OLAT_exam1final

Detailed results

Course	23FS BIO144 Datenanalyse in der Biologie
Test	Mock_exam_2022_updated

This are your test results

🍫 Section 1

> Please answer the following question.

🔒 Question 1

Response

(Multiple choice; 1 point)

Assume data is in the object `dd`, the response variable is `body_mass_index` and the explanatory variable is `age`. Which of these commands would make an appropriate model of this data?

- ☐ `glm(bodymass_index ~ age, data=dd, family=binomial)`

- ☐ `lm(body_mass_index ~ age, data=dd)`

- ☐ `lmer(age ~ bodymass_index, data=dd)`

- ☐ `lm(age ~ body_mass_index, data==dd)`

▼ Solution

(Multiple choice; 1 point)

Assume data is in the object `dd`, the response variable is `body_mass_index` and the explanatory variable is `age`. Which of these commands would make an appropriate model of this data?

- ☐ `glm(bodymass_index ~ age, data=dd, family=binomial)`

- ☑ `lm(body_mass_index ~ age, data=dd)`

- ☐ `lmer(age ~ bodymass_index, data=dd)`

- ☐ `lm(age ~ body_mass_index, data==dd)`

🍫 Section 2

> Please answer the following question.

🔒 Question 2

Response

(Multiple choice; 1 point)

After we make a linear model in R, and check assumptions, we need to interpret the model. Which of these is true, in this context?

☐ The `summary` table can tell us how well the model explains the data.

☐ If the p-value is of the intercept is less than 0.05, we can reject the null hypothesis that there is no relationship.

☐ The p-values in the `summary` table are the only important information to look at.

☐ The `summary` table of the linear model contains the estimates of the slope and intercept.

▾ Solution

(Multiple choice; 1 point)

After we make a linear model in R, and check assumptions, we need to interpret the model. Which of these is true, in this context?

☑ The `summary` table can tell us how well the model explains the data.

☐ If the p-value is of the intercept is less than 0.05, we can reject the null hypothesis that there is no relationship.

☐ The p-values in the `summary` table are the only important information to look at.

☑ The `summary` table of the linear model contains the estimates of the slope and intercept.

📦 Section 3

Please answer the following question.

🔒 Question 3

Response

(Multiple choice; 1 points)

Which of the following are assumptions that are made when fitting a multiple linear regression model $y_i = \beta_0 + \beta_x x_i + \beta_z z_i + e_i$ with $1 \le i \le n$ and $n =$ number of data points?

☐ The residuals e_i follow a χ^2 distribution.

☐ The explanatory variables x_i and z_i are independent of each other (i.e. they are not collinear).

☐ Both x_i and z_i are linearly related to y_i.

☐ The variance of the residuals (σ_e^2) increases with y_i.

▾ Solution

(Multiple choice; 1 points)

Which of the following are assumptions that are made when fitting a multiple linear regression model $y_i = \beta_0 + \beta_x x_i + \beta_z z_i + e_i$ with $1 \le i \le n$ and $n =$ number of data points?

☐ The residuals e_i follow a χ^2 distribution.

☐ The explanatory variables x_i and z_i are independent of each other (i.e. they are not collinear).

☑ Both x_i and z_i are linearly related to y_i.

☐ The variance of the residuals (σ_e^2) increases with y_i.

📦 Section 4

Please answer the following question.

Response

(Multiple choice; 1 point)

The aim of a study was to find out whether people that eat a lot of saturated fat (% of daily energy income, variable `sat_fat`) tend to be more obese (% bodyfat).

```
##
## Call:
## lm(formula = bodyfat ~ sat_fat, data = dd)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -13.0783 -3.4971  0.0675  3.5775  7.7805
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.5697     3.4134   1.046 0.302451
## sat_fat       0.6533     0.1626   4.018 0.000277 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.804 on 37 degrees of freedom
## Multiple R-squared:  0.3037, Adjusted R-squared:  0.2849
## F-statistic: 16.14 on 1 and 37 DF,  p-value: 0.0002768
```

Look at the above model output and find out which of the following statements about the model are true.

☐ Saturated fat explains about 30% of the variability in body fat

☐ The dataset contains 39 individuals.

☐ The p-value in the last row of the summary table is the same as the one for saturated fat, because the respective F- and t-tests are the same when there is only one continuous variable in the model.

☐ The 95% confidence interval for the effect size of saturated fat ranges from 0.49 to 0.82 .

