

第13回 統計的データの整理 知っておくと便利なさまざまな技法

じっさいの調査報告書や統計を使った論文を書く場合、SPSS の出力結果をそのまま貼り付けるのではなく、必要な数値だけコンパクトに示した表やグラフを作成することが望ましい。最終回は、復習もかねて、表やグラフをオリジナルに作成するノウハウを紹介する。

1. 度数分布表

SPSS 出力をそのまま貼り付けた場合

INCOME						
		度数	パーセント	有効パーセント	累積パーセント	
有効	1.00	-300	356	36.1	45.0	45.0
	2.00	300-700	261	26.4	33.0	78.0
	3.00	700-1100	129	13.1	16.3	94.3
	4.00	1100-	45	4.6	5.7	100.0
	合計		791	80.1	100.0	
欠損値	9.00		196	19.9		
合計			987	100.0		

作り直した場合

表1.1 仕事による個人所得の分布(有職者のみ)

所得	%	N
300 万円未満	45.0	356
300 ～ 700 万円未満	33.0	261
700 ～ 1100 万円未満	16.3	129
1100 万円以上	5.7	45
合計	100.0	791

・必要な要素

①表の番号とタイトル、②カテゴリーラベル、③有効パーセント（通常、欠損値を除く百分率のみを表示する）、④サンプル数（通常、断りなく N と表示してよい）。

・スキル

数字をすべて打ち直すのは面倒だし、間違いのもとである。そこで、SPSS 出力結果をエクセルにコピーして、加工するとよい。

① SPSS の出力結果の表を右クリックして、テキストをコピーを選択する（または、Ctrl+C のキー操作）。

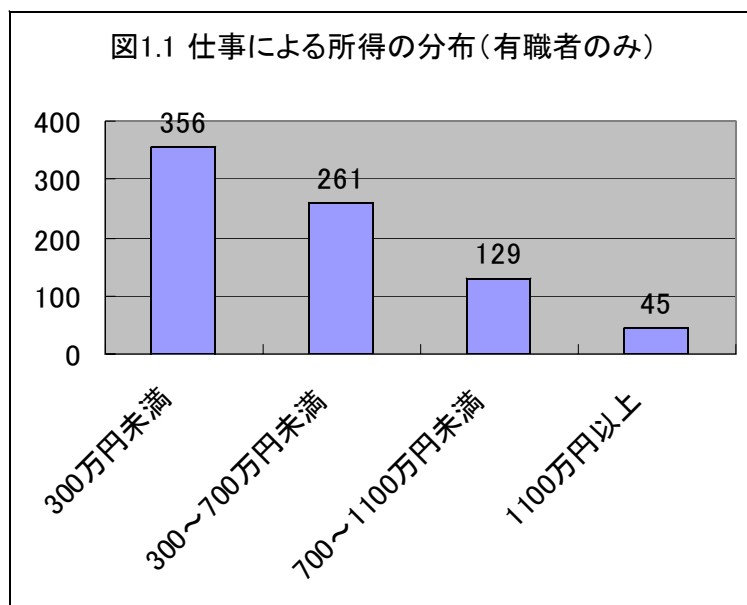
②エクセルを立ち上げて、コピーしたい場所の左上に当たる場所のセルをアクティブにして Ctrl+V のキー操作をする。

→エクセルの各セルに数値が入る。

【課題1】じっさいに income の度数分布表を出力し、エクセル上で表を作ってみましょう。

度数分布表のグラフ表示

一般向け報告書のような場合には、度数分布表をグラフ表示した方が分布がわかりやすい。SPSS にも、グラフ機能があるが、自由に細工ができず、じっさいには使えない。きれいなグラフをつくらうと思うなら、エクセルのグラフ機能を使うことをお勧めします。



