

Statistical Methods - II partial test

November 29, 2023 - 1 hour

To allow evaluation, you must submit the text with name and surname.

Name and Surname	Student ID	1
------------------	------------	---

1 Figure 1 shows the residuals $e_i = y_i - \hat{y}_i$ vs the predictor x of a simple linear regression model. Which of the following assumptions of the linear model seems to be violated?

- ☐ None ☐ Linearity ☐ Normality ☐ Homoscedasticity

2 A Poisson GLM relates the number of yearly shark attack with the yearly average global surface temperature (GST). The intercept of the model is $\hat{\beta}_0 = 2.997$, and the predictor coefficient is $\hat{\beta}_1 = 0.005$. Which of the following statements is false?

- ☐ When the GST is of 0 degrees, the expected number of shark attack is 2.997
☐ The expected number of shark attack for a GST of 25 is 22.692
☐ When the GST increases of 1 degree, there is a multiplicative effect of 1.005 on the expected number of shark attack
☐ For a unit increase of the GST, the average number of shark attacks raises of about 0.5%

3 The following code is meant to perform an F test for comparing model M1 and M2.

```
M1 <- lm(y ~ 1)
M2 <- lm(y ~ x + z)
anova(M1, M2)
```

Complete the statement: “The test can be used to assess whether ...”

- ☐ both coefficients of the x and z predictors are statistically equal to zero
☐ both coefficients of the x and z predictors are statistically different from zero
☐ the coefficient of the x predictor is statistically different from zero
☐ the coefficient of the z predictor is statistically different from zero

4 Consider a normal linear model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$, $i = 1, \dots, n$ with $\varepsilon_i \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$. Which of the following statements on the residuals $e_i = y_i - \hat{y}_i$ is false?

- ☐ The sum of the squares of the residuals is the minimum over all possible $(\beta_0, \beta_1, \beta_2)$
☐ The residuals are uncorrelated
☐ The residual sum of squares allow us to evaluate the proportion of variability not explained by the regression model
☐ The residuals sum up to 0

5 Code 2 reports the summary of a logistic model on the variable **type**, representing diabetic subjects, to evaluate the probability of having diabetes (**type**=1). The covariates describe the plasma glucose concentration (**glu**, continuous), the diabetes pedigree function (**ped**, continuous), the age (**age**, continuous), and the obesity condition (**obese**, dichotomous). Construct a 99% approximate confidence interval for the coefficient glu.

- ☐ 0.052; 2.349 ☐ 0.014; 0.048 ☐ 0.013; 0.105 ☐ 0.259; 3.65

6 Code 1 reports the summary of a linear model that describes the professor evaluation **eval** using the beauty score of the professor (**beauty**, continuous) and the gender (**female**, dichotomous). Select the correct value of the residual sum of squares, i.e. $\sum_i (\text{eval}_i - \widehat{\text{eval}}_i)^2$.

- ☐ 0.927 ☐ 131.918 ☐ 0.073 ☐ 246.07

7 Code 2 reports the summary of a logistic model on the variable **type**, representing diabetic subjects, to evaluate the probability of having diabetes (**type**=1). The covariates describe the plasma glucose concentration (**glu**, continuous), the diabetes pedigree function (**ped**, continuous), the age (**age**, continuous), and the obesity condition (**obese**, dichotomous). Compute the odds ratio of having diabetes for a non-obese subject vs an obese one.

- ☐ 0.833 ☐ 0.301 ☐ 1.201 ☐ 3.322

8 The Akaike information criterion is generally used

- ☐ to evaluate the explained variability by the model
☐ to compare non-nested models
☐ to evaluate the coefficients significance
☐ to evaluate model misspecification

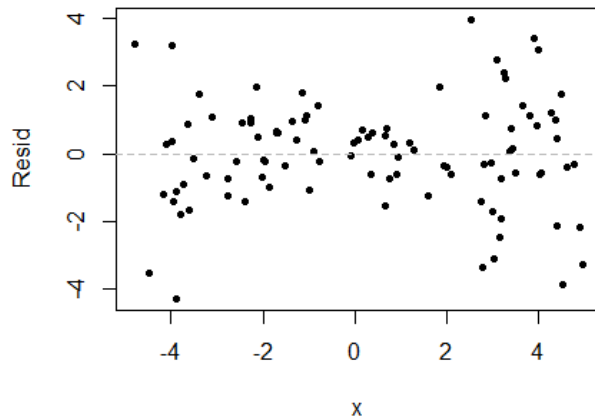
9 Code 1 reports the summary of a linear model that describes the professor evaluation **eval** using the beauty score of the professor (**beauty**, continuous) and the gender (**female**, dichotomous). What is the expected evaluation for a male professor with beauty score -2?

- ☐ 4.104 ☐ 4.074 ☐ 3.899 ☐ 3.703

10 Code 1 reports the summary of a linear model that describes the professor evaluation **eval** using the beauty score of the professor (**beauty**, continuous) and the gender (**female**, dichotomous). With reference to the interaction coefficient, which of the following statement is correct?

- ☐ The probability of observing a value larger than -1.761 is 0.0789
☐ The interaction coefficient is statistically significant at 0.15%
☐ The probability of observing a value larger than -0.113 is 0.0789
☐ The interaction coefficient is statistically significant at 0.05%

Figure 1



Code 1 - Linear regression

```
Call:
lm(formula = eval ~ beauty * female, data = beauty)

Residuals:
    Min       1Q   Median       3Q      Max
-1.83820 -0.37387  0.04551  0.39876  1.06764

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.10364    0.03359  122.158 < 2e-16 ***
beauty         0.20027    0.04333   4.622 4.95e-06 ***
femaleYES     -0.20505    0.05103  -4.018 6.85e-05 ***
beauty:femaleYES -0.11266    0.06398  -1.761  0.0789 .
---
Residual standard error: 0.5361 on 459 degrees of freedom
Multiple R-squared:  0.07256, Adjusted R-squared:  0.0665
F-statistic: 11.97 on 3 and 459 DF, p-value: 1.471e-07
```

Code 2 - Logistic regression

```
Call:
glm(formula = type ~ glu + ped + age + obese, family = binomial,
    data = Pima.tr)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -8.335286    1.177746  -7.077 1.47e-12 ***
glu           0.030631    0.006611   4.634 3.59e-06 ***
ped           1.954256    0.658162   2.969 0.002985 **
age           0.059155    0.017858   3.312 0.000925 ***
obeseYES      1.200593    0.445982   2.692 0.007102 **
---
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 256.41  on 199  degrees of freedom
Residual deviance: 179.13  on 195  degrees of freedom
AIC: 189.13

Number of Fisher Scoring iterations: 5
```