▼ Solution

(Multiple choice; 1 point)

The aim of a study was to find out whether people that eat a lot of saturated fat (% of daily energy income, variable `sat_fat`) tend to be more obese (% bodyfat).

```
##
## Call:
## lm(formula = bodyfat ~ sat_fat, data = dd)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -13.0783 -3.4971  0.0675  3.5775  7.7805
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.5697     3.4134   1.046 0.302451
## sat_fat       0.6533     0.1626   4.018 0.000277 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.804 on 37 degrees of freedom
## Multiple R-squared:  0.3037, Adjusted R-squared:  0.2849
## F-statistic: 16.14 on 1 and 37 DF,  p-value: 0.0002768
```

Look at the above model output and find out which of the following statements about the model are true.

☑ Saturated fat explains about 30% of the variability in body fat

☑ The dataset contains 39 individuals.

☑ The p-value in the last row of the summary table is the same as the one for saturated fat, because the respective F- and t-tests are the same when there is only one continuous variable in the model.

☐ The 95% confidence interval for the effect size of saturated fat ranges from 0.49 to 0.82 .

🧩 Section 5

Please answer the following question.

🔒 Question 5

Response

(Multiple choice; 1 point)

Linear regression makes the assumption of a linear relationship between an explanatory and the response variable. Which one of these can be used to assess if the linearity assumption is justified?

☐ The frequency distribution of the response variable.

☐ The value of the slope of the relationship between the response and explanatory variable.

☐ A scatterplot of the residuals against the fitted values.

☐ The p-value of the explanatory variable.

▼ Solution

(Multiple choice; 1 point)

Linear regression makes the assumption of a linear relationship between an explanatory and the response variable. Which one of these can be used to assess if the linearity assumption is justified?

☐ The frequency distribution of the response variable.

☐ The value of the slope of the relationship between the response and explanatory variable.

☑ A scatterplot of the residuals against the fitted values.

☐ The p-value of the explanatory variable.

📦 Section 6

> Please answer the following question.

🔒 Question 6

Response

(Single choice; 1 point)

You are studying how individual birds respond to different types of threat (e.g. the presence of cats, dogs, snakes, etc). There are 3 types of threat and a control (i.e. a treatment with four levels). In your study you take data on 72 independent birds. For eight of these, you have some missing data, such that these individuals cannot be included in your analysis. If you were to perform one-way analysis of variance with type of threat as the explanatory variable, and the magnitude of the response as the response variable, how many degrees of freedom for error will there be?

Give your answer without decimal places.

[   ]

▼ Solution

(Single choice; 1 point)

You are studying how individual birds respond to different types of threat (e.g. the presence of cats, dogs, snakes, etc). There are 3 types of threat and a control (i.e. a treatment with four levels). In your study you take data on 72 independent birds. For eight of these, you have some missing data, such that these individuals cannot be included in your analysis. If you were to perform one-way analysis of variance with type of threat as the explanatory variable, and the magnitude of the response as the response variable, how many degrees of freedom for error will there be?

Give your answer without decimal places.

[60.0   ]

📦 Section 7

> Please answer the following question.

Response

(Single choice; 1 point)

Look at the estimated coefficients from regression model $y_i = \beta_0 + \beta_x x_i + \beta_z z_i + \beta_w w_i + e_i$ with continuous explanatory variable x and binary explanatory variables z and w, where a total nunber of 188 data points was included.

```
##             Estimate Std. Error    t value    Pr(>|t|)
## (Intercept)  3.3258777  2.0832569  1.5964799 1.120977e-01
## x            0.8738691  0.0779641 11.2086093 1.468118e-22
## z           -0.6192851  0.7657417 -0.8087389 4.197098e-01
## w           -1.0057110  1.1113157 -0.9049733 3.666629e-01
```
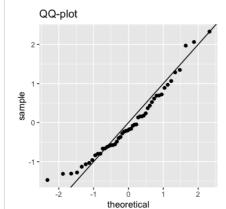
What is the 95% confidence interval of β_z?

○ It is not possible to calculate a confidence interval for binary variables.

○ (0.22, 1.45)

○ (-2.58, 1.34)

○ (-2.13, 0.89)

○ (-1.38, 0.15)

▼ Solution

(Single choice; 1 point)

Look at the estimated coefficients from regression model $y_i = \beta_0 + \beta_x x_i + \beta_z z_i + \beta_w w_i + e_i$ with continuous explanatory variable x and binary explanatory variables z and w, where a total nunber of 188 data points was included.

```
##             Estimate Std. Error    t value    Pr(>|t|)
## (Intercept)  3.3258777  2.0832569  1.5964799 1.120977e-01
## x            0.8738691  0.0779641 11.2086093 1.468118e-22
## z           -0.6192851  0.7657417 -0.8087389 4.197098e-01
## w           -1.0057110  1.1113157 -0.9049733 3.666629e-01
```

What is the 95% confidence interval of β_z?

