1. Here are two matrices $E$ and $F$.

\[
E = \begin{pmatrix}
-1.4 & 2 & 1.6 \\
2.2 & -1 & -1.8 \\
-1.2 & 1 & 2.8
\end{pmatrix}, \quad F = \begin{pmatrix}
0.2 & 0.8 & 0.4 \\
0.8 & 0.4 & -0.2 \\
-0.2 & 0.2 & \_
\end{pmatrix}
\]

(a) (8 pts.) Matrix $F$ is the inverse of matrix $E$. Use that information, to fill in the blank in matrix $F$.

(b) (8 pts.) I want to find three numbers $b_0$, $b_1$, and $b_2$ such that $F \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 3 \\ 5 \end{pmatrix}$ where $F$ is the $F$ matrix above. Find any one of them (no need to do them all).

\[
b_0 = \_\_\_\_ \, \text{ or } \, b_1 = \_\_\_\_ \, \text{ or } \, b_2 = \_\_\_\_
\]

(c) (5 pts.) Find, if possible, a linear combination of columns 1 and 2 of matrix $E$ that equals column 3 of matrix $E$. If not possible, explain.

2. Packages of a certain food product have differing amounts of preservative (variable P) and are stored at different temperatures (variable T) for three weeks. At the end of that time the packages are opened and the amount of decomposition (our response variable Y) is measured. The data are analyzed by regression. As usual, the model is

\[
Y = \beta_0 + \beta_1 T + \beta_2 P + e.
\]

Here is part of a regression printout from PROC REG; MODEL Y = T P / COVB;

### Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>___</td>
<td>(__________)</td>
<td>(__________)</td>
<td>(_______)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>45</td>
<td>343.94</td>
<td>7.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>(___)</td>
<td>1904.58</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > |t| |
|----------|----|-------------------|----------------|--------|------|---|
| Intercept| 1  | 40.00000          | 3.01447        | 13.27  | <.0001 |
| T        | 1  | 0.14000           | 0.06056        | 2.31   | 0.0254 |
| P        | 1  | -1.10000          | (__________)   | (_______) | <.0001 |
The COVB option gives \((X'X)^{-1}(MSE)\). Here it is:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Intercept</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>9.0870140218</td>
<td>-0.168720313</td>
<td>-0.110989126</td>
</tr>
<tr>
<td>T</td>
<td>-0.168720313</td>
<td>0.0036674292</td>
<td>0.0005469816</td>
</tr>
<tr>
<td>P</td>
<td>-0.110989126</td>
<td>0.0005469816</td>
<td>0.0064</td>
</tr>
</tbody>
</table>

Questions:

(a) (28 pts.) Fill in the seven blanks above.
   (If an answer depends on a previous one that you missed, I will recompute
   using your previous number)

(b) (6 pts.) How many packages of food _____ did I use in my experiment?

(c) (12 pts.) Thinking about the population of decay values Y for all packages that might
   ever be stored at T=50 degrees with P=10 units of preservative, estimate the mean
   _______ and variance _______ of that population.

(d) (8 pts.) Estimate the change _______ in the mean of Y that you would expect if T is
   increased by 1 (with everything else held the same)

(e) (25 pts.) On my printout, I see a p-value for my temperature coefficient. It is
   Pr>|t|  = 0.0254. Using the usual 5% level test,

   (i) What is the null hypothesis being tested? H0: ________________

   (ii) Do I reject or fail to reject that null hypothesis? (reject ,    fail to reject)

   (iii) Do I have significant evidence of a temperature effect? ( yes,  no )

   (iii) When I look up the “critical value” of t for this test in the t tables, what
         degrees of freedom _____ do I look for?

   (iv) The “critical value” from the t tables and its negative delimit the middle 95%
         of the t distribution. Is the calculated t statistic for temperature greater than
         this critical value or less than this critical value?

1. Put an x in the blank and multiply these out. You should get an identity matrix so, for
   example, first row third column would be 0. You get (-1.4)(0.4)+2(-0.2) + 1.6x=0
   So x=0.6. Because E has an inverse, it is full rank and you cannot express any of its
   columns as a linear combination of others. If Fb=V where V is the given vector then,
because E is the inverse of F, you have EFb=EV , Ib=EV, b=EV so just multiply E times V where V’=(5,3,5) Ans: (7, -1, 11).

2. df must sum. We have 2 “X variables” so 2 for model, 45+2=47 total. This 47 is n-1 so I used 48 packages.

SS must sum, so missing model SS (1561) plus 344 is 1905.
Mean square is SS/df so 1561/2 = 780 and F is 780/MSE = 102.09.
We see that this is strongly significant. We cannot omit both P and T. To test these individually we would look at the t tests. The missing standard error is just the square root of the bottom right element in that “covariance matrix of parameter estimates”. Of course we also know how SAS computed it from MSE and the inverse of the X’X matrix.
In this case the square root of 0.0064 is 0.08 so t = estimate/0.08 = -1.1/0.08 = -13.75.
The p-value is, of course, very small so we cannot eliminate P from the model. The preservative has a statistically significant effect. Such a big t is clearly significant for any t degrees of freedom in the t table. We really don’t need tables or p-values on this one.

The whole idea here is that for a given P and T, the decay Y will be normally distributed with mean on the true but unknown plane that we are trying to estimate. The variance of that population is \( \sigma^2 \) which we do not know, but can estimate (with MSE=7.64).
The mean of that population is 
\[ \beta_0 + \beta_1(T) + \beta_2(P) \]
which we ESTIMATE as 
\[ b_0 + b_1(T) + b_2(P) = 40 +0.14(50)-1.1(10). \]

Now it does NOT ask for the standard error of this estimate. If it did, we know that we would need to compute the square root of \((1,50,10)V(1,50,10)’\) where V is that covariance matrix of the estimates. We could also compute V if we were given the inverse of X’X as well as MSE.

Note also that an INDIVIDUAL package would not decay at exactly the mean of its population. We would need to add an extra MSE before taking the square root to build a prediction interval for an individual future package’s decay. While we could compute any of these last quantities with some effort, they were NOT what was asked for here.

For a unit increase in any of the X variables (all others being held the same) the increase in the mean of Y is just the coefficient on that variable. We have estimates of our coefficients. The answer is 0.14.

\[ Y = \beta_0 + \beta_1 T + \beta_2 P + e. \]
\[ H_0: \beta_1 = 0 \]
I reject because the p-value is 0.0254 (< 0.05)
Because I reject I conclude that there is a (statistically) significant effect of T.
When I look in the t tables, I get a positive number. That number and its negative are the critical values that delimit (“cut off”) the middle 95% of t. That leaves 0.05 in the two tails together. My calculated t and its negative leave a total area 0.0254 in both tails.
combined and therefore must be farther out than the critical values. My calculated t is positive and so must exceed that value from the t tables. The t, as always, has the same degrees of freedom as my MSE, namely 45.