

READ THIS SECTION CAREFULLY.

This exam is for your work ONLY. You may use:

1. Your notes
2. Your book
3. The class videos (made by me)
4. A calculator

You may not use:

1. Your computer (other than downloading, printing, or reading the exam). Note: If you really want, you can type your answers in Word, but I don't really think you will want to with all of the equations.
2. Excel or any other spreadsheet program
3. The internet except D2L notes and videos
4. Other individuals

Failure to follow these guidelines will result in:

1. A grade of zero on the exam
2. A report of academic misconduct to the Office of Student Standards

Please believe me when I say that I take this seriously. I need to be able to trust that this is your work. Just because this is an online class does not change anything – you should treat this as you would an in-class exam where you are allowed to use your book and notes.

Completion of this exam means you have acknowledged the rules listed above and will act accordingly.

****You will upload your answers either in Word or a pdf the same way as you submit your homework assignments in ONE document.** NO PICTURES! Make sure it is in order!!**

Due date: Tuesday, December 15, 2020 at 11:00 pm. No late exams will be accepted!

Please make sure your answers are written clearly and IN ORDER. Make sure every question is labeled with the proper number. Read each question carefully. For each hypothesis test, you must write both the null and alternative hypotheses, the decision rule, the calculated test statistic, the critical value of the test statistic and then (finally) the decision. You must have ALL parts! (Don't assume that I will find it – make it clear.)

Show all work if you want any partial credit. Point values are in parentheses.

1. Suppose you are interested in determining the factors that affect the violent crime rate in the 100 largest metropolitan areas (MSAs) in the U.S. You gather information and estimate the following model:

$$VC_i = \beta_0 + \beta_1 MedInc_i + \beta_2 Pov_i + \beta_3 UnEmp_i + \beta_4 AvAge_i + \beta_5 Police_i + \varepsilon_i$$

where:

VC_i = violent crime rate (crimes per 1,000 people) in MSA i

$MedInc_i$ = median income in MSA i (in thousands of dollars)

Pov_i = poverty rate in MSA i

$UnEmp_i$ = unemployment rate in MSA i

$AvAge_i$ = average age in MSA i

$Police_i$ = number of police officers per 1,000 residents in MSA i

The results from the estimation are as follows:

Coefficients:

	<u>Estimate</u>	<u>Std. Error</u>
(Intercept)	12.322	2.566
MedInc	-1.455	0.045
Pov	0.745	0.552
UnEmp	0.983	0.794
AvAge	-1.566	0.567
Police	0.549	0.223

R-squared: 0.6402, Adjusted R-squared: 0.5922		
F-statistic: 78.98 on 5 and 846 DF		

(6) a. Test the overall significance of the model at the 5% significance level.

Ho: $\beta_1 = \beta_2 = \beta_3 = \dots = \beta_5 = 0$

H1: At least one $\beta \neq 0$

F-statistic: 78.98 numerator(m)=5, denominator=846 and 846= DF

Reject Ho if F-calculated > F-tabulated

F-tabulated(alpha=0.05,5,846) = 2.2246

F-calculated=78.98 78.98 > 2.382

Reject the null hypothesis and conclude that at least one of the β is not equal to zero.

(6) b. Provide an interpretation of the coefficient on $AvAge_i$.

The coefficient of years of $AvAge_i = (-1.566)$, $s.e = 0.567$. It is negative.

An increase in average age in MSA i by one year, the violent crime rate (crimes per 1,000 people) in MSA i decreases by 1.566 holding median income, poverty rate in MSA, unemployment rate and number of police officers per 1,000 residents constant.

(6) c. It is believed that the impact of median income on the violent crime rate is negative. Test this hypothesis at the 5% level.

