

Analysis of crop data

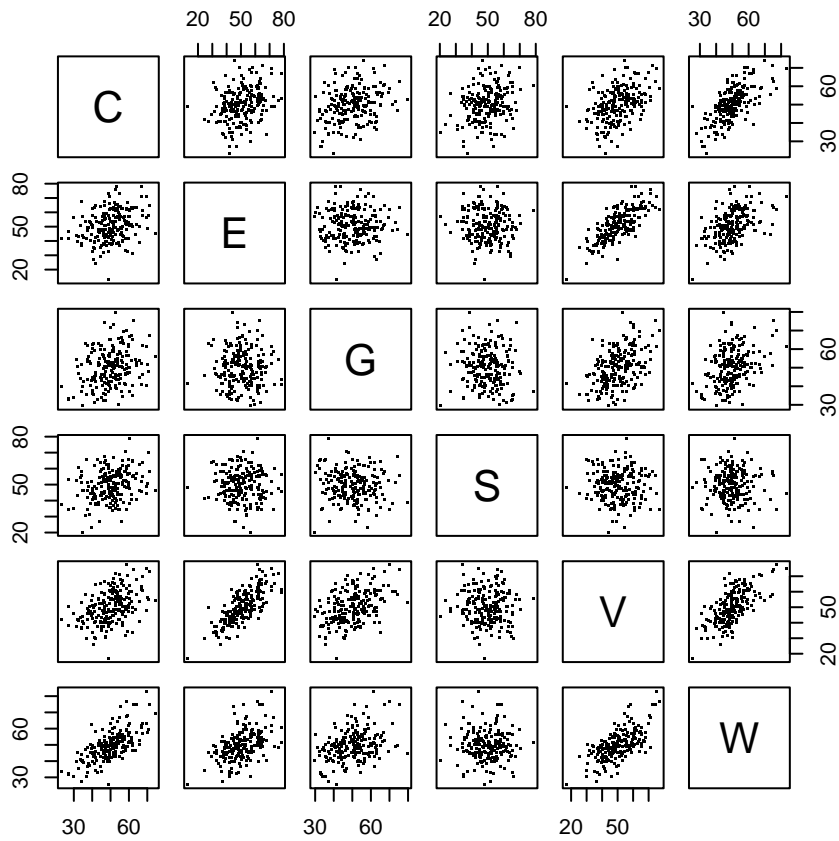
G. Marchetti

2024-03-21

```
source("~/Documents/R_packages/cat_regression_chains/RCG/funs_RCG.R")
library("bnlearn")
library("ggm")
library(mnormt)
```

Read the data

```
setwd("~/Documents/Talks/Firenze PhD 2024/")
crop <- read.table("cropdata.txt", header = TRUE)
crop <- data.frame(crop)
pairs(crop, pch = '.')
```



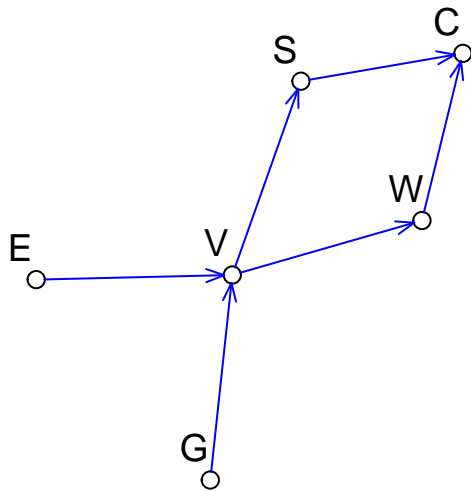
```
G <- DAG(C ~ S + W, S ~ V, W ~ V, V ~ G + E)
G
```

```

  C S W V G E
C 0 0 0 0 0 0
S 1 0 0 0 0 0
W 1 0 0 0 0 0
V 0 1 1 0 0 0
G 0 0 0 1 0 0
E 0 0 0 1 0 0

```

```
drawGraph(G)
```



Fit of the first equation

```
m_full1<- lm(C ~ S+W+V+G+E, data = crop)
round(summary(m_full1)$coef,3)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.001	4.349	0.000	1.000
S	0.276	0.047	5.838	0.000
W	0.706	0.067	10.606	0.000
V	-0.098	0.098	-0.997	0.320
G	0.078	0.062	1.275	0.204
E	0.043	0.079	0.552	0.581

