

第1回 質的変数と量的変数

1. 変数と値

「変数」は、複数の値を取りうるもので、調査票のひとつひとつの問いに対応している。

「値」は、ひとつひとつの回答に対応している。

性別という変数は、男、女の2つの値を取る。

所得という変数は、〇〇円という（原理的には）無数の値を取る。

2. 変数の型

変数は、その値の性質によって、大きく量的変数と質的変数に区別される。

量的変数とは、値が連続的に変化するもので、連続変量ともいう。

このうち原点（0）と単位（1）に意味があるものを絶対尺度、

原点（0）には意味があるが、単位（1）には意味がない（任意である）ものを比率尺度、

原点（0）も単位（1）も任意であるものを間隔尺度という。

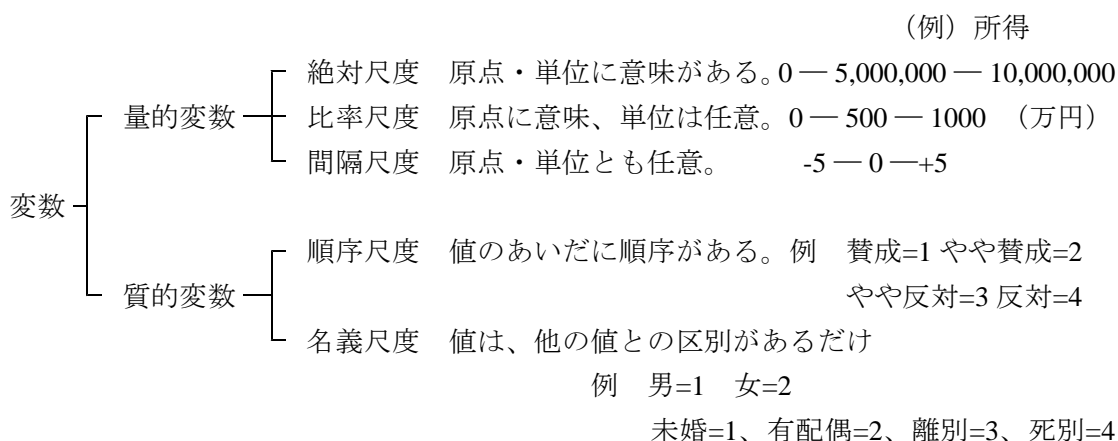
質的変数とは、値と値とのあいだの距離が任意のもので、離散変量ともいう。

値のあいだに順序のあるものは順序尺度*、

値は、たんに他の値と区別する記号に過ぎないものを名義尺度**という。

*順序尺度を質的変数として扱っていない教科書もある。

**カテゴリー変数、カテゴリカル変数ともいう。順序尺度を質的変数に分類しない考え方では、質的変数はカテゴリー変数と同義になる。



ある変数が、量的変数か質的変数かは、測定法（質問の仕方）による。

一般に、量的変数を質的変数に変換することは、可能だが、逆は不可能。

例：年間所得を円単位で訊けば、絶対尺度。これを100万円単位に直せば比率尺度。

大きい順に3分割すれば、順序尺度。これを高所得層、中間所得層、低所得層という名義尺度と見なすことも可。

3. 変数の型と分析法

量的変数と質的変数では、分析の仕方が異なる。

量的変数——平均、標準偏差、相関係数などを基礎とした統計的分析手法が発達している。

質的変数——度数分布表、クロス集計表などの初歩的な分析法のほか、近年、新たな分析法（ロジット・モデルなど）が開発されてきた。

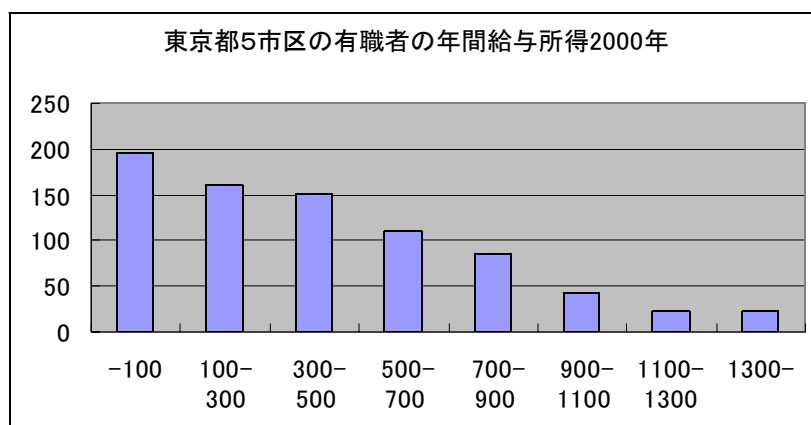
社会学で扱うデータは、質的変数が多い。この授業では、質的変数の分析法から入る。

4. 度数分布表

度数分布表は、どういう回答か何名あったのか、値ごとに個体数を示す最も基礎的な集計表。単純集計表ともいう。

度数分布表の例

q52 本人年収						
			度数	パーセント	有効パーセント	累積パーセント
有効	1	-100	196	19.9	24.8	24.8
	2	100-300	160	16.2	20.2	45.0
	3	300-500	151	15.3	19.1	64.1
	4	500-700	110	11.1	13.9	78.0
	5	700-900	86	8.7	10.9	88.9
	6	900-1100	43	4.4	5.4	94.3
	7	1100-1300	22	2.2	2.8	97.1
	8	1300-	23	2.3	2.9	100.0
	合計		791	80.1	100.0	
欠損値	9		196	19.9		
	合計		987	100.0		



5. 比率の推定

サンプリング調査の場合、度数分布表から母比率を推定することができる。

【母比率の推定の定理（危険率 5 % の場合）】

母比率 P は、

