28

Comparing two samples

Many applications are concerned with *two* groups of observations of the same kind that originate from two possibly different model distributions, and the question is whether these distributions have different expectations. We describe a test for equality of expectations, where we consider normal and nonnormal model distributions and equal and unequal variances of the model distributions.

28.1 Is dry drilling faster than wet drilling?

Recall the drilling example from Sections 15.5 and 16.4. The question was whether dry drilling is faster than wet drilling. The scatterplots in Figure 15.11 seem to suggest that up to a depth of 250 feet the drill time does not depend on depth. Therefore, for a first investigation of a possible difference between dry and wet drilling we only consider the (mean) drill times up to this depth. A more thorough study can be found in [23].

The boxplots of the drill times for both types of drilling are displayed in Figure 28.1. Clearly, the boxplot for dry drilling is positioned lower than the

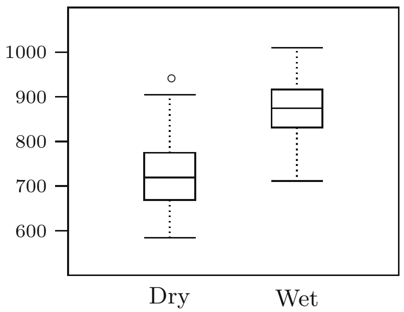

Fig. 28.1. Boxplot of drill times.

one for wet drilling. However, the question is whether this difference can be attributed to chance or if it is large enough to conclude that the dry drill time is shorter than the wet drill time. To answer this question, we model the datasets of dry and wet drill times as realizations of random samples from two distribution functions F and G, one with expected value μ_1 and the other with expected value μ_2. The parameters μ_1 and μ_2 represent the drill times of dry drilling and wet drilling, respectively. We test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 < \mu_2$.

This example illustrates a general situation where we compare two datasets

$$x_1, x_2, \ldots, x_n \quad \text{and} \quad y_1, y_2, \ldots, y_m,$$

which are the realization of independent random samples

$$X_1, X_2, \ldots, X_n \quad \text{and} \quad Y_1, Y_2, \ldots, Y_m$$

from two distributions, and we want to test whether the expectations of both distributions are the same. Both the variance σ_X^2 of the X_i and the variance σ_Y^2 of the Y_j are *unknown*.

Note that the null hypothesis is equivalent to the statement $\mu_1 - \mu_2 = 0$. For this reason, similar to Chapter 27, the test statistic for the null hypothesis $H_0 : \mu_1 = \mu_2$ is based on an estimator $\bar{X}_n - \bar{Y}_m$ for the difference $\mu_1 - \mu_2$. As before, we standardize $\bar{X}_n - \bar{Y}_m$ by an estimator for its variance

$$\operatorname{Var}\left(\bar{X}_n - \bar{Y}_m\right) = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}.$$

Recall that the sample variances S_X^2 and S_Y^2 of the X_i and Y_j, are unbiased estimators for σ_X^2 and σ_Y^2. We will use a combination of S_X^2 and S_Y^2 to construct an estimator for $\operatorname{Var}\left(\bar{X}_n - \bar{Y}_m\right)$. The actual standardization of $\bar{X}_n - \bar{Y}_m$ depends on whether the variances of the X_i and Y_j are the same. We distinguish between the two cases $\sigma_X^2 = \sigma_Y^2$ and $\sigma_X^2 \neq \sigma_Y^2$. In the next section we consider the case of equal variances.

QUICK EXERCISE 28.1 Looking at the boxplots in Figure 28.1, does the assumption $\sigma_X^2 = \sigma_Y^2$ seem reasonable to you? Can you think of a way to quantify your belief?

28.2 Two samples with equal variances

Suppose that the samples originate from distributions with the *same* (but unknown) variance:

$$\sigma_X^2 = \sigma_Y^2 = \sigma^2.$$

In this case we can *pool* the sample variances S_X^2 and S_Y^2 by constructing a linear combination $aS_X^2 + bS_Y^2$ that is an unbiased estimator for σ^2. One particular choice is the weighted average

$$\frac{(n-1)S_X^2+(m-1)S_Y^2}{n+m-2}.$$

It has the property that for normally distributed samples it has the smallest variance among all unbiased linear combinations of S_X^2 and S_Y^2 (see Exercise 28.5). Moreover, the weights depend on the sample sizes. This is appropriate, since if one sample is much larger than the other, the estimate of σ^2 from that sample is more reliable and should receive greater weight.

