

$$4.13 \text{ (a)} \quad E(\bar{X}_w) = E\left(\sum_{i=1}^n w_i X_i\right) = \sum_{i=1}^n w_i E(X_i)$$

$$= \mu \sum w_i$$

Since we want  $E(\bar{X}_w) = \mu$ , we must have  $\sum_{i=1}^n w_i = 1$ .

$$\text{(b)} \quad \text{Var}(\bar{X}_w) = \text{Var}\left(\sum w_i X_i\right) = \sum \text{Var}(w_i X_i)$$

$$= \sum w_i^2 \text{Var}(X_i) = \sigma^2 \sum w_i^2 \quad \left(= |\vec{w}|^2 \sigma^2\right)$$

4.15 (a) Using a probability tree

$$\begin{aligned} p &= \Pr(\text{person answers "True"}) \\ &= \pi\theta + (1-\pi)(1-\theta) \\ &= (2\pi - 1)\theta + (1-\pi) \end{aligned}$$

$$\text{(b) From the formula in (a), } \theta = \frac{p + \pi - 1}{2\pi - 1}.$$

(c) In an i.i.d. sample, the number of "true" responses  $X \sim \text{Binom}(n, p)$ .

We have seen that  $\hat{p}$  is a mean of  $n$  Bernoulli r.v.s, and that its expected value, both in the case of i.i.d. and SRS, is  $p$ , the true probability of a person responding "true".

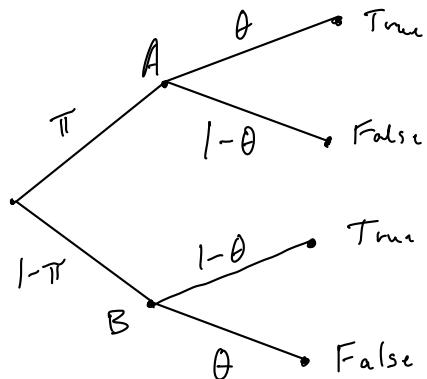
(d) We propose the estimator  $\hat{\theta} = \frac{\hat{p} + \pi - 1}{2\pi - 1}$ . Then

$$\begin{aligned} E(\hat{\theta}) &= E\left(\frac{1}{2\pi-1} \hat{p} + \frac{\pi-1}{2\pi-1}\right) = \frac{1}{2\pi-1} E(\hat{p}) + \frac{\pi-1}{2\pi-1} \\ &= \frac{1}{2\pi-1} p + \frac{\pi-1}{2\pi-1} = \theta, \end{aligned}$$

making  $\hat{\theta}$  an unbiased estimator of  $\theta$ .

(e) We have  $X \sim \text{Binom}(n, p)$ , representing the number of "true" responses, has variance  $\text{Var}(X) = np(1-p)$ . Thus,  $\hat{p} = X/n$ , has

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



(f) As a result of (e),

$$\text{Var}(\hat{\theta}) = \frac{1}{(2\pi - 1)^2} \text{Var}(\hat{p}) = \frac{p(1-p)}{n(2\pi - 1)^2}$$

4.27 (a) There are several ways to attack this one. Under the null hypothesis

$$H_0: \pi = 0.95,$$

the count  $X$  of successes in 10000 runs has a  $\text{Binom}(10000, 0.95)$  distribution. The command to get 0.025- and 0.975-quantiles is

$$\text{qbinom}(c(0.025, 0.975), 10000, 0.95)$$

yielding 9457 and 9542, but

$$\text{pbisnom}(c(9456, 9542), 10000, 0.95)$$

vs.

$$\text{pbisnom}(c(9457, 9541), 10000, 0.95)$$

shows the former to be at the edges of the rejection region. That is, he should reject  $H_0$  if his coverage rate is

0.9456 or lower, or

0.9542 or higher.

This approach, however, does not take advantage of the Central Limit Theorem, which we may do since

$$np = (10000)(0.95) > 10 \quad \text{and}$$

$$n(1-p) = (10000)(0.05) > 10.$$

Under  $H_0$ ,  $\hat{\pi} \sim \text{Norm}(0.95, \sqrt{(0.95)(0.05)/10000})$ , approximately. Solving

$$\frac{\hat{\pi} - 0.95}{\sqrt{(0.95)(0.05)/10000}} \geq 1.96 \Rightarrow \hat{\pi} \geq 0.9543$$

$$\frac{\hat{\pi} - 0.95}{\sqrt{(0.95)(0.05)/10000}} \leq -1.96 \Rightarrow \hat{\pi} \leq 0.9457$$

approximately the same values as above.

4.28 (a) The sample mean is in the center of the confidence interval:

$$\bar{x} = \frac{1}{2}(11.2 + 54.7) = 32.95$$

(b) The hypothesized value for "no weight gain,"  $D$ , is outside the 95% confidence interval. That means that, under  $H_0: \mu = D$ , with  $\alpha = 0.05$ ,  $\bar{x} = 32.95$  is in the rejection region — so the P-value is  $< 0.05$ .

4.29 This is completely wrong. The 95% CI changes with the sample, and shrinks in width as the sample size  $n$  grows. If 10 different studies offered 10 different 95% CIs, your biologist friend would be contradicting himself making such a statement about each of the CIs.

4.32. Given that  $E(\hat{\theta}^2) = \theta^2$ , we have

$$\begin{aligned}\text{Var}(\hat{\theta}) &= E(\hat{\theta}^2) - E(\hat{\theta})^2 \\ \Rightarrow E(\hat{\theta}) &= \sqrt{E(\hat{\theta}^2) - \text{Var}(\hat{\theta})} = \sqrt{\hat{\theta}^2 - \text{Var}(\hat{\theta})},\end{aligned}$$

and this latter expression is  $\hat{\theta}$  only when  $\text{Var}(\hat{\theta}) = 0$  (an uninteresting case).

4.33 We have

$$\begin{aligned}E(\hat{\sigma}^2) &= \frac{1}{n} E\left(\sum (X_i - \mu)^2\right) = \frac{1}{n} \sum E((X_i - \mu)^2) \\ &= \frac{1}{n} \sum_{i=1}^n \sigma^2 = \frac{1}{n} \cdot n\sigma^2 = \sigma^2.\end{aligned}$$

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

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

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

4.48 The margin of error is  $z^* \sigma / \sqrt{n}$ . If we want this to be at most  $\frac{1}{4}$ , then solving  $\frac{1}{4} \geq z^* \frac{\sigma}{\sqrt{n}}$   $\Rightarrow n \geq (4z^* \sigma)^2 = (7.84 \sigma)^2$ .

(a) The lower bound on sample size  $n$  increases with  $\sigma$ . Assuming we cannot guess  $\sigma$  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  $\sigma$  for a single sex. The two sexes very likely have different distributions.

(c) Using 2 for  $\sigma$  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  $6 \times 10^{-6}$ , quite statistically significant.

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

> help(binom.test)

shows Wilson and Score to be implemented the same way, 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)

(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.

➤ same