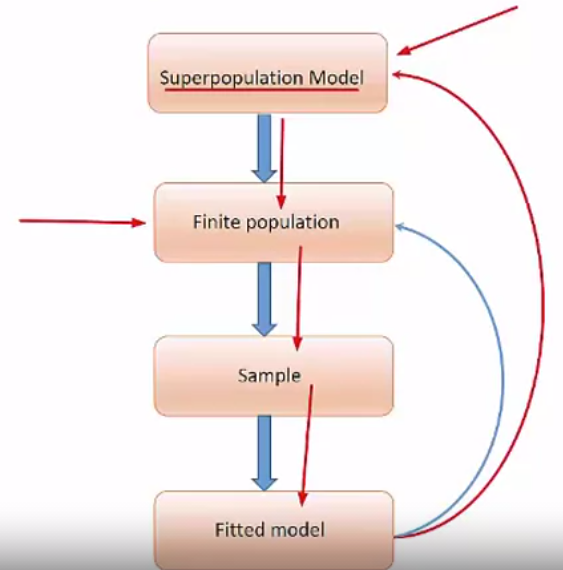
**W2 Models**

**Models**



General Considerations

• One way to think of goal of estimation is to estimate the model that would be fitted if entire population were in hand – “census model”

• If same model is appropriate for sample as for full population, then weights do not have to be used

• Stratification and clustering still need to be accounted for because the affect SEs

• But, using weights does insure that we are aiming at the census model

**Estimation Method**

Pseudo-maximum Likelihood Estimation

• Estimating equation method

• Write down full finite population likelihood

• Derive census estimation equations

- These will often be finite population totals that involve residuals

• Construct sample estimator of census estimating equations



• Solve for estimates of model parameters

• Software will do this for quite a few models

Software Capabilities

• R survey

- linear regression, logistic, probit, complementary log-log, Poisson, Loglinear, Cox proportional hazards model

• Stata

- linear regression, logistic, probit, complementary log-log, Possion, Loglinear, Cox Proportional hazards model

- parametric survival, Multinomial logistic, conditional logit, negative binomial, ordered logistic, probit, ordered probit, structural equation modeling, censored and interval regression, instrumental-variables regression, heckman selection model, probit estimation with selection, nonlinear least squares, multilevel models

API dataset in R survey

• Use academic performance index file from R survey

• API is computed for all California schools based on standardized testing of students

• Several datasets: information for all schools with at least 100 students and for various probability samples of data

• One record per school

Variables & Model syntax

• Specify survey design with *svydesign*

• svyglm(formula = …. , design = …..) to fit the model

• Variables

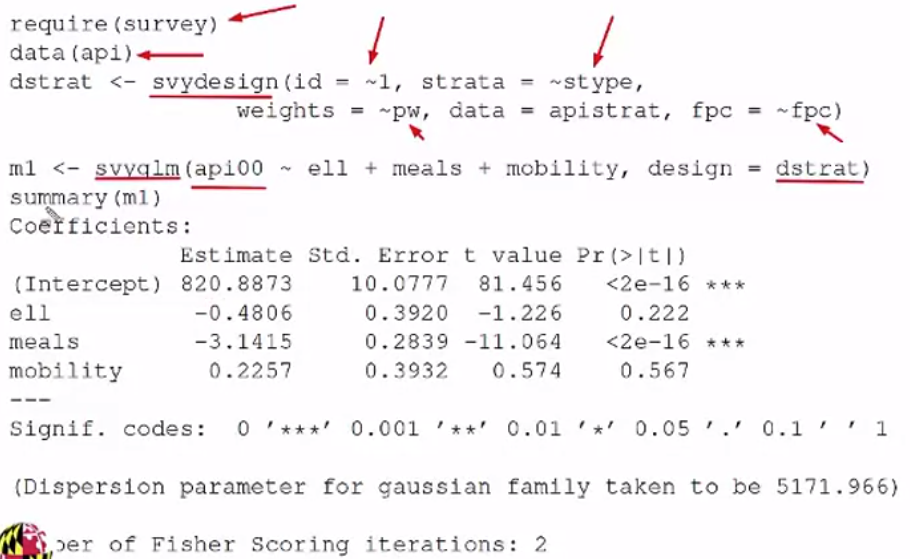
- api00 API in 2000

- ell English language learners (%)

