Final Exam

PSTAT 120C

Summer 2022 Session B

Instructions: This exam is open book and open note and has no strict time limit. You can use any course materials; you are also allowed to use results in the book, lecture notes, and past homework without repeating proofs or derivations. Please do not consult with other students until after the submission deadline has passed.

You may use statistical software, including (but not limited to) R or Python, to help answer these questions, or you may solve them manually. If you do use software, you *must* also submit your code.

By submitting your work, you are acknowledging that your work is entirely your own.

1. A study to determine the effectiveness of a drug, or serum, for the treatment of arthritis resulted in the comparison of two groups, each consisting of 400 arthritic patients. One group was inoculated with the serum, whereas the other received a placebo (an inoculation that appears to contain serum but actually is not active). After a period of time, each person in the study was asked whether their arthritic condition had improved, and the observed results are presented in the accompanying table. The question of interest is: *Do these data present evidence to indicate that the proportion of arthritic individuals who improved differs depending on whether or not they received the drug?*

Condition	Treated	Untreated
Improved	234	148
Not improved	166	252

   (a) Conduct a hypothesis test using the X^2 test statistic, with $\alpha = .05$. Report (i) the null and alternative hypotheses; (ii) the expected cell counts; (iii) the test statistic; (iv) the critical value; (v) the p-value; and (vi) the conclusion.

   (b) Using the Z-statistic, test the hypothesis that the proportion of treated persons who improved is equal to the proportion of untreated persons who improved, with $\alpha = .05$. *Hint:* Express each proportion as a mean. See Section 10.3 of the textbook for a refresher.

   Report (i) the null and alternative hypotheses; (ii) the test statistic; (iii) the critical value; (iv) the p-value; and (v) the conclusion.

   (c) Prove that (assuming α is the same for both tests) the χ^2 statistic X^2 is equivalent to the square of the test statistic Z (Z^2). In other words, prove that the χ^2 test used in part (a) is equivalent to the two-tailed Z-test used in part (b).

   Hint: Use the Z statistic

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}.$$

2. Consider the following model for the responses measured in a randomized block design containing b blocks and k treatments:

$$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij},$$

where:

Y_{ij} = response to treatment i in block j;

μ = overall mean;

τ_i = nonrandom effect of treatment i, where $\sum_{i=1}^{k} \tau_i = 0$;

β_j = random effect of block j, where β_js are independent, normally distributed random variables with $E[\beta_j] = 0$ and $V(\beta_j) = \sigma_B^2$ for $j = 1, 2, ..., b$;

ϵ_{ij} = random error terms where ϵ_{ij}s are independent, normally distributed random variables with $E[\epsilon_{ij}] = 0$ and $V(\epsilon_{ij}) = \sigma_\epsilon^2$ for $i = 1, 2, ..., k$ and $j = 1, 2, ..., b$.

Further assume that the β_js and ϵ_{ij}s are also independent. This model differs from that presented in Section 13.8 of the textbook in that the block effects are assumed to be fixed but unknown constants.

(a) Assuming the model just described is accurate, show that observations taken from different blocks are independent of one another. That is, show that Y_{ij} and $Y_{ij'}$ are independent if $j \neq j'$, as are Y_{ij} and $Y_{i'j'}$ if $i \neq i'$ and $j \neq j'$.

(b) Derive the covariance of two observations from the same block. That is, find $Cov(Y_{ij}, Y_{i'j})$ if $i \neq i'$.

(c) Two random variables that have a joint normal distribution are independent if and only if their covariance is 0. Use the result from part (b) to determine the conditions under which two observations from the same block are independent of one another.

(d) Find the expected value and variance of Y_{ij}.

(e) Let $\bar{Y}_{i*}$ denote the average of all responses to treatment i. Use the model to derive $E\left[\bar{Y}_{i*}\right]$ and $V\left(\bar{Y}_{i*}\right)$.

(f) Calculate the bias of $\bar{Y}_{i*}$. Is it an unbiased estimator of the mean response to treatment i?

3. For a comparison of the academic effectiveness of two junior high schools A and B, an experiment was designed using ten sets of identical twins, where each twin had just completed the sixth grade. In each case, the twins in the same set had obtained their previous schooling in the same classrooms at each grade level. One child was selected at random from each set and assigned to school A. The other was sent to school B. Near the end of the ninth grade, an achievement test was given to each child in the experiment. The results are shown in the accompanying table.

Twin Pair	A	B
1	67	39
2	80	75
3	65	69
4	70	55
5	86	74
6	50	52
7	63	56
8	81	72
9	86	89
10	60	47

(a) Using the sign test, test the hypothesis that the two schools are the same in academic effectiveness, as measured by scores on the achievement test, against the alternative that the schools are not equally effective. What would you conclude with $\alpha = .05$?

(b) Suppose it is suspected that junior high school A has a superior faculty and better learning facilities. Test the hypothesis of equal academic effectiveness against the alternative that school A is superior. What is the p-value associated with this test?

(c) Repeat the test in (a), using the Wilcoxon signed-rank test. Compare your answers.

4. Let $Y_1, Y_2, ..., Y_n$ denote a random sample from an exponentially distributed population with density $f(y|\theta) = \theta e^{-\theta y}$, $0 < y$. (Note that the mean of this population is $\mu = \frac{1}{\theta}$.) Use the conjugate gamma (α, β) prior for θ to find the following:

   (a) The joint density, or $f(y_1, y_2, ..., y_n, \theta)$;

   (b) The marginal density, or $m(y_1, y_2, ..., y_n)$;

   (c) The posterior density for $\theta| (y_1, y_2, ..., y_n)$.