

Notat 1:

Lineære modeller for log(markedssandele).

Formålet med dette notat er at uddybe fortolkningen af visse af de lineære modeller vi skal bruge. Blandt andet kræver det lidt eftertanke at forstå, hvorfor det i visse modeller er fuldstændigt ligemeget om vi som responser tager logaritmerne til de absolutte salg Q_{bt} eller logaritmerne til markedssandelene S_{bt} . Vi bruger samme notation som i artiklen "Regressionsmodeller for Markedssandele".

Vi ser af pædagogiske grunde i øjeblikket kun på to kaffemærker, det kunne f.eks. være Gevalia og Merrild. Vi lader $y_{bt} = \log(Q_{bt})$ betegne logaritmen til det samlede salg af mærke b i uge t , og betragter regressionsmodellen

$$(1) \quad y_{bt} = \delta_t + \alpha_b + \beta_b l_{bt} + \sigma u_{bt}$$

(underforstået: u 'erne er uafhængige, normeret normalfordelte) hvor l_{bt} er logaritmen til uge t 's gennemsnitspris for mærke b . Præcis hvilke kovariater vi har med i modellen er sådan set ligegyldigt, det væsentlige i det følgende er at den additive ugeeffekt δ_t er til stede.

I stedet for denne model for 244 y 'er kunne vi lige så godt opskrive modellen for de 122 summer $s_t = y_{1t} + y_{2t}$ og de 122 differenser $d_t = y_{1t} - y_{2t}$. Summer og differenser er åbenbart stokastisk uafhængige (fordi de er ukorrelerede), og de to modeller bliver

$$(+) \quad s_t = 2\delta_t + (\alpha_1 + \alpha_2) + \beta_1 l_{1t} + \beta_2 l_{2t} + \sigma(u_{1t} + u_{2t})$$

og

$$(-) \quad d_t = (\alpha_1 - \alpha_2) + \beta_1 l_{1t} - \beta_2 l_{2t} + \sigma(u_{1t} - u_{2t})$$

Den første model (+) er imidlertid triviel, i den forstand at middelværdiunderrummet jo (på grund af ugeeffekten) udgør hele observationsrummet. Ganske vist kan man, ved at udnytte at variansen er den samme i de to modeller, benytte modellen for summerne til at estimere ugeeffekten δ_t . Men denne ugeeffekt er vi typisk slet ikke interesseret i. Al information vedrørende de relevante parametre (α 'er og β 'er) må således uddrages af den anden model (-). Denne model for $d_t = \log(Q_{1t}/Q_{2t})$ er en almindelig regressionsmodel med konstantled $\alpha_1 - \alpha_2$ og to forklarende variable l_{1t} og l_{2t} , hvis koefficienter har fået de lidt underlige betegnelser β_1 og $-\beta_2$.

Beregningsmæssigt betyder dette, at analysen af den oprindelige model kan erstattes med analysen af differenserne. Men det betyder *også*, at analysen af den oprindelige model er helt upåvirket af, at man eventuelt erstatter responserne y_{bt} med nogle andre responser af formen $y'_{bt} = y_{bt} + c_t$ for vilkårlige uge-specifikke tal $c_1, \dots, c_{122}$. For eksempel kunne man i stedet for de logaritmerede salgstal $\log(Q_{bt})$ analysere de logaritmerede markedsandele

$$y'_{bt} = \log S_{bt} = \log Q_{bt} - \log(Q_{1t} + Q_{2t}).$$

Dette ville føre til samme estimater af de relevante parametre. Den tilføjede “tidstrend” $-\log(Q_{1t} + Q_{2t})$ er uden betydning, fordi den bliver absorberet af leddet δ_t . Dette kan man indse mere direkte rent beregningsmæssigt.

Umiddelbart lyder det jo meget forbudt at opstille en lineær normalfordelingsmodel for et sæt af reponser, der opfylder en relation som $\exp(y_{1t}) + \exp(y_{2t}) = 1$. Det kan i hvert fald ikke gælde, hvis y -erne også skal være uafhængige og normalfordelte. Men det behøver altså heller ikke at gælde, når bare modellen fortolkes korrekt. Generelt kan man sige, at kravet om uafhængige normalfordelte observationer med samme varians kun gælder “på nær addition af en vektor fra modellens middelværdiunderrum”.

Vi har hermed indset, at modellen (1) giver mening, uanset om responsvektorens værdier er logaritmer til salg eller logaritmer til markedsandele. At modellen så også giver (samme) mening når datasættet omfatter tre eller flere kaffemærker, kræver sådan set ingen yderligere argumentation.

Regneeksempel. Følgende ISU-program udfører de to ovennævnte analyser for mærkerne Gevalia og Merrild. Bemærk hvordan parameterestimaterne i de to modeller og deres estimerede standardafvigelser hører sammen. Bemærk også at variansestimatet i de to modeller afviger med faktoren 2, svarende til at $\text{var}(\sigma(u_{1t} - u_{2t})) = 2\sigma^2$.

Program:

```
delete
get data
x1=ln(prisgev)
x2=ln(prismer)
y=ln(salggev/salgmer)
fitlin y=1+x1+x2
listp

var xx yy 244
transf salggev 1 122 yy 1 122
transf salgmer 1 122 yy 123 244
```

```

yy=ln(yy)
transf x1 1 122 xx 1 122
transf x2 1 122 xx 123 244
fac b 244 2
gen b 122
fac t 244 122
gen t 1
fitlin yy=t+b+b*xx
listp

```

Klip fra (især) LISTPARAMETERS output:

```

FITLINEARNORMAL y=1+x1+x2
ANALYSIS OF VARIANCE TABLE

```

...

```

122 observations, 3 parameters estimated.
Estimated variance and standard deviation for Y.
      0.133173      0.364928

```

LISTPARAMETERS

	Estimate	Std.dev.	T	P
CONSTANT	-0.09	1.342	-0.065	0.948122
X1	-5.026	0.5723	-8.782	0.000000
X2	4.971	0.5991	8.298	0.000000

```

FITLINEARNORMAL yy=t+b+b*xx
ANALYSIS OF VARIANCE TABLE

```

...

```

244 observations, 125 parameters estimated.
Estimated variance and standard deviation for YY.
      0.0665863      0.258043

```

LISTPARAMETERS

	Estimate	Std.dev.	T	P
T[1]	28.12	1.968	14.287	0.000000
T[2]	28.37	1.953	14.521	0.000000
...				
T[122]	28.89	2.008	14.383	0.000000
B[1]	-0.09	1.342	-0.065	0.948123
B[2]	set to zero			
B[1]*XX	-5.026	0.5723	-8.782	0.000000
B[2]*XX	-4.971	0.5991	-8.298	0.000000