Name: ________________________________

Directions: Please work independently. 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. If you use software, or applets, or tables, explain this usage briefly, indicating what you input into the software in order to get an answer out of it. Code and output follows the statements of problems 1 and 2.

You may use the back of the page if you need extra space.
1. An industrial engineer investigates the effect of three different types of computer mouse on repetitive motion disorders (RMD). She is able to recruit $N = 12$ subjects, sampled from a population of interest, and randomly assign them to the three mouse types for a period of two weeks. Each subject records the amount of time they spend using the mouse ($z$, in hours) and their subjective assessment of RMD pain, $y$.

(a) Ignoring hours of usage $z$ altogether, conduct an $F$-test for a treatment (mouse type) effect on RMD pain. Use level $\alpha = .05$. You may use the fact that $F(.05, 2, 9) = 4.26$. (You will need to recompute $MS(E)$.)

(b) Again, test for a mouse type effect on pain, but this time after controlling for hours of usage in an appropriate analysis of covariance model. Report the $p$-value.

(c) Report a table of adjusted and unadjusted means.
(d) Provide an estimate of the mean difference in RMD pain between the conventional mouse and the roll-track mouse, under both the ANCOVA model of part (b) and the naive one-way ANOVA of part (a).

(e) Report standard errors for the difference in part (d) under both models.
proc glm;
  class mousetype;
  model pain=mousetype hours/solution;
  means mousetype;
run;

The GLM Procedure

Class       Levels       Values
mousetype    3            Reflective Rolltrack conventional

Sum of
Source       DF     Squares    Mean Square   F Value   Pr > F
Model         3  4393.110533  1464.370178  15.50   0.0011
Error         8   755.889467   94.486183
Corrected Total 11   5149.000000

Source       DF     Type I SS    Mean Square   F Value   Pr > F
mousetype    2    2558.000000  1279.000000  13.54   0.0027
hours         1    1835.110533  1835.110533  19.42   0.0023

Source       DF     Type III SS  Mean Square    F Value   Pr > F
mousetype    2      937.155809   468.577904  4.96   0.0397
hours         1      1835.110533  1835.110533  19.42   0.0023

Standard
Parameter       Estimate       Error       t Value   Pr > |t|
Intercept     -78.08712614 B  31.48379717   -2.48   0.0381
mousetype Reflective  16.18465540 B  6.89121393    2.35   0.0468
mousetype Rolltrack   -5.66124837 B  7.75626745   -0.73   0.4863
mousetype conventional 0.00000000 B . .
hours          2.18465540  0.49571948    4.41   0.0023

NOTE: The X'X matrix has been found to be singular, and a generalized
inverse was used to solve the normal equations. Terms whose
estimates are followed by the letter ‘B’ are not uniquely estimable.

Level of        ---------pain---------    ---------hours---------
                Mean     Std Dev     Mean    Std Dev
mousetype       N
Reflective      4   73.00000000  20.4613457  61.75000000  8.99536918
Rolltrack       4   37.50000000  11.9303534  55.50000000  2.38047614
conventional   4   59.00000000  17.3973178  62.75000000  6.44851404
2. In an experiment to study the competitive effect of weed growth on grass, \( N = 24 \) plots are randomly assigned to \( t = 6 \) treatments, with \( n = 4 \) plots per treatment. A fescue of interest is allowed to establish itself in all plots, and then a measured amount of a particular weed is introduced after one year. After another year, the fraction of living material in each plot that is fescue is measured, with the results below:

<table>
<thead>
<tr>
<th>Treatment</th>
<th>( N_1 )</th>
<th>( Y_1 )</th>
<th>( N_2 )</th>
<th>( Y_2 )</th>
<th>( N_3 )</th>
<th>( Y_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>94</td>
<td>83</td>
<td>81</td>
<td>68.5</td>
<td>53</td>
<td>49</td>
</tr>
<tr>
<td>sd</td>
<td>3</td>
<td>3.5</td>
<td>3.5</td>
<td>5.3</td>
<td>3.8</td>
<td>4.4</td>
</tr>
</tbody>
</table>

(a) After pooling data across treatments, estimate the variance of grass fraction for a fixed treatment.

(b) Use Tukey’s procedure to identify those sample treatment means which differ significantly at familywise error rate \( \alpha = .05 \). (You may annotate the table above.)

(c) Explain what familywise error rate means here. Be very brief.
(d) Suppose that not all 15 possible pairwise contrasts are important. Rather, a smaller number $k < 15$ of comparisons are of interest. If a data analyst is trying to choose between the Bonferroni and Tukey procedures, for what values of $k$ is the Bonferroni procedure preferred?

(e) Estimate a contrast comparing grass fraction for the average of treatments 1 and 2 with that for the average of treatments 5 and 6. Report a standard error.
3. Consider a completely randomized experiment with $t = 4$ treatments, and $n$ experimental units randomized to each treatment, for a total of $4n$ observations. For the questions below, assume the factorial effects model $Y_{ij} = \mu + \tau_i + E_{ij}$ where $E_{ij}$ are i.i.d. normal with mean 0, unknown variance $\sigma^2$. The null hypothesis is $H_0: \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$. For parts (a), (b) and (c), circle the correct answer.

(a) For fixed error variance $\sigma^2$ and a given alternative hypothesis, $H_1$, what happens to the power of the experiment to reject $H_0$ in favor of $H_1$ as the sample size $n$ becomes larger?
- it goes up
- it goes down
- it stay the same

(b) For fixed $\sigma$ and $n$, what happens to the type II error rate, $\beta$ as one considers treatment means that are further and further apart (larger and larger $|\tau_i|$)?
- it goes up
- it goes down
- it stay the same

(c) Suppose the smallest alternative you’d like to be able to find is one where
$\mu_1 = 65, \mu_2 = 58, \mu_3 = 64, \mu_4 = 49$ so that $\tau_1 = 6, \tau_2 = -1, \tau_3 = 5, \tau_4 = 10$. Assume a population standard deviation of $\sigma = 12$. With $n = 5$ experimental units per treatment, what is the sampling distribution of the statistic $F = MS(Trt)/MS(E)$ under the smallest alternative described above? (Circle one.)
- Central $F$ with 4 and 5 numerator and denominator degrees of freedom
- Central $F$ with 3 and 16 numerator and denominator degrees of freedom
- Non Central $F$ with 4 and 3 numerator and denominator degrees of freedom
- Non Central $F$ with 3 and 16 numerator and denominator degrees of freedom

(d) Use software or the applet to find the power associated with these settings ($H_1$ above, $\sigma = 12, n = 5$). Does your experiment have a 50-50 or greater chance of detecting differences this big or bigger? (Briefly explain any use of software, including what values were input.)