

Setup

Data creation and verification

Distribution in 2017

Spaghetti plot

log transformation

Mixed effect models

Traditional Japanese Diet Score and Breast Cancer 20201021

Chisato Abe

2020/10/21

Setup

Data creation and verification

```
# data input
d1 <- fread("variables.csv")

# Traditional Japanese Dietary score
quin_conv <- function(x) {
  y <-
    ifelse(is.na(x), NA,
           ifelse(x < quantile(x, probs=0.33, na.rm=T), -1,
                  ifelse(x < quantile(x, probs=0.66, na.rm=T), 0, 1)))
  return(y)
}

d1$Rice <- d1$Rice/d1$energy*1000
d1$Fish <- d1$Fish/d1$energy*1000
d1$Soyabeans <- d1$Soyabeans/d1$energy*1000
d1$Vegetables <- d1$Vegetables/d1$energy*1000
d1$Wheat <- d1$Wheat/d1$energy*1000
d1$Milk <- d1$Milk/d1$energy*1000
d1$Meat <- d1$Meat/d1$energy*1000
d1$Eggs <- d1$Eggs/d1$energy*1000
d1$Seaweeds <- d1$Seaweeds/d1$energy*1000

d1$Foods1 <- quin_conv(d1$Rice)
d1$Foods2 <- quin_conv(d1$Fish)
d1$Foods3 <- quin_conv(d1$Soyabeans)
d1$Foods4 <- quin_conv(d1$Vegetables)
d1$Foods5 <- quin_conv(d1$Wheat)*(-1)
d1$Foods6 <- quin_conv(d1$Milk)*(-1)
d1$Foods7 <- quin_conv(d1$Meat)*(-1)
d1$Foods8 <- quin_conv(d1$Eggs)
d1$Foods9 <- quin_conv(d1$Seaweeds)

d1$TJDS <- d1$Foods1 + d1$Foods2 + d1$Foods3 + d1$Foods4 +
  d1$Foods5 + d1$Foods6 + d1$Foods7 + d1$Foods8 + d1$Foods9

d1$population <- d1$population / 1000000
d1 %<>% filter(population > 1) # population more than 1 million
d1$GDP <- d1$GDP / 1000
d1$smoking <- d1$smoking*100
d1$energy <- d1$energy/1000
d1$act <- d1$act/1000
d1 %<>% drop_na() %>% as.data.frame(.)

d1$year <- d1$year - 1990
```

Distribution in 2017

```
d2017 <- d1 %>% filter(year==27)
```

```
# Distribution
```

```
distr <- function(x) {  
  ww <- NULL  
  for (i in (1:ncol(x))) {  
    if (is.numeric(x[[i]])) {  
      mean <- mean(x[[i]])  
      SD <- sd(x[[i]])  
      p05 <- as.numeric(quantile(x[[i]], probs=0.05, na.rm=TRUE))  
      p25 <- as.numeric(quantile(x[[i]], probs=0.25, na.rm=TRUE))  
      p50 <- as.numeric(quantile(x[[i]], probs=0.50, na.rm=TRUE))  
      p75 <- as.numeric(quantile(x[[i]], probs=0.75, na.rm=TRUE))  
      p95 <- as.numeric(quantile(x[[i]], probs=0.95, na.rm=TRUE))  
      w <- data.frame(var=colnames(x[i]), mean, SD, p05, p25, p50, p75, p95)  
      ww <- rbind(ww, w)  
    }  
  }  
  return(ww)  
}
```

```
d2017 %>%
```

