has an F-distribution with ν_1 and ν_2 degrees of freedom. Finally, the square of a t-variable with ν degrees of freedom is $F_{1,\nu}$. Summarizing:

$$\chi_\nu^2 = \sum_{i=1}^{\nu} Z_i^2, \quad t_\nu^2 = F_{1,\nu} = \frac{\chi_1^2/1}{\chi_\nu^2/\nu}.$$

A special case connects all four pivotal variables:

$$Z^2 = t_\infty^2 = \chi_1^2 = F_{1,\infty}.$$

Thus, given the F-table, all the other tables can be generated from it.

For completeness, we now summarize the mean and variance of the four fixed distributions:

Distribution	Symbol	Mean	Variance	
Normal	Z	0	1	
Student t	t_ν	0	$\frac{\nu}{\nu - 2}$	$(\nu > 2)$
Chi-square	χ_ν^2	ν	2ν	
Fisher's F	F_{ν_1,ν_2}	$\frac{\nu_2}{\nu_2 - 2}$	$\frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}$	$(\nu_2 > 4)$

Note 5.4: One-sided Tests and One-sided Confidence Intervals

Corresponding to one-sided (one-tailed) tests are one-sided confidence intervals. A one-sided confidence interval is derived from a pivotal quantity in the same was as a two-sided confidence interval. For example, in the case of a one-sample t-test, a pivotal equation is

$$P\left[-\infty \le \frac{\overline{x} - \mu}{s/\sqrt{n}} \le t_{n-1,1-\alpha}\right] = 1 - \alpha.$$

Solving for μ produces a $100(1 - \alpha)\%$ *upper* one-sided confidence interval for μ: $(\overline{x} - t_{n-1,1-\alpha}s/\sqrt{n}, \infty)$. Similar intervals can be constructed for all the pivotal variables.

PROBLEMS

1. Rickman *et al.* [1974] made a study of changes in serum cholesterol and triglyceride levels of subjects following the "Stillman diet." The diet consists primarily of protein and animal fats, restricting carbohydrate intake. The subjects followed the diet with length of time varying from 3 to 17 days. The mean cholesterol level increased significantly from 215 mg/100 ml at baseline to 248 mg/100 ml at the end of the diet. In this problem, we deal with the triglyceride level.

Subject	Days on Diet	Weight (in kg) Initial	Weight (in kg) Final	Triglyceride (in mg/100 ml) Baseline	Triglyceride (in mg/100 ml) Final
1	10	54.6	49.6	159	194
2	11	56.4	52.8	93	122
3	17	58.6	55.9	130	158
4	4	55.9	54.6	174	154
5	9	60.0	56.7	148	93
6	6	57.3	55.5	148	90
7	3	62.7	59.6	85	101
8	6	63.6	59.6	180	99
9	4	71.4	69.1	92	183
10	4	72.7	70.5	89	82
11	4	49.6	47.1	204	100
12	7	78.2	75.0	182	104
13	8	55.9	53.2	110	72
14	7	71.8	68.6	88	108
15	7	71.8	66.8	134	110
16	14	70.5	66.8	84	81

a. Make a stem-and-leaf diagram of the *changes* in triglyceride levels.

b. Calculate the average change in triglyceride level. Calculate the standard error of the difference.

c. Test the significance of the average change.

d. Construct a 90% confidence interval on the difference.

e. The authors indicate that subjects (5, 6), (7, 8), (9, 10) and (15, 16) were "repeaters"; that is, the same subjects who followed the diet for two sequences. Do you think it is reasonable to include their data on the "second time around" with that of the other subjects? Supposing not, how would you now analyze the data? Carry out the analysis. Does it change your conclusions?

2. The following data are from Dobson *et al.* [1976]. Thirty-six patients with a confirmed diagnosis of phenylketonuria (PKU) were identified and placed on dietary therapy before reaching 121 days of age. The children were tested for IQ (Stanford–Binet) between the ages of four and six; subsequently, their normal siblings of closest age were also tested with the Stanford–Binet. The following are the first 15 pairs listed in the paper:

Pair	1	2	3	4	5	6	7	8
IQ of PKU case	89	98	116	67	128	81	96	116
IQ of sibling	77	110	94	91	122	94	121	114

	9	10	11	12	13	14	15
IQ of PKU case	110	90	76	71	100	108	74
IQ of sibling	88	91	99	93	104	102	82

a. State a suitable null and an alternative hypotheses with regard to these data.

b. Test the null hypothesis.

c. State your conclusions.

d. What are your assumptions?

e. Discuss the concept of power with respect to this set of data using the fact that PKU invariably led to mental retardation until the cause was found and treatment consisting of restricted diet was instituted.

f. The mean difference (PKU case – sibling) in IQ for the full 36 pairs was −5.25; the standard deviation of the difference was 13.18. Test the hypothesis of no difference in IQ for this whole set of data.