基本的な手順

(1) SPSS の出力結果を、エクセルにコピーする。

① SPSS の出力結果の表を右クリックして、テキストをコピーを選択する (または、Ctrl+C のキー操作)。

②エクセルを立ち上げて、コピーしたい場所の左上に当たる場所のセルをアクティブにして Ctrl+V のキー操作をする。

→エクセルの各セルに数値が入る。

③その隣に、グラフに必要なデータのための表を作る (数値などはコピーする)。

300万円未満	356
300~700万円未満	261
700~1100万円未満	129
1100万円以上	45

図1.1 エクセル上で打ち直された表

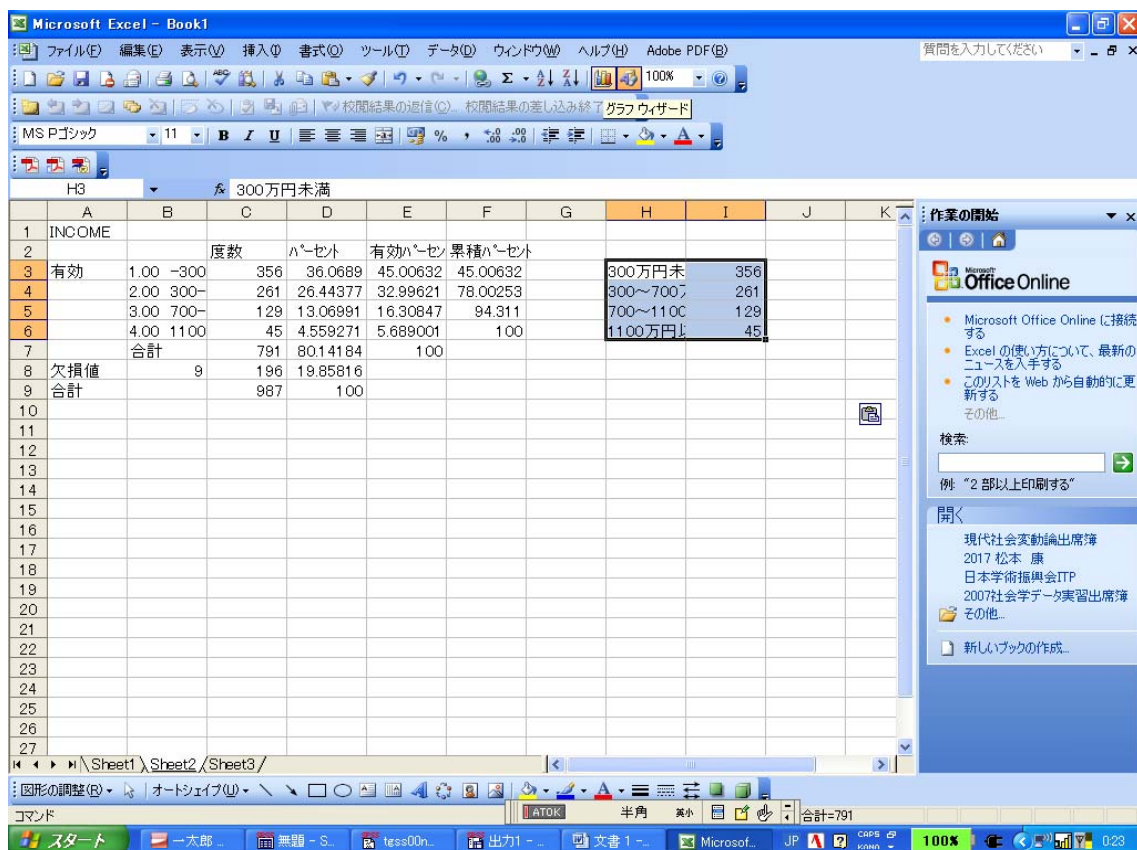
(2)エクセル上でグラフを作成する

①左上 (300 万円未満) のセルをアクティブにしてから、Shift キーを押しながらカーソル

移動キーを使って、表全体をアクティブにする（地が青く表示される）。

②この状態で、グラフウィザードをクリックする（グラフウィザードが表示されていない場合には、「挿入」→「グラフ」）。

図1.2 エクセル上で、グラフウィザードを起動する瞬間



③縦棒→左上のパターンを選択して、「次へ」をクリックして、必要な事項を指定していくと、とりあえず棒グラフができあがる。（いきなり「完了」をクリックしてもよい）。

④グラフを加工する。

1) グラフの余白をクリックして、さらに右クリックして、グラフのオプションを選択すると、グラフタイトル、凡例の有無、値ラベルの有無などを指定することができる。

2) 数値軸（縦軸）をクリックして、さらに右クリックすると、軸の書式設定（目盛り、フォントなど）を変更することができる。

3) 項目軸（横軸）をクリックして、さらに右クリックすると、軸の書式設定を変更することができる。

4) 凡例マーカー（凡例の色の部分）をクリックすると、凡例の色やパターンを変更することができる。

5) プロットエリア（グラフ枠のなかの空白部分）をクリックして、さらに右クリックすると、プロットエリアの書式設定（棒の色など）を変更することができる。

【課題 2】上記の例にならって、仕事による所得(仕事収入)の分布をグラフにしてみましよう。とくに、グラフの加工をいろいろ試みてみましょう。

2. クロス表

同様に、クロス表についても、次のような表に作り直すことが望ましい。

表 2.1 は比較的丁寧なクロス表を作成した例。各セルのパーセントと度数を示してある。

表2.1 学歴別の仕事による個人所得の分布(有職者のみ)

