

Excel による（疑似）ヒストグラムの作成

1. 平成 18 年分の民間給与実態統計調査結果（税務統計から見た民間給与の実態）によると、1 年間を通じて勤務した給与所得者の給与階級別の所得者数は、表 1 のようであった。この表を Excel のワークシート上にデータ化し、その分布の様子を男女別に比較できるようにグラフで示してみよう。ホームページ上の shotoku.xlsx をダウンロードして利用すればよい。合計欄は Excel の計算式（SUM を使用）で求め、表 1 の値と一致することを確認しよう（入力ミスのないことの検証）。

表 1 平成 18 年分・給与階級別所得者数（千人）

給与階級	所得者数（男）	所得者数（女）
100 万円以下	728	2,876
100 万円超～200 万円以下	1,902	4,721
200 万円超～300 万円以下	3,287	3,893
300 万円超～400 万円以下	4,846	2,716
400 万円超～500 万円以下	4,721	1,529
500 万円超～600 万円以下	3,551	762
600 万円超～700 万円以下	2,492	367
700 万円超～800 万円以下	1,815	187
800 万円超～900 万円以下	1,227	102
900 万円超～1,000 万円以下	806	75
1,000 万円超～1,500 万円以下	1,545	109
1,500 万円超～2,000 万円以下	329	35
2,000 万円超～2,500 万円以下	100	12
2,500 万円超	102	9
合計	27,451	17,393

2. とりあえず、データ表が埋まったら、図 2 を見本に、棒グラフで図化してみなさい。
3. 次に、男女それぞれの平均所得額¹はいくらとなるか？Excel 上で計算させてみる。
 ポイント 1：平均所得額は、所得額の合計を所得者数の合計で割って求まる。
 ポイント 2：個々のデータは分からないので、階級ごとの合計所得を求める必要がある。
 ポイント 3：階級内の分布は不明だが、階級の間中値をその階級の代表値と仮定する。
 ”100 万円以下” は下界を 0 と想定して階級の代表値を 50 万円と考える。”2,500 万円超” は、階級の上界が不明であるが、仮に 5,000 万円とみなして、階級の代表値を 3,750 万円としてみる。

¹ 厳密には給与と所得は異なるが、ここでは区別しないで考えておく。

以上の方針に従って、シート中の「階級代表値」の列を、50, 150, …, 3750 と適切な値で埋めたら、「総額 (男)」と「総額 (女)」の列を適切な計算式で埋めてみなさい。
結果、男性の平均所得額が約 540 万円、女性の平均所得額が約 269 万円となることを確認しなさい。

4. 2 の手順で、単純な棒グラフ (図 2) を使って表現してしまうと、分布が不自然に見える。 ” ヒストグラム ” になるように、各階級の人数を階級幅に応じて補正し、図化してみよう。
ポイント 1 : 幅 100 万円の階級幅を 1 としたら、階級幅 500 万円の人数は (幅が 5 倍なので高さを) 1/5 となるように補正する。
ポイント 2 : グラフは、横軸 : 階級の代表値, 縦軸 : 補正後の人数, とした ” 散布図 ” のマーカー付きの折れ線形式を用いる。

「密度 (男/女)」の列は、

$$\underbrace{\text{階級の人数} / \text{合計人数} \times 100}_{\text{合計人数に対する構成割合}} / \underbrace{\text{階級幅}}_{\text{階級幅による補正}}$$

のような計算となるように式を埋めてみなさい。合計人数は、式が複製されてもアドレスが補正されてしまわないように、B\$17 のように、行番号の直前に \$ 記号を入れて仕上げる工夫をしておこう。

5. 「階級代表値」を横軸、「密度 (男)」「密度 (女)」を縦軸に使う、散布図を描く。連続しない範囲をドラッグで指定するには、図 1 のように、2 つめの範囲を、キーボードの **Ctrl** キーを押しながらドラッグ操作すればよい。図 3 を見本に仕上げてみよう。

	D	E	F	G	H	I
	階級幅	階級代表値	総額(男)	総額(女)	密度(男)	密度(女)
76	100	50	36,400	143,800	0.03	0.17
21	100	150	285,300	708,150	0.07	0.27
33	100	250	821,750	973,250	0.12	0.22
6	100	350	1,696,100	950,600	0.18	0.16
29	100	450	2,124,450	688,050	0.17	0.09
32	100	550			0.13	0.04
37	100	650			0.09	0.02
37	100	750			0.07	0.01
22	100	850	1,042,950	86,700	0.04	0.01
75	100	950	765,700	71,250	0.03	0.00
39	500	1,250	1,931,250	136,250	0.01	0.00
35	500	1,750	575,750	61,250	0.00	0.00
2	500	2,250	225,000	27,000	0.00	0.00
8	2500	3,750	382,500	33,750	0.00	0.00
33		合計	14,821,250	4,677,950		
		平均給与	540	269		