Alpha=0.05 d.f.=846 t(0.05,)=1.647

Ho: $\beta_1 = 0$

H1: $\beta_1 > 0$

Decision rule; Reject Ho if $|t\text{-calculated}| > t\text{-critical}$ t calculated = $\beta_1 / SE(\beta_1) = (-1.455) / 0.045 = -32.33$

t-critical=1.647

t-calculated= -2.814

$32.33 > 1.647$, we reject the null hypothesis because the t-calculated is greater than the t-critical. The coefficient of median income is greater than 0. And sign of the co-efficient is positive

(4) d. The R^2 of the model estimated is 0.6402. What does this number tell us?

R^2 Is a measure of goodness-fit of linear regression .It shows the total variation for the dependent variable that could be explained by the independent variables.

$R^2 = 0.6402 = 62\%$. The model explains 62% of the fitted data in the regression model.

(6) e. Both the poverty rate and the unemployment rate are expected to be positively related to the violent crime rates. Unfortunately, the t-statistic for the coefficient on Pov_i is 1.35 and the t-statistic for the coefficient on $Unemp_i$ is 1.24. Do you have any suggestions of why these two coefficients may not be significantly different from zero?

There might be heteroskedasticity, multicollinearity .

(6) f. Name three ways you could check your answer for (b).

Durbin-Watson test

Breusch-Pagan test

Newey-west

- (6) g. How would you go about calculating a variance inflation factor for $Unemp_i$? Be specific about the steps involved. Also, how big does the VIF have to be before we would get concerned?

Conduct multiple regression with $Unemp_i$ as the dependent variable and median income, poverty rate in MSA, unemployment rate and violent crime rate (crimes per 1,000 people) as the independent variables.

We consider the Adjusted R-squared

Calculate the VIF of $Unemp_i$

The variance inflation factors of $Unemp_i = 1/(1 - \text{Adjusted R-squared})$

Obtain the VIF, If $VIF < 5$, we don't have a problem

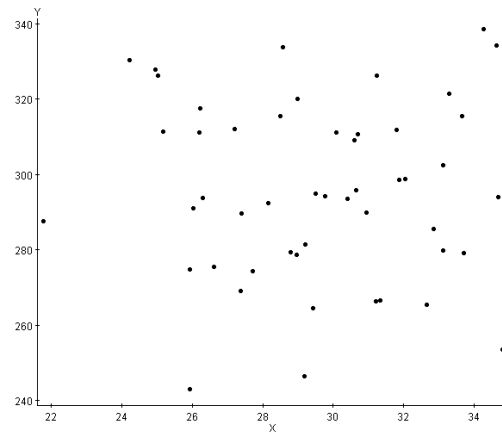
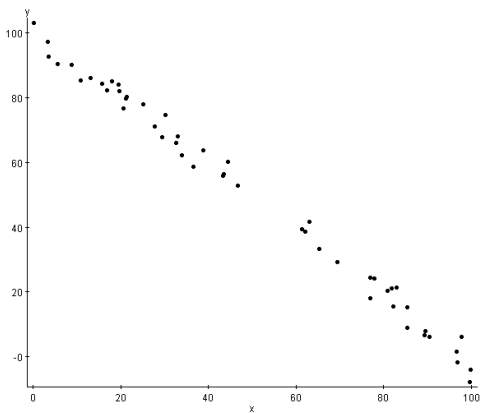
it suggests we possibly dropping one of the variables out of the model or finding some way to consolidate them to capture their joint effect.

2. Let's talk about correlation.

(4) a. What is the range of possible values that a correlation coefficient can take on?

The correlation coefficient can take values range between +1 and -1

(6) b. Below are two scatterplots showing the relationship between variables x and y. For each, provide an estimate of the correlation coefficient. Explain why you chose these values.