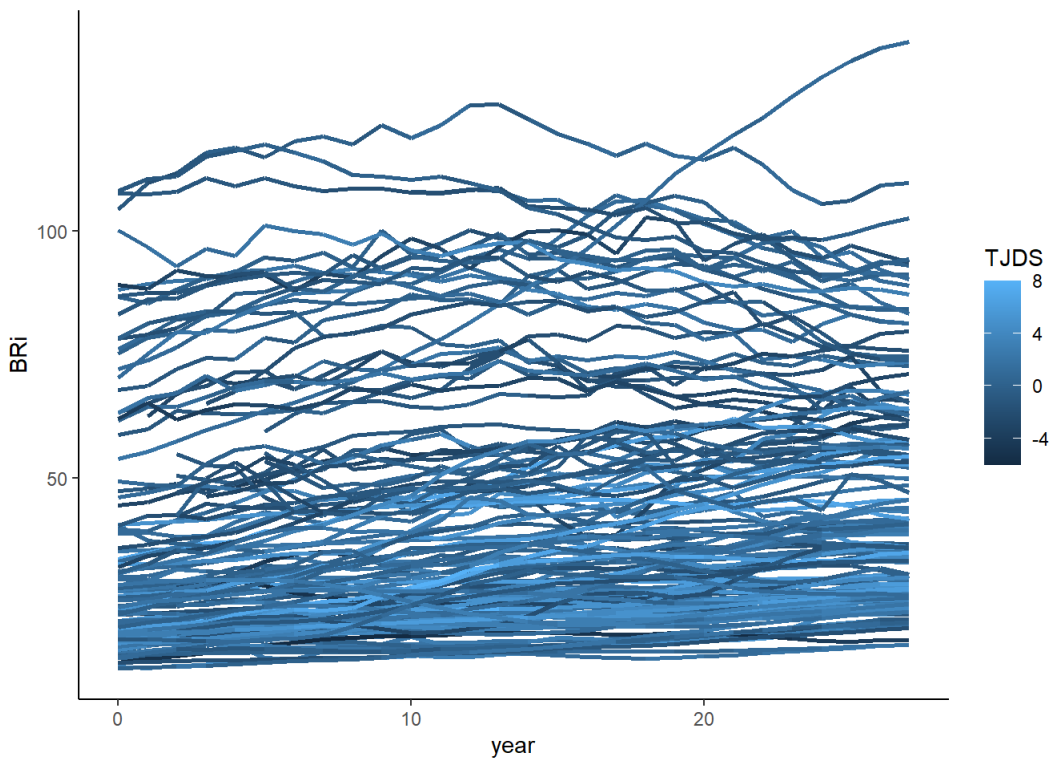
```
select(BRi, BRd, population, GDP, pct65, education, smoking, BMI, act, energy, TJDS) %>%  
distr()
```

	var	mean	SD	p05	p25	p50	p75	p95
1	BRi	46.056333	23.590426	20.4108658	26.833920	38.963149	60.713298	90.838981
2	BRd	16.150789	4.793916	9.4033728	12.618475	16.105394	18.422249	24.170672
3	population	52.479393	167.493474	2.0124775	5.572996	12.462791	38.199924	168.602235
4	GDP	12.831289	17.240510	0.5980085	1.534765	5.189490	15.649846	49.005983
5	pct65	9.124799	6.529541	2.4690682	3.417881	6.541039	14.847727	20.046775
6	education	9.163823	3.431655	3.4109915	6.421555	9.154750	12.478972	13.801763
7	smoking	15.089079	7.271199	4.9579264	9.308494	14.265417	20.798974	27.559401
8	BMI	25.405919	1.969958	22.1325157	23.701233	25.931917	26.839889	28.056667
9	act	5.733286	1.767993	3.2045614	4.267351	5.745459	7.507385	8.454662
10	energy	2.641899	0.407079	1.9784500	2.299000	2.655500	2.969500	3.244850
11	TJDS	1.601449	2.538865	-2.0000000	0.000000	1.500000	3.000000	6.000000

