Final Test, May 4, 1:30pm–3:20pm

Show your work. The test is out of 100 points and you have 110 minutes to finish.

1. The following is part of an article from Newsweek, April 24, 2006. This question concerns the Women's Health Study, described in the second paragraph.

Take an Aspirin and ...

BY JULIE BURING, SC.D., AND NANCY FERRARI

Aspirin is a wonder drug, plain and simple. At high doses, it quells inflammation; at medium doses, it provides effective pain relief; at low doses, it reduces the blood's ability to clot by inhibiting the action of tiny blood cells called platelets. It makes sense, then, that aspirin might help prevent clot-related cardiovascular events such as heart attack and stroke, even in healthy people. In 1988, the Physicians' Health Study showed exactly that. In healthy men, 325mg of aspirin taken every other day for five years reduced the risk of a first heart attack by 44 percent. That was great news. For men.

It wasn't until March 2005 that the Women's Health Study addressed aspirin's benefits for women. Healthy women—who were at least 45 years old at the start of the study—who participated in the study took either 100mg of aspirin or a placebo every other day for 10 years. Surprisingly, the women taking aspirin experienced no reduction in heart-attack risk. However, aspirin takers were 17 percent less likely to have a stroke.

(a) (2 points) Is the study controlled? How do you know?

(b) (2 points) Is the study blind? How do you know?

(c) (2 points) What is a placebo? Why is it used?

(d) (2 points) The article does not say how the women were assigned to the aspirin and placebo groups. What is the best way to do this, and why?

2. (8 points) The following table summarizes the number of milligrams of sodium per serving for 18 types of breakfast cereal. Class intervals include the left endpoint but not the right.

Sodium (milligrams)	Number of cereals
120 to 160	4
160 to 200	3
200 to 240	6
240 to 280	3
280 to 320	2

Draw a histogram for the data, being careful to label the axes correctly.
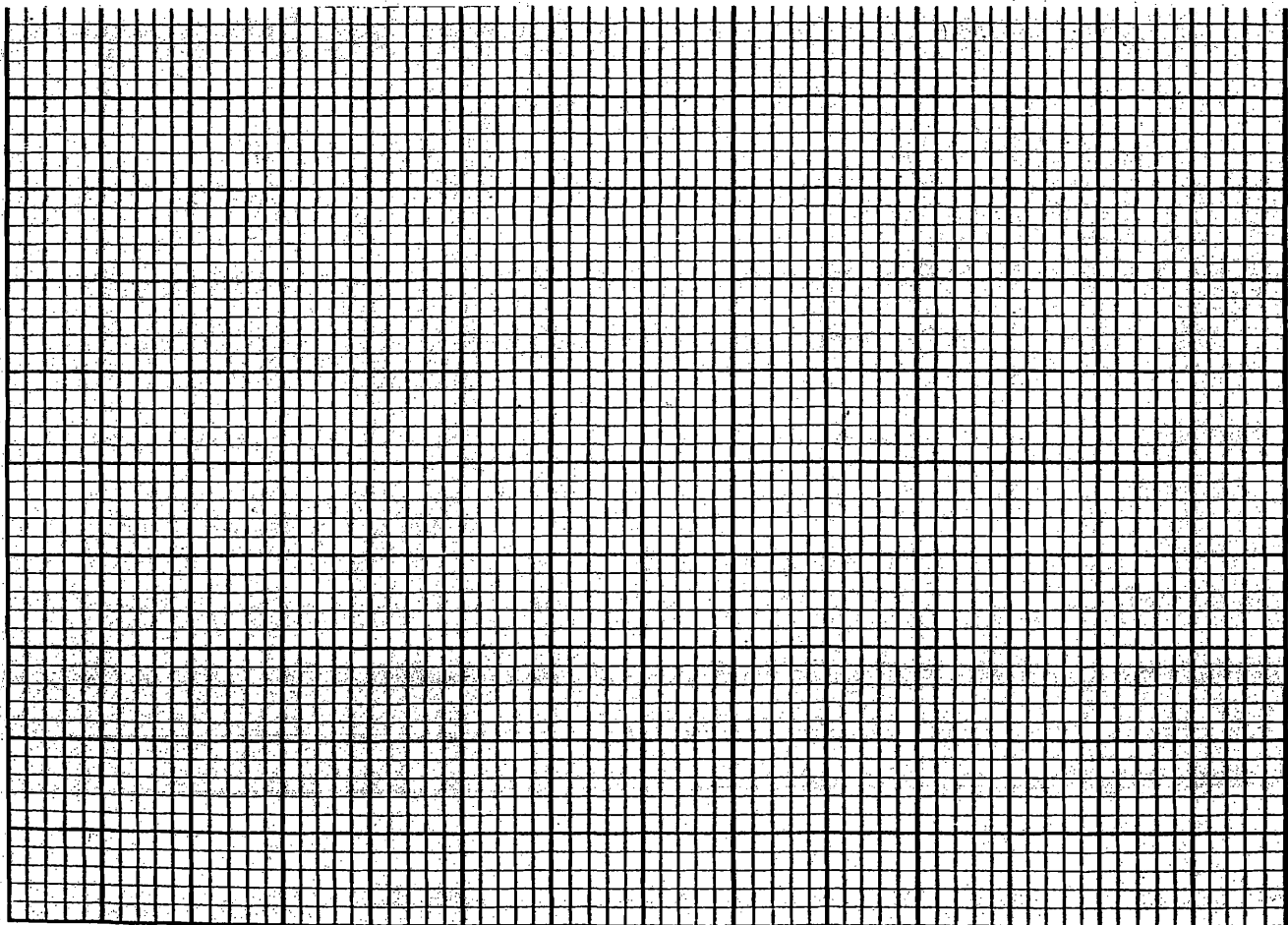
3. (4 points) Suppose a jury of 12 people is randomly chosen from a large city's adult residents, of whom 50% are female. What is the chance that at least one person on the jury is female?

