

Topic 16

Interval Estimation

*The form of this solution consists in determining certain intervals, which I propose to call the confidence intervals..., in which we may assume are contained the values of the estimated characters of the population, the probability of an error is a statement of this sort being equal to or less than $1 - \epsilon$, where ϵ is any number $0 < \epsilon < 1$, chosen in advance. The number ϵ I call the confidence coefficient. - Jerzy Neyman, 1934, On the Two Different Aspects of the Representative Method, *Journal of the Royal Statistical Society**

Our strategy to estimation thus far has been to use a method to find an estimator, e.g., method of moments, or maximum likelihood, and evaluate the quality of the estimator by evaluating its bias and the variance. Often, we know more about the distribution of the estimator and this allows us to take a more comprehensive statement about the estimation procedure.

Interval estimation is an extension to the variety of techniques we have examined. Given data \mathbf{x} , we replace the point estimate $\hat{\theta}(\mathbf{x})$ for the parameter θ by a statistic that is subset $\hat{C}(\mathbf{x})$ of the parameter space. We will consider both the classical and Bayesian approaches to choosing $\hat{C}(\mathbf{x})$. As we shall learn, the two approaches have very different interpretations.

16.1 Classical Statistics

In this case, the random set $\hat{C}(X)$ is chosen to have a prescribed high probability, γ , of containing the true parameter value θ . In symbols,

$$P_{\theta}\{\theta \in \hat{C}(X)\} = \gamma.$$

In this case, the set $\hat{C}(\mathbf{x})$ is called a γ -**level confidence set**. In the case of a one dimensional parameter set, the typical choice of confidence set is a **confidence interval**

$$\hat{C}(\mathbf{x}) = (\hat{\theta}_l(\mathbf{x}), \hat{\theta}_u(\mathbf{x})).$$

Often this interval takes the form

$$\hat{C}(\mathbf{x}) = (\hat{\theta}(\mathbf{x}) - m(\mathbf{x}), \hat{\theta}(\mathbf{x}) + m(\mathbf{x})) = \hat{\theta}(\mathbf{x}) \pm m(\mathbf{x})$$

where the two statistics,

- $\hat{\theta}(\mathbf{x})$ is a **point estimate**, and
- $m(\mathbf{x})$ is the **margin of error**.

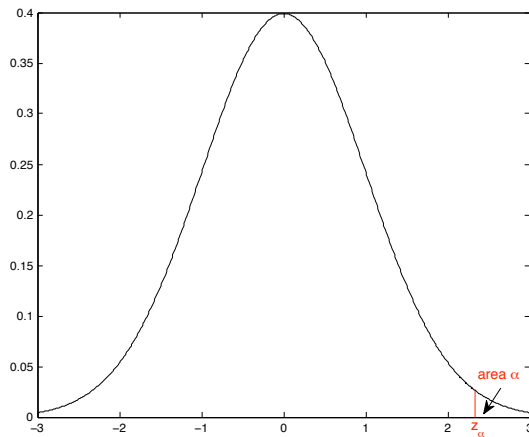


Figure 16.1: Upper tail critical values. α is the area under the standard normal density and to the right of the vertical line at critical value z_{α}

16.1.1 Means

Example 16.1 (1-sample z interval). If X_1, X_2, \dots, X_n are normal random variables with unknown mean μ but known variance σ_0^2 . Then,

$$Z = \frac{\bar{X} - \mu}{\sigma_0/\sqrt{n}}$$

is a standard normal random variable. For any α between 0 and 1, choose z_α so that

$$P\{Z > z_\alpha\} = \alpha \quad \text{or equivalently} \quad P\{Z \leq z_\alpha\} = 1 - \alpha.$$

The value is known as the **upper tail probability with critical value** z_α . We can compute this in \mathbb{R} using, for example

```
> qnorm(0.975)
[1] 1.959964
```

for $\alpha = 0.025$.

If $\gamma = 1 - 2\alpha$, then $\alpha = (1 - \gamma)/2$. In this case, we have that

$$P\{-z_\alpha < Z < z_\alpha\} = \gamma.$$

Let μ_0 is the state of nature. Taking in turn each the two inequalities in the line above and isolating μ_0 , we find that

$$\begin{aligned} \frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}} = Z < z_\alpha \\ \bar{X} - \mu_0 < z_\alpha \frac{\sigma_0}{\sqrt{n}} \\ \bar{X} - z_\alpha \frac{\sigma_0}{\sqrt{n}} < \mu_0 \end{aligned}$$

Similarly,

$$\frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}} = Z > -z_\alpha$$

implies

$$\mu_0 < \bar{X} + z_\alpha \frac{\sigma_0}{\sqrt{n}}$$

Thus

$$\bar{X} - z_\alpha \frac{\sigma_0}{\sqrt{n}} < \mu_0 < \bar{X} + z_\alpha \frac{\sigma_0}{\sqrt{n}}.$$

has probability γ . Thus, for data \mathbf{x} ,

$$\bar{x} \pm z_{(1-\gamma)/2} \frac{\sigma_0}{\sqrt{n}}$$

is a confidence interval with confidence level γ . In this case,

$$\hat{\mu}(\mathbf{x}) = \bar{x} \text{ is the estimate for the mean and } m(\mathbf{x}) = z_{(1-\gamma)/2} \sigma_0/\sqrt{n} \text{ is the margin of error.}$$

We can use the **z -interval** above for the confidence interval for μ for data that is not necessarily normally distributed as long as the central limit theorem applies. For one population intervals for means, $n > 30$ and data not strongly skewed is a good rule of thumb.

Generally, the standard deviation is not known and must be estimated. So, let X_1, X_2, \dots, X_n be normal random variables with unknown mean and unknown standard deviation. Let S^2 be the **unbiased** sample variance. If we are forced to replace the unknown variance σ^2 with its unbiased estimate s^2 , then the statistic is known as t :

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}.$$

The term s/\sqrt{n} which estimates the standard deviation of the sample mean is called the **standard error**. The remarkable discovery by William Gossett is that the distribution of the t statistic can be determined **exactly**. Write

$$T_{n-1} = \frac{\sqrt{n}(\bar{X} - \mu)}{S}.$$

Then, Gossett was able to establish the following three facts:

- The numerator is a standard normal random variable.
- The denominator is the square root of

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

This sum has chi-square distribution with $n - 1$ degrees of freedom.