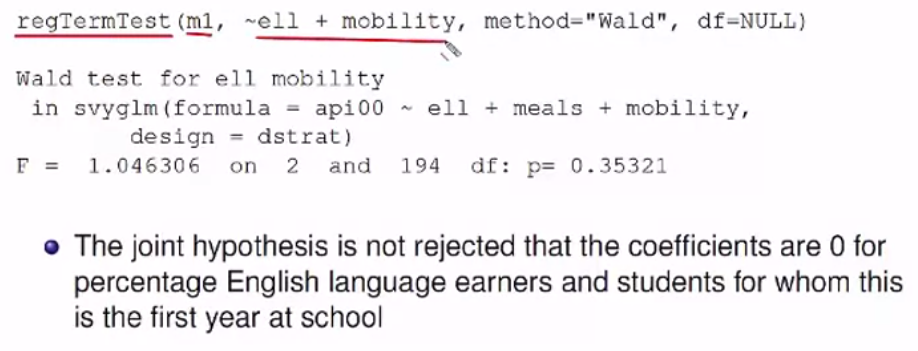
- meals % of students eligible for subsidized meals

- mobility % of students for whom this is the first year at the school

R code



Testing coefficient estimates



**Diagnostics**

Diagnostics

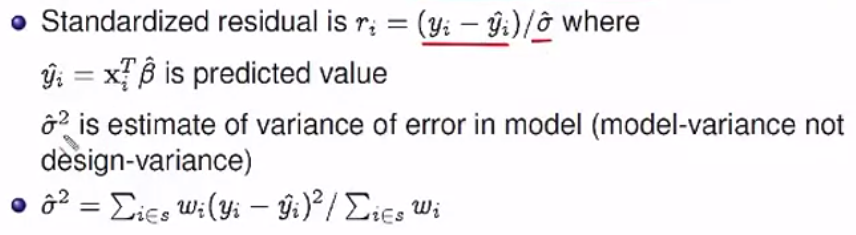
• some literature on adapting standard diagnostics for use with survey data

