# 3 Linear Models I

Note Title 9/7/2011

Today: Quiz
Homework
Multiple Comparisons
Extending R
- agricolae package
- RFS package

} on R script

* Linear models

# Multiple comparisons

In our ANOVA-type tests, we merely were able to conclude that (at least) one population had a different mean.

How boring! (so what?)

It is much more interesting to ask "which?"

Solution: Multiple t-tests

Problem: inflated Type I Error rates

Solution: Bonferroni

Problem: Too too conservative.

Solution: Use different tests

Options:
- Fisher's LSD test
- Tukey's HSD test
- the Student-Newman-Keuls test
- Kruskal's test
- Many others

Which to use?

If you used ANOVA, use HSD

If you used Kruskal-Wallis, use Kruskal

How? In R, library agricolae

then use HSD.test(mmt, grp, df1, df2)

or kruskal(mmt, grp)

Both will output labels that separate the groups into distinct classes

ex

A doves — lone group

B wrens — lone group

C geese } statistically similar

C ducks }

# Linear Models

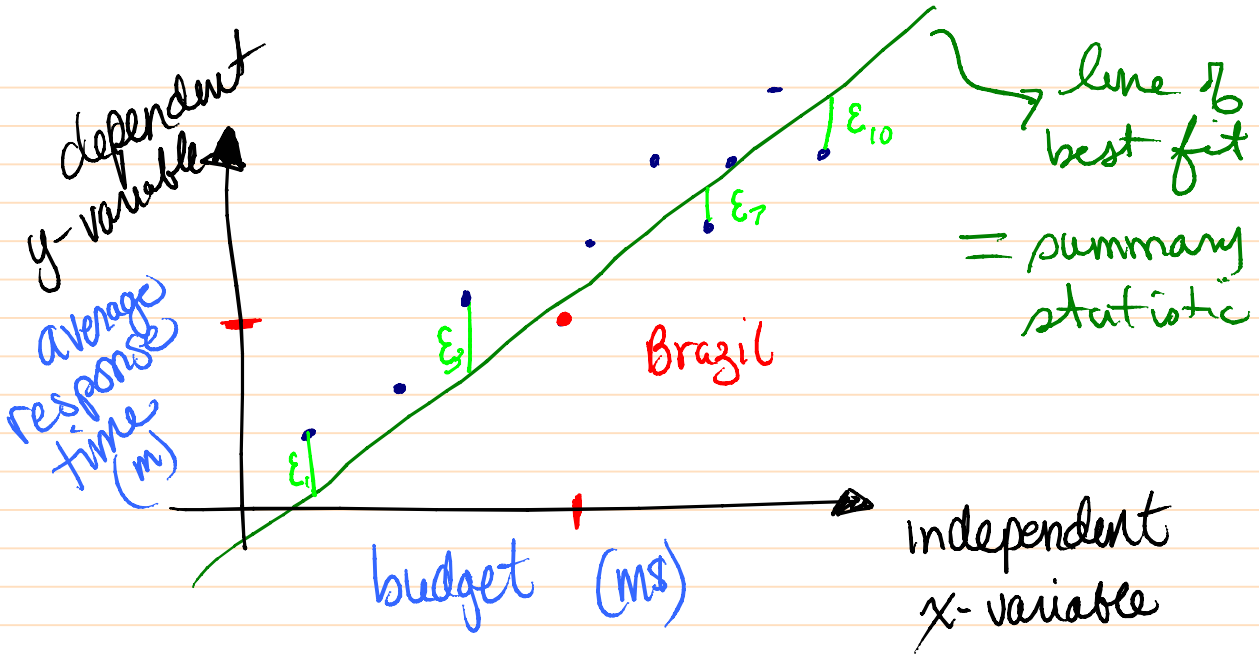
Thus far: dep var = continuous

indep var = discrete

Next: dep var = continuous

ind var = continuous

Usually plotted on a scatterplot




With assumptions, the line of best fit becomes a regression line

Assumptions: 1. Independent variables ARE independent

2. $\varepsilon \sim N(0, \sigma^2)$

These assumptions are needed to draw conclusions on the population using our sample of data.

In R, the function to do linear modeling is lm:

lm( depvar ~ indepvar$_1$ + indepvar$_2$ + ... )

the prediction equation is

$$\text{depvar} = \hat{\beta}_0 + \hat{\beta}_1 \text{indv}_1 + \hat{\beta}_2 \text{indv}_2 + \cdots$$

this is an additive model

$\Rightarrow$ effects are independent of each other

If we believe the effects are NOT indep,
we use an interaction model

$lm(depvar \sim indep_1 * indep_2 + ind_3)$

ex: $lm(happy \sim income * male)$

This means: income effect differs for the different genders

# Testing assumptions

1. Indep vars are independent

   Correlation table & Correlation tests of the independent vars.

   cor( data )  library(RFS)

2. $\varepsilon \sim N(0, \sigma^2)$

1. residuals $= \varepsilon$

$\mu = 0$ : mean(resids)

Normality: shapiro.test(resids)

$\sigma^2$ constant: cor(resid, dep)

cor.test(resid, dep)

residual plot:

plot(resid ~ dep)

plot(model)