- The numerator and denominator are **independent**.

With this, Gossett was able to compute the density of the t **distribution** with $n - 1$ **degrees of freedom**. Gossett, who worked for the the brewery of Arthur Guinness in Dublin, was permitted to publish his results only if it appeared under a pseudonym. Gosset chose the name *Student*, thus the distribution is sometimes known as **Student's t** .

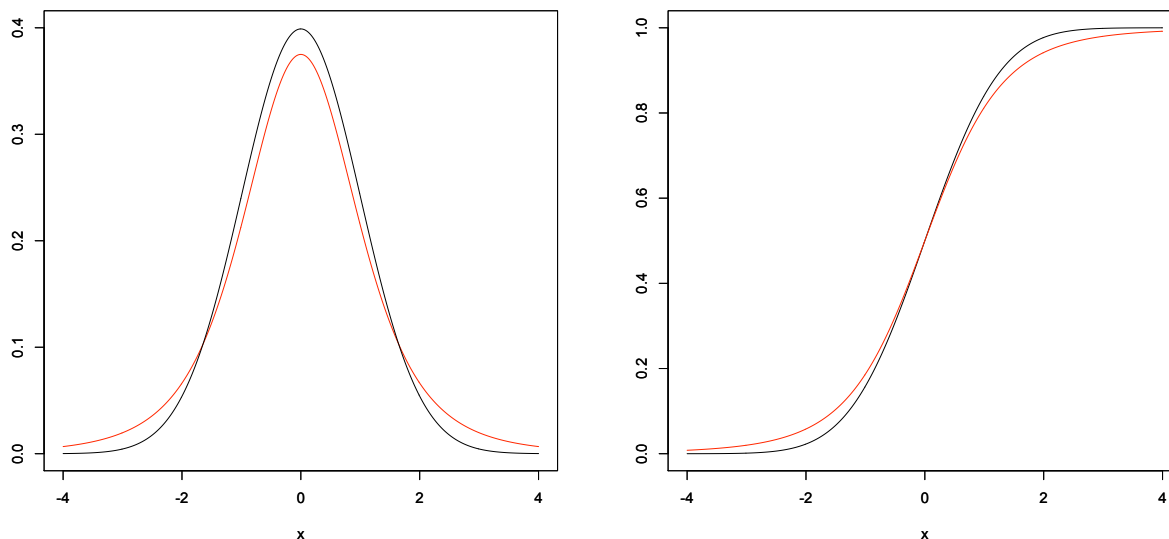


Figure 16.2: The density and distribution function for a standard normal random variable (black) and a t random variable with 4 degrees of freedom (red). The variance of the t distribution is $df/(df - 2) = 4/(4 - 2) = 2$ is higher than the variance of a standard normal. This can be seen in the broader shoulders of the t density function or in the smaller increases in the t distribution function away from the mean of 0.

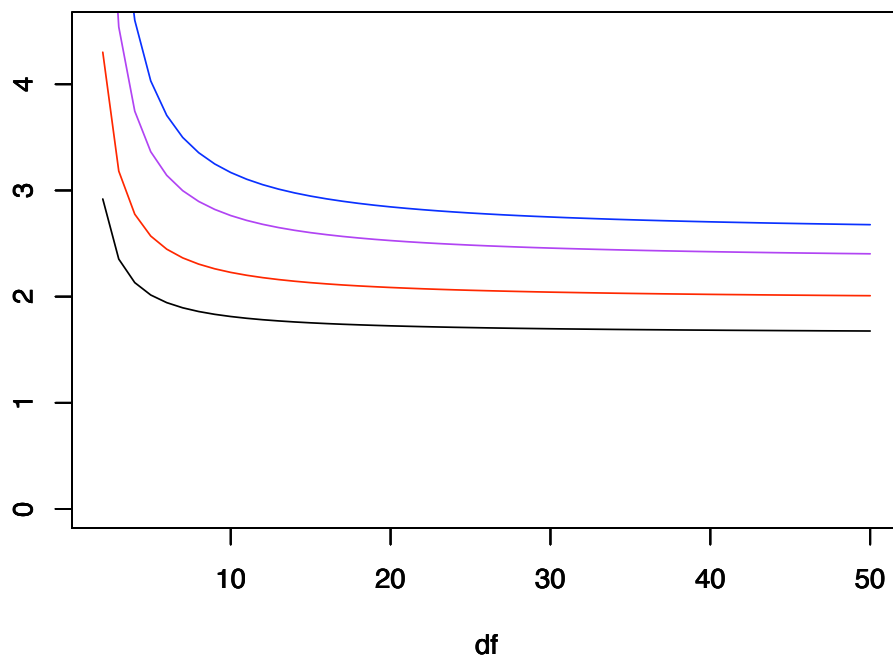


Figure 16.3: Upper critical values for the t confidence interval with $\gamma = 0.90$ (black), 0.95 (red), 0.98 (magenta) and 0.99 (blue) as a function of df , the number of degrees of freedom. Note that these critical values decrease to the critical value for the z confidence interval and increases with γ .

Again, for any α between 0 and 1, let **upper tail probability** $t_{n-1,\alpha}$ satisfy

$$P\{T_{n-1} > t_{n-1,\alpha}\} = \alpha \quad \text{or equivalently} \quad P\{T_{n-1} \leq t_{n-1,\alpha}\} = 1 - \alpha.$$

We can compute this in **R** using, for example

```
> qt(0.975, 12)
[1] 2.178813
```

for $\alpha = 0.025$ and $n - 1 = 12$.

