

Introduction to Multilevel Modeling

Andrew Zieffler

March 21, 2016

Load Libraries

```
library(AICcmodavg)
library(dplyr)
library(foreign)
library(ggplot2)
library(sm)
```

Load Data

When using the `read.spss()` function you need to give the pathname where your file is located. Use `file.choose()` to find the file and obtain the pathname.

```
nbaL1 = read.spss(file = "~/Documents/EPsy-8252/data/nbaLevel1.sav", to.data.frame = TRUE)
```

```
## Warning in read.spss(file = "~/Documents/EPsy-8252/data/nbaLevel1.sav", :
## ~/Documents/EPsy-8252/data/nbaLevel1.sav: Unrecognized record type 7,
## subtype 18 encountered in system file
```

```
## re-encoding from CP1252
```

```
head(nbaL1)
```

```
##   Team_ID Shots_on_five Life_Satisfaction
## 1      01             3             18.804
## 2      01             3             18.000
## 3      01             4             21.000
## 4      01             4             20.500
## 5      01             3             19.000
## 6      01             2             12.100
```

```
nbaL2 = read.spss(file = "~/Documents/EPsy-8252/data/nbaLevel2.sav", to.data.frame = TRUE)
```

```
## Warning in read.spss(file = "~/Documents/EPsy-8252/data/nbaLevel2.sav", :
## ~/Documents/EPsy-8252/data/nbaLevel2.sav: Unrecognized record type 7,
## subtype 18 encountered in system file
```

```
## re-encoding from CP1252
```

```
head(nbaL2)
```

```
## Team_ID Coach_Experience
## 1      01              2
## 2      02              3
## 3      03              2
## 4      04              2
## 5      05              1
## 6      06              2
```

Note you will also get a warning...this is OK. The data still reads in just fine. Here is what Stack Overflow has to say:

As far as warning message is concerned, It does not affect the data. The record type 7 is used to store features in newer SPSS software to make older SPSS software able to read new data. But does not affect data. I have used this numerous times and data is not lost.

Merge the Player- and Team-Level Data

To merge the two files, we will use the `left_join()` function from the **dplyr** package. **dplyr** includes six different join functions. You can read about several different join functions [here](#).

```
nba = left_join(nbaL1, nbaL2, by = "Team_ID")
head(nba)
```

```
## Team_ID Shots_on_five Life_Satisfaction Coach_Experience
## 1      01              3             18.804              2
## 2      01              3             18.000              2
## 3      01              4             21.000              2
## 4      01              4             20.500              2
## 5      01              3             19.000              2
## 6      01              2             12.100              2
```

Fit Linear Models

```
lm.0 = lm(Life_Satisfaction ~ 1, data = nba)
lm.1 = lm(Life_Satisfaction ~ 1 + Shots_on_five, data = nba)
```

Select “Best” Linear Model

```
myAIC = aictab(
  cand.set = list(lm.0, lm.1),
  modnames = c("LM0", "LM1")
)
myAIC
```

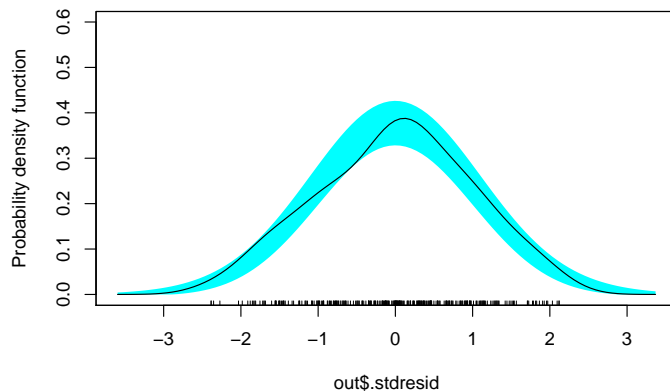
	Modnames	K	AICc	Delta_AICc	ModelLik	AICcWt	LL	Cum.Wt
2	LM1	3	1391.741	0.000	1	1	-692.8301	1
1	LM0	2	1865.773	474.032	0	0	-930.8665	1

Given the data and the two model candidates, the model that includes the predictor `Shots_on_five` should be adopted.

