Inference for Linear Model

Math 463, Spring 2017, University of Oregon

David A. Levin

May 1, 2017

Warm-up

- Download data from `http://pages.uoregon.edu/dlevin/DATA/crime.txt` (Use `read.table` with the option `header=T`)
- This data set contains crime rates and demographic variables for various towns. See `http://pages.uoregon.edu/dlevin/DATA/USCrimeDatafile.html`
- Model `R` as a function of the other variables.
- Find the p-value for the test that the coefficient of `U2` is different from 2.5.

Advertisement

Undergraduate lecture: “Proofs (not) from the Book”, 4pm, Monday, May 1, 2017 in Willamette Hall 100. There will be pizza after the lecture on the second floor of Fenton; undergraduates especially welcome.

This year’s Niven lectures are Monday and Tuesday of this week! 4pm both days, in Willamette 100.

Sergei Tabachnikov (Penn State)

`http://math.uoregon.edu/seminars/niven`

Title: Proofs (not) from the book

Abstract: The eminent mathematician of the 20th century, Paul Erdos, often mentioned “The Book” in which God keeps the most elegant proof of every mathematical theorem. So, attending a mathematical talk, he would say: “This is a proof from The Book!”, or “This is a correct proof, but not from The Book”. M. Aigner and G. Ziegler authored the highly successful “Proofs from THE BOOK” (translated into 13 languages). In this talk, I shall present several proofs that are not included in the Aigner-Ziegler book but that, in my opinion, could belong to “The Book”.

Key points for this unit

- Inference for individual coefficients β_i and linear combination of coefficients $(\sum_i a_i\beta_i)$ using t-distribution.
- Definition of R^2 statistic and other RESIDUAL SUM-OF-SQUARES, REGRESSION SUM-OF-SQUARES.
- Testing if
$$\mathbb{E}(\boldsymbol{Y} \mid \boldsymbol{X}) = \boldsymbol{X}\boldsymbol{\beta} := \boldsymbol{\theta} \in V_0 .$$
  via the F-test.
- Power of above test using non-central F-distribution with parameter $\pi(\boldsymbol{\theta} \mid V \cap V_0^{\perp})$.
- Usefulness of SINGULAR VALUE DECOMPOSITION when design matrix is nearly singular.

Normal Theory Summary (not specific to LINEAR MODEL)

- Want to estimate a **parameter** θ using an estimator $\hat{\theta}$.
- Suppose that $\hat{\theta} \sim N(\theta, v\sigma^2)$, and that $m\hat{\sigma^2}/\sigma^2 \sim \chi^2_m$, independent of $\hat{\theta}$.
- Example: $Y_1, \ldots, Y_n$ are iid $N(\mu, \sigma^2)$ and $\bar{Y} \sim N(\mu, \sigma^2/n)$. Then $\theta = \mu$ and $\hat{\theta} = \bar{Y}$; note that $(n-1)S^2/\sigma^2$ is independent of $\bar{Y}$ and has a chi-squared distribution with $n-1$ degrees of freedom.
- Example: $\hat{\beta}_1$ estimates β_1 and has a $N(\beta_1, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}_{1,1})$ distribution. $(n-k)\mathrm{RSS}/\sigma^2$ has a chi-squared distribution.
- Then
$$\frac{\hat{\theta} - \theta}{\hat{\sigma}\sqrt{v}}$$
has a t-distribution with m degrees of freedom.
- Hypothesis test and confidence intervals on θ easy to build from the above.
- Can apply the above to $\theta = \boldsymbol{a}'\boldsymbol{\beta}$ for any $\boldsymbol{a}$.