Example 16.2. For the data on the lengths of 200 *Bacillus subtilis*, we had a mean $\bar{x} = 2.49$ and standard deviation $s = 0.674$. For a 96% confidence interval $\alpha = 0.02$ and we type in **R**,

```
> qt(0.98, 199)
[1] 2.067298
```

Thus, the interval is

$$2.490 \pm 2.0674 \frac{0.674}{\sqrt{200}} = 2.490 \pm 0.099 \quad \text{or} \quad (2.391, 2.589)$$

Example 16.3. We can obtain the data for the Michaelson-Morley experiment using **R** by typing

```
> data(morley)
```

The data have 100 rows - 5 experiments (column 1) of 20 runs (column 2). The Speed is in column 3. The values for speed are the amounts over 299,000 km/sec. Thus, a t -confidence interval will have 99 degrees of freedom. We can see a histogram by writing `hist(morley$Speed)`. To determine a 95% confidence interval, we find



Figure 16.4: Measurements of the speed of light. Actual values are 299,000 kilometers per second plus the value shown.

```
> mean(morley$Speed)
[1] 852.4
> sd(morley$Speed)
[1] 79.01055
> qt(0.975, 99)
[1] 1.984217
```

Thus, our confidence interval for the speed of light is

$$299,852.4 \pm 1.9842 \frac{79.0}{\sqrt{100}} = 299,852.4 \pm 15.7 \quad \text{or the interval } (299836.7, 299868.1)$$

This confidence interval does not include the presently determined values of 299,792.458 km/sec for the speed of light. The confidence interval can also be found by typing `t.test(morley$Speed)`. We will study this command in more detail when we describe the *t*-test.

Exercise 16.4. Give a 90% and a 98% confidence interval for the example above.

We often wish to determine a sample size that will guarantee a desired margin of error. For a γ -level *t*-interval, this is

$$m = t_{n-1, (1-\gamma)/2} \frac{s}{\sqrt{n}}.$$

Solving this for n yields

$$n = \left(\frac{t_{n-1, (1-\gamma)/2} s}{m} \right)^2.$$

Because the number of degrees of freedom, $n - 1$, for the *t* distribution is unknown, the quantity n appears on both sides of the equation and the value of s is unknown. We search for a **conservative** value for n , i.e., a margin of error that will be no greater than the desired length. This can be achieved by overestimating $t_{n-1, (1-\gamma)/2}$ and s . For the

speed of light example above, if we desire a margin of error of $m = 10$ km/sec for a 95% confidence interval, then we set $t_{n-1, (1-\gamma)/2} = 2$ and $s = 80$ to obtain

$$n \approx \left(\frac{2 \cdot 80}{10} \right)^2 = 256$$

measurements are necessary to obtain the desired margin of error..

The next set of confidence intervals are determined, in those case in which the distributional variance is known, by finding the standardized score and using the normal approximation as given via the central limit theorem. In the cases in which the variance is unknown, we replace the distribution variance with a variance that is estimated from the observations. In this case, the procedure that is analogous to the standardized score is called the **studentized score**.

Example 16.5 (matched pair t interval). *We begin with two quantitative measurements*

$$(X_{1,1}, \dots, X_{1,n}) \quad \text{and} \quad (X_{2,1}, \dots, X_{2,n}),$$

on the same n individuals. Assume that the first set of measurements has mean μ_1 and the second set has mean μ_2 .

If we want to determine a confidence interval for the difference $\mu_1 - \mu_2$, we can apply the t -procedure to the differences

$$(X_{1,1} - X_{2,1}, \dots, X_{1,n} - X_{2,n})$$

to obtain the confidence interval

$$(\bar{X}_1 - \bar{X}_2) \pm t_{n-1, (1-\gamma)/2} \frac{S_d}{\sqrt{n}}$$

where S_d is the standard deviation of the difference.

Example 16.6 (2-sample z interval). *If we have two independent samples of normal random variables*

$$(X_{1,1}, \dots, X_{1,n_1}) \quad \text{and} \quad (X_{2,1}, \dots, X_{2,n_2}),$$

the first having mean μ_1 and variance σ_1^2 and the second having mean μ_2 and variance σ_2^2 , then the difference in their sample means

$$\bar{X}_2 - \bar{X}_1$$

is also a normal random variable with

$$\text{mean } \mu_1 - \mu_2 \quad \text{and} \quad \text{variance } \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Therefore,

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

is a standard normal random variable. In the case in which the variances σ_1^2 and σ_2^2 are known, this gives us a γ -level confidence interval for the difference in parameters $\mu_1 - \mu_2$.

$$(\bar{X}_1 - \bar{X}_2) \pm z_{(1-\gamma)/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

Example 16.7 (2-sample t interval). *If we know that $\sigma_1^2 = \sigma_2^2$, then we can pool the data to compute the standard deviation. Let S_1^2 and S_2^2 be the sample variances from the two samples. Then the **pooled sample variance** S_p^2 is the weighted average of the sample variances with weights equal to their respective degrees of freedom.*

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

This gives a statistic

$$T_{n_1+n_2-2} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

that has a t distribution with $n_1 + n_2 - 2$ degrees of freedom. Thus we have the γ level confidence interval

$$(\bar{X}_1 - \bar{X}_2) \pm t_{n_1+n_2-2, (1-\gamma)/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

for $\mu_1 - \mu_2$.

If we do not know that $\sigma_1^2 = \sigma_2^2$, then the corresponding studentized random variable

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

no longer has a t -distribution.

Welch and Satterthwaite have provided an approximation to the t distribution with **effective degrees of freedom** given by the Welch-Satterthwaite equation

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)}}. \quad (16.1)$$

This give a γ -level confidence interval

$$\bar{x}_1 - \bar{x}_2 \pm t_{\nu, (1-\gamma)/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

For two sample intervals, the number of observations per group may need to be at least 40 for a good approximation to the normal distribution.

Exercise 16.8. Show that the effective degrees is between the worst case of the minimum choice from a one sample t -interval and the best case of equal variances.

$$\min\{n_1, n_2\} - 1 \leq \nu \leq n_1 + n_2 - 2$$

For data on the life span in days of 88 wildtype and 99 transgenic mosquitoes, we have the summary

	observations	mean	standard deviation
wildtype	88	20.784	12.99
transgenic	99	16.546	10.78

Using the conservative 95% confidence interval based on $\min\{n_1, n_2\} - 1 = 87$ degrees of freedom, we use

```
> qt(0.975, 87)
[1] 1.987608
```

to obtain the interval

$$(20.78 - 16.55) \pm 1.9876 \sqrt{\frac{12.99^2}{88} + \frac{10.78^2}{99}} = (0.744, 7.733)$$

Using the the Welch-Satterthwaite equation, we obtain $\nu = 169.665$. The increase in the number of degrees of freedom gives a slightly narrower interval (0.768, 7.710).

