

Applied Statistics Comprehensive Examination

1. Prospective salespeople for an encyclopedia company are now being offered a sales training program. Prior to the existence of the training program, the established standard for average encyclopedia sales was 33 sets per month, with a standard deviation of 8 sets per month.

The training program will be judged effective if the sample mean sales of a random sample of 30 salespeople observed over a one-month period is 35 sets or greater. If the average is less than 35 sets per month the training program will be judged ineffective and will be discontinued.

- (a) Formulate this problem as a statistical test.
 - (b) Find α , the probability of Type I error, for the test you formulated in part (a).
 - (c) Assume that the true mean sales figure after the training program is 35 sets per month. Find β , the probability of Type II error, for the test you formulated in part (a).
2. Suppose 5, 10 and 15 are independent single observations from Poisson distributions with means 2θ , 3θ and 7θ respectively. Find $\hat{\theta}$, the maximum likelihood estimate of θ .
3. In a comparison of the incidence of tumor potential in two strains of rats, 100 rats (50 males, 50 females) were selected from each of two strains and were examined for a period of one year. All of the rats were approximately the same age and were housed and fed under comparable conditions.
- (a) Use the accompanying one-year sample data to test for a difference in the proportions of rats exhibiting tumor potential for the two strains.
 - (b) Construct a 95% confidence interval for the difference in the proportions of rats exhibiting tumor potential for the two strains.

	Strain A	Strain B
Sample size	100	100
Number exhibiting tumor potential	25	15

4. Let X and Y be random variables with uniform distributions defined on the intervals $(-10\sqrt{3}, 10\sqrt{3})$ and $(-30, 30)$ respectively. If X_1, X_2, \dots, X_{100} and Y_1, Y_2, \dots, Y_{100} are independent random samples from these two distributions, respectively, find an approximation to $P(\bar{Y} - \bar{X} < 1)$ using the Central Limit Theorem, where \bar{X} and \bar{Y} are the respective sample means.

5. Consider a completely randomized design for $t = 5$ treatments, with a single covariate x_1 and six observations per treatment.
- Write the complete general linear model under the assumption that the response y is linearly related to the covariate x_1 for each treatment. Identify the parameters in your model.
 - Indicate the relationships among the parameters of the model for each of the following cases; show a graph for each case.
 - The lines are not parallel.
 - The lines are parallel, but do not coincide.
 - The lines are coincident.
6. A study was conducted of the effects of gender and age on learning of a task. The age factor has been defined in terms of three levels, young, middle, and old. The response is the learning time in minutes.

Here are the data. Note that there is only one observation per cell.

		Factor B - Age			Row
Factor A - Gender	Young	Middle	Old	Average	
Male	11	13	18	14	
Female	7	9	14	10	
Column Average	9	11	16	12	

- Draw an interaction plot that graphically illustrates the possible interaction between age and gender.
 - Consider the information contained in your plot. Is the interaction between age and gender important? Do you think it is statistically significant? Give reasons for your answers.
 - Is it possible to conduct a statistical test for the significance of the interaction effect using these data? You do not have to actually carry out the test - just describe how you would test for the significance of the interaction effect.
7. Consider the simple linear regression model

$$y = \beta_0 + \beta_1 x + \epsilon$$

where the intercept β_0 is a known constant that does not need to be estimated.

- Find the least squares estimator of β_1 for this model.
- What is the variance of the estimator you found in part (a)?

Applied Statistics Comprehensive Examination

Monkeys with induced high blood pressure due to a surgical procedure in their kidneys were randomly allocated to one of four antihypertensive drugs. The changes in systolic blood pressure (before – after) are indicated in the following table.

Note: The unequal sample sizes are the result of surgical failures.

Drug 1	Drug 2	Drug 3	Drug 4
15	6	28	19
19	7	30	18
12	10	24	16
21	7	25	21
16	10	29	23
20		28	20
13			18
			17

- (a) What role does randomization play in this experiment?
- (b) Plot the data in an appropriate manner and comment.
- (c) State the appropriate mathematical model for these data and identify all terms in the model.
- (d) Specify the assumptions of the above model. Do these data appear to satisfy the assumptions? If not, take appropriate action.
- (e) Complete the ANOVA table and test the hypothesis that the mean change is the same for all four drugs.
- (f) Estimate the difference in treatment effects for drug 1 and drug 2 (i.e., estimate $\tau_1 - \tau_2$) using a 95% confidence interval.
- (g) Estimate the among monkey (within drug) variance with a 90% confidence interval.
- (h) Evaluate the pattern of missing data. Is the distribution of missing data homogeneous across the four drug groups? If the pattern of missing data is not homogeneous, how does this affect your analysis and conclusions?