3. Data by Mazze et al. [1971] deals with the pre-operative and post-operative creatinine clearance (ml/min) of six patients anesthetized by halothane:

	Patient					
	1	**2**	**3**	**4**	**5**	**6**
Preoperative	110	101	61	73	143	118
Postoperative	149	105	162	93	143	100

a. Why is the paired t-test preferable to the two-sample t-test in this case?

b. Carry out the paired t-test and test the significance of the difference.

c. What is the model for your analysis?

d. Set up a 99% confidence interval on the difference.

e. Graph the data by plotting the pairs of values for each patient.

4. Some of the physiological effects of alcohol are well known. A paper by Squires *et al.* [1978] assessed the acute effects of alcohol on auditory brainstem potentials in humans. Six volunteers (including the three authors) participated in the study. The latency (delay) in response to an auditory stimulus was measured before and after an intoxicating dose of alcohol. Seven different peak responses were identified. In this exercise, we only discuss latency peak no. 3. The measurements of the latency of peak (in milliseconds after the stimulus onset) in the six subjects were as follows:

	Latency of Peak					
	1	**2**	**3**	**4**	**5**	**6**
Before alcohol	3.85	3.81	3.60	3.68	3.78	3.83
After alcohol	3.82	3.95	3.80	3.87	3.88	3.94

a. Test the significance of the difference at the 0.05 level.

b. Calculate the p-value associated with the observed result.

c. Is your p-value based on a one-tail test or a two-tail test? Why?

d. As in the previous problem, graph these data and state your conclusion.

e. Carry out an (incorrect) two-sample test and state your conclusions.

f. Using the observed sample variances s_1^2 and s_2^2 associated with the set of readings before and after, calculate the variance of the difference *assuming* independence (call this Variance 1). How does this value compare with the variance of the difference calculated in the first part of the problem? (Call this Variance 2.) Why do you suppose Variance 1 is so much bigger than Variance 2? The *average* of the differences is the same as the difference in the averages. Show this. Hence, the two-sample t-test differed from the paired t-test only in the divisor. Which of the two tests in more powerful in this case, that is, declares a difference significant when in fact there is one?

5. The following data from Schecter *et al.* [1973] deals with sodium chloride preference as related to hypertension. Two groups, 12 normal and 10 hypertensive subjects, were isolated for a week and compared with respect to Na^+ intake. The following are the average daily Na^+ intakes:

Normal	10.2	2.2	0.0	2.6	0.0	43.1
Hypertensive	92.8	54.8	51.6	61.7	250.8	84.5

Normal	45.8	63.6	1.8	0.0	3.7	0.0
Hypertensive	34.7	62.2	11.0	39.1		

a. Compare the average daily Na^+ intake of the hypertensive subjects with that of the normal volunteers by means of an appropriate t-test.

b. State your assumptions.

c. Assuming that the population variances are not homogenous, carry out an appropriate t-test (see Note 5.2).

6. Kapitulnick *et al.* [1976] compared the metabolism of a drug, zoxazolamine, in placentas from 13 women who smoked during pregnancy and 11 who did not. The purpose of the study was to investigate the presence of the drug as a possible proxy for the rate at which benzo[a]pyrene (a byproduct of cigarette smoke) is metabolized. The following data were obtained in the measurement of zoxazolamine hydroxylase production (nmol $3H_2O$ formed/g/h):

Nonsmoker	0.18	0.36	0.24	0.50	0.42	0.36	0.50
Smoker	0.66	0.60	0.96	1.37	1.51	3.56	3.36

Nonsmoker	0.60	0.56	0.36	0.68		
Smoker	4.86	7.50	9.00	10.08	14.76	16.50

a. Calculate the sample mean and standard deviation for each of the two groups.

b. Test the assumption that the two sample variances came from a population with the same variance.

c. Carry out the t-test using the approximation to the t procedure discussed in Note 5.2. What are your conclusions?

d. Suppose we agree that the variability (as measured by the standard deviations) is proportional to the level of the response. Statistical theory then suggests that the logarithms of the responses should have roughly the same variability. Take logarithms of the data and test, once more, the homogeneity of the variances.

