

✓

Published

[Preview \(https://gatech.instructure.com/courses/189786/quizzes/253611/take?preview=1\)](https://gatech.instructure.com/courses/189786/quizzes/253611/take?preview=1)

[Edit \(https://gatech.instructure.com/courses/189786/quizzes/253611/edit\)](https://gatech.instructure.com/courses/189786/quizzes/253611/edit)

Practice Final Exam Part1 (Closed Book)

⚠ This is a preview of the draft version of the quiz

Important Note: This is the practice final exam, this is NOT the closed-book section of the ACTUAL final exam (Part 1).

Please slowly show both sides of each piece of scratch paper to the camera before beginning. Then show both sides of your formula sheet(s).

Each question in the practice closed book portion of the exam has 1 correct answer. You may use Honorlock's on-screen calculator, your computer's calculator app or a physical calculator but you may not use your phone calculator. Phones are prohibited in both exam sections.

It is your responsibility to keep track of your time and submit before the time limit.

The Honorlock support team is available 24/7. The Honorlock support agents provide best-in-class support, and each one is trained to offer quick and consistent assistance. Whether you're testing at two in the morning or on a weekend evening or over a holiday, you can call at any time and get help.

Quiz Type	Practice Quiz
Points	26
Shuffle Answers	No
Time Limit	90 Minutes
Multiple Attempts	Yes
Score to Keep	Highest
Attempts	Unlimited
View Responses	Always
Show Correct Answers	Immediately
Access Code	HL_NO_EDIT_2rldKY2g
One Question at a Time	No
Require Respondus LockDown Browser	No
Required to View Quiz Results	No
Webcam Required	No

Due	For	Available from	Until
-	Everyone	Jul 19 at 8am	-

[Preview \(https://gatech.instructure.com/courses/189786/quizzes/253611/take?preview=1\)](https://gatech.instructure.com/courses/189786/quizzes/253611/take?preview=1)

Submitted Jul 28 at 3:52pm

Unanswered

Question 1

0 / 1 pts

If the constant variance assumption does not hold in multiple linear regression, we apply a Box-Cox transformation to the predicting variables.

True

False

Correct Answer

False

3.11. Assumptions and Diagnostics

If constant variance or normality assumptions do not hold, we apply a Box-Cox transformation to the response variable.

Unanswered

Question 2

0 / 1 pts

Multicollinearity in multiple linear regression means that the columns in the design matrix are linearly independent.

☐ True

Correct Answer

☐ False**3.13. Model Evaluation and Multicollinearity**

Multicollinearity means there is a dependency between predicting variables which would equate to the columns in the design matrix.

Unanswered

Question 3

0 / 1 pts

In logistic regression, R^2 can be used as a measure of explained variation in the response variable.

☐ True

Correct Answer

☐ False**4.5. Statistical Inference**

When R-squared is used as explained variability: The denominator of the ratio can be thought of as the total variability in the dependent variable, or how much y varies from its mean. The numerator of the ratio can be thought of as the variability in the dependent variable that is predicted by the model. Thus, this ratio is the proportion of the total variability explained by the model.

In logistic regression, the response variable is binary. So the classic R-squared does not make sense in measuring explained variation.

Unanswered

Question 4

0 / 1 pts

The interpretation of the regression coefficients is the same for logistic regression and poisson regression.

☐ True

Correct Answer

☐ False

Topic 4.4 Lesson 15: Poisson Regression: Model Description & Estimation

Interpretation of the regression coefficients of Poisson regression is in terms of log ratio of the rate.

Topic 4.2 Lesson 5: Logistic Regression: Statistical Inference

Interpretation of the regression coefficients of Logistic regression is in terms of log odds.

Unanswered

Question 5

0 / 1 pts

A poisson regression model fit to a dataset with a small sample size will have a hypothesis testing procedure with more Type I errors than expected.

Correct Answer

☐ True☐ False**Topic 4.5 Lesson 17: Poisson Regression: Statistical Inference**

In Poisson regression, if the sample size small, the statistical inference is not reliable. Thus, the hypothesis testing procedure will have a probability of type I error larger than the significance level.