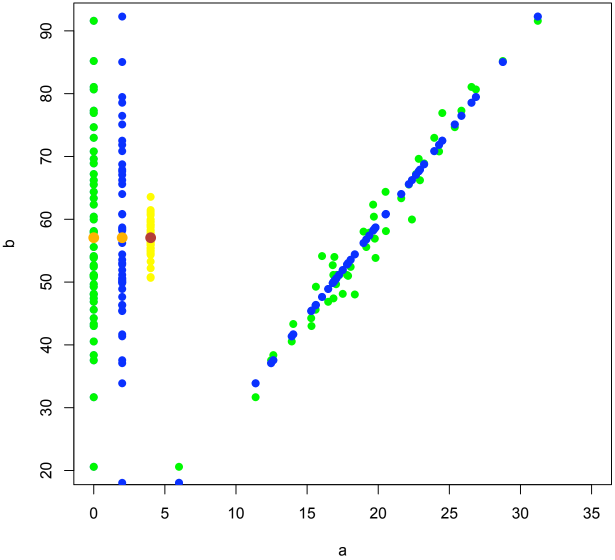
Sum of squares decomposition

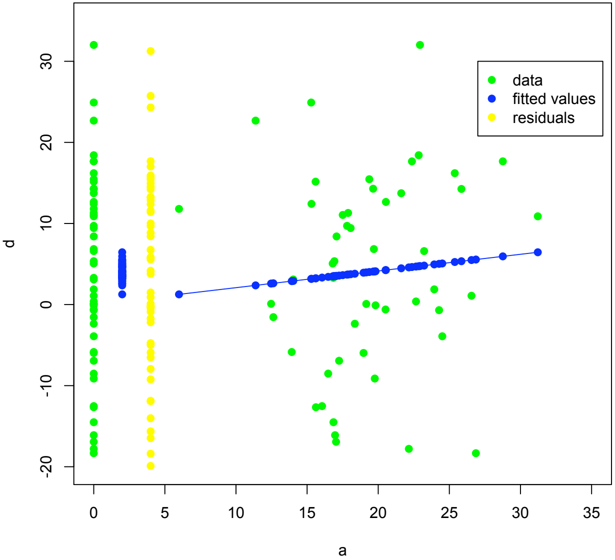
- $$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2$$
- $$\mathrm{SS}_{\mathrm{tot}} = \mathrm{SS}_{\mathrm{reg}} + \mathrm{SS}_{\mathrm{res}}$$
- Set $\hat{\boldsymbol{Y}}_0 = \bar{Y}\mathbf{1} = \pi(\boldsymbol{Y} \mid \mathbf{1})$.

  $$\|\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0\|^2 = \|\hat{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_0\|^2 + \|\boldsymbol{Y} - \hat{\boldsymbol{Y}}\|^2$$

  (Why is $\boldsymbol{Y} - \hat{\boldsymbol{Y}} \perp \hat{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_0$)?

green = data, blue = fitted values, yellow = residuals





data
fitted values
residuals
d
a
-20
-10
0
10
20
30
0
5
10
15
20
25
30
35

- If $\mathrm{SS}_{\mathrm{reg}}$ is small as compared to $\mathrm{SS}_{\mathrm{res}}$, then most of the variability in the Y_i's is due to the $\boldsymbol{\varepsilon}_i$'s.
- If the ratio of $\mathrm{SS}_{\mathrm{reg}}$ to $\mathrm{SS}_{\mathrm{res}}$ is large, then most of the variability in the Y_i's is due to the linear dependence of $\mathbb{E}(\boldsymbol{Y} \mid \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k)$ on $\boldsymbol{x}_j$'s.
- The multiple R^2 is defined as

$$R^2 = \frac{\mathrm{SS}_{\mathrm{reg}}}{\mathrm{SS}_{\mathrm{tot}}}$$

- The value R^2 is in $[0, 1]$, and is interpreted as *the percent of variation in the response data that is explained by the model.*

  Note

$$R^2 = \frac{\mathrm{SS}_{\mathrm{tot}} - \mathrm{SS}_{\mathrm{res}}}{\mathrm{SS}_{\mathrm{tot}}} = 1 - \frac{\mathrm{SS}_{\mathrm{res}}}{\mathrm{SS}_{\mathrm{tot}}}\,.$$

Chi-Squared

- Suppose that $Q = \boldsymbol{W}'\boldsymbol{A}\boldsymbol{W}$ is a quadratic form and $\boldsymbol{W}$ is multivariate Normal with mean vector $\boldsymbol{\theta}$ and positive definite covariance matrix $\boldsymbol{\Sigma}$.
- Note that Σ has a square-root $\Sigma^{1/2}$; let $\boldsymbol{B} = \Sigma^{1/2}\boldsymbol{A}\Sigma^{1/2}$, and $\boldsymbol{V} = \Sigma^{-1/2}\boldsymbol{W}$, so that

$$Q = (\Sigma^{-1/2}\boldsymbol{W})'\Sigma^{1/2}\boldsymbol{A}\Sigma^{1/2}(\Sigma^{-1/2}\boldsymbol{W}) = \boldsymbol{V}'\boldsymbol{B}\boldsymbol{V}.$$

  Note that $\boldsymbol{V} \sim N(\Sigma^{-1/2}\boldsymbol{\theta}, I)$.
- Since $\boldsymbol{B} = \boldsymbol{T}'\Lambda\boldsymbol{T}$ by spectral decomposition, we have

$$Q = \boldsymbol{T}\boldsymbol{V}'\boldsymbol{\Lambda}\boldsymbol{T}\boldsymbol{V}.$$

- Note that $\boldsymbol{Z} := \boldsymbol{T}\boldsymbol{V} \sim N(\boldsymbol{T}\Sigma^{-1/2}\boldsymbol{\theta}, I)$. Thus,

$$Q = \sum_{i=1}^{n} \lambda_i Z_i^2,$$

  where $Z_i^2 \sim \chi_1^2((\boldsymbol{T}\Sigma^{-1/2}\boldsymbol{\theta})_i^2)$

