

COMPREHENSIVE WRITTEN EXAMINATION, PAPER III
FRIDAY AUGUST 13, 2021 9:00 A.M.–1:00 P.M.
STOR 664 Questions (100 points in total)

1. Consider the quadratic regression problem

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i, \quad i = 1, \dots, n \quad (1)$$

where β_0 , β_1 and β_2 are unknown parameters and $\epsilon_i \sim N[0, \sigma^2]$ are independent errors with common unknown variance σ^2 . Assume $\sum x_i = \sum x_i^3 = 0$.

(a) Show that the least squares estimator of β_0 is

$$\hat{\beta}_0 = \frac{\sum y_i \sum x_i^4 - \sum x_i^2 y_i \sum x_i^2}{\Delta} \quad (2)$$

where $\Delta = n \sum x_i^4 - (\sum x_i^2)^2$, and derive similar expressions for $\hat{\beta}_1$ and $\hat{\beta}_2$. **[12 points]**

(b) With \bar{y} the usual sample mean, show that

$$\hat{\beta}_0 - \bar{y} = -\frac{\hat{\beta}_2 \sum x_i^2}{n}. \quad (3)$$

[6 points]

(c) Suppose we are interested in testing $H_0 : \beta_1 = \beta_2 = 0$ against the alternative H_1 where β_0 , β_1 and β_2 are all unrestricted. In standard linear models notation, let SSE_0 , SSE_1 be the error (residual) sums of squares under H_0 and H_1 respectively. Show that

$$SSE_0 - SSE_1 = C_1 \hat{\beta}_1^2 + C_2 \hat{\beta}_2^2 \quad (4)$$

and find explicit expressions for the constants C_1 and C_2 (they are functions of x_1, \dots, x_n but not of y_1, \dots, y_n). **[10 points]**

(d) Using the formulas in parts (a), (b) and (c), write down the F test for testing H_0 against H_1 . Your answer should include an explicit formula for the F statistic, a statement of its distribution when H_0 is true, and a formula for calculating the critical value of the test statistic for a significance test of prescribed level α . [You may express your answer in terms of the R function `qf(p,m,n)` for the p -quantile of an $F_{m,n}$ distribution.] **[10 points]**

(e) Show how to calculate the power of this test when β_1 and β_2 take given non-zero values. [Your answer may be expressed in terms of the R function `pf(c,m,n,ncp=lambda)` where c , m , n are the values of the F statistic and its degrees of freedom, and `lambda` (λ) is the noncentrality parameter; you should give explicit formulas for c , m , n and λ but you are not expected to evaluate any of these terms numerically. Note also that your formulas will depend on x_1, \dots, x_n and σ^2 as well as β_1 and β_2 .] **[12 points]**

Question 2 on the next page

2. The Environmental Protection Agency (EPA) is conducting a study on the distribution of particulate matter (PM), a common air pollutant regulated by the EPA. They set up a network of 16 monitors to measure the long-term average atmospheric concentration of PM. Because not all the monitors are in action at the same times, there are different numbers of readings per monitor. Figure 1 shows a map of the monitor locations, where the numbers indicate the number of observations at each monitor. In total, there are 117 observations.

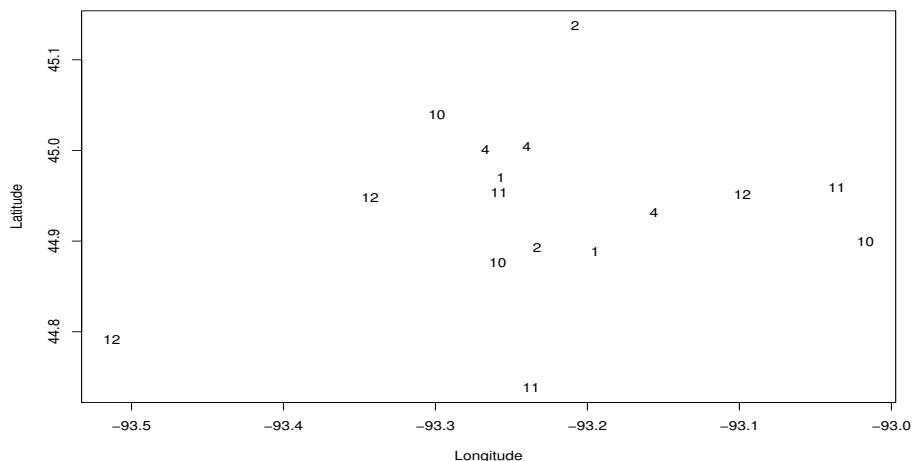


Figure 1. Map of monitor sites. The numbers indicate the number of observations at each monitor.

