ST3131 Cheatsheet
by Yiyang, AY22/23

1. Simple Linear Regression

Simple Regression Model

Consider regressing Y on X:

$$Y = \beta_0 + \beta_1 + \epsilon$$

Here X is called **Covariate**, **Predictor** or **Regressor**, and Y **Response**.
The Regression function:

$$EY = \mathbb{E}[Y|X] = \beta_0 + \beta_1 X$$

Regression coefficients, $\beta_1 = \rho_{xy}\frac{\sigma_y}{\sigma_x}$ and $\beta_0 = \mu_y - \beta_1\mu_x$, minimising $\mathbb{E}(Y - \beta_0 - \beta_1 X)^2$

Observed ~

Assumptions of LRM

1. x_i and ϵ_i independent
2. $\frac{1}{n}\sum_{j=1}^n \epsilon_j = 0$
3. $Cov(\epsilon_i, \epsilon_j) = 0,\ \forall i, j$
4. **Homogenity**, $var\epsilon_j = \sigma^2$ for all j
5. **Normality**, $\epsilon_j \sim \mathcal{N}(\cdot,\cdot)$ for all j

Least Square Estimates of n observations (x_i, y_i) gives

$$\hat\beta_1 = \frac{\sum_{i=1}^n (x_i - \bar X)(y_i - \bar Y)}{\sum_{i=1}^n (x_i - \bar X)^2},\ \hat\beta_0 = \bar Y - \hat\beta_1 \bar X$$

, where $\hat X = \frac{1}{n}\sum_{i=1}^n x_i$ and $\hat Y = \frac{1}{n}\sum_{i=1}^n y_i$, and the estimates minimises

$$Q = Q(\hat\beta_0, \hat\beta_1) = \sum_{i=1}^n (y_i - \hat\beta_0 - \hat\beta_1 x_i)^2$$

Lastly, **Residual Standnard Error** $\hat\sigma^2 = s^2 = \frac{1}{n-2}\text{SSE}$ gives $\hat\sigma$ the LSE for σ.

Analysis of Variance (ANOVA)

Sum of Squares

Sum of Square Errors $\text{SSE} = \sum_{i=1}^n (y_i - \hat y_i)^2 = \sum_{i=1}^n e_i^2$, measures variation of Y due to random errors.
Regression Sum of Squares $\text{SSR} = \sum_{i=1}^n (\hat y_i - \bar Y)^2$, measures variation of Y explained by X.
Total Sum of Squares $\text{SST} = \sum_{i=1}^n (y_i - \bar Y)^2$

$$\text{SST} = \text{SSR} + \text{SSE}$$

Coefficients of Determination

Coefficients of Determination measures how much of Y is explained by X,

$$R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{\beta_1^2\sigma_x^2}{\beta_1^2\sigma_x^2 + \sigma^2}$$

Note: $R^2 \in [0, 1]$ and $R^2 = corr(X, Y)^2 = corr(Y, \hat Y)^2$.

Adjusted R^2 for a RM with p regressors,

$$R_a^2 = \frac{n-1}{n-p-1}R^2 - \frac{p}{n-p-1}$$

Note: [1] R^2 strictly increasing as p increases while R_a^2 does not. [2] Sample R^2 and R_a^2 are both biased estimated for their population counterparts, but latter less biased.

Theoretical Properties of LSE

Unbiasedness of LSE

$\hat\beta_0, \hat\beta_1, s$ are unbiased estimators for their population counterparts.
$\hat Y = \hat\beta_0 + \hat\beta_1 X$ unbiased estimator for $EY = \beta_0 + \beta_1 X$.

Standard Errors

$$\begin{aligned}
var(\hat\beta_0) &= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar X)^2} \\
var(\hat\beta_0) &= var(\bar Y) + \bar X^2 var(\hat\beta_1) \\
&= \frac{\sigma^2}{n} + \frac{\bar X^2}{\sum_{i=1}^n (x_i - \bar X)^2}\sigma^2 \\
var(\hat Y) &= \Big[\frac{1}{n} + \frac{(X - \bar X)^2}{\sum_{i=1}^n (x_i - \bar X)^2}\Big]\sigma^2
\end{aligned}$$

The sample SEs are estimated by substituting σ^2 with s^2.

Inferences for ~

Statistical Tests

Significance Test (ANOVA)
Null: $\beta_1 = 0$. Statistics: $F = \frac{\text{MSR}}{\text{MSE}} \sim F_{1,n-2}$.
Decision: Reject null if $F > f_{1,n-2}(\alpha)$

Test for β_1 (and β_0)
Statistic: $T_{\beta_1} = \frac{\hat\beta_1 - \beta_1}{s(\hat\beta_1)} \sim t_{n-2}$.
Confidence Interval: $\hat\beta_1 \pm t_{n-2}(\alpha/2)\ s(\hat\beta_1)$
Confidence Lower Bound: $\hat\beta_1 - t_{n-2}(\alpha)\ s(\hat\beta_1)$, upper similar.

