

Instructions: This is an open exam in the sense that you may use your notes, books, your computer, statistical software (e.g., R), and access to the internet. However, you are not allowed to communicate with anyone other than your instructor, who can be reached by phone at 515-294-7754 if necessary. Giving or receiving help on this exam is prohibited. Please write your name and answers on your own paper, scan your answers to a single PDF, and upload the PDF file to the STAT 510 Canvas Assignments page before 11:59 A.M. on May 6.

1. Suppose $y_{ij} \sim N(\mu_i, \sigma_i^2)$ for $i = 1, 2$ and $j = 1, \dots, n_i$, where μ_1 and μ_2 are unknown real parameters, σ_1^2 and σ_2^2 are unknown positive parameters, and n_1 and n_2 are known positive integers. Furthermore, suppose that all y_{ij} values are independent.

(a) Consider the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$. Provide an expression for a test statistic that can be used to test this null hypothesis. Choose your test statistic so that it has an F distribution when the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ is true.

(b) State the distribution of the test statistic in part (a) under the alternative hypothesis

$$H_A : \sigma_1^2 = 5\sigma_2^2.$$

(c) Consider the null hypothesis $H_0 : \mu_1 = \mu_2$. Provide an expression for a test statistic that can be used to test this null hypothesis. Choose your test statistic so that it is approximately distributed according to a t distribution when the null hypothesis $H_0 : \mu_1 = \mu_2$ is true.

(d) Provide an expression for the degrees of freedom associated with the test statistic in part (c).

2. Researchers placed 100 insect traps in a large field. After 48 hours, they collected the traps and counted the number of insects in each trap. For $i = 1, \dots, 100$, let y_i be the number of insects in trap i . Suppose $y_1, \dots, y_{100} \stackrel{iid}{\sim} \text{Poisson}(\lambda)$. Given that the average number of insects per trap was 10.2, use the method depicted on slide 31 of slide set 22 to find a confidence interval for λ with coverage level approximately equal to 0.95.

3. A small field experiment was conducted to compare the yield of three plant genotypes (1, 2, and 3) using a randomized complete block design. The three genotypes were randomly assigned to three plots within each of five blocks. Yield data are provided in the following table.

Block	Genotype 1	Genotype 2	Genotype 3
1	16.3	25.6	22.9
2	19.3	19.8	20.3
3	22.3	32.2	28.1
4	11.1	16.7	14.4
5	17.0	27.8	22.1

Genotype 3 is a hybrid that resulted from a cross between genotypes 1 and 2. The researchers who conducted the experiment would like to know if the mean yield for genotype 3 is different from the average of the mean yields for genotypes 1 and 2. Conduct one test to answer this question. Compute a test statistic, determine a p -value, and state a conclusion.

4. Researchers investigated the effects of 20 different chemical compounds on the level of a protein in the blood of mice. On each of 5 days, 20 mice were randomly assigned to the 20 chemical compounds with one mouse for each compound. Each mouse was injected with its assigned compound, and then

blood samples were taken from each mouse at 4 time points: 1, 2, 3, and 4 hours after injection. The same process was repeated each day with 20 different mice, so a total of 100 mice were used in the experiment. The level of the protein of interest was measured in each of the 400 blood samples.

For $i = 1, \dots, 5$, $j = 1, \dots, 20$, and $k = 1, \dots, 4$, let y_{ijk} be the protein level measurement on day i for chemical compound j at time k . For $i = 1, \dots, 5$, $j = 1, \dots, 20$, and $k = 1, \dots, 4$, consider the model

$$y_{ijk} = \mu_{jk} + d_i + e_{ijk},$$

where μ_{jk} terms are unknown fixed parameters and the other terms are random effects defined as follows. Let $\mathbf{d} = [d_1, \dots, d_5]'$. For $i = 1, \dots, 5$ and $j = 1, \dots, 20$, let $\mathbf{e}_{ij} = [e_{ij1}, \dots, e_{ij4}]'$. Suppose

$$\mathbf{d} \sim N(\mathbf{0}, \sigma_d^2 \mathbf{I}_{5 \times 5}),$$

and

$$\mathbf{e}_{ij} \sim N(\mathbf{0}, \Sigma_e) \text{ for } i = 1, \dots, 5 \text{ and } j = 1, \dots, 20,$$

