Quiz 3 (web)

Due Oct 11 at 10:30am **Points** 24 **Questions** 7
Available Oct 6 at 8am - Oct 11 at 10:30am **Time Limit** 60 Minutes **Allowed Attempts** 2

Instructions

This quiz is open book, open notes, "open R/Python". The expected duration is 60 minutes. Two attempts are allowed. If both attempts are taken, the score for the second attempt will "overwrite" that from the first attempt (regardless if it is higher or lower). Even though the quiz has 24 points, it will be graded out of 20 points.

You are allowed to use any of the class materials from our SML class, but no other materials (no internet browsing or communication with other parties online/offline).

Even if a question is asking for a numerical value or True/False answer, in order to receive full credit (if your "guess" is correct) or partial credit (if appropriate, if your "guess" is incorrect), please provide your rationale as comments in the uploaded file requested at the end of the quiz.

If you solved some of the questions analytically, provide a clear scan of your work (possibly, with a smartphone app, not a raw photo), in pdf format. In case of multiple files, create a folder, place all files in it, make an archive (zip or rar) and submit.

Only if, for whatever reason, the form at the end of the quiz is not working, please submit the supplementary files via Dropbox file request link,

tinyurl.com/nbliznyuk-submit-files

Take the Quiz Again

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	13 minutes	12 out of 24 *

* Some questions not yet graded

Correct answers will be available on Oct 11 at 10:30am.

Score for this attempt: **12** out of 24 *
Submitted Oct 7 at 8:51pm
This attempt took 13 minutes.

Question 1 4 / 4 pts

Q1

Consider the standard multiple linear regression setup with quantitative predictors (no qualitative predictors, no interactions, no polynomial regression). Let X be an n-by-p design matrix with n > p (predictors are in the columns; all predictors are quantitative); assume that the column of ones is the first column of X. Let Y be the corresponding vector of responses (of length n). The goal is to estimate the vector of coefficients beta by solving the ordinary least squares (OLS) problem (this is the same as the "least squares problem" presented in class).

If the columns of X are linearly dependent, then <multiple choice question>

Make sure to provide your rationale/brief justification as upload at the end of the quiz.

If you are unable to provide the answer in the "general p" case,

assume that p=3 and that X[,3] = a1*X[,1] + a2*X[,2] for nonzero constants a1 and a2.

- ○ the OLS problem has no solution.
- ○ the OLS problem has a unique solution.
- ● the OLS problem has multiple solutions (in fact, infinitely many).

  The sum of squared errors/residuals objective function is continuous and is bounded below (by 0). Hence it has a global minimum. Let bhat be one such solution. Since the columns of X are linearly dependent, we can find a non-zero vector c such that X*c = 0. Then (bhat + lambda*c) is another global solution, where lambda is any real number. Hence, infinitely many solutions exist.
- ○ the question cannot be answered since not enough information was provided.

Incorrect

Question 2 0 / 4 pts

Q2a

Solve exactly one question out of Q2a or Q2b. If you solve both, only Q2a will be graded.

Let X be an n-by-p design matrix with n > p (predictors are in the columns; all predictors are quantitative).

True/False: if predictors/covariates are linearly independent, then they are uncorrelated. Put differently (but equivalently): if columns of X are linearly independent, then the correlations between pairs of columns of X are zero.

(To get full credit, please include your brief justification (or a code example) in the supplementary file for Q7.)

- ● True
- ○ False

Last Attempt Details:

Time:	13 minutes
Current Score:	12 out of 24 *
Kept Score:	12 out of 24

* Some questions not yet graded

1 More Attempt available

Take the Quiz Again

(Will keep the latest of all your scores)

Incorrect

Question 3

0 / 4 pts

Q2b

Solve exactly one question out of Q2a or Q2b. If you solve both, only Q2a will be graded.

Let X be an n-by-p design matrix with n > p (predictors are in the columns; all predictors are quantitative). Assume that the column of ones is included.

True/False: if columns of X are linearly dependent, then the correlations between one or more pairs of columns of X is 1 or -1.

(To get full credit, please include your brief justification (or a code example) in the supplementary file for Q7.)

- (selected) True
- False

Question 4

4 / 4 pts

Q3

True/False: in multiple linear regression, the coefficient of determination R^2 is equal to the squared sample correlation coefficient between the response vector Y and the vector of fitted values $\widehat{Y} = X\widehat{\beta}$.

(You do not need to provide a theoretical/analytical answer here. Rather, please use your "experience" from the three test problems in 3.1-3.3 (regression with an intercept) of homework 3 (based on the models fitted by the lm function).

- (selected) True

If the fit0 is the object produced by lm, then

yhat = fit0$fitted.values; # extract fitted values

cor(y, yhat)^2 # examine the squared correlation coefficient and compare it with the (unadjusted) R^2 reported by

summary(fit0)

- False

Unanswered

Question 5

Not yet graded / 4 pts

Q4

Consider the standard multiple linear regression setup with quantitative predictors (no qualitative predictors, no interactions, no polynomial regression). Let X be an n-by-p design matrix with n > p (predictors are in the columns; all predictors are quantitative); assume that the column of ones is the first column of X.

Suppose you have access to a function myRsq(Y ~ C1 + ... Ck) that uses the same syntax as the lm() function but outputs only the coefficient of determination R^2 but no other outputs that lm produces.

Briefly explain how to use myRsq function to determine the (column) rank of X.

If you have difficulties, first explain how to use myRsq to identify "redundant" columns of X (i.e., those that may be represented as a linear combination of the remaining columns of X).

Your Answer:

Solution (in pseudocode):

line 1: call out = myRsq(X[,i] ~ X[,-i]) for i=2,...,# of columns of X until out = 1;

line 2: if out = 1, remove X[,i] from X and return to line 1

line 3: else (if out never becomes 1) terminate; the rank of X is the number of columns left or, equivalently, p minus the number of columns removed.

Question 6

4 / 4 pts

Q5

Consider the standard simple linear regression setup with a (single) quantitative predictor (no qualitative predictors, no interactions, no polynomial regression). Let X be an n-by-2 design matrix with n > 2 (predictors are in the columns; all predictors are quantitative); assume that the column of ones is the first column of X and that the columns of X are linearly independent. Let Y be the corresponding vector of responses (of length n). The goal is to estimate the vector of coefficients beta (i) by solving the ordinary least squares (OLS) problem and (ii) by the method of maximum likelihood estimation (MLE) under the model that
$Y_i \; = \; x_i^T \beta + \epsilon_i, \; where \; \epsilon_i \; are \; iid \; Normal \; (0, \sigma^2) \; .$

True/False: the estimate of beta produced by OLS matches that produced by MLE. (Please answer based on your hw3 problem 4 experience.)

- (selected) True

You explored this in problems 4.1-4.2 of hw3.

The result holds generally in multiple linear regression (so long as there is no "multicollinearity"/redundant predictors/design matrix is of full column rank).

False

Unanswered

Question 7

Not yet graded / 0 pts

Upload your supplementary file with code and additional explanations here.

Do NOT upload via Dropbox if you are able to upload through this question.

Quiz Score: **12** out of 24