1. The effects of two gasoline additives on improving mileage were examined in the following study. In this study, 600 gallons of gasoline were obtained from a single supplier and divided into four containers, each containing 150 gallons of gasoline. No additive was added to the gasoline in one container(control treatment), 0.5 ounce/gal of additive A was added to another container, 1.0 ounce/gal of additive A was added to the third container, and 1.0 ounce/gal of additive B was added to the fourth container. Twelve cars were used in the study, each car was of the same make, same model and same year. Three cars were assigned to use the gasoline from one container. The number of miles that a car could travel on 50 gallons of gasoline was recorded. This was divided by 50 gallons to get a mileage rate reported in the miles per gallon (mpg). Each car was driven by a different driver, i.e., each of the 12 drivers used in the study drove just one of 12 cars.

   (a) Describe how randomization should be used in this experiment.

   (b) Explain the consequences of applying the randomization that you described in Part(a).

   (c) Observed results (in miles per gallon) are shown in the following table. Each sample mean is the average of the results for three cars.

Additive	Sample Mean	Sample Variance
None(Control)	25	1.8
0.5 oz/gal of additive A	29	2.4
1.0 oz/gal of additive A	31	1.5
1.0 oz/gal of additive B	27	2.3

   Construct an appropriate ANOVA table.

Source of Variation	d.f.	Sum of Squares	Mean Square	F
Corrected Total				

   (d) State the null hypothesis and alternative hypothesis corresponding to the F-test in ANOVA table. Define any parameters or symbols used in this answer.

(e) Construct a set of orthogonal contrasts that would be of interest to the researchers. Give the formula for each contrast, its value, and its sum of squares. Give a one sentence interpretation of each of your contrasts.

2. A researcher who is trying to breed pigs to be resistent to a certain disease wants to compare the mean levels of a certain antibody in the blood of pigs from two different breeds, called breed A and breed B. He plans to take a random sample of n_1 healthy pigs from breed A and measure the antibody levels in blood samples taken from each pig. He will take a second random sample of n_2 healthy pigs from breed B and measure the antibody levels in blood samples taken from each of those pigs.

   (a) Show how you would test the null hypothesis that the mean antibody levels are the same for these two breeds of pigs against the alternative that breed A has a higher mean antibody level than breed B. Give a formula for a t-test, its degree of freedom, and indicate how you would make a decision. You may assume that the population variances for antibody levels are the same for these two breeds.

   (b) The researcher wants to determine how many pigs should be included in the study. He intended to examine $n_1 = n_2 = n$ pigs from each breed, and perform the t-test in part (a) with an $\alpha = .05$ type I error level. He wants the probability of obtaining a significant difference to be at least 0.90 in the situation where the mean antibody level for breed A is 1 mg/l greater than the mean antibody level for breed B. He knows from previous studies of this type that the standard deviation for antibody levels is about $\sigma = 2$ mg/l for each breed. How large must n be?

   (c) Explain why it is a good idea to use equal sample sizes: $n_1 = n_2 = n$.

   (d) Clearly state the statistical model underlying the use of the t-test in part (a) and (b).

3. In a study of a certain drug used to control pain, six hours were injected with the drug and after 48 hours tissue sample were taken from a target tendon. The concentration of the drug in the sample of tendon tissue was measured in units of mg per kg of tissue. The dosage levels(X) and measured concentration of the drug in tendon tissue (Y) are given below for each horse.

Horse	Dosage (mg) X	Concentration in tendon tissue (mg/kg) Y
1	1	3
2	2	9
3	2	10
4	2	14
5	4	17
6	4	19
Sample Means	2.5	12.0

(a) Find the least squares estimates of the regression coefficients in the model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

where ϵ_i represents a random error.

(b) Complete the following ANOVA table for the model in part (a)

Source of Variation	d.f.	Sum of Squares	Mean Square	F
Regression				
Residual				
Corrected Total				

(c) Use the estimates from part (a) to predict the concentration of drug that will be present in the target tendon 48 hours after another horse is injected with a dosage of $3mg$.

(d) Construct a 95% prediction interval for the potential outcome in part (c).

(e) The construction of the 95% prediction interval in part (d) is based on some assumptions about the distribution of the random errors for the model in part (a). List those properties.

(f) If the model you reported is correct, what can you say about the distribution and other properties of b_1, the least squares estimator of β_1?

4. Perform a lack of fit test for the model considered in problem 1. Use $\alpha = 0.05$ and state your conclusion.

5. Consider a multiple regression model in which 4 explanatory variables are used to explain the variation in Y.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \epsilon_i.$$

   (a) Explain how you would interpret β_3 in this model.

