AN INTRODUCTION TO STATISTICS

WITH

DATA ANALYSIS

by

SHELLEY RASMUSSEN

Department of Mathematical Sciences
Olney 428T
University of Massachusetts/Lowell
Lowell, MA 01854

Originally published by Brooks/Cole Publishing Company,
Division of Wadsworth, Inc.

ISBN 0-534-13578-1

Shelley_Rasmussen@uml.edu

Minitab is a statistical package, a computer program that performs many statistical procedures. The versions of Minitab now available for use on personal computers are menu-driven and much easier to use than the main-frame version originally discussed in this text. Those sections are not included in this online edition of the text. At this time, the most recent version is Minitab 15, available at very reasonable prices for purchase or rental from:

www.e-academy.com/minitab

System Requirements

Processor:	PC with a 1 GHz 32- or 64-bit processor
Memory:	512 MB or more of available RAM
Disk Space:	125 MB free space available
Operating System:	Microsoft Windows 2000, XP, or Vista.
Display:	A display capable of 1024 X 768 or higher resolution
Software:	Adobe Acrobat Reader 5.0 or higher for Meet Minitab

Additional Exercises

For each exercise, plot the observations in any ways that seem helpful. Describe the population (whether real or hypothetical) sampled.

Consider the forms of statistical inference (if any) that are appropriate. For some exercises, no formal statistical analysis may be appropriate. For others, several types of analysis may be reasonable.

For hypothesis testing, state the null and alternative hypotheses. For hypothesis testing and interval estimation, state the assumptions that make the statistical analysis appropriate. Do the assumptions seem reasonable? State conclusions based on your analysis.

When more than one analysis is possible, try each procedure. Use similar confidence levels for confidence intervals calculated under differing assumptions. Compare and discuss results.

Discuss possible extraneous factors that might have influenced results. What additional information would you want about the experiment or how the data were collected? Discuss reasons for caution in interpreting results of the analysis.

Give a complete discussion of the questions that can be answered by the data set. Discuss the information provided by the results of the experiment.

EXERCISE R-1

Experimenters wanted to investigate the effects of four different baking temperatures and five different recipes on the size of cakes. They baked two cakes for each recipe/temperature combination. The recorded response for each cake was the area of a cross-section in square inches. The results are shown below (Johnson, 1976; from Li, 1964). Discuss the results of this experiment.

	Temperature (°C)			
	149	163	190	218
Recipe A:	3.01, 4.08	4.63, 4.63	4.59, 4.45	4.26, 4.49
Recipe B:	3.87, 3.74	4.56, 4.91	4.75, 5.10	5.35, 5.39
Recipe C:	4.13, 4.03	4.80, 4.86	5.30, 5.57	5.67, 5.67
Recipe D:	3.98, 4.11	4.79, 4.88	5.00, 5.02	5.30, 5.67
Recipe E:	4.16, 4.35	4.65, 4.80	5.41, 5.29	5.52, 5.80

EXERCISE R-2

In this dental study, investigators wanted to compare a new treatment with placebo for effects on dental hygiene. The response variable is an index of oral hygiene. The 34 volunteers in group A used the placebo for a fixed period and later used the new treatment for a fixed period. The 30 volunteers in group B used the new treatment for the first fixed period and then used the placebo later during the second fixed period. We call such a study, in which each subject is treated over two or more periods, a *crossover study*. The index of oral hygiene is listed below for each period, for the volunteers in both groups (Brown, 1980; Varma and Chilton, 1974; from Zinner, Duany, and Chilton, 1970).

Group A			Group B		
Volunteer code number	Period 1 Placebo	Period 2 New treatment	Volunteer code number	Period 1 New treatment	Period 2 Placebo
1	.83	1.83	1	1.67	.33
2	1.00	2.17	2	2.50	.50
3	.67	1.67	3	1.00	−.17
4	.50	1.50	4	1.67	.50
5	.50	2.33	5	1.83	.50
6	.83	1.83	6	.50	.33
7	1.00	.50	7	1.33	.67
8	.67	.33	8	1.33	.00
9	.67	.50	9	.50	.17
10	.33	.67	10	2.17	.83
11	.00	.83	11	1.67	.33
12	1.17	1.33	12	1.50	.00
13	.00	.67	13	1.33	.50
14	.50	1.83	14	1.50	.50
15	.33	1.50	15	1.33	.00
16	.33	1.50	16	.67	−.17
17	.50	1.17	17	1.67	.50
18	1.00	1.67	18	2.50	.67
19	.00	1.33	19	1.83	.00
20	.50	1.50	20	.83	.67
21	−.50	2.83	21	2.33	.17
22	.17	2.33	22	1.17	.50
23	1.00	1.33	23	1.33	.00
24	1.00	1.67	24	1.33	.83
25	1.33	.67	25	.33	1.33
26	.33	.83	26	2.17	1.17
27	2.00	1.00	27	1.00	.33
28	4.00	.17	28	.33	1.00
29	.83	1.67	29	1.17	.17
30	.50	1.33	30	.50	.50
31	.50	1.50			
32	.50	1.67			
33	2.17	1.33			
34	.67	1.17			

In any crossover study, we worry that during later treatment periods there may be a residual or carryover effect from earlier treatments. Discuss this concern with respect to this dental study. Discuss the results of this experiment.

