2.8 Exercises

2.8.1. Verify, via simulation, the level of the wilcox.test when sampling from a standard normal distribution. Use n = 30 and levels of $\alpha = 0.1, 0.05, 0.01$. Based on the resulting estimate of α, the empirical level, obtain a 95% confidence interval for α.

2.8.2. Redo Exercise 2.8.1 for a t-distribution using 1,2,3,5,10 degrees of freedom.

2.8.3. Redo Example 2.4.1 without a for loop and using the apply function.

2.8.4. Redo Example 2.3.2 without a for loop and using the apply function.

2.8.5. Suppose in a poll of 500 registered voters, 269 responded that they would vote for candidate P. Obtain a 90% percentile bootstrap confidence interval for the true proportion of registered voters who plan to vote for P.

2.8.6. For Example 2.3.1 obtain a 90% two-sided confidence interval for the treatment effect.

2.8.7. Write an R function which computes the sign analysis. For example, the following commands compute the statistic S+, assuming that the sample is in the vector x.

```
xt <- x[x!=0]; nt <- length(xt); ind <- rep(0,nt);
ind[xt > 0] <-1; splus <- sum(ind)
```

2.8.8. Calculate the sign test for the nursery school example, Example 2.3.1. Show that the p-value for the one-sided sign test is 0.1445.

2.8.9. The data for the nursery school study were drawn from page 79 of Siegel (1956). In the data table, there is an obvious typographical error. In the 8th set of twins, the score for the the twin that stayed at home is typed as 82 when it should be 62. Rerun the signed-rank Wilcoxon and t-analyses using the typographical error value of 82.

2.8.10. The contaminated normal distribution is frequently used in simulation studies. A standardized variable, X, having this distribution can be written as

$$X = (1 - I_\varepsilon)Z + cI_\varepsilon Z,$$

where $0 \le \varepsilon < 1$, I_ε has a binomial distribution with n = 1 and probability of success ε, Z has a standard normal distribution, c > 1, and I_ε and Z are independent random variables. When sampling from the distribution of X, $(1-\varepsilon)100\%$ of the time the observations are drawn from a N(0, 1) distribution but ε 100% of the time the observations are drawn from a $N(0, c^2)$. These later observations are often outliers. The distribution of X is a mixture distribution; see, for example, Section 3.4.1 of Hogg et al. (2013). We say that X has a $CN(c, \varepsilon)$ distribution.

1. Using the R functions rbinom and rnorm, write an R function which obtains a random sample of size n from a contaminated normal distribution $CN(c, \varepsilon)$.
2. Obtain samples of size 100 from a N(0, 1) distribution and a CN(16, 0.25) distribution. Form histograms and comparison boxplots of the samples. Discuss the results.

2.8.11. Perform the simulation study of Example 2.3.2 when the population has a CN(16, 0.25) distribution. For the alternatives, select values of θ so the spread in empirical powers of the signed-rank Wilcoxon test ranges from approximately 0.05 to 0.90.

2.8.12. The ratio of the expected squared lengths of confidence intervals is a measure of efficiency between two estimators. Based on a simulation of size 10,000, estimate this ratio between the Hodges–Lehmann and the sample mean for $n = 30$ when the population has a standard normal distribution. Use 95% confidence intervals. Repeat the study when the population has a t-distribution with 2 degrees of freedom.

2.8.14. Let p be the probability of success. Suppose it is of interest to test

$$H_0 : p = 0.30 \text{ versus } H_A : p < 0.30.$$

Let S be the number of successes out of 75 trials. Suppose we reject H_0, if $S \leq 16$.

(a) Determine the significance level of the test.

(b) Determine the power of the test if the true p is 0.25.

(c) Determine the power function for the test for the sequence for the probabilities of success in the set {0.02, 0.03, . . ., 0.35}. Then obtain a plot of the power curve.