**STAT 501 – Final Exam – Fall 2015**

**Instructions:** Use Word to type your answers within this document. Then, submit your answers in the appropriate dropbox in ANGEL by the due date and **within 3 hours of downloading the exam**. The point distribution is located next to each question.

1. **(10x2 = 20 points)** State which of the following statements are true and which are false. For the statements that are false, explain why they are false.
2. In weighted least squares, the “ideal” weights to use, in theory, will equalσi, the population standard deviation of the ith error term.

False. We define the reciprocal of each variance, *σi2*, as the weight, *wi*=1/*σ*i2

1. Weighted least squares is used to adjust for autocorrelated errors.

False. The method of weighted least squares can be used when the ordinary least squares assumption of constant variance in the errors is violated.

1. In the context of time series, suppose we wish to estimate the relationship between the variables and In these transformations, r represents the correlation between εt and εt-1, where these ε-values are errors from the regression between yt and xt.

~~False. A multiple (time series) regression model can be written as:~~

*~~yt~~*~~=~~**~~X~~***~~tβ~~*~~+~~*~~ϵt~~*

~~Further,~~ *~~ϵt~~*~~=~~*~~ρϵ~~~~t~~*~~−1~~~~+~~*~~ω~~~~t~~ ~~ρ is the~~* ~~is called the~~ **~~autocorrelation parameter for the error~~**

~~In this context the yt = y\* and xt = x\* as defined above in the question.~~

1. A very large value of Cook's Di for a data point indicates that if the point is removed, the overall set of fits for the regression will likely change.

True. Di directly summarizes how much all of the fitted values change when the ith observation is deleted.

1. Suppose that X1 and X2 have been used to predict Y; then, X3 is added to the model. The value of R2 could possibly decrease due to adding X3.

False. We know *R*2=*SSR/SSTO* = 1− *SSE/SSTO*

Also we know that SSR only increases as we add more variables. In worst case it may remain same or very slightly increase and therefore R2 will never decrease upon addition of a predictor.

1. In binary logistic regression, the success event logit function (log odds of success) is the negative of the failure event logit function (log odds of failure).

True

1. In simple linear regression analysis, the width of a prediction interval for a future response of *Y* based on a single predictor *X* does not vary with the value of *X*.

False. The prediction interval is given by:



We can see that the value depends on xh

1. The error terms in a regression with autoregressive AR(1) errors have zero mean.

True

1. In model selection, the *MSE* (or *S*) criterion minimizes confidence/prediction interval widths.

True

1. In model selection, the *PRESS* criterion evaluates model unbiasedness.

False. PRESS is calculated by omitting each observation individually and then the remaining *n* - 1 observations are used to calculate a regression equation which is used to predict the value of the omitted response value. We then calculate the *i*th PRESS residual as the difference *yi*−*y*^*i*(*i*). Then, the formula for PRESS is given by



In general, the smaller the PRESS value, the better the model's predictive ability.

1. **(3+4+3+3 = 13 points)** The following ANOVA table is abstracted from a regression run for the model: *Y* = *β*0 + *β*1 *X*1 + *β*2 *X*2 + *β*3 *X*3 + … + *β*10 *X*10 + ε.

|  |  |  |
| --- | --- | --- |
| ***Source*** | ***DF*** | ***SS*** |
| ***Regression*** | 10 | 110.53 |
| ***Residual Error*** | 28 | 39.47 |
| ***Total*** | 38 | 150 |

|  |  |  |
| --- | --- | --- |
| ***Source*** | ***DF*** | ***Seq SS*** |
| ***X1*** | 1 | 0.10 |
| ***X2*** | 1 | 40 |
| ***X3*** | 1 | 1.0 |
| ***X4*** | 1 | 55 |
| ***X5*** | 1 | 2.5 |
| ***X6*** | 1 | 0.08 |
| ***X7*** | 1 | 6.5 |
| ***X8*** | 1 | 4.0 |
| ***X9*** | 1 | 0.4 |
| ***X10*** | 1 | 0.95 |