○ It is not possible to calculate a confidence interval for binary variables.

○ (0.22, 1.45)

○ (-2.58, 1.34)

◉ (-2.13, 0.89)

○ (-1.38, 0.15)

🧊 Section 8

Please answer the following question.
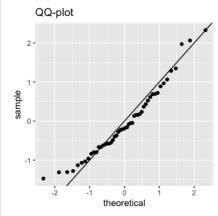
🔒 Question 8

Response

(Multiple choice; 1 point)

QQ-plot

When examining if the assumptions of your linear model are badly violated, you see this graph. What could you conclude?

☐ There are some extreme outliers.

☐ There is evidence of non-constant variance in the residuals.

☐ The residuals appear to be reasonably normally distributed.

☐ There is evidence of non-linearity in the relationship between the response and explanatory variable of the model.

▼ Solution

(Multiple choice; 1 point)

QQ-plot

When examining if the assumptions of your linear model are badly violated, you see this graph. What could you conclude?

☐ There are some extreme outliers.

☐ There is evidence of non-constant variance in the residuals.

☑ The residuals appear to be reasonably normally distributed.

☐ There is evidence of non-linearity in the relationship between the response and explanatory variable of the model.

📦 Section 9

Please answer the following question.

🔒 Question 9

Response

(First question of tree age exercise; Multiple choice; 1 point)
Determining the age of trees can be quite laborious, so researchers wondered if age could be accurately determined from measures of tree height and trunk width. For a group of trees, accurate actual age was known, and height and width were measured.

The dataset therefore contains these three variables:

- `Age` (continuous)
- `Height` (continuous)
- `Width` (continuous)

The data is in the file `tree_age1.csv` .

Which of these is correct?

☐   There are no missing values in the dataset.

☐   Error degrees of freedom for a model with both explanatory variables and no interaction term should be 55.

☐   Analyses of this dataset can inform about effects, i.e. causation.

☐   The two explanatory variables are correlated.

▼ Solution

(First question of tree age exercise; Multiple choice; 1 point)
Determining the age of trees can be quite laborious, so researchers wondered if age could be accurately determined from measures of tree height and trunk width. For a group of trees, accurate actual age was known, and height and width were measured.

The dataset therefore contains these three variables:

- `Age` (continuous)
- `Height` (continuous)
- `Width` (continuous)

The data is in the file `tree_age1.csv` .

Which of these is correct?

☑   There are no missing values in the dataset.

☐   Error degrees of freedom for a model with both explanatory variables and no interaction term should be 55.

☐   Analyses of this dataset can inform about effects, i.e. causation.

☑   The two explanatory variables are correlated.

🎲 Section 10

Please answer the following question.

🔒 Question 10

Response

(Second question of tree age exercise; single choice; 1 point)
Using the same dataset as in the previous question, what is the adjusted proportion of variance explained when both explanatory variables (but no interaction between them) are included in a linear model?

Give your answer rounded to **two** decimal places using the `round` function.

[   ]

▼ Solution

(Second question of tree age exercise; single choice; 1 point)
Using the same dataset as in the previous question, what is the adjusted proportion of variance explained when both explanatory variables (but no interaction between them) are included in a linear model?

Give your answer rounded to **two** decimal places using the `round` function.

[0.22   ]

🎲 Section 11

Please answer the following question.

Question 11

Response

(Third question of tree age exercise; single choice; 1 point)
Using the same dataset as in the previous question, with the goal of being able to accurately predict tree age in mind, which one of these is true?

○ We need to measure both tree height and width.

○ We need only measure tree width.

○ We need only measure tree height.

○ We can use the sum of the tree height and width.

○ We should include an interaction between the two explanatory variables.

▼ Solution

○ We need to measure both tree height and width.

○ We need only measure tree width.

◉ We need only measure tree height.

○ We can use the sum of the tree height and width.

○ We should include an interaction between the two explanatory variables.

🎲 Section 12

Please answer the following question.

🔒 Question 12

Response

(Fourth question of tree age exercise; single choice; 1 point)
Using the same dataset as in the previous question, considering unadjusted R-squared, what amount of the total variance explained by both explanatory variables in combination is shared (i.e. could be attributed to either).

Give your answer rounded to **two** decimal places using the `round` function.

[   ]

▼ Solution

[0.2   ]