We find that the *pooled-variance*:

$$S_p^2 = \frac{(n-1)S_X^2+(m-1)S_Y^2}{n+m-2}\left(\frac{1}{n}+\frac{1}{m}\right)$$

is an unbiased estimator for

$$\mathrm{Var}\big(\bar{X}_n-\bar{Y}_m\big)=\sigma^2\left(\frac{1}{n}+\frac{1}{m}\right).$$

This leads to the following test statistic for the null hypothesis $H_0:\mu_1=\mu_2$:

$$T_p=\frac{\bar{X}_n-\bar{Y}_m}{S_p}.$$

As before, we compare the estimator $\bar{X}_n-\bar{Y}_m$ with 0 (the value of $\mu_1-\mu_2$ under the null hypothesis), and we standardize by dividing by the estimator S_p for the standard deviation of $\bar{X}_n-\bar{Y}_m$. Values of T_p close to zero are in favor of the null hypothesis $H_0:\mu_1=\mu_2$. Large positive values of T_p suggest that $\mu_1>\mu_2$, whereas large negative values suggest that $\mu_1<\mu_2$.

The next step is to determine the distribution of T_p. Note that under the null hypothesis $H_0:\mu_1=\mu_2$, the test statistic T_p is the *pooled studentized mean difference*

$$\frac{(\bar{X}_n-\bar{Y}_m)-(\mu_1-\mu_2)}{S_p}.$$

Hence, *under the null hypothesis*, the probability distribution of T_p is the *same* as that of the pooled studentized mean difference. To determine its distribution, we distinguish between normal and nonnormal data.

Normal samples

In the same way as the studentized mean of a single normal sample has a $t(n-1)$ distribution (see page 349), it is also a fact that if two independent samples originate from normal distributions, i.e.,

$$\begin{aligned} &X_1,X_2,\ldots,X_n \text{ random sample from } N(\mu_1,\sigma^2)\\ &Y_1,Y_2,\ldots,Y_m \text{ random sample from } N(\mu_2,\sigma^2), \end{aligned}$$

then the pooled studentized mean difference has a $t(n+m-2)$ distribution. Hence, under the null hypothesis, the test statistic T_p has a $t(n+m-2)$

distribution. For this reason, a test for the null hypothesis $H_0 : \mu_1 = \mu_2$ is called a *two-sample t-test*.

Suppose that in our drilling example we model our datasets as realizations of random samples of sizes $n = m = 50$ from two normal distributions with equal variances, and we test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 < \mu_2$ at level 0.05. For the data we find $\bar{x}_{50} = 727.78$, $\bar{y}_{50} = 873.02$, and $s_p = 13.62$, so that

$$t_p = \frac{727.78 - 873.02}{13.62} = -10.66.$$

We compare this with the left critical value $-t_{98,0.05}$. This value is not in Table B.2, but $-1.676 = -t_{50,0.05} < -t_{98,0.05}$. This means that $t_p < -t_{98,0.05}$, so that we reject $H_0 : \mu_1 = \mu_2$ in favor of $H_1 : \mu_1 < \mu_2$ at level 0.05. The p-value corresponding to $t_p = -10.66$ is the left tail probability $\mathrm{P}(T \leq -10.66)$. From Table B.2 we can only see that this is smaller than 0.0005 (a statistical software package gives $\mathrm{P}(T \leq -10.66) = 2.25 \cdot 10^{-18}$). The data provide overwhelming evidence against the null hypothesis, so that we conclude that dry drilling is faster than wet drilling.

QUICK EXERCISE 28.2 Suppose that in the ball bearing example of Quick exercise 27.2, we test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$, where μ_1 and μ_2 represent the diameters of a ball bearing from the first and second production line. What are the critical values corresponding to level $\alpha = 0.01$?