EXERCISE R-3 In this experiment, investigators studied a method of determining aflatoxin levels in a lot of contaminated peanuts. (This problem is important in sampling inspection. Inspectors want to protect consumers while not rejecting too many good peanuts.) They ground the peanuts into meal and divided the meal into separate samples. They blended each sample in a chemical solution. For each sample, the investigators divided the blend equally among 16 centrifuge bottles. They measured aflatoxin concentration for each bottle. For three samples, one observation was lost, leaving 15 aflatoxin determinations for each of those samples. The determination of aflatoxin concentration for each bottle (units not given) is shown below (Quesenberry, Whitaker, and Dickens, 1976; from Waltking, Bleffert, and Kiernan, 1968). Discuss your findings as they relate to the problem of quality control and consumer protection.

Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7	Sample 8
22.35	30.02	10.84	28.60	34.30	16.71	7.63	52.56
32.54	26.26	19.31	32.59	27.90	18.11	6.37	112.86
23.23	26.48	13.12	37.24	34.87	11.19	6.97	69.89
68.31	36.38	13.20	24.61	32.56	11.98	5.28	100.15
27.51	48.01	12.39	24.71	31.67	24.86	6.74	87.14
27.44	49.50	17.20	35.83	29.06	20.18	6.74	83.42
28.53	15.68	12.61	48.93	32.00	17.49	12.06	82.88
29.49	30.79	18.07	37.99	33.40	9.23	7.53	64.63
51.98	22.03	18.19	29.19	30.55	15.35	8.81	73.98
28.87	26.54	16.65	28.89	31.87	21.12	13.58	112.11
21.80	22.74	15.85	32.46	29.17	17.23	12.64	97.59
29.12	35.36	13.53	40.31	26.05	18.67	10.71	85.36
36.65	51.61	11.06	35.88	36.50	23.17	5.41	82.19
40.76	27.53	15.25	31.04	27.51	17.05	7.46	94.55
23.81	36.71	16.25	32.05	30.67	15.43	4.39	60.24
36.19		12.39		31.76	16.82	11.85	

EXERCISE R-4 In this experiment, investigators exposed adult flour beetles to several doses of gaseous carbon disulphide. The results are shown below (Prentice, 1976; from Bliss, 1935). Discuss the relationship between dose and percentage of beetles killed.

Dose (logarithm base-10 of CS_2 in mg/liter)	Number of beetles	Number killed
1.6907	59	6
1.7242	60	13
1.7552	62	18
1.7842	56	28
1.8113	63	52
1.8369	59	53
1.8610	62	61
1.8839	60	60

EXERCISE R-5 Mental retardation is associated with some metabolic diseases. In this experiment, investigators studied metabolism of tyrosine among 36 mentally handicapped patients (Geertsema and Reinecke, 1984). The measured response was the total amount of tyrosine catabolites excreted in the urine (in μmoles per 100 ml urine). Ten separate measurements for each patient, plus the average and standard deviation of those ten measurements, are listed below. The patients are listed in order of increasing average tyrosine determination.

	Observation										Patient	Standard
ID	1	2	3	4	5	6	7	8	9	10	average	deviation
1	.325	.317	.375	.325	.508	.117	.150	.317	.275	.383	.309	.106
2	.333	.283	.342	.325	.250	.358	.283	.392	.450	.267	.328	.058
3	.208	.483	.317	.300	.217	.217	.433	.392	.575	.408	.355	.118
4	.458	.592	.133	.600	.292	.542	.467	.300	.067	.233	.368	.181
5	.225	.317	.492	.617	.217	.308	.425	.508	.425	.258	.379	.128
6	.100	.317	.675	.625	.133	.542	.150	.642	.183	.442	.381	.219
7	.233	.267	.300	.217	.258	.700	.483	.667	.408	.300	.383	.169
8	.433	.508	.350	.275	.342	.333	.667	.300	.425	.275	.391	.116
9	.133	.800	.575	.617	.283	.433	.275	.308	.283	.225	.393	.198
10	.500	.417	.358	.708	.283	.233	.483	.300	.258	.567	.411	.146
11	.250	.442	.233	.183	.275	.367	.408	.692	.883	.667	.440	.221
12	.133	.200	.417	.750	.083	.833	.183	.842	.767	.233	.444	.301
13	.175	.458	.358	.533	.242	.400	.308	.650	.567	.783	.447	.179
14	.650	.567	.500	.808	.508	.442	.400	.442	.108	.217	.464	.190
15	.217	.183	.667	.767	.900	.183	.608	.700	.233	.592	.505	.259
16	.625	.533	.292	.367	.683	.667	.433	.783	.200	.625	.521	.180
17	.442	.567	.412	.625	.200	.683	.892	.175	.883	.658	.554	.236
18	.642	.408	.500	.733	.317	.700	.617	.667	.683	.667	.593	.131
19	.900	1.067	.950	.442	.542	.317	.183	.483	.942	.300	.613	.305
20	.733	.433	.383	.550	.383	.650	.717	.833	.817	.700	.620	.163
21	.617	.667	.683	.308	.567	.858	.358	.483	.833	.890	.628	.194
22	.842	.808	.408	.492	.683	.500	.350	.833	1.083	.500	.650	.224
23	.683	.700	.633	.850	.775	.750	.067	.867	.950	.467	.674	.240
24	.733	1.025	1.183	.417	.542	.833	.275	.417	.483	1.058	.697	.300
25	.167	.792	.833	.400	1.142	.142	.683	.542	1.375	.867	.699	.378
26	.367	1.075	.700	1.350	.467	.200	.700	.167	.475	1.650	.715	.471
27	1.012	.442	1.000	.583	.633	.783	.383	1.550	.417	.442	.725	.353
28	.983	.600	.417	.450	.667	.800	.550	.750	1.242	1.083	.754	.261
29	.708	2.200	.292	1.633	1.417	.450	.233	.233	.367	.650	.818	.654
30	.583	.575	.533	.292	.408	.783	.625	1.333	1.550	1.667	.835	.469
31	.767	.492	.933	.933	1.358	.683	.692	.700	.992	1.133	.868	.241
32	1.083	.542	.400	.558	.708	1.333	.550	.883	2.100	1.792	.995	.550
33	.542	1.650	.983	.892	.333	1.267	.750	1.142	1.625	1.808	1.099	.468
34	1.550	1.433	.550	.583	.808	1.400	1.133	1.467	1.300	.925	1.115	.355
35	.850	.900	1.242	.917	.092	1.267	.983	1.483	2.367	1.750	1.185	.576
36	2.167	1.600	.717	1.675	1.525	.758	.808	2.283	2.333	3.067	1.693	.743

