Take Home Exam Agreement Form

Date: ____5/5/2017_____

I (type your name here-on the line) _____Carlos Sathler_____ have received a copy of the final exam for SPEA V506.

Directions: Read and initial next to each statement, type your name below as indicated. Your initial located next to each item below signifies your understanding and compliance with the instructions for this take home examination.

_css_____ **I understand that I may not copy or distribute this exam to anyone.**

_css_____ **I understand that I must type the exam on a computer and that hand-written (scanned) exams will not be accepted.**

_css_____ **I understand that I must return the examination via Canvas, by the stated deadline May 5, 2017 and that exams will not be accepted via email.**

_css_____ **I understand this is a test and I cannot copy answers from other students. I understand that I may not discuss this exam with other students (current or previous). Furthermore, I cannot seek assistance, of any kind, from ANYONE.**

_css_____ **I understand that I CAN consult resources including the textbook for the class, lecture notes, my own hand-written notes, the course PowerPoint slides, or any of the resource on the Canvas course site. I understand that I CANNOT consult the web beyond Canvas, or use SAS to complete for proof my answers.**

_css_____ **I understand that violation of this agreement will result in an F on this exam and it cannot be replaced, it will be averaged in (as a 0%) with my other class scores. Violations of this agreement will be handled according to IU's policies on academic misconduct.**

Print your name here: _____Carlos Sathler_____

Name: __Carlos Sathler_____

Final Exam - SPEA V506 (Spring 2017)

Part I - Multiple Choice: Circle the correct answer. (40 pts, each question is worth 2 pts)

1. The method of *least squares* refers to choosing values of coefficient estimates that:

A. minimize the sum of the residuals
B. maximize the sum of squares of the error term
C. minimize the sum of the squares of the vertical distances between the actual Y values and the predicted values of Y
D. B and C
E. A, B, and C

2. An *F* statistic is:

A. a ratio of two means
B. a ratio of two variances
C. used to test the null hypothesis of equal population variances
D. is constructed by putting the larger sample variance in the denominator
E. B and C
F. B, C, and D

3. The *Pearson product-moment correlation coefficient* is used to:

A. assess multicollinearity issues between independent variables
B. measure the strength and direction of the relationship between two variables
C. test for heteroscedasticity
D. A and B
F. A, B, and C'

4. The primary difference(s) between the two variable linear model and the general linear model is (are):

A. the number of independent variables and the assumption of a correctly specified equation
B. the number of independent variables and the assumption of a model that is linear in the parameters
C. the number of independent variables and the assumption of no statistical dependence between the independent variables
D. the number of independent variables and the assumption of no linear dependence between the independent variables
E all of the above

5. The difference between the residual term and the error term in regression theory and analysis is that:

A. the residual is derived from estimating the regression equation for the population whereas the error term is derived from estimating the regression equation for the sample
B. the residual is derived from estimating the regression equation for the sample whereas the error term is derived from estimating the regression equation for the sampling distribution

C. the residual represents the variation in the dependent variable that is not accounted for in the sample regression function, whereas the error term represents the random component associated with the dependent variable in the population
D. the error term represents the variation in the dependent variable that is not accounted for in the sample regression function, whereas the residual represents the random component associated with the dependent variable in the population
E. the residual represents random effects such as measurement error which generally cancel out, whereas the error term represents the effect of incorrect specification of the sample regression function

6. The standard error of a simple linear regression represents:

A. the average error in predicting the value of y
B. the standard deviation of the sampling distribution for b
C. the standard deviation of the sampling distribution for a
D. the standard deviation of the sampling distribution for u
E. the average error in predicting the mean of x

Use the following data for questions 7-9. A recent study of the relationship between social activity and education for a sample of corporate executives showed the following results.

	Social Activity		
Education	Above Average	Average	Below Average
College	30	20	10
High School	20	40	90
Grade School	10	50	130

7. Using 0.05 as the significance level, what is the critical value for the test statistic?

A. 9.488
B. 5.991
C. 7.815
D. 3.841

8. What is the value of the chi-square test statistic?

A. 100
B. 83.67
C. 50
D. 4.94

9. Based on the analysis in questions 7-8, what can be concluded?

A. Social activity and education are correlated.
B. Social activity and education are not related.
C. Social activity and education are related.
D. No conclusion is possible.