Based on these 117 observations and assuming the standard assumptions for a linear model, the EPA’s statisticians have fitted the following five models:

- M0: Common mean for each monitor
- M1: Each monitor has a different mean with no predetermined relationship (unbalanced one-way ANOVA experiment)
- M2: The mean at each monitor is given by a linear function of latitude and longitude
- M3: The mean at each monitor is given by a quadratic function of latitude and longitude
- M4: The mean at each monitor is given by a cubic function of latitude and longitude

A table that shows the error sum of squares (ESS) for each model, the corresponding degrees of freedom (DF), the F statistic against the null model M0, and the associated p-value, is as follows:

Model	ESS	DF	F statistic	p-value
M0	90.983	116	—	—
M1	28.369	101	???	2.6e-19
M2	84.489	114	4.381	???
M3	51.880	???	16.733	2.7e-12
M4	???	107	12.89	1e-13

The table entries marked by ??? are left blank for you to fill in (note that the F statistic and p-value in the first line are not defined because you cannot test a model against itself).

- (a) Fill in the four blanks in this table. [You are not expected to provide numerical answers, but state explicitly the expressions that must be evaluated. For the p-value, your answer will be expressed as a tail probability of an F distribution.] [20 points]

For the model M4 (fitted in R by a command of the form `m4=lm(y~X, ...)`), the command `plot(m4)` produces the plot in Figure 2. In addition, R produces the following output (the three-digit numbers represent the order of observations in the original dataset of which this is a subset):

```
> sort(rstandard(m4))
      112      224      222      219      116      168      149      121      154      161      141
-2.935651346 -2.066516118 -2.009799803 -1.993808519 -1.940971966 -1.805696805 -1.581142489 -1.493646167 -1.463965123 -1.450926171 -1.409126547
      177      214      212      225      181      126      201      227      180      173
-1.320739667 -1.312233868 -1.233646507 -1.228039154 -1.133856699 -1.106239324 -1.037364310 -0.992198013 -0.920763158 -0.869367578 -0.856642779
      188      208      134      152      206      148      131      129      185      191      136
-0.802464401 -0.797691097 -0.777072841 -0.765254116 -0.705360889 -0.687459241 -0.683804300 -0.535833312 -0.525553517 -0.512747757 -0.503267642
      160      125      190      163      157      205      164      146      145      169      183
-0.502645595 -0.495715084 -0.491301018 -0.393569520 -0.393412522 -0.377852767 -0.368493577 -0.306625550 -0.281886182 -0.253398161 -0.241058072
      159      124      143      166      184      142      144      187      209      165      176
-0.227490147 -0.205912242 -0.198733956 -0.186369671 -0.151499669 -0.147665795 -0.109892213 -0.093834428 -0.068925241 -0.063645971 -0.061282298
      216      175      113      167      156      217      203      202      194      195      147
-0.003369503 0.008857137 0.025148754 0.028285294 0.046191037 0.073456237 0.076009409 0.100968972 0.116349036 0.131846108 0.150940531
      130      155      115      158      226      117      170      174      204      137      197
0.164323420 0.180745577 0.207236818 0.231765655 0.271372899 0.271700427 0.282883962 0.297902641 0.331812297 0.333930686 0.338864088
      211      123      182      200      193      135      186      139      140      212      215
0.347091705 0.382976906 0.439042147 0.440691124 0.463335726 0.497328236 0.509976812 0.512830708 0.527525637 0.540600650 0.601413973
      178      198      171      189      196      150      127      213      133      218      207
0.604528526 0.611224134 0.621240772 0.670014832 0.670792573 0.712225153 0.751528475 0.752389851 0.763890745 0.774530398 0.824451396
      132      199      153      122      162      138      179      114      210      120      221
0.917177569 0.985922834 1.042116845 1.091438824 1.094941244 1.291785388 1.319577183 1.322522676 1.360671980 1.377898281 1.852082462
      220      192      118      151      128      111      223
1.863846165 1.899568158 1.970913013 1.975139274 2.139963451 2.423544043 2.566352478

> sort(rstudent(m4))
      112      224      222      219      116      168      149      121      154      161      141
-3.047191105 -2.099153120 -2.039247688 -2.022393415 -1.966815871 -1.825264355 -1.592450057 -1.502395306 -1.471923877 -1.458549726 -1.415723952
      177      119      214      172      225      181      126      201      227      180      173
-1.325401483 -1.316725586 -1.236694654 -1.230992901 -1.135387457 -1.107408809 -1.037737007 -0.992125276 -0.920102851 -0.868367892 -0.855569297
      188      208      134      152      206      148      131      129      185      191      136
-0.801120060 -0.796326158 -0.775624807 -0.763762683 -0.703695018 -0.685755380 -0.682093453 -0.534040528 -0.523749966 -0.510974260 -0.501504308
      160      125      190      163      157      205      164      146      145      169      183
-0.500882972 -0.493960755 -0.489552321 -0.392009939 -0.391853337 -0.376334114 -0.367000546 -0.305323528 -0.280670099 -0.252286988 -0.239994164
      159      124      143      166      184      142      144      187      209      165      176
-0.226479390 -0.204988395 -0.197839629 -0.185526855 -0.150806240 -0.146989126 -0.109383665 -0.093398763 -0.068603927 -0.063349061 -0.060996330
      216      175      113      167      156      217      203      202      194      195      147
-0.003353721 0.008815655 0.025031035 0.028152914 0.045975143 0.073114021 0.075655433 0.100500834 0.115811399 0.131239220 0.150249542
      130      155      115      158      226      117      170      174      204      137      197
0.163574391 0.179926459 0.206307556 0.230738020 0.270194822 0.270521153 0.281664317 0.296630343 0.330428176 0.332539922 0.337458016