Nonnormal samples

Similar to the one-sample t-test, if we *cannot* assume normal model distributions, then we can *no longer* conclude that our test statistic has a $t(n + m - 2)$ distribution under the null hypothesis. Recall that under the null hypothesis, the distribution of our test statistic is the same as that of the pooled studentized mean difference (see page 417).

To approximate its distribution, we use the empirical bootstrap simulation for the pooled studentized mean difference

$$\frac{(\bar{X}_n - \bar{Y}_m) - (\mu_1 - \mu_2)}{S_p}.$$

Given datasets $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_m$, determine their empirical distribution functions F_n and G_m as estimates for F and G. The expectations corresponding to F_n and G_m are $\mu_1^* = \bar{x}_n$ and $\mu_2^* = \bar{y}_m$. Then repeat the following two steps many times:

1. Generate a bootstrap dataset $x_1^*, x_2^*, \ldots, x_n^*$ from F_n and a bootstrap dataset $y_1^*, y_2^*, \ldots, y_m^*$ from G_m.
2. Compute the pooled studentized mean difference for the bootstrap data:

$$t_p^* = \frac{(\bar{x}_n^* - \bar{y}_m^*) - (\bar{x}_n - \bar{y}_m)}{s_p^*},$$

where $\bar{x}_n^*$ and $\bar{y}_m^*$ are the sample means of the bootstrap datasets, and

$$(s_p^*)^2 = \frac{(n-1)(s_X^*)^2 + (m-1)(s_Y^*)^2}{n+m-2}\left(\frac{1}{n}+\frac{1}{m}\right)$$

with $(s_X^*)^2$ and $(s_Y^*)^2$ the sample variances of the bootstrap datasets.

The reason that in each iteration we subtract $\bar{x}_n - \bar{y}_m$ is that $\mu_1 - \mu_2$ is the difference of the expectations of the two model distributions. Therefore, according to the bootstrap principle we should replace this by the difference $\bar{x}_n - \bar{y}_m$ of the expectations corresponding to the two empirical distribution functions.

We carried out this bootstrap simulation for the drill times. The result of this simulation can be seen in Figure 28.2, where a histogram and the empirical distribution function are displayed for one thousand bootstrap values of t_p^*. Suppose that we test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 < \mu_2$ at level 0.05. The bootstrap approximation for the left critical value is $c_l^* = -1.659$. The value of $t_p = -10.66$, computed from the data, is much smaller. Hence, also on the basis of the bootstrap simulation we reject the null hypothesis and conclude that the dry drill time is shorter than the wet drill time.

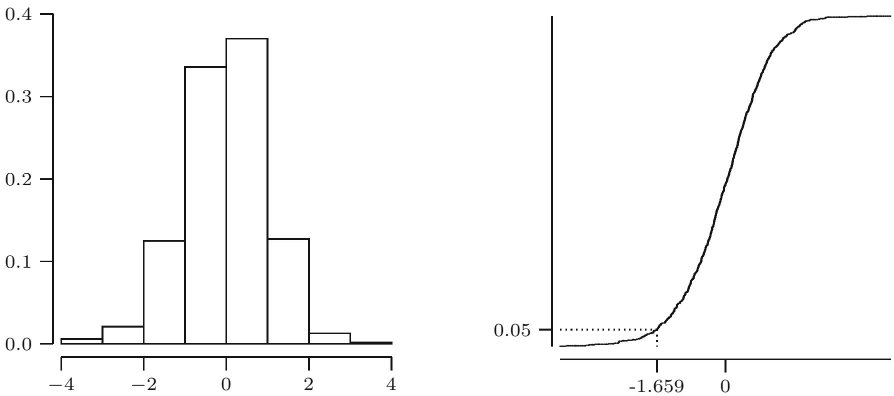

Fig. 28.2. Histogram and empirical distribution function of 1000 bootstrap values for T_p^*.

28.3 Two samples with unequal variances

During an investigation about weather modification, a series of experiments was conducted in southern Florida from 1968 to 1972. These experiments were designed to investigate the use of massive silver-iodide seeding. It was

Table 28.1. Rainfall data.