	300万円未満	300～700万円	700～1100万円	1100万円以上	計
中学卒	62.3 38	34.4 21	3.3 2	0.0 0	100.0 61
高校卒	53.5 151	30.1 85	14.9 42	1.4 4	100.0 282
短大・高専卒	55.1 70	37.8 48	7.1 9	0.0 0	100.0 127
大学卒	28.9 87	34.2 103	23.9 72	13.0 39	100.0 301
計	44.9 346	33.3 257	16.2 125	5.6 43	100.0 771

$$\chi^2 = 103.41, p < .001, V = .211$$

表 2.2 は、必要最低限の情報に限定した例。クロス表はスペースをとるので、このように簡略化して示されることが多い。

表2.2 学歴別の仕事による個人所得の分布(有職者のみ)

	300万円未満	300～700万円	700～1100万円	1100万円以上	計	N
中学卒	62.3	34.4	3.3	0.0	100.0	61
高校卒	53.5	30.1	14.9	1.4	100.0	282
短大・高専卒	55.1	37.8	7.1	0.0	100.0	127
大学卒	28.9	34.2	23.9	13.0	100.0	301
計	44.9	33.3	16.2	5.6	100.0	771

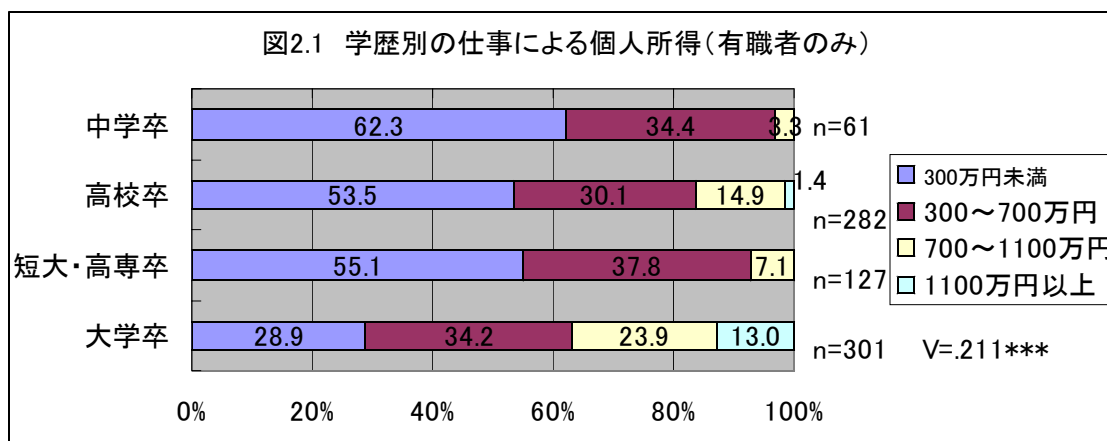
$$V = .211^{***}$$

【解説】各セルの度数は省略し、行の合計のみを表示。行 100%になるようにパーセントを計算していることは、表を見れば分かるから、セルの度数を知りたい読者は、Nにパーセントを掛ければ、度数が計算できる。

χ^2 自乗検定結果と Cramer の V 係数の表示は、かなり簡略化されている (χ^2 二乗値そのものは表示されていない)。この場合、本文のどこかに(表中のVはクラメールのV係数、 χ^2 自乗検定による有意確率は、*** $p < .001$, ** $p < .01$, * $p < .05$ 、以下同じ)と書き込んでおくか、表にそのような注をつけ、同じタイプの表が出てきたときに(表 2.2 の注を参照)などと付け加えておくことよい。有意でない場合はn.s.と表示する。n.s.とは non significant (非有意)の略号である。

クロス表のグラフ表示

一般の読者にわかりやすいようにクロス表をグラフ表示するとすれば、図 2.1 のようなグラフをエクセルで作成することができる（これは、表 2.2 と情報量が同じになるように、かなり手間をかけて作成したものである）。各自、工夫してみよ。



【課題 3】 SPSS の出力結果をエクセルに貼り付けて加工し、図 2.1 のようなグラフをつくってみましょう。

3. 相関係数

相関係数を単独で使用する場合には、いちいち表にせずに直接本文中に、相関係数 $r = .233$ ($p < .01$) とか、相関係数 $r = .233^{**}$ といった表記で済ませるのが一般的。しかし、相関係数をいくつも参照するときは、次のような表を作るとわかりやすい。