10. Assume the least squares equation is $\hat{Y} = 10 + 20X$. What does the value of 10 in the equation indicate?

A. When X = 0, Y = 10.
B. X increases by 10 for each unit increase in Y.
C. Y increases by 10 for each unit increase in X.
D. It is the error of estimation.

11. A sales manager for an advertising agency believes that there is a relationship between the number of contacts that a salesperson makes and the amount of sales dollars earned. What is the dependent variable?

A. Salesperson
B. Number of contacts
C. Amount of sales dollars
D. Sales manager

12. Given the least squares regression equation, $\hat{Y} = 1{,}202 + 1{,}133X$, when X = 3, what does $\hat{Y}$ equal?

A. 5,734
B. 8,000
C. 4,601
D. 4,050

13. Using the following information, estimate the value of $\hat{Y}$ when X = 4.

	Coefficients			
Intercept	–12.8094			
Independent Variable	2.179463			
ANOVA				
	df	*SS*	*MS*	*F*
Regression	1	12323.56	12323.56	90.04814
Residual	8	1094.842	136.8552	
Total	9	13418.4		

A. 10.45
B. 3.73
C. 8.718
D. -4.092

14. A manager at a local bank analyzed the relationship between monthly salary and three independent variables: length of service (measured in months), gender (0 = female, 1 = male), and job type (0 = clerical, 1 = technical). The following table summarizes the regression results:

ANOVA				
Source of Variation	df	Sum of Squares	Mean Square	F
Regression	3	1004346.771	334782.257	5.96
Residual	26	1461134.596	56197.48445	
Total	29	2465481.367		
	Coefficients	Standard Error	t Stat	p-value
Intercept	784.92	322.25	2.44	0.02
Service	9.19	3.20	2.87	0.01
Gender	222.78	89.00	2.50	0.02
Job	-28.21	89.61	-0.31	0.76

The results for the variable gender show that:

A. Males average $222.78 more than females in monthly salary
B. Females average $222.78 more than males in monthly salary
C. Gender is not related to monthly salary
D. Gender and months of service are correlated

15. A sales manager for an advertising agency believes that there is a relationship between the number of contacts that a salesperson makes and the amount of sales dollars earned.

A regression ANOVA shows the following results:

ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1.00	13555.42	13555.42	156.38	0.00
Residual	8.00	693.48	86.68		
Total	9.00	14248.90			

What is the value of the coefficient of determination?

A. -0.9513
B. 0.9754
C. 0.6319
D. 0.9513

16. A multiple regression model includes the interaction term $(X_1)(X_2)$. The term implies that:

A. The independent variables are correlated
B. there is only one independent variable in the regression model
C. the effect of X_1 on the dependent variable is independent of the value of X_2
D. the effect of X_1 on the dependent variable depends on the value of X_2.

17. The variance inflation factor can be used to reduce multicollinearity by:

A. eliminating variables for a multiple regression model
B. decreasing homoscedasticity
C. evaluating the distribution of residuals
D. testing the null hypothesis that all regression coefficients equal zero

18. When the variance of the differences between the actual and the predicted values of the dependent variable vary depending on the value of the independent variable, the differences are said to exhibit _____.

A. Homoscedasticity
B. Heteroscedasticity
C. Multicollinearity
D. All of the Above
E. None of the Above

19. It has been hypothesized that overall academic success for college freshmen as measured by grade point average (GPA) is a function of IQ scores, X1, hours spent studying each week, X2, and one's high school average, X3. Suppose the regression equation is: $\hat{Y} = 6.9 + 0.055X_1 + 0.107X_2 + 0.0853X_3$

What is the predicted GPA for a student with an IQ of 108, 32 hours spent studying per week, and a high school average of 82?

A. 3.789
B. 2.652
C. 3.145
D. 2.256
E. None of the Above

20. A multiple regression analysis showed the following results

	Coefficients	*Standard Error*	*t-stat*	*p-value*
Intercept	139.577	84.235	1.657	0.104
X_1	–0.135	5.698	–0.024	0.981
X_2	3.734	1.220	3.062	0.004
X_3	2.448	26.524	–0.376	0.709
X_4	–9.976	1.044	2.345	0.024

X_4 is a qualitative (dummy) variable. If this variable is equal to one, what is the predicted effect on the dependent variable?