Unanswered

Question 6

0 / 1 pts

LASSO regression will *a/ways* select the same number or more predicting variables than Ridge and Elastic-Net regression.

☐ True

Correct Answer

☐ False**5.7. Regularized Regression: Approaches**

Because LASSO can eliminate predicting variables using the penalty while Ridge and Elastic Net retain coefficients, LASSO will have the same number or LESS predicting variables.

Unanswered

Question 7

0 / 1 pts

Forward stepwise variable selection starts with the simpler model and selects the predicting variable that increases the R-squared the most, unless the R-squared cannot be increased any further by adding variables.

☐ True

Correct Answer

☐ False

5.12. Variable Selection

R-squared is not compared during stepwise variable selection. Variables are selected if they *reduce* the *AIC* or *BIC* of a model.

Unanswered

Question 8

0 / 1 pts

In the balance of Bias-Variance tradeoff, adding variables to our model tends to increase our variance and decrease our bias.

Correct Answer

☐ True☐ False**5.3. Prediction Risk Estimate**

Adding more variables will increase the variability and possibly induce multicollinearity. Adding more variables also reduces the bias in the model since it has an additional predictor to conform to which keeps the model from favoring one of the original predictors.

Unanswered

Question 9

0 / 1 pts

In stepwise regression, we accept/remove variables that produce larger AICs or BICs.

☐ True

Correct Answer

☐ False**5.4. Model Search**

We desire the model that has the **smallest** AIC or BIC.

Unanswered

Question 10

0 / 1 pts

BIC variable selection criteria favors simpler models.

Correct Answer

☐ True☐ False

5.3. Prediction Risk Estimate

BIC penalizes complexity more than other approaches.

Unanswered

Question 11

0 / 1 pts

For a linear regression under normality, the variance used in the Mallows's Cp penalty is the true variance, not an estimate of variance.

☐ True

Correct Answer

☐ False**5.3. Prediction Risk Estimate**

For linear regression under normality, the variance used in the Mallows's Cp penalty is the estimated variance from the full model.

Unanswered

Question 12

0 / 1 pts

Stepwise regression is a greedy search algorithm that is not guaranteed to find the model with the best score.

☐ True

Correct Answer

☐ False**5.4. Model Search**

"Stepwise regression is a greedy search algorithm. It does not guarantee to find the model with the best score."

Unanswered

Question 13

0 / 1 pts

Ridge regression corrects for the impact of multicollinearity by re-weighting the regression coefficients.

☐ True

Correct Answer

☐ False

5.7. Regularized Regression: Approaches

Ridge regression has been developed to correct for the impact of multicollinearity. If there is multicollinearity in the model, all predicting variables are considered to be included in the model but ridge regression will allow for re-weighting the regression coefficients in a way that those corresponding to correlated predictor variables share their explanatory power and thus minimizing the impact of multicollinearity on the estimation and statistical inference of the regression coefficients.

Unanswered

Question 14

0 / 1 pts

Elastic Net often underperforms LASSO regression in terms of prediction accuracy because it considers both L1 and L2 penalties together.

☐ True

Correct Answer

☐ False**5.7. Regularized Regression: Approaches**

Elastic net often outperforms the lasso in terms of prediction accuracy. The difference between lasso and elastic net is the addition of a penalty just like the one used in ridge regression. By considering both penalties, L1 and L2 together, we have the advantages of both lasso and ridge regression.

Unanswered

Question 15

0 / 1 pts

Random sampling is computationally less expensive than the K-fold cross validation.

☐ True

Correct Answer

☐ False**4.9. Classification**

Random sampling is computationally more expensive than the K-fold cross validation, with no clear advantage in terms of the accuracy of the estimation classification error rate. K fold cross validation is preferred at least from a computation standpoint.

Part 2: Multiple Choice

An experiment was conducted to determine the effect of gamma radiation on the numbers of chromosomal abnormalities observed in cells. A multiple linear regression model was fitted to estimate the effect of the number of cells, amount of the radiation dose (Grays), and the rate of the radiation dose (Grays/hour) on the number of chromosomal abnormalities observed. The data frame has 27 observations.

