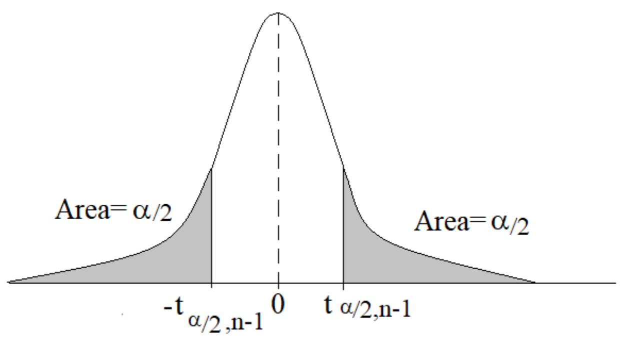
8.3.1 Confidence Interval for a Normal Mean when the Variance Is Unknown

Suppose now that $X_1, X_2, \ldots, X_n$ is a sample from a *normal distribution* with unknown mean μ and unknown variance σ^2, and that we wish to construct a $100(1-\alpha)$ percent confidence interval for μ. Since σ is unknown, we can no longer base our interval on the fact that $\sqrt{n}(\bar{X}-\mu)/\sigma$ is a standard normal random variable. However, by letting $S^2 = \sum(X_i - \bar{X})^2/(n-1)$ denote the sample variance, then from Corollary (section 7.2.4b), it follows that

$$\sqrt{n}\frac{(\bar{X}-\mu)}{S} \tag{8.25}$$

is a *t*-random variable with $n-1$ degrees of freedom. Hence, from the symmetry of the *t*-density function, we have that for any $\alpha \in (0, 1/2)$,



$$\Pr\left(-t_{\alpha/2,n-1} < \sqrt{n}\frac{(\bar{X}-\mu)}{S} < t_{\alpha/2,n-1}\right) = 1-\alpha \tag{8.26}$$

$$\Pr\left(-t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < \bar{X}-\mu < t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right) = 1-\alpha$$

$$\Pr\left(\bar{X}-t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < \mu < \bar{X}+t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right) = 1-\alpha \tag{8.27}$$

If it is observed that $\bar{X} = \bar{x}$ and $S = s$, then “with $100(1-\alpha)$ percent confidence”

$$\mu \in \left(\bar{x}-t_{\alpha/2,n-1}\frac{s}{\sqrt{n}},\ \bar{x}+t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right) \tag{8.28}$$

Exercise 8 Let consider Exercise 4 but let us now suppose that when the value μ is transmitted at location *A* then the value received at location *B* is normal with mean μ and variance σ^2 but with σ^2 being unknown. If 9 successive values are, as in Exercise 4: 5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5, compute a 95 percent confidence interval for μ.

A simple calculation yields that $\bar{x} = 9$ and

$$s^2 = \frac{\sum x_i^2 - 9(\bar{x})^2}{8} = 9.5$$

$$s = 3.082$$

Hence, as $t_{0.025,8} = 2.306$,

```
qt(0.025,df=8,lower.tail = FALSE)
[1] 2.306004
```

a 95 percent confidence interval for μ is

$$\left[9 - 2.306\frac{(3.082)}{3},\ 9 + 2.306\frac{(3.082)}{3}\right] = (6.63,\ 11.37)$$

a larger interval than obtained in Exercise 4 ($(7.69, 10.31)$). The reason why the interval just obtained is larger than the one in Exercise 4 is twofold. The primary reason is that we have a larger estimated variance than in Exercise 4. In Exercise 4 we assumed that σ^2 was known to equal 4, whereas in this example we assumed it to be unknown and our estimate of it turned out to be 9.5, which resulted in a larger confidence interval. In fact, the confidence interval would have been larger than in Exercise 4 even if our estimate of σ^2 was again 4 because by having to estimate the variance we need to utilize the *t*-distribution, which has a greater variance and thus a larger spread than the standard normal (which can be used when σ^2 is assumed known). For instance, if it had turned out that $\bar{x} = 9$ and $s^2 = 4$, then our confidence interval would have been

$$\left[9 - 2.306\frac{2}{3},\ 9 + 2.306\frac{2}{3}\right] = (7.46,\ 10.54)$$

which is larger than that obtained in Exercise 4.

