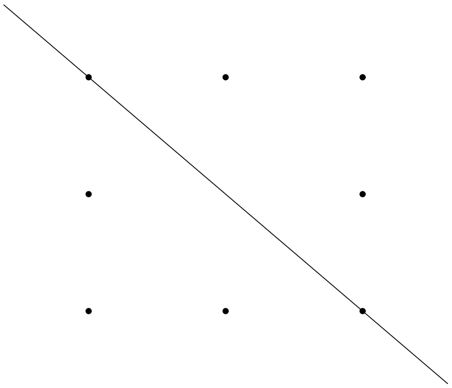
STOR 664: FALL 2021
Final Exam, December 9, 2021

This is an open-book, remote-learning exam. Access to course materials and standard computational tools (in particular, R) is allowed; communication with other students or with anybody via the internet, other than the instructor, is not. The university Honor Code is in effect at all times. Answers may be given in any of the following formats (including combinations of more than one): R Markdown, Word, Latex or handwritten pages scanned or photographed; if handwritten, it is recommended you use blue or black ink on plain sheets of white paper. They should then be uploaded in gradescope. The exam is worth 100 points total, 50 for each of the two questions. Points for each part-question are stated below. Although the questions are intended to be answered in sequence, you may write out your answers in any order and errors in one part-question will not prevent you gaining full credit in other parts of the same question. Attempt all questions.

1. Consider a response surface design with the center point removed:



The diagonal line is to indicate that we will be interested in predicting the surface at points $(t, -t)$ where t is arbitrary.

(a) Consider the standard response surface model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{12} x_{i1} x_{i2} + \beta_{22} x_{i2}^2 + \epsilon_i, \ i = 1, \ldots, 8 \qquad (1)$$

with $\epsilon_i \sim N[0, \sigma^2]$ (mutually independent) as usual. Here, x_{i1} takes the values $(-1, 0, 1, -1, 1, -1, 0, 1)$ in some order, and x_{i2} is similar.

Writing this model in the form $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, specify X, $X^T X$ and $(X^T X)^{-1}$. (For the inverse, you are allowed and encouraged to use R or some other language for matrix computations, such as matlab.) **[5 points]**

(b) Consider the same model with β_{12} removed:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \epsilon_i,\ i = 1, \ldots, 8. \tag{2}$$

For this model also, specify X, $X^T X$ and $(X^T X)^{-1}$. **[5 points]**

(c) Now suppose we want to predict the response surface at a point $x_1^* = t,\ x_2^* = -t$, assuming the equation (1) with $x_{i1},\ x_{i2},\ \epsilon_i$ replaced by $x_1^*,\ x_2^*,\ \epsilon^*$ with ϵ^* independent of $\epsilon_i,\ i = 1, \ldots, 8$. Show that under model (1), the prediction error variance is $\frac{(27t^4 - 32t^2 + 27)\sigma^2}{12}$, and find the corresponding formula under model (2). **[5 points]**

(d) Suppose model (1) is the true model with $\beta_{12} \neq 0$ but we erroneously use model (2) for the analysis. What will be the bias of the predictor at the point $x_1^* = t,\ x_2^* = -t$? **[5 points]**

(e) Hence show that the mean squared error (sum of squared bias and variance) is smaller when using model (2) than model (1) if and only if $4\beta_{12}^2 < \sigma^2$. **[5 points]**

(f) Now suppose the data as follows:

i	x_{i1}	x_{i2}	y_i
1	–1	–1	0.78
2	–1	0	–0.09
3	–1	1	1.62
4	0	–1	2.03
5	0	1	2.05
6	1	–1	5.69
7	1	0	4.18
8	1	1	4.49

See file `respsurf.csv` on sakai. Either directly from the standard formula for $\hat{\boldsymbol{\beta}}$, or using the `lm` command in R, calculate the estimates $\hat{\boldsymbol{\beta}}$ and write down an estimate for the covariance matrix of $\hat{\boldsymbol{\beta}}$. (There is no need to write out the covariance matrix in full: you can express it in terms of the expression for $(X^T X)^{-1}$ given earlier.) **[5 points]**

(g) Suppose we want to estimate the response function at a point $x_1^* = t,\ x_2^* = -t$, as in the earlier parts of the question. Show that the response function is of the form $A + Bt + Ct^2$ where $A,\ B,\ C$ are functions of $\beta_0, \beta_1, \ldots, \beta_{22}$ (which you should specify). Based on your regression estimates and their estimated covariance matrix, calculate estimates $\hat{A}$ for A, $\hat{B}$ for B, $\hat{C}$ for C, together with the variances and covariances. **[6 points]**

