QMS 210: Multiple Linear Regression

TRAIN TO LEARN EFFECTIVELY: TIP SHEETS

WHAT IS MULTIPLE LINEAR REGRESSION?

- The multiple linear regression model is used to predict the change in a single dependent variable (Y) based on **more than one** independent variable (X).[1] The dependent variable is a measurement and the independent variables are different factors that affect this measurement

DIFFERENCES FROM SIMPLE LINEAR REGRESSION

- Cannot be performed on a calculator and requires the use of **SPSS**
- T-test is performed on **multiple independent variables.**
- An **adjusted coefficient of determination (adjusted r^2)**, as opposed to a coefficient of determination (r^2), is used because of the effect of the multiple independent variables in the multiple linear regression model

MULTIPLE LINEAR REGRESSION MODEL

- $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_k X_k + \varepsilon$
- Where:
  - Y = Dependent variable
  - β_0 = Y-intercept
  - β_1 = Slope coefficient of X_1, representing the relationship between X_1 and Y**
  - β_2 = Slope coefficient of X_2, representing the relationship between X_2 and Y**
  - β_k = Slope coefficient of any future independent variable*
  - X_1 = Independent Variable 1
  - X_2 = Independent Variable 2
  - X_k = Independent Variable k (future independent variable)
  - ε = random error in Y

1. All information in this tip sheet is based on Chapter 16 of: Chua, C., & Smith, D. (2020). Multiple Linear Regression. In Business Statistics, Fifteenth Custom Edition for Ryerson University, Ted Rogers School of Management (pp. 693–713). essay, Pearson Learning Solutions. Retrieved 2025, from https://plus.pearson.com/courses/vanhilten02713/products/156133/pages/1?locale=&platformId=1030&lms=Y.

QMS 210: Multiple Linear Regression

TRAIN TO LEARN EFFECTIVELY: TIP SHEETS

*The sample multiple linear regression equation replaces β with b (b is an **estimate** of β) and **does not have** an error term.

Assuming the other independent variables are **held constant (no effect on dependent variable)

LINE ASSUMPTIONS

- **L**inearity: A **linear relationship** exists between the dependent variable and **each** independent variable.
- **I**ndependence of errors: There is **no relationship** between the **errors** (residuals) of the multiple regression model.
- **N**ormality of error: The **errors** (residuals) in the multiple regression model are **normally distributed.**
- **E**qual Variance: The **errors** (residuals) in the multiple regression model **do not change** (have the **same variation**) for different values of the **dependent variable**.

SPSS - COEFFICIENTS TABLE

- **Unstandardized coefficients,** located in Column **B**, are used to develop the **sample multiple linear regression equation** and represent the **slope coefficients** of the independent variables**.**
    - The sample multiple linear regression equation is used to **predict** the value of the dependent variable **based on** the values of the independent variables in a **sample.** It is an **estimate** of the multiple linear regression model
    - The y-intercept is known as **constant** and the unstandardized coefficient for the constant is the **value** of the y-intercept
    - **Disregard** the standardized coefficients, represented by Column Beta, as these are **not** used to develop the sample regression equation

1. All information in this tip sheet is based on Chapter 16 of: Chua, C., & Smith, D. (2020). Multiple Linear Regression. In Business Statistics, Fifteenth Custom Edition for Ryerson University, Ted Rogers School of Management (pp. 693–713). essay, Pearson Learning Solutions. Retrieved 2025, from https://plus.pearson.com/courses/vanhilten02713/products/156133/pages/1?locale=&platformId=1030&lms=Y.

QMS 210: Multiple Linear Regression

TRAIN TO LEARN EFFECTIVELY: TIP SHEETS

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	**87.830**	6.385		13.756	.000	75.155	100.506
	age	**-.165**	.063	-.176	-2.633	.010	-.290	-.041
	weight	**-.385**	.043	-.677	-8.877	.000	-.471	-.299
	heart_rate	**-.118**	.032	-.252	-3.667	.000	-.182	-.054
	gender	**13.208**	1.344	.748	9.824	.000	10.539	15.877

a. Dependent Variable: VO2max

(Laerd Statistics, n.d.)