- The non-central chi-squared distribution is the distribution of $\sum_{i=1}^{m} Z_i^2$, where $Z_i \sim N(\delta_i, 1)$. The degrees of freedom equals m and the non-centrality parameter is $\delta = \sum_{i=1}^{m} \delta_i^2$.
- Special case: $\Sigma = \sigma^2 I$, and $\boldsymbol{A}$ is the projection onto V_1, with $\dim({}_1 V) = r$
- Let $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ be an orthonormal basis such that $\boldsymbol{v}_1, \dots, \boldsymbol{v}_r$ span V_1. Note that $\boldsymbol{v}_1, \dots, \boldsymbol{v}_r$ are eigenvectors of A with eigenvalue 1, and $\boldsymbol{v}_{r+1}, \dots \boldsymbol{v}_n$ are eigenvectors with eigenvalue 0. Thus if $\boldsymbol{T}' = [\boldsymbol{v}_1 \boldsymbol{v}_2 \dots \boldsymbol{v}_n]$, then

$$\boldsymbol{B} = \sigma^2 \boldsymbol{T}' \text{diag}(1, 1, \dots, 1, 0, 0, \dots, 0) \boldsymbol{T} .$$

and

$$Q = \sigma^2 \sum_{i=1}^{r} Z_i^2 .$$

Note that

$$(\boldsymbol{T\theta})_i = \langle \boldsymbol{t}_i, \boldsymbol{\theta} \rangle ,$$

and

$$\sum_{i=1}^{4} (\boldsymbol{T\theta})_i = \frac{1}{\sigma^2} \|\pi(\boldsymbol{\theta} \mid V)\|^2$$

- Thus
$$\frac{Q}{\sigma^2} \sim \chi^2_r\left(\frac{\|\pi(\boldsymbol{\theta} \mid V_1)\|^2}{\sigma^2}\right).$$
- If $V_1 = V \cap V_0^{\perp}$, then $r = \dim(V_1) = k - \dim(V_0)$.

- If $\boldsymbol{U}$ and $\boldsymbol{V}$ are random vectors, then

$$\mathrm{Cov}(\boldsymbol{U}, \boldsymbol{V})_{i,j} = \mathrm{Cov}(U_i, V_j)\,.$$

- Note

$$\mathrm{Cov}(A\boldsymbol{U}, B\boldsymbol{V}) = A\mathrm{Cov}(U, V)B'\,.$$

- Note that if V and W are perpendicular subspaces and $\Sigma = \sigma^2 I$, then

$$\mathrm{Cov}(\Pi_V \boldsymbol{Y}, \Pi_W \boldsymbol{Y}) = \Pi_V \Pi_W = 0\,.$$

- Let $V_0 \subset V$. Suppose $\dim(V_0) = r < k = \dim(V)$. Then

$$\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0 = \boldsymbol{Y} - \hat{\boldsymbol{Y}} + \hat{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_0$$

- Under normality assumption on errors,

$$\begin{aligned} \|\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0\|^2 &= \|\boldsymbol{Y} - \hat{\boldsymbol{Y}}\|^2 + \|\hat{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_0\|^2 \\ \chi^2_{n-r} &= \chi^2_{n-k} + \chi^2_{k-r} . \end{aligned}$$

  Also, the two on the right are independent (why?)
- Note the non-centrality parameter for the first on the right is 0 (always), and the non-centrality parameter for the second is $\pi(\boldsymbol{\theta} \mid V \cap V_0^{\perp})/\sigma^2$.
- If $\boldsymbol{\theta} \in V_0$ then the second non-centrality parameter is 0.
- We can rewrite this as

$$\text{RSS}_0 = \text{RSS} + \text{RegSS}$$

Thus

$$F = \frac{\mathrm{RegSS}/(k-r)}{\mathrm{RSS}/(n-k)}$$

has a $F_{k-r,n-k}$ distribution when $\boldsymbol{\theta} \in V_0$.
In general, F has a non-central F distribution with non-centrality parameter equal to $\boldsymbol{\pi}(\boldsymbol{\theta} \mid V \cap V_0^{\perp})$.
Example: $V_0 = \mathcal{L}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r)$. Testing $\boldsymbol{\theta} \in V_0$ is equivalent to testing $\beta_i = 0$ for $i > r$.

Singular Value Decomposition

$$\boldsymbol{X} = \boldsymbol{U}\,\boldsymbol{D}\,\boldsymbol{V}'$$

where $\boldsymbol{U}$, $\boldsymbol{V}$ are orthogonal, $\boldsymbol{D}$ is diagonal. Then

$$\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' = \boldsymbol{U}\,\boldsymbol{D}\,\boldsymbol{V}'\,\boldsymbol{V}\,\boldsymbol{D}^{-2}\,\boldsymbol{V}'\,\boldsymbol{V}\,\boldsymbol{D}\,\boldsymbol{U}' = \boldsymbol{U}\,\boldsymbol{U}'$$

To find $\boldsymbol{x}_i^{\perp}$, helpful to use `model.matrix` and `svd`.