The investigators selected 1.0 μmole per 100 ml urine as a cutoff for classifying a patient's tyrosine metabolism as clinically negative (less than 1.0, no metabolic problem) or clinically positive (greater than 1.0, a metabolic disorder). As part of your discussion, address the following questions:

a. Is it reasonable to use a single measurement to classify a person as clinically negative or positive for tyrosine metabolism?

b. How consistent are separate determinations in a single patient?

c. How different are the standard deviations for these patients?

d. How different are the tyrosine determinations between patients?

e. Each of these 36 patients had been classified as clinically positive for a tyrosine metabolic disorder after a first screening. Discuss this finding in relation to the later results listed above.

f. Does it appear that the authors use the population or sample formula to calculate their standard deviations? Which would you consider appropriate here? Does this affect your answer to part (c)?

EXERCISE R-6

Investigators estimated specific activity of the enzyme sucrase using samples of intestine from 24 patients who had had intestinal bypass surgery. For each patient, the investigators determined sucrase activity in two ways. The investigators believe the first method, using pellet fractions, is accurate, but it is very time-consuming. The second method, from homogenates, is easier and faster to use. The two recorded levels of sucrase activity (units not given) are listed below for each patient (Carter, 1981; data provided by Dr. Helen Lane). Assess the idea of using the homogenate measurement of sucrase activity as an estimate of the actual (or pellet) level of sucrase activity.

Patient	Homogenate	Pellet	Patient	Homogenate	Pellet
1	18.88	70.00	13	60.78	277.30
2	7.26	55.43	14	77.92	331.50
3	6.50	18.87	15	51.29	133.74
4	9.83	40.41	16	77.91	221.50
5	46.05	57.43	17	36.65	132.93
6	20.10	31.14	18	31.17	85.38
7	35.78	70.10	19	66.09	142.34
8	59.42	137.56	20	115.15	294.63
9	58.43	221.20	21	95.88	262.52
10	62.32	276.43	22	64.61	183.56
11	88.55	316.00	23	37.71	86.12
12	19.50	75.56	24	100.82	226.55

EXERCISE R-7

In this experiment, investigators wanted to compare antagonistic behavior in mice with different brain weights. They bred mice selectively for brain weight (small, medium, large). Under each of two environmental conditions, the experimenters raised seven pairs of mice with each brain weight. After the mice reached maturity, the investigators housed the two mice in a pair together, and noted their fighting behavior. The response variable is a measure of aggressive behavior in mice: a score of seconds of tail rattling per second of fighting. The results are shown below (Scheirer, Ray, and Hare, 1976; from Hahn, Haber, and Fuller, 1973). Discuss the results of this experiment.

	Brain weight		
	Low	Medium	Large
Environmental condition A	9.60	1.98	.77
	8.00	1.81	.66
	5.14	1.37	.37
	3.50	.98	.40
	3.23	.27	.19
	2.66	.00	.13
	1.44	.00	.00
Environmental condition B	4.36	2.61	2.95
	1.49	2.09	2.73
	1.45	1.76	.91
	1.35	1.27	.82
	.33	.84	.69
	.20	.64	.11
	.00	.00	.00

EXERCISE R-8 Investigators carried out a 6-month double-blind randomized study to compare a new treatment with a standard treatment for patients with acute rheumatoid arthritis. At the end of the study, the investigators classified the condition of each patient into one of five categories. The results are summarized in the following frequency table (Mehta, Patel, and Tsiatis, 1984). Does there appear to be a difference in effectiveness between the two treatments or do they appear to be therapeutically equivalent?

