Solutions of Suggested problems of ch8

The mean ±1 *sd* of ln [calcium intake (mg)] among 25 females, 12 to 14 years of age, below the poverty level is 6.56 ± 0.64. Similarly, the mean ± 1 *sd* of ln [calcium intake (mg)] among 40 females, 12 to 14 years of age, above the poverty level is 6.80 ± 0.76.

8.2 Test for a significant difference between the variances of the two groups.

8.3 What is the appropriate procedure to test for a significant difference in means between the two groups?

8.4 Implement the procedure in Problem 8.3 using the critical-value method.

8.5 What is the *p*-value corresponding to your answer to Problem 8.4?

8.6 Compute a 95% CI for the difference in means between the two groups.

8.1 The symbol to denote this percentile is $F_{14,7,.025}$. It is obtained from: $F_{14,7,.025} = \frac{1}{F_{7,14,.975}} = \frac{1}{3.38} = 0.296$.

8.2 Test the hypothesis $H_0: \sigma_1^2 = \sigma_2^2$ versus $H_1: \sigma_1^2 \neq \sigma_2^2$. We have the test statistic $F = \frac{s_2^2}{s_1^2} = \frac{0.76^2}{0.64^2} = 1.410 \sim F_{39,24}$ under H_0. Since $F < F_{\infty,30,.975} = 1.79 < F_{39,24,.975}$, it follows that H_0 is accepted at the 5% level and we conclude that there is no significant difference between the variances.

8.3 Since H_0 was accepted in Problem 8.2, use the two-sample *t*-test for independent samples with equal variances.

8.4 First compute the pooled variance estimate:

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(25-1)(0.64)^2 + (40-1)(0.76)^2}{25+40-2} = \frac{32.357}{63} = 0.514$$

Compute the test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{6.56-6.80}{\sqrt{0.514\left(\frac{1}{25}+\frac{1}{40}\right)}} = \frac{-0.24}{0.183} = -1.314$$

The critical value is given by $t_{63,.975} > t_{120,.975} = 1.980 > |t|$. Therefore, accept H_0 at the 5% level.

8.5 The *p*-value is given by $2 \times \Pr(t_{63} < -1.314) = 2 \times P(t_{63} > 1.314)$. If there were 60 *df*, then since $t_{60,.9} = 1.296$, $t_{60,.95} = 1.671$, and $1.296 < 1.314 < 1.671$, it would follow that $2 \times (1-.95) < p < 2 \times (1-.9)$ or $.10 < p < .20$. If there were 120 *df*, then since $t_{120,.9} = 1.289$, $t_{120,.95} = 1.658$, and $1.289 < 1.314 < 1.658$, it would follow that $.10 < p < .20$. Since we reach the same conclusion with either 60 or 120 *df* and $60 < 63 < 120$, it follows that $.10 < p < .20$. The exact *p*-value obtained by computer is $2 \times \Pr(t_{63} < -1.314) = 0.194$.

```
> 2*pt(-1.314, 63)
[1] 0.1936108
```

8.6 The 95% confidence interval is given by $\bar{x}_1 - \bar{x}_2 \pm t_{63,.975}\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$. We use R to estimate $t_{63,.975}$ as follows:

```
> qt(.975, 63)
[1] 1.998341
```

Therefore, the 95% confidence interval is given by $-0.24 \pm 1.9984(0.183) = -0.24 \pm 0.365 = (-0.61,\ 0.13)$.

Refer to the data in Table 2.13.

8.14 Test for a significant difference in the variances of the initial white blood cell count between patients who did and patients who did not receive a bacterial culture.

8.15 What is the appropriate test procedure to test for significant differences in mean white blood cell count between people who do and people who do not receive a bacterial culture?

8.16 Perform the procedure in Problem 8.15 using the critical-value method.

8.17 What is the *p*-value corresponding to your answer to Problem 8.16?

8.18 Compute a 95% CI for the true difference in mean white blood cell count between the two groups.

TABLE 2.13 Hospital-stay data

ID no.	Duration of hospital stay	Age	Sex 1 = M 2 = F	First temp. following admission	First WBC ($\times 10^3$) following admission	Received antibiotic? 1 = yes 2 = no	Received bacterial culture? 1 = yes 2 = no	Service 1 = med. 2 = surg.
1	5	30	2	99.0	8	2	2	1
2	10	73	2	98.0	5	2	1	1
3	6	40	2	99.0	12	2	2	2
4	11	47	2	98.2	4	2	2	2
5	5	25	2	98.5	11	2	2	2
6	14	82	1	96.8	6	1	2	2
7	30	60	1	99.5	8	1	1	1
8	11	56	2	98.6	7	2	2	1
9	17	43	2	98.0	7	2	2	1
10	3	50	1	98.0	12	2	1	2
11	9	59	2	97.6	7	2	1	1
12	3	4	1	97.8	3	2	2	2
13	8	22	2	99.5	11	1	2	2
14	8	33	2	98.4	14	1	1	2
15	5	20	2	98.4	11	2	1	2
16	5	32	1	99.0	9	2	2	2
17	7	36	1	99.2	6	1	2	2
18	4	69	1	98.0	6	2	2	2
19	3	47	1	97.0	5	1	2	1
20	7	22	1	98.2	6	2	2	2
21	9	11	1	98.2	10	2	2	2
22	11	19	1	98.6	14	1	2	2
23	11	67	2	97.6	4	2	2	1
24	9	43	2	98.6	5	2	2	2
25	4	41	2	98.0	5	2	2	1

