ST 563 601 – SPRING 2025 – POST
Exam #1 (Paper)

Student's Name: **Ahmed, Ameen** AZ

Date of Exam: Thursday, February 6, 2025 - Friday, February 7, 2025
Time Limit: ~~75 minutes~~ 225 minutes **(3x time)**
Appointment Start Time: 12:30pm, Monday, Feb 24, 2025
Appointment End Time: 4:15pm, Monday, Feb 24, 2025
Allowed Materials: None (closed book & closed notes)
Student Accommodations: Individual Testing Room, Extended Time (3T), Computer for Paper-Based exams

Student – NC State University Pack Pledge

I, Ameen Ahmed have neither given nor received unauthorized aid on this exam or assignment. I have read the instructions and acknowledge that this is the correct exam.
Student's Printed Name

Thomas Martin for Ameen Ahmed — 2/24/2025
Student Signature — *Date*

Exam must be turned in by: 4:50 — AA
Exam End Time — *Student's Initial Agreement*
4:52

NOTE: Failure to turn in exam on time may result in penalties at the instructor's discretion.

1.

Statistical inference is when you have a model and you are trying to do inference on it. In this case doing confidence intervals and hypothesis testing. A real world example is there a statistical significance between Drug A and Drug B. The null and alternative hypotheses is:
H_0: mu_1 = mu_2

H_a: mu_1 \neq mu_2

Predictive modeling is when you are doing regression analysis, statistical learning / machine learning. A real world example is predicting how a team would perform on the next season based on the previous season. An example model would be:

\hat{y} = \beta_0 + \beta_1x_1 + \beta_2x_2 + ... + \beta_px_p

Pattern finding is when the model does not have a response variable.

2.

a.

A very flexible model would have low bias. When the bias decreases, the variance increases. This is known as the bias variance trade off.

b.

A very flexible model is when the variance is high. As the variance increases, the bias would decrease. This is known as the bias variance trade off.

c.

A very flexible model would have a low training error. When the model has a low training error, it would have a high test error. This would lead to the model overfitting.

d.

A very flexible model would have a high test error because when the model memorizes the training data very well, it would perform poorly on unseen data, leading to overfitting.

3.

A tuning parameter has values like MSE and MAE. We create a tuning grid that has tuning parameters on it. Which ever tuning parameter has the lowest MSE, we would use that tuning parameter to fit the model on the entire training set.

A regular parameter is something like regression coefficients or least square estimates. These are used to define a regression model.

5.

The four model selections are forward selection, backwards selection, stepwise procedure, and best subset selection.

6.

a. True

b. False

c. True

d. False

e. False

f. True

g. False

h. True

7.

1. We first create a 80% training set and 20% test set.

2. We create a tuning grid of tuning parameters. Lambda for lasso regression model and k for KNN regression model.

3. If we are doing a 10 fold cross validation, we partition the first 9 folds for the training set and we do the 10^{th} fold on the test set.

4. We compute the MSE value for the first 10^{th} fold.

5. We partition the second 9 folds on the training set and the second 10^{th} fold on the test set.

6. We compute the MSE for the second 10^{th} fold.

7. We repeat this process 10 times by partitioning 9 folds on the training set and the 10^{th} fold on the test set and computing the MSE for the remaining 8 folds.

8. We compute the average MSE for lambda for the lasso regression model and k for KNN regression model.

9. We repeat this process for the remaining lambda values for lasso regression model and k for KNN regression model by partitioning 9 folds for the training set and the 10^{th} fold on the test set and computing the MSE 10 times and taking the average MSE.

10. We select the lambda value for the lasso regression model and k for the KNN regression model that has the lowest MSE value.

11. We then fit the lasso regression model and KNN regression model on the entire training set.

12. We then compute the RMSE value for both the lasso regression model and the KNN regression model on the test set.

13. Which ever model has the lowest RMSE value, would be the best model.

14. Then we fit that best model on the entire dataset.

8.

a. The benefits of fitting a ridge regression model compared to an OLS regression model is ridge regression is a L2 regularization model. It fits better on high dimensional data than OLS regression model. Ridge regression model works very well with multicollinearity. It prevents the model from overfitting. It shrinks the parameters close to 0, but keeps all predictors.

b. for a large tuning parameter, the coefficient estimates gets close to 0, but keeps all predictors. For a small tuning parameter that is near 0, the coefficient estimates remains unchanged in the model. In otherwards the coefficient estimates stays in the model.

9.

a.

$\hat{y} = 25.293 - 19.780x_1 - 31.760x_2 + 2.846x_3 - 0.024x_3^2 + 2.024x_1x_3 + 2.230x_2x_3 - 0.025x_1(x_3)^2 - 0.032x_2(x_3)^2$

marital_status is the indicator variable.

b.

The use of the t-value is finding how many standard errors above and below the regression coefficients. It is also used for finding the p-value.

c.

$\hat{y} = 25.293 - 19.780 + 2.846 * 30 - 0.024 * (30)^2 + 2.024 * 30 - 0.025 * (30)^2$

d.

$\hat{y} = 25.293 + 2.846 * 30 - 0.024 * 30^2$

f.