16.1.2 Linear Regression

For ordinary linear regression, we have given least squares estimates for the slope β and the intercept α . For data $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, our model is

$$y_i = \alpha + \beta x_i + \epsilon_i$$

where ϵ_i are independent $N(0, \sigma)$ random variables. Recall that the estimator for the slope

$$\hat{\beta}(x, y) = \frac{\text{cov}(x, y)}{\text{var}(x)}$$

is unbiased.

Exercise 16.9. Show that

$$\text{Var}_{(\alpha, \beta)}(\hat{\beta}) = \frac{\sigma^2}{(n-1)\text{var}(x)}.$$

$$\text{Var}_{(\alpha, \beta)}(\hat{\alpha}) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{(n-1)\text{var}(x)} \right) = \sigma^2 \left(\frac{(n-1)\text{var}(x) + n\bar{x}^2}{n(n-1)\text{var}(x)} \right) = \frac{\sigma^2 \bar{x}^2}{(n-1)\text{var}(x)},$$

and

$$\text{Cov}_{(\alpha, \beta)}(\hat{\alpha}, \hat{\beta}) = -\bar{x}\text{Var}_{(\alpha, \beta)}(\hat{\beta})$$

The last equality for $\text{Var}_{(\alpha, \beta)}(\hat{\alpha})$ uses formula (2.2) for the sample variance.

Notice that $\text{Var}_{(\alpha, \beta)}(\hat{\alpha})$ increases with the distance that the mean of the x values is from 0. The correlation

$$\rho_{(\alpha, \beta)}(\hat{\alpha}, \hat{\beta}) = \frac{\text{Cov}_{(\alpha, \beta)}(\hat{\alpha}, \hat{\beta})}{\sqrt{\text{Var}_{(\alpha, \beta)}(\hat{\alpha})\text{Var}_{(\alpha, \beta)}(\hat{\beta})}} = -\bar{x} \sqrt{\frac{\text{Var}_{(\alpha, \beta)}(\hat{\beta})}{\text{Var}_{(\alpha, \beta)}(\hat{\alpha})}} = -\frac{\bar{x}}{\sqrt{\bar{x}^2}},$$

which does not depend on the data. If the mean of the explanatory variable $\bar{x} > 0$, then $\hat{\alpha}$ and $\hat{\beta}$ are negatively correlated. For example, if we underestimate $\hat{\beta}$ for $\beta > 0$, then the line is more shallow and we will likely overestimate $\hat{\alpha}$.

Exercise 16.10. Explore the fact that the correlation between $\hat{\alpha}$ and $\hat{\beta}$ does not depend on the data.

If σ is known, this suggests a z -interval for a γ -level confidence interval

$$\hat{\beta} \pm z_{(1-\gamma)/2} \frac{\sigma}{s_x \sqrt{n-1}}.$$

Generally, σ is unknown. However, the variance of the residuals,

$$s_u^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - (\hat{\alpha} + \hat{\beta}x_i))^2 \quad (16.2)$$

is an unbiased estimator of σ^2 and s_u/σ has a t distribution with $n-2$ degrees of freedom. This gives the t -interval

$$\hat{\beta} \pm t_{n-2, (1-\gamma)/2} \frac{s_u}{s_x \sqrt{n-1}}.$$

As the formula shows, the margin of error is proportional to the standard deviation of the residuals. It is inversely proportional to the standard deviation of the x measurement. Thus, we can reduce the margin of error by taking a broader set of values for the explanatory variables.

For the data on the humerus and femur of the five specimens of *Archeopteryx*, we have $\hat{\beta} = 1.197$, $s_u = 1.982$, $s_x = 13.2$, and $t_{3, 0.025} = 3.1824$. Thus, the 95% confidence interval is 1.197 ± 0.239 or $(0.958, 1.436)$.

16.1.3 Sample Proportions

Example 16.11 (proportions). For n Bernoulli trials with success parameter p , the sample proportion \hat{p} has

$$\text{mean } p \text{ and variance } \frac{p(1-p)}{n}.$$

The parameter p appears in both the mean and in the variance. Thus, we need to make a choice \tilde{p} to replace p in the confidence interval

$$\hat{p} \pm z_{(1-\gamma)/2} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}. \quad (16.3)$$

One simple choice for \tilde{p} is simply to take the sample proportion \hat{p} . Based on extensive numerical experimentation, one more recent popular choice is

$$\tilde{p} = \frac{x+2}{n+4}$$

where x is the number of successes.

For population proportions, we ask that the **mean number of successes** np and the **mean number of failures** $n(1-p)$ each be at least 10. We have this requirement so that a normal random variable is a good approximation to the appropriate binomial random variable.

Example 16.12. For Mendel's data the F_2 generation consisted 428 for the dominant allele green pods and 152 for the recessive allele yellow pods. Thus, the sample proportion of green pod alleles is

$$\hat{p} = \frac{428}{428+152} = 0.7379.$$

The confidence interval, using

$$\tilde{p} = \frac{428+2}{428+152+4} = 0.7363$$

is

$$0.7379 \pm z_{(1-\gamma)/2} \sqrt{\frac{0.7363 \cdot 0.2637}{580}} = 0.7379 \pm 0.0183 z_{(1-\gamma)/2}$$

For $\gamma = 0.98$, $z_{0.01} = 2.326$ and the confidence interval is $0.7379 \pm 0.0426 = (0.6953, 0.7805)$. Note that this interval contains the predicted value of $p = 3/4$.

A comparable formula gives confidence intervals based on more than two independent samples

Example 16.13. For the difference in two proportions p_1 and p_2 based on n_1 and n_2 independent trials. We have, for the difference $p_1 - p_2$, the confidence interval

$$\hat{p}_1 - \hat{p}_2 \pm \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

Example 16.14 (transformation of a single parameter). If

$$(\hat{\theta}_\ell, \hat{\theta}_u)$$

is a level γ confidence interval for θ and g is an increasing function, then

$$(g(\hat{\theta}_\ell), g(\hat{\theta}_u))$$

is a level γ confidence interval for $g(\theta)$

Exercise 16.15. For the example above, find the confidence interval for the yellow pod genotype.