REMARK

The confidence interval for μ when σ is known is based on the fact that $\sqrt{n}\left(\bar{X}-\mu\right)/\sigma$ has a standard normal distribution. When σ is unknown, the foregoing approach is to estimate it by S and then use the fact that $\sqrt{n}\left(\bar{X}-\mu\right)/S$ has a t-distribution with $n-1$ degrees of freedom.

A *one-sided upper confidence interval* can be obtained by noting that

$$P\left\{\sqrt{n}\frac{\left(\bar{X}-\mu\right)}{S}<t_{\alpha,n-1}\right\}=1-\alpha$$

$$P\left\{\bar{X}-\mu<\frac{S}{\sqrt{n}}t_{\alpha,n-1}\right\}=1-\alpha$$

$$P\left\{\mu>\bar{X}-\frac{S}{\sqrt{n}}t_{\alpha,n-1}\right\}=1-\alpha$$

Hence, if it is observed that $\bar{X}=\bar{x}$, $S=s$, then we can assert "with $100(1-\alpha)$ percent confidence" that

$$\mu\in\left(\bar{x}-\frac{s}{\sqrt{n}}t_{\alpha,n-1}\ ,\ \infty\right)$$

Similarly, a $100(1-\alpha)$ *lower confidence interval* would be

$$\mu\in\left(-\infty\ ,\ \bar{x}+\frac{s}{\sqrt{n}}t_{\alpha,n-1}\right)$$

Finding Confidence Intervals with *R*

Exercise 9 Determine a 95 percent confidence interval for the average resting pulse of the members of a health club if a random selection of 15 members of the club yielded the data 54, 63, 58, 72, 49, 92, 70, 73, 69, 104, 48, 66, 80, 64, 77. Also determine a 95 percent lower confidence interval for this mean.

```
> x<-c(54, 63, 58, 72, 49, 92, 70, 73, 69, 104, 48, 66, 80, 64, 77)
> mean(x)
[1] 69.26667
> sd(x)
[1] 15.16795
```

```
> qt(0.975,df=14)
```
$t_{\alpha/2,n-1}$
```
[1] 2.144787
> qt(0.975,df=14)*sd(x)/sqrt(15)
```
$t_{\alpha/2,n-1}s/\sqrt{n}$
```
[1] 8.39973
> mean(x)-qt(0.975,df=14)*sd(x)/sqrt(15)
[1] 60.86694
> mean(x)+qt(0.975,df=14)*sd(x)/sqrt(15)
[1] 77.6664
```

The 95% confidence interval for the mean is (60.865, 77.6683)

```
> qt(0.95,df=14)
```
$t_{\alpha,n-1}$
```
[1] 1.76131
> qt(0.95,df=14)*sd(x)/sqrt(15)
```
$t_{\alpha,n-1}s/\sqrt{n}$
```
[1] 6.897903
> mean(x)+qt(0.95,df=14)*sd(x)/sqrt(15)
[1] 76.16457
```

The 95% lower confidence interval for the mean is (-infinity, 76.1662)

Our derivations of the $100(1-\alpha)$ percent confidence intervals for the population mean μ have assumed that the population distribution is normal. However, even when this is not the case, if the sample size is reasonably large then the intervals obtained will still be approximate $100(1-\alpha)$ percent confidence intervals for μ. This is true because, by the central limit theorem, $\sqrt{n}\left(\bar{X}-\mu\right)/\sigma$ will have approximately a normal distribution, and $\sqrt{n}\left(\bar{X}-\mu\right)/S$ will have approximately a *t*-distribution.

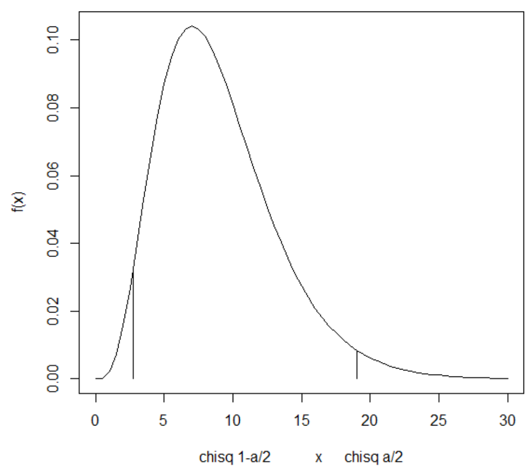
8.3.2 Confidence Intervals for the Variance of a Normal Distribution