表3.1 所得と教育年数、年齢、性別(ダミー)の積率相関係数

	所得	(N)
教育年数	.291***	(771)
年齢	.143***	(781)
性別(男性 = 0)	-.488***	(782)

*** $p < .001$

4. 重回帰分析

授業では、重回帰方程式の意味を理解するために、わざわざ係数を入れた式をつくったが、専門的な論文では、たんに表で示すことが多い。次の例は、複数の重回帰式の解をひとつの表にまとめて、考察しやすくした例である。

表4.1 仕事による個人所得を予測する回帰式における偏回帰係数、標準誤差、標準化された偏回帰係数、決定係数および重相関係数

	モデル 1 (N=761)			モデル 2 (N=761)		
	B	S.E	β	B	S.E	β
教育年数	59.684***	5.673	.371	48.132***	5.120	.299
年齢	6.803***	0.932	.257	5.053***	0.839	.191
性別	-	-	-	-304.689***	21.649	-.429
定数	-703.284***	100.188		-312.312**	93.478	
R ²			.146***			.323***
R			.382***			.568***

*** $p < .001$, ** $p < .01$

この表に対応する、本文の記述は、たとえばつぎのようになる。

「個人所得を教育年数と年齢に回帰させたモデル 1 では、教育年数と年齢はともに有意で、偏回帰係数 B はそれぞれ 59.684***、6.803***、標準化された偏回帰係数 β はそれぞれ .371、.257 であった。モデル 1 は、個人所得の分散の 14.6 % を説明している。また、モデル 1 に性別を加えたモデル 2 では、教育年数、年齢、性別がいずれも有意であり、偏回帰係数 B は、順に 48.132***、5.053***、-304.689***、標準化された偏回帰係数 β は、順に .299、.191、-.429 であった。モデル 2 は、個人所得の分散の 32.3 % を説明しており、モデル 1 に比べて、説明力が 18 % も上昇した。…」

5. 平均値の比較

平均値の比較についても、複数の表をつぎのように簡略化して 1 枚にまとめることができる。(この表は、SPSS の出力結果をエクセルにコピーし、エクセル上で表を作り直して、テキストのみを一太郎にコピーして、整理したものである)

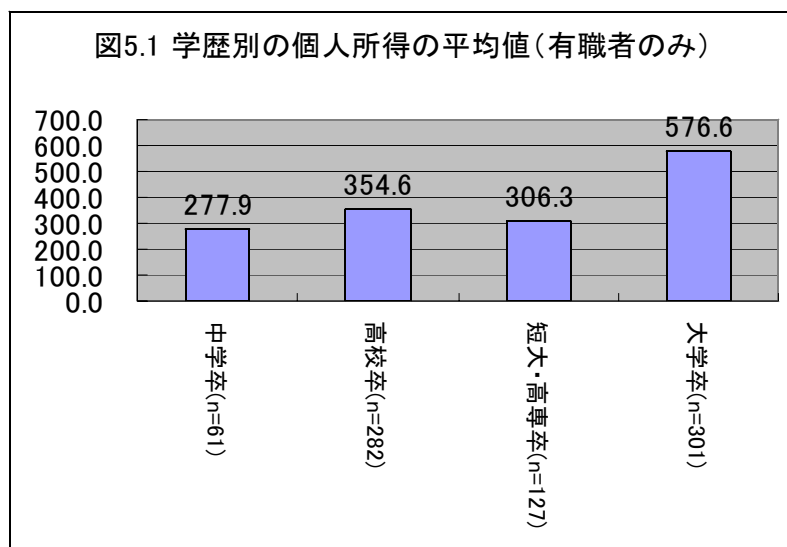
表5.1 学歴別、年齢10歳階級別、男女別の個人所得の平均値(万円)

	平均値	N	相関比
学歴			
中学卒	277.9	61	
高校卒	354.6	282	
短大・高専卒	306.3	127	
大学卒	576.6	301	
	427.2	771	.343***
年齢			
20-29	250.0	101	
30-39	389.7	165	
40-49	463.2	155	
50-59	560.2	192	
60-69	387.8	168	
	427.7	781	.270***
性別			
男性	600.9	392	
女性	252.7	390	
	427.2	782	.488***

*** $p < .001$

平均値の比較のグラフ表示

エクセルを使って、たとえばつぎのようなグラフをつくることできる。それほど難しくくない。



【課題4】 図5.1のようなグラフをつくってみましょう。