

UNIVERSITI SAINS MALAYSIA

Peperiksaan Semester Pertama
Sidang 1989/1990

Oktober/November 1989

BOO 284/4 Biostatistik

Masa: [3 jam]

Jawab **LIMA** daripada **ENAM** soalan.

Tiap-tiap soalan bernilai 20 markah.

1. Bagi setiap kajian yang dihuraikan di bawah
 - (i) Nyatakan ujian statistik yang wajar digunakan untuk analisis data yang dihasilkan.
 - (ii) Tuliskan hipotesis nul dan hipotesis alternatif.
 - (iii) Berikan formula bagi statistik ujian yang perlu dihitung.
 - (iv) Tulis dengan ayat lengkap kesimpulan ujian statistik itu sekiranya hipotesis nul dapat ditolak.

Terangkan semua simbol yang digunakan.

Kajian A

Kajian ini dijalankan ke atas lima ekor anak burung yang menetas daripada pengeraman telur yang sama di dalam sarang burung. Bagi setiap anak burung direkodkan:

- (a) Jujukan penetasannya, iaitu sama ada ia merupakan anak burung yang pertama menetas, atau yang kedua, dan seterusnya.
- (b) Kedominanan daripada kejayaannya bersaing untuk mendapatkan makanan daripada ibunya.

Data yang diperolehi digunakan untuk menguji sama ada jujukan penetasan anak burung berkait dengan kedominanan.

Kajian B

Kajian ini dijalankan untuk menentukan sama ada penyemburan larutan baja nitrogen pada daun dapat mengatasi kekurangan nitrogen pada tumbuhan kacang soya. Tujuh pasu kacang soya yang menunjukkan klorosis (iaitu kekuningan daun kerana kekurangan nitrogen yang serius) disemur dengan larutan baja nitrogen satu kali sehari. Satu minggu kemudian, direkodkan sama ada daun kacang soya menjadi lebih hijau, makin menguning atau tiada perubahan. Data ini digunakan untuk analisis statistik.

Kajian C

Sepuluh ekor tikus digunakan untuk kajian kesan ubat perangsang terhadap kebolehan belajar mencari jalan keluar melalui "maze". Lima ekor tikus dipilih secara rawak untuk diberi makan ubat perangsang (kumpulan A), dan lima ekor dijadikan kawalan (kumpulan B). Data yang diperolehi adalah pangkat-pangkat (dalam skala 1 hingga 10) mengikut betapa cepat sesuatu tikus dapat belajar mencari jalan keluar. Data ini digunakan untuk menguji sama ada ubat perangsang dapat mempercepatkan masa pembelajaran tikus.

Kajian D

Kadar percambahan biji benih satu varieti kacang soya ialah 90%, iaitu nisbah biji benih yang bercambah kepada biji benih dan yang tak bercambah dijangka 9:1. Lima puluh biji benih kacang soya yang disimpan selama 6 bulan direndam dan dibiarkan bercambah. Bilangan biji benih yang bercambah dan yang tidak bercambah dihitung selepas 3 hari. Ujian statistik dijalankan untuk menentukan sama ada kadar percambahan biji benih kacang soya itu masih tak terubah selepas penyimpanan selama enam bulan.

(20 markah)

2. Dalam satu kajian, tenaga yang digunakan oleh burung pipit untuk metabolisme ditentukan pada suhu yang berbeza. Untuk menjalankan kajian ini, burung pipit perlu disimpan dalam keadaan suhu udara yang berbeza, iaitu pada 25°C (suhu biasa) dan pada suhu 20°C.
 - (a) Huraikan rekabentuk eksperimen supaya anda boleh menggunakan ujian t bagi kes dua sampel berpasangan untuk analisis data kelak. Misalkan sepuluh replikat digunakan. Apakah kebaikan rekabentuk eksperimen ini berbanding

(BOO 284/4)

dengan dengan kes dua sampel tak bersandaran?