The first scatter plot estimate to be -0.98 which shows a strong negative correlation

The first scatter plot estimate to be slope of almost Zero -0.18 close to zero which shows a weak negative correlation and the line will have

3. The following model was estimated using OLS and a time series data set with 86 observations:

$$RCons_t = \beta_0 + \beta_1 RGDP_t + \varepsilon_t$$

where:

$RCons_t$ = Consumption expenditures (in constant dollars) in year t

$RGDP_t$ = Real GDP in year t

The output from the regression model is as follows:

Coefficients:

	<u>Estimate</u>	<u>Std. Error</u>
(Intercept)	-0.01879	32.03
RGDP	0.6752	0.004082

R-squared: 0.9969, Adjusted R-squared: 0.9969

F-statistic: 2.735e+04 on 1 and 84 DF

(6) a. Test whether the coefficient on $RGDP_t$ is positive at the 5% level of significance.

Alpha=0.05 N=50 d.f.=84

$t(0.05,)=1.6787$

$t \text{ calculated} = \beta_1 / SE(\beta_1) = (0.6752 / 0.004082) = 165.40$

Decision rule; Reject H_0 if $|t\text{-calculated}| > t\text{-critical}$ and the sign positive

$t\text{-critical} = 1.66$

$H_0: \beta_1 = 0$

$H_1: \beta_1 > 0$

$165.40 > 1.66$. We reject the null hypothesis because the $t\text{-calculated}$ is greater than the $t\text{-critical}$ and that the coefficient of $RGDP_t$ is positive. It is greater than Zero.

(6) b. Provide an interpretation of the coefficient on $RGDP_t$.

The coefficient of years of $RGDP_t = (0.6752)$, $s.e = 0.004082$. It is positive.

An increase in average age in Real GDP in year by one unit, the Consumption expenditures increases by 67 cents.

(8) c. A Durbin-Watson test was performed for this model. The results are:

Durbin-Watson test
data: model1
DW = 0.12803

Test this model for positive serial correlation at the 5% level of significance.

N= 86 k=1 df =N-1-1=84 Alpha=0.05

Ho: $\rho = 0$

H1: $\rho > 0$

alternative hypothesis: true autocorrelation is greater than 0

Decision rule; If Durbin-Watson (d)<dl we reject Ho , If d>du do not reject Ho

From the table we get dl=(1.481+1.496)/2=1.4885

Du=(1.529+1.541)/2= 1.535

0.12803<1.4885 we reject Ho and conclude that there is a positive correlation

(6) d. Explain what it means for a regression to have positive serial correlation.

A positive serial correlation means that the values are likely to change in future time periods in the same way, or direction, that they have in recent past time periods.

4. Economists are interested in determining the sales price of used tractors sold at an auction. The following model was estimated using OLS and a cross section data set with 276 observations:

$$Price_i = \beta_0 + \beta_1 Horsepower_i - \beta_2 Age_i - \beta_3 EngHours_i + \beta_4 Cab_i + \beta_5 JohnDeere_i + \varepsilon_i$$

where:

$Price_i$ = price paid for tractor i (in dollars)

$Horsepower_i$ = horsepower of tractor i

Age_i = the number of years since tractor i was manufactured

$EngHours_i$ = number of hours of use recorded on tractor i

Cab_i = dummy variable that = 1 if tractor i has an enclosed cab; = 0 otherwise

$JohnDeere_i$ = dummy variable that =1 if tractor i is a John Deere; = 0 otherwise

The results from the estimation are as follows:

Coefficients:

	<u>Estimate</u>	<u>Std. Error</u>
(Intercept)	11363.5419	2308.3146
Horsepower	223.0509	14.8812
Age	-790.4896	126.7347
EngHours	-1.9608	0.3926
Cab	8089.1044	2483.6226
JohnDeere	13731.7988	2899.2501

R-squared: 0.6512, Adjusted R-squared: 0.6447		
F-statistic: 100.8 on 5 and 270 DF		

(6) a. Provide an interpretation of the coefficient on $JohnDeere_i$.

The coefficient of years of JohnDeere =(13731.7988),s.e=2899.2501.It is positive.

