STAT 151A Lecture 35: Quiz 3 Review

Henry Liev

17 November 2025

Remark 0.1 (Roadmap)

- Heteroskedastic Linear model
- Diagnostics
  - Influential points
  - Checking model assumptions
- Model selection
  - Crieria
  - Search Strategies
  - Ridge
  - LASSO
- Logistic regression
  - Model (vs LPM)
  - Asymptotic Inference

Remark 0.2 (HLM)

What is it?
$y_i = x_i^\intercal \beta + \varepsilon_i, \mathrm{Var}(\varepsilon_i) = \sigma_i^2, \mathbb{E}(\varepsilon_i|x_i) = 0, \varepsilon_i \perp\!\!\!\perp \varepsilon_j, i \neq j$
$\mathbb{E}(\hat{\beta}|\mathbf{X}) = \beta, \mathrm{Cov}(\hat{\beta}|\mathbf{X})$ different from NLM
NLM: $\sigma^2(\mathbf{X}^\intercal\mathbf{X})^{-1}$
HLM: $(\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\Omega\mathbf{X}(\mathbf{X}^\intercal\mathbf{X})^{-1}, \Omega = \mathrm{diag}(\sigma_i^2) \approx \mathrm{diag}(e_i^2)$
How to do inference?
Need $n \to \infty$. As this happens
$\hat{\beta} \overset{n\to\infty}{\sim} \mathcal{N}(\beta, (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\Omega\mathbf{X}(\mathbf{X}^\intercal\mathbf{X})^{-1})$
Can use this to do Z-tests for $H_0 : \beta_j = 0$, build confidence intervals, model comparisons with F-statistics (χ^2 distributions)

Remark 0.3 (Diagnostics)

How can a point be unusual?

(1) Regression Outlier, Value of y_i is surprising compared to $\mathbb{E}(y_i|x_i)$
Measure via:

- Residual: e_i
- Standardized Residual $\tilde{e}_i = \frac{e_i}{\hat{\sigma}^2\sqrt{1-h_{ii}}}, \mathrm{Cov}(\vec{e}) = \sigma^2(\mathbb{I} - \mathbb{H})$
- Studentized Residual $e_i^* = \frac{e_i}{\hat{\sigma}_{(-i)}\sqrt{1-h_{ii}}} = \tilde{e}_i\sqrt{\frac{n-p-2}{n-p-1-\tilde{e}_i^2}} \overset{NLM}{\sim} t_{n-p-2}$

Can test for large residuals (but remember Bonferroni Correction)

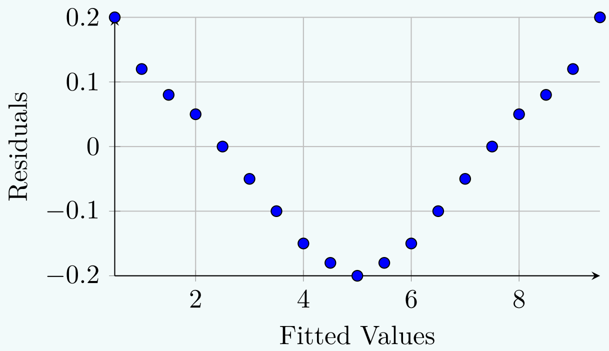
(2) Leverage − all about covariates, unusual in its $\vec{x}_i-$ value
To measure: leverage $h_i i = \mathbf{H}_{ii} = \mathbf{X}(\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal, \frac{1}{n} \leq h_{ii} \leq 1$
Can interpret h_{ii} as a weighted distance of $\vec{x}_i$ from vector $\bar{x}$

(3) Influence: How much does removing this point change $\hat{\beta} \to$ Cook's Distance, $\hat{\sigma}^2 \to$ COVRATIO
Cook's distance: $D_i = \frac{(\hat{\beta}-\hat{\beta}_{(-i)})^\intercal(\mathbf{X}^\intercal\mathbf{X})(\hat{\beta}-\hat{\beta}_{(-i)})}{(p+1)\hat{\sigma}^2} = \frac{(\tilde{e}_i)^2}{p+1} \cdot \frac{h_{ii}}{1-h_{ii}}$
COVRATIO: $\frac{1}{(1-h_{ii})\left(\frac{n-p-2-{e_i^*}^2}{n-p-1}\right)^{p+1}}$
If COVRATIO > 1, omitting i makes CI for $\hat{\beta}$ wider
If COVRATIO < 1, omitting i makes CI for $\hat{\beta}$ narrower

