0.1666\cdots)$ の 2 項分布 $B(180, 1/6)$ に従う」ということになります．

セル F7 に「=D4*I1」と入力し，セル H7 に「=D4*I1*(1-I1)」と入力します．

「帰無仮説を仮定すると，このサイコロを 180 回振ったときに 1 の出る回数 X は，期待値 30，分散 25 の正規分布 $N(30, 25)$ に従う」ということになります．

問題 12-2

セル D9 に，「=COUNTIF(A2:A181,1)」と入力して，サイコロを 180 回振ったときに 1 が出た回数を求めます．33 であることがわかります．

セル D10 に，「=(D9-F7)/SQRT(H7)」と入力し，1 が出た回数 33 を標準化します．z 値は 0.6 であることがわかります．

セル F12 に「=NORM.S.INV(1-0.025)」と入力し，標準正規分布における上側 2.5%点を求めます．約 1.96 であることがわかります．

ここで，確率分布が左右対称であり，上側 2.5%点が約 1.96 であることから，下側 2.5%点は約 −1.96 であることもわかります．

セル E14 に「=IF(ABS(D10)>F12,” 棄却する”,” 棄却しない”)」と入力し，z 値 0.6 が標準正規分布において，上側 2.5%点より大きい，または，下側 2.5%点より小さいなら「棄却する」と表示し，そうでないなら「棄却しない」と表示するように設定しましょう．

「棄却しない」と表示されます．

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	標本		帰無仮説：	このサイコロを振ったとき1の出る確率（母比率）は					0.166667	である（このサイコロは正しいサイコロである）						
2	6		対立仮説：	このサイコロを振ったとき1の出る確率（母比率）は					0.166667	でない（このサイコロは正しいサイコロでない）						
3	1															
4	6		標本サイズn	180												
5	5															
6	5		帰無仮説を仮定すると，1の出る回数Xは，			試行回数	180	，成功率	0.166667	の2項分布に従う						
7	1		したがって，1の出る回数Xは，		期待値	30	，分散	25	の正規分布に従う							
8	2															
9	6		1が出た回数X	33												
10	4		z	0.6	（検定統計量）											
11	6															
12	5		標準正規分布における上側2.5%点			1.959964										
13	3															
14	5		有意水準5%の両側検定		棄却しない			成り立てば帰無仮説を棄却しない →				-1.95996	≤	0.6	≤	1.959964
15	4															
16	2															
17	6															

Sheet1

以上より，帰無仮説を棄却しないことにします．つまり，「このサイコロは正しいサイコロでない」とは判断しないことにします．

12.1 の確認テスト

問 1

(1) 帰無仮説，棄却

(2) 対立仮説

(3) 棄却域

(4) 両側検定

(5) 有意水準

問 2

(1) 1/2，1/2

(2) 5

(3) 100，1/2，50，25，25

(4) 2.5，2.5

(5) 50，2

(6) 2，2.5

12.2　母比率の検定

問題 12-3

セル G7 に「=I6」と入力し，セル I7 に「=I6*(1-I6)/D4」と入力します．

帰無仮説を仮定すると，「このサイコロを 180 回振ったときに 1 の出る比率は，期待値 $1/6$ $(= 0.1666\cdots)$，分散 $(1/6 \cdot (1-1/6)/180 =)$ $1/1296$ $(=$ 約$0.000772)$ の正規分布 $N(1/6, 1/1296)$ に従う」ということになります．

セル D9 に，「=COUNTIF(A2:A181,1)/D4」と入力して，サイコロを 180 回振ったときに 1 が出た比率（標本比率）を求めます．(33/180 =) $0.18333\cdots$ であることがわかります．

セル D10 に，「=(D9-G7)/SQRT(I7)」と入力し，1 が出た比率 $0.18333\cdots$ を標準化します．z 値は 0.6 であることがわかります．

セル F12 に「=NORM.S.INV(1-0.025)」と入力し，標準正規分布における上側 2.5%点を求めます．約 1.96 であることがわかります．

ここで，確率分布が左右対称であり，上側 2.5%点が約 1.96 であることから，下側 2.5%点は約 -1.96 であることもわかります．

セル E14 に「=IF(ABS(D10)>F12,"棄却する","棄却しない")」と入力し，z 値 4.1 が標準正規分布において，上側 2.5%点より大きい，または，下側 2.5%点より小さいなら「棄却する」と表示し，そうでないなら「棄却しない」と表示するように設定しましょう．

