Total: 100 points

1. **[20 points]**

To investigate the percentage of abused children in a certain population, doctors examine a random sample of 1000 children from this population, and among this sample, 67 children are classified by doctors as abused. However, doctors are not perfect: They may classify an abused child (A) by saying that the child is not abused (ND), or they may classify a nonabused child (N) as abused (AD). According to a pilot study, these error rates are P(ND|A) = 10% and P(AD|N) = 5%, respectively, where P(ND|A) is the conditional probability that a child is classified as nonabused given that the child is indeed abused, and P(AD|N) is the conditional probability that a child is classified as abused given that the child is indeed nonabused. Please use these data to estimate the percentage of abused children in the population, P(A).

2. **[20 points]**

Let P equal the proportion of all college students who would say yes to the question, "Would you drink from the same glass after your friend had used it if you suspected that this friend was an AIDS virus carrier?" Given a random sample of size N, let Y equal the number of students who would say yes to this question.

(a) What is the approximate probability distribution of $\frac{Y-NP}{\sqrt{NP(1-P)}}$ when N is sufficiently large? Please also briefly state the theorem on which your answer is based. [10 points]

(b) How large is a sample size N needed so that we are 90% confident that P is within $\frac{Y}{N} \pm 3\%$? Use the fact that $P(1-P) \leq 0.25$ for all values of P. [10 points]

3. **[20 points]**

In the following contingency table, 60 individuals are classified by gender and by whether they favor, oppose, or have no opinion on a complete ban on smoking in public places. You are asked to perform a χ^2 (chi-square) test to test whether gender and opinion on smoking in public places are independent.

	Ban on Smoking in Public Places			
Gender	Favor	Oppose	No Opinion	Total
Male	8	10	12	30
Female	12	8	10	30
Total	20	18	22	60

(a) Specify the null hypothesis of this test. [5 points]

接背面

(b) Determine the expected frequencies in each gender-opinion cell under the null hypothesis. [6 points]

(c) What is the conclusion of this test at the 10% significance level? [9 points]

4. [40 points]

Suppose that 20 obesity patients were enrolled into a randomized clinical trial for comparing the effects of two different drugs (A and B) on weight reduction. They were randomized to two groups of equal sample size. The 10 patients in Group 1 took the tested Drug A ($X = 1$) one tablet daily for six months and the other 10 patients in Group 2 took the standard Drug B ($X = 0$) two tablets daily for six months too. We recorded their body weights (kg) in the research clinic before they began taking the assigned drug (Y_0) and at the end of the six month period (Y_1) respectively. The data are listed below:

		Weight (kg)				Weight (kg)	
I.D.	Drug	Before	After	I.D.	Drug	Before	After
1	A	105	92	11	B	96	91
2	A	112	90	12	B	107	94
3	A	127	101	13	B	100	89
4	A	99	82	14	B	119	101
5	A	97	85	15	B	98	90
6	A	108	91	16	B	137	116
7	A	132	99	17	B	121	105
8	A	124	100	18	B	99	89
9	A	101	87	19	B	128	103
10	A	140	108	20	B	110	99

The following linear regression models have been considered in analyzing this data set.

- **Model 1:** $\widehat{Y}_1 - Y_0 = \widehat{\beta}_0$
- **Model 2:** $\widehat{Y}_1 - Y_0 = \widehat{\beta}_0 + \widehat{\beta}_1 Y_0$
- **Model 3:** $\widehat{Y}_1 = \widehat{\beta}_0 + \widehat{\beta}_1 Y_0$
- **Model 4:** $\widehat{Y}_1 = \widehat{\beta}_0 + \widehat{\beta}_1 X$
- **Model 5:** $\widehat{Y}_1 - Y_0 = \widehat{\beta}_0 + \widehat{\beta}_1 X$
- **Model 6:** $\widehat{Y}_1 - Y_0 = \widehat{\beta}_0 + \widehat{\beta}_1 X + \widehat{\beta}_2 Y_0$
- **Model 7:** $\widehat{Y}_1 = \widehat{\beta}_0 + \widehat{\beta}_1 X + \widehat{\beta}_2 Y_0$

Note that (1) $\widehat{\beta}_0$, $\widehat{\beta}_1$, and $\widehat{\beta}_2$ are least squares estimates of the corresponding regression coefficients, (2) $\widehat{Y}_1$ is the predicted value of Y_1, and (3) X and Y_0 are treated as fixed covariates in fitting these linear regression models to the data or a subset of the data. The medical investigator is interested in the following three questions: (a), (b), and (c).