$$p - 1.96\sqrt{\frac{p(1-p)}{n}} < P < p + 1.96\sqrt{\frac{p(1-p)}{n}}$$

である。

ただし、 p は標本における比率、 n は標本数。（標本は無作為抽出標本でなくてはならない）

推定誤差の早見表

サンプル数・比率と推定誤差

n/p	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
100	0.043	0.059	0.078	0.090	0.096	0.098	0.096	0.090	0.078	0.059	0.043
200	0.030	0.042	0.055	0.064	0.068	0.069	0.068	0.064	0.055	0.042	0.030
300	0.025	0.034	0.045	0.052	0.055	0.057	0.055	0.052	0.045	0.034	0.025
400	0.021	0.029	0.039	0.045	0.048	0.049	0.048	0.045	0.039	0.029	0.021
500	0.019	0.026	0.035	0.040	0.043	0.044	0.043	0.040	0.035	0.026	0.019
600	0.017	0.024	0.032	0.037	0.039	0.040	0.039	0.037	0.032	0.024	0.017
700	0.016	0.022	0.030	0.034	0.036	0.037	0.036	0.034	0.030	0.022	0.016
800	0.015	0.021	0.028	0.032	0.034	0.035	0.034	0.032	0.028	0.021	0.015
900	0.014	0.020	0.026	0.030	0.032	0.033	0.032	0.030	0.026	0.020	0.014
1000	0.014	0.019	0.025	0.028	0.030	0.031	0.030	0.028	0.025	0.019	0.014
1500	0.011	0.015	0.020	0.023	0.025	0.025	0.025	0.023	0.020	0.015	0.011
2000	0.010	0.013	0.018	0.020	0.021	0.022	0.021	0.020	0.018	0.013	0.010
2500	0.009	0.012	0.016	0.018	0.019	0.020	0.019	0.018	0.016	0.012	0.009
3000	0.008	0.011	0.014	0.016	0.018	0.018	0.018	0.016	0.014	0.011	0.008
5000	0.006	0.008	0.011	0.013	0.014	0.014	0.014	0.013	0.011	0.008	0.006
10000	0.004	0.006	0.008	0.009	0.010	0.010	0.010	0.009	0.008	0.006	0.004

標本数 1000 で、 $p=0.2$ だと、危険率 5 % で、 ± 2.5 % の誤差がある。

6.SPSS による度数分布表の作成

コマンドは、

freq 変数名.