If $X_1, X_2, \ldots, X_n$ is a sample from a normal distribution having unknown parameters μ and σ^2, then we can construct a confidence interval for σ^2 by using the fact that

$$(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1} \tag{8.29}$$

Hence,

$$P\left\{\chi^2_{1-\alpha/2,n-1} \le (n-1)\frac{S^2}{\sigma^2} \le \chi^2_{\alpha/2,n-1}\right\} = 1-\alpha$$



$$P\left\{\chi^2_{1-\alpha/2,n-1}\frac{1}{(n-1)S^2} \le \frac{1}{\sigma^2} \le \chi^2_{\alpha/2,n-1}\frac{1}{(n-1)S^2}\right\} = 1-\alpha$$

$$P\left\{\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right\} = 1-\alpha \tag{8.30}$$

Hence when $S^2 = s^2$, a $100(1-\alpha)$ percent confidence interval for σ^2 is

$$\sigma^2 \in \left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}\right) \tag{8.31}$$

Exercise 10 A standardized procedure is expected to produce washers (saibe) with very small deviation in their thicknesses. Suppose that 10 such washers were chosen and measured. If the thicknesses of these washers were, in inches,

0.123 0.133 0.124 0.125 0.126 0.128 0.120 0.124 0.130 0.126

what is a *90 percent confidence interval* for the standard deviation of the thickness of a washer produced by this procedure?

A computation gives that

```
wa<-c(0.123,0.133,0.124,0.125,0.126,0.128,0.120,0.124,0.130,0.126)
> var(wa)
[1] 1.365556e-05
```

$$s^2 = 1.366 \times 10^{-5}$$

90 percent confidence interval $\Rightarrow \quad \alpha = 0.1$

Because $\chi^2_{\alpha/2} = \chi^2_{0.05,9} = 16.917$ and $\chi^2_{1-\alpha/2} = \chi^2_{0.95} = 3.334$,

```
alfa<-0.1
> qchisq(alfa/2,df=9,lower.tail = FALSE)
[1] 16.91898
> qchisq(1-alfa/2,df=9,lower.tail = FALSE)
[1] 3.325113
```

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} = \frac{9 \times 1.366 \times 10^{-5}}{16.917} = 7.267 \times 10^{-6}$$

$$\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} = \frac{9 \times 1.366 \times 10^{-5}}{3.334} = 36.875 \times 10^{-6}$$

it follows that, with confidence 90%,

$$\sigma^2 \in (7.267 \times 10^{-6},\ 36.875 \times 10^{-6})$$

Taking square roots yields that, with confidence 0.90,

$$\sigma \in (2.696 \times 10^{-3}, 6.072 \times 10^{-3})$$

One-sided confidence intervals for σ^2 are obtained by similar reasoning and are presented in the Table, which sums up the results of this section.

Table: $100(1-\alpha)$ Percent Confidence Intervals

$$X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$$

$$\bar{X} = \sum_{i=1}^{n} X_i / n \qquad S = \sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 / (n-1)}$$

Assumption	Parameter	Confidence Inteval	Lower Interval	Upper Interval
σ^2 known	μ	$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$	$\left(-\infty, \bar{X} + z_\alpha \frac{\sigma}{\sqrt{n}}\right)$	$\left(\bar{X} - z_\alpha \frac{\sigma}{\sqrt{n}}, \infty\right)$
σ^2 unknown	μ	$\bar{X} \pm t_{\alpha/2,n-1} \frac{S}{\sqrt{n}}$	$\left(-\infty, \bar{X} + t_{\alpha,n-1} \frac{S}{\sqrt{n}}\right)$	$\left(\bar{X} - t_{\alpha,n-1} \frac{S}{\sqrt{n}}, \infty\right)$
μ unknown	σ^2	$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right)$	$\left(-\infty, \frac{(n-1)S^2}{\chi^2_{1-\alpha,n-1}}\right)$	$\left(\frac{(n-1)S^2}{\chi^2_{\alpha,n-1}}, \infty\right)$

8.4 ESTIMATING THE DIFFERENCE IN MEANS OF TWO NORMAL POPULATIONS

Let $X_1, X_2, \dots, X_n$ be a sample of size *n* from a normal population having mean μ_1 and variance σ_1^2 and let $Y_1, Y_2, \dots, Y_m$ be a sample of size *m* from a different normal population having mean μ_2 and variance σ_2^2 and suppose that the two samples are independent of each other. We are interested in estimating $\mu_1 - \mu_2$.