16.1.4 Summary of Standard Confidence Intervals

The confidence interval is an extension of the idea of a point estimation of the parameter to an interval that is likely to contain the true parameter value. A level γ confidence interval for a population parameter θ is an interval computed from the sample data having probability γ of producing an interval containing θ .

For an estimate of a population mean or proportion, a level γ confidence interval often has the form

$$\text{estimate} \pm t^* \times \text{standard error}$$

where t^* is the upper $\frac{1-\gamma}{2}$ critical value for the t distribution with the appropriate number of degrees of freedom. If the number of degrees of freedom is **infinite**, we use the **standard normal** distribution to determine the critical value, usually denoted by z^* .

The margin of error $m = t^* \times \text{standard error}$ decreases if

- γ , the confidence level, decreases
- the standard deviation decreases
- n , the number of observations, increases

The procedures for finding the confidence interval are summarized in the table below.

procedure	parameter	estimate	standard error	degrees of freedom
one sample	μ	\bar{x}	$\frac{s}{\sqrt{n}}$	$n - 1$
two sample	$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	See (16.1)
pooled two sample	$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	$n_1 + n_2 - 2$
one proportion	p	\hat{p}	$\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}$, $\tilde{p} = \frac{x+2}{n+4}$	∞
two proportion	$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	∞
linear regression	β	$\hat{\beta} = \text{cov}(x, y)/\text{var}(x)$	$\frac{s_u}{s_x \sqrt{n-1}}$	$n - 2$

The first confidence interval for $\mu_1 - \mu_2$ is the two-sample t procedure. If we can assume that the two samples have a common standard deviation, then we pool the data to compute s_p , the pooled standard deviation. Matched pair procedures use a one sample procedure on the difference in the observed values.

For these intervals, we need a sample size large enough so that the central limit theorem is a sufficiently good approximation. For one population tests for means, $n > 30$ and data not strongly skewed is a good rule of thumb. For two population tests, 40 observations for each group may be necessary. For population proportions, we ask that the mean number of successes np and the mean number of failures $n(1-p)$ each be at least 10.

For the standard error for $\hat{\beta}$ in linear regression, s_u is defined in (16.2) and s_x is the standard deviation of the values of the explanatory variable.

16.1.5 Interpretation of the Confidence Interval

The confidence interval for a parameter θ is based on two statistics - $\hat{\theta}_l(\mathbf{x})$, the lower end of the confidence interval and $\hat{\theta}_u(\mathbf{x})$, the upper end of the confidence interval. As with all statistics, these two statistics cannot be based on the value of the parameter. In addition, the formulas for these two statistics are determined in advance of having the actual data. The term confidence can be related to the production of confidence intervals. We can think of the situation in

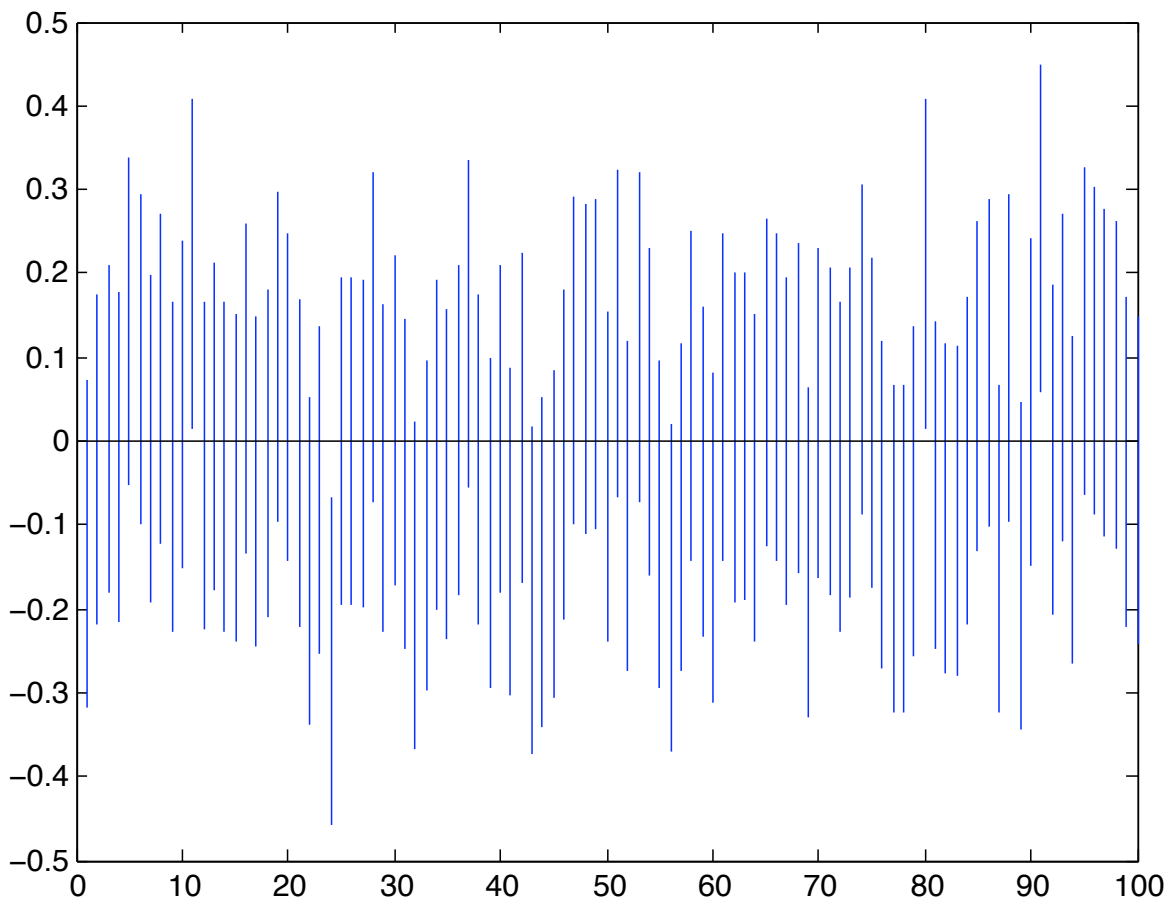


Figure 16.5: One hundred confidence build from repeatedly simulating 100 standard normal random variables and constructing 95% confidence intervals for the mean value - 0. Note that the 24th interval is entirely below 0 and so does not contain the actual parameter. The 11th, 80th and 91st intervals are entirely above 0 and again do not contain the parameter.

which we produce independent confidence intervals repeatedly. Each time, we may either succeed or fail to include the true parameter in the confidence interval. In other words, the inclusion of the parameter value in the confidence interval is a Bernoulli trial with success probability γ .

