Name: _________________________________

Directions: Answer questions as directed. Please show work. Give expressions for answers where possible, as partial credit may be awarded in cases where expressions are correct, but numerical answers are not.

You may use the back of the page if you need extra space.
An NCSU entomologist selects $N = 30$ homes from a large population of local houses that are similar in age and occupancy, then randomly assigns them to three treatment groups (below) and measures the reductions in trap counts over a two week period.

<table>
<thead>
<tr>
<th>Group</th>
<th>Symbol</th>
<th>Extermination strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C</td>
<td>control (no treatment)</td>
</tr>
<tr>
<td>2</td>
<td>KB</td>
<td>poison applied in kitchen and bathroom</td>
</tr>
<tr>
<td>3</td>
<td>W</td>
<td>poison applied in whole house (same amount as KB)</td>
</tr>
</tbody>
</table>

Let $y_{ij}$ denote the observed reduction for house $j$ receiving strategy $i$. Then

\[
\sum_{i=1}^{3} \sum_{j=1}^{10} (\bar{y}_i - \bar{y}_+)^2 = 71.5 \quad \bar{y}_1^+ = 1.85 \quad s_1^2 = 6.86
\]
\[
\sum_{i=1}^{3} \sum_{j=1}^{10} (y_{ij} - \bar{y}_j)^2 = 216.0 \quad \bar{y}_2^+ = 2.98 \quad s_2^2 = 7.40
\]
\[
\sum_{i=1}^{3} \sum_{j=1}^{10} (y_{ij} - \bar{y}_i)^2 = 144.5 \quad \bar{y}_3^+ = 5.54 \quad s_3^2 = 1.82
\]

(a) Compose an analysis of variance (ANOVA) table with four columns: source of variation, degrees of freedom, sum of squares, and mean square. Test the hypothesis that all three strategies lead to the same average reduction in roach counts. Report a $p$-value, using an $F$-table, or an applet, or software. Draw a brief conclusion.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>2</td>
<td>71.5</td>
<td>35.75</td>
</tr>
<tr>
<td>Error</td>
<td>27</td>
<td>144.5</td>
<td>5.35</td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td>216</td>
<td></td>
</tr>
</tbody>
</table>

In testing $H_0: \mu_C = \mu_{KB} = \mu_W$, we observe $F = 35.75/5.35 = 6.68$ with a $p$-value of .0044 on $df = 2, 27$. Therefore, the 3 sample treatment means differ significantly. That is, there is evidence of a treatment effect is highly significant.

(b) Consider a model for the mean reduction in roach counts that uses multiple linear regression with two indicator variables, one for the conventional strategy, $X_{KB}$ and another for the whole house strategy, $X_W$. This model was fit using PROC REG and output for the regression coefficients is included on the next page.

i. Estimate three mean differences: between $C$ and $W$, between $C$ and $KB$ and between $W$ and $KB$.

<table>
<thead>
<tr>
<th>Difference</th>
<th>Estimate</th>
<th>$\hat{SE}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W - C$</td>
<td>3.69</td>
<td>1.03</td>
</tr>
<tr>
<td>$KB - C$</td>
<td>1.13</td>
<td>1.03</td>
</tr>
<tr>
<td>$W - KB$</td>
<td>3.69-1.13=2.56</td>
<td>1.03</td>
</tr>
</tbody>
</table>

ii. Standard error may also be obtained from

\[
\hat{SE}(\bar{y}_+ - \bar{y}_j^+) = \sqrt{2MS[E]/10}
\]

or

\[
\hat{SE}(\hat{\beta}_W - \hat{\beta}_{KB}) = \sqrt{0,1,-1,0,1,-1}
\]
2. The relationship between fuel efficiency (mpg) and \( p = 10 \) automotive characteristics is modelled using multiple linear regression with a survey of \( n = 32 \) different cars.