7. Sometime you may be asked to do a two-sample t-test knowing only the mean, standard deviation, and sample sizes. A paper by Holtzman *et al.* [1975] illustrates the problem. This paper dealt with terminating the phenylalanine-restricted diet in 4-year-old children with phenylketonuria (PKU). The purpose of the diet is to reduce the phenylalanine level. A high level is associated with mental retardation. After obtaining informed consent, eligible children of 4 years of age were randomly divided into two groups. Children in one group had their restricted diet terminated while children in the other group were continued on the restricted diet. At 6 years of age, the phenylalanine levels were tested in all children, and the following data reported:

	Diet Terminated	**Diet Continued**
Number of children	5	4
Mean phenylalanine level (mg/dl)	26.9	16.7
Standard deviation	4.1	7.3

a. State a reasonable null hypothesis and alternative hypothesis.

b. Calculate the pooled estimate of the variance s_p^2.

c. Test the null hypothesis of Part (a) of this problem. Is your test one-tail, or two? Why?

d. Test the hypothesis that the sample variances came from two populations with the same variance.

e. Construct a 95% confidence interval on the difference in the population phenylalanine levels.

f. Interpret the interval constructed in Part (e).

g. "This set of data has little power," someone says. What does it mean? Interpret the implications of a Type II error in this example.

h. What is the interpretation of a Type I error in this example? Which, in your opinion, is more serious in this example: a Type I error, or a Type II error?

i. On the basis of this data, what would you recommend to a parent with a 4-year-old PKU child?

j. Can you think of some additional information that would make the analysis more precise?

8. Several population studies have demonstrated an inverse correlation of Sudden Infant Death Syndrome (SIDS) rate with birthweight. The occurrence of SIDS in one of a pair of twins provides an opportunity to test the hypothesis that birthweight is a major determinant of SIDS. The following set of data collected by Dr. D. R. Peterson of the Department of Epidemiology, University of Washington, consists of the birthweights (in grams) of each of 22 dizygous twins, and each of 19 monozygous twins:

Dizygous Twins		Monozygous Twins	
SID	**Non-SID**	**SID**	**Non-SID**
1474	2098	1701	1956
3657	3119	2580	2438
3005	3515	2750	2807
2041	2126	1956	1843
2325	2211	1871	2041
2296	2750	2296	2183
3430	3402	2268	2495
3515	3232	2070	1673
1956	1701	1786	1843
2098	2410	3175	3572
3204	2892	2495	2778
2381	2608	1956	1588
2892	2693	2296	2183
2920	3232	3232	2778
3005	3005	1446	2268
2268	2325	1559	1304
3260	3686	2835	2892
3260	2778	2495	2353
2155	2552	1559	2466
2835	2693		
2466	1899		
3232	3714		

a. With respect to the dizygous twins, test the above hypothesis. State the null hypothesis.

b. Make a similar test on the monozygous twins.

c. Discuss your conclusions.

9. A pharmaceutical firm claims that a new analgesic drug relieves mild pain under standard conditions for 3 h with a standard deviation of 1 h. Sixteen patients are tested under the same conditions and have an average pain relief of 2.5 h. The hypothesis that the population mean of this sample is also 3 h is to be tested against the hypothesis that the population mean is in fact less than 3 h; $\alpha = 0.5$.

a. What is an appropriate test?

b. Set up the appropriate critical region.

c. State your conclusion.

d. Suppose the sample size is doubled. State precisely how the non-rejection region for the null hypothesis is changed.

10. Consider Problem 3.9, dealing with the treatment of essential hypertension. Compare treatments A and B by means of an appropriate t-test. Set up a 99% confidence interval on the reduction of blood pressure under treatment B as compared to treatment A.

63:161 - Chapter 5 Problems

1. b. The sample mean change (final-baseline) is -15.56 with sample s.d. 51.79 and s.e. 12.95.
   c. The paired t statistic is -1.20 with 15 df (two-sided p-value is $0.2<p<0.5$).
   d. The 90% confidence interval for the true mean change is $-15.56 \pm 1.75(12.95)$ or (-38.22,7.10).