4. For 167 college students the average handspan size is 20.9 inches, with an SD of 1.9 inches.

    (a) (5 points) Using the normal curve, approximately what percentage of the students have a handspan of more than 23 inches?

    (b) (5 points) If 90 percent of the students have a smaller handspan than the teacher, what is the teacher's handspan?

5. For the 167 college students in question 4, the relationship between height and handspan size is summarized as follows:

height: average = 68.0 inches   SD = 4.0 inches $r = 0.75$
handspan size: average = 20.9 inches   SD = 1.9 inches

    (a) (4 points) Six scatter diagrams are printed on the back of the formula sheet. Which of the scatter diagrams is the correct one for these data? Circle the correct one below:

<div align="center">A   B   C   D   E   F</div>

    (b) (4 points) Using the summary statistics above, what is the regression estimate for handspan for a student who is 60 inches tall?

    (c) (3 points) Find the rms error for your answer in part (b).

    (d) (3 points) What would the correlation coefficient be if we changed all the handspan measurements to centimeters? (There are 2.54 centimeters in an inch).

<div align="center">3</div>

6. (4 points) I have 20 batteries, of which 15 work and the other 5 don't. I choose 4 batteries at random from all 20 to put in my camera. What is the chance that all 4 batteries work?

7. (8 points) In 2002, 17.5% of US tax forms reported incomes under $25,000. If the IRS takes a random sample of 500 tax forms from 2002, what is the chance that between 17% and 18% of the sampled forms reported incomes of under $25,000?

8. The 2002 General Social Survey asked "What do you think is the ideal number of children for a family to have?" For the 497 females who responded, the average was 3.02 with a standard deviation of 1.81. Suppose these 497 women were like a simple random sample of adult American females.

   (a) (7 points) Find a 95% confidence interval for the average response of all adult American females.

   (b) (2 points) Is your confidence interval valid even though the answers don't follow the normal curve? Explain briefly.

9. (10 points) The following table comes from a simple random sample of high school seniors from a large city, each of whom was asked whether they had ever smoked cigarettes and whether they had ever drunk alcohol.

| | | Alcohol | | |
		No	Yes	
Cigarettes	No	281	500	781
	Yes	46	1449	1495
		327	1949	2276

Test to see whether there is an association between smoking and drinking alcohol. You should clearly state the null and the alternative hypotheses, find a test statistic and an approximate P-value, and state your conclusions in everyday language.

10. (10 points) In the aspirin study described in question 1, there were 19,934 women in the aspirin group and 19,942 in the placebo group. There were 477 major cardiovascular events in the aspirin group and 522 in the placebo group. Assuming the women were assigned to the aspirin and placebo groups appropriately, perform a statistical significance test to determine whether or not aspirin prevents major cardiovascular events for women like these. You should clearly state the null and the alternative hypotheses, find a test statistic and an approximate P-value, and state your conclusions in everyday language.

