Review Test Submission: MBC638 Quiz #5 - Simple Linear Regression (due Sunday, Sept. 30, 10:00pm)

User	David Forteguerre
Course	MBC.638.M001.FALL18.Data Anls & Decisn Making
Test	MBC638 Quiz #5 - Simple Linear Regression (due Sunday, Sept. 30, 10:00pm)
Started	9/30/18 4:02 PM
Submitted	9/30/18 4:30 PM
Status	Completed
Attempt Score	100 out of 100 points
Time Elapsed	28 minutes out of 1 hour
Results Displayed	All Answers, Submitted Answers, Correct Answers, Feedback, Incorrectly Answered Questions

Question 1

20 out of 20 points

Use the data file **Housing prices Boston.xlsx** . The data provides information on a sample of 506 communities in the Boston area. We are interested in understanding how the median housing price of a community is explained by various community characteristics. The table below summarizes the variables in the spreadsheet.

Variable Name	Description	Units of Measurement
Price	Median housing price in the community	U.S. dollars ($)
Crime	Crime rate	Number of crimes per 100,000 population
Nitrogen Oxide	Amount of nitrogen oxide in the air	Parts per million
Rooms	Average number of rooms in houses in the community	Number of rooms
Distance	Weighted distance of the community from five employment centers	Miles
Property Tax	Property tax rate	Percent (%)
Student Teacher Ratio	Average student-teacher ratio of schools in the community	Number of students per teacher

Estimate a simple linear regression equation involving Price (the dependent variable) and Distance (the explanatory variable).

a) **How good is the linear model?** Provide a relevant number that summarizes the answer: **[a]**

b) **Interpret the intercept** of your linear model. When the distance from five employment centers equals **[g]** miles, the median housing price is predicted to be **$[c]**.

c) **Interpret the slope** of your linear model. Every additional **[b]** of **[d]** is predicted to increase **[e]** by **$[f]**.

Selected Answer:

Use the data file **Housing prices Boston.xlsx** . The data provides information on a sample of 506 communities in the Boston area. We are interested in understanding how the median housing price of a community is explained by various community characteristics. The table below summarizes the variables in the spreadsheet.

Variable Name	Description	Units of Measurement
Price	Median housing price in the community	U.S. dollars ($)
Crime	Crime rate	Number of crimes per 100,000 population
Nitrogen Oxide	Amount of nitrogen oxide in the air	Parts per million
Rooms	Average number of rooms in houses in the community	Number of rooms
Distance	Weighted distance of the community from five employment centers	Miles
Property Tax	Property tax rate	Percent (%)
Student Teacher Ratio	Average student-teacher ratio of schools in the community	Number of students per teacher

Estimate a simple linear regression equation involving Price (the dependent variable) and Distance (the explanatory variable).

a) **How good is the linear model?** Provide a relevant number that summarizes the answer: **0.0622**

b) **Interpret the intercept** of your linear model. When the distance from five employment centers equals **0** miles, the median housing price is predicted to be **$ 18,373**.

c) **Interpret the slope** of your linear model. Every additional **mile(s)** of **distance** is predicted to increase **median price** by **$ 1,090**.

Answers:

Use the data file **Housing prices Boston.xlsx** . The data provides information on a sample of 506 communities in the Boston area. We are interested in understanding how the median housing price of a community is explained by various community characteristics. The table below summarizes the variables in the spreadsheet.

Variable Name	Description	Units of Measurement
Price	Median housing price in the community	U.S. dollars ($)
Crime	Crime rate	Number of crimes per 100,000 population
Nitrogen Oxide	Amount of nitrogen oxide in the air	Parts per million
Rooms	Average number of rooms in houses in the community	Number of rooms
Distance	Weighted distance of the community from five employment centers	Miles
Property Tax	Property tax rate	Percent (%)
Student Teacher Ratio	Average student-teacher ratio of schools in the community	Number of students per teacher

Estimate a simple linear regression equation involving Price (the dependent variable) and Distance (the explanatory variable).

a) **How good is the linear model?** Provide a relevant number that summarizes the answer: **0.0622**

b) **Interpret the intercept** of your linear model. When the distance from five employment centers equals **0** miles, the median housing price is predicted to be **$ 18,373**.

c) **Interpret the slope** of your linear model. Every additional **mile(s)** of **distance** is predicted to increase **median price** by **$ 1,090**.