where σ_d^2 is an unknown positive variance parameter and

$$\Sigma_e = \sigma_e^2 \begin{bmatrix} 1 & \phi & \phi^2 & \phi^3 \\ \phi & 1 & \phi & \phi^2 \\ \phi^2 & \phi & 1 & \phi \\ \phi^3 & \phi^2 & \phi & 1 \end{bmatrix}$$

for some unknown variance parameter $\sigma_e^2 > 0$ and some unknown correlation parameter ϕ . Finally, suppose that \mathbf{d} and $\mathbf{e}_{11}, \dots, \mathbf{e}_{520}$ are all independent. In terms of model parameters, give a simplified expression for the variance of the generalized least squares estimator of each of the following:

- (a) μ_{11}
- (b) $\bar{\mu}_1$
- (c) $\mu_{11} - \mu_{21}$
- (d) $\mu_{11} - \mu_{14}$

5. An experiment was conducted to study the effects of two diets (1 vs. 2) and variable amounts of a drug (0, 2, 5, and 10 units per week) on pig litter size. A total of 72 female pigs were housed in 18 pens with 4 pigs per pen. The 18 pens were randomly divided into two groups of 9 pens each. The pigs in the 9 pens associated with one group received diet 1. The pigs in the other 9 pens received diet 2. The 4 pigs within each pen were randomly assigned to the drug amounts 0, 2, 5, and 10 units per week. During the four weeks leading up to conception and the first six weeks of gestation, pigs were fed their assigned diets and injected weekly with their assigned doses of the drug. After the first six weeks of gestation, pigs were separately housed and treated identically until they gave birth. Litter size (i.e., the number of piglets born) was recorded for each mother pig. Let y_{ijk} be the litter size for the mother pig associated with diet i , pen j , and dose x_k , where $i = 1, 2$, $j = 1, \dots, 9$, $k = 1, 2, 3, 4$, $x_1 = 0$, $x_2 = 2$, $x_3 = 5$, and $x_4 = 10$.

Suppose litter size is appropriately modeled using a generalized linear mixed-effects model with a Poisson response. The R code provided at the end of this exam fits three versions of such a model. Use the code and output to complete parts (a) through (d) below.

- (a) One of the fixed-effect estimates from the fit of model `m1` is labeled `x` in the output and takes the value `0.02038`. Is this value significantly different from zero? Conduct a hypothesis test to answer this question. Compute a test statistic, determine the p -value, and state a conclusion.


```

Formula: y ~ diet + (1 | pen)
      logLik   deviance
-197.3743  394.7486
Random effects:
  Groups Name      Std.Dev.
  pen      (Intercept) 0.09924
Number of obs: 72, groups:  pen, 18
Fixed Effects:
(Intercept)      diet2
      2.5228      0.1524

> ##### Model m1 #####
> m1 = glmer(y ~ diet + x + (1 | pen),
+           family = poisson(link = "log"))
> m1
Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [glmerMod]
Family: poisson ( log )
Formula: y ~ diet + x + (1 | pen)
      logLik   deviance
-194.4344  388.8687
Random effects:
  Groups Name      Std.Dev.
  pen      (Intercept) 0.09924
Number of obs: 72, groups:  pen, 18
Fixed Effects:
(Intercept)      diet2      x
      2.43322      0.15235      0.02038

> ##### Model m2 #####
> m2 = glmer(y ~ diet + drug + diet:drug + (1 | pen),
+           family = poisson(link = "log"))
> m2
Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [glmerMod]
Family: poisson ( log )
Formula: y ~ diet + drug + diet:drug + (1 | pen)
      logLik   deviance
-192.9113  385.8227
Random effects:
  Groups Name      Std.Dev.
  pen      (Intercept) 0.09924
Number of obs: 72, groups:  pen, 18
Fixed Effects:
(Intercept)      diet2      drug2      drug3      drug4
      2.49811      0.07063      -0.08536      -0.04652      0.20479
diet2:drug2  diet2:drug3  diet2:drug4
      0.21251      0.12786      0.00125

```