(5 markah)

- (b) Jadual 1 memberikan data yang diperolehi daripada kajian yang dihuraikan di atas, bagi kes sampel berpasangan. Jalankan ujian statistik untuk menentukan sama ada burung pipit perlu menggunakan lebih tenaga metabolisme pada suhu yang rendah.

(10 markah)

- (c) Dengan menggunakan data yang sama, tentukan sama ada pertalian linear antara tenaga metabolisme yang digunakan oleh burung pipit pada suhu 25°C dan 20°C .

(5 markah)

...6/-

Jadual 1. Tenaga (kalori/jam) yang digunakan oleh burung pipit pada suhu 20°C (x_1) dan pada suhu 25°C (x_2)

x_1	x_2
18.7	15.2
16.0	16.2
17.6	15.5
18.0	15.0
16.9	17.0
19.2	16.9
16.5	16.9
17.9	15.4

$$\begin{aligned} \bar{x}_1 &= 17.6000 & \bar{x}_2 &= 16.0125 \\ s_1 &= 1.0876 & s_2 &= 0.8374 \\ \Sigma x_{1i} &= 140.8000 & \Sigma x_{2i} &= 128.1000 \\ \Sigma x_{1i}^2 &= 2486.3600 & \Sigma x_{2i}^2 &= 2056.1100 \end{aligned}$$

$$\Sigma x_{1i} x_{2i} = 2252.5300$$

3. Satu kajian dijalankan untuk menentukan kesan kehadiran bahan gula terhadap percambahan biji benih pea. Data berikut adalah panjang radikel (akar pertama) dalam ukuran mm yang muncul satu hari selepas percambahan biji benih pea bagi kawalan (tanpa gula) dan dua olahan gula, iaitu pada 2% glukosa dan 2% fruktosa.

(BOO 284/4)

Jadual 2. Panjang radikel (mm) satu hari selepas percambahan biji benih pea

	Kawalan	2% glukosa	2% fruktosa
	7.5	5.7	5.8
	6.7	5.8	6.1
	7.0	6.0	5.6
	7.5	5.9	5.8
	6.5	6.2	5.7
	7.1	6.0	5.6
	6.7	6.0	6.1
	6.7	5.7	6.0
	7.6	5.9	5.7
	6.8	6.1	5.8
Min olahan	7.01	5.93	5.82
Jumlah olahan	70.1	59.3	58.2

$$\sum \sum x_{ij} = 187.6000$$

$$\sum \sum x_{ij}^2 = 1,183.7600$$

- (a) ① Ujikan sama ada terdapat kesan olahan yang bererti ke atas percambahan biji benih pea.
- ② Lakukan perbandingan antara semua olahan untuk menentukan sama ada terdapat perbezaan yang bererti.

(15 markah)

...8/-

(BOO 284/4)

(b) Andaikan setiap baris data dalam jadual di atas adalah bagi biji benih pea daripada lenggai yang sama. Lakukan analisis statistik untuk menentukan

(i) Kesan olahan gula

(ii) Kesan lenggai

ke atas pertambahan biji benih pea.

(5 markah)

4. Berikut adalah gambarajah yang menunjukkan susunatur plot di lapangan bagi satu kajian untuk menentukan kesan urea (iaitu satu jenis baja N) ke atas hasil cili. Olahan baja ditetapkan pada dua aras:

n_1 , iaitu 1 ton/ha, dan

n_2 , iaitu 2 ton/ha

Satu lagi faktor diambil kira, iaitu cara meletak baja ke dalam plot:

a_1 , iaitu baja disembur pada permukaan tanah sahaja

a_2 , iaitu baja dimasukkan ke dalam tanah dan dicampurkan dengan merata

Hasil cili (kg/plot) bagi setiap plot diberikan juga dalam gambarajah.