An increase in average age in JohnDeere by one unit, the price paid for tractor i (in dollars) increases by 13731 dollars,80 cents horsepower of tractor i , Age_i , number of hours of use recorded on tractor i , dummy variable that = 1 if tractor i has an enclosed cab; = 0 otherwise constant.

(6) b. To test for heteroskedasticity, a White test was performed. Here are the results of the

estimation:

Coefficients:

	<u>Estimate</u>	<u>Std. Error</u>
(Intercept)	5.726e+07	1.795e+08
Horsepower	8.909e+05	2.843e+06
Age	-7.604e+06	9.317e+06
EngHours	-9.565e+04	5.610e+04
Cab	1.362e+08	2.322e+08

JohnDeere	6.678e+08	2.751e+08	
Horsepower2	-2.788e-02	1.922e-02	
Age2	3.877e-02	2.196e-02	'''
EngHours2	2.876e-02	1.786e-02	
Cab2	-3.227e-02	2.665e-02	
JohnDeere2	-5.337e-02	3.652e-02	
Horsepower:Age	-1.741e+04	7.740e+04	
Horsepower:EngHours	-4.384e+02	1.777e+02	
Horsepower:Cab	5.984e+06	2.743e+06	
Horsepower:JohnDeere	3.346e+06	1.551e+06	
Age:EngEours	8.212e+03	1.785e+03	
Age:Cab	-4.132e+07	1.298e+07	
Age:JohnDeere	-2.237e+07	1.467e+07	
EngHours:Cab	-6.960e+03	4.216e+04	
EngHours:JohnDeere	-1.430e+05	4.509e+04	
Cab:JohnDeere	7.921e+08	3.040e+08	

R-squared: 0.5113, Adjusted R-squared: 0.4831			
F-statistic: 18.14 on 20 and 255 DF			

Calculate the White test statistic and test for heteroskedasticity at the 5% level of significance.

H0: $\alpha_1 = \alpha_2 = \alpha_3 = \dots = \alpha_9 = 0$ (Homoskedasticity)

H1: Heteroskedasticity

Decision rule: Reject H0 χ^2 -calculated > χ^2 critical $df = N - k - 1 = N - 20 - 1, N = df + 20 + 1$

χ^2 -calculated = $N \times \text{Multiple R-squared}$ $N = 255 + 20 + 1 = 276$

$= 276 \times 0.5113 = 140.12$

χ^2 critical ($\chi^2, df = 20$ $\alpha = 0.05$) = **31.41**, $df = 18 = 28.86$ which are lower

$140.12 > 31.41$ we reject the null hypothesis and conclude that there is heteroskedasticity

- (6) c. The researcher is considering adding another independent variable to the analysis: *Diesel*, which is a dummy variable = 1 if tractor *i* runs on diesel fuel and = 0 otherwise. What variables will have to be added to the White regression testing for heteroskedasticity? (List them all.)

Horsepower

Age

JohnDeere

Diesel

Horsepower2

Age2

EngHours2

Cab2

JohnDeere2

*Diesel*²

Horsepower:Age

Horsepower:EngHours

Horsepower:Cab

Horsepower:JohnDeere

Age:EngHours

Age:Cab

Age:JohnDeere

EngHours:Cab

EngHours:JohnDeere

Diesel:Age

Diesel: EngHours

Diesel: Cab

Diesel: JohnDeere

Cab:JohnDeere

- (6) d. Explain why it is important to test for heteroskedasticity. If a researcher fails to correct for heteroskedasticity, what may happen?

A researcher should correct for heteroskedasticity, corrected standard error so to obtain a less biased. Presence of residual variance in the model shows that the scattering of the model is dependent on at least one independent one independent variable.

Consequences of heteroskedasticity is that the estimators and the regression predictions based on them remains unbiased and constant.

The estimators are no longer best linear unbiased estimator because they are no longer efficient so the predictions would result to insignificant.

Because of the inconsistency of covariance matrix of the estimated regression coefficients, the tests hypotheses are no longer valid.