Y2.1 An investigation is to be carried out into the spreading properties of two brands of paint, Brand A and Brand B. Samples of 5 cans (of the same size) of each type are analysed, and the area of wall covered by the paint in each can is measured (in square metres), with the following results:

Brand A	55.4	53.2	56.0	50.1	51.8
Brand B	49.2	47.9	52.2	50.8	48.3

(i) Test whether the variances of the 2 brands can be considered to be equal. [8]

(ii) Based on your answer to part (i), test the hypothesis that Brand A covers a greater area than Brand B, against the hypothesis that both brands are equally effective. State the probability value of your test statistic. [6]

(iii) It is decided that the assumption of normality is not appropriate. Repeat part (ii) using an appropriate non-parametric test without resampling. [10]

[Total 24]

Y2.2 The aggregate claims X each year, from a portfolio of insurance policies, are assumed to have a normal distribution with unknown mean θ and variance $\tau^2 = 400$. Prior information is such that θ is assumed to have a normal distribution with mean $\mu = 270$ and variance $\sigma^2 = 225$. Independent claim amounts over the past $n = 5$ years have been obtained.

The Bayesian estimate for θ under quadratic loss for claims with a sample mean of $\bar{x}$ is:

$$\frac{\frac{n\bar{x}}{\tau^2} + \frac{\mu}{\sigma^2}}{\frac{n}{\tau^2} + \frac{1}{\sigma^2}}$$

Simulate 1,000 samples of 5 aggregate claims using a seed value of 13 and hence calculate the empirical Bayesian estimate for θ under quadratic loss for each sample. Use these to obtain a single estimate based on all 1,000 samples. [10]

Y2.3 A statistician is analysing the fall in fertility rates that occurred in Switzerland in 1888. The standardised fertility measure for each of 47 French-speaking provinces is obtained along with five other socio-economic factors:

- Agriculture (percentage of males involved in agriculture as occupation)
- Examination (percentage of draftees who received the highest mark on army examination)
- Education (percentage of draftees who received education beyond primary school)
- Catholic (percentage of province who were catholic)
- Infant.Mortality (percentage of live births who lived less than 1 year)

This information can be found in the built-in dataset, `swiss`.

(i) Show that Education has the second strongest Spearman's correlation with Fertility. [6]

(ii) Plot a labelled scattergraph of Fertility against Education. [6]

(iii) (a) Fit a linear regression model, using Fertility as the response variable and Education as the explanatory variable. State the intercept and gradient parameters and comment on their *p*-values.

(b) Plot a red dotted fitted regression line to the scattergraph in part (i).

(c) By considering the coefficient of determination and the fitted line from part (b), explain the limitations of this model. [12]

Neuchatel, the 42nd swiss province has 17.6% of men occupied in agriculture, 35% of draftees receiving the highest mark on the army examination, 32% of draftees receiving education beyond primary school, 16.92% who are catholic and 23.0% of live births who live less than 1 year.

(iv) Calculate the residual and a 90% confidence interval for the fertility rate of the Neuchatel province, based on the model in part (iii). [8]

[continued over]

To improve the model it is decided to use forward selection to add other variables and interaction terms that meet the following criteria:

- Main effects are considered before interaction terms
- Only interactions between education and one other variable (that was included as a main effect) are to be considered
- The variable that most improves the adjusted R^2 out of the remaining possibilities is to be added first
- The variable is then only kept if all the resulting parameters in the model are significant.

(v) Derive the best model for fertility that meets all these criteria, recording your adjusted R^2 for each model considered. Comment on how each model meets the criteria. [26]

(vi) (a) Repeat part (iv) for your model from part (v).

(b) Hence, comment on the fit of the second model compared to the first. [8]

[Total 66]

END OF PAPER