NMSA407 Linear Regression
Exam Details, winter term 2021–22

1 Admittance

Student is admitted to the exam only if (s)he has passed the exercise class *final test*, his/her first homework was accepted and the second homework was either accepted or a revision was requested and the revised version has not been evaluated yet. Obtained grade (of any type) will be recorded in SIS only after getting the *course credit* (if that happens).

---

2 Exam Terms

Exam terms are/will be open for enrollment in SIS.

---

3 Written part

Written part lasts **120** minutes. Leaving the lecture room is only possible after handing in the solutions to the examiner. It is then not possible to return to the lecture room during the written part of the exam.

One A4 page covered on its one side by whatever is allowed to be used during the written part of the exam.

3.1 Assignments

Assignments of the written part of the exam will be provided in **English**. Questions concerning clarification of the assignment might be asked in English or Czech/Slovak. Answer will follow in the same language as the question.[1] Solutions can be worked out in either of languages: English, Czech, Slovak.

The written part will be composed of three main problems/assignments:

Problem 1 Specification of a linear model for a "practical" situation ("fairytale"). The model must conform to all assumptions mentioned in the description of the situation.

Problems 2 and 3 Proof or derivations being directly *motivated* by theoretical parts of the lecture. The assignment might be motivated by the model proposed in Problem 1.

[1] Czech and Sloval languages are considered as equivalent for this purpose.

3.2 Evaluation of the problems

Each main problem will be graded on a classical scale 1, 2, 3, 4:

- **1**: excellent solution with perhaps few minor mistakes/omissions/....
- **2**: solution with some mistakes/omissions/... but still being safely acceptable.
- **3**: solution which is close to being non-acceptable.
- **4**: non-acceptable and/or wrong solution (substantially incomplete, gross errors, ...).

3.3 Recursion

In a solution of the written exam assignment, it is necessary to define all involved quantities. Note that statements of theorems and lemmas in the lecture notes often include quantities that were defined separately and earlier. Also those must be defined or their meaning explained in the solution!

If proof or derivation is requested then it is necessary to proof/derive all those arguments that were provided during the *NMSA407 Linear Regression* lecture unless explicitely stated that some parts of the proof/derivation can be omitted. That is, if proof of Theorem A uses arguments of Lemma B and both theorems/lemmas A and B are covered by the *NMSA407* lecture then it is necessary to prove both theorems A and B. It is not necessary to prove theorems provided in the Appendix of the lecture notes. In other words, knowledge gathered outside the *NMSA407* lecture can be taken for granted, knowledge gathered during the *NMSA407* lecture must be justified (unless stated otherwise).

---

4 Oral part

Oral part of the exam will be carried out in English or in Czech/Slovak depending on the student's option. It will involve discussion about solutions to the problems assigned for the written part and possibly also discussion about solutions of an additional assignment given to student. Such assignment would be given at the beginning of the oral part after which student gets a certain amount of time to prepare his/her answers on a paper that will then serve as a basis for discussion with the examiner.

An exact "technical" form of the oral part will be specified separately for each exam term depending on current antiCOVID regulations. The oral part might also take place on-line using the ZOOM platform (camera will be requested).

---

5 Overall evaluation

Grades from the written part will serve as a ground basis for the overall evaluation of the exam to be made after the oral part. The overall exam result will always be **4** if any of the main problems from the written

part is graded by **4**. The overall exam result will also be **4** if the performance during the oral part will be graded by **4** (irrespective of the grades obtained for the written part).

Grade "Fail" (4) means that both the written and the oral part of the exam must be repeated.

---

6 Content

Exam assignments (for both written and oral part) can be based on everything that was said during both lectures and practical exercises except those items that were mentioned as being "for information only" (historical remarks, exact form of the $\mathbb{C}$ matrices for the orthonormal polynomial contrasts with specific values of G, density of the multivariate t-distribution, ...) or "not for exam". Assignments can also be based on items that were left as home exercise (parts of some proofs or calculations). The whole lecture is summarized in the lecture notes, see

http://msekce.karlin.mff.cuni.cz/~komarek/vyuka/2021_22/nmsa407/2021-NMSA407-notes.pdf

Materials from the appendices will not be explicitely examined. It is only expected to be able to use the corresponding knowledge if necessary (in proofs/calculations requested for the exam).

The list below provides all theorems and lemmas stated during the lecture. It also mentions which proofs *are not requested* for exam and where by-heart knowledge of an expression that appears in the theorem/lemma *is not appreciated* (it is however assumed that student is able to derive corresponding expression if needed/requested in which case also appropriate time will be provided).

Next to the items mentioned below, knowledge of related definitions, notions, etc. and their understanding is assumed as well.

1 Linear Model

1. Lemma 1.1: Conditional mean and covariance matrix of the response vector.
2. Lemma 1.2: Moments of the error terms.

2 Least Squares Estimation

1. Lemma 2.1: Least squares estimator.
2. Lemma 2.2: Moments of the least squares estimator.
3. Lemma 2.3: Algebraic properties of fitted values, residuals and related projection matrices.
4. Theorem 2.4: Gauss–Markov.
5. Theorem 2.5: Gauss–Markov for linear combinations.
6. Lemma 2.6: Altenative expressions of residuals and residual sum of squares.
7. Lemma 2.7: Moments of residuals and residual sum of squares.

3 Basic Regression Diagnostics

1. Lemma 3.1: Moments of standardized residuals under normality.
   Proof not requested.
2. *Working knowledge of materials from Sections 3.1–3.3 is assumed.*

4 Parameterizations of Covariates

For majority of the material of Chapter 4, only knowledge allowing to use it in practical analyses is expected. This means, among the other things, being able to propose a model, being able to interprete a model or being able to compare critically suggested approaches.