Spaghetti plot

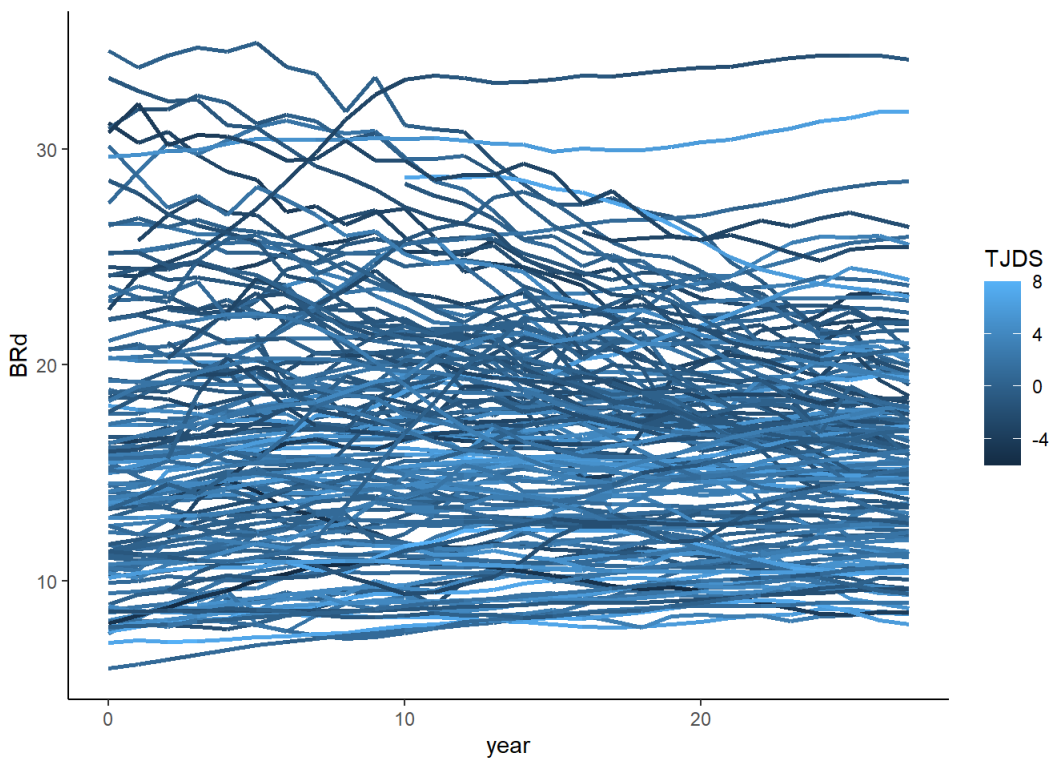
```
## Breast cancer incidence
```

```
p <-  
ggplot(data=d1, aes(x=year, y=BRi)) +  
  geom_line(aes(group=location, color=TJDS), size=1) +  
  labs(x="year", y="BRi")  
p
```



Breast cancer mortality

```
p <-
  ggplot(data=d1, aes(x=year, y=BRd)) +
  geom_line(aes(group=location, color=TJDs), size=1) +
  labs(x="year", y="BRd")
p
```



log transformation

```
d1log <- d1
```

```
d1log$GDP <- d1log$GDP * 1000  
d1log$smoking <- d1log$smoking/100  
d1log$energy <- d1log$energy*1000  
d1log$act <- d1log$act*1000
```

```
d1log$BRi <- log(d1log$BRi)*1000  
d1log$BRd <- log(d1log$BRd)*1000  
d1log$GDP <- log(d1log$GDP)  
d1log$pct65 <- log(d1log$pct65)  
d1log$education <- log(d1log$education)  
d1log$smoking <- log(d1log$smoking)  
d1log$BMI <- log(d1log$BMI)  
d1log$act <- log(d1log$act)  
d1log$energy <- log(d1log$energy)
```

Mixed effect models

```
library(nlme)
```

```
Attaching package: 'nlme'
```

```
The following object is masked from 'package:dplyr':
```

```
collapse
```

```
library(AICcmodavg)  
library(directlabels)
```

```
Attaching package: 'directlabels'
```

```
The following object is masked from 'package:nlme':
```

```
gapply
```

```
library(texreg)
```

```
Version: 1.37.5  
Date: 2020-06-17  
Author: Philip Leifeld (University of Essex)
```

```
Consider submitting praise using the praise or praise_interactive functions.  
Please cite the JSS article in your publications -- see citation("texreg").
```

```
Attaching package: 'texreg'
```

```
The following object is masked from 'package:magrittr':
```

```
extract
```

```
The following object is masked from 'package:tidyr':
```

```
extract
```

```
# Breast cancer incidence
```

```
g0 <- lme(fixed = BRi~year, data=d1log, random=~1+year|location, correlation=corCompSymm())  
summary(g0)
```