(h) Now suppose we are trying to find a value t^* that minimizes $A + Bt + Ct^2$. Based on your answer to part (g), find an estimate of t^*, an approximation to its standard error using the delta method, and hence an approximate 95% confidence interval for t^*. **[7 points]**

(i) Use Fieller's method to calculate an alternative 95% confidence interval for t^*, and compare with your result in (h). **[7 points]**

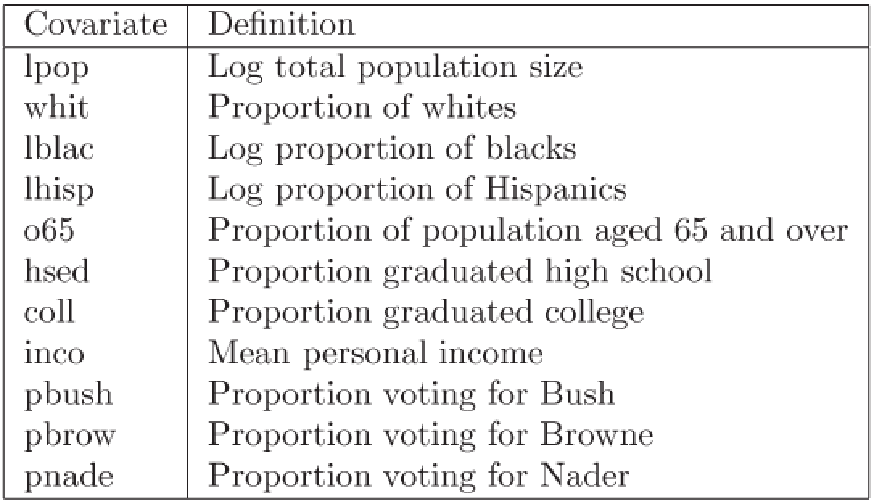
2. This example is based on the presidential election vote in Florida in 2000. The example is discussed extensively in Chapter 6 of Smith & Young, but you don't need to look at that because the analysis you are asked to do here is totally different.

Download the dataset `fldat.csv` from sakai and load it through some command of the form (insert your own path name)

```
fl=read.csv('.../fldat.csv')
```

One of the numerous controversies generated by this election was the vote for the Reform Party candidate Pat Buchanan in Palm Beach County, which at 3,407 votes was far larger than he could have expected based on his overall showing in the election. This was widely attributed to the notorous "butterfly ballot" which allegedly misled voters who intended to vote for the Democrat Al Gore to vote by mistake for Buchanan. In this study, we aim to predict Buchanan's vote in Palm Beach from the votes in the other 65 counties.

The data we are going to examine here consist of the following variables:



Covariate	Definition
lpop	Log total population size
whit	Proportion of whites
lblac	Log proportion of blacks
lhisp	Log proportion of Hispanics
o65	Proportion of population aged 65 and over
hsed	Proportion graduated high school
coll	Proportion graduated college
inco	Mean personal income
pbush	Proportion voting for Bush
pbrow	Proportion voting for Browne
pnade	Proportion voting for Nader

Table 6.12 List of covariates used in the analysis

The response variable will be the "arc sine square root" transformation of the proportion of vote for Buchanan: $y = \arcsin\sqrt{\frac{V_{\text{buch}}}{V_{\text{tot}}}}$ where V_{buch} and V_{tot} are, respectively, the number of votes for Buchanan and the total number of votes cast in a given county. In the case of Palm Beach County, these numbers are respectively 3,407 and 432,286. (An explanation of this transformation is given on page 231 of Smith & Young, but again, it's not necessary for you to look at that. This is different from any of the transformations considered in Chapter 6 of Smith & Young.)

Palm Beach county is row 50 of the dataset. The objective of the exercise is to fit a regression model for y from the other 65 counties (i.e. omitting row 50), and then use that to predict y and hence the number of Buchanan votes in Palm Beach County itself, as if nothing unusual had happened in Palm Beach. All preceding analyses of this dataset produced predicted votes that were well under 3,407: your task here is to see whether that remains true under the proposed new analyses.

For each of parts (a) through (f), show how the indicated method of analysis is applied and use it to predict Buchanan's vote in Palm Beach county based on the other 65 counties. Your prediction should be accompanied by a 95% (or other percent) prediction interval for the methods where such a calculation is appropriate. Finally you are asked to summarize your conclusions.