...9/-

Lakukan ujian statistik yang wajar. Apakah kesimpulan anda? *RCBD*

Blok I	Blok II	Blok III	Blok IV
a_{1n_1} 9.8	a_{2n_1} 9.5	a_{2n_2} 33.5	a_{1n_2} 31.3
a_{2n_1} 11.1	a_{1n_1} 10.5	a_{1n_2} 30.5	a_{1n_1} 12.6
a_{2n_2} 36.1	a_{2n_2} 32.8	a_{2n_1} 13.3	a_{2n_1} 9.7
a_{1n_2} 30.2	a_{1n_2} 31.7	a_{1n_1} 8.5	a_{2n_2} 35.1

*↓
Hina blok*

C.F. = 592.6200

5. Data berikut diambil daripada satu kajian untuk menentukan pengecutan otot jantung siput remis di bawah pengaruh sejenis ester pada dosej yang berlainan.

Dosej, ug	Pengecutan otot, mm
30	1.5
100	3.2
300	4.4
1000	6.1

- (a) Plotkan pengecutan otot melawan dosej ester dan perhatikan aliran pertalian antara dua variabel ini. Jelmakan dosej kepada log (dosej) supaya terdapat pertalian linear antara log (dosej) dan pengecutan otot.

(5 markah)

(BOO 284/4)

- (b) Dengan menggunakan kaedah statistik yang sesuai, anggarkan dosej ester yang akan menyebabkan pengecutan otot sejauh 3.5 mm.

(15 markah)

6. Anda diberi tugas mendapatkan sampel 50 plot sawah padi untuk menentukan hasil padi di suatu kawasan penanaman padi di Kedah. Anda telah memilih satu mukim di daerah Sik di Kedah, yang mana terdapat 500 plot sawah padi.

- (a) Huraikan bagaimana anda akan menggunakan skema penyampelan rawak ringkas, dengan bantuan sifir angka-angka rawak, untuk mendapatkan sampel anda.

(5 markah)

- (b) Anda dapati bahawa 500 plot sawah padi itu dimiliki oleh 100 petani. Huraikan bagaimana anda akan menggunakan skema penyampelan berkelompok untuk mendapatkan sampel anda.

(5 markah)

...11/-

- (c) Misalkan daripada penyampelan yang dijalankan, didapati bahawa min hasil padi daripada 50 plot ialah 4.54 ton/ha dan sishan piawai ialah 0.92 ton/ha. Apakah penganggar selang untuk min hasil padi di mukim itu pada 95% keyakinan?

(5 markah)

- (d) Anda dapati bahawa 60 petani telah menanam padi secara tabur terus, manakala 40 petani menggunakan kaedah pindah tanam. Huraikan bagaimana anda akan memperolehi 25 plot sawah yang ditanam secara tabur terus dan 25 plot lagi yang ditanam secara pindah tanam untuk membandingkan hasil padi bagi dua kaedah penanaman padi tersebut.

(5 markah)

LAMPIRAN 1

(BOO 284/4)

FORMULA-FORMULA PANDUAN

1. Ujian-t bagi sampel takbersandaran

Formula panduan bagi anggaran varians populasi:-

$$i. s_p^2 = \frac{\Sigma(x_{1i} - \bar{x}_1)^2 + \Sigma(x_{2i} - \bar{x}_2)^2}{n_1 + n_2 - 2} \text{ bagi } n_1 \neq n_2$$

$$\text{atau } s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$$

$$ii. s_p^2 = \frac{s_1^2 + s_2^2}{2} \text{ bagi } n_1 = n_2 = n$$

$$s_p^2 \bar{x}_1 - \bar{x}_2 = s_p^2 \left[\frac{n_1 + n_2}{n_1 n_2} \right]$$

2. Anggaran kecerunan garis regresi linear

$$\hat{b} = \frac{n \Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n \Sigma x_i^2 - (\Sigma x_i)^2} \text{ atau } \hat{b} = \frac{\Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}}{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}}$$

3. Anggaran koefisien korelasi Pearson

$$r = \frac{n \Sigma x_i y_i - \Sigma x_i \Sigma y_i}{\sqrt{[n \Sigma x_i^2 - (\Sigma x_i)^2][n \Sigma y_i^2 - (\Sigma y_i)^2]}}$$

4. Formula kebarangkalian bagi taburan Poisson.

$$f(n) = \frac{a^n e^{-a}}{n!}$$

Jadual 2. Sifir bagi taburan χ^2 (ujian dua hujung) (BOO 284/4)

TABLE B. TABLE OF CRITICAL VALUES OF χ^2

df	Level of significance for one-tailed test					.0005
	.10	.05	.025	.01	.005	
	Level of significance for two-tailed test					
	.20	.10	.05	.02	.01	.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.355	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.452	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.674	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

* Table B is abridged from Table III of Fisher and Yates: *Statistical tables for biological, agricultural, and medical research*, published by Oliver and Boyd Ltd., Edinburgh, by permission of the authors and publishers.

