

For graduating students : typically need a C (not C-) and may need a higher grade to fulfill "avg grade in major" requirement.

Guide to where you may stand now

NOTE: much of the grade (MT2, Final) is yet to be determined.

Because of dropping of lowest scores, most people get a most of the HW credit.

The tests make a bigger difference than one might think.

Guidance for MT1: avg. 8 std. dev posted.

avg score $\approx$ B -

± 1 std. dev : A - or C - (approx.).

Midterm 2 : In class (80 minutes)

Monday April 28.

① Things that can be on the midterm which are NOT on the practice midterm

② Things in the book that will NOT be on the midterm.

③ Questions about practice midterm.

① (A) Box-Cox procedure for selecting a transformation $Y \mapsto Y^\lambda$. (Ch. 9)

(B) Interpretation of regression models with categorical variables, or $\log Y \sim \beta_0 + \beta_1 X + \epsilon$
(See homework assignment.)

(C) Model Selection (Ch. 10)
Adjusted- R^2 , AIC, BIC, C_p .

Previously: F-test.

Forward & Backward & Stepwise & "Best Subsets".

② (A) Goodness-of-Fit tests.

(B) Generalized Additive Models (9.6)

(C) Roughness Penalty.

(D) "Broken Stick" Regression.

& more from earlier chapters, e.g.
"confidence ellipse" for $\hat{\beta}_0, \hat{\beta}_1$.

Note that the exam will have a multiple choice section. You can tell your favorite LLM,

"make up some multiple choice questions about _____"
and tell me the correct answers, giving a reason."

Box-Cox procedure: look for max in graph.

Select closest interpretable value of λ to max.

if $\lambda = 0$, answer is " $\log Y$ ".

Strategy for studying for questions of type 1B:

Feed the homework to your LLM.

Tell it to make up some similar questions.

For questions of type 1C:

You need to know what the words mean and what the procedures do.

Example: We have 5 explanatory variables and select a model with 3 vars using "forward selection" and "Backwards Elimination". Are the models necessarily the same?

Answer: No.

Note that ① has to be a relatively small part of the exam.

The most important part of the exam will be stuff from Ch. 6-7.

- Diagnostic Plots
- Attenuation Bias & Omitted Var Bias.
- What is a high-leverage point?

How does that differ from an influential point or an outlier?

Note: whether a point is a high-leverage point does NOT depend on $\hat{y}$.

high-leverage points are extreme in the X (predictor) space. (p. 83).

$$\frac{\partial \hat{y}_i}{\partial y_i} = h_{ii} \text{ is the "justification".}$$

Also fundamental: what are the $\hat{y}$ diagnostic plots that R shows you and what do they diagnose?

Note: mostly these plots are diagnosing whether the assumptions of linear regression are satisfied.

So the residuals vs. fitted plot is ~~the~~ testing assumptions about the variance of errors.

Hardest question: Some variation on the formula of p. 100.