8.14 There are 6 persons who received a bacterial culture with mean WBC of 9.50 and standard deviation of 3.39. There are 19 persons who did not receive a bacterial culture with mean WBC of 7.32 and standard deviation of 3.06. We assume that the distribution of WBC is normal in each group. We test the hypothesis $H_0{:}\sigma_1^2 = \sigma_2^2$ versus $H_1{:}\sigma_1^2 \neq \sigma_2^2$. We use the test statistic $F = \frac{s_1^2}{s_2^2} = \frac{3.39^2}{3.06^2} = 1.23 \sim F_{5,18}$ under H_0. Since $F_{5,18,.975} = 3.38 > 1.23$, it follows that $p > .05$ and the variances are not significantly different.

8.15 The two-sample t-test for independent samples with equal variances.

8.16 We test the hypothesis $H_0{:}\mu_1 = \mu_2$ versus $H_1{:}\mu_1 \neq \mu_2$. First, we compute the pooled variance estimate

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

$$= \frac{(6-1)(3.39)^2 + (19-1)(3.06)^2}{6+19-2} = 9.809$$

Thus, we compute the test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$= \frac{9.50-7.32}{\sqrt{9.809\left(\frac{1}{6}+\frac{1}{19}\right)}}$$

$$= \frac{2.18}{1.467} = 1.49 \sim t_{23} \text{ under } H_0$$

Since $t_{23,.975} = 2.069 > 1.49$, we accept H_0 at the 5% level.

8.17 Since $t_{23,.9} = 1.319 < 1.49 < t_{23,.95} = 1.714$, it follows that $2\times(1-.95) < p < 2\times(1-.9)$ or $.10 < p < .20$. The exact p-value, obtained by computer, is $2\times \Pr(t_{23} > 1.49) = 0.15$.

8.18 $2.18 \pm t_{23,.975}(1.467) = 2.18 \pm 2.069(1.467) = 2.18 \pm 3.03 = (-0.85, 5.22)$.

Ten patients with advanced diabetic nephropathy (kidney complications of diabetes) were treated with captopril over an 8-week period [9]. Urinary protein was measured before and after drug therapy, with results listed in Table 8.16 in both the raw and ln scale.

TABLE 8.16 Changes in urinary protein after treatment with captopril

	Raw scale urinary protein (g/24 hr)		ln scale urinary protein ln (g/24 hr)	
Patient	Before	After	Before	After
1	25.6	10.1	3.24	2.31
2	17.0	5.7	2.83	1.74
3	16.0	5.6	2.77	1.72
4	10.4	3.4	2.34	1.22
5	8.2	6.5	2.10	1.87
6	7.9	0.7	2.07	–0.36
7	5.8	6.1	1.76	1.81
8	5.4	4.7	1.69	1.55
9	5.1	2.0	1.63	0.69
10	4.7	2.9	1.55	1.06

***8.44** What is the appropriate statistical procedure to test whether mean urinary protein has changed over the 8-week period?

***8.45** Perform the test in Problem 8.44 using both the raw and ln scale, and report a *p*-value. Are there any advantages to using the raw or the ln scale?

8.44 Since each person is being used as his or her own control, the paired t test is the appropriate test procedure, where we test the hypothesis $H_0{:}\ \mu_d = 0$ versus $H_1{:}\ \mu_d \neq 0$ and μ_d = true mean change in urinary protein over the 8-week period is tested.

8.45 Calculate the difference scores in the raw scale as follows:

Difference scores in the raw scale			
i	d_i	i	d_i
1	−15.5	6	−7.2
2	−11.3	7	+0.3
3	−10.4	8	−0.7
4	− 7.0	9	−3.1
5	− 1.7	10	−1.8

Then compute the test statistic:

$$t = \frac{d}{s_d / \sqrt{n}}$$
$$= \frac{-5.84}{5.294 / \sqrt{10}} = \frac{-5.84}{1.674} = -3.49 \sim t_9 \text{ under } H_0.$$

Since $t_{9,.995} = 3.250$, $t_{9,.9995} = 4.781$, and $3.250 < 3.49 < 4.781$, it follows that $2 \times (1-.9995) < p < 2 \times (1-.995)$, or $.001 < p < .01$ (exact p-value = .007 by computer). Calculate the difference scores in the ln scale as follows:

Difference scores in the ln scale			
Person(i)	d_i	Person(i)	d_i
1	−0.93	6	−2.43
2	−1.09	7	+0.05
3	−1.05	8	−0.14
4	−1.12	9	−0.94
5	−0.23	10	−0.49

Then compute the test statistic:

$$t = \frac{\bar{d}}{s_d / \sqrt{n}}$$
$$= \frac{-0.837}{0.708 / \sqrt{10}} = \frac{-0.837}{0.224} = -3.74 \sim t_9 \text{ under } H_0.$$

Since $t_{9,.995} = 3.250$, $t_{9,.9995} = 4.781$, and $3.250 < 3.74 < 4.781$, it follows that $2 \times (1-.9995) < p < 2 \times (1-.995)$, or $.001 < p < .01$ (exact p-value = .005 by computer). Thus, there is a significant decline in mean ln(urinary protein) over the 8-week period. Since the difference scores appear to be related to the initial urinary protein, the ln scale is preferable here.