The examination consists of 10 short-answer questions. Choose any 10 questions to answer from the 12 questions given. Each question counts 10 points.

The exam is closed-book. You are permitted to consult (both sides of) a notebook-sheet of formulas, and you will be provided with two pages of relevant tables. Use a calculator wherever appropriate (especially wherever it is necessary to calculate precisely in order to find table entries).

**MAIN TOPICS:** (A) Central Limit Theorem and Poisson Limit of Binomial Probabilities, (B) Simple Linear Regression & Correlation, (C) Multinomial Random Variables, Goodness of fit tests, (D) Method of Maximum Likelihood, Goodness of fit with estimated parameters, (E) Hypothesis Testing & Confidence Intervals, one- and two-sample (F) \( t \), two-sample \( t \), ANOVA tests, Multiple Comparisons, (G) Law of large numbers, interpreting simulations, graphics including empirical distribution function, QQplot, histogram, boxplots.

(1). Suppose that we generate 300 independent Normal(6,4) random variables on a computer and store them in a 100 \( \times \) 3 matrix as \( Z_{i,j} \), \( i = 1, \ldots, 100 \), \( j = 1, 2, 3 \). Define

\[
\bar{Z}_i = \frac{1}{3}(Z_{i,1} + Z_{i,2} + Z_{i,3}), \quad S_i^2 = \frac{1}{2} \sum_{j=1}^{3} (Z_{i,j} - \bar{Z}_i)^2
\]

(a) What is the probability distribution of \( S_i^2 \)? What is the probability that \( S_i^2 \geq 3 \)?

(b) What is the probability that at most 6 of the variables \( S_i^2 \) (for \( i = 1, \ldots, 100 \)) exceed 3?

(2). Suppose that a random sample of data \( X_1, \ldots, X_{100} \) is assumed to come from the density \( f(x|\theta) = 2x/\theta^2 \) for \( 0 \leq x \leq \theta \). Find the Method of Moments estimator \( \hat{\theta} \) of \( \theta \) based on these data, and give its variance if the correct value of \( \theta \) is 2.0.

(3). Independent observations \( Y_{i,j} \) for \( i = 1, \ldots, 3 \), \( j = 1, \ldots, 10 \) are assumed to be Normally distributed with means \( \mu_i \) depending upon \( i \) but
not \( j \), and constant variance \( \sigma^2 \). Suppose that for \( i = 1, 2, 3 \), the sample mean \( \bar{Y}_i \) and corresponding sample variance based on \( Y_{i,1}, \ldots, Y_{i,10} \) are respectively 10.7 and 4.2 for \( i = 1 \), 6.4 and 3.5 for \( i = 2 \), and 11.8 and 4.9 for \( i = 3 \).

(a) Find a 95% confidence interval for \( \mu_1 - \mu_3 \).

(b) Construct an ANOVA table, and test at level \( \alpha = 0.05 \) the hypothesis that all of the three means \( \mu_i \) are equal.

(4). Let \( S^2_Y \) be the sample variance based on 100 independent \( \mathcal{N}(2, 4) \) random variables \( Y_i \). Approximate as closely as you can the probability that \( |S^2_Y - 4| \geq 0.47 \).

(5). Observations were made on the number of ovaries formed in each of 1388 female fruit-flies in an experiment on induced sterility. The observed count of flies with 0 ovaries was 1212, with 1 ovary was 118, with the remaining 58 flies developing 2 ovaries. Test the hypothesis that each of 2 ovaries in each fly develops independently of the other ovary, with some probability \( p \) the same for all ovaries and all flies.

(6). A dataset of 21 normally distributed observations (with unknown mean \( \mu \) and variance \( \sigma^2 \)) yield sample mean 87.3, sample variance 14.7.

(a) Find a 90% two-sided confidence interval for each of \( \mu, \sigma^2 \).

(b) Based on the data given, bracket as closely as you can the p-value for the hypothesis test of \( H_0 : \mu = 95.0 \) versus the two-sided alternative.

(7). Suppose that \( (N_1, N_2, N_3) \) is a multinomially distributed vector of random counts based on \( n \) trials and probabilities \( (p, 2p, 1-3p) \). Find the Maximum Likelihood Estimator \( \hat{p} \) of \( p \), and give its asymptotic variance for large \( n \).

(8). Two samples of data each consist of the yield of corn from 15 plots, with corn raised by identical methods; the soil/fertilizer combination was identical within each sample of 15 plots, but different across the two samples. The data are summarized by: sample 1, sample mean and sample standard deviation (in bushels) were 20.5 and 3.3; in sample 2, sample mean and sample standard deviation were 23.5 and 2.5.

(a). Using the method of the two-sample t-test, test whether there is a difference in mean yields between the two types of plots (i.e., those in sample 1 versus sample 2).
(b). Using the method of ANOVA, test whether there is a difference in mean yields between the two types of plots (i.e., those in sample 1 versus sample 2).

(c). Did you require different assumptions in your answers to (a) and (b) ?

(9). Suppose that measurements of 81 independent random variables $X_i$ with density $f_X(x) = \lambda^2 xe^{-\lambda x}$ for $x > 0$ yield sample average $\bar{X} = 0.25$. Give an approximate 95% confidence interval for the positive unknown parameter $\lambda$ based on its maximum likelihood estimator.

(10) Suppose that independent discrete random variable values $Y_i$ have been observed, for $i = 1, \ldots, 64$, and that of these 64 observations, 30 were equal to 0, 25 were equal to 1, and 9 were equal to 2. Find the chi-square statistic value and degrees of freedom for testing the goodness of fit of these data to the model $Y \sim \text{Binom}(2, p)$ for some $p$ (where you must estimate $p$).

(11) Suppose that $(W_j, V_j)$ for $j = 1, \ldots, 100$ are independent pairs of independent $\text{Uniform}[0, 1]$ random variables. Let $M$ be the number of indices $j = 1, \ldots, 100$ for which simultaneously $W_j \geq 0.4$ and $V_j \leq 0.5$, and let $L$ be the number of the $W_j$’s which are $\leq 0.05$. Find the approximate probabilities that (a) $M \leq 38$, and (b) $L \leq 9$.

(12). One problem may well ask you to define terms like empirical d.f. or QQplot or histogram, or to interpret pictures of these types. Review the book material on interpreting plots!