StatsI — Exercise 5

1. Let $X_1, \cdots, X_n$ be a sample from Bernoulli(p), and $Y_1, \cdots, Y_m$ be a sample from Bernoulli(q), and the two samples are independent with each other.

   (a) Find a reasonable estimator for $p - q$ and its standard error.

   (b) Find an approximate 95% confidence interval for $p - q$ when both n and m are large.

2. 100 people are given a standard antibiotic to treat an infection and another 100 are given a new antibiotic. In the first group, 90 people recover; in the second group, 85 people recover. Let p_1 be the probability of recovery with the standard antibiotic and p_2 be the probability of recovery with the new antibiotic. We are interested in estimating $\theta = p_1 - p_2$. Provide an estimate, standard error, an 80% confidence interval, and a 95% confidence interval for θ.

3. Let $X_1, \cdots, X_n$ be a sample from a distribution F and let $\widehat{F}_n$ be the empirical distribution. Let x and y be two distinct points, and $x < y$.

   (a) Compute $E\{\widehat{F}(x)\}$ and $\mathrm{Var}\{\widehat{F}(x)\}$.

   (b) Compute $\mathrm{Cov}\{\widehat{F}(x),\ \widehat{F}(y)\}$.

   (c) Based on the CLT, find the limiting distribution of $\widehat{F}(x)$. Furthermore, find an approximate $1 - \alpha$ confidence interval for $F(x)$.

   (d) Suppose we are interested in estimating $\theta \equiv F(y) - F(x)$, and we use $\widehat{\theta} \equiv \widehat{F}(y) - \widehat{F}(x)$ as our estimate. Find the standard deviation and the standard error of $\widehat{\theta}$.

4. Let $X_1, \cdots, X_n$ be a sample from $N(\mu, 1)$. Let $\theta = e^{\mu}$, and we estimate θ by $\widehat{\theta} = e^{\bar{X}}$, where $\bar{X} = n^{-1} \sum_{1 \le i \le n} X_i$.

   (a) State how to obtain a bootstrap estimator v^* for $v \equiv \mathrm{Var}(\widehat{\theta})$.

   (b) Draw a sample of size $n = 400$ from $N(5, 1)$ in R. Compute v^* with the bootstrap replication $B = 2000$ times. Plot the histogram of the bootstrap estimates $\theta_1^*, \cdots, \theta_B^*$. This is a bootstrap estimator for the distribution of $\widehat{\theta}$.

   (c) Draw 2000 samples (with fixed size $n = 400$) from $N(5, 1)$. For each sample, compute the estimate $\widehat{\theta}_i$, $i = 1, \cdots, 2000$. Draw histogram of those $\widehat{\theta}_i$. This may be regarded as the true distribution of $\widehat{\theta}$. Compare it with its bootstrap estimator obtained in (b) above.

5. Let $Y_1, \cdots, Y_n$ be a sample from $N(\mu, \sigma^2)$. Let $\tau = \mu/\sigma$. We estimate τ by $\widehat{\tau} = \bar{Y}/S$, where $\bar{Y}$ and S are, respectively, the sample mean and the sample standard deviation.

   (a) Based on the estimator $\widehat{\tau}$, construct the three types of 95% bootstrap confidence intervals for τ.

   (b) Set $n = 100$ and $B = 200$, repeat the estimation 500 times, and check the relative frequencies of the three intervals covering the true value of τ when $\mu = 1$ and $\sigma^2 = 1$.