A. Decrease of 9.976, holding all the other independent variables constant.
B. Increase of 9.976, holding all the other independent variables constant.
C. Increase of 1.044, holding all the other independent variables constant.
D. Decrease of 1.044, holding all the other independent variables constant.
E. Y is equal to 139.577 when all the other independent variables are equal to zero.

Part II – Short Answer Essays and Problems [60 pts; show work; circle numerical answers]

21. One of the first regression efforts in estimating the returns to schooling – in other words, the causal effect of going to college on future earning – was made by economist, Jacob Mincer, as described in Angrist and Pischke's chapter on "The Wages of Schooling". Briefly explain what Mincer found in relation to how potential experience affects the coefficient estimate of the effect of schooling. (5 pts)

According to Chapter 6 class handout, Mincer created 2 regression equations with log of income as a dependent variable, using a sample of about 31k nonfarm white man in the 60's census. Equation 1 only used years spent studying as an independent variable (S). Equation 2 used years spent studying (S) *and* potential years of work experience (W) as independent variables. We repeat below the equations from the book:

1: $\ln Y_i = \alpha + 0.07 * S_i + e_i$

2: $\ln Y_i = \alpha + 0.107 * S_i + 0.081 * X_i + 0.00012 * X_i^2 + e_i$

As equation 1 shows, one unit increase in years of study (all other factors remaining constant) should generate a 7% increase in the log of income, on average. Equation 2, on the other hand, shows that one unit increase in years of study (all other factors remaining constant) should generate a 10.7% increase in log of income, on average.

Therefore, Mincer created a model (equation 2) showing that potential experience affects the coefficient estimate of the effect of schooling by increasing it from 7% to 10.7%, which is a 52.9% increase. Schooling has a much higher impact on a person's income if we also consider the potential work experience of that person.

22. What are the assumptions of linear-regression, and what implications would violations of each assumptions have for regression analysis? What techniques are used to test for violations of each assumption, and what are some possible ways to address violations for each assumption, should they occur? (15 pts)

Assumptions of linear-regression (per the class slides):

1. There is a linear relationship between the dependent variable and set of independent variables. The linear regression equation Y = b0 + b1 * X1 + … + bn*Xn + u is the "correct specification".
2. The variation of residuals (y – yhat), is unrelated to whether yhat is large or small. In other words, the spread of the residuals is constant along the regression line/plane. (Homoscedasticity)
3. The residuals are normally distributed with a mean approximately equal to zero.
4. The independent variables should not be themselves highly correlated. (Absence of multicollinearity)
5. The residuals are independent, i.e., successive observations of the dependent variable are not correlated. (Absence of autocorrelation)

What implications of each assumption have for regression analysis?

In a general sense together these assumptions ensure the best method for the estimation of a linear regression equation is the Ordinary Least Square (OLS) method. If any of them won't hold then using the OLS method won't yield the best fit for the data and the inferences made with the Global Test and the individual test of significance of coefficients will be unreliable/incorrect. The same can be said about predictions of dependent variable based on new values of the independent variables.

1. If the relationship is not linear the linear regression equation won't provide a good fit for the data, and predictions using a linear regression equation will be unreliable or incorrect. As an extreme example, that would happen if we would tried to fit a plane in the three-dimensional space to data that follows a spherical pattern.
2. If the variation of residuals changes with the value of the predicted value the variances of the regression coefficients will not be constant and could not be determined by the OLS formulas we learned in class. The results we obtained in SAS for example for the standard errors of coefficients

using the PROC REG would be incorrect, since the standard error for the coefficient will vary depending on what section of the regression line/plane is being analyzed.

3. If residuals are not normally distributed with mean zero that implies the sample values for the dependent variable are not normally distributed around the regression line/plane, with an approximate mean falling on the line/plane. For linear regression the observed data should be spread around the regression line according to the normal distribution and should have means approximately on the regression line/plane. If that assumption is violated the linear regression equation is not a good model for the observed data.
4. If there is multicollinearity one of the highly correlated variables may appear in the regression equation with a negative coefficient even though it has a positive correlation with the dependent variable. Or it may have a coefficient that is not statistically significant even thought the independent variable is highly correlated to the dependent variable. Additionally, there may be a drastic change in remaining regression coefficients in the presence of multicollinearity when a variable is added or removed.
5. If residuals show signs of not being independent we are again using linear regression when it's not best suited to fit the data at hand. There may be some seasonality in the data that explains the autocorrelation and would require some transformation before we could fit a linear regression model.