図 1 離れた範囲の同時選択の手順

(仕上げの見本)

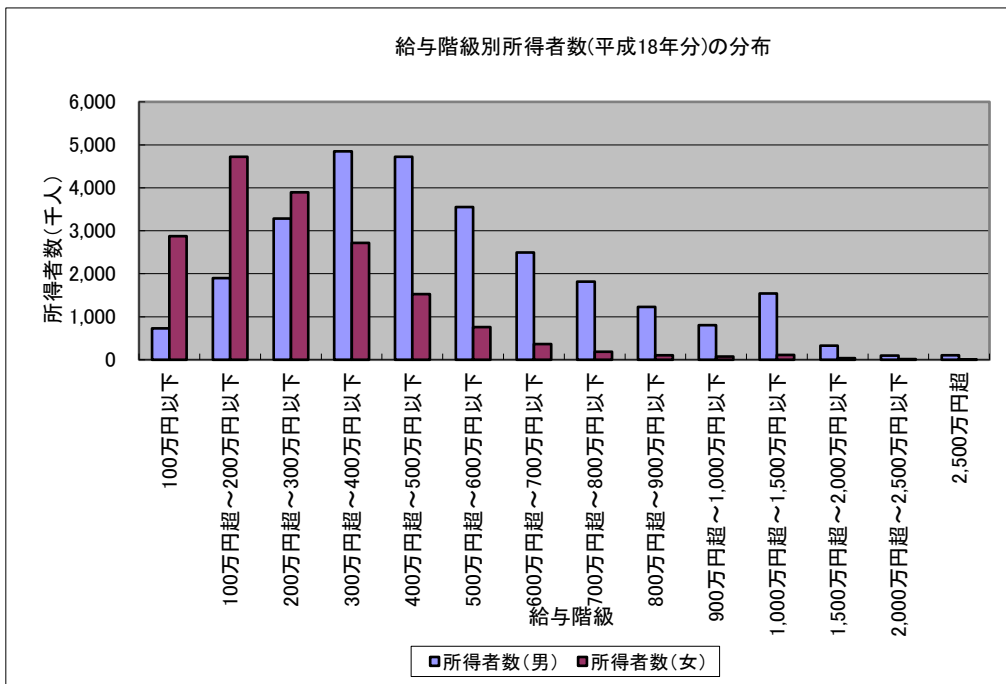


図 2 棒グラフによる分布の図化例

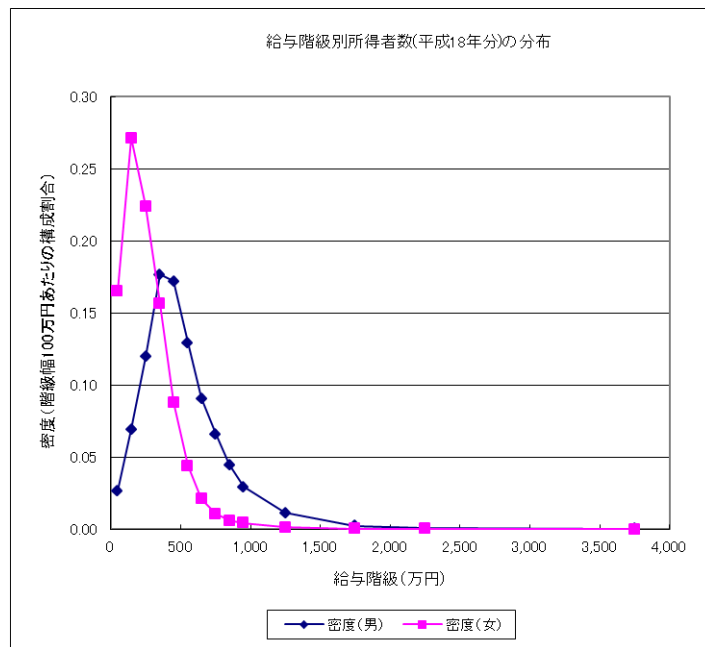


図 3 散布図を使用した擬似ヒストグラムによる分布の図化例

作業した結果を，図 4 のように Excel シート 1 枚で印刷できるようにレイアウトしてみよう。

