4.33 We have

$$E(\hat{\sigma}^2) = \frac{1}{n} E\left(\sum (X_i - \mu)^2\right) = \frac{1}{n} \sum E((X_i - \mu)^2)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \sigma^2 = \frac{1}{n} \cdot n\sigma^2 = \sigma^2.$$

4.35 We need

$$Pr(|\bar{X} - \bar{Y}| \le t^* s/\sqrt{n}) = Pr\left(\frac{|\bar{X} - \bar{Y}|}{s\sqrt{2}/\sqrt{n}} \le t^* \frac{s/\sqrt{n}}{s\sqrt{2}/\sqrt{n}}\right) = Pr\left(|t| \le t^*/\sqrt{2}\right)$$

The command, which depends on n, to produce this probability, can be

```
> pt( qt(0.975, n-1)/sqrt(2), n-1) - pt(qt(0.025, n-1)/sqrt(2), n-1)
```

It might also be

```
> 1 - 2 * pt( qt(0.025, n-1)/sqrt(2), n-1)
```

4.36 (a) We have $\bar{X}_1 - \bar{X}_2$ is normal, centered on $\mu_1 - \mu_2$, with variance

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

$$\Rightarrow \bar{X}_1 - \bar{X}_2 \sim Norm\left(\mu_1 - \mu_2, \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}\right)$$

To obtain a level C confidence interval, find $z_{\alpha/2} = -qnorm((1-C)/2)$.

Then obtain boundaries of the CI via

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

4.39 (a) $|\vec{u}_1|^2 = \left(\frac{1}{\sqrt{n}}\right)^2 + \cdots + \left(\frac{1}{\sqrt{n}}\right)^2 = n \cdot \left(\frac{1}{n}\right) = 1 \quad \Rightarrow |\vec{u}_1| = 1.$

For $i = 2, 3, \cdots, n$,

$$|\vec{u}_i|^2 = \frac{1}{i(i-1)}\left[(i-1)^2 + \sum_{j=1}^{i-1} 1^2\right] = \frac{1}{i(i-1)}\left[(i-1)^2 + (i-1)\right]$$

$$= \frac{i-1}{i(i-1)}\left[(i-1) + 1\right] = \frac{1}{i}(i) = 1 \quad \Rightarrow |\vec{u}_i| = 1.$$

Moreover, for $1 < i < j$,

$$\vec{u}_i \circ \vec{u}_j = 1 - i + \sum_{m=1}^{i-1} 1 = (1-i) + (i-1) = 0.$$

It is more transparent that each $\vec{u}_i \circ \vec{u}_1 = 0$, $i > 1$.

(b) Let $\vec{x} = \langle x_1, x_2, \ldots, x_n \rangle$. Then,

$$\vec{x} \circ \vec{u}_1 = \frac{1}{\sqrt{n}}(x_1 + x_2 + \cdots + x_n) = \sqrt{n} \cdot \frac{1}{n}\sum x_i = \bar{x}\sqrt{n}.$$

(c) With $\vec{x}$ as in part (b),

$$(\vec{x} \circ \vec{u}_1)\vec{u}_1 = (\bar{x}\sqrt{n}) \cdot \frac{1}{\sqrt{n}} \langle 1, 1, \ldots, 1 \rangle = \langle \bar{x}, \bar{x}, \ldots, \bar{x} \rangle.$$

(d) For $\vec{x} = \langle 3, 4, 4, 7, 7 \rangle$, $\bar{x} = \frac{1}{5}(25) = 5$, and

$$\vec{v} = \langle 3, 4, 4, 7, 7 \rangle - \langle 5, 5, 5, 5, 5 \rangle = \langle -2, -1, -1, 2, 2 \rangle.$$

Thus,

i	$\vec{u}_i$	$\vec{x} \circ \vec{u}_i$	$\vec{v} \circ \vec{u}_i$	$\lvert \text{proj}(\vec{x} \to \vec{u}_i) \rvert$	$\lvert \text{proj}(\vec{v} \to \vec{u}_i) \rvert$
1	$1/\sqrt{5}\langle 1, 1, 1, 1, 1 \rangle$	$5^{3/2}$	0	$5^{3/2}$	0
2	$1/\sqrt{2}\langle 1, -1, 0, 0, 0 \rangle$	$-1/\sqrt{2}$	$-1/\sqrt{2}$	$1/\sqrt{2}$	$1/\sqrt{2}$
3	$1/\sqrt{6}\langle 1, 1, -2, 0, 0 \rangle$	$-1/\sqrt{6}$	$-1/\sqrt{6}$	$1/\sqrt{6}$	$1/\sqrt{6}$
4	$1/\sqrt{12}\langle 1, 1, 1, -3, 0 \rangle$	$-10/\sqrt{12}$	$-10/\sqrt{12}$	$10/\sqrt{12}$	$10/\sqrt{12}$
5	$1/\sqrt{20}\langle 1, 1, 1, 1, -4 \rangle$	$-10/\sqrt{20}$	$-10/\sqrt{20}$	$10/\sqrt{20}$	$10/\sqrt{20}$