(a) Simple regression of y on the 11 covariates. Your analysis should include some discussion of model fit and diagnostics but need not go into great detail about that. [**5 points**]

(b) Variable selection, i.e. start with all 11 covariates and reduce them by one of the recognized methods of variable selection. There is no requirement to use a particular method but you should describe carefully what you do. [**8 points**]

(c) Principal components regression. Use cross-validation or some other technique (which you should describe) to determine the number of components. [**8 points**]

(d) Partial least squares regression. Use cross-validation or some other technique (which you should describe) to determine the number of components. [**8 points**]

(e) Ridge regression. Use cross-validation or some other technique (which you should describe) to determine the optimal tuning parameter. [**8 points**]

(f) Lasso regression. Use cross-validation or some other technique (which you should describe) to determine the optimal tuning parameter. [**8 points**]

(g) Briefly summarize your conclusions from this whole exercise. Do all the methods lead to comparable answers? Would any one of them stand out as best in your opinion? [**5 points**]

SOLUTIONS

1. (a) We have

$$X = \begin{pmatrix} 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 0 & 1 & 0 & 0 \\ 1 & -1 & 1 & 1 & -1 & 1 \\ 1 & 0 & -1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}, \; X^TX = \begin{pmatrix} 8 & 0 & 0 & 6 & 0 & 6 \\ 0 & 6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 & 0 \\ 6 & 0 & 0 & 6 & 0 & 4 \\ 0 & 0 & 0 & 0 & 4 & 0 \\ 6 & 0 & 0 & 4 & 0 & 6 \end{pmatrix}, \; (X^TX)^{-1} = \frac{1}{12} \begin{pmatrix} 15 & 0 & 0 & -9 & 0 & -9 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ -9 & 0 & 0 & 9 & 0 & 3 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ -9 & 0 & 0 & 3 & 0 & 9 \end{pmatrix}. \tag{3}$$

(b) Writing X_1 instead of X to keep the two models separate, X_1 is the same as X but with the fifth column deleted, then

$$X_1^TX_1 = \begin{pmatrix} 8 & 0 & 0 & 6 & 6 \\ 0 & 6 & 0 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 \\ 6 & 0 & 0 & 6 & 4 \\ 6 & 0 & 0 & 4 & 6 \end{pmatrix}, \; (X_1^TX_1)^{-1} = \frac{1}{12} \begin{pmatrix} 15 & 0 & 0 & -9 & -9 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ -9 & 0 & 0 & 9 & 3 \\ -9 & 0 & 0 & 3 & 9 \end{pmatrix}. \tag{4}$$

Note that all the entries of $(X_1^TX_1)^{-1}$ are the same as those of $(X^TX)^{-1}$, but with the fifth row and the fifth column deleted. This is because the fifth column of X is orthogonal to all the other columns, therefore, deleting that column does not affect the other entries of $(X^TX)^{-1}$.

(c) The prediction error variance is $(1 + \mathbf{c}^T(X^TX)^{-1}\mathbf{c})\sigma^2$ where

$$\begin{aligned} \mathbf{c}^T(X^TX)^{-1}\mathbf{c} &= \frac{1}{12} \begin{pmatrix} 1 & t & -t & t^2 & -t^2 & t^2 \end{pmatrix} \begin{pmatrix} 15 & 0 & 0 & -9 & 0 & -9 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ -9 & 0 & 0 & 9 & 0 & 3 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ -9 & 0 & 0 & 3 & 0 & 9 \end{pmatrix} \begin{pmatrix} 1 \\ t \\ -t \\ t^2 \\ -t^2 \\ t^2 \end{pmatrix} \\ &= \frac{1}{12} \begin{pmatrix} 15-18t^2 & 2t & -2t & -9+12t^2 & -3t^2 & -9+12t^2 \end{pmatrix} \begin{pmatrix} 1 \\ t \\ -t \\ t^2 \\ -t^2 \\ t^2 \end{pmatrix} \\ &= \frac{15 - 18t^2 + 2t^2 + 2t^2 - 9t^2 + 12t^4 + 3t^4 - 9t^2 + 12t^4}{12} \\ &= \frac{15 - 32t^2 + 27t^4}{12}. \end{aligned}$$

Adding 1 and multiplying by σ^2, we get the formula given.

The corresponding result under model (2) has exactly the same terms except for the fifth row and column, which contributed $+3t^2$ in the second last line above. Therefore, the result in this case is $\frac{(24t^4-32t^2+27)\sigma^2}{12}$.