Unseeded					
1202.6	830.1	372.4	345.5	321.2	244.3
163.0	147.8	95.0	87.0	81.2	68.5
47.3	41.1	36.6	29.0	28.6	26.3
26.1	24.4	21.7	17.3	11.5	4.9
4.9	1.0				
Seeded					
2745.6	1697.8	1656.0	978.0	703.4	489.1
430.0	334.1	302.8	274.7	274.7	255.0
242.5	200.7	198.6	129.6	119.0	118.3
115.3	92.4	40.6	32.7	31.4	17.5
7.7	4.1				

Source: J. Simpson, A. Olsen, and J.C. Eden. A Bayesian analysis of a multiplicative treatment effect in weather modification. *Technometrics*, 17:161–166, 1975; Table 1 on page 162.

hypothesized that under specified conditions, this leads to invigorated cumulus growth and prolonged lifetimes, thereby causing increased precipitation. In these experiments, 52 isolated cumulus clouds were observed, of which 26 were selected at random and injected with silver-iodide smoke. Rainfall amounts (in acre-feet) were recorded for all clouds. They are listed in Table 28.1. To investigate whether seeding leads to increased rainfall, we test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 < \mu_2$, where μ_1 and μ_2 represent the rainfall for unseeded and seeded clouds.

In Figure 28.3 the boxplots of both datasets are displayed. From this we see that the assumption of equal variances may not be realistic. Indeed, this is confirmed by the values $s_X^2 = 77\,521$ and $s_Y^2 = 423\,524$ of the sample variances of the datasets. This means that we need to test $H_0 : \mu_1 = \mu_2$ without the assumption of equal variances. As before, the test statistic will be a standardized version of $\bar{X}_n - \bar{Y}_m$, but S_p^2 is no longer an unbiased estimator for

$$\mathrm{Var}\big(\bar{X}_n - \bar{Y}_m\big) = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}.$$

However, if we estimate σ_X^2 and σ_Y^2 by S_X^2 and S_Y^2, then the *nonpooled variance*

$$S_d^2 = \frac{S_X^2}{n} + \frac{S_Y^2}{m}$$

is an unbiased estimator for $\mathrm{Var}\big(\bar{X}_n - \bar{Y}_m\big)$. This leads to test statistic

$$T_d = \frac{\bar{X}_n - \bar{Y}_m}{S_d}.$$

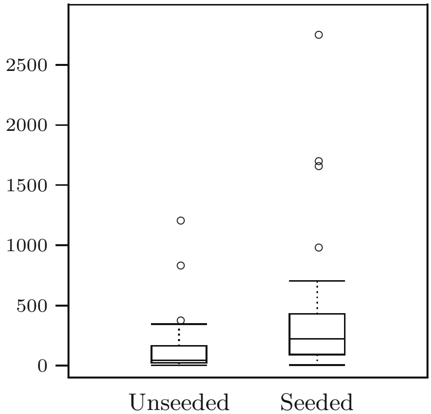

Fig. 28.3. Boxplots of rainfall.

Again, we compare the estimator $\bar{X}_n - \bar{Y}_m$ with zero and standardize by dividing by an estimator for the standard deviation of $\bar{X}_n - \bar{Y}_m$. Values of T_d close to zero are in favor of the null hypothesis $H_0 : \mu_1 = \mu_2$.

Quick exercise 28.3 Consider the ball bearing example from Quick exercise 27.2. Compute the value of T_d for this example.

Under the null hypothesis $H_0 : \mu_1 = \mu_2$, the test statistic

$$T_d = \frac{\bar{X}_n - \bar{Y}_m}{S_d}$$

is equal to the *nonpooled studentized mean difference*

$$\frac{(\bar{X}_n - \bar{Y}_m) - (\mu_1 - \mu_2)}{S_d}.$$

Therefore, the distribution of T_d under the null hypothesis is the same as that of the nonpooled studentized mean difference. Unfortunately, its distribution is not a t-distribution, not even in the case of normal samples. This means that we have to approximate this distribution.

