

Lab 1

Daniel Dema

2025-05-25

```
{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)
```

Lab 1

```
library( spida2 )
library(nlme)
library(car)
library(lattice)
library(latticeExtra)
layer <- latticeExtra::layer
library( p3d )

?hsfull
dim( hsfull )
some( hsfull )    # a random selection of rows

hsup <- up(hsfull, ~ school)
some(hsup)
dim(hsup)
xqplot(hsup)
gicc(hsfull, ~ school )
gicc(hsfull, ~ school, method = 'raw' )    # 'real' intra-class
correlation (look it up)

Tab( hs, ~ school)          # school size: 'Tab' is a variant of 'tab'
with no marginal totals
tab( Tab( hs, ~ school ))    # schools by school size
xqplot( hsup )

hsup <- up (hsfull, ~ school)  # as above
dim(hsup)                    # one row per school

selected <- sample(1:nrow(hsup), round(nrow(hsup) / 2)) # row indices
# for half the schools
hsu1 <- hsup[selected, ]      # first half of schools
hsu2 <- hsup[ -selected, ]    # other half

hs1 <- merge( hsu1, hsfull )   # long data for 1st half of schools
hs2 <- merge( hsu2, hsfull )   # long data for other half
```

```

dim(hs1)
dim(hs2)

search()
rm( hs1 )

search()

dd <- hs1

dim( dd )
some(dd)

summary( dd )

sum(is.na(dd))

xqplot( dd )          # 1 dim summary
xqplot( dd ,ptype = 'n')      # 1 dim summary

sum(is.na(dd))

```

QUESTIONS 1

```

#
# Looking at univariate distributions for this data:
#
# Any problems with:
#     NAs?
#     Highly skewed distributions?
#     Possible univariate outliers?
#     Skewed factors: i.e. very rare category(ies)
#
# Any actions to take?

```

Question 1 Solution:

We see that the sum of NA is 0, so there are no NA problems.

Size is highly right skewed, and ses is highly left skewed.

There's a clear low outlier for ses. Moreover, we see small clusters of possible high outliers for DISCLIM and for size.

We see that minority is highly skewed towards no.

In terms of actions to take - I'm not certain but perhaps a log transformation for size and ses to account for the skews?

```
wald( fitc )
```

```
L <- list( 'Effect of ses' = rbind(  
  "Within-school" = c( 0,1,0,0,0),  
  "Contextual"    = c( 0,0,0,1,0),  
  "Compositional" = c( 0,1,0,1,0)))
```

```
wald( fitc )
```

```
L
```

```
wald ( fitc , L )
```

```
### QUESTIONS 5
```

```
-----
```

```
#  
# 1) What is the relationship among the 3 estimated values?  
#  
# 2) Why does the F test have 2 numDFs although there are  
#     three lines in the L matrix?  
#
```

Question 5 Solution:

- 1) The compositional estimate is the sum of the within-school and contextual estimates.
- 2) The matrix has 3 rows but the row space is 2-dimensional since $(0, 1, 0, 1, 0) = (0, 1, 0, 0, 0) + (0, 0, 0, 1, 0)$.