1999 Honors Examination in Statistics

Department of Mathematics and Statistics
Swarthmore College

Instructions: This exam consists of a total of five questions. Number the questions clearly in your work and start each question on a new page. You must show your work to make it clear how you obtained your answers. Answers without any work may lose credit even if they are correct, and will receive no credit if incorrect.

This is a closed-book three-hour exam. You may not refer to your notes or textbooks. You may use the normal distribution and t-distribution tables attached, if necessary.

1. A large number, N, of people are subject to a blood test for a specific disease. The probability that a person has this disease is p. Assume this probability is the same for everyone and whether or not one person has the disease is independent of whether or not anyone else has the disease. The blood tests can be administered in two ways:

   (i) Each person can be tested separately. In this case, a total of N tests are required. The test will be positive with probability 0.9 if a person has the disease. (There is a 0.1 probability that the test is incorrectly negative if the person has the disease.) If a person does not have the disease, the test will be negative with probability 0.99. (There is a 0.01 probability of a false positive result.)

   (ii) The blood samples of k people can be pooled and analyzed together. If none of the people in the group has the disease then the test will be negative with probability 0.99 and this *one* test suffices for k people. If at least one person in a group has the disease then the test will be positive with probability 0.9, and each of the k persons must be tested separately, and, in all, $k+1$ tests are required for these k people.

   Assume that the results from each test are independent and that the groups are formed at random.

   a. In plan (ii), what is the probability that the test for a pooled sample of k people will be positive?

   b. Let X =number of blood tests necessary under plan (ii). Find the expected value of X.

   c. In terms of minimizing the expected number of blood tests to be performed on the N people, which plan, (i) or (ii), would be preferred if it is known that p is close to 0? Justify your answer using the expression derived in part (b).

2. Let X and Y be the (percentage) returns from two stocks A and B, respectively. Assume that X and Y have a joint bivariate normal distribution with $E(X)$=12, $E(Y)$=10, Var(X)=800, Var(Y)=250, and Cov(X,Y)=200. You decide to invest one quarter of your money in A and three quarters of your money in B. A friend tosses a fair coin twice. If he gets two heads he invests all his money in A, otherwise he invests all his money in B. What is the probability that you will have a greater percentage return than your friend?

3. Let $X_1, \ldots, X_n$ be a random sample from a population with the following density function

$$f(x|\theta) = \theta x^{\theta-1} \quad 0 < x < 1, \quad 0 < \theta < \infty$$

   a. Find a sufficient statistic for θ.

   b. Find the maximum likelihood estimate of θ, and show that its asymptotic variance $\to 0$ as $n \to \infty$.

   c. Find the method of moments estimator of θ.

4. The label from a Cheerfree laundry detergent box (see next page) implies that the box should contain 41.4 ounces of detergent (18 scoopfuls where each scoopful is 2.3 ounces). Notice, however, that there is also a disclaimer stating:
   "Individual packages of Cheerfree may weigh slightly more or less than the marked weight. This is due to normal variations incurred with high speed packaging machines. However, each day's production of Cheerfree will average slightly above the marked weight."

   Suppose, that after several loads of laundry you suspect that the Cheerfree boxes tend to contain systematically less (and never more) detergent than advertised. You decide to perform a statistical

test to settle the matter.

Suppose you measure the detergent in a random sample of 10 Cheerfree detergent boxes. (How you'd go about actually selecting this sample is an interesting question in itself.) You find that the average amount of detergent in these 10 boxes is 40.8 ounces.

Let's assume that the distribution of the weight of detergent in a Cheerfree box is normally distributed with a **known** standard deviation of 1.5 ounces.

a. Test the following hypothesis at a 0.05 level of significance:
$H_0 : \mu = 41.4$ vs $H_A : \mu < 41.4$
where μ is the (population) average amount of detergent in a Cheerfree box.

b. How large a sample size is needed so that the test in (a) has a power of 0.80? (Assume the true mean is 41.2 ounces.)

5. Assume you have a single observation the following density function:

$f(x|\theta) = 2e^{\theta^2} x e^{-x^2} \quad x \geq \theta$

Find an unbiased estimate for θ.
Hint: use the definition of unbiasedness.