(d) If model (2) is assumed when model (1) is correct, then each of the estimators $\hat{\beta}_0, \hat{\beta}_1, \ldots$ will be exactly the same as they would be under (1) and therefore unbiased (this is an important part of the answer, and is true because of the orthogonality between the fifth column and the other columns of X). Therefore, the only bias arises from the β_{12} term, where we insert $x_{i1} = t,\ x_{i2} = -t$ in the expected response $\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{12} x_{i1} x_{i2} + \beta_{22} x_{i2}^2$. Hence the bias is $\beta_{12} t^2$.

(e) If the assumed model is (1), then there is no bias and the mean squared prediction error (MSPE) is the same as the prediction error variance, i.e. $\frac{(27t^4-32t^2+27)\sigma^2}{12}$. If the assumed model is (2), then the MSPE is the sum of the prediction error variance and squared bias, i.e. $\frac{(24t^4-32t^2+27)\sigma^2}{12} + \beta_{12}^2 t^4$. Model (2) is preferred when $\frac{(24t^4-32t^2+27)\sigma^2}{12} + \beta_{12}^2 t^4 < \frac{(27t^4-32t^2+27)\sigma^2}{12}$

(f) Fitting by the standard lm command yields the output

```
lm(y~x1+x2+I(x1^2)+I(x1*x2)+I(x2^2))
...
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.94000    0.19600   4.796  0.04083 *
x1             2.00833    0.07157  28.061  0.00127 **
x2            -0.05667    0.07157  -0.792  0.51149
I(x1^2)        1.10500    0.15182   7.278  0.01836 *
I(x1 * x2)    -0.51000    0.08765  -5.818  0.02829 *
I(x2^2)        1.10000    0.15182   7.245  0.01852 *
...
Residual standard error: 0.1753 on 2 degrees of freedom
```

so the coefficient estimates are $\hat{\beta}_0 = 0.94,\ \hat{\beta}_1 = 2.00833, \ldots$ and the estimated covariance matrix of the estimators is $0.1753^2 (X^T X)^{-1}$ where $(X^T X)^{-1}$ is as in (3). [In the event that you chose to fit model (2), the parameter estimates would be the same except for `I(x1 * x2)` and the covariance matrix would be given by $(X_1^T X_1)^{-1} s^2$ with $(X_1^T X_1)^{-1}$ from (4) and now $s = 0.6060436$.]

(g) $\beta_0 + \beta_1 t - \beta_2 t + \beta_{11} t^2 - \beta)12 t^2 + \beta_{22} t^2$ so $A = \beta_0,\ B - \beta_1 - \beta_2,\ C = \beta_{11} - \beta_{12} + \beta_{22}$. Hence the estimators are

$$\begin{pmatrix} \hat{A} \\ \hat{B} \\ \hat{C} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 0.94 \\ 2.00833 \\ -0.05667 \\ 1.105 \\ -0.51 \\ 1.1 \end{pmatrix} = \begin{pmatrix} 0.94 \\ 2.065 \\ 2.715 \end{pmatrix}$$

with estimated covar1ance matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 1 \end{pmatrix} (X^TX)^{-1} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 1 \end{pmatrix}^T s^2$$

$$= \begin{pmatrix} 0.03841667 & 0 & -0.0461 \\ 0 & 0.0102444 & 0 \\ -0.0461 & 0 & 0.06915 \end{pmatrix}.$$

(h) The minimium of $A + Bt + Ct^2$ (assuming $C > 0$) is achieved at $t^* = -\frac{B}{2C}$ so this is estimated as $-\frac{2.065}{2\times 2.715} = -0.3802947$. By the delta method, the variance is $\left(\frac{1}{2C}\right)^2 \text{Var}(\hat{B}) - 2 \cdot \frac{1}{2C} \cdot \frac{B}{2C^2} \cdot \text{Cov}(\hat{B}, \hat{C}) + \left(\frac{B}{2C^2}\right)^2 \text{Var}(\hat{C})$ which is estimated (substituting $\hat{B}$ for B, $\hat{C}$ for C) as $\left(\frac{1}{2\times 2.715}\right)^2 \times 0.0102444 + \left(\frac{2.065}{2\times 2.715^2}\right)^2 \times 0.06905 = 0.0413^2$. Noting also that `qt(0.975,2)=4.303`, we have an approximate 95% confidence interval for t^* which is $-0.3803 \pm 4.303 \times 0.0413 = (-0.558, -0.203)$.