「棄却しない」と表示されます．

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	標本		帰無仮説：	このサイコロを振ったとき1の出る確率（母比率）は					0.166667	である（このサイコロは正しいサイコロである）						
2	6		対立仮説：	このサイコロを振ったとき1の出る確率（母比率）は					0.166667	でない（このサイコロは正しいサイコロでない）						
3	1															
4	6		標本サイズn	180												
5	5															
6	5		帰無仮説を仮定すると，1の出る回数Xは，			試行回数	180	，成功率	0.166667	の2項分布に従う						
7	1		したがって，1の出る比率（標本比率）は，			期待値	0.166667	，分散	0.000772	の正規分布に従う						
8	2															
9	6		標本比率	0.183333												
10	4		z	0.6	（検定統計量）											
11	6															
12	5		標準正規分布における上側2.5%点			1.959964										
13	3															
14	5		有意水準5%の両側検定		棄却しない			成り立てば帰無仮説を棄却しない →				-1.95996	≤	0.6	≤	1.959964
15	4															
16	2															
17	6															

Sheet1

以上より，帰無仮説を棄却しないことにします．つまり，「このサイコロは正しいサイコロでない」とは判断しないことにします．

標本比率が従う正規分布を考えることによりおこなってみても，得られている標本から算出される検定統計量は，問題 12-2 と同じ値（0.6）になりました．つまり，問題 12-3 の検定は，1 の出る回数が従う正規分布を考えることによりおこなった問題 12-2 の検定と同じ結果になることが確認されました．

問題 12-4

アンケートをとった N 大学の学生 1000 人のパスポート保有率，つまり，標本比率である「0.32」を，セル B9 に入力しなおします．

セル C14 は「棄却しない」と表示されます．

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	**帰無仮説：**	大学生のパスポート保有率（母比率）は				0.3	である						
2	**対立仮説：**	大学生のパスポート保有率（母比率）は				0.3	でない						
3													
4	**標本サイズn**	1000											
5													
6	帰無仮説を仮定すると，保有人数Xは，**試行回数**				1000	，**成功率**	0.3	**の2項分布**に従う					
7	よって，保有率（標本比率）は，**期待値**			0.3	，**分散**	0.00021	**の正規分布**に従う						
8													
9	**標本比率**	0.32											
10	**z**	1.380131	（検定統計量）										
11													
12	**標準正規分布における上側2.5%点**			1.959964									
13													
14	**有意水準5%の両側検定**		棄却しない		成り立てば帰無仮説を棄却しない →				-1.95996	≤	1.380131	≤	1.959964
15													
16													

以上より，帰無仮説を棄却しないことにします．つまり，「大学生のパスポート保有率（母比率）が 0.3 でない」とは判断しないことにします．

問題 12-5

アンケートをとった N 大学の学生 1000 人のパスポート保有率，つまり，標本比率である「0.34」を，セル B9 に入力しなおします．

セル C14 は「棄却する」と表示されます．

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	**帰無仮説：**	大学生のパスポート保有率（母比率）は				0.3	である						
2	**対立仮説：**	大学生のパスポート保有率（母比率）は				0.3	でない						
3													
4	**標本サイズn**	1000											
5													
6	帰無仮説を仮定すると，保有人数Xは，**試行回数**				1000	，**成功率**	0.3	**の2項分布**に従う					
7	よって，保有率（標本比率）は，**期待値**			0.3	，**分散**	0.00021	**の正規分布**に従う						
8													
9	**標本比率**	0.34											
10	**z**	2.760262	（検定統計量）										
11													
12	**標準正規分布における上側0.5%点**			2.575829									
13													
14	**有意水準1%の両側検定**		棄却する		成り立てば帰無仮説を棄却しない →				-2.57583	≤	2.760262	≤	2.575829
15													
16													

以上より，帰無仮説を棄却し，対立仮説を採用することにします．つまり，「大学生のパスポート保有率（母比率）は 0.3 でない」といえそうと判断することにします．

12.2 の確認テスト

問 1

(1) 1/2，1/2

(2) 5

(3) 100，1/2，1/2，1/400，1/2

(4) 2.5，2.5

(5) 1/400，-1.8

(6) -1.8，2.5，2.5

問 2

(1) 1/10，1/10

(2) 1

(3) 10000，1/10，1/10，9/1000000，9/1000000

(4) 0.5，0.5

(5) 1/10，$-8/3$ (= 約 -2.67)

(6) $-8/3$，0.5

第 13 章　母平均の検定

13.1　母平均の検定（母分散既知）

問題 13-1

B クラス 70 名の生徒の平均点，つまり，標本平均である「47.7」を，セル B10 に入力しなおします．

セル B11 の値が変更され，z 値は約 -1.92 であることがわかります．

セル C15 は「棄却しない」と表示されます．

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	**帰無仮説：**	このテストの平均点（母平均）は			50	である							
2	**対立仮説：**	このテストの平均点（母平均）は			50	でない							
3													
4	**母分散**	100											
5													
6	**標本サイズn**	70											
7													
8	帰無仮説を仮定すると，標本平均は，**期待値**				**50**	**，分散**	**1.428571**	**の正規分布**に従う					
9													
10	**標本平均**	47.7											
11	**z**	**-1.92432**	（検定統計量）										
12													
13	**標準正規分布における上側2.5%点**			1.959964									
14													
15	**有意水準5%の両側検定**		棄却しない		成り立てば帰無仮説を棄却しない →				-1.95996	≤	**-1.92432**	≤	1.959964
16													