TABLE C. TABLE OF CRITICAL VALUES OF CHI SQUARE*

df	Probability under H_0 that $\chi^2 \geq$ chi square													
	.99	.98	.95	.90	.80	.70	.50	.30	.20	.10	.05	.02	.01	.001
1	.00016	.00063	.0039	.016	.064	.15	.46	1.07	1.64	2.71	3.84	5.41	6.64	10.83
2	.02	.04	.10	.21	.45	.71	1.39	2.41	3.22	4.60	5.99	7.82	9.21	13.82
3	.12	.18	.35	.58	1.00	1.42	2.37	3.66	4.64	6.25	7.82	9.84	11.34	16.27
4	.30	.43	.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	11.67	13.28	18.46
5	.55	.75	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	13.16	15.09	20.52
6	.87	1.13	1.64	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	15.03	16.81	22.46
7	1.24	1.56	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	16.62	18.48	24.32
8	1.65	2.03	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	18.17	20.09	26.12
9	2.09	2.53	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	19.68	21.67	27.88
10	2.56	3.06	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	21.16	23.21	29.52
11	3.05	3.61	4.58	5.58	6.99	8.15	10.34	12.60	14.63	17.28	19.68	22.31	24.72	31.22
12	3.57	4.18	5.23	6.30	7.81	9.03	11.34	13.58	15.81	18.55	21.03	23.65	26.22	32.91
13	4.11	4.76	5.89	7.04	8.63	9.93	12.34	14.51	16.95	19.81	22.36	25.47	27.69	34.53
14	4.66	5.37	6.57	7.79	9.47	10.82	13.34	15.62	18.15	21.06	23.68	26.79	29.14	36.12
15	5.23	5.98	7.26	8.55	10.31	11.72	14.34	16.72	19.31	22.31	25.00	28.26	30.58	37.70
16	5.81	6.61	7.96	9.31	11.15	12.62	15.34	17.82	20.46	23.54	26.30	29.63	32.00	39.29
17	6.41	7.26	8.67	10.08	12.00	13.53	16.34	19.51	22.62	25.77	28.59	31.00	33.41	40.75
18	7.02	7.91	9.39	10.86	12.86	14.44	17.34	20.60	24.76	28.59	32.35	34.82	35.72	42.31
19	7.63	8.57	10.12	11.65	13.72	15.35	18.34	21.69	26.90	30.30	33.69	36.19	37.43	43.82
20	8.26	9.24	10.85	12.44	14.58	16.27	19.34	22.78	28.04	32.41	35.41	37.57	38.58	45.32
21	8.90	9.92	11.59	13.24	15.44	17.18	20.34	23.86	29.17	33.62	37.36	39.38	40.45	46.80
22	9.54	10.60	12.34	14.04	16.31	18.10	21.24	24.94	30.30	34.81	38.58	40.79	41.64	48.27
23	10.20	11.29	13.09	14.85	17.19	19.02	22.34	26.02	31.43	35.97	39.91	41.64	42.93	49.73
24	10.86	11.99	13.85	15.66	18.06	19.94	23.34	27.10	32.53	37.15	41.64	42.93	44.18	51.18
25	11.52	12.70	14.61	16.47	18.94	20.87	24.34	28.17	33.63	38.37	42.93	44.18	45.43	52.62
26	12.20	13.41	15.38	17.29	19.82	21.79	25.34	29.25	34.73	39.51	43.83	45.43	46.68	54.05
27	12.88	14.12	16.15	18.11	20.70	22.72	26.34	30.32	35.81	40.60	44.74	46.68	47.93	55.48
28	13.56	14.85	16.93	18.94	21.59	23.65	27.34	31.39	36.94	41.73	45.81	47.93	49.18	56.89
29	14.26	15.57	17.71	19.77	22.48	24.58	28.34	32.42	38.09	42.94	46.99	49.18	50.43	58.30
30	14.95	16.31	18.49	20.60	23.36	25.51	29.34	33.53	39.25	44.03	48.18	50.43	51.68	59.70