Here is the model summary and Cook's Distance plot.

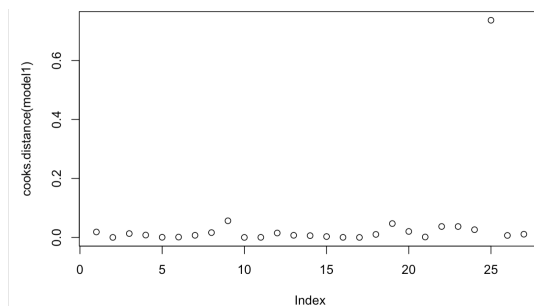
Coefficient	Estimate	SE	t-value	Pr(> t)
(Intercept)	-74.15392	42.24544	-1.755	0.092518
cells	0.06871	0.02196	3.129	0.004709**
doseamt	41.33160	9.13907	4.523	0.000153***
doserate	20.28402	8.29071	2.447	0.022482*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 54.05 on **X** degrees of freedom

Multiple R-squared: 0.5213, Adjusted R-squared: 0.4588

F-statistic: 8.348 on **Y** and **X** DF, p-value: 0.0006183



Unanswered

Question 16

0 / 1 pts

Suppose you wanted to test if the coefficient for *doseamt* is equal to 50. What t-value would you use for this test?

☐ 1.54

☐ -0.948

☐ 0.692

☐ -0.882

Correct Answer

$$t\text{-stat} = (\text{Coefficient} - \text{comparison value}) / \text{SE} = (41.33160 - 50) / 9.13907 = -0.948$$

Unanswered

Question 17

0 / 1 pts

Based on the cook's distance plot and the rule of thumb $4/n$, how many data points may be outliers potentially influencing the model?

☐ None

Correct Answer

☐ One☐ Two☐ Three**3.3 .11: Assumptions and Diagnostics**

There are 27 observations hence the rule of thumb is $4/n = 4/27 = 0.15$. From the graph there is only one observation with a Cook's distance measurement >0.15 .

Unanswered

Question 18

0 / 1 pts

How does an increase in 1 unit in *doserate* affect the expected number of chromosome abnormalities, given that the other predictors in the model are held constant?

☐ Increase of 8.291☐ Decrease of 41.331

Correct Answer

☐ Increase of 20.284☐ Decrease of 9.134**Topic 3.2 Lesson 6: Inference for Regression Parameters**

The estimated coefficient for *doserate* is 20.284.

Unanswered

Question 19

0 / 1 pts

For an F-test of overall significance of the regression model, what degrees of freedom would be used?

☐ 3, 24☐ 2, 26

Correct Answer

☐ 3, 23☐ 3,26

$$F(p, n-p-1) = F(3, 23)$$

$n = \# \text{ of observations} = 27$

$p = \# \text{ of predicting variables} = 3$

$p = 3$

$n-p-1 = 27-3-1=23$

You were hired to consult on a study for the attendance behavior of high school students at two different schools. The data set you were given contains for each 316 students: the number of days he/she was absent in an academic year (*daysabs*), his/her math scores (*math*), his/her language arts scores (*langarts*), and whether the student is male or female (1 = *male*, 0 = *female*).

A Poisson regression model was fitted to evaluate the relationship between the number of days of absence in an academic year and all the predictors. The R output for the model summary is as follows:

Coefficient	Estimate	SE	z value	$Pr(> z)$
(Intercept)	2.687666	0.072651	36.994	<2e-16
math	-0.003523	0.001821	-1.934	0.0531
langarts	-0.012152	0.001835	-6.623	3.52e-11
male	-0.400921	0.048412	-8.281	<2e-16

Also, assume the average language arts scores (across all students) is 50, and the average math scores (across all students) is 45.5.

Unanswered

Question 20

0 / 1 pts

For students with average math and language arts scores, how many more days on average are female students absent compared to their male counterparts?