<table>
<thead>
<tr>
<th>Variable</th>
<th>meaning</th>
<th>Variable</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td>Weight (lbs/1000)</td>
<td>raxratio</td>
<td>Rear axle ratio</td>
</tr>
<tr>
<td>qmsec</td>
<td>1/4 mile time</td>
<td>cyl</td>
<td>Number of cylinders</td>
</tr>
<tr>
<td>standard</td>
<td>Trans. (0=auto, 1=manual)</td>
<td>straight</td>
<td>engine (0=v, 1=straight)</td>
</tr>
<tr>
<td>disp</td>
<td>Displacement (cu.in.)</td>
<td>gears</td>
<td>Number of forward gears</td>
</tr>
<tr>
<td>hp</td>
<td>Gross horsepower</td>
<td>carbs</td>
<td>Number of carburetors</td>
</tr>
</tbody>
</table>

A multiple linear regression of mpg on all \( p = 10 \) explanatory variables is fit (last page):

\[
\mu(\text{wt}, \text{qmsec}, \text{standard}, \text{disp}, \text{hp}, \text{raxratio}, \text{cyl}, \text{straight}, \text{gears}, \text{carbs}) = \beta_0 + \beta_1 \text{wt} + \beta_2 \text{qmsec} + \beta_3 \text{standard} + \beta_4 \text{disp} + \beta_5 \text{hp} + \beta_6 \text{raxratio} + \beta_7 \text{cyl} + \beta_8 \text{straight} + \beta_9 \text{gears} + \beta_{10} \text{carbs}
\]

In terms of matrices, the model may be written \( Y = X'\beta + E \) where \( X \) is a design matrix and \( E \) is a vector of independent normal errors with constant variance, \( \sigma^2 \).

(a) Specify the dimension of each matrix: \( Y, X \) and \( (X'X)^{-1} \).

\[
dim(Y) = (32 \times 1), \quad dim(X) = (32 \times 11), \quad dim((X'X)^{-1}) = (11 \times 11)
\]

(b) Report the transpose of the vector \( (X'X)^{-1}X'Y \).

\[
\hat{\beta}' = (12.3, -3.7, 0.8, 2.5, .01, -.02, .79, -.11, .32, .66, -.2)
\]

(c) Are any of the partial regression coefficients significant? Pick one of them, report the \( p \)-value and interpret it. (Explain what hypothesis is being tested by this \( p \)-value and clarify what models are being compared. Use level \( \alpha = .05 \))

The \( p \)-value for a test of the partial regression coefficient for weight, (wt), \( \beta_1 = 0 \) is not less than .05, so there is no evidence that mean fuel efficiency depends on weight, after controlling for the other nine predictors.

(d) Consider the hypothesis where efficiency does not depend on any of these \( p = 10 \) characteristics. Give the null hypothesis \( (H_0) \) in terms of regression coefficients. Report the \( F \)-ratio and \( p \)-value for a test of this hypothesis.

\[
H_0 : \beta_1 = \beta_2 = \cdots = \beta_{10} = 0
\]

\[
F = 13.9, p < .0001
\]
(e) Consider a reduced model with only three predictors: \textit{wt} and \textit{qmsec} and \textit{standard}.

i. Conduct a statistical test to compare this model with the full model. In doing so, specify the null hypothesis clearly. Say what level of significance you choose. Draw a conclusion regarding the comparison.

\[ H_0 : \beta_4 = \beta_5 = \cdots = \beta_{10} = 0 \]

\[ F = \frac{(SS[R]_f - SS[R]_r) / (10 - 3)}{MS[E]_f} = \frac{(978.6 - 956.8) / 7}{7.02} = 0.44 (df = 7, 21) \]

After controlling for \textit{wt} and \textit{qmsec} and \textit{standard}, there is no evidence of dependence of mean efficiency on any of the other predictors.

ii. Using the output entitled “MODEL2”, test the hypothesis that efficiency does not depend on any of the 3 characteristics in the reduced model. Report an \( F \)-ratio and \( p \)-value and draw a brief conclusion.