11. The long-held standard for normal body temperature is 98.6 degrees. A group of researchers believe that normal body temperature is actually lower than 98.6 degrees. To test this, they take the temperatures of 10 healthy people. They find the following:

98.2 97.8 99.0 98.6 97.8 98.4 99.7 98.2 97.4 97.6

Suppose this is a simple random sample of all healthy people and that the temperatures of healthy people follow the normal curve.

(a) (2 points) State the appropriate null and alternative hypotheses.

(b) (3 points) Find a test statistic.

(c) (2 points) How big is the P-value?

(d) (2 points) Do you reject the null hypothesis? Why/why not?

(e) (2 points) Based on the data, are the researchers correct? Explain briefly.

(f) (2 points) If it was believed that the temperatures of healthy people did not follow the normal curve, would your hypothesis test still be valid? Explain briefly.

Please note that these are provided for your convenience, but it is your responsibility to know how and when to use them.

$$\text{rms error} = \sqrt{1 - r^2} \times SD_Y$$

$$\text{slope} = r \times \frac{SD_Y}{SD_X}$$

$$\text{intercept} = \text{ave}_Y - \text{slope} \times \text{ave}_X$$

$$SD^+ = \sqrt{\frac{\text{number of draws}}{\text{number of draws} - 1}} \times SD$$

$$SD_{\text{box}} = \sqrt{\text{fraction of 0's} \times \text{fraction of 1's}}$$

$$EV_{\text{sum}} = \text{number of draws} \times \text{ave}_{\text{box}}$$
$$SE_{\text{sum}} = \sqrt{\text{number of draws}} \times SD_{\text{box}}$$

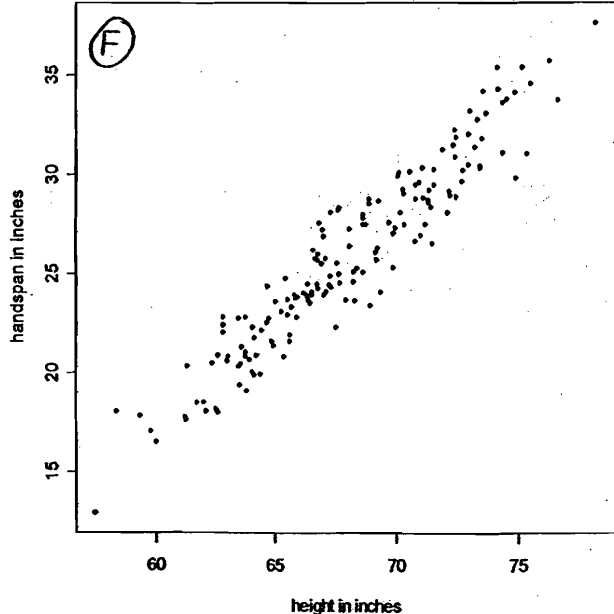$$EV_{\text{ave}} = \text{ave}_{\text{box}}$$
$$SE_{\text{ave}} = \frac{SE_{\text{sum}}}{\text{number of draws}}$$

$$EV_\% = \text{\% of 1's in the box}$$
$$SE_\% = \left(\frac{SE_{\text{sum}}}{\text{number of draws}}\right) \times 100\%$$

$$SE_{\text{diff}} = \sqrt{a^2 + b^2} \quad \text{where } a \text{ is the SE for the first quantity,}$$
$$b \text{ is the SE for the second quantity, and the two quantities are independent}$$

7

A — handspan in inches vs height in inches
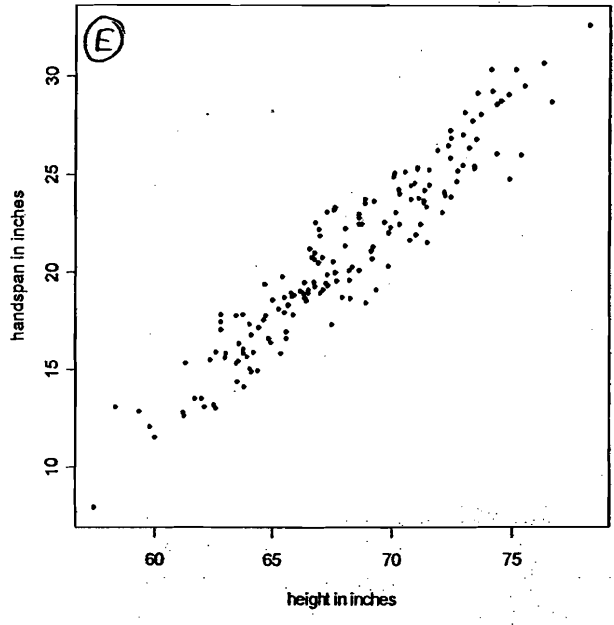B — handspan in inches vs height in inches
C — handspan in inches vs height in inches
D — handspan in inches vs height in inches
E — handspan in inches vs height in inches
F — handspan in inches vs height in inches