* Table C is abridged from Table IV of Fisher and Yates: *Statistical tables for biological, agricultural, and medical research*, published by Oliver and Boyd Ltd., Edinburgh, by permission of the authors and publishers.

Nilai χ^2 yang dihitung adalah bererti pada sesuatu paras keertian jika ia sama dengan atau melebihi nilai genting bagi χ^2 yang ditunjukkan di dalam sifir ini.

Jadual 4. Sifir bagi ρ , koefisien korelasi Pearson.
(ujian dua hujung)

TABLE A 11
CORRELATION COEFFICIENTS AT THE 5% AND 1% LEVELS OF SIGNIFICANCE

Degrees of Freedom	5%	1%	Degrees of Freedom	5%	1%
1	.997	1.000	24	.388	.496
2	.950	.990	25	.381	.487
3	.878	.959	26	.374	.478
4	.811	.917	27	.367	.470
5	.754	.874	28	.361	.463
6	.707	.834	29	.355	.456
7	.666	.798	30	.349	.449
8	.632	.765	35	.325	.418
9	.602	.735	40	.304	.393
10	.576	.708	45	.288	.372
11	.553	.684	50	.273	.354
12	.532	.661	60	.250	.325
13	.514	.641	70	.232	.302
14	.497	.623	80	.217	.283
15	.482	.606	90	.205	.267
16	.468	.590	100	.195	.254
17	.456	.575	125	.174	.228
18	.444	.561	150	.159	.208
19	.433	.549	200	.138	.181
20	.423	.537	300	.113	.148
21	.413	.526	400	.098	.128
22	.404	.515	500	.088	.115
23	.396	.505	1.000	.062	.081

Portions of this table were taken from Table VA in *Statistical Methods for Research Workers* by permission of Professor R. A. Fisher and his publishers, Oliver and Boyd.

(BOO 284/4)

Jadual 5: Sifir bagi taburan z

TABLE A 3
CUMULATIVE NORMAL FREQUENCY DISTRIBUTION
(Area under the standard normal curve from 0 to Z)

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000									

...17/-

Although it is easy to define a simple random sample, it is not always obvious how such a sample is to be drawn from an actual population.

Drawing chips from a bowl. We now consider the most straightforward situation in which the population is finite and in which the elements are easily identified and can be numbered. For example, suppose a college freshman class has 100 students and we wish to draw a simple random sample of ten of these students without replacement. We could assign numbers from 1 to 100 to each of the students and place these numbers on physically similar disks (or balls, slips of paper, etc.), and put them in a bowl. We could shake the bowl to mix the disks thoroughly and then draw the sample. On drawing the first disk, we record the number written on it. We then shake the bowl again, draw the second disk, and record the result. We could repeat the process until we have drawn ten numbers. The students corresponding to these ten numbers would then constitute the required simple random sample.

Tables of random numbers. If the population size were very large, the foregoing procedure would become quite unwieldy and time-consuming. Furthermore, it may introduce biases if the disks are not thoroughly mixed. Therefore, in recent years, there has been a marked tendency to use tables of random digits for drawing such samples. These tables are also useful for the selection of other types of probability samples.

A table of random digits is simply a table of digits generated by a random process. Usually, the digits are combined, for example, into groups of five, for ease of use. Thus, a table of random digits could be generated by drawing chips from a bowl, in a process similar to the one just described. The digits 0, 1, 2, . . . , 9 could be written on disks, the disks placed in a bowl, and then drawn, one at a time, with the selected disk replaced after each drawing. Thus, on each selection, the population would consist of the ten digits. The recorded digits would constitute a particular sequence of random digits. These tables are now usually produced by a computer, which has been programmed to generate random sequences of digits.

