Applied Multivariate Statistics

Exam 3 by RBS 2018-02-05

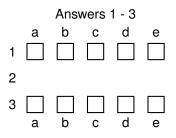## Personal Data

Family Name:

Given Name:

Signature:

checked

Registration Number

	0 1 2 3 4 5 6 7 8 9

In this section **no** changes or modifications must be made!

Type: 205

Exam ID(Statistics): 18020500001

Please mark the boxes carefully: ☒ Not marked: ☐ or ■

This document is scanned automatically. Please keep clean and do not bend or fold. For filling in the document please use a **blue or black pen**.
Only clearly marked and positionally accurate crosses will be processed!

Answers 1 - 3

　　a　b　c　d　e

1 ☐ ☐ ☐ ☐ ☐

2

3 ☐ ☐ ☐ ☐ ☐

　　a　b　c　d　e

In this section **no** changes or modifications must be made!

2 ☐ ☐ ☐ ☐ ☐

Type

999

Exam ID

18020500001

In the following please fill in your answers.

2

1. *(4 points)* What distinguishes the GLM from the linear regression model?

    (a) can model normal error distribution

    (b) can model mean as a function of the variance

    (c) can model variance as function of the mean

    (d) can model non-normal error distribution

    (e) statistical power

2. *(2 points)* The annual mean concentration of sulphur dioxide, in micrograms per cubic metre, is a measure of the air pollution of a city. Related measurements have been taken in 52 US cities along with the following variables that may determine the SO2 levels:

    - `temp`: average annual temperature in Fahrenheit.
    - `manu`: number of manufacturing enterprises employing 20 or more workers.
    - `popul`: population size (1970 census); in thousands.
    - `wind`: average annual wind speed in miles per hour.
    - `precip`: average annual precipitation in inches.
    - `predays`: average number of days with precipitation per year.

    A multiple linear regression model has been fitted that returned the following summary information:

    ```
    ##
    ## Call:
    ## lm(formula = SO2 ~ ., data = final_data)
    ##
    ## Residuals:
    ##     Min      1Q  Median      3Q     Max
    ## -24.439  -7.500   1.420   7.752  23.362
    ##
    ## Coefficients:
    ##             Estimate Std. Error t value Pr(>|t|)
    ## (Intercept) 65.94902   31.76346   2.076  0.04361 *
    ## temp        -0.68571    0.41118  -1.668  0.10233
    ## manu         0.05321    0.01031   5.160 5.39e-06 ***
    ## popul       -0.02820    0.01028  -2.742  0.00874 **
    ## wind        -2.93647    1.44476  -2.032  0.04803 *
    ## precip       0.08808    0.24965   0.353  0.72589
    ## predays      0.18272    0.11353   1.609  0.11452
    ## ---
    ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
    ##
    ## Residual standard error: 10.96 on 45 degrees of freedom
    ## Multiple R-squared:  0.8092, Adjusted R-squared:  0.7838
    ## F-statistic: 31.82 on 6 and 45 DF,  p-value: 1.245e-14
    ```

    How many explanatory variables in the model are statistically significant?

3. *(2 points)* Which statements are correct with respect to PCA?

    (a) eigenvalues give the variance that each principal component represents

    (b) eigenvalues give the variance that each variable explains of the principal component

    (c) the eigenvectors resulting from the analysis are only stretched or shrunk with the eigenvalues when multiplied with the raw data entering the analysis

    (d) the eigenvalue problem is solved by converting the covariance matrix into a matrix with variables that maximise variation

    (e) the eigenvalue problem is solved by converting the covariance matrix into a matrix with linear independent variables