	Treatment	
Patient status	New	Standard
Much improved	24	11
Improved	37	51
No change	21	22
Worse	19	21
Much worse	6	7

EXERCISE R-9 At a United States industrial site, public health workers used grab-sample techniques to measure concentration of airborne chlorine. They made 15 chlorine determinations over the course of one working day (Owen and DeRouen, 1980):

Chlorine determination in parts per million (ppm):	6 .5	0 2	6 2	9 0	6.5 0	0 1	0	0	1

At the time, federal guidelines stated that the maximum allowable exposure for chlorine was an average of 1 ppm over the work day. Was this site in compliance with the federal occupational safety standard for chlorine? You may

wish to use the information that the minimum detectable level of chlorine was .25 ppm.

EXERCISE R-10 In this experiment, investigators recorded the length of time for patients with headaches to feel relief. Each patient received a standard treatment and a new treatment, on different occasions. The time until relief from headache (in minutes) is shown below for each patient and treatment (Gross and Lam, 1981; from Gross and Clark, 1975, page 232). Does there appear to be a difference between the two treatments in time until relief?

Patient	New treatment	Standard treatment
1	6.9	8.4
2	6.8	7.7
3	10.3	10.1
4	9.4	9.6
5	8.0	9.3
6	8.8	9.1
7	6.1	9.0
8	7.4	7.7
9	8.0	8.1
10	5.1	5.3

EXERCISE R-11 In a study of cyanotic heart disease in children, investigators recorded the age at first word and the Gesell Adaptive Score for each of 21 children (Ellenberg, 1976; from Mickey, Dunn, and Clark, 1967). How would you describe the relationship between these two variables for this group of children?

Child	Age at first word (months)	Gesell Adaptive Score	Child	Age at first word (months)	Gesell Adaptive Score
1	15	95	12	9	96
2	26	71	13	10	83
3	10	83	14	11	84
4	9	91	15	11	102
5	15	102	16	10	100
6	20	87	17	12	105
7	18	93	18	42	57
8	11	100	19	17	121
9	8	104	20	11	86
10	20	94	21	10	100
11	7	113			

EXERCISE R-12 Investigators carried out a double-blind multiclinic trial to compare a new agent with aspirin in treatment of patients with rheumatoid arthritis. A frequency table summarizing the investigators' assessment of therapeutic effect after 26 weeks of treatment is shown below (Gould, 1980).

Response category	New agent	Aspirin
Excellent	28	23
Satisfactory	44	33
No change	8	10
Worse	2	0
Withdrawn from study	49	72

As we see from the last line of the table, many patients withdrew before the end of the study. Patients may withdraw for reasons related to treatment (feeling cured or feeling worse) or reasons unrelated to treatment (loss of interest, moving away, illness or death unrelated to the condition under study).

a. Compare withdrawal rates for the two treatment groups.

b. Ignoring withdrawals, compare responses for the two treatment groups.

c. Based on your results in parts (a) and (b), how would you assess the relative effectiveness of these two treatments for rheumatoid arthritis?

d. Suppose you learn that all withdrawals were due to treatment intolerance or lack of therapeutic effect. Now how would you assess the relative effectiveness of these two treatments?

EXERCISE R-13 In a study of heart disease in Framingham, Massachusetts, investigators measured serum cholesterol of participants and noted participants who suffered from coronary heart disease 6 years later (Leung and Kupper, 1981; from Walter, 1978):

Coronary heart disease six years later	Initial serum cholesterol (mg %)	
	<220	≥220
Yes	20	72
No	553	684

What can you say about the association between initial serum cholesterol level and coronary heart disease 6 years later for participants of this study?

EXERCISE R-14 Around 1950 medical workers began to administer continuous high concentrations of oxygen to premature babies who showed difficulty breathing. (Premature babies sometimes died because of impaired breathing.) Also around 1950, many premature infants developed retrolental fibroplasia, which led to blindness. Some workers suspected a link between the high doses of oxygen administered and subsequent blindness in these babies. In one study to investigate this idea, experimenters randomly divided 85 premature babies into two treatment groups: higher-dose oxygen and lower-dose oxygen. The researchers noted whether each baby survived to 3 months of age. They classified sur-

viving babies by blindness following treatment. The results are summarized below (Meier, 1979; from Lanman et al., 1954).

Survival:	Treatment group	Number of babies	Number alive at age 3 months
	Higher-dose oxygen	45	36
	Lower-dose oxygen	40	28

Blindness:	Treatment group	Alive at 3 months	Number blind
	Higher-dose oxygen	36	8
	Lower-dose oxygen	28	0

Discuss the results of this study. Because of studies such as this one, medical workers stopped using high doses of oxygen on premature babies and blindness from retrolental fibroplasia became once again a rare condition.

EXERCISE R-15 An investigator carried out this experiment to study the specific retention volume of the organic liquid methylene chloride in the polymer polyethylene terephthalate, at several temperatures. The results are shown below (Gallant, 1977; from P. O. Hsiung, "Study of the Interaction of Organic Liquids with Polymers by Means of Gas Chromatography," unpublished Ph.D. dissertation, North Carolina State University, 1974). Describe the relationship between reciprocal of temperature and the natural logarithm of specific volume seen in this experiment.