Similar to the previous section, we use the empirical bootstrap simulation for the nonpooled studentized mean difference. The only difference with the procedure outlined in the previous section is that now in each iteration we compute the nonpooled studentized mean difference for the bootstrap datasets:

$$t_d^* = \frac{(\bar{x}_n^* - \bar{y}_m^*) - (\bar{x}_n - \bar{y}_m)}{s_d^*},$$

where $\bar{x}_n^*$ and $\bar{y}_m^*$ are the sample means of the bootstrap datasets, and

$$(s_d^*)^2 = \frac{(s_X^*)^2}{n} + \frac{(s_Y^*)^2}{m}$$

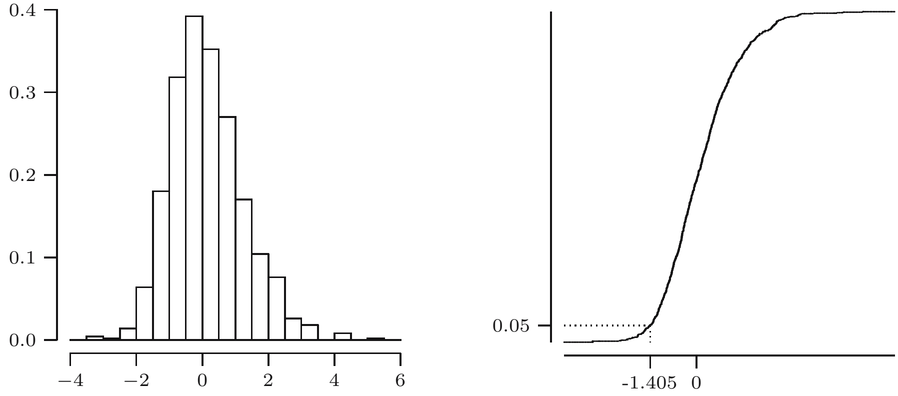

Fig. 28.4. Histogram and empirical distribution function of 1000 bootstrap values of T_d^*.

with $(s_X^*)^2$ and $(s_Y^*)^2$ the sample variances of the bootstrap datasets.

We carried out this bootstrap simulation for the cloud seeding data. The result of this simulation can be seen in Figure 28.4, where a histogram and the empirical distribution function are displayed for one thousand values t_d^*. The bootstrap approximation for the left critical value corresponding to level 0.05 is $c_l^* = -1.405$. For the data we find the value

$$t_d = \frac{164.59 - 441.98}{138.92} = -1.998.$$

This is smaller than c_l^*, so we reject the null hypothesis. Although the evidence against the null hypothesis is not overwhelming, there is some indication that seeding clouds leads to more rainfall.

28.4 Large samples

Variants of the central limit theorem state that as n and m both tend to infinity, the distributions of the pooled studentized mean difference

$$\frac{(\bar{X}_n - \bar{Y}_m) - (\mu_1 - \mu_2)}{S_p}$$

and the nonpooled studentized mean difference

$$\frac{(\bar{X}_n - \bar{Y}_m) - (\mu_1 - \mu_2)}{S_d}$$

both approach the standard normal distribution. This fact can be used to approximate the distribution of the test statistics T_p and T_d under the null hypothesis by a standard normal distribution.

We illustrate this by means of the following example. To investigate whether a restricted diet promotes longevity, two groups of randomly selected rats were put on the different diets. One group of $n = 106$ rats was put on a restricted diet, the other group of $m = 89$ rats on an ad libitum diet (i.e., unrestricted eating). The data in Table 28.2 represent the remaining lifetime in days of two groups of rats after they were put on the different diets. The average lifetimes are $\bar{x}_n = 968.75$ and $\bar{y}_m = 684.01$ days. To investigate whether a restricted diet promotes longevity, we test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 > \mu_2$, where μ_1 and μ_2 represent the lifetime of a rat on a restricted diet and on an ad libitum diet, respectively.

If we may assume equal variances, we compute

$$t_p = \frac{968.75 - 684.01}{32.88} = 8.66.$$

This value is larger than the right critical value $z_{0.0005} = 3.291$, which means that we would reject $H_0 : \mu_1 = \mu_2$ in favor of $H_1 : \mu_1 > \mu_2$ at level $\alpha = 0.0005$.