Techniques used to test for violations of each assumption:

1. Analysis of scatter plots (dependent variable in y(z) axis, independent variable in the x(x,y) axis) will confirm linearity of relationship between dependent variable and independent variable in the 2(3) dimensional space. Residual (scatter) Plots with residuals in the y axis and predicted values on the x axis will confirm if distribution of residuals follow a random pattern. Residual Plots are useful for n-dimensional space problems.
2. Residual plots that exhibit a random pattern will offer evidence of homoscedasticity. If pattern is not random, we say there is evidence of heteroscedasticity. For example, a residual plot could show a cone shaped pattern, where residuals increase as predicted value increases.
3. A histogram of residuals and/or the Normal Probability Plot will allow confirmation that the residuals are normally distributed. The histogram should show a bell shape with mean zero and the residuals in the Normal Probability Plot should adhere closely to the line in the plot.
4. The best technique to test for multicollinearity is the Variance Inflation Factor (VIF). If VIF is greater than 10 for any variable in the model then multicollinearity among independent variables is confirmed. Another way to test for multicollinearity is to analyze a correlation matrix; the existence of one or more independent variable pairs that show a correlation greater than 0.7 or smaller than -0.7 is an indication of multicollinearity.
5. Autocorrelation can be identified with scatter plots showing residuals in the y axis and predicted values in the x axis. For example, a linear pattern in the plot may indicate seasonality in the data. There is also a test of autorrelation called Durgin-Watson that is mentioned in the Lind book.

Possible ways to address violations for each assumption:

1. Data transformation is a good way to address violations of assumptions 1-3. For example, we could take the log of the dependent variable and obtain a good fit for the data. By obtaining a better fit for the data using data transformation we may also eliminate issues with heteroscedasiticity and obtain a distribution of residuals that is normal.
2. The way to address issues of multicollinearity is by removing one of the variables in a par of variables that exhibit high correlation. We should also use VIF to identify variables to remove. If VIF is greater than 10 the variable should be removed.
3. Autocorrelation could be addressed for example by exploring if the data is time related. There are techniques we did not study in class that can be used to factor seasonality. For example in chapter 18 of the book there is an example that shows a regression line being fitted to "deseasonlized" sales data.

23. In a multiple regression equation, two independent variables are considered, and the sample size is 25. The regression coefficients and the standard errors are as follows.

$b_1 = 2.676 \qquad s_{b1} = 0.56$
$b_2 = -0.880 \qquad s_{b2} = 0.71$

Conduct a test of hypothesis to determine whether either independent variable has a coefficient equal to zero. (15 pts)

A. State the null and alternative hypotheses for each coefficient. (6 pts)

For b_1	For b_2
$\boldsymbol{H_0: b_1 = 0}$	$\boldsymbol{H_0: b_2 = 0}$
$\boldsymbol{H_1: b_1 \neq 0}$	$\boldsymbol{H_1: b_2 \neq 0}$

B. Calculate the *t* statistic and interpret the results for each variable. Use the .05 significance level. (6 pts)

$$t = \frac{b - 0}{s_b} \Rightarrow \qquad t_1 = \frac{b_1 - 0}{s_{b_1}} = \frac{2.676}{0.56} = \mathbf{4.7786} \qquad t_2 = \frac{b_2 - 0}{s_{b_2}} = \frac{-0.880}{0.71} = \mathbf{-1.2394}$$

Step 1: See part A above
Step 2: alpha = 0.05
Step 3: We will use the t-statistics.
This is a two-tailed test for both coefficients.
$n = number\ of\ observations = 25$
$k = number\ of\ independent\ variables = 2$
Degrees of freedom = $df = n - (k + 1) = 25 - (2 + 1) = \mathbf{22}$
Step 4: Reject H_0 if $|t| > t_{\alpha/2,df} \Rightarrow |t| > t_{0.025,22} \Rightarrow \boldsymbol{|t| > 2.074}$
Step 5: Arrive at decision
$|t_1| = 4.7786 > 2.074 \Rightarrow Reject\ H_0: b_1 = 0$
$|t_2| = 1.2394 < 2.074 \Rightarrow Fail\ to\ reject\ H_0: b_2 = 0$
Step 6: Interpretation of results
Given a t-value of **4.7786** and a critical value of **2.074**, we can ***reject*** H_0 for independent variable 1 at the 0.05 level of significance, indicating that there is statistically significant evidence of a relationship between the independent variable 1 and the dependent variable in the multiple regression equation.
Given a t-value of **-1.2394** and a critical value of **2.074**, we ***fail to reject*** H_0 for independent variable 2 at the 0.05 level of significance, indicating that there is no statistically significant evidence of a relationship between the independent variable 2 and the dependent variable in the multiple regression equation.

