1. Twenty patients participate in a one-year clinical trial to compare four drugs on the effects of diastolic blood pressure. In random order, each patient takes each drug separately for 3 months and the patient's diastolic blood pressure is measured at the end of each 3 month period. The experiment results in 4 diastolic blood pressure values for each patient, ie one value for each drug. The data are far from normal and are quite variable both within and across patients.

   (a) Briefly describe the non-parametric method and its procedures that would be most appropriate for this experiment to test equality of medians across the drugs.

   (b) Drugs 1 and 2 are standard blood pressure drugs while 3 and 4 are experimental drugs. The major interest of the study is (i) to determine if the "average of the standard drugs differ from that of the experimental drugs" and (ii) to "compare the experimental drugs". Briefly explain a non-parametric approach that you would use to test these two implied hypotheses.

2. Crime rates per 100,000 for each of the fifty states were analyzed using eigenanalysis on the covariance matrix of seven standardized variables with the following results.

	Eigenvalue	Difference	Proportion	Cumulative
1	4.33346629	3.19570829	0.6191	0.6191
2	1.13775799	0.45615889	0.1625	0.7816
3	0.68159910	0.39666725	0.0974	0.8790
4	0.28493185	0.02602219	0.0407	0.9197
5	0.25890966	0.08318843	0.0370	0.9567
6	0.17572123	0.04810733	0.0251	0.9818
7	0.12761389		0.0182	1.0000

Eigenvectors

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7
Murder	0.299158	0.654158	0.080100	0.591332	-.184727	0.118197	0.279995
Rape	0.418487	0.163720	-.339125	-.001443	0.402966	-.701888	-.167417
Robbery	0.387757	0.065273	0.551609	-.482186	-.445636	-.305669	0.128775
Assault	0.404581	0.296312	-.208856	-.522719	0.297183	0.585827	-.012086
Burglary	0.424644	-.248581	-.171260	0.225639	-.442282	0.183291	-.669652
Larceny	0.351904	-.507383	-.396479	0.068960	-.166001	0.031496	0.654390
Auto_Theft	0.342304	-.366034	0.586651	0.298191	0.541679	0.148890	-.013625

   (a) Summarize the meaning of these results in a way that any educated adult would understand. In your answer be sure to explain how well your summary results fit the data.

   (b) How would you proceed if you wanted to see how a particular state compared with the rest?

3. Let (X, Y) be a bivariate random vector such that $Y \sim$ Gamma $(\alpha, 1), \alpha > 2$ and $X|y \sim$ Exp(y^{-1}), i.e., the probability density function $f_Y(y)$ of Y and the conditional probability density function of X given Y at $Y = y$ are, respectively, given by

$$f_Y(y) = \begin{cases} \frac{1}{\Gamma(\alpha)} y^{\alpha-1} \exp(-y) & \text{if } y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

and for $y > 0$,

$$f_{X|y}(x) = \begin{cases} y \exp(-xy) & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the marginal pdf of X.

(b) Find EY^r where r is a given real number satifying $r > -\alpha$.

(c) Find $E(XY)$.

(d) Find $E(X - Y)^2$.

(e) Let $U = (X + 1)Y$ and $V = Y$.

i. Using the transformation technique, find the joint probability density function of (U, V).

ii. Show that $U \sim$ Gamma$(\alpha + 1, 1)$.

4. Suppose we observe independent random variables $X_1, X_2, ..., X_n$, and for each $k \geq 1$, X_k is normally distributed with mean $k\theta$ and variance 1, where $-\infty < \theta < \infty$. Let $\hat{\theta}$ be the maximum likelihood estimator of θ.

(a) Find a one dimensional sufficient statistic for θ.

(b) Find $\hat{\theta}$.

(c) Find E$(\hat{\theta})$ and Var$(\hat{\theta})$.