A NORMAL TABLE

Area (percent)
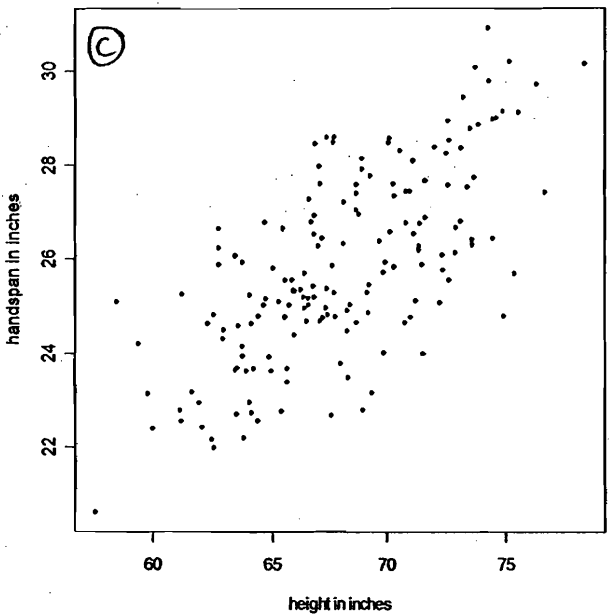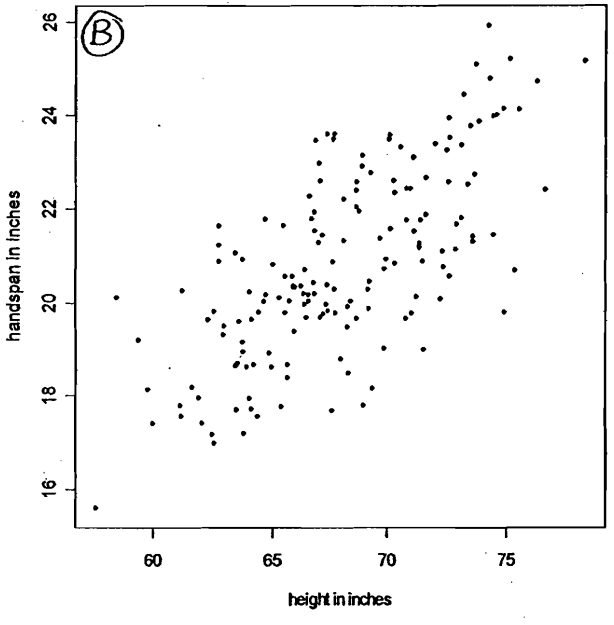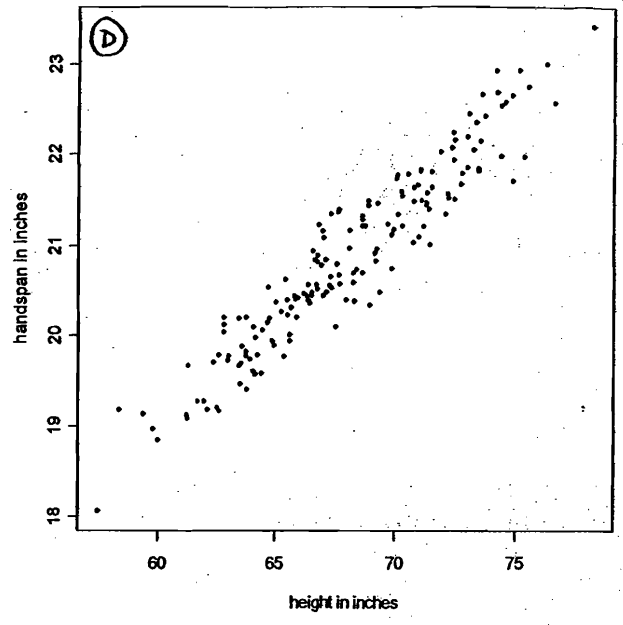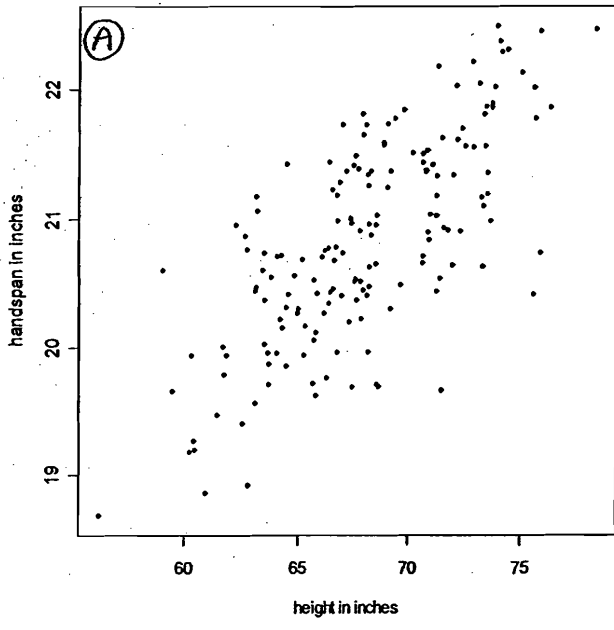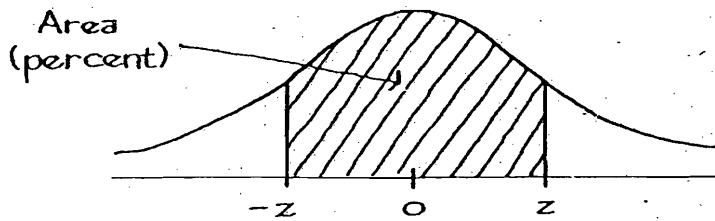
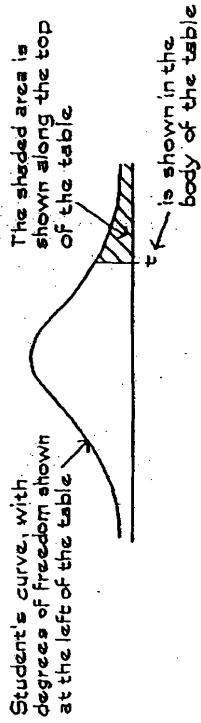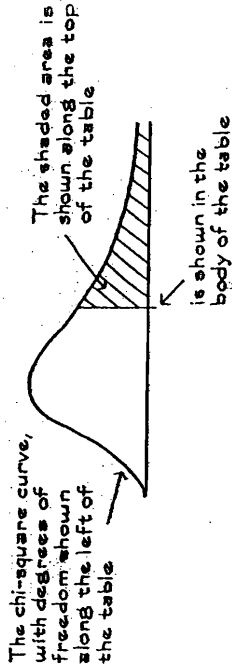| z	Area	z	Area	z	Area
0.00	0	1.50	86.64	3.00	99.730
0.05	3.99	1.55	87.89	3.05	99.771
0.10	7.97	1.60	89.04	3.10	99.806
0.15	11.92	1.65	90.11	3.15	99.837
0.20	15.85	1.70	91.09	3.20	99.863
0.25	19.74	1.75	91.99	3.25	99.885
0.30	23.58	1.80	92.81	3.30	99.903
0.35	27.37	1.85	93.57	3.35	99.919
0.40	31.08	1.90	94.26	3.40	99.933
0.45	34.73	1.95	94.88	3.45	99.944
0.50	38.29	2.00	95.45	3.50	99.953
0.55	41.77	2.05	95.96	3.55	99.961
0.60	45.15	2.10	96.43	3.60	99.968
0.65	48.43	2.15	96.84	3.65	99.974
0.70	51.61	2.20	97.22	3.70	99.978
0.75	54.67	2.25	97.56	3.75	99.982
0.80	57.63	2.30	97.86	3.80	99.986
0.85	60.47	2.35	98.12	3.85	99.988
0.90	63.19	2.40	98.36	3.90	99.990
0.95	65.79	2.45	98.57	3.95	99.992
1.00	68.27	2.50	98.76	4.00	99.9937
1.05	70.63	2.55	98.92	4.05	99.9949
1.10	72.87	2.60	99.07	4.10	99.9959
1.15	74.99	2.65	99.20	4.15	99.9967
1.20	76.99	2.70	99.31	4.20	99.9973
1.25	78.87	2.75	99.40	4.25	99.9979
1.30	80.64	2.80	99.49	4.30	99.9983
1.35	82.30	2.85	99.56	4.35	99.9986
1.40	83.85	2.90	99.63	4.40	99.9989
1.45	85.29	2.95	99.68	4.45	99.9991