For example, after having seen these 100 intervals in Figure 5, we can conclude that the lowest and highest intervals are much less likely than 95% of containing the true parameter value. This phenomena can be seen in the presidential polls for the 2012 election. Three days before the election we see the following spread between Mr. Obama and Mr. Romney

$$0\% \quad -1\% \quad 0\% \quad 1\% \quad 5\% \quad 0\% \quad -5\% \quad -1\% \quad 1\% \quad 1\%$$

with the 95% confidence interval having a margin of error $\sim 3\%$ based on a sample of size ~ 1000 . Because these values are highly dependent, the values of $\pm 5\%$ is less likely to contain the true spread.

Exercise 16.16. Perform the computations needed to determine the margin of error in the example above.

The following example, although never likely to be used in an actual problem, may shed some insight into the difference between confidence and probability.

Example 16.17. Let X_1 and X_2 be two independent observations from a uniform distribution on the interval $[\theta - 1, \theta + 1]$ where θ is an unknown parameter. In this case, an observation is greater than θ with probability $1/2$, and less than θ with probability $1/2$. Thus,

- with probability $1/4$, both observations are above θ ,
- with probability $1/4$, both observations are below θ , and
- with probability $1/2$, one observation is below θ and the other is above.

In the third case alone, the confidence interval contains the parameter value. As a consequence of these considerations, the interval

$$(\hat{\theta}_l(X_1, X_2), \hat{\theta}_u(X_1, X_2)) = (\min\{X_1, X_2\}, \max\{X_1, X_2\})$$

is a 50% confidence interval for the parameter.

Sometimes, $\max\{X_1, X_2\} - \min\{X_1, X_2\} > 1$. Because any subinterval of the interval $[\theta - 1, \theta + 1]$ that has length at least 1 must contain θ , the midpoint of the interval, this confidence interval must contain the parameter value. In other words, sometimes the 50% confidence interval is certain to contain the parameter.

Exercise 16.18. For the example above, show that

$$P\{\text{confidence interval has length} > 1\} = 1/4.$$

Hint: Draw the square $[\theta - 1, \theta + 1] \times [\theta - 1, \theta + 1]$ and shade the region in which the confidence interval has length greater than 1.

16.1.6 Extensions on the Use of Confidence Intervals

Example 16.19 (delta method). For estimating the distribution μ by the sample mean \bar{X} , the delta method provides an alternative for the example above. In this case, the standard deviation of $g(\bar{X})$ is approximately

$$\frac{|g'(\mu)|\sigma}{\sqrt{n}}. \quad (16.4)$$

We replace μ with \bar{X} to obtain the confidence interval for $g(\mu)$

$$g(\bar{X}) \pm z_{\alpha/2} \frac{|g'(\bar{X})|\sigma}{\sqrt{n}}.$$

Using the notation for the example of estimating α_3 , the coefficient of volume expansion based on independent length measurements, Y_1, Y_2, \dots, Y_n measured at temperature T_1 of an object having length ℓ_0 at temperature T_0 .

$$\frac{\bar{Y}^3 - \ell_0^3}{\ell_0^3 |T_1 - T_0|} \pm z_{(1-\gamma)/2} \frac{3\bar{Y}^2 \sigma_Y}{n}$$

Exercise 16.20. Recall that the odds of an event having probability p is

$$\psi = \frac{p}{1-p}. \quad (16.5)$$

Use the delta method to show that

$$\sigma_{\hat{\psi}}^2 \approx \frac{\psi(\psi+1)^2}{n}.$$

In the example above on green peas,

$$\hat{\psi} = \frac{\hat{p}}{1 - \hat{p}} = \frac{0.7379}{1 - 0.7379} = 2.8153.$$

Using (16.4), we obtain a 98% confidence interval

$$\hat{\psi} \pm z_{0.01} \sqrt{\sigma_{\hat{\psi}}^2} \approx 2.8153 \pm 2.326 \sqrt{\frac{2.8153(1 + 2.8153)^2}{580}} = (2.1970, 3.4337)$$

which includes the predicted value $\psi = 3$.

If we transform the 98% confidence interval (0.6953, 0.7805) for p to a confidence interval for the odds ψ using the transformation (16.5), we obtain an interval (2.2819, 3.5558) that is slightly shifted upward from the confidence interval obtained by the delta method.

For multiple independent samples, the simple idea using the transformation in the Example 12 no longer works. For example, to determine the confidence interval using \bar{X}_1 and \bar{X}_2 above, the confidence interval for $g(\mu_1, \mu_2)$, the delta method gives the confidence interval

$$g(\bar{X}_1, \bar{X}_2) \pm z_{(1-\gamma)/2} \sqrt{\left(\frac{\partial}{\partial x} g(\bar{X}_1, \bar{X}_2)\right)^2 \frac{\sigma_1^2}{n_1} + \left(\frac{\partial}{\partial y} g(\bar{X}_1, \bar{X}_2)\right)^2 \frac{\sigma_2^2}{n_2}}.$$

Example 16.21. Let's return to the example of n_ℓ and n_h measurements x and y of, respectively, the length ℓ and the height h of a right triangle with the goal of giving the angle

$$\theta = g(\ell, h) = \tan^{-1}\left(\frac{h}{\ell}\right)$$

between these two sides. Here are the measurements:

```
> x
[1] 10.224484 10.061800  9.945213  9.982061  9.961353 10.173944  9.952279  9.855147
[9]  9.737811  9.956345
> y
[1] 4.989871 5.090002 5.021615 4.864633 5.024388 5.123419 5.033074 4.750892 4.985719
[10] 4.912719 5.027048 5.055755
> mean(x) ; sd(x)
[1] 9.985044
[1] 0.1417969
> mean(y) ; sd(y)
[1] 4.989928
[1] 0.1028745
```