* 1. Calculate the three missing values in the upper table.
  2. For the 10-predictor model, perform a hypothesis test at significance level 0.05 to determine whether predictors *X7*, *X8*, *X9*,and *X10* are significantly linearly related to Y upon controlling for the remaining predictors *X1*-*X6* using a general linear *F* test. Write the null and alternative hypotheses, the value of the test statistic, the decision rule, and the conclusion. *[Note: an F-distribution table is provided on the last page of the exam.]*

Here we test the hypotheses:

H0 : β7 = β8 = β9 = β10 =0

HA : At least one βi ≠ 0 (for i = 7,8,9,10)

The full model is the model containing all of the possible predictors: *Y* = *β*0 + *β*1 *X*1 + *β*2 *X*2 + *β*3 *X*3 + … + *β*10 *X*10 + ε

The reduced model is: *Y* = *β*0 + *β*1 *X*1 + *β*2 *X*2 + *β*3 *X*3 + … + *β*6 *X*6 + ε

The general linear statistic: 

Since we have the been given Sequential SS, we can use:

*F\* = MSR*(*x*7,*x*8,x9,x10| *x*1,*x*2,*x*3, *x*4,*x*5,*x*6)/*MSE*(*x*1,*x*2,*x*3, …, x10)

*MSR*(*x*7,*x*8,x9,x10| *x*1,*x*2,*x*3, *x*4,*x*5,*x*6) = (6.5 + 4 + 0.4 + 0.95) / 4 = 2.9625

MSE(all) = 39.47/ 28 = 1.41

F\* = 2.9625 / 1.41 = 2.101

The P-value is the probability — if the null hypothesis were true — that we would observe a partial F-statistic more extreme than 2.101

Comparing our partial F-statistic to an F-distribution with 4 numerator degree of freedom and 28 denominator degrees of freedom =2.71

We have test statistic F\* = 2.101< critical F=2.701

We fail to reject the null hypothesis and conclude that there is insufficient evidence to conclude that Y is significantly related to X7,X8, X9 and X10 — after taking into account the other predictors.

* 1. Later it was decided to consider a regression of Y on the **first 4 predictors ONLY**. Use information from both tables to calculate adjusted R2 for the model with only the first 4 predictors?

We have SSTO = 150

With the first 4 predictors ONLY, we have SSR = 0.10+40+1.0+55 = 96.1

R2= 96.1/150 = 0.641

**SSE for 4-predictor model = 150 – (0.10+40+1+55) = 150 – 96.1 = 53.9.**

**MSE for 4-predictor model = 53.9 / (39 – 5) = 53.9/34.**

**Adj R2 for 4-predictor model = 1 – 38(53.9/34)/150 = 0.598.**

* 1. Given only the information in both tables, is it possible to test whether *X1* and *X3* can be dropped from the 4-predictor model? Give a brief argument supporting your answer. *[You do not have to do a test, even if one is possible.]*

It will not be possible since we know that the order matters for SS. In general, the number appearing in each row of the table is the sequential sum of squares for the row's variable given all the other variables that come before it in the table.So we have SS for X3 when X2 and X1 are already in the model. However to test to test whether *X1* and *X3* can be dropped from the 4-predictor model we need the sequence where x1 and x3 come before x2 and x4.

1. **(4x3 = 12 points)** Answer the following questions showing calculations and/or brief justification.
   1. If Correlation(X, Y) = 0.853, what proportion of variation in Y is *not* explained by the regression of Y on X?

R-sq = 0.853^2 = 0.728

So now we know that 72.8 % of of variation in Y is explained by the regression of Y on X. Therefore 27.2 % of variation in Y is *not* explained by the regression of Y on X

* 1. Consider regressing Y on X1 and X2. If SSR(X1 | X2) = SSR(X2 | X1), what can you say about the correlation between X1 and X2?

~~This gives us no information about the correlation. If we had known that SSR(X~~~~1~~ ~~| X~~~~2~~~~) = SSR(X~~~~1~~~~) than we could say that it leads us to believe that there is no (or little) correlation between X1 and X2.~~

**X1 and X2 are uncorrelated. The order in which X1 and X2 enter the model is not important only if they are uncorrelated!**

* 1. Consider regressing Y on 7 predictors. When all 7 predictors are included in the model, R2 = 70% and adjusted R2 = 60%. When only the 5 most significant predictors are included in the model, R2 = 65% and adjusted R2 = 62%. Of these two competing models, which one will you prefer and why?