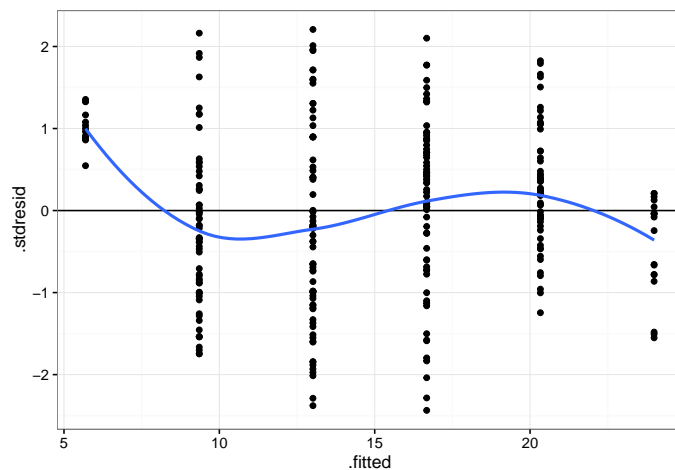
Examine Residuals

```
# Obtain the fortified data frame
out = fortify(lm.1)

# Normality
sm.density(out$.stdresid, model = "normal")
```



```
# All other assumptions
ggplot(data = out, aes(x = .fitted, y = .stdresid)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  theme_bw() +
  geom_smooth(se = FALSE)
```

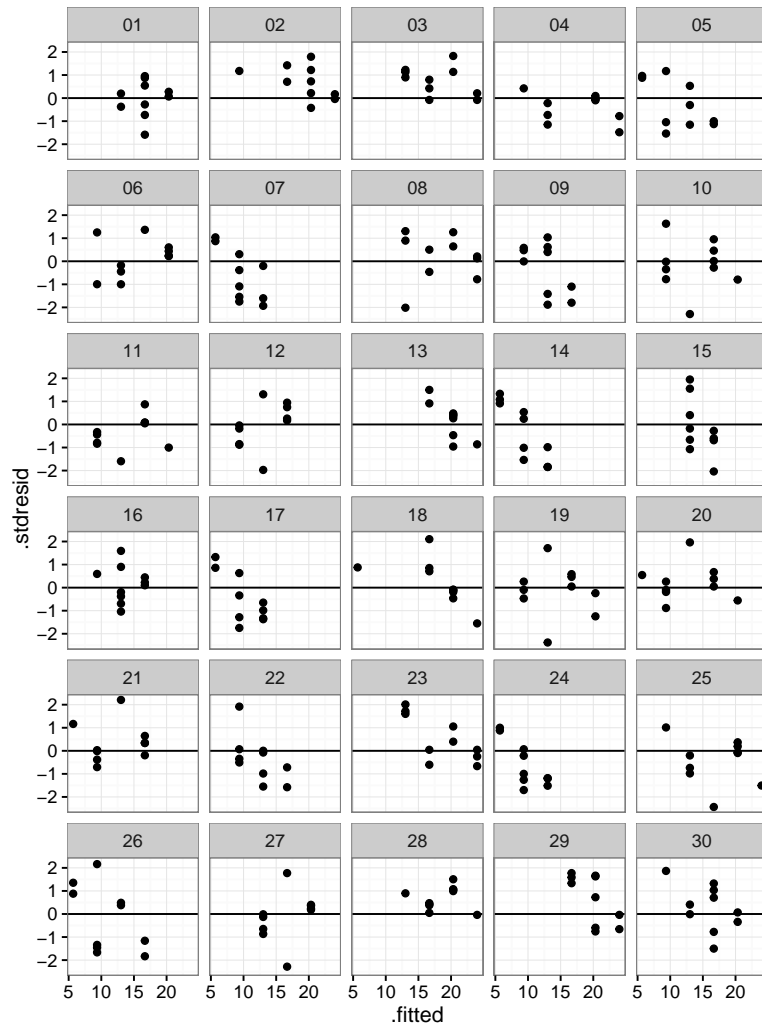


The assumptions of linearity, normality, and homoscedasticity seem reasonably satisfied. The assumption of independence is probably not tenable. The life satisfaction scores (and thus the residuals) are probably more correlated within teams than between teams—this is a violation of independence.

We can examine this by plotting the residuals separately for each team.

```
# Add Team_ID variable to fortified data
out$Team_ID = nba$Team_ID
#head(out)
```

```
### Show residuals by team
ggplot(data = out, aes(x = .fitted, y = .stdresid)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  theme_bw() +
  facet_wrap(~Team_ID, nrow = 6)
```



The residuals are systematically over or under 0 within teams (e.g., Team 11, Team 28). This is a sign of non-independence of the residuals.