Predictions

Confidence Interval for $E[Y|X = x_h]$:

$$\hat y_h \pm t_{n-2}(\alpha/2)\cdot\sqrt{\hat\sigma^2\Big(\frac{1}{n} + \frac{(x_h - \bar X)^2}{\sum_{i=1}^n (x_i - \bar X)^2}\Big)}$$

Prediction Interval for Y when $X = x_h$:

$$\hat y_h \pm t_{n-2}(\alpha/2)\cdot\sqrt{\hat\sigma^2\Big(1 + \frac{1}{n} + \frac{(x_h - \bar X)^2}{\sum_{i=1}^n (x_i - \bar X)^2}\Big)}$$

2. Multiple Linear Regression

Multiple Regression Model

Objective: To investigate relationship between Y and p predictors $\vec X = (X_1, \ldots, X_p)$ with n observations.

Least Square Estimation

Let $\boldsymbol\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T \in \mathbb{R}^{p+1}$,

$$\hat{\boldsymbol\beta} = (X^T X)^{-1}X^T\boldsymbol y$$
$$s^2 = \frac{\|\boldsymbol y - H\boldsymbol y\|^2}{n-p-1}$$

Here $X = X_{n\times(p+1)}$ is the **Design Matrix** where a column of 1 is joined to the left of observed predictors.

Hat Matrix

The **Hat Matrix** $H = H_{n\times n}$ is the projection matrix of linear space spanned by column vectors of X.

$$H = X(X^T X)^{-1}X^T$$

Properties of H:

- $HX = X$
- Idempotent, $H^2 = H$, so does $I - H$
- $H\mathbb{1} = \mathbb{1}$ for $\mathbb{1} = \text{One}(n, 1)$
- $H\boldsymbol{x_j} = \boldsymbol{x_j}$ for $\boldsymbol{x_j} = (x_{1j}, \ldots, x_{nj})^T$
- $\hat{\boldsymbol y} = H\boldsymbol y$ is the fitted values, and $\boldsymbol e = (I - H)\boldsymbol y$ is the residuals.

Explicit Expression for One Coefficient

$$\hat\beta_j = \frac{\boldsymbol{x_j}^T(I - H_{-j}\boldsymbol y)}{\boldsymbol{x_j}^T(I - H_{-j}\boldsymbol{x_j})}$$

, where X_{-j} is the Sub-Design Matrix with column $\boldsymbol{x_j}$ removed, and H_{-j} the corresponding Hat Matrix.

ANOVA

Sum of Squares

$$\mathrm{SST} = \boldsymbol{y}^T H_T \boldsymbol{y},\ \ H_T = I - \frac{1}{n}\mathbf{1}\mathbf{1}^T$$
$$\mathrm{SSR} = \boldsymbol{y}^T H_R \boldsymbol{y},\ \ H_R = H - \frac{1}{n}\mathbf{1}\mathbf{1}^T$$
$$\mathrm{SSE} = \boldsymbol{y}^T H_E \boldsymbol{y},\ \ H_E = I - H$$

Distributions of Sum of Squares

$$\frac{\mathrm{SSE}}{\sigma^2} \sim \chi^2_{n-p-1},\ \ \frac{\mathrm{SSR}}{\sigma^2} \sim \chi^2_p$$

ANOVA Table

	df	SS	MS	F
Regression	p	SSR	MSR	MSR/MSE
Error	n-p-1	SSE	MSE	
Total	n-1	SST		

Inferences for ~

Statistical Tests

Significance Test (ANOVA)
Null: $\boldsymbol{\beta} = \mathbf{0}$. Statistics, $F = \frac{\mathrm{MSR}}{\mathrm{MSE}} \sim F_{p,n-p-1}$.

Individual t-Test
Null: $\beta_i = 0$. Statistic: $T_{\beta_i} = \frac{\hat{\beta}_i}{s(\hat{\beta}_i)} \sim t_{n-p-1}$.

General Linear Hypothesis Test
For a given linear hypothesis $\boldsymbol{c} = (c_0, c_1, \ldots, c_p)^T$,
Null: $\boldsymbol{c}^T\boldsymbol{\beta} = 0$ Statistics: $T = \frac{\boldsymbol{c}^T\hat{\boldsymbol{\beta}}}{\sqrt{\boldsymbol{c}^T\hat{\Sigma}\boldsymbol{c}}} \sim t_{n-p-1}$

Predictions

Confidence Interval for $E[Y|\vec{X} = \boldsymbol{x_h}]$:

$$\hat{y}_h \pm \hat{\sigma}^2_F(\hat{y}_h) \times t_{n-p-1}(\alpha/2)$$

where the estimated variance is

$$\hat{\sigma}^2_F(\hat{y}_h) = \boldsymbol{x_h}^T Var(\hat{\boldsymbol{\beta}})\boldsymbol{x_h} = \boldsymbol{x_h}^T(X^TX)^{-1}\boldsymbol{x_h}\hat{\sigma}^2$$

Prediction Interval for Y when $\vec{X} = \boldsymbol{x_h}$:

$$\hat{y}_h \pm \hat{\sigma}^2_P(\hat{y}_h) \times t_{n-p-1}(\alpha/2)$$

where the estimated variance is

$$\hat{\sigma}^2_P(\hat{y}_h) = \left[1 + \boldsymbol{x_h}^T(X^TX)^{-1}\boldsymbol{x_h}\right]\hat{\sigma}^2 = \hat{\sigma}^2 + \hat{\sigma}^2_F$$