Reciprocal $\times 10^3$ of temperature in degrees Kelvin	Natural logarithm of specific volume in cc per gm
2.54323	1.16323
2.60960	1.10458
2.67952	.98832
2.75330	.87471
2.79173	.62060
2.82965	.51175
2.87026	.35371
2.91120	.66954
2.94637	.85555
3.00030	1.07086
3.04228	1.22272
3.09214	1.29113
3.13971	1.38480
3.19081	1.46728

EXERCISE R-16 Listed below are weight/height ratios by age for preschool boys, from a nutritional study of preschool children in the north central United States (Gallant, 1977; Gallant and Fuller, 1973; from Eppright et al., 1972). Age is age in months and W/H is weight divided by height, in pounds per inch. Describe the relationship between weight/height ratio and age for preschool boys suggested by this study.

Age	W/H	Age	W/H	Age	W/H	Age	W/H
.5	.46	18.5	.81	36.5	.87	54.5	.93
1.5	.47	19.5	.78	37.5	.87	55.5	.98
2.5	.56	20.5	.87	38.5	.85	56.5	.95
3.5	.61	21.5	.80	39.5	.90	57.5	.97
4.5	.61	22.5	.83	40.5	.87	58.5	.97
5.5	.67	23.5	.81	41.5	.91	59.5	.96
6.5	.68	24.5	.88	42.5	.90	60.5	.97
7.5	.78	25.5	.81	43.5	.93	61.5	.94
8.5	.69	26.5	.83	44.5	.89	62.5	.96
9.5	.74	27.5	.82	45.5	.89	63.5	1.03
10.5	.77	28.5	.82	46.6	.92	64.5	.99
11.5	.78	29.5	.86	47.5	.89	65.5	1.01
12.5	.75	30.5	.82	48.5	.92	66.5	.99
13.5	.80	31.5	.85	49.5	.96	67.5	.99
14.5	.78	32.5	.88	50.5	.92	68.5	.97
15.5	.82	33.5	.86	51.5	.91	69.5	1.01
16.5	.77	34.5	.91	52.5	.95	70.5	.99
17.5	.80	35.5	.87	53.5	.93	71.5	1.04

EXERCISE R-17 In a study of metabolism, investigators incubated isolated liver cells from starved rats with lactate. They then noted accumulation of pyruvate (a first step in glucose production) over time. The results of duplicate observations at each time are shown below (James and Conyers, 1985). Describe the relationship between pyruvate level and time in this experiment.

Time (minutes)	Pyruvate (μmoles)
.05	.0135, .0115
3.05	.3320, .3500
6.05	.4755, .4885
9.05	.5560, .5430
12.05	.5945, .6025
15.05	.6130, .6195
18.05	.6135, .6225
21.05	.6095, .5640
24.05	.5690, .5505
27.05	.5400, .5250
30.05	.5105, .4965
33.05	.5050, .4630
36.05	.4400, .4305
39.05	.3900, .4190

EXERCISE R-18

In a National Cancer Institute animal carcinogenesis experiment, investigators divided male and female mice and rats into three treatment groups. The control groups received none of the drug tolazamide. Animals in the low-dose groups received tolazamide as .5% of their diet. Animals in the high-dose groups received tolazamide as 1% of their diet. The investigators noted development of leukemia or lymphoma in the animals. The results are shown below as number developing leukemia or lymphoma/number in group (Tarone and Gart, 1980). Discuss the results of this experiment.

	Treatment group		
Sex/Species	Control	Low dose	High dose
Female mice	6/15	2/33	4/34
Male mice	4/14	5/35	1/34
Female rats	4/15	3/33	2/35
Male rats	2/15	1/35	4/35

EXERCISE R-19

In this study of the association between hyperglycemia and relative hyperinsulinemia, investigators administered standard glucose tolerance tests to 13 control and 20 obese patients on the Pediatric Clinical Research Ward, University of Colorado Medical Center. As part of the study, the investigators determined plasma inorganic phosphate levels (mg/dl) from blood samples taken 0, $\frac{1}{2}$, 1, $1\frac{1}{2}$, 2, 3, 4, and 5 hours after a standard-dose oral glucose challenge. The results are listed below (Zerbe, 1979). Discuss the results of this experiment.

	Hours after glucose challenge							
	0	$\frac{1}{2}$	1	$1\frac{1}{2}$	2	3	4	5
Control patients								
1	4.3	3.3	3.0	2.6	2.2	2.5	3.4	4.4
2	3.7	2.6	2.6	1.9	2.9	3.2	3.1	3.9
3	4.0	4.1	3.1	2.3	2.9	3.1	3.9	4.0
4	3.6	3.0	2.2	2.8	2.9	3.9	3.8	4.0
5	4.1	3.8	2.1	3.0	3.6	3.4	3.6	3.7
6	3.8	2.2	2.0	2.6	3.8	3.6	3.0	3.5
7	3.8	3.0	2.4	2.5	3.1	3.4	3.5	3.7
8	4.4	3.9	2.8	2.1	3.6	3.8	4.0	3.9
9	5.0	4.0	3.4	3.4	3.3	3.6	4.0	4.3
10	3.7	3.1	2.9	2.2	1.5	2.3	2.7	2.8
11	3.7	2.6	2.6	2.3	2.9	2.2	3.1	3.9
12	4.4	3.7	3.1	3.2	3.7	4.3	3.9	4.8
13	4.7	3.1	3.2	3.3	3.2	4.2	3.7	4.3