Table 28.2. Rat data.

Restricted									
105	193	211	236	302	363	389	390	391	403
530	604	605	630	716	718	727	731	749	769
770	789	804	810	811	833	868	871	875	893
897	901	906	907	919	923	931	940	957	958
961	962	974	979	982	1001	1008	1010	1011	1012
1014	1017	1032	1039	1045	1046	1047	1057	1063	1070
1073	1076	1085	1090	1094	1099	1107	1119	1120	1128
1129	1131	1133	1136	1138	1144	1149	1160	1166	1170
1173	1181	1183	1188	1190	1203	1206	1209	1218	1220
1221	1228	1230	1231	1233	1239	1244	1258	1268	1294
1316	1327	1328	1369	1393	1435				

Ad libitum									
89	104	387	465	479	494	496	514	532	536
545	547	548	582	606	609	619	620	621	630
635	639	648	652	653	654	660	665	667	668
670	675	677	678	678	681	684	688	694	695
697	698	702	704	710	711	712	715	716	717
720	721	730	731	732	733	735	736	738	739
741	743	746	749	751	753	764	765	768	770
773	777	779	780	788	791	794	796	799	801
806	807	815	836	838	850	859	894	963	

Source: B.L. Berger, D.D. Boos, and F.M. Guess. Tests and confidence sets for comparing two mean residual life functions. *Biometrics*, 44:103–115, 1988.

The p-value is the right tail probability $P(T_p \geq 8.66)$, which we approximate by $P(Z \geq 8.66)$, where Z has an $N(0,1)$ distribution. From Table B.1 we see that this probability is smaller than $P(Z \geq 3.49) = 0.0002$. By means of a statistical package we find $P(Z \geq 8.66) = 2.4 \cdot 10^{-16}$.

If we repeat the test without the assumption of equal variances, we compute

$$t_d = \frac{968.75 - 684.01}{31.08} = 9.16,$$

which also leads to rejection of the null hypothesis. In this case, the p-value $P(T_d \geq 9.16) \approx P(Z \geq 9.16)$ is even smaller since $9.16 > 8.66$ (a statistical package gives $P(Z \geq 9.16) = 2.6 \cdot 10^{-18}$). The data provide overwhelming evidence against the null hypothesis, and we conclude that a restricted diet promotes longevity.

28.5 Solutions to the quick exercises

28.1 Just by looking at the boxplots, the authors believe that the assumption $\sigma_X^2 = \sigma_Y^2$ is reasonable. The lengths of the boxplots and their IQRs are almost the same. However, the boxplots do not reveal how the elements of the dataset vary around the center. One way of quantifying our belief would be to compare the sample variances of the datasets. One possibility is to compare the ratio of both sample variances; a ratio close to one would support our belief of equal variances (in case of normal samples, this is a standard test called the F-test).

28.2 In this case we have a right and left critical value. From Quick exercise 27.2 we know that $n = m = 10$, so that the right critical value is $t_{18,0.005} = 2.878$ and the left critical value is $-t_{18,0.005} = -2.878$.

28.3 We first compute $s_d^2 = (0.0290)^2/10 + (0.0428)^2/10 = 0.000267$ and then $t_d = (1.0194 - 1.0406)/\sqrt{0.000267} = -1.297$.

28.6 Exercises

28.1 ⊡ The data in Table 28.3 represent salaries (in pounds Sterling) in 72 randomly selected advertisements in the *The Guardian* (April 6, 1992). When a range was given in the advertisement, the midpoint of the range is reproduced in the table. The data are salaries corresponding to two kinds of occupations ($n = m = 72$): (1) creative, media, and marketing and (2) education. The sample mean and sample variance of the two datasets are, respectively:

$$\begin{array}{ll} (1) & \bar{x}_{72} = 17\,410 \text{ and } s_x^2 = 41\,258\,741, \\ (2) & \bar{y}_{72} = 19\,818 \text{ and } s_y^2 = 50\,744\,521. \end{array}$$

Table 28.3. Salaries in two kinds of occupations.