接次頁

(a) Does the tested Drug A have a positive effect on weight reduction? In other words, can the tested Drug A significantly reduce patients' weights?

i. Please conduct a paired t test with significance level $\alpha = 0.05$ to answer this question. [10 points]

ii. What assumption(s) do you need for this statistical test? [4 points]

iii. Which linear regression model(s) on the above list can be used for the patients in Group 1 to conduct a statistical test (on one or more regression coefficients) that is equivalent to the paired t test? (You should select at least one appropriate model from that list and state your reason to justify your answer. Yet, if you think all these models are not appropriate for this question, then please write down your own regression model(s) and state your reason(s) to justify your answer.) [6 points]

(b) More importantly, is the effect of the tested Drug A stronger than that of the standard Drug B on weight reduction? Which linear regression model(s) on the above list can be used for the patients in these two Groups to answer this question? (You should select at least one appropriate model from that list and state your reason to justify your answer. Yet, if you think all these models are not appropriate for this question, then please write down your own regression model(s) and state your reason(s) to justify your answer.) [10 points]

(c) However, as the medical investigator suspects, heavier patients have more weights to loose so that they may be able to loose more weights in a weight reduction program. Then, which linear regression model(s) on the above list would be preferred to take this information into account in addition to the potential drug's effect? (You should select at least one appropriate model from that list and state your reason to justify your answer. Yet, if you think all these models are not appropriate for this question, then please write down your own regression model(s) and state your reason(s) to justify your answer.) [10 points]

(Several statistical tables are attached for your information.)

接背面

Critical Values, $P(Z \geq z_\alpha) = \alpha$

α	.10	.05	.025	.01	.005	.001	.0005	.0001
z_α	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902	3.2905	3.7190

α	.00009	.00008	.00007	.00006	.00005	.00004	.00003	.00002	.00001
z_α	3.7455	3.7750	3.8082	3.8461	3.8906	3.9444	4.0128	4.1075	4.2649

TABLE 5 Percentage points of the t distribution ($t_{\alpha u}$)[a]

Degrees of freedom, d	u = .75	.80	.85	.90	.95	.975	.99	.995	.9995
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.598
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
60	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660	3.460
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

[a]The uth percentile of a t distribution with d degrees of freedom.

[Table 5 is taken from Table III of Fisher and Yates: "Statistical Tables for Biological, Agricultural and Medical Research," published by Longman Group Ltd., London (previously published by Oliver and Boyd Ltd., Edinburgh) and by permission of the authors and publishers.]

接次頁

TABLE 6 Percentage points of the chi-square distribution ($\chi^2_{d,u}$)[a]

	u													
d	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995	.999
1	0.0^4393[b]	0.0^3157[c]	0.0^3982[d]	0.00393	0.02	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.88	10.83
2	0.0100	0.0201	0.0506	0.103	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60	13.81
3	0.0717	0.115	0.216	0.352	0.58	1.21	2.37	4.11	6.25	7.81	9.35	11.34	12.84	16.27
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86	18.47
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75	20.52
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55	22.46
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28	24.32
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.09	21.95	26.12
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.67	23.59	27.88
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19	29.59
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.76	31.26
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.22	28.30	32.91
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82	34.53
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.68	26.12	29.14	31.32	36.12
15	4.60	5.23	6.27	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80	37.70
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27	39.25
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72	40.79
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.16	42.31
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58	43.82
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.57	40.00	45.32
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40	46.80
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.92	36.78	40.29	42.80	48.27
23	9.26	10.20	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18	49.73
24	9.89	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.36	42.98	45.56	51.18
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.31	46.93	52.62
26	11.16	12.20	13.84	15.38	17.29	20.84	25.34	30.43	35.56	38.89	41.92	45.64	48.29	54.05
27	11.81	12.88	14.57	16.15	18.11	21.75	26.34	31.53	36.74	40.11	43.19	46.96	49.64	55.48
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34	32.62	37.92	41.34	44.46	48.28	50.99	56.89
29	13.12	14.26	16.05	17.71	19.77	23.57	28.34	33.71	39.09	42.56	45.72	49.59	52.34	58.30
30	13.79	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67	59.70
40	20.71	22.16	24.43	26.51	29.05	33.66	39.34	45.62	51.81	55.76	59.34	63.69	66.77	73.40
50	27.99	29.71	32.36	34.76	37.69	42.94	49.33	56.33	63.17	67.50	71.42	76.15	79.49	86.66
60	35.53	37.48	40.48	43.19	46.46	52.29	59.33	66.98	74.40	79.08	83.30	88.38	91.95	99.61
70	43.28	45.44	48.76	51.74	55.33	61.70	69.33	77.58	85.53	90.53	95.02	100.42	104.22	112.32
80	51.17	53.54	57.15	60.39	64.28	71.14	79.33	88.13	96.58	101.88	106.63	112.33	116.32	124.84
90	59.20	61.75	65.65	69.13	73.29	80.62	89.33	98.64	107.56	113.14	118.14	124.12	128.30	137.21
100	67.33	70.06	74.22	77.93	82.36	90.13	99.33	109.14	118.50	124.34	129.56	135.81	140.17	149.45

[a] $\chi^2_{d,u}$ = *u*th percentile of a χ^2 distribution with *d* degrees of freedom.
[b] = 0.0000393
[c] = 0.000157
[d] = 0.000982
(Reproduced in part with permission of the Biometrika Trustees, from Table 3 of "Biometrika Tables for Statisticians," Volume II, edited by E. S. Pearson and H. O. Hartley, published for the Biometrika Trustees, Cambridge University Press, Cambridge, England, 1972.)