

## Self-Test 2

**Question 1:** The correlation between indoor temperature and thermic comfort is determined two times, one time whereby the temperature is measured in degrees Celsius (case 1) and another time whereby the temperature is measured in degrees Fahrenheit (case 2). What do you know about the covariance and correlation coefficient in the two cases?

- a. **The covariances differ, the correlation coefficients are the same**
- b. The covariances are the same, the correlation coefficients differ
- c. The covariances are the same and the correlation coefficients are the same
- d. The covariances differ and the correlation coefficients differ

**Question 2:** the correlation between two variables is perfect (correlation coefficient is equal to 1 or to -1). Which statement is true?

- a. The variables have the same mean and standard deviation
- b. **The variables have the same z-values**
- c. The two variables only differ by a constant
- d. None of the above

**Question 3:** Which of the following is a requirement for correlation analysis

- a. There must be a causal relationship between the variables
- b. One variable is independent and the other is dependent
- c. **The relationship between the variables is linear**
- d. The variance of the variables must be constant

**Question 4:** If the estimated coefficient of some independent variable in a multiple regression analysis is significant, can we then conclude that there is a *causal* effect of the variable on the dependent variable?

- a. **No, we cannot determine causality from a regression analysis**
- b. Yes, because we control for other independent variables in the regression analysis
- c. Yes, because we have tested the significance of the coefficient
- d. No, because the estimation is based on a sample

**Question 5:** someone includes gender as a variable in a regression analysis and codes it as 0 = man and 1 = woman. Is it allowed to use this variable in a regression analysis?

- a. **Yes as independent variable, Not as dependent variable**
- b. Not as independent variable, Yes as dependent variable
- c. Yes as independent variable and Yes as dependent variable
- d. Not as independent variable and Not as dependent variable

**Question 6:** Why do we strive for parsimony in the selection of independent variables for a regression model?

- a. To avoid multicollinearity
- b. To avoid overfitting**
- c. To improve the goodness-of-fit of the model
- d. To facilitate interpretation of the model

**Question 7:** What does  $R^2$  of a regression model represent?

- a. The sum of squares of the model
- b. The determination coefficient of the model
- c. The standard error of the model
- d. The ratio of variance explained by the model**

**Question 8:** Which of the following is NOT a condition for simple regression analysis

- a. Y is normally distributed for fixed values of X
- b. X is normally distributed**
- c. The variance of Y is constant over the range of X
- d. The relationship between X and Y is linear

**Question 9:** A researcher is interested in the question whether the average income differs between home owners and home renters. Is it possible to use regression analysis for answering this question?

- a. No this can only be tested in an independent samples t-test
- b. No this can only be tested in a paired samples t-test
- c. Yes by means of dummy regression**
- d. Yes by running a regression analysis for each group separately

**Question 10:** If multi-collinearity occurs due to a strong correlation between two predictor variables, what can one do to solve the problem?

- a. Select one variable and exclude the other from the regression analysis (a)
- b. Combine the two variables by summing the two variables (b)
- c. Both methods a and b are possible methods**
- d. None of the methods a and b are possible

**Question 11:** which X-variable in a regression model is the most important independent variable for predicting or explaining the dependent variable (Y)

- a. The variable with the largest t-value
- b. The variable with the smallest standard error
- c. The variable with the largest unstandardized beta
- d. The variable with the largest standardized beta**

**Question 12:** What do you know of a regression model that has been obtained by the stepwise method?

- a. All excluded predictor variables have a zero or small correlation with the dependent variable
- b. All independent variables included in the model are significant**
- c. It is expected that all predictor variables are significant but this is not necessarily the case
- d. The  $R^2$  of the model cannot be further improved by adding other predictor variables

In a survey, people have scored the accessibility by bicycle of a train station on a 7-point scale (1=very poor and 7=very good). A researcher wants to test whether the age and income of a person has an effect on the accessibility score and conducts a regression analysis. Age is measured in years and income is a dichotomous variable:

Low\_inc = 0      modal income or higher  
 Low\_inc = 1      lower than modal income

SPSS provides the following output of the regression analysis:

<b>Coefficients<sup>a</sup></b>					
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	2,590	,321		8,073	,000
age	-,015	,008	-,110	-1,897	,059
low_inc	,534	,196	,157	2,723	,007

a. Dependent Variable: accessibility bike

**Question 13:** What is according to the model the predicted accessibility score for a person who is 23 years old and has a lower than modal income.

- a. 2.779**
- b. 0.217
- c. 1.711
- d. None of the above

**Question 14.** We want to predict the number of visits per month to a shopping centre based on the distance to the shopping centre. We conduct a (simple) regression analysis based on the data of a sample. The below table gives the SPSS output table with the estimation results.

**Coefficients<sup>a</sup>**

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	7,133	,385		18,549	,000
distance to shopping center	-,555	,069	-,671	-7,996	,000

a. Dependent Variable: shopping average shopping visits per month

Based on this results, what is the correct way of writing the regression equation (Y=number of visits per month and X = distance to shopping centre)?

- a.  $Y = -0.555 + 7.133 \times X + e$
- b.  $Y = 7.133 - 0.555 \times X + e$**
- c.  $Y = -0.555 + 7.133 \times X$
- d.  $Y = 7.133 - 0.555 \times X$