All Answer Choices

- 0
- 0.0622
- 0.2493
- 2.5120
- 0.05703
- median price
- distance
- mile(s)
- 18,373
- 1,090 thousand
- 1,090
- 2,512

Response Feedback:

Question 2 20 out of 20 points

In a regression model to predict the **selling price of a house** using the **square footage** as a predictor, the following coefficients were obtained:

- Intercept = 85,526.84
- Slope = 47,290.97

The measurement units of the two variables are:

- Selling price: $
- Square feet: thousands of square feet

A house with 2,500 square feet is currently on the market for $175,000. It seems **underpriced**. TRUE or FALSE?

Selected Answer: True
Answers: True
False

Response Feedback:

Question 3 20 out of 20 points

The Excel data **Catalog Marketing.xlsx** contains data on 1,000 HyTex customers.

- Age: coded as 1 for 30 or younger, 2 for 31 to 55, 3 for 56 or older
- Gender: coded as 1 for males, 0 for females
- OwnHome: coded as 1 if the customer owns a home, 0 otherwise
- Married: coded as 1 if the customer is currently married, 0 otherwise
- Close: coded as 1 if the customer lives reasonably close to a shopping area that sells similar merchandise, 0 otherwise
- Salary: combined annual salary of the customer and spouse (if any)
- Children: number of children living with the customer
- History: coded as "NA" if the customer had no dealings with HyTex before this year, 1 if the customer was a low-spending customer last year, 2 if medium-spending, 3 if high-spending
- Catalogs: number of catalogs sent to the customer this year
- FirstPurchase: date of the customer's first purchase with HyTex
- AmountSpent: total amount of purchases made by the customer this year
- In addition, the variables Region, State, and City indicate where the customer resides.

Unfortunately, due to a virus that has infected my computer, the original Excel file got lost...

However, the following table of sample **covariances** was recovered, along with the historic sample **average** values:

	Salary	Catalogs	AmountSpent
Salary	936,421,374.79		
Catalogs	37,181.14	43.82	
AmountSpent	20,564,792.27	3,005.42	922,742.20

	Average
Salary	56,103.90
Catalogs	14.68
AmountSpent	1,216.77

Based on this information, can you **predict the Amount Spent** ($) when the **number of catalogs** sent to the customer this year is **10**? Round your answer to 2 decimal places. Don't write the $ symbol. (For example, if the predicted amount spent is is $950 then write 950.00)

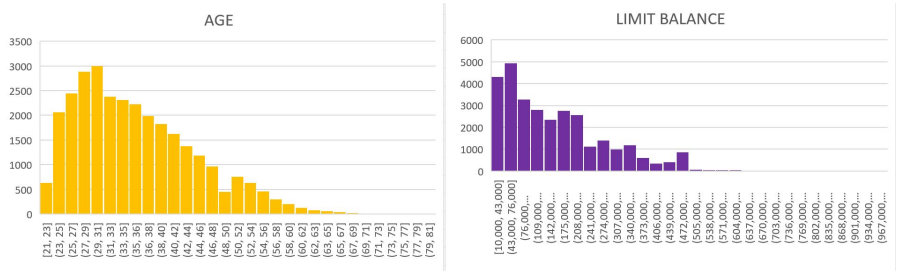
Selected Answer: 895.79
Correct Answer: 895.64
Answer range +/- 3 (892.64 - 898.64)
Response Feedback:

Question 4 20 out of 20 points

The Excel data **default-of-credit-card-clients.xlsx** contains information on default payments, demographic factors, credit data, history of payments, and bill statements of 30,000 credit card clients in Taiwan from April 2005 to September 2005. There are 25 variables in the dataset. Some relevant variables are:

Variable Name	Description
LIMIT_BAL	Amount of given credit in NT dollars (includes individual and family/supplementary credit)
SEX	Gender (1=male, 2=female)
EDUCATION	Graduate school (1), university (2), high school (3), others (4), unknown (5 and 6)
AGE	Age in years
Default.payment.next.month	A dummy variable: 1=yes, 0=no

You are a bank examiner and **would like to understand how the age of credit card clients affects their amount of given credit**. You notice that the distributions of the **AGE** and **LIMIT_BAL** variables are both right-skewed and potentially contain outliers:



Using the LN( ) command in Excel, you convert the **AGE** variable into **logarithmic form** but keep the **LIMIT_BAL** variable in its **original form**, and run a linear regression model. The regression output is displayed below.