\[ H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \]

\[ F = 52.8, p < .0001, df = 3, 28 \]

Therefore, efficiency is related to at least one of the three predictors.

iii. Using the reduced model, estimate the mean fuel efficiency among cars that weigh 2000 lbs, take 20 seconds for the quarter mile and have an automatic transmission.

\[
\hat{\mu}(wt = 2, qmsec = 20, standard = 0) = \hat{\beta}_0 + \hat{\beta}_1(2) + \hat{\beta}_2(20)
\]

\[
= 9.62 - 3.92(2) + 1.23(20) + 2.94(0)
\]

\[ = 26.4 mpg \]

iv. Provide a matrix expression that may be evaluated to obtain a standard error for your answer to part (e), iii.

\[
\text{SE}(\hat{\mu}) = \sqrt{MS[E](1, 2, 20, 0)(X'X)^{-1} \begin{pmatrix} 1 \\ 2 \\ 20 \\ 0 \end{pmatrix}}
\]

v. Fill in the two missing elements in the estimated variance-covariance matrix of the regression coefficients, \( \Sigma \) given in the output (AAAA and BBBB).

\[ AAAA = (1.41)^2, BBBB = -6.867. \] (First is \( SE(\hat{\beta}_2)^2 \), second is by symmetry.

(f) Think about a plot of the observed efficiencies against the predicted values. What is the squared correlation from such a plot? \( r^2 = 957/1126 = .85 \)

(g) It seems the reduced model enables us to detect some important explanatory variables but the full model does not. What is the problem with the full model? Use between 1 and 20 words in your answer. The full model is overfit and there is multicollinearity.
proc reg;
   model mpg=wt qmsec standard disp hp raxratio cyl straight gears carbs;
   model mpg=wt qmsec standard/covb;
run;

The REG Procedure
   Model: MODEL1

              Source       DF      Sum of    Mean
                 Squares    Square    F Value    Pr > F
Model          10     978.55276  97.85528    13.93    <.0001
Error          21     147.49443   7.02354
Corrected Total 31      1126.04719

         Parameter     Standard
      Variable      DF    Estimate      Error   t Value    Pr > |t|
Intercept       1   12.30337   18.71788    0.66    0.5181
        wt           1   -3.71530   1.89441   -1.96    0.0633
       qmsec         1    0.82104   0.73084    1.12    0.2739
      standard      1    2.52023   2.05665    1.23    0.2340
         disp         1    0.01334   0.01786    0.75    0.4635
        hp           1   -0.02148   0.02177   -0.99    0.3350
       raxratio      1    0.78711   1.63537    0.48    0.6353
        cyl          1   -0.11144   1.04502   -0.11    0.9161
straight        1    0.31776   2.10451    0.15    0.8814
       gears         1    0.65541   1.49326    0.44    0.6652
       carbs         1   -0.19942   0.82875   -0.24    0.8122

Model: MODEL2

              Source       DF      Sum of    Mean
                 Squares    Square    F Value    Pr > F
Model          3     956.76126 318.92042   52.75    <.0001
Error          28     169.28593  6.04593
Corrected Total 31      1126.04719

         Parameter     Standard
      Variable      DF    Estimate      Error   t Value    Pr > |t|
Intercept       1    9.61778    6.95959    1.38    0.1779
        wt           1   -3.91650    0.71120   -5.51    <.0001
       qmsec         1    1.22589    0.28867    4.25    0.0002
      standard      1    2.93584    1.41090    2.08    0.0467

Covariance of Estimates

   Variable  Intercept   wt  qmsec  standard
Intercept  48.435934514  -3.681623712  -1.8831754  -6.867614794
  wt       -3.681623712    0.5058077651   0.0976336178   0.7672015854
qmsec      -1.8831754    0.0976336178    0.0833301114   0.2011700148
standard  BBBB BBBBBBB    0.7672015854    0.2011700148    AAAAAA