- Using this coefficients table describing the relationship between oxygen volume (dependent variable) and body characteristics (independent variables) as an example, it can be seen that the sample multiple linear regression equation is:
    - $Y = 87.830 - 0.165X_1 - 0.385X_2 - 0.118X_3 - 13.208X_4$
    - Where:
        - X_1 = age (independent variable)
        - X_2 = weight (independent variable)
        - X_3 = heart_rate (independent variable)
        - X_4 = gender (independent variable)
        - 87.830 = Y-intercept
        - Y = Dependent variable
- The coefficients table can also be used to perform a **2 tailed, 1 sample t-test on each independent variable** in the regression equation
    - H_0: The independent variable slope coefficient **is 0** and there is **no linear relationship** between the independent variable and the dependent variable
    - H_a: The independent variable slope coefficient **is not 0** and signifies that a linear relationship between the independent and dependent variable **exists**

1. All information in this tip sheet is based on Chapter 16 of: Chua, C., & Smith, D. (2020). Multiple Linear Regression. In Business Statistics, Fifteenth Custom Edition for Ryerson University, Ted Rogers School of Management (pp. 693–713). essay, Pearson Learning Solutions. Retrieved 2025, from https://plus.pearson.com/courses/vanhilten02713/products/156133/pages/1?locale=&platformId=1030&lms=Y.

- The t-test statistic is provided in the column titled **t** & the p-value of the t-test is provided in the column titled **Sig.**
- As can be seen in the previous coefficients table about the relationship between oxygen volume (dependent variable) and body characteristics (independent variable):
  - The p value of weight, heart rate, and gender are all 0.000, which is **less** than the alpha of 0.05 (alpha is 0.05 unless indicated otherwise)
  - The null hypothesis for these independent variables is **rejected** at the 5% level of significance. Thus, a linear relationship between these independent variables and oxygen volume **exists**.

SPSS - ANOVA TABLE (F TEST)

- The F-test is used to determine if there is a **linear relationship** between the dependent variable and the **entire** multiple linear regression model
  - H_0: **All** the independent variables have a **slope of 0**, signifying there is **no** linear relationship between the dependent variable and the multiple regression model
  - H_a: **At least one** of the independent variables has a **non-zero slope**, signifying that a linear relationship between the dependent variable and the multiple regression model **exists**
    - Note: This is the same type of null and alternative hypothesis as a regular ANOVA, changing the population mean with the **independent variable slope**.
  - The F-test statistic is provided in the column of the ANOVA table titled **F** & the p-value of the F-test is provided in the column of the ANOVA table titled **Sig.**

1. All information in this tip sheet is based on Chapter 16 of: Chua, C., & Smith, D. (2020). Multiple Linear Regression. In Business Statistics, Fifteenth Custom Edition for Ryerson University, Ted Rogers School of Management (pp. 693–713). essay, Pearson Learning Solutions. Retrieved 2025, from https://plus.pearson.com/courses/vanhilten02713/products/156133/pages/1?locale=&platformId=1030&lms=Y.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4196.483	**4**	1049.121	**32.393**	**.000[b]**
	Residual	3076.778	**95**	32.387		
	Total	7273.261	99			

a. Dependent Variable: VO2max

b. Predictors: (Constant), gender, age, heart_rate, weight

(Laerd Statistics, n.d.)

- Using the ANOVA table about the relationship between oxygen volume (dependent variable) and body characteristics (independent variable):
  - The p-value is 0.000, which is **less** than the alpha of 0.05
  - The F-test statistic is 32.393, which is **greater** than the F critical value of 2.467 (calculated through InvF, setting n:df as 4, d:df as 95, area as 0.05)
    - Thus, the null hypothesis of the F-test is **rejected** at the 5% level of significance. This means that a linear relationship **exists** between at least one of the independent variables in the multiple regression model and the dependent variable

Example:

Test scores for 200 high school students were collected in the following subjects: reading, social sciences, math, and science. As well, information on whether the students were female or male were collected (1 indicated a female and 0 indicated a male). A multiple linear regression model was created from this data, with SPSS output provided below

1. All information in this tip sheet is based on Chapter 16 of: Chua, C., & Smith, D. (2020). Multiple Linear Regression. In Business Statistics, Fifteenth Custom Edition for Ryerson University, Ted Rogers School of Management (pp. 693–713). essay, Pearson Learning Solutions. Retrieved 2025, from https://plus.pearson.com/courses/vanhilten02713/products/156133/pages/1?locale=&platformId=1030&lms=Y.

