

共分散分析の説明・補足

データの説明

`http://phi.med.gunma-u.ac.jp/grad/sample3.dat` は、都道府県別のタブ区切りテキストデータファイルである。変数としては、都道府県名 (PREF)、日本の東西 (REGION)、1990 年の 100 世帯あたり乗用車台数 (CAR1990)、1989 年の人口 10 万人当たり交通事故死者数 (TA1989)、1985 年の国勢調査による人口集中地区居住割合 (DIDP1985) が含まれている (REGION の 1 は東日本、2 は西日本を意味する)。

解析プログラム

通常は、共分散分析などをする前に、個別の変数の記述統計や図示をしてデータの性状をつかんでおくことが必須である。ここでは共分散分析に直接関わる部分のみ解説する。

実は、東日本と西日本では、世帯当たりの乗用車所有台数が有意に異なり、東日本の方が多 (Rcmdr を使って t 検定で容易に確かめることができる)。しかし、乗用車所有台数は、人口が集中して住んでいるところよりも、散らばって住んでいるところの方が多 (期待されるので、その影響を調整しても東日本の方が多) と言えるのか検討することが、共分散分析の目的である。

東西日本別々の回帰分析

まず、東西日本別々に、人口集中地区居住割合のばらつきによって乗用車所有台数が説明されるかをみるため、単回帰分析を行う。

```
plot(CAR1990~DIDP1985,pch=as.integer(REGION),xlab="人口集中地区居住割合 (%)",ylab="車所有台数 (/100 世帯)")
title("東日本 ( ), 西日本 ( ) 別, 人口集中地区居住割合\ n と車所有台数の関係")
abline(rE <- lm(CAR1990[REGION=="East"]~DIDP1985[REGION=="East"]),col="red",lty=1)
abline(rW <- lm(CAR1990[REGION=="West"]~DIDP1985[REGION=="West"]),col="blue",lty=2)
summary(rE)
summary(rW)
```

車所有台数の人口集中地区居住割合への回帰は、東西どちらでも有意である。したがって、その影響を調整することに意味はあると思われる。そこで、次に、傾きに差があるかを解析する。資料にも示したとおり、~ の右辺の独立変数群を * でつなぐと、その両側の変数単独と、その両者の交互作用のすべてを意味する。

```
rA <- lm(CAR1990 ~ REGION*DIDP1985)
summary(rA)
```

結果をみると、交互作用効果は有意でない (2 本の回帰直線の傾きに有意差はない) ことがわかる。そこで今度は、人口集中地区居住割合で調整した乗用車所有台数の修正平均に差があるかどうかをみるため、交互作用項を除いて回帰を行う。

```
rB <- lm(CAR1990 ~ REGION+DIDP1985)
summary(rB)
```

この結果、REGION の効果は有意なので、人口集中地区居住割合で調整しても、東日本では西日本よりも一世帯あたり乗用車所有台数が多い傾向があることがわかる。