The response variable (wage), would depend on the interaction term between marital_status and age and the interaction term between marital_status and age^2. As age increases or decreases, it is directly going to effect marital_status. It makes it more difficult to interpret the predicted value by including interaction terms in the model.

G.

i.

H0: \beta_1 = beta_2 = beta_3 = beta_4 = beta_5 = beta_6 = beta_7 = beta_8

H_a At least one of the betas are not 0

ii.

For an F-test, we are testing the entire model by including all predictors in the model, where as a t-test, we are testing the predictors separately.

h.

We would look at the plot residuals vs predicted values to see if there is a constant variance.

Exam 1

Please write your answers below each question. You should not have access nor use any materials during this exam.

A reminder that, by taking this exam, you are required to uphold the NC State honor pledge:

"I have neither given nor received unauthorized aid on this test or assignment."

1. In the statistical learning paradigm, we discussed three major goals: statistical inference, predictive modeling, and pattern finding.

   Give a brief real world example for each of these goals. Specify a possible model or method we discussed in class that would help answer the question from each real world example.

- Statistical Inference (4 pts)

- Predictive Modeling (4 pts)

- Pattern Finding (4 pts)

2. Consider having models characterized by flexibility with the scale going from not very flexible to very flexible.
    a. What type of relationship between flexibilty and squared bias would we expect? Why? (4 pts)
    b. What type of relationship between flexibilty and variance would we expect? Why? (4 pts)
    c. What type of relationship between flexibilty and training error would we expect? Why? (4 pts)
    d. What type of relationship between flexibilty and test error would we expect? Why? (4 pts)
3. What is a tuning parameter or hyperparameter? How does this differ from a ‘regular’ parameter in a parametric model? (4 pts)

5. In the multiple linear regression setting, we discussed a number of model selection methods. State four model selection methods that can be used in the $p > n$ situation. (4 pts)

6. State true or false (no need to explain). (3 pts each)
   a. Ordinary least squares performs variable selection.
   b. Ordinary least squares performs shrinkage of coefficient estimates.
   c. Best subset selection performs variable selection.
   d. Best subset selection performs shrinkage of coefficient estimates.
   e. Ridge Regression performs variable selection.
   f. Ridge Regression performs shrinkage of coefficient estimates.
   g. LASSO performs variable selection.
   h. LASSO performs shrinkage of coefficient estimates.

7. Suppose we have a large data set where we want to perform a regression task. We want to determine the best overall model between a LASSO model and a kNN regression model. We want to use a train test split and compare the best kNN and LASSO model on the test set. We wish to determine the appropriate tuning parameters on the training set only using cross-validation. Fully outline the process for splitting the data, tuning, comparing, and fitting a final overall best model. (10 pts)

8. Consider the Ridge Regression procedure for fitting a multiple linear regression model. With this model we minimize the following criterion (recall $\lambda \geq 0$):

$$\sum_{i} \left(Y_i - \beta_0 - X_{i1}\beta_1 - \cdots - X_{ip}\beta_p\right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

a. What are the benefits of fitting a Ridge Regression model as compared to an ordinary least squares model? (4 pts)

b. What happens to our coefficient estimates for a 'large' value of the tuning parameter? What happens for a tuning parameter value near 0? (4 pts)

9. Suppose we fit a multiple linear regression model to data about how much people earn. Our response variable is the `wage` (in 1000's of dollars) and our predictors are `marital_status` (`married`, `never_married`, or `divorced`), and `age`.

   We fit a linear and quadratic term for age and include an interaction between `marital_status` and `age` and an interaction between `marital_status` and `age` squared in the model. Output for the model is given below.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	25.293	38.116	0.664	0.507
marital_statusmarried	-19.780	40.405	-0.490	0.624
marital_statusnever_married	-31.760	40.992	-0.775	0.439
age	2.846	1.611	1.767	0.077
I(age^2)	-0.024	0.017	-1.470	0.142
marital_statusmarried:age	2.024	1.716	1.179	0.238
marital_statusnever_married:age	2.230	1.820	1.225	0.221
marital_statusmarried:I(age^2)	-0.025	0.018	-1.412	0.158
marital_statusnever_married:I(age^2)	-0.032	0.020	-1.607	0.108

   a. Write down the fitted equation for $\hat{y}$. Define any indicator variables as needed. (4 pts)

   b. One column of the output represents the t-value or t-statistic. What is the usefulness of this t-value? (2 pts)

c. Write down the form of a predicted value for somone that is `married` and has an age of 30. No need to simplify. (2 pts)

d. Write down the form of a predicted value for somone that is `divorced` and has an age of 30. No need to simplify. (2 pts)

f. Conceptually, what does including an interaction between `marital_status` and age and an interaction between `marital_status` and age squared do to our model as compared to a model without those interactions (that still includes a main effect for `marital_status` and a linear and quadratic term for age)? (3 pts)

g. The F-statistic for the global model test is 46.26 on 8 numerator and 2991 denominator degrees of freedom. The p-value for the test is very close to zero.

    i. Write down the null and alternative hypotheses for this global test. (3 pts)

    ii. We see a significant global test but none of the coefficient tests are significant. What do you think could be causing this issue? (3 pts)

h. What type of plot might we look at to investigate the homogenous error variance (i.e. the assumption of equal error variance)? (3 pts)