C. Would you consider deleting either variable from the regression equation? Why or why not? (3 pts)
I would consider deleting independent variable 2. The t-value for variable 2 is not in the critical region for the level of significance for the problem, and therefore b_2 could be zero. Based on this result, we conclude that variable 2 is not a significant predictor of the dependent variable and I would remove it from the regression equation.

24. Suppose that the sales manager of a large automotive parts distributor wants to estimate as early as April the total annual sales.

According to the manager of the distribution warehouse, several factors are related to annual sales (measured in millions of dollars) (*sales*), including the number of retail outlets in the region stocking the company's parts (*outlets*), the number of automobiles in the region registered as of April 1 (measured in millions) (*cars*), the total personal income for the first quarter of the year (measured in billions of dollars) (*income*), the average age of automobiles in years (*age*), and the number of supervisors at the distribution warehouse (*bosses*). The data for all these variables were gathered for a recent year.

Consider the following correlation matrix.

	sales	*outlets*	*cars*	*income*	*age*
outlets	0.899				
cars	0.605	0.775			
income	0.964	0.825	0.409		
age	-0.323	-0.489	-0.447	-0.349	
bosses	0.286	0.183	0.395	0.155	0.291

A. Which single variable has the strongest correlation with the dependent variable? Is there evidence of multicollinearity? If so, between what variables? (3 pts)

Income has the strongest correlation with the dependent variable (0.964).
Yes, there is evidence of multicollinearity since a few variables show a correlation outside the range (-0.7,0.7).
The following dependent variables violate the rule for multicollinearity:
Cars and outlets (correlation = 0.775 > 0.7)
Income and outlets (correlation = 0.825 > 0.7)

Using the data, the following multivariate regression equation was estimated:
sales = -19.7 – 0.00063 *outlets* – 1.74 *cars* + 0.410 *income* + 2.04 *age* – 0.034 *bosses*

The output for all five variables is shown below.

Predictor	Coef	SE Coef	T	P
Constant	-19.672	5.422	-3.63	0.022
Outlets	-0.000629	0.002638	-0.24	0.823
Cars	-1.7399	0.5530	3.15	0.035
Income	0.40994	0.04385	9.35	0.001
Age	2.0357	0.8779	2.32	0.081
bosses	-0.0344	0.1880	-0.18	0.864

Analysis of Variance

SOURCE	DF	SS	MS	F	P
Regression	5	1593.81	318.76	140.36	0.000
Residual Error	4	9.08	2.27		
Total	9	1602.89			

B. State the null hypothesis concerning the statistical significance of the overall regression, test this hypothesis, and interpret the results. (use a .05 level of significance) (6 pts)

We will use the Global test.

Step 1: State the hypotheses

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

$H_1: At\ least\ one\ \beta_i\ is\ not\ zero$

Step 2: Level of significance alpha = 0.05

Step 3: Select the appropriate statistic

We will use the F statistic, and select that information from the ANOVA table.

Step 4: Formulate the decision rule

Decision Rule: Reject H_0 if $F > F_{\alpha,k,n-(k+1)}$

Where alpha = 0.05, k = independent variables, n = sample size; n-(k+1) = degrees of freedom. Since we don't have n (the sample size), we will use the p-value for the F statistic from the ANOVA table for our test.

Decision Rule: Reject if p-value < α

Step 5: Arrive at a decision

Using the p-value from the ANOVA table, we see that the F statistic for the sample (140.36) is associated with a p-value of 0.000, which is less than 0.05. Hence F is in the rejection region and therefore we reject the null hypothesis.

Step 6: Interpret the results

The result of the Global Test of the multiple regression show us that (at the 0.05 significance level) at least one of the independent variables has a coefficient different than zero and is therefore related to the dependent variable *Sales*. In other words, one or more independent variables in the regression equation explain the variation in the dependent variable.