Suppose there were 9241 undergraduates at a large university and we wished to draw a simple random sample of 300 of these students. Each of the 9241 students could be assigned a four-digit number, say, for convenience, from 0001 to 9241. This list of names and numbers would constitute the sampling frame. We now turn to a table of random digits, such as Table 6-1, to select a simple random sample of 300 such four-digit numbers. We may begin on any page in the table and proceed in any systematic manner to draw the sample. Assume we decided to use the first four columns of each group of five, beginning at the upper left and reading downward. The first group of five digits on the left-hand side in Table 6-1 designates the line number, so we ignore it. Starting with the second group of digits, we find the sequence 98389. Since we are using the first four digits, we have the number 9838, which

exceeds the largest number in our population, 9241. Therefore, we ignore this number and read down to pick up the next four-digit number, 1724. This, then, is the number of the first student in the sample. Reading down consecutively, we find 0128, 9818 (which we ignore), 5928, and so forth until 300 four-digit numbers between 0001 and 9241 have been specified. If any previously selected number is repeated, we simply ignore the repeated appearance, and continue. In this illustration, we read downward on the page. We could have read laterally, diagonally, or in any other systematic fashion. What is important is that each four-digit number has an equal probability of selection, regardless of the systematic method of drawing used and regardless of the preceding numbers.

Methods are available for drawing samples other than simple random ones, and, even for situations where the elements have not been prelisted. Many tables include instructions for their use, so we will not pursue the subject further.

In this section, we have concentrated on simple random sampling. Since the theoretical structure of statistical inference is largely based on this type of

TABLE 6-1
Random Digits

19300	98389	95130	36323	33381	98930	60278	33338	45778	86643	78214
19301	17245	58145	89635	19473	61690	33549	70476	35153	41736	96170
19302	01289	68740	70432	43824	98577	50959	36855	79112	01047	30005
19303	98182	43535	79938	72575	13602	44115	11316	55879	78224	96740
19304	59266	39490	21582	09389	93679	26320	51754	42930	93609	06815
19305	42162	43375	78975	89654	71446	77779	95460	01250	01551	42552
19306	50357	15046	27813	34984	32297	57063	65413	79579	23870	00982
19307	11326	67204	56708	28022	80243	51848	06119	59285	86325	02977
19308	55636	06783	60962	12436	75218	38374	43797	65961	52366	83357
19309	31149	06588	27838	17511	02935	69747	88322	70360	77368	04222
19310	25055	23402	60275	81173	21950	63463	09389	83095	90744	44178
19311	35150	34706	08125	35809	57489	51799	01665	13834	97714	55167
19312	61486	33467	28352	58951	70174	21360	99318	69504	65556	02721
19313	44444	86623	23371	23287	36548	33503	78550	21593	27517	63104
19314	14825	81523	62729	36417	67047	16506	76410	42372	55040	27411
19315	59079	46755	72348	69595	53408	92708	67110	68260	77920	91123
19316	48391	76486	60421	69414	37271	89276	07577	43880	04133	07194
19317	67072	33693	19176	68018	89363	39340	93294	82290	05422	44129
19318	86050	07311	89994	36265	62134	47301	25352	01167	51641	41113
19319	84428	40439	57595	37715	16839	08343	00144	10294	64312	11201
19320	41048	28126	02664	23909	50517	65201	07369	79368	77991	40289
19321	30335	84930	99485	68202	79272	91220	76515	23902	29430	42749
19322	33524	27859	20526	52412	86213	60787	70235	36975	26660	90793
19323	26784	20591	20308	75604	94285	46100	13120	16694	63017	83112
19324	85741	22643	16202	48470	97412	65416	36996	52391	81122	95157

Source: RAND Corporation, A Million Random Digits with One Hundred Thousand Normal Deviate Influences, 2nd Edition, 1955, excerpt from page 347. Used by permission.