The angle θ is the arctangent, here estimated using the mean and given in radians

```
> (thetahat<-atan(mean(y)/mean(x)))
[1] 0.4634398
```

Using the delta method, we have estimated the standard deviation of these measurements.

$$\sigma_{\hat{\theta}} \approx \frac{1}{h^2 + \ell^2} \sqrt{h^2 \frac{\sigma_\ell^2}{n_\ell} + \ell^2 \frac{\sigma_h^2}{n_h}}.$$

We estimate this with the sample means and standard deviations

$$s_{\hat{\theta}} \approx \frac{1}{\bar{y}^2 + \bar{x}^2} \sqrt{\bar{y}^2 \frac{s_x^2}{n_\ell} + \bar{x}^2 \frac{s_y^2}{n_h}} = 0.0030.$$

This gives a γ level z -confidence interval

$$\hat{\theta} \pm z_{(1-\gamma)/2} s_{\hat{\theta}}.$$

For a 95% confidence interval, this is $0.4634 \pm 0.0059 = (0.4575, 0.4693)$ radians or $(26.22^\circ, 26.89^\circ)$.

We can extend the Welch and Satterthwaite method to include the delta method to create a t -interval with effective degrees of freedom

$$\nu = \frac{\left(\frac{\partial g(\bar{x}, \bar{y})^2}{\partial x} \frac{s_1^2}{n_1} + \frac{\partial g(\bar{x}, \bar{y})^2}{\partial y} \frac{s_2^2}{n_2} \right)^2}{\frac{\partial g(\bar{x}, \bar{y})^4}{\partial x} \frac{s_1^4}{n_1^2 \cdot (n_1 - 1)} + \frac{\partial g(\bar{x}, \bar{y})^4}{\partial y} \frac{s_2^4}{n_2^2 \cdot (n_2 - 1)}}.$$

We compute to find that $\nu = 19.4$ and then use the t -interval

$$\hat{\theta} \pm t_{\nu, (1-\gamma)/2} s_{\hat{\theta}}.$$

For a 95% confidence, this is slightly larger interval $0.4634 \pm 0.0063 = (0.4571, 0.4697)$ radians or $(26.19^\circ, 26.91^\circ)$.

Example 16.22 (maximum likelihood estimation). The Fisher information is the main tool used to set an confidence interval for a maximum likelihood estimation. Two choices are typical. Let $\hat{\theta}$ be the maximum likelihood estimate for the parameter θ .

First, we can use the Fisher information $I(\theta)$ and recall that $\hat{\theta}$ is approximately normally distributed, mean θ , standard deviation $1/\sqrt{nI(\theta)}$. Replacing θ by its estimate gives a confidence interval

$$\hat{\theta} \pm z_{\alpha/2} \frac{1}{\sqrt{nI(\hat{\theta})}}.$$

More recently, the more popular method is to use the **observed information** based on the observations $\mathbf{x} = (x_1, x_2, \dots, x_n)$.

$$J(\theta) = -\frac{\partial^2}{\partial \theta^2} \log L(\theta|\mathbf{x}) = -\sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \log f_X(x_i|\theta).$$

This is the second derivative of the score function evaluated at the maximum likelihood estimator. Then, the confidence interval is

$$\hat{\theta} \pm z_{\alpha/2} \frac{1}{\sqrt{J(\hat{\theta})}}.$$

To compare the two approaches, first note that $E_{\theta} J(\theta) = nI(\theta)$, the Fisher information for n observations. Thus, by the law of large numbers,

$$\frac{1}{n} J(\theta) \rightarrow I(\theta) \quad \text{as } n \rightarrow \infty.$$

If the estimator is consistent and I is continuous at θ , then

$$\frac{1}{n} J(\hat{\theta}) \rightarrow I(\theta) \quad \text{as } n \rightarrow \infty.$$

16.2 The Bootstrap

The confidence regions have been determined using aspects of the distribution of the data. In particular, these regions have often been specified by appealing to the central limit theorem and normal approximations. The notion behind **bootstrap** techniques begins with the concession that the information about the source of the data is insufficient to perform the analysis to produce the necessary description of the distribution of the estimator. This is particularly true for small data sets or highly skewed data.

The strategy is to take the data and treat it as if it were the distribution underlying the data and to use a resampling protocol to describe the estimator. For the example above, we estimated the angle in a right triangle by estimating ℓ

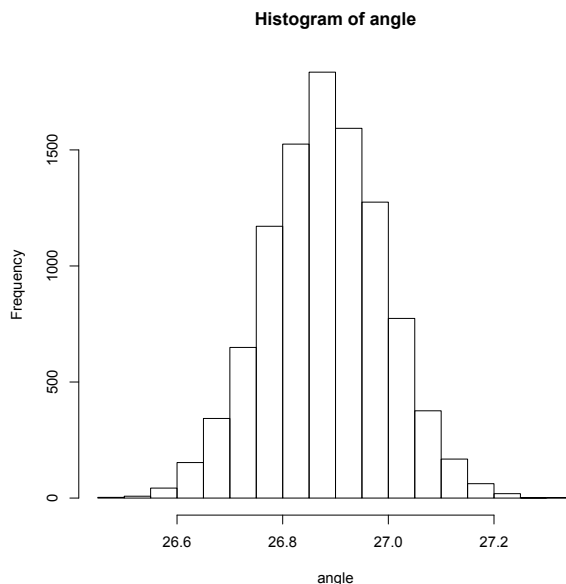


Figure 16.6: Bootstrap distribution of $\hat{\theta}$.

and h , the lengths two adjacent sides by taking the mean of our measurements and then computing the arctangent of the ratio of these means. Using the delta method, our confidence interval was based on a normal approximation of the estimator.