• few options available now in packages 🡺 program your own

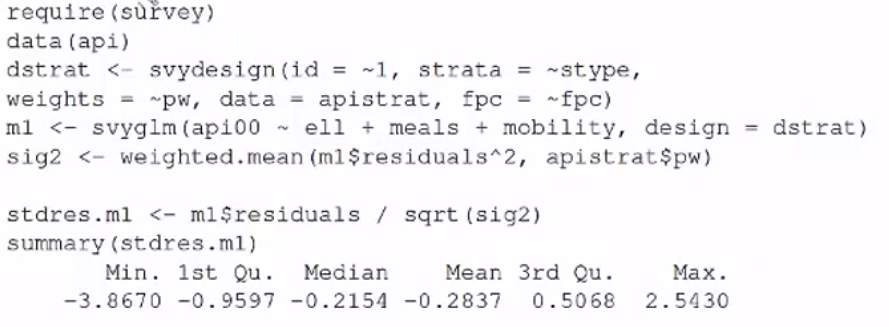
Compute Standardized residuals

• same model as in previous video

• standardized residuals have mean 0, variance 1 under the model

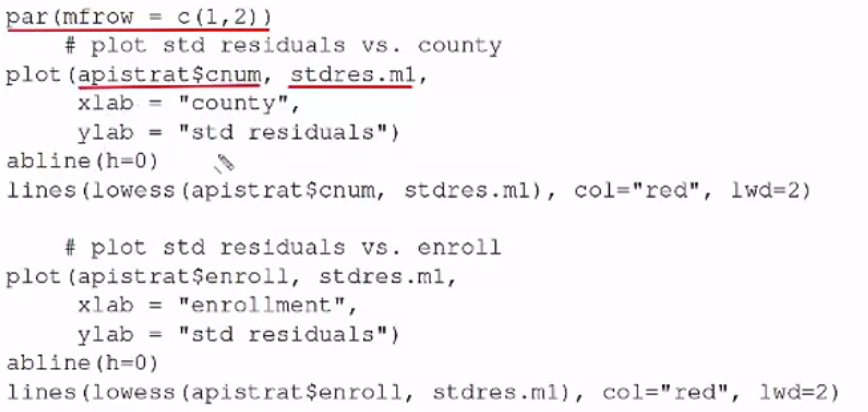


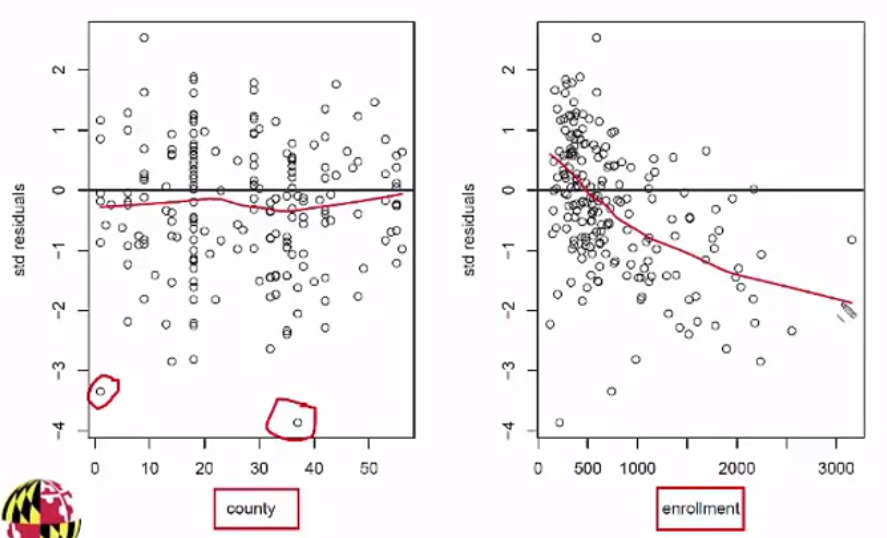
R code to compute standardized residuals



Standardized residual plots

Plotting v.s. variables not in model is a way to look for omitted covariates



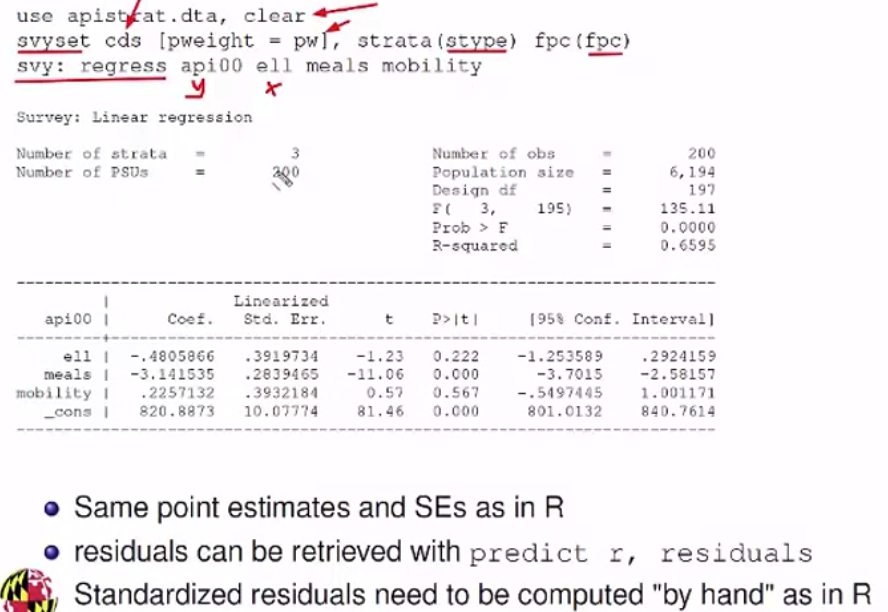


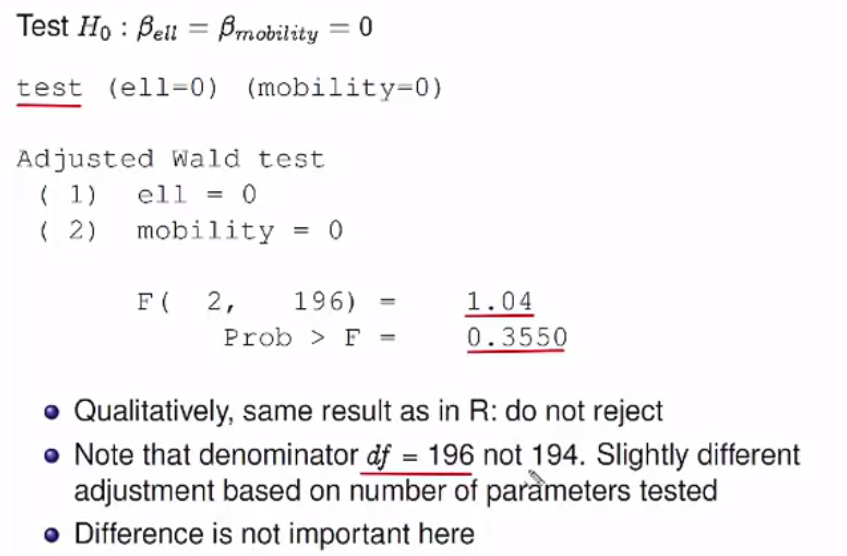
**Linear Models in Stata**

Fit same model as in R

• Regress school academic performance indexes (*api00*) on percentage English Language Learners (*ell*),

Percentage of students eligible for subsidized meals (*meals*), and percentage of students for whom this is the first year at the school (*mobility*)





•