Since $\bar{X} = \sum X_i / n$ and $\bar{Y} = \sum Y_i / m$ are the maximum likelihood estimators of μ_1 and μ_2 it seems intuitive (and can be proven) that $\bar{X} - \bar{Y}$ is the maximum likelihood estimator of $\mu_1 - \mu_2$.

To obtain a confidence interval estimator, we need the distribution of $\bar{X} - \bar{Y}$. Because

$$\bar{X} \sim N(\mu_1, \sigma_1^2 / n)$$

$$\bar{Y} \sim N(\mu_2, \sigma_2^2 / m)$$

it follows from the fact that the sum of independent normal random variables is also normal, that

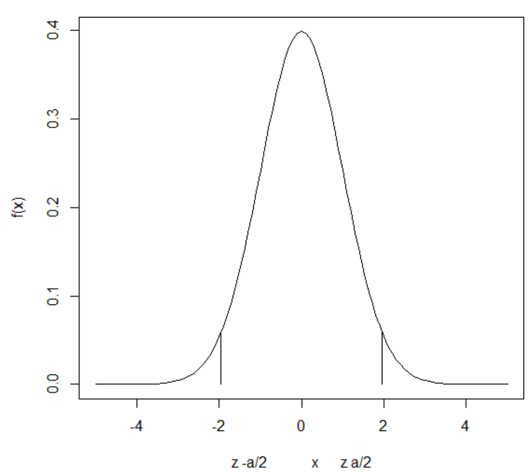
$$\bar{X}-\bar{Y} \sim N\left(\mu_1-\mu_2, \frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}\right) \quad (8.32)$$

Hence, assuming σ_1^2 and σ_2^2 are known, we have that

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} \sim N(0,1) \quad (8.33)$$

and so

$$\Pr\left(-z_{\alpha/2} < \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} < z_{\alpha/2}\right) = 1-\alpha \quad (8.34)$$



$$\Pr\left(-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}} < \bar{X}-\bar{Y}-(\mu_1-\mu_2) < z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\right) = 1-\alpha$$

$$\Pr\left(\bar{X}-\bar{Y}-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}} < \mu_1-\mu_2 < \bar{X}-\bar{Y}+z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\right) = 1-\alpha \quad (8.35)$$

Hence, if $\bar{X}$ and $\bar{Y}$ are observed to equal $\bar{x}$ and $\bar{y}$, respectively, then a $100(1-\alpha)$ two-sided confidence interval estimate for $\mu_1-\mu_2$ is

$$\mu_1-\mu_2 \in \left(\bar{x}-\bar{y}-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}, \quad \bar{x}-\bar{y}+z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}} \right) \qquad (8.36)$$

One-sided confidence intervals for $\mu_1-\mu_2$ are obtained in a similar fashion, and for exemple, a $100(1-\alpha)$ percent lower one-sided interval is given by

$$\mu_1-\mu_2 \in \left(-\infty,\ \bar{x}-\bar{y}+z_{\alpha}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}} \right)$$

Exercise 11 Two different types of electrical cable insulation have recently been tested to determine the voltage level at which failures tend to occur. When specimens were subjected to an increasing voltage stress in a laboratory experiment, failures for the two types of cable insulation occurred at the following voltages:

Type A		Type B	
36	54	52	60
44	52	64	44
41	37	38	48
53	51	68	46
38	44	66	70
36	35	52	62
34	44		