The bootstrap takes another approach. We take the data

$$x_1, x_2, \dots, x_{n_1}, \quad y_1, y_2, \dots, y_{n_2},$$

the **empirical distribution** of the measurements of ℓ and h and act as if it were the **actual distribution**. The next step is the use the data and randomly create the results of the experiment many times over. In the example, we choose, *with replacement* n_1 measurements from the x data and n_2 measurements from the y data. We then compute the bootstrap means

$$\bar{x}_b \quad \text{and} \quad \bar{y}_b$$

and the estimate

$$\hat{\theta}(\bar{x}_b, \bar{y}_b) = \tan^{-1} \left(\frac{\bar{y}_b}{\bar{x}_b} \right).$$

Repeating this many times gives an empirical distribution for the estimator $\hat{\theta}$. This can be accomplish in just a couple lines of R code.

```
> angle<-numeric(10000)
> for (i in 1:10000) {xb<-sample(x,length(x),replace=TRUE);
  yb<-sample(y,length(y),replace=TRUE);angle[i]<-atan(mean(yb)/mean(xb))*180/pi}
> hist(angle)
```

We can use this bootstrap distribution of $\hat{\theta}$ to construct a confidence interval.

```
> q<-c(0.005,0.01,0.025,0.5,0.975,0.99,0.995)
> quantile(angle,q)
  0.5%    1%    2.5%    50%    97.5%    99%    99.5%
26.09837 26.14807 26.21860 26.55387 26.86203 26.91486 26.95065
```

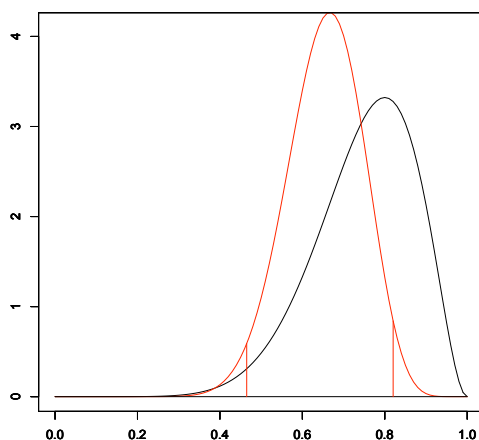


Figure 16.7: Bayesian credible interval. The 95% credible interval based on a $Beta(9, 3)$ prior distribution and data consisting of 8 heads and 6 tails. This interval has probability 0.025 both above and below the end of the interval. Because the density is higher for the upper value, a narrow credible interval can be obtained by shifting the values upward so that the densities are equal at the endpoints of the interval and (16.6) is satisfied.

A 95% confidence interval $(26.21^\circ, 26.86^\circ)$ can be accomplished using the 2.5th percentile as the lower end point and the 97.5th percentile as the upper end point. This confidence interval is very similar to the one obtained using the delta method.

Exercise 16.23. Give the 98% bootstrap confidence interval for the angle in the example above.

Exercise 16.24. Bootstrap confidences are based on a simulation. So, the answer will vary with each simulation. Repeat the bootstrap above and compare.

16.3 Bayesian Statistics

A Bayesian interval estimate is called a **credible interval**. Recall that for the Bayesian approach to statistics, both the data and the parameter are random. Thus, the interval estimate is a statement about the posterior probability of the parameter θ .

$$P\{\tilde{\Theta} \in C(X) | X = x\} = \gamma. \quad (16.6)$$

Here $\tilde{\Theta}$ is the random variable having a distribution equal to the prior probability π . We have choices in defining this interval. For example, we can

- choose the narrowest interval, which involves choosing those values of highest posterior density.
- choosing the interval in which the probability of being below the interval is as likely as being above it.

We can look at this by examining the two examples given in the Introduction to Estimation.

Example 16.25 (coin tosses). *In this example, we began with a beta prior distribution. Consequently, the posterior distribution will also be a member of the beta family of distributions. We then flipped a coin $n = 14$ times with 8 heads. Here, again, is the summary.*

		prior		data		posterior		
α	β	mean	variance	heads	tails	α	β	mean
6	6	1/2	1/52	8	6	14	12	7/13
9	3	3/4	3/208	8	6	17	9	17/26
3	9	1/4	3/208	8	6	11	15	11/26

We use the R command `qbeta` to find the credible interval. For the second case in the table above and with $\gamma = 0.95$, we find values that give the lower and upper 2.5% of the posterior distribution.

```
> qbeta(0.025, 17, 9)
[1] 0.4649993
> qbeta(0.975, 17, 9)
[1] 0.8202832
```

This gives a 95% credible interval of (0.4650, 0.8203). This is indicated in the figure above by the two vertical lines. Thus, the area under the density function from the vertical lines outward totals 5%.

The narrowest credible interval is (0.4737, 0.8276). At these values, the density equals 0.695. The density is lower for more extreme values and higher between these values. The beta distribution has a probability 0.0306 below the lower value for the credible interval and 0.0194 above the upper value satisfying the criterion (16.6) with $\gamma = 0.95$.

Example 16.26. For the example having both a normal prior distribution and normal data, we find that we also have a normal posterior distribution. In particular, if the prior is normal, mean θ_0 , variance $1/\lambda$ and our data has sample mean \bar{x} and each observation has variance 1.

The classical statistics confidence interval

$$\bar{x} \pm z_{\alpha/2} \frac{1}{\sqrt{n}}.$$

For the Bayesian credible interval, note that the posterior distribution is normal with mean

$$\theta_1(\mathbf{x}) = \frac{\lambda}{\lambda + n} \theta_0 + \frac{n}{\lambda + n} \bar{x}.$$

and variance $1/(n + \lambda)$. Thus the credible interval is

$$\theta_1(\mathbf{x}) \pm z_{\alpha/2} \frac{1}{\sqrt{\lambda + n}}.$$

Thus, the center of the interval is influenced by the prior mean. The prior variance results in a narrower interval.

16.4 Answers to Selected Exercises

16.4. Using R to find upper tail probabilities, we find that

```
> qt(0.95, 99)
[1] 1.660391
> qt(0.99, 99)
[1] 2.364606
```

For the 90% confidence interval

$$299,852.4 \pm 1.6604 \frac{79.0}{\sqrt{100}} = 299852.4 \pm 13.1 \quad \text{or the interval } (299839.3, 299865.5).$$

For the 98% confidence interval

$$299,852.4 \pm 2.3646 \frac{79.0}{\sqrt{100}} = 299852.4 \pm 18.7 \quad \text{or the interval } (299833.7, 299871.1).$$

16.8. Let

$$c = \frac{s_1^2/n_1}{s_2^2/n_2}. \quad \text{Then, } \frac{s_2^2}{n_2} = c \frac{s_1^2}{n_1}.$$