Stat 511 Exam 1

February 24, 2003 Prof. Vardeman

1. In the calibration of a scientific instrument, "true" values x are known and produce experimental readings y on the instrument. Suppose that we are willing to assume that the mean value of y is proportional to x, so that

$$y_i = x_i b + e_i$$

where for $\mathbf{e} = (e_1, e_2, \ldots, e_n)'$, $\mathrm{E}\,\mathbf{e} = \mathbf{0}$. A particular calibration experiment produces $n = 4$ data points as per

x	3	4	5	6
y	3	6	11	14

Initially suppose that $\mathrm{Var}\,\mathbf{e} = s^2\mathbf{I}$.

a) Find a matrix $\mathbf{P_X}$ so that $\hat{\mathbf{Y}} = \mathbf{P_X}\mathbf{Y}$.

b) By the criterion of "size of the hats, h_{ii}" which of the 4 observations is "most influential" in the fitting of the linear model here?

c) Give 90% two-sided confidence limits for s in the normal version of this model. (No need to simplify.)

d) Give 90% two-sided prediction limits for a new y for $x = 10$. (No need to simplify.)

Now suppose that it is plausible that not only is the mean value of y is proportional to x, but that so too is the standard deviation of y. That is, suppose that $\text{Var}\,\mathbf{e} = s^2 \,\text{diag}(9,16,25,36)$.

e) Give a matrix $\mathbf{T}$ such that $\mathbf{TY}$ follows a Gauss-Markov model. What is the model matrix for $\mathbf{TY}$?

$\mathbf{T}$: Model Matrix:

f) Evaluate an appropriate point estimate of b under these model assumptions.

g) Give a standard error (an estimated standard deviation) for your estimate of b in part f) under these heteroscedastic model assumptions.

2. So-called "mixture experiments" are run to investigate how the composition of a substance (as measured by fractions of it that are of "pure component" types $i = 1, 2, \ldots, r$) affect some physical property y. For example, y might be an octane rating for a gasoline blended from r "pure" components like butane, alkylate, cat cracked, etc. Notice that in a mixture study

$$x_1 + x_2 + \cdots + x_r = 1$$

In this problem, we consider an $r = 4$ component mixture problem. Consider the linear model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

for

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & .5 & .5 & 0 & 0 \\ 1 & .5 & 0 & .5 & 0 \\ 1 & .5 & 0 & 0 & .5 \\ 1 & 0 & .5 & .5 & 0 \\ 1 & 0 & .5 & 0 & .5 \\ 1 & 0 & 0 & .5 & .5 \\ 1 & .3\overline{3} & .3\overline{3} & .3\overline{3} & 0 \\ 1 & .3\overline{3} & .3\overline{3} & 0 & .3\overline{3} \\ 1 & .3\overline{3} & 0 & .3\overline{3} & .3\overline{3} \\ 1 & 0 & .3\overline{3} & .3\overline{3} & .3\overline{3} \\ 1 & .25 & .25 & .25 & .25 \end{pmatrix} = (\mathbf{1} \,|\, \mathbf{x}_1 \,|\, \mathbf{x}_2 \,|\, \mathbf{x}_3 \,|\, \mathbf{x}_4) \text{ and } \boldsymbol{\beta} = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix}$$

a) For an arbitrary composition vector (x_1, x_2, x_3, x_4) (with each $x_i \geq 0$ and $\sum x_i = 1$) argue carefully that the corresponding mean response $b_0 + \sum_i b_i x_i$ is estimable.

b) The parameter b_0 is not estimable in this model. Argue this point carefully.

Now consider a full rank restricted version of the original mixture model of the form

$$\mathbf{Y} = \mathbf{X}^{*}\mathbf{?} = (\mathbf{x}_1 \,|\, \mathbf{x}_2 \,|\, \mathbf{x}_3 \,|\, \mathbf{x}_4)\mathbf{?} + \mathbf{e} \quad \text{for} \quad \mathbf{?} = \begin{pmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \end{pmatrix}$$

c) In this restricted model and the mixture context, what is the interpretation of the parameter g_1? What is $g_1 - g_2$?

g_1:

$g_1 - g_2$:

d) Give a matrix $\mathbf{C}$ and a vector $\mathbf{d}$ so that the hypothesis that "pure component #1 has mean response 3 and simultaneously an 'equal parts mixture of components' has mean response 12" in the form $\mathrm{H}_0 : \mathbf{C?} = \mathbf{d}$.