We will prefer the model with 5 predictors because adding 2 extra predictors is only increasing the R-sq by 5%. We have to note that R-sq always increases (or worst case stays same) when adding new variables and we have to pay the cost in terms of loosing more degrees of freedom. R2-adj is a better measure for comparison since it keeps track of this cost:



Since the adjusted R2 has decreased when going from 5 – 7 predictors, we should stay with the 5 predictor model.

* 1. The figure below gives the sample PACF for a time series data of monthly sales (in thousands of dollars). Use the PACF plot to propose an appropriate time series model for yt = sales during the tth month.



The proposed model based on PACF is:

*yt*=*β*0 + *β*1*yt*−1 + *β*2*yt*−2 + *ϵt*

1. **(4+8+4+4 = 20 points)** Data from a local supermarket revealed that the deli usage of customers depends on their grocery bill and also on the time of shopping. To understand the link between these variables, a logistic regression model was fitted based on data from 890 sales records, which yielded the following.

Source DF Adj Dev Adj Mean Chi-Square P-Value

Regression 2 17.532 8.7660 17.53 0.000

bill 1 10.824 10.8241 10.82 0.001

lunch 1 5.549 5.5489 5.55 0.018

Error 887 534.290 0.6024

Total 889 551.822

Odds Ratio 95% CI

Bill 1.0760 (1.0305, 1.1236)

Odds Ratio for lunch=1 relative to lunch=0

Odds Ratio 95% CI

0.5771 (0.3651, 0.9124)

Here is the estimated probability of deli usage, bill is the amount of the grocery bill and lunch is a binary variable that equals 1 for a store visit at lunchtime and 0 for a store visit at other times.

* 1. Is their any statistical evidence that shopping time is related to the odds of deli usage, and if so, does lunchtime have a higher odds of deli usage than a visit at other times?

The p-values are less than 0.05 for both bill and lunch. Therefore there is evidence that both x-variables are useful for predicting the probability of deli usage.

Lunchtime has lower odds according to this model

* 1. Write the regression equations for estimating the logit of the probability of:
     1. deli usage in terms of both predictors bill and lunch;
     2. deli usage for lunchtime shoppers;
     3. deli usage for others;
     4. NO deli usage in terms of both predictors bill and lunch.

The logit transformation is written as:



1. deliUsageLogitProbabilityHat = -3.672+0.0733(bill)- 0.550(lunch)
2. deliUsageLogitProbabilityHat = -3.672+0.0733(bill)- 0.550 = -4.222 +0.0733(bill)
3. deliUsageLogitProbabilityHat = -3.672+0.0733(bill)
4. noDeliUsageLogitProbabilityHat = 3.672 - 0.0733(bill) + 0.550(lunch)
   1. Use your answer in (b)(ii) above to find the value of bill at which the probability of using the deli is 0.80 for a lunchtime shopper.

deliUsageLogitProbabilityHat = -4.222 +0.0733(bill)

We have been given p=0.8

So odds = 0.8/(1-0.8) = 4

Logit transformation =log(odds) = log(4) = 1.386

So we have -4.222 +0.0733(bill) = 1.386

Therefore bill = (1.386 + 4.222) / 0.0733 = 76.50750341

* 1. Write a sentence that interprets the coefficient estimate for the predictor variable bill.

The interetation of coefficient estimate for the predictor variable bill can be stated through its Odds Ratio which is derived from the coefficient estimate.

In the results given above, we see that the estimate of the odds ratio is 1.0760 for bill. This is given under Odds Ratio in the table of coefficients, standard errors and so on. The sample odds ratio was calculated as e^coeff est. The interpretation of the odds ratio is that for each increase of one unit (dollar) of the amount of the grocery bill, the predicted odds of using deli are multiplied by 1.0760

1. **(5+15 = 20 points)** The linear regression results below are for a response variable, Y, and three predictor variables, X1, X2, X3, for a dataset of ten observations. Assume that the intercept term is β0 and the coefficients for the predictor variables are β1, β2 and β3, respectively,.
   1. Interpret the t-test result provided for X2 in the Minitab output given below.