5 Multiple Regression

For majority of the material of Chapter 5, only knowledge allowing to use it in practical analyses is expected. This means, among the other things, being able to propose a model, being able to interprete a model or being able to compare critically suggested approaches.

6 Normal Linear Model

1. Lemma 6.1: Error terms in a normal linear model.
2. Theorem 6.2: Least squares estimators under the normality.
3. Consequence of Theorem 6.2: LSE of the regression coefficients in a full-rank normal linear model.
4. Theorem 6.3: Confidence interval for the model based mean, prediction interval.
5. Theorem 6.4: Distribution of the linear hypotheses test statistics under the alternative.
   Skipped completely, not requested for exam.
6. Theorem 6.5: Distribution of the linear hypotheses test statistics under the alternative.
   Skipped completely, not requested for exam.

7 Coefficient of Determination

1. Lemma 7.1: Model with intercept only.
2. Lemma 7.2: Identity in a linear model with intercept.
3. Lemma 7.3: Breakdown of the total sum of squares in a linear model with intercept.

8 Submodels

1. Theorem 8.1: On a submodel.
2. Theorem 8.2: On submodels.
3. Lemma 8.3: Effect of omitting some covariates.
4. Theorem 8.4: On a submodel given by linear constraints.
   By-heart knowledge of expressions of $\boldsymbol{b}^0$ and $\widehat{\mathbf{Y}}^0$ not appreciated.

5. Lemma 8.5: Overall F-test.
   By-heart knowledge of the identity $F_0 = \frac{R^2}{1-R^2}\,\frac{n-r}{r-1}$ *not appreciated.*

9 Checking Model Assumptions

1. Lemma 9.1: Model with added regressors.
   By heart knowledge of the expressions within Lemma 9.1 not appreciated.
2. Lemma 9.2: Property of partial residuals.
3. *Working knowledge of materials from Sections 9.2–9.6 is assumed.*

10 Consequences of a Problematic Regression Space

1. Lemma 10.1: Bias in estimation of the squared norms.
2. Theorem 10.2: Estimated variances of the LSE of the regression coefficients.
   Proof/derivations not requested. With respect to the statement, it is sufficient to know that the estimated variance of the LSE of β_j *is a product of two terms: the first that does not depend on degree of collinearity, and the second which equals* $1/(1-R_j^2)$ *and is then defined as* VIF_j*. It is assumed that student is able to explain why multicollinearity is a problem and what are the consequences.*
3. Lemma 10.3: Variance of the LSE in the two models.
   Proof requested only for the variances of the fitted values.
4. Lemma 10.4: Mean squared error of the BLUP in a linear model.
5. Lemma 10.5: Properties of the LSE in a model with omitted regressors.
   By-heart knowledge of expressions in Lemma 10.5 not appreciated. Derivation of $\mathbb{E}\left[MS_{e,X} \,\middle|\, \mathbb{Z}\right]$ *not requested.*

11 Unusual Observations

1. Lemma 11.1: Three equivalent statements.
   Proof not requested.
2. Lemma 11.2: Equivalence of the outlier model and the leave-one-out model.
3. Lemma 11.3: Quantities of the outlier and leave-one-out model expressed using quantities of the original model.
   By heart knowledge of the expressions within Lemma 11.3 not appreciated.
4. Lemma 11.4: On studentized residuals.
   Not necessary to know and derive that $\text{var}(T_t) = \frac{n-r-1}{n-r-3}$.
5. *Section 11.4 (Influential diagnostics) will not be examined.*

12 Model Building

For the material of Chapter 12, only knowledge allowing to use it in practical analyses is expected.

13 Analysis of Variance

1. Lemma 13.1: Least squares estimation in one-way ANOVA linear model.
2. Lemma 13.2: Least squares estimation in two-way ANOVA linear models.
3. Consequence of Theorem 13.2: LSE of the means of the means in the interaction and the additive model with balanced data.
4. Lemma 13.3: Breakdown of the total sum of squares in a balanced two-way classification.

14 Simultaneous Inference in a Linear Model

1. Lemma 14.1: Studentized range.
2. Theorem 14.2: Tukey's pairwise comparisons theorem, balanced version.
3. Theorem 14.3: Tukey's pairwise comparisons theorem, general version.
   Proof not requested.
4. Theorem 14.4: Tukey's honest significance differences.
5. Theorem 14.5: Hothorn-Bretz-Westfall MCP for linear hypotheses in a normal linear model.
6. Theorem 14.6: Confidence band for the regression function.
   Not requested for exam.

15 General Linear Model

The whole Chapter 15 has been skipped and it is not requested for exam.

16 Asymptotic Properties of the LSE and Sandwich Estimator

1. Lemma 16.1: Consistent estimator of the second and first mixed moments of the covariates.
2. Theorem 16.2: Strong consistency of LSE.
3. Theorem 16.3: Strong consistency of the mean squared error.
4. Theorem 16.4: Asymptotic normality of LSE in homoscedastic case.
5. Consequence of Theorem 16.4: Asymptotic distribution of t- and F-statistics.
   Including explanation why in which sense the classical inference holds even without normality.
6. Theorem 16.5: Asymptotic normality of LSE in heteroscedastic case.
7. Theorem 16.6: Sandwich estimator of the covariance matrix.
8. Consequence of Theorem 16.5 and 16.6: Heteroscedasticity consistent asymptotic inference.
   Including explanation how to adjust the classical inference in case of heteroscedasticity.