Suppose that it is known that the amount of voltage that cables having type A insulation can withstand is normally distributed with unknown mean μ_A and known variance $\sigma_A^2=40$, whereas the corresponding distribution for type B insulation is normal with unknown mean μ_B and known variance $\sigma_B^2=100$. Determine a 95 percent confidence interval for $\mu_A-\mu_B$. Determine a value that we can assert, with 95 percent confidence, exceeds $\mu_A-\mu_B$.

```
> n<-14
> m<-12
> A<-c(36,54,44,52,41,37,53,51,38,44,36,35,34,44)
> B<-c(52,60,64,44,38,48,68,46,66,70,52,62)
> mean(A)
[1] 42.78571
> mean(B)
[1] 55.83333

> alfa<-0.05
> za2<-qnorm(alfa/2, lower.tail = FALSE)
> za2
[1] 1.959964

> [ll<-mean(A)-mean(B)-za2*sqrt(40/n+100/m)
>
[1] -19.60412
> [ul<-mean(A)-mean(B)+za2*sqrt(40/n+100/m)
>
[1] -6.491114
```

The 95% confidence interval for the mean is (-19.6056, -6.4897)

```
> za<-qnorm(alfa, lower.tail = FALSE)
> za
[1] 1.644854

> ul<-mean(A)-mean(B)+za*sqrt(40/n+100/m)

[1] -7.723731
```

The 95% lower confidence interval for the mean is (-infinity, -7.7237)

Let us suppose now that we again desire an interval estimator of $\mu_1 - \mu_2$ but that the population variances σ_1^2 and σ_2^2 are unknown. In this case, it is natural to try to replace σ_1^2 and σ_2^2 in Equation (8.36)

$$\mu_1 - \mu_2 \in \left(\bar{x} - \bar{y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}, \quad \bar{x} - \bar{y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \right)$$

by the sample variances

$$S_1^2 = \sum_{i=1}^{n} \frac{\left(X_i - \bar{X}\right)^2}{n-1} \qquad S_2^2 = \sum_{i=1}^{n} \frac{\left(Y_i - \bar{Y}\right)^2}{m-1} \tag{8.37}$$

That is, it is natural to base our interval estimate on

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}}} \qquad (8.38)$$

However, to utilize the foregoing to obtain a confidence interval, *we need its distribution* and it must not depend on any of the unknown parameters σ_1^2 and σ_2^2. Unfortunately, this distribution is both complicated and does indeed depend on the unknown parameters σ_1^2 and σ_2^2. In fact, it is only in the special case when $\sigma_1^2=\sigma_2^2$ that we will be able to obtain an interval estimator. So let us suppose that the population variances, though unknown, are equal and let σ^2 denote their common value. Now, from Theorem (section 7.2.4b), it follows that

$$(n-1)\frac{S_1^2}{\sigma^2}\sim\chi_{n-1}^2 \text{ and } (m-1)\frac{S_2^2}{\sigma^2}\sim\chi_{m-1}^2 \qquad (8.39)$$

Also, because the samples are independent, it follows that these two chi-square random variables are independent. Hence, from the additive property of chi-square random variables, which states that the sum of independent chi-square random variables is also chi-square with a degree of freedom equal to the sum of their degrees of freedom, it follows that

$$(n-1)\frac{S_1^2}{\sigma^2}+(m-1)\frac{S_2^2}{\sigma^2}\sim\chi_{n+m-2}^2 \qquad (8.40)$$

Also, since

$$\bar{X}-\bar{Y}\sim N\left(\mu_1-\mu_2,\frac{\sigma^2}{n}+\frac{\sigma^2}{m}\right) \qquad (8.41)$$

we see that

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma^2}{n}+\frac{\sigma^2}{m}}}\sim N(0,1) \qquad (8.42)$$

Now it follows from the fundamental result that in normal sampling $\bar{X}$ and S^2 are independent (Theorem (section 7.2.4b)), that $\bar{X}_1$, S_1^2, $\bar{X}_2$, S_2^2 are independent random variables. Hence, using the definition of a *t*-random variable (as the ratio of two independent random variables, the numerator being a standard normal and the denominator being the square root of a chi-square random variable divided by its degree of freedom parameter), it follows from Equations (8.40) and (8.42) that if we let