(*continued*)

(continued)

	Hours after glucose challenge							
	0	½	1	1½	2	3	4	5
Obese patients								
1	4.3	3.3	3.0	2.6	2.2	2.5	2.4	3.4
2	5.0	4.9	4.1	3.7	3.7	4.1	4.7	4.9
3	4.6	4.4	3.9	3.9	3.7	4.2	4.8	5.0
4	4.3	3.9	3.1	3.1	3.1	3.1	3.6	4.0
5	3.1	3.1	3.3	2.6	2.6	1.9	2.3	2.7
6	4.8	5.0	2.9	2.8	2.2	3.1	3.5	3.6
7	3.7	3.1	3.3	2.8	2.9	3.6	4.3	4.4
8	5.4	4.7	3.9	4.1	2.8	3.7	3.5	3.7
9	3.0	2.5	2.3	2.2	2.1	2.6	3.2	3.5
10	4.9	5.0	4.1	3.7	3.7	4.1	4.7	4.9
11	4.8	4.3	4.7	4.6	4.7	3.7	3.6	3.9
12	4.4	4.2	4.2	3.4	3.5	3.4	3.9	4.0
13	4.9	4.3	4.0	4.0	3.3	4.1	4.2	4.3
14	5.1	4.1	4.6	4.1	3.4	4.2	4.4	4.9
15	4.8	4.6	4.6	4.4	4.1	4.0	3.8	3.8
16	4.2	3.5	3.8	3.6	3.3	3.1	3.5	3.9
17	6.6	6.1	5.2	4.1	4.3	3.8	4.2	4.8
18	3.6	3.4	3.1	2.8	2.1	2.4	2.5	3.5
19	4.5	4.0	3.7	3.3	2.4	2.3	3.1	3.3
20	4.6	4.4	3.8	3.8	3.8	3.6	3.8	3.8

EXERCISE R-20 In a life-testing experiment, investigators recorded the time to breakdown of an insulating fluid under two elevated levels of voltage stress. (They used elevated stress levels because time to breakdown under voltages ordinarily used would be too long to measure.) The results are shown below (Nair, 1984; from Nelson, 1982). Compare the times to breakdown under these two voltage levels.

Voltage level	Time to breakdown (minutes)							
32 Kv	.27	.40	.69	.79	2.75	3.91	9.88	13.95
	15.93	27.80	53.24	82.85	89.29	100.58	215.10	
36 Kv	.35	.59	.96	.99	1.69	1.97	2.07	2.58
	2.71	2.90	3.67	3.99	5.35	13.77	25.50	

EXERCISE R-21 Scientists report using a sensitive genetic test to detect HIV, the AIDS virus, in newborn babies (I. Wickelgren, "Test diagnoses AIDS in newborns," *Science News,* volume 135, June 24, 1989, page 389). Such a test would have great usefulness in diagnosing babies born to HIV-positive mothers; babies needing intensive treatment would be identified earlier and babies not needing the treatment would be spared its toxic effects.

Using their genetic test, the researchers detected HIV DNA in five of seven newborns who later developed AIDS. The researchers detected HIV DNA in one of eight newborns who later showed symptoms of possible HIV infection. The researchers detected no HIV DNA in any of nine newborns who were still healthy after 16 months.

Discuss these experimental results.

EXERCISE R-22 In the 1974 and 1975 General Social Surveys of the National Opinion Research Center, interviewers asked men the following question: "Do you agree with this statement?—Women should take care of running their homes and leave running the country to men." The accompanying table classifies the respondents by years of education and response to the question (Haberman, 1982). What relationship (if any) do you see between years of education and response to this question, for these men?

Years of education	Agree	Disagree
0	4	2
1	2	0
2	4	0
3	6	3
4	5	5
5	13	7
6	25	9
7	27	15
8	75	49
9	29	29
10	32	45
11	36	59
12	115	245
13	31	70
14	28	79
15	9	23
16	15	110
17	3	29
18	1	28
19	2	13
20+	3	20

EXERCISE R-23 Investigators wanted to compare four formulations of the drug verapamil. They randomly divided 26 healthy male volunteers into four groups. They treated each group with one of the drug formulations. The recorded response was area under the plasma time curve. In addition, the investigators recorded the age, height, and weight of each volunteer. The results (no units provided) are listed below (these are the first-period data from a larger experiment reported in Chinchilli, Schwab, and Sen, 1989; data provided by William H. Barr of the

Pharmacy and Pharmaceutics Department, Virginia Commonwealth University). Discuss the results of this experiment.

