Practice questions

Q1

Consider a regression model for the yearly wage (measured in 1000 EUR) with work experience as the predictor (measured in years). Based on a random sample, you obtain the following estimated model

$$\widehat{Wage} = 42.60 + 2.40 \cdot Exper + (-0.05) \cdot Exper^2$$

Which statement is correct?

A. For an individual with one year of work experience, predicted yearly wage is 45000 EUR.

B. For an individual with one year of work experience, predicted yearly wage is 45050 EUR.

C. For an individual with zero work experience, predicted yearly wage is 42600 EUR.

D. For an individual with zero work experience, predicted yearly wage is 44950 EUR.

Q2

Let Z denote a standard normal variable. What is $P(-1 < Z < 1)$?

A. Approximately 10 %

B. Approximately 68 %

C. Approximately 95 %

D. Approximately 99 %

Q3

In a small survey, 10 individuals were asked about which German car brand they prefer. The possible answers are: *Mercedes*, *Audi*, *VW* and *BMW*. The 10 answers are compiled in the following table:

Individual	Car brand
1	Mercedes
2	Mercedes
3	VW
4	BMW
5	Audi
6	Mercedes
7	BMW
8	BMW
9	VW
10	Mercedes

What is the measurement scale of the variable *Car brand*?

A. Nominal scale.

B. Ordinal scale.

C. Interval scale.

D. Ratio scale.

Q4

Use the information and the data in **Q3**. Let Y be a variable defined in the following way:

$$Y = \begin{cases} 1 & \text{if the individual prefers Mercedes} \\ 0 & \text{otherwise} \end{cases}$$

What is the average of Y?

A. 0.1

B. 0.2

C. 0.4

D. 0.6

Q5

Use the information and the data in **Q3** and **Q4**. Suppose you estimate a logistic regression model with Y as the dependent variable and no predictor variables in the specification. What is the estimated value of B_0?

A. -0.67

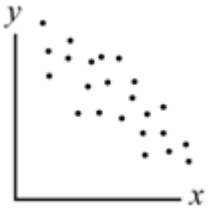
B. -0.41

C. 0.29

D. 0.67

Q6

Let $\{y_i, x_i\}_{i=1}^{n}$ denote a set of n pairs of observations on y and x. Consider the following scatter plot of the data:



You formulate and estimate a regression model of the form:

$$y = B_0 + B_1 x + \varepsilon$$

The estimated regression line is

$$\hat{y} = b_0 + b_1 x$$

Which of the following statements is generally **not** correct?

A. b_0 is an estimate of B_0.

B. $b_0 > 0$.

C. The estimate b_1 is less than zero.

D. The estimated correlation between y and x is negative.

Q7

Consider a regression model for the yearly wage (measured in 1000 NOK) with age, education and work experience as the predictors (all measured in years). Based on a random sample, you obtain the following estimated model

$$\log(\widehat{Wage}) = 4.813 + 0.001 \cdot Age + 0.047 \cdot Educ + 0.061 \cdot Exper$$

First, consider an individual with a predicted yearly wage of 460 000 NOK. Next, assume an individual with one more year of work experience, but with otherwise equal characteristics as the first individual (same age and education). Based on the model, what is the predicted yearly wage of the second individual?

A. It is not possible to compute the predicted wage based on the given information.

B. Approximately 434 000 NOK.

C. Approximately 460 000 NOK.

D. Approximately 488 000 NOK.

Q8

Consider a dataset consisting of 13 items. You decide to run an exploratory factor analysis (EFA) using most of the items in the data. The results of the analysis are displayed below:

Communalities

	Initial	Extraction
item1	1.000	.490
item2	1.000	.563
item3	1.000	.516
item4	1.000	.391
item5	1.000	.626
item6	1.000	.726
item7	1.000	.548
item8	1.000	.614
item9	1.000	.499
item11	1.000	.577
item12	1.000	.468
item13	1.000	.479

Extraction Method: Principal Component Analysis.

Total Variance Explained

Component	Initial Eigenvalues			Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.407	28.391	28.391	3.407	28.391	28.391	3.028	25.235	25.235
2	1.932	16.103	44.494	1.932	16.103	44.494	2.146	17.884	43.119
3	1.157	9.642	54.136	1.157	9.642	54.136	1.322	11.017	54.136
4	.952	7.935	62.071						
5	.894	7.454	69.525						
6	.708	5.904	75.429						
7	.624	5.197	80.626						
8	.589	4.912	85.538						
9	.503	4.193	89.731						
10	.484	4.035	93.766						
11	.393	3.275	97.041						
12	.355	2.959	100.000						

Extraction Method: Principal Component Analysis.