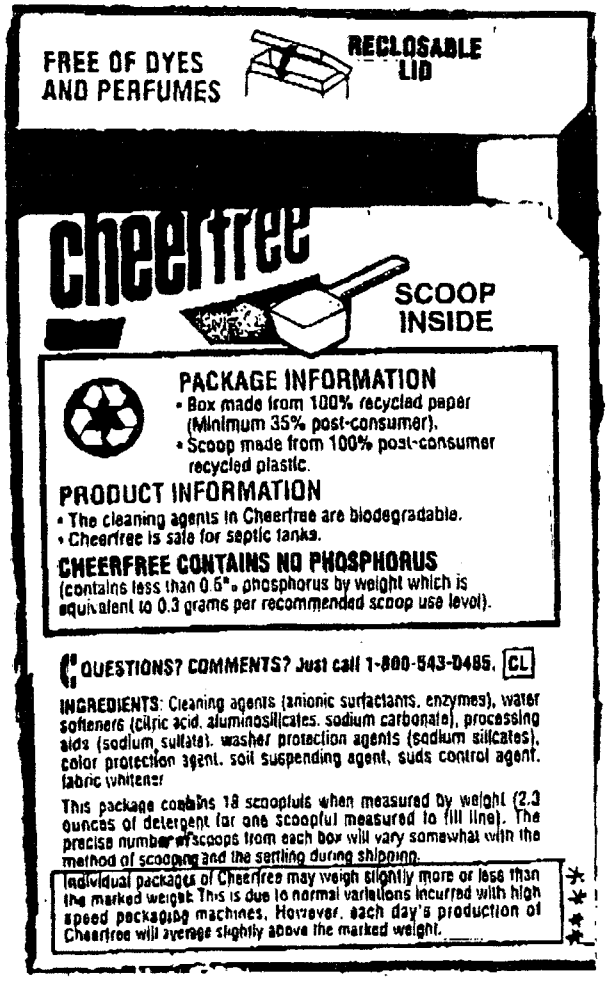

TABLE 2

Cumulative Normal Distribution—Values of P Corresponding to z_p for the Normal Curve

z is the standard normal variable. The value of P for $-z_p$ equals 1 minus the value of P for $+z_p$; for example, the P for -1.62 equals $1 - .9474 = .0526$.

z_p	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

TABLE 4

Percentiles of the t Distribution

df	$t_{.60}$	$t_{.70}$	$t_{.80}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$
1	.325	.727	1.376	3.078	6.314	12.706	31.821	63.657
2	.289	.617	1.061	1.886	2.920	4.303	6.965	9.925
3	.277	.584	.978	1.638	2.353	3.182	4.541	5.841
4	.271	.569	.941	1.533	2.132	2.776	3.747	4.604
5	.267	.559	.920	1.476	2.015	2.571	3.365	4.032
6	.265	.553	.906	1.440	1.943	2.447	3.143	3.707
7	.263	.549	.896	1.415	1.895	2.365	2.998	3.499
8	.262	.546	.889	1.397	1.860	2.306	2.896	3.355
9	.261	.543	.883	1.383	1.833	2.262	2.821	3.250
10	.260	.542	.879	1.372	1.812	2.228	2.764	3.169
11	.260	.540	.876	1.363	1.796	2.201	2.718	3.106
12	.259	.539	.873	1.356	1.782	2.179	2.681	3.055
13	.259	.538	.870	1.350	1.771	2.160	2.650	3.012
14	.258	.537	.868	1.345	1.761	2.145	2.624	2.977
15	.258	.536	.866	1.341	1.753	2.131	2.602	2.947
16	.258	.535	.865	1.337	1.746	2.120	2.583	2.921
17	.257	.534	.863	1.333	1.740	2.110	2.567	2.898
18	.257	.534	.862	1.330	1.734	2.101	2.552	2.878
19	.257	.533	.861	1.328	1.729	2.093	2.539	2.861
20	.257	.533	.860	1.325	1.725	2.086	2.528	2.845
21	.257	.532	.859	1.323	1.721	2.080	2.518	2.831
22	.256	.532	.858	1.321	1.717	2.074	2.508	2.819
23	.256	.532	.858	1.319	1.714	2.069	2.500	2.807
24	.256	.531	.857	1.318	1.711	2.064	2.492	2.797
25	.256	.531	.856	1.316	1.708	2.060	2.485	2.787
26	.256	.531	.856	1.315	1.706	2.056	2.479	2.779
27	.256	.531	.855	1.314	1.703	2.052	2.473	2.771
28	.256	.530	.855	1.313	1.701	2.048	2.467	2.763
29	.256	.530	.854	1.311	1.699	2.045	2.462	2.756
30	.256	.530	.854	1.310	1.697	2.042	2.457	2.750
40	.255	.529	.851	1.303	1.684	2.021	2.423	2.704
60	.254	.527	.848	1.296	1.671	2.000	2.390	2.660
120	.254	.526	.845	1.289	1.658	1.980	2.358	2.617
∞	.253	.524	.842	1.282	1.645	1.960	2.326	2.576