

コルモゴロフ=スミルノフ検定量の計算方法について

平成 22 年国勢調査 神奈川県 年齢 5 歳階級別人口（男女計）

	国勢調査 調査結果	匿名 データ	累積相対比率 (国勢調査)	累積相対比率 (匿名データ)	差率絶対値
0～4 歳	386,632	3,730	0.042978	0.042444	0.000534
5～9 歳	394,466	3,738	0.086826	0.084979	0.001848
10～14 歳	406,645	3,818	0.132029	0.128424	0.003605
15～19 歳	421,017	3,903	0.178829	0.172836	0.005993
20～24 歳	497,253	4,730	0.234103	0.226659	0.007444
25～29 歳	560,556	5,515	0.296414	0.289414	0.007000
30～34 歳	649,305	6,444	0.368590	0.362741	0.005850
35～39 歳	781,964	7,693	0.455513	0.450279	0.005233
40～44 歳	725,746	7,175	0.536186	0.531924	0.004262
45～49 歳	615,874	5,907	0.604646	0.599140	0.005506
50～54 歳	513,356	4,928	0.661710	0.655216	0.006495
55～59 歳	555,978	5,357	0.723512	0.716173	0.007339
60～64 歳	667,808	6,577	0.797745	0.791013	0.006733
65～69 歳	572,060	5,770	0.861335	0.856670	0.004665
70～74 歳	458,921	4,673	0.912348	0.909844	0.002504
75～79 歳	354,820	3,556	0.951790	0.950308	0.001482
80～84 歳	235,284	2,370	0.977944	0.977276	0.000668
85～89 歳	126,188	1,269	0.991971	0.991716	0.000255
90 歳以上	72,230	728	1.000000	1.000000	0.000000
合計	8,996,103	87,881			

コルモゴロフ = スミルノフ検定量 (KS 統計量) の計算方法は以下の通りである。事例として「平成 22 年国勢調査 神奈川県 年齢 5 歳階級別人口 (男女計)」を用いる。

なお、年齢不詳については累積がないため、計算から除外している。

KS 統計量の定義式は以下の通り。

$$\text{KS 統計量} = D \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

n_1 : 第 1 組の標本数 (国勢調査調査結果の総数)

n_2 : 第 2 組の標本数 (匿名データの標本数)

D : 累積相対比率の差の絶対値の中で最大の値

標本数の n_1 : 第 1 組の標本数 (国勢調査調査結果の総数) n_2 : 第 2 組の標本数 (匿名データの標本数) の取得

事例の表より $n_1 = 8,996,103$ 、 $n_2 = 87,881$ を得る。

国勢調査調査結果、匿名データのそれぞれの各階級において、累積相対度数を計算する。

で求めた累積相対度数について、各階級の差を求めその絶対値を計算する。

で求めた各階級の差の絶対値の中での最大値を取得する。この事例では 20 ~ 24 歳階級の 0.007444 が最大値であるから、 $D = 0.007444$ となる。

と で得た数値を KS 統計量の定義式に代入し、KS 統計量を計算する。

$$\text{KS 統計量} = 0.007444 \sqrt{\frac{8996103 \cdot 87881}{8996103 + 87881}} \approx 2.196$$

で得た KS 統計量について、以下の臨界値の表に従って評価する。この事例では $2.196 > 1.63$ のため、1%有意水準で有意差ありと判断する。

KS 統計量の臨界値 (両側検定) ()

有意水準	10%	5%	1%
臨界値	1.22	1.36	1.63

() 武藤眞介 (1995) 『統計解析ハンドブック』朝倉書店 P450 より引用。