QMS 210: Multiple Linear Regression

TRAIN TO LEARN EFFECTIVELY: TIP SHEETS

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	12.325	3.194		3.859	.000	6.027	18.624
	math score	.389	.074	.368	5.252	.000	.243	.535
	female	-2.010	1.023	-.101	-1.965	.051	-4.027	.007
	social studies score	.050	.062	.054	.801	.424	-.073	.173
	reading score	.335	.073	.347	4.607	.000	.192	.479

a. Dependent Variable: science score

(UCLA Statistical Methods and Data Analytics, n.d.)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	9543.721	4	2385.930	46.695	.000[a]
	Residual	9963.779	195	51.096		
	Total	19507.500	199			

a. Predictors: (Constant), reading score, female, social studies score, math score

b. Dependent Variable: science score

(UCLA Statistical Methods and Data Analytics, n.d.)

Questions:

1. Create a sample multiple linear regression equation and predict the science test score resulting from a male student having a reading test score of 50%, a math test score of 85%, and a social studies test score of 70%.

1. All information in this tip sheet is based on Chapter 16 of: Chua, C., & Smith, D. (2020). Multiple Linear Regression. In Business Statistics, Fifteenth Custom Edition for Ryerson University, Ted Rogers School of Management (pp. 693–713). essay, Pearson Learning Solutions. Retrieved 2025, from https://plus.pearson.com/courses/vanhilten02713/products/156133/pages/1?locale=&platformId=1030&lms=Y.

- As seen in the coefficients table, the **dependent variable** is science test score and the **independent variables** are the math test score, the social studies test score, the reading test score, and the sex of the student.
    a. Thus, the sample multiple linear regression equation is: $Y = 12.325 + 0.389X_1 + 0.050X_2 + 0.335X_3 - 2.010X_4$
        i. Inputting a math test score of 85%, a social studies test score of 70%, a reading test score of 50%, and the fact that the student is male (corresponding to a value of 0) into the sample multiple linear regression equation yields the following science tests score:
            1. $Y = 12.325 + 0.389(85) + 0.050(70) + 0.335(50) - 2.010(0)$
            2. $Y = 65.64\%$

2. Is there a linear relationship between the multiple linear regression model and science test scores at the 5% level of significance?

- Since this question is asking about whether or not there is a **linear relationship** between the **entire multiple linear** regression model and science test scores (the **dependent variable**), the F-test for multiple linear regression (and thus the ANOVA table) will be used.

- The F-test for multiple linear regression is **upper tailed**, with the the null and alternative hypotheses being:
    a. H_0: **All** the independent variables of the multiple linear regression model (math test score, social studies test score, reading test score, and sex of the student) have a **slope of 0**, signifying there is **no** linear relationship between the dependent variable and the multiple regression model

    b. H_a: **At least one** of the independent variables of the multiple linear regression model has a **non-zero slope**, signifying that a linear relationship between the dependent variable and the multiple regression model **exists**

- Looking at the **ANOVA table**, the p-value (written as **sig.**) is 0.000. This is far less than the alpha of 0.05 provided in the question. As well, the F-test statistic is 46.695. This is far greater than the F-critical value of 2.418.

1. All information in this tip sheet is based on Chapter 16 of: Chua, C., & Smith, D. (2020). Multiple Linear Regression. In Business Statistics, Fifteenth Custom Edition for Ryerson University, Ted Rogers School of Management (pp. 693–713). essay, Pearson Learning Solutions. Retrieved 2025, from https://plus.pearson.com/courses/vanhilten02713/products/156133/pages/1?locale=&platformId=1030&lms=Y.

c. Both of these facts provide strong evidence to **reject the null hypothesis** of the F-test at the 5% level of significance
    i. Thus, at least one of the independent variables of the multiple linear regression model has a non-zero slope. This signifies that a linear relationship between the science test score and the multiple linear regression model exists at the 5% level of significance

3. Which of the test scores or student sex information have a linear relationship with the science test score at the 5% level of significance?