Occupation (1)			Occupation (2)		
17703	13796	12000	25899	17378	19236
42000	22958	22900	21676	15594	18780
18780	10750	13440	15053	17375	12459
15723	13552	17574	19461	20111	22700
13179	21000	22149	22485	16799	35750
37500	18245	17547	17378	12587	20539
22955	19358	9500	15053	24102	13115
13000	22000	25000	10998	12755	13605
13500	12000	15723	18360	35000	20539
13000	16820	12300	22533	20500	16629
11000	17709	10750	23008	13000	27500
12500	23065	11000	24260	18066	17378
13000	18693	19000	25899	35403	15053
10500	14472	13500	18021	17378	20594
12285	12000	32000	17970	14855	9866
13000	20000	17783	21074	21074	21074
16000	18900	16600	15053	19401	25598
15000	14481	18000	20739	15053	15053
13944	35000	11406	15053	15083	31530
23960	18000	23000	30800	10294	16799
11389	30000	15379	37000	11389	15053
12587	12548	21458	48000	11389	14359
17000	17048	21262	16000	26544	15344
9000	13349	20000	20147	14274	31000

Source: D.J. Hand, F. Daly, A.D. Lunn, K.J. McConway, and E. Ostrowski. *Small data sets.* Chapman and Hall, London, 1994; dataset 385. Data collected by D.J. Hand.

Suppose that the datasets are modeled as realizations of normal distributions with expectations μ_1 and μ_2, which represent the salaries for occupations (1) and (2).

a. Test the null hypothesis that the salary for both occupations is the same at level $\alpha = 0.05$ under the assumption of equal variances. Formulate the proper null and alternative hypotheses, compute the value of the test statistic, and report your conclusion.

b. Do the same without the assumption of equal variances.

c. As a comparison, one carries out an empirical bootstrap simulation for the nonpooled studentized mean difference. The bootstrap approximations for the critical values are $c_l^* = -2.004$ and $c_u^* = 2.133$. Report your conclusion about the salaries on the basis of the bootstrap results.

28.2 The data in Table 28.4 represent the duration of pregnancy for 1669 women who gave birth in a maternity hospital in Newcastle-upon-Tyne, England, in 1954.

Table 28.4. Durations of pregnancy.

Duration	Medical	Emergency	Social
11		1	
15		1	
17	1		
20		1	
22	1	2	
24	1	3	
25		2	1
26		1	
27	2	2	1
28	1	2	1
29	3	1	
30	3	5	1
31	4	5	2
32	10	9	2
33	6	6	2
34	12	7	10
35	23	11	4
36	26	13	19
37	54	16	30
38	68	35	72
39	159	38	115
40	197	32	155
41	111	27	128
42	55	25	64
43	29	8	16
44	4	5	3
45	3	1	6
46	1	1	1
47	1		
56		1	

Source: D.J. Newell. Statistical aspects of the demand for maternity beds. *Journal of the Royal Statistical Society, Series A*, 127:1–33, 1964.

The durations are measured in complete weeks from the beginning of the last menstrual period until delivery. The pregnancies are divided into those where an admission was booked for medical reasons, those booked for social reasons (such as poor housing), and unbooked emergency admissions. For the three groups the sample means and sample variances are

Medical:	775 observations with $\bar{x} = 39.08$ and $s^2 = 7.77$,
Emergency:	261 observations with $\bar{x} = 37.59$ and $s^2 = 25.33$,
Social:	633 observations with $\bar{x} = 39.60$ and $s^2 = 4.95$.

Suppose we view the datasets as realizations of random samples from normal distributions with expectations μ_1, μ_2, and μ_3 and variances σ_1^2, σ_2^2, and σ_3^2, where μ_i represents the duration of pregnancy for the women from the ith group. We want to investigate whether the duration differs for the different groups. For each combination of two groups test the null hypothesis of equality of μ_i. Compute the values of the test statistic and report your conclusions.