Volunteer	Treatment	Age	Height	Weight	Area under the plasma time curve
1	A	25	68.0	145.0	224.29
2	B	29	66.5	140.0	231.35
3	C	29	68.5	155.0	253.88
4	D	23	70.0	188.0	327.95
5	A	22	66.5	140.0	326.06
6	B	22	72.0	170.0	259.53
7	D	23	68.0	165.0	347.43
8	A	24	72.0	203.0	270.10
9	B	24	71.0	160.0	618.61
10	C	24	63.0	135.0	476.27
11	D	22	68.5	149.5	337.45
12	A	28	71.0	210.0	483.25
13	B	23	70.0	150.0	223.04
14	C	26	72.0	159.0	399.92
15	D	22	70.0	150.0	117.45
16	B	26	70.5	160.0	183.20
17	C	28	75.0	240.0	344.18
18	D	24	72.0	173.0	181.75
19	A	31	70.0	145.0	94.25
20	B	31	69.0	170.0	195.67
21	C	25	78.0	193.0	458.89
22	D	29	72.0	205.0	383.64
23	A	24	74.0	150.0	413.53
24	B	26	72.0	161.0	132.88
25	C	23	73.0	172.0	245.21
26	D	22	70.0	165.0	298.06

EXERCISE R-24

A frequency table summarizing the sex of each child (by birth order) in completed families of 2, 3, 4, and 5 children in the United States is shown below (Crouchley and Pickles, 1984). Discuss what you find from studying this data set.

Number of children	Sex by birth order	Number of families	Number of children	Sex by birth order	Number of families
2	MM	4,862	3	MMM	2,467
	MF	4,854		MMF	2,504
	FM	5,133		MFM	2,432
	FF	4,432		MFF	2,172
				FMM	2,298
				FMF	2,178
				FFM	2,373
				FFF	1,988

Number of children	Sex by birth order	Number of families	Number of children	Sex by birth order	Number of families
4	MMMM	1,133	5	MMMMM	2,242
	MMMF	1,140		MMMMF	2,187
	MMFM	1,106		MMMFM	2,122
	MMFF	1,046		MMMFF	2,092
	MFMM	1,105		MMFMM	2,062
	MFMF	1,049		MMFMF	1,970
	MFFM	1,094		MMFFM	1,968
	MFFF	982		MMFFF	1,878
	FMMM	1,085		MFMMM	2,187
	FMMF	1,019		MFMMF	2,019
	FMFM	1,071		MFMFM	2,018
	FMFF	935		MFMFF	1,857
	FFMM	1,028		MFFMM	2,011
	FFMF	1,010		MFFMF	1,874
	FFFM	952		MFFFM	1,810
	FFFF	913		MFFFF	1,800
				FMMMM	2,088
				FMMMF	1,948
				FMMFM	1,931
				FMMFF	1,882
				FMFMM	1,957
				FMFMF	1,903
				FMFFM	1,923
				FMFFF	1,770
				FFMMM	1,995
				FFMMF	1,820
				FFMFM	1,859
				FFMFF	1,765
				FFFMM	1,879
				FFFMF	1,760
				FFFFM	1,788
				FFFFF	1,732

EXERCISE R-25 In this study, investigators recorded fill weights of bottles filled by six heads of a multiple-head machine, at five different times. The value recorded is bottle fill weight in grams minus 1,200. The results are shown below (Snee, 1982; from Ott and Snee, 1973). Discuss the results of this experiment.

Time	Head 7	Head 8	Head 9	Head 10	Head 11	Head 12
9:15	68	65	75	57	32	70
12:55	56	52	55	48	65	47
1:55	40	51	52	36	49	45
2:55	84	87	88	73	34	70
3:55	50	52	52	50	45	61

EXERCISE R-26 In this experiment, eight calves from a single herd of Angus cattle arrived at a feedlot. Investigators weighed the calves and divided them among three diet regimens. After a fixed period, the calves were slaughtered. The investigators recorded the slaughter weight for each calf. The slaughter weight and weight on entering feedlot (no units given) are shown below for each calf (Urquhart, 1982).

Energy level in diet	Slaughter weight (Weight on entering feedlot)
Low	1,046 (690)
	1,027 (685)
	1,018 (690)
Medium	874 (665)
	874 (635)
High	1,018 (680)
	874 (605)
	970 (665)

As another part of the experiment, investigators weighed eight calves from a single herd of Brangus cattle and divided them among the same three diet regimens. Weights for these calves are shown below.

Energy level in diet	Slaughter weight (Weight on entering feedlot)
Low	1,133 (765)
	989 (750)
	970 (645)
	1,162 (730)
Medium	1,190 (840)
	1,104 (755)
High	1,104 (755)
	1,114 (755)

Discuss the results of this experiment.

EXERCISE R-27 Fifty-nine women participated in this study to compare an active drug with a placebo for treatment of rheumatoid arthritis. Identification number, age, and response are shown below for the women in each treatment group (Koch

et al., 1982). The responses are ranked from best to worst in this order: excellent, good, moderate, fair, poor. Discuss the results of this study.