Linear mixed-effects model fit by REML

Data: d1log
AIC BIC logLik
41819.32 41862.86 -20902.66

Random effects:

Formula: ~1 + year | location
Structure: General positive-definite, Log-Cholesky parametrization
StdDev Corr
(Intercept) 633.97115 (Intr)
year 11.65929 -0.606
Residual 53.71001

Correlation Structure: Compound symmetry

Formula: ~1 | location
Parameter estimate(s):

Rho
0

Fixed effects: BRi ~ year

	Value	Std.Error	DF	t-value	p-value
(Intercept)	3428.786	53.8154	3575	63.71384	0
year	11.581	0.9967	3575	11.61983	0

Correlation:

(Intr)
year -0.605

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-5.0250136	-0.4586621	0.0209707	0.4758529	4.7192405

Number of Observations: 3715

Number of Groups: 139

```
g1 <- lme(BRi~TJDS*year, data=d1log, random=~1+year|location, correlation=corCompSymm())  
summary(g1)
```

Linear mixed-effects model fit by REML

Data: d1log
AIC BIC logLik
41799.23 41855.2 -20890.61

Random effects:

Formula: ~1 + year | location
Structure: General positive-definite, Log-Cholesky parametrization
StdDev Corr
(Intercept) 642.64275 (Intr)
year 12.21995 -0.636
Residual 53.47831

Correlation Structure: Compound symmetry

Formula: ~1 | location
Parameter estimate(s):

Rho
0

Fixed effects: BRi ~ TJDS * year

	Value	Std.Error	DF	t-value	p-value
(Intercept)	3418.574	54.60460	3573	62.60598	0e+00
TJDS	9.254	2.43326	3573	3.80299	1e-04
year	12.449	1.05866	3573	11.75925	0e+00
TJDS:year	-0.717	0.14569	3573	-4.91895	0e+00

Correlation:

(Intr) TJDS year
TJDS -0.044
year -0.632 0.117
TJDS:year 0.040 -0.833 -0.163

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-5.05416684	-0.46275719	0.01278031	0.48365530	4.70474921

Number of Observations: 3715

Number of Groups: 139

```
g2 <- lme(BRi~TJDS*year+GDP, data=d1log, random=~1+year|location, correlation=corCompSymm())
summary(g2)
```

Linear mixed-effects model fit by REML

Data: d1log

	AIC	BIC	logLik
	41796.12	41858.31	-20888.06

Random effects:

Formula: ~1 + year | location

Structure: General positive-definite, Log-Cholesky parametrization

	StdDev	Corr
(Intercept)	639.11090	(Intr)
year	12.21021	-0.638
Residual	53.49853	

Correlation Structure: Compound symmetry

Formula: ~1 | location

Parameter estimate(s):

Rho

0

Fixed effects: BRi ~ TJDS * year + GDP

	Value	Std.Error	DF	t-value	p-value
(Intercept)	3397.662	61.01242	3572	55.68805	0.0000
TJDS	9.278	2.43416	3572	3.81159	0.0001
year	12.277	1.08233	3572	11.34357	0.0000
GDP	2.925	3.88976	3572	0.75191	0.4522
TJDS:year	-0.716	0.14572	3572	-4.91356	0.0000

Correlation:

	(Intr)	TJDS	year	GDP
TJDS	-0.046			
year	-0.455	0.112		
GDP	-0.456	0.014	-0.211	
TJDS:year	0.032	-0.833	-0.162	0.009

Standardized Within-Group Residuals:

	Min	Q1	Med	Q3	Max
	-5.055033466	-0.465006997	0.009659856	0.484599467	4.726733147

Number of Observations: 3715

Number of Groups: 139

```
g3 <- lme(BRi~TJDS*year+GDP+pct65+education+smoking+act+BMI+energy,
data=d1log, random=~1+year|location, correlation=corCompSymm())
summary(g3)
```

Linear mixed-effects model fit by REML

Data: d1log

	AIC	BIC	logLik
	41296.85	41396.32	-20632.42

Random effects:

Formula: ~1 + year | location

Structure: General positive-definite, Log-Cholesky parametrization

	StdDev	Corr
(Intercept)	459.46995	(Intr)
year	10.69331	-0.733
Residual	51.30193	

Correlation Structure: Compound symmetry

Formula: ~1 | location

Parameter estimate(s):

Rho

0

Fixed effects: BRi ~ TJDS * year + GDP + pct65 + education + smoking + act + BMI + energy

	Value	Std. Error	DF	t-value	p-value
(Intercept)	2488.7330	925.9545	3566	2.687749	0.0072
TJDS	5.9060	2.3340	3566	2.530477	0.0114
year	-2.4455	1.2856	3566	-1.902178	0.0572
GDP	3.4139	3.9041	3566	0.874442	0.3819
pct65	134.0515	21.2661	3566	6.303543	0.0000
education	571.0119	49.5089	3566	11.533510	0.0000
smoking	-24.6132	19.0611	3566	-1.291281	0.1967
act	-640.0066	73.4079	3566	-8.718495	0.0000
BMI	1643.2621	181.2937	3566	9.064086	0.0000
energy	-5.4311	25.7217	3566	-0.211148	0.8328
TJDS:year	-0.4528	0.1380	3566	-3.281438	0.0010

Correlation:

	(Intr)	TJDS	year	GDP	pct65	eductn	smokng	act	BMI	energy
TJDS	-0.115									
year	0.258	0.107								
GDP	0.228	-0.025	-0.109							
pct65	0.000	-0.004	0.003	-0.017						
education	-0.082	-0.068	-0.565	0.034	-0.290					
smoking	0.053	0.005	0.222	-0.039	-0.072	-0.111				
act	-0.804	0.095	-0.157	-0.131	-0.075	0.166	0.010			
BMI	-0.683	0.076	-0.173	-0.210	0.077	-0.188	-0.012	0.177		
energy	-0.114	0.036	-0.014	-0.198	-0.048	0.008	-0.006	0.026	-0.180	
TJDS:year	0.079	-0.833	-0.169	0.035	-0.007	0.090	-0.001	-0.079	-0.045	-0.018

Standardized Within-Group Residuals:

	Min	Q1	Med	Q3	Max
	-5.58572151	-0.47723311	0.01552767	0.49981239	4.52093920

Number of Observations: 3715

Number of Groups: 139

```
screenreg(list(g0, g1, g2, g3), digits=3, single.row=T)
```

	Model 1	Model 2	Model 3	Model 4
(Intercept)	3428.786 (53.815) ***	3418.574 (54.605) ***	3397.662 (61.012) ***	2488.733 (925.955) **
year	11.581 (0.997) ***	12.449 (1.059) ***	12.277 (1.082) ***	-2.445 (1.286)
TJDS		9.254 (2.433) ***	9.278 (2.434) ***	5.906 (2.334) *
TJDS:year		-0.717 (0.146) ***	-0.716 (0.146) ***	-0.453 (0.138) **
GDP			2.925 (3.890)	3.414 (3.904)
pct65				134.051 (21.266) ***
education				571.012 (49.509) ***
smoking				-24.613 (19.061)
act				-640.007 (73.408) ***
BMI				1643.262 (181.294) ***
energy				-5.431 (25.722)
AIC	41819.325	41799.228	41796.123	41296.847
BIC	41862.862	41855.199	41858.311	41396.322
Log Likelihood	-20902.662	-20890.614	-20888.061	-20632.424
Num. obs.	3715	3715	3715	3715
Num. groups: location	139	139	139	139