28.3 ⊡ In a seven-day study on the effect of ozone, a group of 23 rats was kept in an ozone-free environment and a group of 22 rats in an ozone-rich environment. From each member in both groups the increase in weight (in grams) was recorded. The results are given in Table 28.5. The interest is in whether ozone affects the increase of weight. We investigate this by testing $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$, where μ_1 and μ_2 denote the increases of weight for a rat in the ozone-free and ozone-rich groups. The sample means are

$$\text{Ozone-free:} \quad \bar{x}_{23} = 22.40$$
$$\text{Ozone-rich:} \quad \bar{y}_{22} = 11.01.$$

The pooled standard deviation is $s_p = 4.58$, and the nonpooled standard deviation is $s_d = 4.64$.

Table 28.5. Weight increase of rats.

Ozone-free			Ozone-rich		
41.0	38.4	24.4	10.1	6.1	20.4
25.9	21.9	18.3	7.3	14.3	15.5
13.1	27.3	28.5	−9.9	6.8	28.2
−16.9	17.4	21.8	17.9	−12.9	14.0
15.4	27.4	19.2	6.6	12.1	15.7
22.4	17.7	26.0	39.9	−15.9	54.6
29.4	21.4	22.7	−14.7	44.1	−9.0
26.0	26.6		−9.0		

Source: K.A. Doksum and G.L. Sievers. Plotting with confidence: graphical comparisons of two populations. *Biometrika*, 63(3):421–434, 1976; Table 10 on page 433. By permission of the Biometrika Trustees.

a. Perform the test at level 0.05 under the assumption of normal data with equal variances, i.e., compute the test statistic and report your conclusion.

b. One also carries out a bootstrap simulation for the test statistic used in **a**, and finds critical values $c_l^* = -1.912$ and $c_u^* = 1.959$. What is your conclusion on the basis of the bootstrap simulation?

c. Also perform the test at level 0.05 without the assumption of equal variances, where you may use the normal approximation for the distribution of the test statistic under the null hypothesis.

d. A bootstrap simulation for the test statistic in **c** yields that the right tail probability corresponding to the observed value of the test statistic in this case is 0.014. What is your conclusion on the basis of the bootstrap simulation?

28.4 Show that in the case when $n = m$, the random variables T_p and T_d are the same.

28.5 ⊞ Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be independent random samples from normal distributions with variances σ^2. It can be shown that

$$\operatorname{Var}\left(S_X^2\right) = \frac{2\sigma^4}{n-1} \quad \text{and} \quad \operatorname{Var}\left(S_Y^2\right) = \frac{2\sigma^4}{m-1}.$$

Consider linear combinations $aS_X^2 + bS_Y^2$ that are unbiased estimators for σ^2.

a. Show that a and b must satisfy $a + b = 1$.

b. Show that $\operatorname{Var}\left(aS_X^2 + (1-a)S_Y^2\right)$ is minimized for $a = (n-1)/(n+m-2)$ (and hence $b = (m-1)/(n+m-2)$).

28.6 Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be independent random samples from distributions with (possibly unequal) variances σ_X^2 and σ_Y^2.

a. Show that

$$\operatorname{Var}\left(\bar{X}_n - \bar{Y}_m\right) = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}.$$

b. Show that the pooled variance S_p^2, as defined on page 417, is a biased estimator for $\operatorname{Var}\left(\bar{X}_n - \bar{Y}_m\right)$.

c. Show that the nonpooled variance S_d^2, as defined on page 420, is the only unbiased estimator for $\operatorname{Var}\left(\bar{X}_n - \bar{Y}_m\right)$ of the form $aS_X^2 + bS_Y^2$.

d. Suppose that $\sigma_X^2 = \sigma_Y^2 = \sigma^2$. Show that S_d^2, as defined on page 417, is an unbiased estimator for $\operatorname{Var}\left(\bar{X}_n - \bar{Y}_m\right) = \sigma^2(1/n + 1/m)$.

e. Is S_d^2 also an unbiased estimator for $\operatorname{Var}\left(\bar{X}_n - \bar{Y}_m\right)$ in the case $\sigma_X^2 \neq \sigma_Y^2$? What about when $n = m$?