      211      123      182      200      193      135      186      139      140      212      215
0.345660621 0.381444618 0.437379874 0.439025590 0.461628852 0.495571917 0.508206153 0.511057127 0.525738888 0.538804871 0.599611340
      178      198      171      189      196      150      127      213      133      218      207
0.602727172 0.609426090 0.619449123 0.668279934 0.669058935 0.710575532 0.749990431 0.750854614 0.762394500 0.773072770 0.823208668
      132      199      153      122      162      138      179      114      210      120      221
0.916491385 0.985792848 1.042539885 1.092424712 1.095969956 1.295879435 1.324215588 1.327220500 1.366169814 1.383776172 1.873685530
      220      192      118      151      128      111      223
1.885984493 1.923380018 1.998289515 2.002736439 2.177036564 2.481255588 2.625827182

> sort(hatvalues(m4))
      122      132      143      156      167      179      190      199      207      216      223      111      169
0.04100220 0.04100220 0.04100220 0.04100220 0.04100220 0.04100220 0.04100220 0.04100220 0.04100220 0.04100220 0.04100220 0.04364829 0.05336682
      181      141      154      165      177      126      136      149      161      173      185      194      203
0.05336682 0.05778113 0.05778113 0.05778113 0.05778113 0.05858191 0.05858191 0.05858191 0.05858191 0.05858191 0.05858191 0.05858191 0.05858191
      211      218      225      116      128      138      151      118      121      131      142      155      166
0.05858191 0.05858191 0.05858191 0.05858191 0.07231535 0.07231535 0.07231535 0.07231535 0.08091843 0.08091843 0.08091843 0.08091843 0.08091843
      178      189      198      206      112      114      124      134      147      159      171      183      192
0.08091843 0.08091843 0.08091843 0.08091843 0.08091843 0.08273334 0.08273334 0.08273334 0.08273334 0.08273334 0.08273334 0.08273334 0.08273334
      201      209      217      224      119      129      139      152      163      175      187      196      204
0.08273334 0.08273334 0.08273334 0.08273334 0.08332363 0.08332363 0.08332363 0.08332363 0.08332363 0.08332363 0.08332363 0.08332363 0.08332363
      213      220      227      117      127      137      150      162      174      186      195      212      219
0.08332363 0.08332363 0.08332363 0.08603352 0.08603352 0.08603352 0.08603352 0.08603352 0.08603352 0.08603352 0.08603352 0.08603352 0.08603352
      226      145      157      168      180      120      130      140      153      164      176      188      197
0.08603352 0.08876747 0.08876747 0.08876747 0.08876747 0.09084161 0.09084161 0.09084161 0.09084161 0.09084161 0.09084161 0.09084161 0.09084161
      205      215      222      144      113      123      133      146      158      170      182      191      200
0.09084161 0.09084161 0.09084161 0.09432233 0.09856016 0.09856016 0.09856016 0.09856016 0.09856016 0.09856016 0.09856016 0.09856016 0.09856016
      208      125      135      148      160      172      184      193      202      210      115      214      221
0.09856016 0.09980983 0.09980983 0.09980983 0.09980983 0.09980983 0.09980983 0.09980983 0.09980983 0.09980983 0.09980983 0.49731930 0.49731930
```

```

> sort(cooks.distance(m4))
      216      175      167      113      156      165      203      176      209      217      187
4.854241e-08 7.130810e-07 3.420668e-06 6.915087e-06 9.122314e-06 2.484140e-05 3.595139e-05 3.752454e-05 4.284908e-05 4.866786e-05 8.003438e-05
194      202      144      195      143      142      147      184      130      155      166
8.423771e-05 1.130355e-04 1.257693e-04 1.636334e-04 1.688627e-04 1.919788e-04 2.054928e-04 2.544851e-04 2.698014e-04 2.876264e-04 3.058045e-04
169      124      159      115      183      158      226      117      211      145      174
3.619900e-04 3.824276e-04 4.667777e-04 4.761819e-04 5.241170e-04 5.873038e-04 6.932188e-04 6.948932e-04 7.496688e-04 7.740557e-04 8.353839e-04
170      204      146      190      137      197      164      163      205      157      136
8.749461e-04 1.000777e-03 1.027972e-03 1.032012e-03 1.049663e-03 1.147351e-03 1.356766e-03 1.407975e-03 1.426561e-03 1.507721e-03 1.576083e-03
123      185      182      200      193      139      186      129      125      135      212
1.603651e-03 1.718638e-03 2.107546e-03 2.123407e-03 2.380294e-03 2.390563e-03 2.448151e-03 2.609826e-03 2.724603e-03 2.742365e-03 2.751000e-03
140      160      191      207      178      198      171      132      215      218      189
2.780561e-03 2.801320e-03 2.874565e-03 2.906161e-03 3.217563e-03 3.289231e-03 3.481006e-03 3.596636e-03 3.614034e-03 3.732999e-03 3.952412e-03
196      131      199      206      173      150      122      213      148      127      152
4.090050e-03 4.116774e-03 4.155999e-03 4.380424e-03 4.566469e-03 4.774985e-03 5.093172e-03 5.145624e-03 5.240020e-03 5.316530e-03 5.323086e-03
134      188      133      188      126      181      208      180      179      227      225      201
5.446386e-03 6.380095e-03 6.434233e-03 6.696434e-03 6.899026e-03 6.957195e-03 7.362605e-03 7.444905e-03 7.706337e-03 8.000135e-03 8.879359e-03
177      153      162      141      138      161      154      149      119      114      172
1.069717e-02 1.085121e-02 1.128545e-02 1.217683e-02 1.300801e-02 1.310001e-02 1.314304e-02 1.555690e-02 1.565217e-02 1.577579e-02 1.672105e-02
120      121      210      116      111      223      118      151      220      220      192
1.897054e-02 1.964214e-02 2.052797e-02 2.344334e-02 2.680720e-02 2.794030e-02 3.028064e-02 3.041064e-02 3.157710e-02 3.176247e-02 3.254578e-02
128      219      224      222      112      214      221
3.569792e-02 3.742004e-02 3.851789e-02 4.035997e-02 7.587564e-02 1.505652e-01 3.393624e-01
>