Coefficients

Term Coef SE Coef T-Value P-Value VIF

Constant 0.208 0.374 0.55 0.599

x1 0.946 0.293 3.23 0.018 1.09

x2 0.697 0.223 3.13 0.020 1.38

x3 0.301 0.506 0.60 0.573 1.31

The test and its interpretation is:

Null hypothesis H0 : β2 = 0

Alternative hypothesis Ha: β2 ≠ 0

Test Statistic t\* = Coef / SECoef = 0.697/0.223 = 3.13

Since the reported p-value is 0.020 < alpha (assumed 0.05).

Because the P-value is less than alpha, we reject the null hypothesis and conclude that β2 ≠ 0

We have sufficient evidence to reject the null hypothesis that slope β2 = 0. We conclude that there is a statistically significant linear relationship between X2 and Y

* 1. Use the following Minitab output, which gives the ANOVA tables for several models, to answer the questions below.

\* Regression Equation: y = 0.208 + 0.946 x1 + 0.697 x2 + 0.301 x3

Analysis of Variance

Source DF Adj SS Adj MS F-Value P-Value

Regression 3 31.63 10.543 11.01 0.007

Error 6 5.75 0.958

Total 9 37.38

\* Regression Equation: y = 0.191 + 0.914 x1 + 0.761 x2

Analysis of Variance

Source DF Adj SS Adj MS F-Value P-Value

Regression 2 31.29 15.645 17.99 0.002

Error 7 6.09 0.870

Total 9 37.38

\* Regression Equation: y = 0.966 + 1.172 x1

Analysis of Variance

Source DF Adj SS Adj MS F-Value P-Value

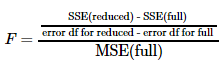
Regression 1 16.75 16.753 6.50 0.034

Error 8 20.62 2.578

Total 9 37.38

* + 1. Test, at α = 0.05, H0: β2 = β3 = 0 vs. Ha: at least one of β2, β3 are not equal to 0.

To calculate the F-statistic, we see that SSE(full) = 5.75 with error df = 6, MSE(full) = 0.958, and SSE(reduced) = 20.62 with error df = 8. Thus,



= ((20.62-5.75)/2) / 0.958 = 7.761

with df 2 and 6 (i.e., this F-statistic comes from an F2,6 distribution)

Critical F(95% alpha and df 2 and 6) = 5.14

We have test statistic F\* = 7.761 > critical F=5.14

We reject the null hypothesis and conclude that there is sufficient evidence to conclude that Y is significantly related to X2 and X3 — after taking into account the other predictors.

* + 1. Test, at α = 0.05, H0: β3 = 0 vs. Ha: β3 is not equal to 0. *[Make sure you use the ANOVA tables to answer this question, although you may use the Coefficient Table in part (a) to check your answer.*

We test the hypotheses:

H0 : β3 = 0

HA : β3 ≠ 0

The full model is the model containing all of the possible predictors: yi=(β0+β1xi1+β2xi2+β3xi3)+ϵi

The reduced model is: yi=(β0+β1xi1+β2xi2)+ϵi

The general linear statistic:

= ((SSE(X1, X2) – SSE(X1, X2, X3)) / 1) / MSE(X1, X2, X3)

= (6.09 – 5.75) / 0.958 = 0.355

with df 1 and 6 (i.e., this F-statistic comes from an F2,6 distribution)

Critical F(95% alpha and df 1 and 6) = 5.99

We have test statistic F\* = 0.355 < critical F=5.99

We fail to reject the null hypothesis and conclude that there is insufficient evidence to conclude that Y is significantly related to X3 — after taking into account the other predictors.

* + 1. Calculate R2Y,2|1 (note this is a partial R2) and interpret the value.

SSR(X2|X1) = SSE(X1) - SSE(X1, X2) = 20.62 – 6.09 = 14.53

R2Y,2|1 = SSR(X2 | X1) / SSE(X1) = 14.53 / 20.62 = 0.705

It is the proportion of variation in Y that is explained by the predictor X2 that cannot be explained by the predictor X1.

Stated differently, it measures the proportionate reduction in the variation in Y remaining after X1 is included in the model that is gained by also including X2 in the model.