$\mathbf{C}$:

$\mathbf{d}$:

e) Is the hypothesis in d) testable? Explain.

There is some R output attached to this exam. The first part of it concerns this mixture problem. Use it to help you answer the following questions.

f) For which of the (x_1, x_2, x_3, x_4) mixtures in the data set is the mean of y most precisely estimated? Say why your answer agrees with intuition.

g) Give 90% two sided confidence limits for $g_1 - g_2$. (Plug in, but you need not simplify.)

h) Notice in this model that if $\mathrm{H}_0\!: g_1 = g_2 = g_3 = g_4$ is true, then the fact that in the mixture context $\sum x_i = 1$ implies that $\mathrm{E}\,\mathbf{Y} = g\mathbf{1}$ for some g, that is, the mean response is constant. Give the value of and degrees of freedom for an F statistic for testing this hypothesis.

$F =$ ____________________ $df =$ __________ , ____________

```
> X
          [,1]      [,2]      [,3]      [,4]
 [1,] 1.0000000 0.0000000 0.0000000 0.0000000
 [2,] 0.0000000 1.0000000 0.0000000 0.0000000
 [3,] 0.0000000 0.0000000 1.0000000 0.0000000
 [4,] 0.0000000 0.0000000 0.0000000 1.0000000
 [5,] 0.5000000 0.5000000 0.0000000 0.0000000
 [6,] 0.5000000 0.0000000 0.5000000 0.0000000
 [7,] 0.5000000 0.0000000 0.0000000 0.5000000
 [8,] 0.0000000 0.5000000 0.5000000 0.0000000
 [9,] 0.0000000 0.5000000 0.0000000 0.5000000
[10,] 0.0000000 0.0000000 0.5000000 0.5000000
[11,] 0.3333333 0.3333333 0.3333333 0.0000000
[12,] 0.3333333 0.3333333 0.0000000 0.3333333
[13,] 0.3333333 0.0000000 0.3333333 0.3333333
[14,] 0.0000000 0.3333333 0.3333333 0.3333333
[15,] 0.2500000 0.2500000 0.2500000 0.2500000
> V<-ginv(t(X)%*%X)
> V
            [,1]        [,2]        [,3]        [,4]
[1,]  0.53218391 -0.08850575 -0.08850575 -0.08850575
[2,] -0.08850575  0.53218391 -0.08850575 -0.08850575
[3,] -0.08850575 -0.08850575  0.53218391 -0.08850575
[4,] -0.08850575 -0.08850575 -0.08850575  0.53218391
> H<-X%*%V%*%t(X)
> H
             [,1]        [,2]        [,3]        [,4]        [,5]        [,6]
 [1,]  0.53218391 -0.08850575 -0.08850575 -0.08850575  0.22183908  0.22183908
 [2,] -0.08850575  0.53218391 -0.08850575 -0.08850575  0.22183908 -0.08850575
 [3,] -0.08850575 -0.08850575  0.53218391 -0.08850575 -0.08850575  0.22183908
 [4,] -0.08850575 -0.08850575 -0.08850575  0.53218391 -0.08850575 -0.08850575
 [5,]  0.22183908  0.22183908 -0.08850575 -0.08850575  0.22183908  0.06666667
 [6,]  0.22183908 -0.08850575  0.22183908 -0.08850575  0.06666667  0.22183908
 [7,]  0.22183908 -0.08850575 -0.08850575  0.22183908  0.06666667  0.06666667
 [8,] -0.08850575  0.22183908  0.22183908 -0.08850575  0.06666667  0.06666667
 [9,] -0.08850575  0.22183908 -0.08850575  0.22183908  0.06666667 -0.08850575
[10,] -0.08850575 -0.08850575  0.22183908  0.22183908 -0.08850575  0.06666667
[11,]  0.11839080  0.11839080  0.11839080 -0.08850575  0.11839080  0.11839080