(d) Show that $\hat{\theta})$ is consistent for θ. You may use the fact that

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

(e) Show that the size α likelihood ratio test for

$$H_0 : \theta = 0 \text{ vs } H_1 : \theta \neq 0$$

rejects H_0 when $|\hat{\theta}| > c_1$, where c_1 is a constant determined by n and α. Express c_1 in terms of α and the cumulative distribution function for the standard normal distribution.

(f) Suppose that $Y_1 = X_1 + W_1, ..., Y_n = X_n + W_n$, where $W_1, ..., W_n$ are independent random variables of mean 0 and variance 1, and $(W_1, ..., W_n)$ are independent of $(X_1, ..., X_n)$. Suppose now we can observe $Y_1, ..., Y_n$, but cannot observe $X_1, ..., X_n$. Find a consistent estimator of θ and show that the estimator is consistent.

5. Let Y_{ij} be mutually independent random variables, $i = 1, \ldots, A$ and $j = 1, \ldots, n$, where $Y_{ij} \sim \text{beta}(\theta_i, 1)$. Recall that the pdf of a beta(α, β) is

$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}; \quad 0 < x < 1.$$

Note: You may wish to simplify the pdf for the case when $\beta = 1$.

(a) Derive the MLE of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_A)$.

(b) Derive the asymptotic size α LRT of the hypothesis that $\theta_i = i\tau$. That is $\theta_1 = \tau$, $\theta_2 = 2\tau, \ldots$.

6. In a study of the effects of a heart medication on blood pressure, two pigs were randomly selected from each of four litters. For each litter, one of the pigs was randomly assigned to no medication and the other pig received the medication. Each pig was 20 weeks old at the beginning of the study. The blood pressure was measured for each pig after administering the drug daily for 4 weeks. The data contains two missing points because two animals were affected by a disease. The data table is shown below:

	Litter 1	Litter 2	Litter 3	Litter 4
No medication (i=1)	Y_{11}	Y_{12}	Y_{13}	–
Heart medication (i=2)	Y_{21}	Y_{22}	–	Y_{24}

Consider the model

$$Y_{ij} = \mu + \alpha_i + \gamma_j + \epsilon_{ij}$$

where $\epsilon_{ij} \sim NID(0, \sigma_\epsilon^2)$ and $\gamma_j \sim NID(0, \sigma_\gamma^2)$ and the ϵ_{ij}'s are independent of the γ_j's.

(a) This model equation can also be expressed in a matrix form as $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Z\gamma} + \boldsymbol{\epsilon}$. If $\boldsymbol{Y} = (Y_{11}\ Y_{21}\ Y_{12}\ Y_{22}\ Y_{13}\ Y_{24})^T$ specify the other terms of the model in matrix/vector form.

(b) Give a definition of estimability for a linear combination of parameters

$$c^T\beta = c_1\mu_1 + c_2\alpha_1 + c_3\alpha_2.$$

(c) Using your definition from the previous part, show that $\alpha_1 - \alpha_2$ is estimable. Give an interpretation of $\alpha_1 - \alpha_2$ in the context of the study.

(d) Give a numerical value for a generalized inverse of X^TX.

(e) Use your solution in part (d) to solve the equations $X^TY = X^TX\beta$ and obtain a formula for the least squares estimator for $\alpha_1 - \alpha_2$.

(f) Give a matrix formula for the generalized least squares estimator for $\alpha_1 - \alpha_2$.

(g) Describe how the REML estimators for σ_ϵ^2 and σ_γ^2 would be computed.

(h) Let $S^2 = [(Y_{11} - Y_{12} - Y_{21} + Y_{22})^2]/4$ and let $d = [(Y_{11} + Y_{12}) - (Y_{21} + Y_{22})]/2$.

i. Show that S^2 is an unbiased estimator of σ_ϵ^2.

ii. Show that $F = d^2/S^2$ has a non-central F distribution. Report the df and NCP.

iii. Identify the H_0 that can be tested with $F = d^2/S^2$. Identify the H_a.