2. a. H_0: the mean IQ difference between the PKU case and sibling is zero. The two-sided alternative hypothesis is of interest (since the problem does not state a specific alternative).
   b. The mean change (PKU IQ - sibling IQ) is -4.13 with a standard deviation of 15.90. The paired t statistic is -1.01 with 14 df. The two-sided p-value is $0.2<p<0.5$.
   c. At the 0.05 level of significance, we do not reject the null hypothesis.
   d. This test is valid if the IQ difference is normally-distributed, and if the 15 cases in the study are a random sample from the population of PKU-sibling pairs.
   f. The paired t statistic is -2.39 with 35 df; the two-tailed p-value is $0.02<p<0.05$.

3. a. Because the same six patients are studied preoperatively and postoperatively.
   b. The mean change (post-pre) is 24.3333 with a standard deviation of 42.2122. The paired t statistic is 1.41 with 5 df; the two-tailed p-value is $0.2<p<0.5$.
   c. The change in creatinine clearance is normally-distributed; the six patients are a random sample of patients anesthetized by halothane.
   d. The 99% confidence interval is $24.3333 \pm 4.03(17.23)$ or (-45.12, 93.78).

4. a. The mean change (after-before) is 0.118333 with a s.d. of 0.083287. The paired t statistic is 3.48 with 5 df. Since 3.48 exceeds 2.57 ($t_{5,0.975}$), we reject the null hypothesis.
   b. The two-tailed p-value is $0.01<p<0.02$.

5. a. The (mean, variance) in normal and hypertensive groups are (14.4167, 512.608) and (74.32, 4401.4818). Using the two-sample t test, the pooled estimate of variance is 2262.602 and the t statistic is -2.94. Since $0.005<p<0.01$ (two-sided), we reject H_0 at $\alpha=0.05$.
   b. We assume NA+ is normally distributed with the same variance in both populations. If we test for variance equality, $F=8.59>3.59$ ($F_{9,11,0.975}$) and we reject H_0 at $\alpha=0.05$. Since 8.59 exceeds $F_{9,11,0.999}$, $p<0.002$.

6. a. The (mean, s.d.) in nonsmokers and smokers are (0.432727, 0.152124) and (5.747692, 5.423356).
   b. The F statistic is 1270.97 with 12,10 df. Since this exceeds $F_{12,10,0.975}=3.52$, we reject the hypothesis of variance homogeneity at significance level $\alpha=0.05$. Since 1270.97 exceeds $F_{12,10,0.999}$, $p<0.002$.

7. a. Let μ_1 (μ_2) denote the population mean phenylalanine level in children whose diet is terminated (continued). We wish to test $H_0:\mu_1=\mu_2$. Since it appears that the focus is on determining if discontinuing the diet increases the mean phenylalanine level, the alternative of interest is $H_A:\mu_1>\mu_2$.
   b. The pooled estimate of variance is 32.444.
   c. The two-sample t statistic is 2.67. With reference to the t_7 distribution, $0.01<p<0.025$ (one-sided).
   d. $F=(7.3/4.1)^2=3.17$ has the $F_{3,4}$ distribution under $H_0:\sigma_1^2=\sigma_2^2$. Since $F_{3,4,0.975}=9.98$, we do not reject.
   e. A 95% confidence interval for the mean difference is $(26.9-16.7) \pm 2.37(3.82)$ or (1.14, 19.26).
   f. If we repeatedly select random samples of sizes 5 and 4 from the "terminated" and "continued" populations, then 95% of the confidence intervals we construct will straddle the true mean difference.
   h. Type I error occurs if we conclude diet termination has an adverse effect, when it does not; Type II error occurs if we conclude termination has no effect, when it does. A Type II error is more serious.
   i. Do not terminate the diet of your child.

8. a. In dizygous twins, the mean difference (SID - Non-Sid) is -43.9091 with a standard deviation of 365.1833. The paired t statistic is -0.564 with 21 df ($p>0.5$).
   b. In monozygous twins, the mean difference (SID - Non-Sid) is -59.7368 with a standard deviation of 370.1531. The paired t statistic is -0.7035 with 18 df ($0.2<p<0.5$).
   c. In both groups of twins, there is no evidence that birthweight is a major determinant of SIDS.