以上より，帰無仮説を棄却しないことにします．つまり，「このテストの平均点（母平均）は 50 でない」（「B クラスの生徒の点数はふつうでない」）とは判断しないことにします．

問題 13-2

C クラス 70 名の生徒の平均点，つまり，標本平均である「46.9」を，セル B10 に入力しなおします．

セル B11 の値が変更され，z 値は約 -2.59 であることがわかります．

セル C15 は「棄却する」と表示されます．

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	帰無仮説：	このテストの平均点（母平均）は			50	である							
2	対立仮説：	このテストの平均点（母平均）は			50	でない							
3													
4	母分散	100											
5													
6	標本サイズn	70											
7													
8	帰無仮説を仮定すると，標本平均は，期待値				50	，分散	1.428571	の正規分布に従う					
9													
10	標本平均	46.9											
11	z	-2.59365	（検定統計量）										
12													
13	標準正規分布における上側0.5%点			2.5758293									
14													
15	有意水準1%の両側検定		棄却する		成り立てば帰無仮説を棄却しない →				-2.57583	≤	-2.59365	≤	2.575829
16													

Sheet1

以上より，帰無仮説を棄却し，対立仮説を採用することにします．つまり，「このテストの平均点（母平均）は 50 でない」（「C クラスの生徒の点数はふつうでない」）といえそうと判断することにします．

13.1 の確認テスト

問 1

(1) 100， 100

(2) 5

(3) 100， 1/4， 1/4

(4) 2.5， 2.5

(5) 100， −2.2

(6) −2.2， 2.5

問 2

問 1 の手順 2 において，有意水準を 5%にしましたが，本問では，有意水準を 1%にします．

問 1 の手順 5 で求めた統計量 −2.2 は「標準正規分布における上側 0.5%点」より大きくもなく，下側 0.5%点より小さくもないので，帰無仮説を棄却しません．

つまり，「これらの製品の重さはふつうでない」とは判断しないことにします．

13.2 母平均の検定（母分散未知）

問題 13-3

E クラスの児童の身長の平均値，つまり，標本平均である「135」を，セル B10 に入力しなおします．

セル B11 の値が変更され，T への変換後の値は約 1.26 であることがわかります．

セル C15 は「棄却しない」と表示されます．

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	帰無仮説：	小学4年生の身長の平均値（母平均）は				134	cmである						
2	対立仮説：	小学4年生の身長の平均値（母平均）は				134	cmでない						
3													
4	標本サイズn	40											
5													
6	不偏分散	25											
7													
8	帰無仮説を仮定すると，標本平均を T へ変換したものは，自由度					39	のt分布に従う						
9													
10	標本平均	135											
11	t	1.264911	（検定統計量）										
12													
13	自由度	39	のt分布における上側2.5%点			2.022691							
14													
15	有意水準5%の両側検定		棄却しない		成り立てば帰無仮説を棄却しない →				-2.02269	≦	1.264911	≦	2.022691
16													

Sheet1

以上より，帰無仮説を棄却しないことにします．つまり，「小学 4 年生の身長の平均値（母平均）は 134 cm でない」（「E クラスの児童の身長はふつうでない」）とは判断しないことにします．

問題 13-4

F クラスの児童の身長の不偏分散である「21」を，セル B6 に入力しなおします．

セル B11 の値が変更され，T への変換後の値は約 2.76 であることがわかります．

セル C15 は「棄却する」と表示されます．

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	帰無仮説：	小学4年生の身長の平均値（母平均）は				134	cmである						
2	対立仮説：	小学4年生の身長の平均値（母平均）は				134	cmでない						
3													
4	標本サイズn	40											
5													
6	不偏分散	21											
7													
8	帰無仮説を仮定すると，標本平均は，自由度				39	のt分布に従う							
9													
10	標本平均	136											
11	t	2.760262	（検定統計量）										
12													
13	自由度	39	のt分布における上側0.5%点			2.707913							
14													
15	有意水準1%の両側検定		棄却する		成り立てば帰無仮説を棄却しない →				-2.70791	≦	2.760262	≦	2.707913
16													

Sheet1

以上より，帰無仮説を棄却し，対立仮説を採用することにします．つまり，「小学 4 年生の身長の平均値（母平均）は 134 cm でない」（「F クラスの児童の身長はふつうでない」）といえそうと判断することにします．

13.2 の確認テスト

問 1

(1) 300，300

(2) 5

(3) 25，9
(4) 9，9，2.5，2.5
(5) 300，3.4
(6) 3.4，9，2.5

問 2

問 1 の手順 2 において，有意水準を 5%にしましたが，本問では，有意水準を 1%にします．

問 1 の手順 5 で求めた統計量 3.4 は「自由度 9 の t 分布における上側 0.5%点」より大きいので，帰無仮説を棄却します．

これより，対立仮説を採用することにします．つまり，「これらのりんごの重さはふつうでない」といえそうと判断することにします．