C. What percent of the variation is explained by the regression equation? (2 pts)

The Coefficient of Determination will give us that answer.

$$R^2 = \frac{SSR}{SS\ Total} = \frac{1593.81}{1602.89} = \mathbf{0.9943}$$

We conclude that 99.43% of the variation in the dependent variable *Sales* is explained by variations in the independent variables in the regression equation.

D. Interpret the results (both statistical significance and magnitude of effect) for each of the independent variables in the model. (Use a .05 level of significance) (10 pts)

Significance Tests

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

Given a t-value of **-0.24** and a p-value of **0.823**, we fail to reject H_0 at the **0.05** level of significance, indicating that there is no statistically significant evidence of a relationship between **Outlets** and **Sales** holding the effects of other independent variables constant. The standard error of β_1 indicates the average error in estimating β_1 is **0.002638**.

$H_0: \beta_2 = 0$

H_1: $\beta_2 \neq 0$

Given a t-value of **3.15** and a p-value of **0.035**, we reject H_0 at the **0.05** level of significance, indicating that there is statistically significant evidence of a relationship between **Cars** and **Sales** holding the effects of other independent variables constant. The standard error of β_2 indicates the average error in estimating β_2 is **0.5530**.

H_0: $\beta_3 = 0$
H_1: $\beta_3 \neq 0$

Given a t-value of **9.35** and a p-value of **0.001**, we reject H_0 at the **0.05** level of significance, indicating that there is statistically significant evidence of a relationship between **Income** and **Sales** holding the effects of other independent variables constant. The standard error of β_3 indicates the average error in estimating β_3 is **0.04385**.

H_0: $\beta_4 = 0$
H_1: $\beta_4 \neq 0$

Given a t-value of **2.32** and a p-value of **0.081**, we fail to reject H_0 at the **0.05** level of significance, indicating that there is no statistically significant evidence of a relationship between **Age** and **Sales** holding the effects of other independent variables constant. The standard error of β_4 indicates the average error in estimating β_4 is **0.8779**.

H_0: $\beta_5 = 0$
H_1: $\beta_5 \neq 0$

Given a t-value of **-0.18** and a p-value of **0.864**, we fail to reject H_0 at the **0.05** level of significance, indicating that there is no statistically significant evidence of a relationship between **bosses** and **Sales** holding the effects of other independent variables constant. The standard error of β_3 indicates the average error in estimating β_3 is **0.1880**.

Magnitude of effect:

$\hat{\beta}_1$ = **-0.000629.** A one unit increase in **Outlets** would result in a **0.000629 decrease** in ***Sales***, holding the effects of other independent variables constant.

$\hat{\beta}_2$ = **-1.7399**. A one unit increase in **Cars** would result in a **1.7399 decrease** in ***Sales***, holding the effects of other independent variables constant.

$\hat{\beta}_3$ = **0.40994**. A one unit increase in **Income** would result in a **0.40994 increase** in ***Sales***, holding the effects of other independent variables constant.

$\hat{\beta}_4$ = **2.0357.** A one unit increase in **Age** would result in a **2.0357 increase** in ***Sales***, holding the effects of other independent variables constant.

$\hat{\beta}_5$ = **-0.0344**. A one unit increase in **bosses** would result in a **0.0344 decrease** in ***Sales***, holding the effects of other independent variables constant.

E. What would be the projected value in annual sales if the following were true?
outlets = 1739, *cars* = *9.27*, *income* = 85.4, *age* = 3.5, and *bosses* = 9.0
If these values are outside the range of values used for the regression, would this be a reliable forecast? Why or why not? (4 pts)

$$\text{sales} = -19.7 - 0.00063\,\text{outlets} - 1.74\,\text{cars} + 0.410\,\text{income} + 2.04\,\text{age} - 0.034\,\text{bosses}$$
$$\text{sales} = -19.7 - 0.00063 * 1739 - 1.74 * 9.27 + 0.410 * 85.4 + 2.04 * 3.5 - 0.034 * 9$$
$$\text{sales} = 4.92263$$

No, this would not be a reliable forecast if the values are outside the range of values used in the regression because one or more assumptions of multiple regression could be violated outside the range of values used to derive the regression equation. For example, the relationship between dependent and independent variables could follow a non-linear patter outside the range of values used to generate the model.