

Applied Statistics Comprehensive Examination**Statistical Methods I & II**

1. (15 points) A pessimistic librarian at Villanova believes that students in recent years have become unlikely to return library books on time, if at all. In a conversation with a coworker, the librarian hypothesizes that only half of the loaned books are returned on time and, of those not returned on time, half are never returned. To test this theory, the coworker obtains a random sample of 80 books loaned out by the library over the past several years and constructs the following table for the number of books in each category:

Returned on time	Returned late	Never returned
46	26	8

Use the data in this table to test the librarian's theory at the 0.05 level.

2. (30 points) The Internal Revenue Service (IRS) wants to know if the amount of federal income tax paid by corporations in the U.S. tends to change the year after an audit (after adjusting for inflation). To answer this question, the IRS considers the following tax payment amounts (in millions) from 7 randomly chosen U.S. corporations the year before and the year after an audit (adjusted for inflation):

Corporation	1	2	3	4	5	6	7	Mean	Std. Dev.
Year before audit	10.0	12.0	0.3	2.4	80.0	71.0	1.2	25.27	34.69
Year after audit	10.4	11.8	1.6	3.6	82.0	71.7	2.1	26.17	34.97
Difference	0.4	-0.2	1.3	1.2	2.0	0.7	0.9	0.90	0.70

- (10 points) Create a 99% confidence interval for the mean difference in the amount of taxes paid by U.S. corporations to the federal government in the years before and after an audit.
- (5 points) Explain what the term "confidence" means in terms of this problem.
- (5 points) Based on the confidence interval in part a, can the IRS conclude at the 0.01 level that, on average, there is a difference in the tax payment amounts for U.S. corporations the year before and after an audit? Explain briefly.
- (10 points) Suppose now that the IRS knows that the true standard deviation of the differences in tax payments is 0.4 million dollars. Then, for a two-sided $\alpha=0.01$ -level test of the mean difference being 0, based on a sample of size of 7, find the power for detecting that, on average, corporations pay 0.2 million more in federal taxes in the year after an audit.

3. (15 points) A pharmaceutical company making a generic version of Tylenol intends to produce 200 milligram (mg) drug capsules. The FDA will shut down the production of this generic version of Tylenol if the standard deviation of mgs per capsule is shown to be more than 1 mg. The FDA obtains a random sample of 16 capsules and finds a sample standard deviation of 1.2 mg.

Should the FDA shut down production? Test the appropriate hypothesis at the 0.10 level. You may assume that mgs/capsule are normally distributed.

4. (30 points) Adhesive force on gummed material was determined under three fixed humidity conditions and three fixed temperature conditions. Four readings were made under each set of conditions. The experiment was completely randomized, and partial results are in the ANOVA table below.

Source	df	SS	MS	F
Humidity		9.07		
Temperature		8.66		
Humidity \times Temperature		6.07		
Residual				
Total		52.30		

a. (10 points) State the complete mathematical model for this experiment including all assumptions.

b. (10 points) Complete the ANOVA table and make all appropriate inferences based on a 5% level of significance.

c. (10 points) Describe two types of plots which should accompany this analysis and indicate their purpose.

5. (10 points) As an alternative to dangerous insecticides, a chemist is working on a synthetic pheromone (a type of hormone involved in mating behavior) to be used as a bait to attract destructive insects into traps. In a field experiment, 6 different levels of the synthetic hormone are used, with 10 traps per level. The 60 traps are placed at random in a peach orchard and the observed values below represent the number of Mediterranean fruit flies trapped during the same 4 hour period.

	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
	2	2	4	9	16	17
	2	3	6	11	17	24
	2	5	7	12	18	27
	3	5	7	15	19	33
	3	6	8	15	24	33
	4	7	11	17	25	36
	4	9	11	21	36	41
	5	10	12	28	38	54
	10	12	17	33	42	76
	15	16	22	39	65	109
Average	5.0	7.5	10.5	20.0	30.0	45.0
Std. Dev.	4.2	4.3	5.5	10.1	15.5	28.0

Determine a suitable transformation to stabilize the variance of these six sample distributions. Justify your choice of transformation. Do **not** perform the transformation.