Active treatment			Placebo treatment		
ID	Age (years)	Response	ID	Age (years)	Response
1	23	Poor	28	23	Fair
2	32	Poor	29	30	Fair
3	37	Moderate	30	30	Fair
4	41	Good	31	31	Moderate
5	41	Fair	32	32	Poor
6	48	Good	33	33	Good
7	48	Poor	34	37	Poor
8	55	Excellent	35	44	Poor
9	55	Good	36	45	Poor
10	56	Good	37	46	Poor
11	57	Good	38	48	Fair
12	57	Good	39	49	Poor
13	57	Good	40	51	Poor
14	58	Poor	41	53	Poor
15	59	Good	42	54	Good
16	59	Excellent	43	54	Poor
17	60	Excellent	44	54	Poor
18	61	Good	45	55	Good
19	62	Good	46	57	Moderate
20	62	Moderate	47	57	Fair
21	66	Excellent	48	58	Moderate
22	67	Good	49	59	Excellent
23	68	Moderate	50	59	Moderate
24	68	Excellent	51	61	Fair
25	69	Moderate	52	63	Moderate
26	69	Poor	53	64	Poor
27	70	Moderate	54	65	Excellent
			55	66	Moderate
			56	66	Fair
			57	66	Poor
			58	68	Moderate
			59	74	Good

EXERCISE R-28 Sixty-nine girls completed this 2-year study to compare three treatments in reducing incidence of dental caries. The three treatments were stannous fluoride, acid phosphate fluoride, and distilled water, all applied topically. The investigators measured the number of decayed, missing, or filled teeth before and after the study. The accompanying table provides this information, in addition to an identification number for each girl, her age in years, and a code for the institution (place) where she was treated (Quade, 1982; from Cartwright, Lindahl, and Bawden, 1968). Note that the recorded result for one girl shows a larger value before than after the study. Discuss the results of this experiment.

ID	Place	Age	Before	After	ID	Place	Age	Before	After
				Distilled water treatment					
1	1	13	7	11	18	2	7	3	4
2	1	17	20	24	19	2	11	4	7
3	1	16	21	25	20	2	15	4	9
4	1	13	1	2	37	3	8	2	4
5	1	10	3	7	38	3	16	13	18
6	1	17	20	23	39	3	14	9	12
7	1	13	9	13	40	3	16	15	18
8	1	9	2	4	41	3	12	13	17
16	2	16	10	14	42	3	8	2	5
17	2	16	13	17	43	3	14	9	12
				Stannous fluoride treatment					
9	1	14	11	13	30	2	10	4	6
10	1	14	15	18	44	3	9	4	6
21	2	14	15	18	45	3	15	10	14
22	2	11	6	8	46	3	14	7	11
23	2	9	4	6	47	3	13	14	15
24	2	17	18	19	48	3	12	7	10
25	2	14	11	12	49	3	12	3	6
26	2	13	9	9	50	3	14	9	12
27	2	9	4	7	51	3	13	8	10
28	2	9	5	7	52	3	14	19	19
29	2	15	11	14	53	3	14	10	13
				Acid phosphate fluoride					
11	1	11	7	10	57	3	9	5	8
12	1	15	17	17	58	3	11	1	3
13	1	11	9	11	59	3	12	8	9
14	1	7	1	5	60	3	14	4	5
15	1	11	3	7	61	3	10	4	7
31	2	10	4	4	62	3	12	14	14
32	2	15	7	7	63	3	11	8	10
33	2	11	0	4	64	3	11	3	5
34	2	9	3	3	65	3	16	11	12
35	2	9	0	1	66	3	15	16	18
36	2	16	8	8	67	3	10	8	8
54	3	14	10	12	68	3	6	0	1
55	3	11	7	11	69	3	7	3	4
56	3	14	13	12					

EXERCISE R-29 In a trial of a vaccine for autoimmune deficiency syndrome (AIDS), 15 of 40 people treated showed an antibody response within 2 months after treatment (*Science News,* 1988). Find an interval estimate for the proportion of the population who would respond to this vaccine.

EXERCISE R-30 Ten volunteers participated in this study of the effects of aerobic exercise on individuals who had amputation of a lower limb (Pitetti et al., 1987). Weight (including prosthesis, trousers, and shoes), resting heart rate, and resting

blood pressure were recorded for each volunteer before and after a 15-week aerobic exercise program. The results of the experiment are shown below.

Volunteer	Age (years)	Weight (kilograms)		Resting heart rate (beats/minute)	
		Before	After	Before	After
1	39	81.4	79.1	70	64
2	36	101.2	103.4	114	108
3	33	103.4	105.5	91	80
4	36	91.9	91.2	73	66
5	41	109.1	110.5	70	61
6	61	81.2	81.2	65	62
7	35	94.1	89.1	68	51
8	36	72.3	79.6	75	72
9	35	83.2	83.2	92	61
10	41	101.8	97.7	93	76

Volunteer	Resting blood pressure (mm Hg)	
	Before	After
1	120/75	120/75
2	130/108	130/100
3	133/75	140/85
4	160/87	137/75
5	130/85	125/85
6	140/80	136/80
7	140/85	118/74
8	135/93	135/87
9	125/90	125/85
10	145/75	140/75

a. Is there evidence that the exercise program had an effect on the weight of the volunteers?

b. Is there evidence that the exercise program had an effect on the resting heart rate of the volunteers?

c. Is there evidence that the exercise program had an effect on the resting blood pressure of the volunteers?

d. In parts (a), (b), and (c) we have based several statistical inferences on data from the same experiment. How does this add to our usual need for caution in interpreting results of analyses (p-values and confidence levels in particular)?