*** p < 0.001; ** p < 0.01; * p < 0.05

Breast cancer mortality

```
g0 <- lme(fixed = BRd~year, data=d1log, random=~1+year|location, correlation=corCompSymm())
summary(g0)
```

Linear mixed-effects model fit by REML

Data: d1log

	AIC	BIC	logLik
	41192.45	41235.99	-20589.22

Random effects:

Formula: ~1 + year | location

Structure: General positive-definite, Log-Cholesky parametrization

	StdDev	Corr
(Intercept)	419.66875	(Intr)
year	11.33703	-0.715
Residual	50.16250	

Correlation Structure: Compound symmetry

Formula: ~1 | location

Parameter estimate(s):

Rho

0

Fixed effects: BRd ~ year

	Value	Std.Error	DF	t-value	p-value
(Intercept)	2773.401	35.65219	3575	77.79049	0.0000
year	-0.523	0.96858	3575	-0.53998	0.5892

Correlation:

	(Intr)
year	-0.714

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-5.6305341	-0.4320980	0.0142341	0.4496728	5.6047087

Number of Observations: 3715

Number of Groups: 139

```
g1 <- lme(BRd~TJDS*year, data=d1log, random=~1+year|location, correlation=corCompSymm())
summary(g1)
```


Linear mixed-effects model fit by REML

Data: d1log

	AIC	BIC	logLik
	41174.89	41230.86	-20578.45

Random effects:

Formula: ~1 + year | location

Structure: General positive-definite, Log-Cholesky parametrization

	StdDev	Corr
(Intercept)	426.99882	(Intr)
year	11.81936	-0.734
Residual	49.96197	

Correlation Structure: Compound symmetry

Formula: ~1 | location

Parameter estimate(s):

Rho

0

Fixed effects: BRd ~ TJDS * year

	Value	Std. Error	DF	t-value	p-value
(Intercept)	2763.7620	36.34481	3573	76.04283	0.0000
TJDS	8.9084	2.28338	3573	3.90143	0.0001
year	0.2358	1.02256	3573	0.23058	0.8177
TJDS:year	-0.6404	0.13650	3573	-4.69174	0.0000

Correlation:

	(Intr)	TJDS	year
TJDS	-0.062		
year	-0.730	0.114	
TJDS:year	0.057	-0.836	-0.158

Standardized Within-Group Residuals:

	Min	Q1	Med	Q3	Max
	-5.630420109	-0.437186670	0.005192734	0.457281541	5.599472815

Number of Observations: 3715

Number of Groups: 139

```
g2 <- lme(BRd~TJDS*year+GDP, data=d1log, random=~1+year|location, correlation=corCompSymm())
summary(g2)
```

Linear mixed-effects model fit by REML

Data: d1log

	AIC	BIC	logLik
	41144.1	41206.29	-20562.05

Random effects:

Formula: ~1 + year | location

Structure: General positive-definite, Log-Cholesky parametrization

	StdDev	Corr
(Intercept)	442.08330	(Intr)
year	11.87127	-0.749
Residual	49.73425	

Correlation Structure: Compound symmetry

Formula: ~1 | location

Parameter estimate(s):

Rho

0

Fixed effects: BRd ~ TJDS * year + GDP

	Value	Std.Error	DF	t-value	p-value
(Intercept)	2901.1961	45.45973	3572	63.81904	0.0000
TJDS	8.8041	2.27217	3572	3.87477	0.0001
year	1.3695	1.04806	3572	1.30671	0.1914
GDP	-19.2282	3.57035	3572	-5.38552	0.0000
TJDS:year	-0.6499	0.13565	3572	-4.79135	0.0000

Correlation:

	(Intr)	TJDS	year	GDP
TJDS	-0.057			
year	-0.491	0.108		
GDP	-0.561	0.012	-0.201	
TJDS:year	0.038	-0.836	-0.156	0.013

Standardized Within-Group Residuals:

	Min	Q1	Med	Q3	Max
	-5.44405564	-0.44014170	0.02026679	0.45952087	5.45660061

Number of Observations: 3715

Number of Groups: 139

```
g3 <- lme(BRd~TJDS*year+GDP+pct65+education+smoking+act+BMI+energy,  
data=d1log,random=~1+year|location,correlation=corCompSymm())  
summary(g3)
```

Linear mixed-effects model fit by REML

Data: d1log

	AIC	BIC	logLik
	40945.29	41044.77	-20456.65

Random effects:

Formula: ~1 + year | location

Structure: General positive-definite, Log-Cholesky parametrization

	StdDev	Corr
(Intercept)	403.59152	(Intr)
year	10.48917	-0.471
Residual	48.52981	

Correlation Structure: Compound symmetry

Formula: ~1 | location

Parameter estimate(s):

Rho

0

Fixed effects: BRd ~ TJDS * year + GDP + pct65 + education + smoking + act + BMI + energy

	Value	Std.Error	DF	t-value	p-value
(Intercept)	4714.121	938.8926	3566	5.020938	0.0000
TJDS	6.040	2.2377	3566	2.699027	0.0070
year	-6.161	1.2284	3566	-5.015366	0.0000
GDP	-18.157	3.7136	3566	-4.889304	0.0000
pct65	134.440	21.7490	3566	6.181408	0.0000
education	313.911	44.8063	3566	7.005944	0.0000
smoking	-16.151	18.6807	3566	-0.864601	0.3873
act	-447.254	77.4246	3566	-5.776641	0.0000
BMI	309.298	179.5986	3566	1.722162	0.0851
energy	33.255	24.4881	3566	1.357996	0.1746
TJDS:year	-0.455	0.1346	3566	-3.379342	0.0007

Correlation:

	(Intr)	TJDS	year	GDP	pct65	eductn	smokng	act	BMI	energy
TJDS	-0.154									
year	0.254	0.085								
GDP	0.228	-0.027	-0.113							
pct65	-0.034	0.018	-0.046	0.007						
education	-0.066	-0.048	-0.528	0.037	-0.262					
smoking	0.032	0.020	0.222	-0.040	-0.045	-0.117				
act	-0.812	0.127	-0.130	-0.142	-0.064	0.148	0.030			
BMI	-0.659	0.099	-0.194	-0.199	0.118	-0.191	-0.006	0.151		
energy	-0.110	0.031	-0.007	-0.200	-0.056	0.006	-0.006	0.042	-0.189	
TJDS:year	0.130	-0.837	-0.140	0.037	-0.036	0.064	-0.021	-0.120	-0.076	-0.011

Standardized Within-Group Residuals:

	Min	Q1	Med	Q3	Max
	-5.67149591	-0.44554476	0.02627328	0.47230319	5.37119658

Number of Observations: 3715

Number of Groups: 139

screenreg(list(g0, g1, g2, g3), digits=3, single.row=T)

	Model 1	Model 2	Model 3	Model 4
(Intercept)	2773.401 (35.652) ***	2763.762 (36.345) ***	2901.196 (45.460) ***	4714.121 (938.893) ***
year	-0.523 (0.969)	0.236 (1.023)	1.370 (1.048)	-6.161 (1.228) ***
TJDS		8.908 (2.283) ***	8.804 (2.272) ***	6.040 (2.238) **
TJDS:year		-0.640 (0.136) ***	-0.650 (0.136) ***	-0.455 (0.135) ***
GDP			-19.228 (3.570) ***	-18.157 (3.714) ***
pct65				134.440 (21.749) ***
education				313.911 (44.806) ***
smoking				-16.151 (18.681)
act				-447.254 (77.425) ***
BMI				309.298 (179.599)
energy				33.255 (24.488)
AIC	41192.449	41174.892	41144.102	40945.294
BIC	41235.986	41230.864	41206.290	41044.769
Log Likelihood	-20589.224	-20578.446	-20562.051	-20456.647
Num. obs.	3715	3715	3715	3715
Num. groups: location	139	139	139	139

*** p < 0.001; ** p < 0.01; * p < 0.05