Multiple Regression for LIMIT_BAL *Summary*	Multiple R	R-Square	Adjusted R-square	Std. Err. of Estimate	Rows Ignored	Outliers
	0.1816	0.0330	0.0329	127594.9132	0	0

ANOVA Table	Degrees of Freedom	Sum of Squares	Mean of Squares	F	p-Value
Explained	2	1.66518E+13	8.32591E+12	511.4050544	< 0.0001
Unexplained	29997	4.88365E+14	16280461880		

Regression Table	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95% Lower	Confidence Interval 95% Upper
Constant	-179,871.28	10926.93956	-16.46126759	< 0.0001	-201288.5482	-158454.0039
SEX	11,048.98	1512.739047	7.303953933	< 0.0001	8083.942625	14014.01
LN(age)	93,197.25	2941.571315	31.68281075	< 0.0001	87431.64081	98962.85376

Interpret the coefficient of LN(age). Pick one from the list below.

Selected Answer:
For both genders, every 1% increase in age is predicted to increase balance limit by 932 dollars.

Answers:
For both genders, every 1% increase in age is predicted to increase balance limit by 932 dollars.

For both genders, every 1% increase in age is predicted to increase balance limit by 0.93 %.

For both genders, every 1 year increase in age is predicted to increase balance limit by 932 dollars.

For both genders, every 1% increase in age is predicted to increase balance limit by 93,197 dollars.

For both genders, every 1 year increase in LN-age is predicted to increase balance limit by 93,197 dollars.

Response Feedback:

Question 5

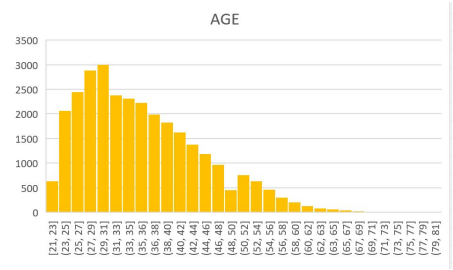
20 out of 20 points

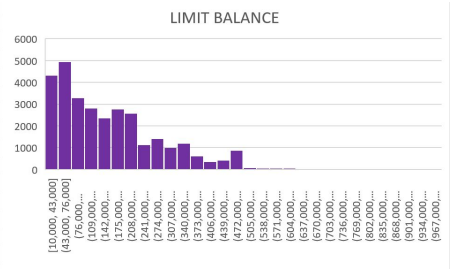
The Excel data **default-of-credit-card-clients.xlsx** contains information on default payments, demographic factors, credit data, history of payments, and bill statements of 30,000 credit card clients in Taiwan from April 2005 to September 2005. There are 25 variables in the dataset. Some relevant variables are:

Variable Name	Description
LIMIT_BAL	Amount of given credit in NT dollars (includes individual and family/supplementary credit)
SEX	Gender (1=male, 2=female)
EDUCATION	Graduate school (1), university (2), high school (3), others (4),

	unknown (5 and 6)
AGE	Age in years
Default.payment.next.month	A dummy variable: 1=yes, 0=no

You are a bank examiner and **would like to understand how the age of credit card clients affects their amount of given credit**. You notice that the distributions of the **AGE** and **LIMIT_BAL** variables are both right-skewed and potentially contain outliers:





Using the LN( ) command in Excel, you convert the **LIMIT_BAL** variable into **logarithmic form** but keep the **AGE** variable in its **original form**, and run a linear regression model. The regression output is displayed below.

Multiple Regression for LN(limit_bal) Summary	Multiple R	R-Square	Adjusted R-square	Std. Err. of Estimate	Rows Ignored	Outliers
	0.1600	0.0256	0.0255	0.929045353	0	0

ANOVA Table	Degrees of Freedom	Sum of Squares	Mean of Squares	F	p-Value
Explained	2	679.9894261	339.9947131	393.9112031	< 0.0001
Unexplained	29997	25891.16869	0.863125269		

Regression Table	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95% Lower	Confidence Interval 95% Upper
Constant	10.87958343	0.028933368	376.0220187	< 0.0001	10.82287278	10.93629408
SEX	0.165818218	0.011011858	15.05815112	< 0.0001	0.144234502	0.187401933
AGE	0.014583146	0.000584322	24.95739347	< 0.0001	0.01343785	0.015728442

Interpret the coefficient of AGE. Pick one from the list below.

Selected Answer:
For both genders, every 1 year increase in age is predicted to increase balance limit by 1.46%.

Answers:
For both genders, every 1 year increase in age is predicted to increase balance limit by 1.46%.

For both genders, every 1 year increase in age is predicted to increase balance limit by 0.0146%.

For both genders, every 1 year increase in age is predicted to increase balance limit by 1.46 dollars.

For both genders, every 1% increase in age is predicted to increase balance limit by 1.46%.

For both genders, every 1% increase in age is predicted to increase balance limit by 1.46 dollars.

Response Feedback:

Wednesday, November 21, 2018 1:13:56 PM EST

← OK