

MA2261 - DLI, Linear Statistical Models

Year 2022-2023

Revision checklist

1 Exam Format

- 1) In-person exam to be done within 2 hours.
- 2) Four questions on the following topics,
 - Probability and Statistical inference
 - Simple linear regression
 - Simple linear regression with repeated observations
 - Multiple linear regression
 - One-way ANOVA
- 3) Each question is worth 25 points and consists of 4-5 sub-questions.
- 4) All the material presented in the slides will be examinable. It's crucial to revise the slides and the coursework, and to practice exercises from the workbook without immediately checking the solutions. The emphasis should be on understanding, not just memorizing.
- 5) About 30% of challenging questions to distinguish high first.
- 6) Formula sheet and statistical tables will be provided.
- 7) **Caution: The topics listed on this revision checklist are not the only ones that will appear on the exam! It's important to also review other seemingly trivial items!**

2 Probability and Statistical Inference

To be revised on your own.

3 Simple Linear Regression

3.1 Some definitions and theoretical results

- 1) Definition: Simple linear regression model.
- 2) Definition: Explanatory variable and response variable.
- 3) Theorem: Maximum likelihood estimation of the model parameters a and b in the simple linear regression model.
- 4) Definition: Fitted values, Residuals, Standardized residuals.
- 5) Definition: Residuals sum of squares, RSS.
- 6) Corollary: $\sum_{i=1}^n r_i = \sum_{i=1}^n x_i r_i = 0$.
- 7) Definition: Sampling distributions of \hat{a} , \hat{b} .
- 8) Theorem: $\hat{b} \sim N\left(b, \frac{\sigma^2}{S_{xx}}\right)$, $\hat{a} \sim N\left(a, \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right) \sigma^2\right)$.
- 9) Definition: Correlation coefficient, Sample correlation coefficient.
- 10) Definition: Total variation SST, Error variation SSE, Model variation SSM.
- 11) Definition: Pure error sum of squares SSW, Between groups variation SSB, Lack of fit sum of squares SSL.
- 12) Theorem: $SST=SSM+SSE$, $SST=SSW+SSB$, $SSE=SSW+SSL$.

3.2 Some typical theoretical questions

- 1) Show that the sampling distributions of \hat{a} , \hat{b} are normal.
- 2) Show that if random variables X , Y are independent, then $\rho = 0$ but not conversely.
- 3) Show that the sum of the residuals in the simple linear regression model is zero.
- 4) Show that $\hat{b} = \frac{S_{xy}}{S_{xx}}$, $\hat{a} = \bar{y} - \hat{b}\bar{x}$.
- 5) State the assumptions of the simple linear regression model.
- 6) Define the correlation coefficient between two random variables X and Y .
- 7) Show that \hat{a} , \hat{b} are unbiased estimators of a , b .

3.3 Some typical computational questions

- 1) Calculate the point estimates of the model parameters.
- 2) Test the hypothesis that the slope parameter is equal to zero.
- 3) Test the hypothesis that the intercept parameter is equal to zero.
- 4) Calculate the 95% C.I. for the correlation coefficient.
- 5) Calculate the proportion of the total variation in the response explained by the explanatory variable.
- 6) Establish if the explanatory variable is statistically significant in affecting the mean response.
- 7) Calculate the 95% C.I. for the mean response corresponding to a new observation x_0 .
- 8) Calculate the 95% prediction interval corresponding to a new observation x_0 .
- 9) Is the simple linear regression model a valid model for the dataset?
- 10) Calculate the ANOVA Table.

Simple linear regression with repeated observations

Note: All the computational questions above can also be asked in this case.

- 1) Calculate the ANOVA Table with pure error and lack of fit.
- 2) Test the hypothesis that the simple linear regression model is true.

4 Multiple Linear Regression

- 1) Definition: Multiple linear regression model.
- 2) Definition: Polynomial regression model.
- 3) What is the model equation? How many parameters are there?
- 4) Write the model and the model assumptions in matrix form.
- 5) Show that $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$.
- 6) Calculate the least squares estimates of the parameters.
- 7) Show that the estimator $\hat{\beta}$ is unbiased.
- 8) Derive the distribution of the estimator $\hat{\beta}$.

- 9) Are any of the predictors statistically significant in affecting the mean response?
- 10) Hypothesis testing for a single predictor variable, and the C.I.
- 11) Calculate the C.I. for future expected value and prediction interval for future observed value.
- 12) Definition: Coefficient of multiple determination.
- 13) Calculate the coefficient of multiple determination.
- 14) Calculate the ANOVA Table.
- 15) Calculate the extra sum of squares and make model comparisons.
- 16) Sequential sum of squares.

5 One-way ANOVA

- 1) Definition: Categorical variables
- 2) Assumptions for one-way ANOVA.
- 3) Write the model in matrix form.
- 4) Least squares estimates of the parameters.
- 5) Compute the ANOVA table.
- 6) Determine whether there is a statistically significant difference between the means of each group.
- 7) Theorem: $SST = SSB + SSE$.
- 8) Definition of contrast and its meaning.
- 9) Point estimate, standard error and 95% C.I. of the contrast.