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2} \tag{8.43}$$

Then

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma^2}{n} + \dfrac{\sigma^2}{m}}} \div \sqrt{\frac{S_p^2}{\sigma^2}} = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma^2}{n} + \dfrac{\sigma^2}{m}}} \div \sqrt{\frac{(n-1)S_1^2 + (m-1)S_2^2}{\sigma^2(n+m-2)}} = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{S_p^2\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}}$$

$$(8.44)$$

has a *t*-distribution with $n+m-2$ degrees of freedom. Consequently,

$$\Pr\left(-t_{\alpha/2,n+m-2} \le \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{S_p^2\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}} \le t_{\alpha/2,n+m-2}\right) = 1-\alpha \tag{8.45}$$

$$\Pr\left(-t_{\alpha/2,n+m-2}\sqrt{S_p^2\left(\frac{1}{n} + \frac{1}{m}\right)} \le \bar{X} - \bar{Y} - (\mu_1 - \mu_2) \le t_{\alpha/2,n+m-2}\sqrt{S_p^2\left(\frac{1}{n} + \frac{1}{m}\right)}\right) = 1-\alpha$$

$$\Pr\left(\bar{X} - \bar{Y} - t_{\alpha/2,n+m-2}\sqrt{S_p^2\left(\frac{1}{n} + \frac{1}{m}\right)} \le \mu_1 - \mu_2 \le \bar{X} - \bar{Y} + t_{\alpha/2,n+m-2}\sqrt{S_p^2\left(\frac{1}{n} + \frac{1}{m}\right)}\right) = 1-\alpha \tag{8.46}$$

Therefore, when the data result in the values $\bar{X} = \bar{x}$, $\bar{Y} = \bar{y}$, $S_p = s_p$, we obtain the following $100(1-\alpha)$ percent confidence interval for $\mu_1 - \mu_2$:

$$\left(\bar{x} - \bar{y} - t_{\alpha/2,n+m-2}s_p\sqrt{\frac{1}{n} + \frac{1}{m}},\ \bar{x} - \bar{y} + t_{\alpha/2,n+m-2}s_p\sqrt{\frac{1}{n} + \frac{1}{m}}\right) \tag{8.47}$$

One-sided confidence intervals are similarly obtained.

Exercise 12 There are two different techniques a given manufacturer can employ to produce batteries. A random selection of 12 batteries produced by technique I and of 14 produced by technique II resulted in the following capacities (in ampere hours):

Technique I		Technique II	
140	132	144	134
136	142	132	130
138	150	136	146
150	154	140	128
152	136	128	131
144	142	150	137
		130	135

Determine a 90 percent level two-sided confidence interval for the difference in means, assuming a common variance. Also determine a 95 percent upper confidence interval for $\mu_I - \mu_{II}$.

We run Program R.

```
> alfa<-0.1
> n<-12
> m<-14
> I<-c(140,132,136,142,138,150,150,154,152,136,144,142)
> II<-c(144,134,132,130,136,146,140,128,128,131,150,137,130,135)
> mean(I)
[1] 143
> mean(II)
[1] 135.7857

> ta2<-qt(alfa/2,df=n+m-2,lower.tail = FALSE)
> ta2
[1] 1.710882

> sp<-sqrt(((n-1)*var(I)+(m-1)*var(II))/(n+m-2))
> sp
[1] 7.007012

> [ll<- mean(I)-mean(II)-ta2*sp*sqrt(1/n+1/m)
>
[1] 2.498164
> [ul<- mean(I)-mean(II)+ta2*sp*sqrt(1/n+1/m)
>
[1] 11.93041
```

The 90% confidence interval for the mean difference is (2.4971, 11.9315)

The 95% upper confidence interval for the mean difference is (2.4971, infinity)

REMARK

The confidence interval given by Equation (8.47) was obtained under the assumption that the population variances are equal; with σ^2 as their common value, it follows that

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma^2}{n}+\frac{\sigma^2}{m}}}=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}$$

has a standard normal distribution. However, since σ^2 is unknown this result cannot be immediately applied to obtain a confidence interval; σ^2 must first be estimated. To do so, note that both sample variances are estimators of σ^2; moreover, since S_1^2 has $n-1$ degrees of freedom and S_2^2 has $m-1$, the appropriate estimator is to take a weighted average of the two sample variances, with the weights proportional to these degrees of freedom. That is, the estimator of σ^2 is the pooled estimator

$$S_p^2=\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2} \qquad (8.48)$$

and the confidence interval is then based on the statistic

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{S_p^2}\sqrt{\frac{1}{n}+\frac{1}{m}}} \qquad (8.49)$$

which, by our previous analysis, has a *t*-distribution with $n+m-2$ degrees of freedom.