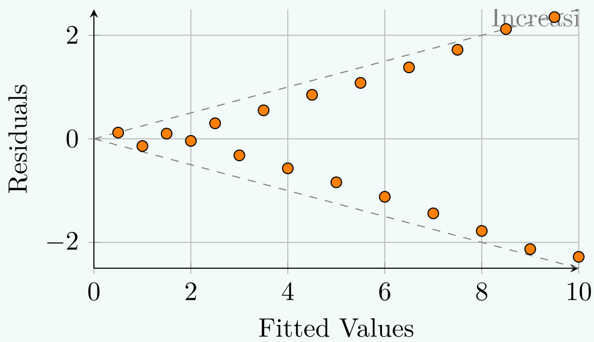
Remark 0.4 (Checking Assumptions)

Residual vs Fitted

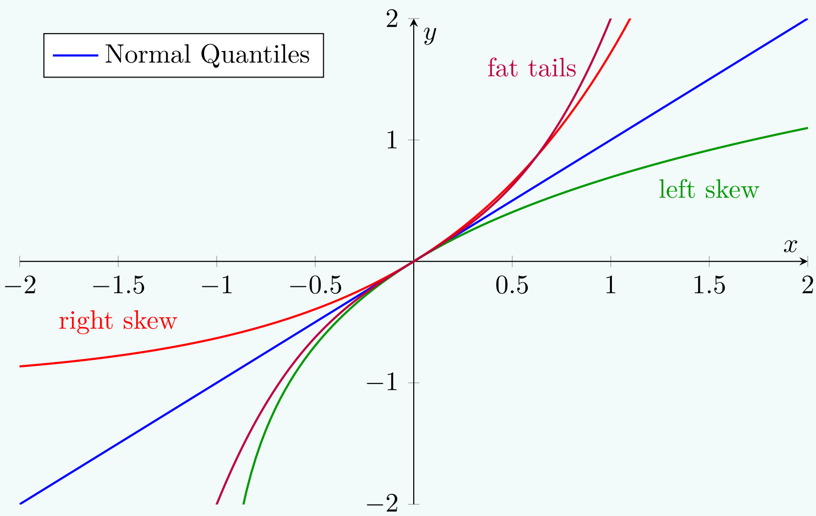
- Look for trends/patterns: Suggests violations of linearity $\mathbb{E}(\varepsilon_i|x_i) \neq 0$, problem for both NLM and HLM



- Also look for uneven spread, suggests violation of homoskedasticity, fix with a transform or use HLM or bootstrap cases



- Q-Q Plot $\rightarrow$ check normality
  Hope to see points along $y = x$
  If we have a line that looks like $\log x$, we can anticipate left skew, while something that looks like e^x is more right skewed, something that is like x^3 will have fat tails



Remark 0.5 (Model Selection)

Why? Less overfitted, simpler
Why not? No inference in some data you used to select model, leaving out variables may create hidden cofounders
Pick a criterion:

- Adj R^2
- Mallow's C_p
- CV Error
- AIC
- BIC, big preface, value of ErrSS against number of predictors p

Pick a search strategy
All subsets:

- All subsets (General, but slower)
  - Fit all possible models
  - Comptue scores
  - Pick best one
- Forward, backward, or stepwise search (Fast, but greedy)
  - Start with a model
  - Look at all ways to add/drop/both a variable
  - Choose the one that improves criteria best
  - Repeat

Remark 0.6 (Shrinkage)

Model selection, but not variable selection
$\min_\beta \|\vec{y} - \mathbf{X}\beta\|^2 + \lambda \mathcal{S}(\beta)$
Ridge: $S(\beta) = \sum \beta_i^2 = \|\beta\|_2^2$ or Lasso: $S(\beta) = \sum |\beta_i| = \|\beta\|_1$
No intercepts are allowed in the β_i calculation for shrinkage, center and standardize variables before fitting
Easy to compute (Can pick λ by CV error)
Ridge: Has a closed form solution $\hat{\beta}_m = (\mathbf{X}^\intercal \mathbf{X} + \lambda \mathbb{I})^{-1} \mathbf{X}^\intercal \vec{y}$