   (b) An ANOVA table for the model is shown below. Complete the ANOVA table.

Source of Variation	d.f.	Sum of Squares	Mean Square	F
Model			10.71	
Residual			1.85	
Corrected Total				

   (c) State the null and alternative hypotheses for the F-test in part (b). State your conclusion.

   (d) What is the R^2 value for this model? What does R^2 measure?

   (e) The least squares estimates for the model in part (a) are shown below with their standard errors and t-values. Carefully state the conclusions you would reach from this information.

Variable	Parameter Estimate	Standard Error	p-value for t-test	Type I Sum of Squares
Intercept	5.5	1.6244	.0009	3.57
X_1	0.0514	0.0293	0.083	3.57
X_2	0.0015	0.0019	0.42	36.57
X_3	0.0007	0.0024	0.76	0.53
X_4	0.0163	0.0145	0.26	2.36

6. Children infected with a certain virus develop a red rash on their face, arms, and legs that typically lasts for several weeks. This condition is generally not life threatening, but is can cause severe discomfort and itching. An experiment was performed to determine if the effectiveness of ointments used to treat this condition could be increased by administration of antihistamines.

   The experiment was performed at five clinics that were selected from a larger set of clinics that could accommodate this study. Eight patients were recruited at each clinic and they were randomly assigned to the eight combinations of the ointment and antihistamines factors. two ointments were used, called ointment I and ointment II. Antihistamines was administered in pill form and the levels of the factor were

- no antihistamines (placebo)
- $10mg$ of antihistamines A each day
- $10mg$ of antihistamines B each day
- a combination of $5mg$ of antihistamines A and $5mg$ of antihistamines B each day

After 72 hours of treatment the percent reduction in skin area covered by the rash was recorded for each patient. The observed means for percent reduction in rash area are recorded in the following table. each mean is an average of results for 5 patients.

	no antihistamines	$10mg$ antihistamine A	$10mg$ antihistamine B	$5mg$ of each antihistamines
ointment I	66	69	67	82
ointment II	44	65	65	78

(a) Consider the model

$$Y_{ijk} = \mu + \beta_i + \tau_j + \alpha_k + (\tau\alpha)_{jk} + \epsilon_{ijk},$$

where

$$\beta_i \sim NID\left(0, \sigma_\beta^2\right) \text{ is a random block effect}$$

$$\epsilon_{ijk} \sim NID\left(0, \sigma_\epsilon^2\right) \text{ is a random error}$$

$$\tau_j \text{ is a fixed effect associated with the } j\text{-th ointment}$$

$$\alpha_k \text{ is a fixed effect associated with the } k\text{-th level of the antihistamine factor}$$

and β_i is independent of any ϵ_{ijk} and $\tau_2 = \alpha_4 = 0$ and $(\tau\alpha)_{2k} = 0$ for any k and $(\tau\alpha)_{j4} = 0$ for any j. What do the following quantities represent?

$$\mu, \quad \alpha_3 - \alpha_2, \quad (\tau\alpha)_{12}$$

(b) Compute the value of the sum of squares for the ointment main effects.

(c) What is the degrees of freedom for the error mean squares in the ANOVA table?

7. Thirty six patients were enrolled in a study to examine the relative effectiveness of three different physical therapy program for restoring strength in leg muscles. Nine therapists participated in the study. Each of the program was used by different set of three therapists. each therapist worked with four of the patients, and the patients were randomly assigned to the therapists. A measure of the gain in strength in the leg muscles was recorded for each patient after six months of therapy.

   (a) Report sources of variation and degrees of freedom for the ANOVA table you would use to analyze the data from this experiment.

   (b) Report a formula for the standard error of the difference between the sample means for program A and B, averaging across nurses and patients.

8. An experiment was performed to examine the effects of levels of salt and use of irradiation in preserving pork sausage. A batch of ground pork was divided into three equal parts. The standard level of salt was mixed into one part, twice the standard level of salt was mixed into a second part, and no salt was mixed into the third part. The three parts of the batch were randomly assigned to the levels of salt. After the salt was mixed in, each of these parts was further split into two equal halves and one half was randomly chosen to be subjected to irradiation and nothing further was done to the other half. Each of these halves was made into four sausages. The same procedure was repeated for five different batches of pork, and a total of 120 sausages were made. All of the sausages were stored under the same conditions for 90 days. Then a measurement of the level of activity of certain microbes was made for each of the 120 sausages.

   Show sources of variation and degrees of freedom for the ANOVA table you would construct from these data.