For your convenience a table of moment generating functions has been included on the last page of the exam.

1. An entomologist has conducted a study to evaluate the effects of 4 levels of irrigation and two bio-control treatments on bean leaf beetle infestation in soybeans. The experiment is set out in 12 infested fields as a CRD split-plot experiment where the irrigation levels are the whole plot treatments and the bio-control treatments are the split-plot treatments. Each irrigation level was randomized to three different fields and the two bio-control treatments were randomized to half of each field. Counts of bean leaf beetles were obtained for each half field in mid-August. Given the complexity of beetle behavior in this experiment, the entomologist does not want to assume any particular type of distribution for the beetle counts but is sure the data are not normally distributed.

   Briefly describe a non-parametric approach and its procedure that would be most appropriate for this experiment to test for:

   (a) equality of median beetle counts across the irrigation levels.

   (b) equality of the median beetle counts across the bio-control treatments

   (c) Interaction between the bio-control treatments and the irrigation levels.

   Assume the researcher has had a good design of experiments class and is familiar with statistical computing software. Be sure to give enough detail in your answer so the entomologist can compute all the needed quantities herself.

2. Let (X, Y) have a bivariate normal distribution with parameters $\mathrm{E}(X) = 0, \mathrm{E}(Y) = 1, \mathrm{Var}(X) = 1, \mathrm{Var}(Y) = 1$ and $\mathrm{Corr}(X, Y) = 1/2$.

   (a) State the marginal distributions of X and Y

   (b) State the conditional distribution of Y given X

   (c) Find $\mathrm{E}(X^4)$.

   (d) Let $U = X^3 + Y$ and $V = Y$.

      i. Find $\mathrm{Cov}(U, V)$.

      ii. Find the joint pdf of (U, V)

3. Let $X_1, ..., X_n (n \geq 2)$ be i.i.d. observations from a population with mean μ and variance σ^2.

   (a) Show that the sample variance $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ is an unbiased estimator of σ^2

   Assume from now on that $\mu = 0$ and the distribution of X_i is $N(0, \sigma^2)$.

   (b) Let σ_0^2 be a given positive number. Find the likelihood ratio test of size α, $(0 < \alpha < 1)$, for $H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 \neq \sigma_0^2$.

4. Let $X_1, ...X_n$ and $Y_1, ..., Y_n$ be two independent i.i.d. samples from $N(\mu_X, \sigma^2)$ and $N(\mu_Y, \sigma^2)$ respectively. Here $-\infty < \mu_X, \mu_Y < \infty$ and $\sigma^2 > 0$ are unknown parameters. One is interested in testing the hypothesis $H_0 : \mu_Y = 2\mu_X$ against $H_1 : \mu_Y \neq 2\mu_X$. Let $\bar{X}_n, \bar{Y}_n, S_X^2, S_Y^2$ denote the sample means and sample variances respectively.

   (a) Identify the distribution of $W = \bar{Y}_n - 2\bar{X}_n$.

   (b) Let $S^2 = (S_X^2 + S_Y^2)/2$. Show that S^2 is a consistent estimator of σ^2.

   (c) Show that S^2 is also the UMVUE estimator of σ^2.

   (d) Is W independent of S^2? Explain your answer briefly.

   (e) Construct a test statistics T for the hypothesis of interest based on W and S^2. Specify the critical region of size α.

   (f) Construct a two-sided 95% confidence interval for $\mu_Y - 2\mu_X$.

5. Let $X_1, X_2, ...$ be a sequence of i.i.d. distributed random variables with common probability mass function given by

$$f(x) = \begin{cases} \frac{|x|}{6} & \text{if } x = -1, 2, 3 \\ 0 & \text{otherwise.} \end{cases}$$

   (a) Show that $\mathrm{E}(X_1) = 2$ and $\mathrm{Var}(X_1) = 2$.