☐ 4.8545

☐ 3.5729

Correct Answer

☐ 2.2525

☐ 0.6697
Topic 4.4 Lesson 15: Poisson Regression: Model Description & Estimation

$$E(Y|X_{\text{math}} = 45.5, X_{\text{langarts}} = 50, X_{\text{male}} = 0) = \exp(2.68766 - 0.003523 * (45.5) - 0.012152 * (50) - 0.400921 * (0)) = 6.$$

$$E(Y|X_{\text{math}} = 45.5, X_{\text{langarts}} = 50, X_{\text{male}} = 1) = \exp(2.68766 - 0.003523 * (45.5) - 0.012152 * (50) - 0.400921 * (1)) = 4.$$

$$E(Y|X_{\text{math}} = 45.5, X_{\text{langarts}} = 50, X_{\text{male}} = 0) - E(Y|X_{\text{math}} = 45.5, X_{\text{langarts}} = 50, X_{\text{male}} = 1) = 2.252423$$

Unanswered

Question 21

0 / 1 pts

How does an increase in 1 unit in *langarts* affect the expected number of days missed, given that the other predictors in the model are held constant?

- ☐ Increase by 0.012152 days
- ☐ Increase by 0.9879 days
- ☐ Increase by 1.22%
- ☐ Decrease by 1.21%

Correct Answer

Topic 4.4 Lesson 15: Poisson Regression: Model Description & Estimation

The estimated coefficient for *langarts* is -0.012152. A one unit increase in *langarts* gives us $\exp(-0.012152) = 0.9879215$. It is interpreted as the expected number of days missed decreasing by 1.21% (1-0.9879). Hence, given that the other predictors in the model are held fixed, one unit increase in *langarts* results in the expected number of days missed decreasing by 1.21%.

Unanswered

Question 22

0 / 1 pts

What is the expected number of days missed for a female student with a *langarts* of 48 and a math score of 50 based on the model?

- ☐ 6.8773
- ☐ 1.9106
- ☐ 6.6363
- ☐ 4.5251

Correct Answer

Topic 4.4 Lesson 15: Poisson Regression: Model Description & Estimation

$$E(Y|X_{math} = 50, X_{langarts} = 48, X_{male} = 0) = \exp(2.68766 - 0.003523 * (50) - 0.012152 * (48) - 0.400921 * (0)) = 6.877$$

Unanswered

Question 23

0 / 1 pts

How many regression coefficients including the intercept are statistically significant at the significance level 0.05?

- ☐ All
- ☐ None
- ☐ Two

Correct Answer

☐ Three
Topic 4.4 Lesson 15: Poisson Regression: Model Description & Estimation

The coefficients of *intercept*, *langarts* and *male* are statistically significant at $\alpha = 0.05$.

Unanswered

Question 24

0 / 1 pts

The approximated distribution of the deviance test statistic (D) is ____ with ____ degrees of freedom.

☐ Normal, 315

Correct Answer

☐ Chi-Squared, 312

☐ Chi-Squared, 315

☐ t, 312
Topic 4.5 Lesson 20: Poisson Regression: Goodness of Fit Assessment Data Example

Hypothesis Testing Procedure:

Under null hypothesis, the approximated distribution of the deviance test statistic $D \sim \chi^2_{df}$ with $df = n - p - 1 = 316 - 3 - 1 = 312$, where $n = \#$ of observations, $p = \#$ of predicting variables.

Unanswered

Question 25

0 / 1 pts

The ____ penalty is referred to as ____ regression which performs variable selection by ____.

☐ L1, ridge, forcing beta coefficients to zero

☐ L2, ridge, reducing effectiveness of beta coefficients

Correct Answer

☐ L1, LASSO, forcing beta coefficients to zero

☐ L2, LASSO, reducing effectiveness of beta coefficients
5.7. Regularized Regression: Approaches

Ridge regression does NOT perform variable selection because the variables remain in the model. LASSO performs variable selection because coefficients will be forced to 0, which removes the variable from the model.

LASSO represents the L1 penalty.

Unanswered

Question 26

0 / 1 pts

If there are 5 quantitative predicting variables, how many different models can be created?

☐ 10☐ 16☐ 25

Correct Answer

☐ 32

A model with p predictors has 2^p different model combinations.