Biology 4605/7220 Name_____Key_________________
Exam 2b 10 November 2004

For each of the following situations (1 and 2):

(A) Define variables in a tabular format, as follows. name symbol scale
scale = nominal, ordinal, or cardinal
cardinal = interval or ratio scale.

(B) Using the symbols, write a general linear model relating the response variable
to explanatory variable(s) and interaction terms (if appropriate).

(C) Complete the first two columns of the ANOVA table source df

(D) Write a Minitab glm statement to carry out the analysis (omit residuals and fits subcommands)
Fill in the covariate command line only when appropriate.

(E) State the name of the analysis, from the following list.
t-test, one-way ANOVA, two-way ANOVA, three-way ANOVA
paired comparisons, randomized blocks,
hierarchical (nested) ANOVA
regression, multiple regression,
ANCOVA (= at least 1 nominal and at least 1 cardinal scale explanatory variable)
none of the above.

1. Rao (1988 *Statistical Research Methods in the Life Sciences*, Duxbury Press, p283) reports soil bulk density (g cm^{-3}) at 4 sites with continuous grazing, at 4 sites with 2-week grazing and 1-week rest, and at 4 sites with 2-week grazing and 2-week rest.
Does soil compaction depend on grazing practice ? A = [2rows] C = [3rows]

A.

name	symbol	scale
Density	D	cardinal
Grazing	Gr	nominal

C.

source	df
S	2
error	9
total	11

B. _____D___ = _______β_o___+___β_{Gr} * Gr_____________________ + error [2terms]

D. MTB > glm ____D___ = ___ Gr____________________ [2terms]

SUBC> covariate ______blank_________ [1]

E. One-way ANOVA [1]

2. Augustin and Clark (1991 *Journal of Dairy Research* 58: 219-229) investigated the effects of three variables on calcium ion activity (mM) of milk manufactured from powder: the pH of the milk, percent solids (9%, 19.6%, and 26% total solids), and preheat treatment during manufacture (none, low heat, medium heat, high heat, indirect UHT, and direct UHT). pH was measured on each of 36 batches of milk manufactured from powder. Assume that there are no interactive effects, except that preheat treatment interacts with level of solids in its effect on calcium ion activity.

A = [4rows] C = [6rows]

A.

name	symbol	scale
Ca ion activity	Ca	cardinal
pH	PH	cardinal
% Solid	S	nominal
Pre-heat treatment	Tr	nominal

C.

source	df
pH	1
S	2
Tr	5
S*Tr	10
error	17
total	35

B. __Ca__ = $\beta_o + \beta_{pH}$ * pH + β_S * S + β_{Tr} * Tr + β_{S*Tr} * S*TR + error [5terms]

D. MTB > glm ____Ca___ = ___pH S Tr S*Tr ___________ [5terms]

SUBC> covariate ______pH_________ [1]

E. ANCOVA [1]

Compare Quiz 9c in 2002, as follows.

1. Austin and Clarke (1991 *Journal of Dairy Research* 58:219-229) investigated the calcium ion activity in cooled and aged reconstituted and recombined milks. They measured calcium ion activity and pH for 5 samples taken in each of 18 categories resulting from 3 categories of milk composition and 6 categories of heat treatment. Does calcium ion activity depend on pH, composition, and heat treatment? Assign symbols to variables. Assuming no interaction terms, write a general linear model to address this question.

Name	Symbol
Calcium ion activity	___[Ca]__
_milk composition____	__Mtype__
_heat treatment____	___trt____
______pH__________	___pH____

Source	df
Mtype	2
trt	5
pH	1
error	81
total	89

GLM ____ [Ca] = β_o + β_{Mtype}*Mtype + β_{trt}*trt + β_{pH}*pH + error _____

Complete the first two columns of the ANOVA table (above).

3a. Define a symbol for litter size of cavies belonging to strain 13 (Sokal and Rohlf, 1995, p 443), then define a symbol for the observed (sample) mean and the true (population) mean . . . . [3]

N = litter size

observed mean $\overline{N}$

population mean E(N), or μ_N

3b. For the following data, compute the observed mean. ____2.646______ [1]

litter size = [2.36 2.41 2.39 2.85 2.82 2.73 2.58 2.89 2.78] cavies

3c. Using your symbols from 4a, write a probability statement for the confidence limits that include the true mean 95% of the time. [2]

$P\{L_1 \leq \mu_N \leq L_2\} = 1-\alpha = 95\%$ carry symbol from above

3d. To compute the 95% confidence limits on your estimate, which t-value should you use?

t = __2.306___ [1]

```
MTB > invcdf c1;    SUBC> t 8.
     0.0100   -2.8965
     0.0250   -2.3060 <--2.5% in each tail to obtain 95% limit     Why ?        [2]
     0.0500   -1.8595
     0.1000   -1.3968
     0.9000    1.3968
     0.9500    1.8595
     0.9750    2.3060 <--2.5% in each tail to obtain 95% limit
     0.9900    2.8965
```

4. Draw a residual versus fit plot from a single way ANOVA with 5 classes, for which the assumption of homogeneity of error is violated. [2]

Show cone, spindle, or any other pattern other than uniform band [1]

Show 5 stacks of residuals [1]

5a. Complete the following one-way ANOVA table for which 25% of the variability in the response variable is due to treatment effects (control vs one treatment). . . . . [3]

Source	df	SS	MS	F
Tr	1	250	250	16
error	48	750	15.625	
total	49	1000		

5b. Add a regression variable (M = body mass) to the ANOVA table you constructed.
Show the new ANOVA table. . . . . [13]
For the regression variable, df = 1 and SS = 100.
Assume: no interactive effect of explanatory variables on the response variable.
SS due to the treatment variable remains unchanged
total SS remains unchanged

Source	df	SS	MS	F
Tr	1	250	250	18.077
Regr (M)	1	100	100	7.231
error	47	650	13.830	
total	49	1000		

5c. Circle the effect (increase/decrease) of adding a regression variable (M = body mass) to the ANOVA table you completed in 5a. . . . . [3]

increase **decrease** in MS error

increase decrease in F-ratio for Treatment

increase **decrease** in p-value for Treatment