```
m_red1 <- lm(C ~ S + W, data = crop)
anova(m_red1, m_full1, test = "F")
```

Analysis of Variance Table

Model 1: C ~ S + W

Model 2: C ~ S + W + V + G + E

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	197	7851.9				
2	194	7770.3	3	81.565	0.6788	0.566

LR test (asymptotic)

```
`LRtest` <- function(m_red, m_full) {  
  l0 <- logLik(m_red)[1]  
  lsat <- logLik(m_full)[1]  
  w <- 2 * (lsat - l0)  
  df <- m_red$df.residual - m_full$df.residual  
  p <- 1 - pchisq(w, df)  
  c(w = w, df = df, p = p)  
}
```

```
LRtest(m_red1, m_full1)
```

w	df	p
2.0884556	3.0000000	0.5542518

Second equation

```
m_full2 <- lm(S ~ W + V + G + E, data = crop)  
round(summary(m_full2)$coef, 3)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	54.374	5.313	10.235	0.000
W	0.020	0.101	0.199	0.842
V	0.014	0.148	0.097	0.923
G	-0.081	0.093	-0.874	0.383
E	-0.048	0.119	-0.402	0.688

```
m_red2 <- lm(S ~ 1, data = crop)  
anova(m_red2, m_full2, test = "F")
```

Analysis of Variance Table

Model 1: S ~ 1

Model 2: S ~ W + V + G + E

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	199	18031				
2	195	17915	4	116	0.3157	0.8673

LR test

```
LRtest(m_red2, m_full2)
```

	w	df	p
	1.2909126	4.0000000	0.8629154

Third equation

```
m_full3 <- lm(W ~ V + G + E, data = crop)
round(summary(m_full3)$coef, 3)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	18.702	3.518	5.317	0.000
V	0.593	0.096	6.168	0.000
G	0.062	0.066	0.947	0.345
E	-0.035	0.084	-0.420	0.675

```
m_red3 <- lm(W ~ V, data = crop)
anova(m_red3, m_full3, test = "F")
```

Analysis of Variance Table

Model 1: $W \sim V$

Model 2: $W \sim V + G + E$

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	198	9134.5				
2	196	9032.6	2	101.8	1.1045	0.3334

LR test

```
LRtest(m_red3, m_full3)
```

	w	df	p
	2.2415432	2.0000000	0.3260281

Fourth equation

```
m_full4 <- lm(V ~ G + E, data = crop)
round(summary(m_full4)$coef,3)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-10.455	2.500	-4.182	0
G	0.455	0.036	12.501	0
E	0.743	0.033	22.357	0

```
m_red4 <- m_full4
```

Last equation

```
m_full5 <- lm(G ~ E, data = crop)
round(summary(m_full5)$coef,3)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	49.692	3.365	14.766	0.000
E	0.009	0.065	0.137	0.891

```
m_red5 <- lm(G ~ 1, data = crop)
anova(m_red5, m_full5, test = "F")
```

Analysis of Variance Table

Model 1: G ~ 1

Model 2: G ~ E

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	199	19108				
2	198	19106	1	1.8208	0.0189	0.8909

LR test

```
LRtest(m_red5, m_full5)
```

```
           w           df           p  
0.01905948 1.00000000 0.89019611
```

Fit DAG

```
ord <- colnames(G)  
S <- cov(crop[,ord])  
lapply(fitDag(G, S, n= 200), function(x) round(x,3))
```

\$Shat

	C	S	W	V	G	E
C	84.539	22.814	56.638	42.640	17.156	33.639
S	22.814	90.605	-2.860	-4.848	-1.950	-3.824
W	56.638	-2.860	83.725	64.107	25.792	50.574
V	42.640	-4.848	64.107	108.656	43.716	85.719
G	17.156	-1.950	25.792	43.716	96.019	0.000
E	33.639	-3.824	50.574	85.719	0.000	115.416

\$Ahat

	C	S	W	V	G	E
C	1	-0.273	-0.686	0.000	0.000	0.000
S	0	1.000	0.000	0.045	0.000	0.000
W	0	0.000	1.000	-0.590	0.000	0.000
V	0	0.000	0.000	1.000	-0.455	-0.743
G	0	0.000	0.000	0.000	1.000	0.000
E	0	0.000	0.000	0.000	0.000	1.000

\$Dhat

	C	S	W	V	G	E
	39.457	90.388	45.902	25.090	96.019	115.416

\$dev

[1] 5.159

\$df

[1] 9