2020

STATISTICS — HONOURS

Paper : CC-12

(Linear Models and Regression)

Full Marks : 50

The figures in the margin indicate full marks.

Candidates are required to give their answers in their own words
as far as practicable.

1. Answer **any ten** questions : 1×10

    (a) State the Gauss Markov theorem.

    (b) Assuming a Gauss Markov set up, write down the estimability condition for a linear parametric function.

    (c) For the one-way ANOVA model, suggest an estimable function of the parameters.

    (d) Mention one drawback of using the standard linear model in case of a binary outcome.

    (e) When is it advisable to use adjusted R^2 rather than R^2 to assess model fit?

    (f) Give an example of a situation where it would be preferable to use a mixed effects model rather than a two-way fixed effects model.

    (g) For the standard linear model, explain why the normal equations are always consistent i.e. solvable.

    (h) For a two-way ANOVA in standard notation, find $\mathrm{cov}\left(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j, \hat{e}_{ij}\right)$.

    (i) Mention any criterion which can be used to choose between alternative logistic models fitted to the same dataset.

    (j) Three treatment groups and one control group are used in an experiment to identify the best treatment for improving depression scores. In addition, information on the age of subjects is also available. Write down a suitable linear model.

    (k) For the standard linear model, with full column rank of the design matrix, write down the expression for an unbiased estimator of variance.

    (l) For a one-way random effects ANOVA, write down the expression for the estimator of variance of the random effects.

    (m) Write down the valid error for testing the interaction term in two-way fixed effects ANOVA model with equal number of observations per cell.

    (n) Suppose a logistic regression of a binary outcome on a single binary covariate leads to an estimated slope of 0.5. Interpret this in terms of Odds Ratio.

    (o) Mention one drawback of using the standard linear model for a count outcome.

2. Answer **any four** questions : 5×4

    (a) For the standard linear model, find the linear predictor for the response corresponding to a new value of the covariate vector together with the prediction interval.

Please Turn Over

(b) Consider a simple linear regression of a continuous response on a single continuous covariate. Given the results from two different slop estimates, corresponding to two different datasets, indicate how to test that the regression lines are parallel. State any underlying assumptions.

(c) Write down the estimators of the group effects and ANOVA table for a one-way classified fixed effects ANOVA model. How do these change if in addition, it is necessary to also account for the effect of an additional continuous predictor?

(d) Suppose you are given data on the number of annual pedestrian deaths and population density for k different Indian cities. Write down a suitable regression equation which could be used to predict the former. Indicate how to estimate the parameters of the model.

(e) (i) Write down the regression equation for a logistic model relating a binary outcome to a single binary covariate.

(ii) Derive the interpretation of the intercept and slope terms in the model.

(iii) How does the interpretation change if a second binary covariate is included in the model?

(f) Consider the standard multiple regression model.

(i) Indicate how to test for the overall fit of the model. State any assumptions necessary for the validity of the test.

(ii) Write the test statistic in (i) in terms of the R^2 criterion.

3. Answer **any two** questions : 10×2

(a) Suppose that for a group of 50 employees, data are collected on the number of days of sick leave in the last six months together with explanatory variables such as age, gender, years of experience and educational level.

(i) Write down a suitable prediction model for the data.

(ii) What is the interpretation of the coefficient corresponding to age?

(iii) How will you test whether educational level is an important explanatory factor?

(b) A researcher wishes to analyse how type of employment (part time/full time) and marital status (single/married/divorced/widowed) of a person affects his total annual claims on health insurance.

(i) Indicate how to perform a univariate test of whether employment type affects total claims. Mention all underlying assumptions.

(ii) Write down a two-way ANOVA model which the researcher can use to simultaneously test for the effect of both employment and marital status. Assuming a balanced design, derive the orthogonal splitting of the Sum of Squares necessary for the validity of the F statistics.

(iii) Write down the standard null and alternative hypotheses for the effect of employment status and indicate how to proceed if the null hypothesis is rejected.

(c) Consider a one-way ANOVA layout.

(i) When is a random effects model more appropriate than a fixed effects model?

(ii) Write down the appropriate hypothesis to test for the absence of group effect.

(iii) Derive a suitable test statistic for testing the above hypothesis.

———————————