Rotated Component Matrix [a]

	Component		
	1	2	3
item1	-.083	-.247	.650
item2	.743	-.058	.090
item3	.268	.104	.658
item4	.201	.278	.523
item5	-.022	.763	.207
item6	.058	.850	-.023
item7	.703	.143	-.181
item8	.143	.763	-.110
item9	.685	.116	.131
item11	.727	.062	.211
item12	.634	.242	.083
item13	.663	-.106	.167

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 4 iterations.

Component Transformation Matrix

Component	1	2	3
1	.888	.382	.255
2	-.352	.923	-.157
3	-.295	.050	.954

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Ignoring loadings with absolute value of 0.3 or less, which statement about the results of the analysis is **not** correct?

A. According to the rule: *choose the number of factors according to how many eigenvalues that are larger than one*, a three factor solution is appropriate.

B. The choice of rotation method dictates a factor solution with uncorrelated factors.

C. item2, item7, item9, item11, item12 and item13 seem to be associated with a common factor, item5, item6 and item8 seem to be associated with another common factor, while item1, item3 and item4 seem to be associated with a third common factor.

D. The rotation method applied in the analysis is an oblique type of method.

Q9

Consider a linear regression model of the form:

$$Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + \varepsilon$$

You want to evaluate the following hypothesis using the F-test:

$$H_0 : B_2 = B_3$$
$$H_A : B_2 \neq B_3$$

Recall that applying the F-test involves estimating an unrestricted model and a restricted model. What is a correct formulation of the restricted model?

A. $Y = B_0 + B_1X_1 + B_2X_3 + \varepsilon$

B. $Y = B_0 + B_1X_1 + (B_2 + B_3)X_2 + \varepsilon$

C. $Y = B_0 + B_1X_1 + \varepsilon$

D. $Y = B_0 + B_1X_1 + B_3(X_2 + X_3) + \varepsilon$

Q10

Based on a random sample ($n = 33$), you estimate a model of the form:

$$\log(Y) = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + \varepsilon$$

Suppose you want to evaluate the hypothesis:

$$H_0 : B_1 = 0$$
$$H_A : B_1 \neq 0$$

You find the t-statistic to be 2.4. Which statement about the p-value is correct?

A. $0.000 < p\text{-value} \leq 0.010$

B. $0.010 < p\text{-value} \leq 0.050$

C. $0.050 < p\text{-value} \leq 0.100$

D. $p\text{-value} > 0.100$

Q11

As part of a study, you collect a random sample of $n = 14$ observations from a normal distribution with mean μ and variance σ^2. Based on the sample, you use a one-sample t-test for the mean to evaluate the hypothesis:

$$H_0 : \mu = 0$$
$$H_A : \mu > 0$$

After carefully computing the t-statistic, you find the critical value of the test to be 2.16. Approximately, at what frequency do you expect to reject a true null-hypothesis?

A. In 1 out of 200.

B. In 2 out of 200.

C. In 5 out of 200.

D. In 20 out of 200.

Q12

Let Y denote the random variable *number of smoked cigarettes per day*. What scale characterizes the highest measurement level of Y?

A. Nominal scale.

B. Ordinal scale.

C. Interval scale.

D. Ratio scale.

Q13

Consider a regression model for the yearly wage with gender as the only predictor. Gender is here a dummy variable that takes the value 1 for males and the value 0 for females. Based on a random sample, you obtain the following estimated model using the OLS estimator

$$\widehat{Wage} = b_0 + b_1 \cdot gender$$

Which statement is correct?

A. b_0 is the average wage among males.

B. $b_0 + b_1$ is the average wage among females.

C. if $b_1 > 0$, then the average wage among males is less than the average wage among females.

D. If b_1 is approximately 0, then the average wage among males is approximately the same as the average wage among females.

Q14

Consider a linear regression model of the form

$$Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + \varepsilon$$

Recall that testing multiple restrictions using the F-test involves estimating an unrestricted model and a restricted one. Let the restricted model be given by

$$Y^* = B_0 + B_2X_2 + \varepsilon$$

where $Y^* = Y - X_3$. What are the restrictions?

A. $B_1 = 0, B_2 = 1$

B. $B_1 = B_2, B_3 = 1$

C. $B_1 = 0, B_3 = 1$

D. $B_1 = 0, B_3 = 0$

Q15

Let F denote an F-distributed random variable with $df_1 = 12$ and $df_2 = 10$. What is the probability that $F > 4.71$?

A. 0.5%

B. 1%

C. 2.5%

D. 5%