```

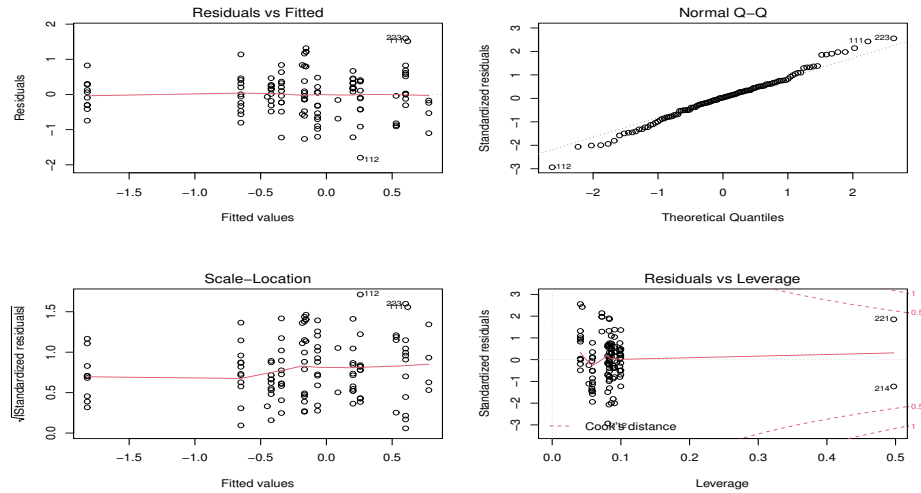


Figure 2: Results of the `plot(m4)` command for model M4.

- (b) What is the distinction between `rstandard` and `rstudent`? State which one would be more suitable for testing the presence of an outlier, and describe how to perform such a test. [6 points]
- (c) For the model M4, show how to interpret the above diagnostics with specific reference to (i) presence of outliers, (ii) suitability of the linear model, (iii) normal distribution of errors, (iv) constancy of variance, (v) points of high leverage, (vi) influential values. [As in earlier parts of the exam, you are not expected to make explicit numerical calculations, but if you refer to any calculated statistics or tests, you should state the appropriate formulas.] [24 points]