A t-TABLE

Student's curve, with degrees of freedom shown at the left of the table

The shaded area is shown along the top of the table

t ← is shown in the body of the table

Degrees of freedom	25%	10%	5%	2.5%	1%	0.5%
1	1.00	3.08	6.31	12.71	31.82	63.66
2	0.82	1.89	2.92	4.30	6.96	9.92
3	0.76	1.64	2.35	3.18	4.54	5.84
4	0.74	1.53	2.13	2.78	3.75	4.60
5	0.73	1.48	2.02	2.57	3.36	4.03
6	0.72	1.44	1.94	2.45	3.14	3.71
7	0.71	1.41	1.89	2.36	3.00	3.50
8	0.71	1.40	1.86	2.31	2.90	3.36
9	0.70	1.38	1.83	2.26	2.82	3.25
10	0.70	1.37	1.81	2.23	2.76	3.17
11	0.70	1.36	1.80	2.20	2.72	3.11
12	0.70	1.36	1.78	2.18	2.68	3.05
13	0.69	1.35	1.77	2.16	2.65	3.01
14	0.69	1.35	1.76	2.14	2.62	2.98
15	0.69	1.34	1.75	2.13	2.60	2.95
16	0.69	1.34	1.75	2.12	2.58	2.92
17	0.69	1.33	1.74	2.11	2.57	2.90
18	0.69	1.33	1.73	2.10	2.55	2.88
19	0.69	1.33	1.73	2.09	2.54	2.86
20	0.69	1.33	1.72	2.09	2.53	2.85
21	0.69	1.32	1.72	2.08	2.52	2.83
22	0.69	1.32	1.72	2.07	2.51	2.82
23	0.69	1.32	1.71	2.07	2.50	2.81
24	0.68	1.32	1.71	2.06	2.49	2.80
25	0.68	1.32	1.71	2.06	2.49	2.79

A CHI-SQUARE TABLE

The chi-square curve, with degrees of freedom shown along the left of the table

The shaded area is shown along the top of the table

← is shown in the body of the table

Degrees of freedom	99%	95%	90%	70%	50%	30%	10%	5%	1%
1	0.00016	0.0039	0.016	0.15	0.46	1.07	2.71	3.84	6.64
2	0.020	0.10	0.21	0.71	1.39	2.41	4.60	5.99	9.21
3	0.12	0.35	0.58	1.42	2.37	3.67	6.25	7.82	11.34
4	0.30	0.71	1.06	2.20	3.36	4.88	7.78	9.49	13.28
5	0.55	1.14	1.61	3.00	4.35	6.06	9.24	11.07	15.09
6	0.87	1.64	2.20	3.83	5.35	7.23	10.65	12.59	16.81
7	1.24	2.17	2.83	4.67	6.35	8.38	12.02	14.07	18.48
8	1.65	2.73	3.49	5.53	7.34	9.52	13.36	15.51	20.09
9	2.09	3.33	4.17	6.39	8.34	10.66	14.68	16.92	21.67
10	2.56	3.94	4.86	7.27	9.34	11.78	15.99	18.31	23.21
11	3.05	4.58	5.58	8.15	10.34	12.90	17.28	19.68	24.73
12	3.57	5.23	6.30	9.03	11.34	14.01	18.55	21.03	26.22
13	4.11	5.89	7.04	9.93	12.34	15.12	19.81	22.36	27.69
14	4.66	6.57	7.79	10.82	13.34	16.22	21.06	23.69	29.14
15	5.23	7.26	8.55	11.72	14.34	17.32	22.31	25.00	30.58
16	5.81	7.96	9.31	12.62	15.34	18.42	23.54	26.30	32.00
17	6.41	8.67	10.09	13.53	16.34	19.51	24.77	27.59	33.41
18	7.00	9.39	10.87	14.44	17.34	20.60	25.99	28.87	34.81
19	7.63	10.12	11.65	15.35	18.34	21.69	27.20	30.14	36.19
20	8.26	10.85	12.44	16.27	19.34	22.78	28.41	31.41	37.57

Source: Adapted from p. 112 of Sir R. A. Fisher, *Statistical Methods for Research Workers* (Edinburgh: Oliver & Boyd. 1958).