For $i \geq 2$, $\vec{x} \circ \vec{u}_i = \vec{v} \circ \vec{u}_i$. This is true for other $\vec{x} \in \mathbb{R}^5$, too, since

$$\vec{v} \circ \vec{u}_i = (\vec{x} - \overrightarrow{\bar{x}}) \circ \vec{u}_i = \vec{x} \circ \vec{u}_i - \overrightarrow{\bar{x}} \circ \vec{u}_i = \vec{x} \circ \vec{u}_i - 0$$

since $\overrightarrow{\bar{x}}$, being parallel to $\vec{u}_1$, is orthogonal to each $\vec{u}_i$ with $i \geq 2$.

4.47 (a) For a 1-sided level C confidence interval of the form $(-\infty, L)$ when σ is unknown, take t^* as the output from

```
> qt( C/100, df = n-1 )
```

and then set $L = \bar{x} + t^* \cdot \frac{s}{\sqrt{n}}$.

4.48 The margin of error is $z^* \sigma/\sqrt{n}$. If we want this to be at most $\frac{1}{4}$, then solving

$$\frac{1}{4} \geq z^* \frac{\sigma}{\sqrt{n}} \quad \Rightarrow \quad n \geq (4 z^* \sigma)^2 = (7.84\,\sigma)^2.$$

(a) The lower bound on sample size n increases with σ. Assuming we cannot guess σ perfectly, it is better to estimate it on the high side.

(b) If it does not thwart your purposes to obtain gender-specific confidence intervals, then it should be easier to estimate σ for a single sex. The two sexes very likely have different distributions.

(c) Using 2 for σ in the formula from (a),

$$n \geq [(7.84)(2)]^2 = 245.86.$$

Since sample size must be an integer, take $n \geq 246$.

4.50 (a) Using commands

```
> mySt = favstats( ~ weight | feed, data = chickwts )
> with( mySt, mean - qt(0.975, df = (n-1)) * sd / sqrt(n))
> with( mySt, mean + qt(0.975, df = (n-1)) * sd / sqrt(n))
```

I obtain lists of lower/upper bounds for the 6 confidence intervals:

feed	95% confidence interval
casein	(282.64, 364.52)
horsebean	(132.57, 187.83)
linseed	(185.56, 251.94)
meatmeal	(233.31, 320.51)
soybean	(215.18, 277.68)
sunflower	(297.89, 359.95)

(b) With no overlap between intervals for horsebean and soybean, and again between soybean and casein, it does seem reasonable to believe a real difference exists between population mean weights across these three feeds.

(c) It may not be appropriate to use 1-sample t methods, given the sample sizes range from 10 to 14. As weights seem to depend on many biological factors, it may be appropriate to assume underlying populations that are normal, in which case sample sizes are not an issue.

4.58 (a) The command

```
> t.test( ~ (vitamin - placebo), data = Endurance )
```

reveals that the average difference is negative (about 48 fewer repetitions to get fatigued when receiving the vitamin), but this difference is not statistically significant, having a P-value of 0.4553.

(b) For

```
> t.test( ~ (log(vitamin) - log(placebo)), data = Endurance )
```

the P-value is 0.07868.

(c) The attempt

> t.test(~(vitamin / placebo), data = Endurance)

gives P-value 6×10^{-6}, quite statistically significant.

(d) Here

> t.test(~(1/vitamin − 1/placebo), data = Endurance)

gives P-value 0.03022, also statistically significant.

(e) The commands

> nrow(filter(Endurance, vitamin - placebo == 0))

> nrow(filter(Endurance, vitamin - placebo < 0))

show there are no ties, and in 11 of 15 cases, the endurance level was less under the vitamin treatment. If π represents the proportion of times vitamin endurance is less, we test

$$H_0: \pi = 0.5 \quad \text{vs.} \quad H_a: \pi \neq 0.5$$

with the sign test

> binom.test(11, 15)

The P-value is 0.1185, so we cannot reject H_0.

4.59 (a) The text suggests that Wilson and Plus4 are the same, but

> help(binom.test)

shows Wilson and Score to be the same thing. That is borne out by the results:

command	interval
binom.test(115, 200, ci.method = "Wald")	(0.50649, 0.64351)
binom.test(115, 200, ci.method = "Score")	(0.50571, 0.64146)
binom.test(115, 200, ci.method = "Wilson")	(0.50571, 0.64146)
binom.test(115, 200, ci.method = "Plus4")	(0.50566, 0.64140)

Same

(b)

command	interval
prop.test(115, 200, correct = TRUE)	(0.50321, 0.64386)
prop.test(115, 200, correct = FALSE)	(0.50571, 0.64146)

The second of these matches the Score method above.