   (b) For $n \geq 1$, define the random variable

$$R_n = \sqrt{n}\left(\left[2 + \bar{X}_n\right]^{-1} - \frac{1}{4}\right)$$

   where $\bar{X}_n = (X_1 + ... + X_n)/n$.
   Using the Delta method or otherwise, find the limiting distribution or R_n.

6. The size of baby turtles is affected by nest temperature among many other factors. A study is carried out to measure this effect. Twelve eggs are taking from each of four adult female turtles. The 12 turtle eggs from each female are randomly assigned to four treatment groups (3 to each group) corresponding to temperature controlled nests with temperatures 27, 27.5, 28, 28.5 (degrees Celsius). The data are analyzed as a randomized block experiment; the average birth weight for the 3 offspring in a given treatment (from a given parent) are provided in the table below. Assume a random block effect.

	Treatments			
Parent	27 degrees	27.5 degrees	28 degrees	28.5 degrees
1	7.9	10.9	11.9	14.1
2	8.5	14.0	13.5	15.0
3	7.6	11.4	14.1	11.7
4	5.7	11.0	12.2	11.6

(a) i. Write down the statistical model you would use to analyze the data from the randomized block experiment that is presented in the table Identify each term in the model.

ii. Write out the model in a matrix form and define each term. Is your X matrix full column rank? Why or why not?

iii. According to your model provide 2 quantities that are estimable. Show why. Provide 2 quantities that are not estimable. Show why.

iv. Find a simplified expression for the BLUP of the birth weight of a turtle when nest temperature is 27 degrees.

(b) The following analysis of variance table was produced by SAS's Proc GLM.

source	df	SS
block	3	13.95
temperature	3	85.66
error	9	9.76
total	15	109.38

i. Is there a significant treatment (temperature) effect? Explain your answer.

ii. Was the blocking effective? Explain.

iii. Explain why it would not be appropriate to use individual turtle 'birth weights' as responses for the model that produced this ANOVA table.

(c) Next we try to say a bit more about the nature of the temperature effect.

i. Test a linear contrast in the temperatures and state your conclusion.

ii. Identify (but do not test) a second contrast that is orthogonal to the linear contrast you used.

(d) For this part of the problem we address the biologically important question of whether the temperature effect is the same for all parents.

i. Explain why this question cannot be addressed from the previous analysis.

ii. Recall that we actually have three measurements of each parent/treatment combination. An analysis of variance is run for a different statistical model using the individual turtle measurements as the response instead of the mean of the three turtles. The resulting ANOVA table:

source	df	SS
block	3	41.85
temperature	3	256.98
block*temp	9	29.28
error	32	120.31
total	47	448.42

Use this ANOVA table to address the question of interest.

(e) For this part of the problem suppose that we decide to treat the block (parent) effect as a fixed effect.

i. What does treating the block effect as a fixed effect imply about the conclusions we can draw from our analysis?

ii. How would we test the significance of the treatment effect if the block effect is treated as a fixed effect? Is your answer different from what it would be for a random block effect?

(f) Temperature also has an effect on the proportion of male offspring born to a parent. For the same study we also recorded $Z_{ijk} = 1$ if the kth offspring of parent i in the treatment j was a male and $Z_{ijk} = 0$ for a female.

i. Would it be appropriate to analyze the Z_{ijk}'s using the methodology that was used in previous parts? Explain.

ii. Regardless of your answer, propose another method for analyzing the data to assess the effect of temperature on the gender of baby turtles.

Discrete		Continuous	
Distribution	mgf	Distribution	mgf
Binomial	$\left[pe^t + (1-p)\right]^n$	Uniform	$\frac{e^{t\theta_2} - e^{t\theta_1}}{t(\theta_2 - \theta_1)}$
Geometric	$\frac{pe^t}{1-(1-p)e^t}$	Normal	$\exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)$
Poisson	$\exp\left[\lambda\left(e^t - 1\right)\right]$	Exponential	$(1-\beta t)^{-1}$
Negative Binomial	$\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$	Gamma	$(1-\beta t)^{-\alpha}$
		Chi-square	$(1-\beta t)^{-\nu/2}$