1. **(5x3 = 15 points)** In an experiment, a researcher compares three different metal alloys (say, A, B, and C) used to weld pipes. For each alloy, 10 welds are made. The response variable is strength of the weld (Y). In addition to the type of alloy (A, B, or C), a quantitative predictor, diameter of weld (X), will be used in a regression model for predicting Y.
   1. Write a population regression model (with no interaction) for predicting Y using X and alloy (A, B, or C) as predictors. Clearly define any necessary indicator variables.

Use Categorical variables are:

AlloyA = 1 if alloy A is used, 0 otherwise

AlloyB = 1 if alloy B is used, 0 otherwise

E(*Yi* ) = 𝛽0 + 𝛽1 *\* Dia +* 𝛽2 \* AlloyA + 𝛽3 \* AlloyB

* 1. Explain precisely what each regression coefficient measures in the model that you wrote for part (a).

β1 represents the change in the mean response μY for each additional unit increase in the quantitative predictor Dia ... for all groups (A,B and C)

β2 represents how much higher (or lower) the mean response function of the group that uses alloy A is than that of the group that uses alloy C

β3 represents how much higher (or lower) the mean response function of the group that uses alloy B is than that of the group that uses alloy C

β0 represents the overall "average" intercept ignoring group

* 1. What null hypothesis would be tested to determine whether there are differences among the alloys? Write the hypothesis in terms of the regression coefficients in part (a).

To perform the test of whether or not there is a significant difference between alloys:

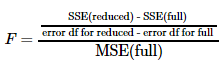
The null hypothesis that makes this happen is H0 : β2 = β3 = 0

HA : at least one of β2, β3 are not equal to 0H0: β2 = β3 = 0 vs. Ha:

* 1. Explain how you would carry out the test of hypothesis described in part (c). Do not forget to write down the degrees of freedom for the test you propose.
     1. H0: β2 = β3 = 0 vs. Ha: at least one of β2, β3 are not equal to 0.

H0: β2 = β3 = 0 vs. Ha: at least one of β2, β3 are not equal to 0

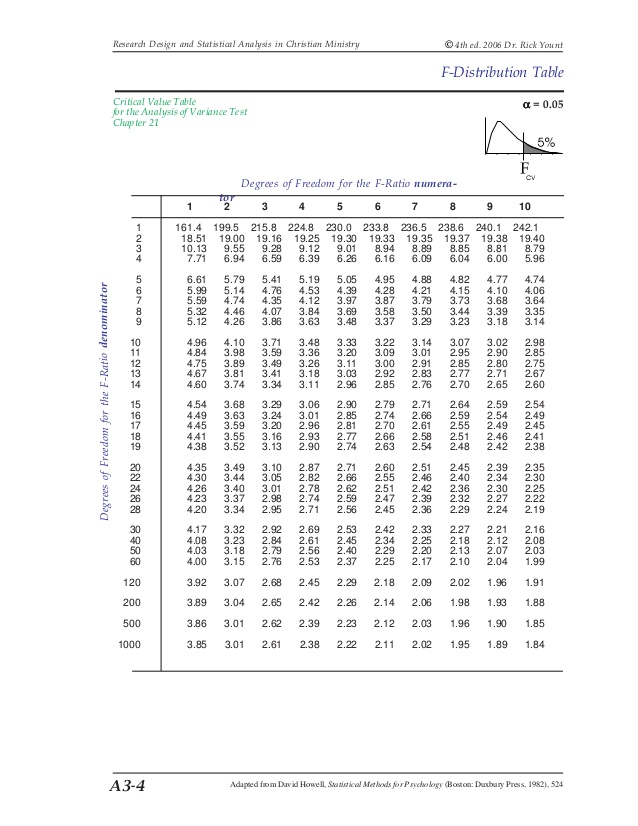
To calculate the F-statistic



with df 2 and 26 (i.e., this F-statistic comes from an F2,26 distribution)

* 1. What term(s) should be added to the model to create a model with interactions? Write down the model.

E(Yi ) = 𝛽0 + 𝛽1 \* Dia + 𝛽2 \* AlloyA + 𝛽3 \* AlloyB + 𝛽4 \* DiaAlloyA + 𝛽5 \* DiaAlloyB

****