[12,]  0.11839080  0.11839080 -0.08850575  0.11839080  0.11839080  0.01494253
[13,]  0.11839080 -0.08850575  0.11839080  0.11839080  0.01494253  0.11839080
[14,] -0.08850575  0.11839080  0.11839080  0.11839080  0.01494253  0.01494253
[15,]  0.06666667  0.06666667  0.06666667  0.06666667  0.06666667  0.06666667
             [,7]        [,8]        [,9]       [,10]       [,11]       [,12]
 [1,]  0.22183908 -0.08850575 -0.08850575 -0.08850575  0.11839080  0.11839080
 [2,] -0.08850575  0.22183908  0.22183908 -0.08850575  0.11839080  0.11839080
 [3,] -0.08850575  0.22183908 -0.08850575  0.22183908  0.11839080 -0.08850575
 [4,]  0.22183908 -0.08850575  0.22183908  0.22183908 -0.08850575  0.11839080
 [5,]  0.06666667  0.06666667  0.06666667 -0.08850575  0.11839080  0.11839080
 [6,]  0.06666667  0.06666667 -0.08850575  0.06666667  0.11839080  0.01494253
 [7,]  0.22183908 -0.08850575  0.06666667  0.06666667  0.01494253  0.11839080
 [8,] -0.08850575  0.22183908  0.06666667  0.06666667  0.11839080  0.01494253
 [9,]  0.06666667  0.06666667  0.22183908  0.06666667  0.01494253  0.11839080
[10,]  0.06666667  0.06666667  0.06666667  0.22183908  0.01494253  0.01494253
[11,]  0.01494253  0.11839080  0.01494253  0.01494253  0.11839080  0.04942529
[12,]  0.11839080  0.01494253  0.11839080  0.01494253  0.04942529  0.11839080
[13,]  0.11839080  0.01494253  0.01494253  0.11839080  0.04942529  0.04942529
[14,]  0.01494253  0.11839080  0.11839080  0.11839080  0.04942529  0.04942529
[15,]  0.06666667  0.06666667  0.06666667  0.06666667  0.06666667  0.06666667
```

```
            [,13]       [,14]      [,15]
 [1,]  0.11839080 -0.08850575 0.06666667
 [2,] -0.08850575  0.11839080 0.06666667
 [3,]  0.11839080  0.11839080 0.06666667
 [4,]  0.11839080  0.11839080 0.06666667
 [5,]  0.01494253  0.01494253 0.06666667
 [6,]  0.11839080  0.01494253 0.06666667
 [7,]  0.11839080  0.01494253 0.06666667
 [8,]  0.01494253  0.11839080 0.06666667
 [9,]  0.01494253  0.11839080 0.06666667
[10,]  0.11839080  0.11839080 0.06666667
[11,]  0.04942529  0.04942529 0.06666667
[12,]  0.04942529  0.04942529 0.06666667
[13,]  0.11839080  0.04942529 0.06666667
[14,]  0.04942529  0.11839080 0.06666667
[15,]  0.06666667  0.06666667 0.06666667
> Y
      [,1]
 [1,] 11.5
 [2,] 12.5
 [3,]  4.7
 [4,] 16.5
 [5,] 11.3
 [6,]  6.6
 [7,] 13.0
 [8,]  6.4
 [9,] 13.3
[10,] 10.4
[11,] 10.2
[12,] 12.4
[13,] 11.4
[14,] 13.4
[15,]  9.6
> t(Y)%*%(diag(rep(1,15))-H)%*%Y
         [,1]
[1,] 16.47676
> V%*%t(X)%*%Y
          [,1]
[1,] 10.807586
[2,] 11.873103
[3,]  4.466207
[4,] 16.373103
> C<-matrix(c(1,-1,0,0,1,0,-1,0,1,0,0,-1),3,byrow=T)
> C
     [,1] [,2] [,3] [,4]
[1,]    1   -1    0    0
[2,]    1    0   -1    0
[3,]    1    0    0   -1
> t(C%*%V%*%t(X)%*%Y)%*%ginv(C%*%V%*%t(C))%*%(C%*%V%*%t(X)%*%Y)
         [,1]
[1,] 116.4872
> t(Y)%*%(H-(1/15)*matrix(c(rep(1,225)),15))%*%Y
         [,1]
[1,] 116.4872
```