(i) For a given t, a hypothesis test of $H_0: \ B + 2Ct = 0$ would reject when

$$\frac{|\hat{B} + 2\hat{C}t|}{\sqrt{\hat{\text{Var}}(\hat{B}) + 4t^2 \hat{\text{Var}}(\hat{C})}} \quad > \quad qt(0.975, 2) \qquad (5)$$

where $\hat{\text{Var}}$ is estimated variance (we ignore the covariance between $\hat{B}$ and $\hat{C}$ because we know that is 0). Substituting the various estimate, equality in (5) is achieved when

$$(2.065 + 2 \times 2.2715 \times t)^2 \quad = \quad 4.303^2(0.0102444 + 4 \times t^2 \times 0.06915)$$

which reduced to the quadratic equation

$$24.3634t^2 + 22.4259t + 4.07455 \quad = \quad 0$$

and this quadratic equation has roots `(-22.4259+c(-1,1)*sqrt(22.4259*22.4259-4*24.3634*4.07455))/(2*24.3634)=-0.6713723` and `-0.2491026`.

So the 95% confidence interval by Fieller's method is (–0.671,–0.249), which does differ somewhat from the delta method interval given in (h).

Remark. I've written out this calculation "longhand" to show exactly how the method works, but the following R code is one way to get the answer much more quickly:

```
lm1=lm(y~x1+x2+I(x1^2)+I(x1*x2)+I(x2^2))
summary(lm1)
B=lm1$coef[2]-lm1$coef[3]
C=lm1$coef[4]-lm1$coef[5]+lm1$coef[6]
est=-B/(2*C)
V=solve(t(X)%*%X)
vbb=V[2,2]+V[3,3]-2*V[2,3]
vcc=V[4,4]+V[5,5]+V[6,6]-2*V[4,5]-2*V[5,6]+2*V[4,6]
```

```
vbc=V[2,4]-V[2,5]+V[2,6]-V[3,4]-V[3,5]+V[3,6]
# delta method
sest=sqrt(vbb/(4*C*C)+vcc*(B/(2*C^2))^2)*summary(lm1)$sigma
print(est+c(-qt(0.975,2),0,qt(0.975,2))*sest)
# Fieller method
q=qt(0.975,2)
s=summary(lm1)$sigma
a=4*(C*C-q*q*s*s*vcc)
b=4*B*C
c=B*B-q*q*s*s*vbb
(-b+c(-1,1)*sqrt(b*b-4*a*c))/(2*a)
```

which gives almost the same answer. Your answer will almost inevitably differ slightly from mine because of rounding errors, but the major difference between the two methods is not due to rounding error.

2. (a) Reading the data frame `fl` as stated in the question, the commands

```
lm1=lm(y~lpop+whit+lblac+lhisp+o65+hsed+coll+inco+pbush+pbrow+pnade,fl,subset=-50)
pr1=predict(newdata=fl[50,],lm1,interval='prediction')
sin(pr1)^2*432286
```

produce a predicted Buchanan vote of 329.8 and a 95% prediction interval of (0,1482) (compared with actual value of 3407). In this case the predicted lower bound of y is negative, which is impossible, so I have set to 0 for the prediction interval calculation.

[Side comment here: in fact this model does not fit the data so well, since a residual plot confirms the residuals are heavily right-skewed. I don't actually think the arc since square root transformation is very good here, which is why I didn't use it in the original paper that I published on this dataset, but I wanted to try something different for an exercise here.]

(b) The simplest approach is just to use the `step` command in R, when

```
lm2=step(lm1)
pr2=predict(newdata=fl[50,],lm2,interval='prediction')
sin(pr2)^2*432286
```

reduces to the model with variables `lpop + whit + lhisp + inco + pnade`, and an estimated Buchanan vote of 276.8 and a 95% prediction interval (0,1219) (the lower bound is 0 for the same reason as in the first analysis). Of course, alternative methods of variable selection are also fully justified.

(c) Use of `pcr(...,validation='LOO')` and `RMSEP` within `library(pls)` leads to

# Comp.	1	2	3	4	5	6	7	8	9	10
Prediction	160	131	231	277	289	205	300	331	249	306
RMSEP ×100	1.649	1.653	1.639	1.590	1.621	1.340	1.292	1.293	1.302	1.338

(d) Use of `plsr(...,validation='LOO')` and `RMSEP` within `library(pls)` leads to

# Comp.	1	2	3	4	5	6	7	8	9	10
Prediction	160	223	332	270	272	236	310	326	249	311
RMSEP ×100	1.649	1.614	1.583	1.553	1.402	1.308	1.290	1.292	1.302	1.339