- Since this question is asking **which independent variable** (math test score, social studies test score, reading test score, and sex of the student) has a **linear relationship** with the **dependent variable** (science test score), the t-test for multiple linear regression (and thus the coefficients table) will be used for **each** independent variable

- The t-test for multiple linear regression is **two tailed**, with the the null and alternative hypotheses being:
    - H_0: The independent variable slope coefficient is **0** and there is **no** linear relationship between the independent variable and the dependent variable
    - H_a: The independent variable slope coefficient is **not 0** and signifies that a linear relationship between the independent and dependent variable **exists**

- Looking at the **coefficients table**:
    - The math test score has a p-value of 0.000 and a t-test statistic of 5.252. The p value is far less than the alpha of 0.05. As well, the t-test statistic is far greater than the upper t-critical value of 1.972
        - Both of these facts provide strong evidence to **reject** the null hypothesis of the t-test for the math test at the 5% level of significance
    - The social studies test score has a p-value of 0.424 and a t-test statistic of 0.801. The p value is far greater than the alpha of 0.05. As well, the t-test statistic is

1. All information in this tip sheet is based on Chapter 16 of: Chua, C., & Smith, D. (2020). Multiple Linear Regression. In Business Statistics, Fifteenth Custom Edition for Ryerson University, Ted Rogers School of Management (pp. 693–713). essay, Pearson Learning Solutions. Retrieved 2025, from https://plus.pearson.com/courses/vanhilten02713/products/156133/pages/1?locale=&platformId=1030&lms=Y.

greater than the lower t-critical value of -1.972 and less than the upper t-critical value of 1.972
- Both of these facts provide strong evidence to **not reject** null hypothesis of the t-test for the social studies test score at the 5% level of significance

- The reading test score has a p-value of 0.000 and a t-test statistic of 4.607. The p value is far less than the alpha of 0.05. As well, the t-test statistic is far greater than the upper t-critical value of 1.972
  - Both of these facts provide strong evidence to **reject** the null hypothesis of the t-test for the reading test score at the 5% level of significance

- The sex of the student has a p-value of 0.051 and a t-test statistic of -1.965. The p value is greater than the alpha of 0.05. As well, the t-test statistic is greater than the lower t-critical value of -1.972 and less than the upper t-critical value of 1.972
  - Both of these facts provide evidence to **not reject** the null hypothesis of the t-test for the sex of the student at the 5% level of significance

Thus, it can be concluded that the **math and reading test score** slope coefficients are not zero and **have a linear relationship** with the science test score exists at the 5% level of significance

References:

1. All information in this tip sheet is based on Chapter 16 of: Chua, C., & Smith, D. (2020). Multiple Linear Regression. In Business Statistics, Fifteenth Custom Edition for Ryerson University, Ted Rogers School of Management (pp. 693–713). essay, Pearson Learning Solutions. Retrieved 2025, from https://plus.pearson.com/courses/vanhilten02713/products/156133/pages/1?locale=&platformId=1030&lms=Y.

QMS 210: Multiple Linear Regression

TRAIN TO LEARN EFFECTIVELY: TIP SHEETS

Multiple regression analysis using SPSS statistics. How to perform a Multiple Regression Analysis in SPSS Statistics | Laerd Statistics. (n.d.). https://statistics.laerd.com/spss-tutorials/multiple-regression-using-spss-statistics.php

Regression Analysis | *SPSS Annotated Output*. UCLA Statistical Methods and Data Analytics. (n.d.). https://stats.oarc.ucla.edu/spss/output/regression-analysis/

1. All information in this tip sheet is based on Chapter 16 of: Chua, C., & Smith, D. (2020). Multiple Linear Regression. In Business Statistics, Fifteenth Custom Edition for Ryerson University, Ted Rogers School of Management (pp. 693–713). essay, Pearson Learning Solutions. Retrieved 2025, from https://plus.pearson.com/courses/vanhilten02713/products/156133/pages/1?locale=&platformId=1030&lms=Y.