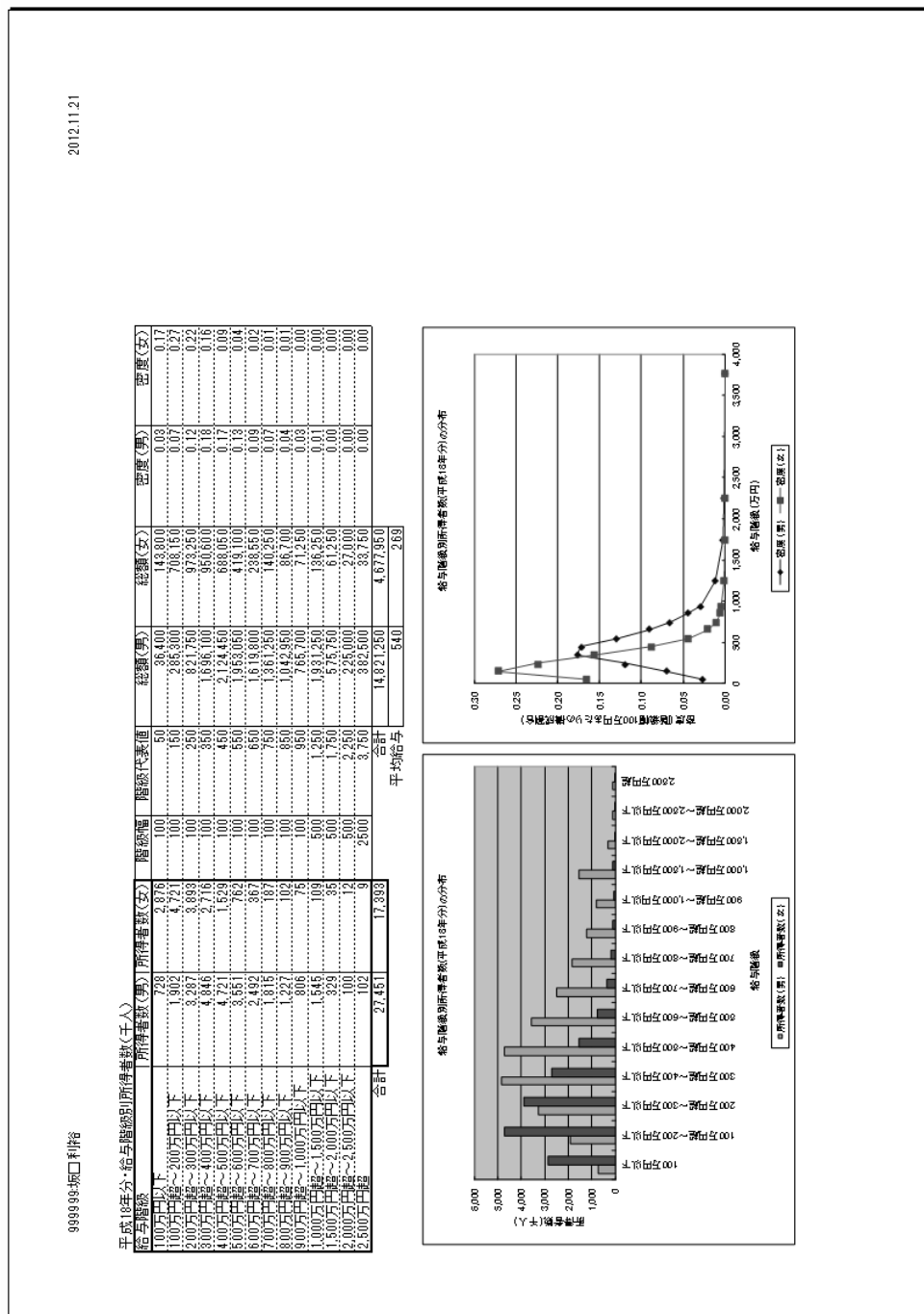


図 4 仕上げ見本

余力のある者は，shotoku.xlsx に含まれている平成 23 年度分および平成 28 年度分のデータについても同様の作業²を行い，結果を比較してみよう。可能であれば，それらを 1 枚のグラフにまとめて表示するように工夫してみなさい。

² ヒント：計算の過程は同じなのだから，「シート間でも式のコピペが可能」であることを利用すればより効率的に作業ができるはず