Confidence Intervals

1. Example (variance known). The fat content of a hotdog is assumed to follow $N(\mu, 4^2)$. We randomly select 10 hotdogs:

   ```r
   > x<-c(25.2, 21.3, 22.8, 17.0, 29.8, 21.0, 25.5, 16.0, 20.9, 19.5)
   > mean(x)
   [1] 21.9
   ```

   We are interested in estimating the mean fat content. We can estimate $\mu$ with the sample mean $\bar{X} = 21.9$. The point estimate does not provide the precision of the estimation. Alternately, we will report an entire interval of possible values $(\hat{\mu}_L, \hat{\mu}_U)$ that contains possible values of $\mu$.

2. A confidence interval (CI) is an interval used to estimate the likely size of a population parameter. A confidence level is a measure of the degree of reliability of the confidence interval. Most commonly used confidence levels are the 90%, 95% and 99% confidence intervals that have 0.90, 0.95 and 0.99 probabilities respectively of containing the parameter. For population parameter $\mu$, 95% confidence interval $(\hat{\mu}_L, \hat{\mu}_U)$ of $\mu$ is an interval that satisfies

   $$P(\hat{\mu}_L \leq \mu \leq \hat{\mu}_U) = 0.95.$$ 

   We usually make the interval centered so that

   $$P(\hat{\mu}_L \leq \mu \leq \hat{\mu}_U) = P(\mu \leq \hat{\mu}_U) - P(\mu \geq \hat{\mu}_L) = 0.95.$$ 

3. If $X_i \sim N(\mu, \sigma^2)$ with known $\sigma^2$ and unknown $\mu$. 100(1 $-$ $\alpha$)% confidence interval for $\mu$ is,

   $$\hat{\mu}_L = \bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \hspace{1cm} \hat{\mu}_U = \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},$$

   where quantile $z_{\alpha/2}$ is given by $P(Z > z_{\alpha/2}) = \alpha/2$. For instance, a 95% confidence interval for $\mu$ is

   $$\hat{\mu}_L = \bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \hspace{1cm} \hat{\mu}_U = \bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}.$$

4. A 95% confidence interval can be interpreted probabilistically as an interval that can contain true unknown parameter 95% of time in the long run.

5. Example (variance unknown). Reconsider the hotdog example. Assume the fat content of a hotdog is assumed to follow $N(\mu, \sigma^2)$ with unknown $\sigma$. We are interested in constructing a 95% confidence interval for $\mu$.

6. $T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1},$

   a t distribution with $n - 1$ degrees of freedom. As $n \to \infty$, $t_n \to N(0, 1)$. $t$ distributions have thicker tails than the standard normal distribution. For large sample size $n$, we can approximate $t$ distribution with $N(0, 1)$.

7. Critical value $t_{\alpha,n-1}$ is defined as the point that gives $P(T > t_{\alpha,n-1}) = \alpha$. We can construct CI based on

   $$P(-t_{\alpha/2,n-1} < T < t_{\alpha/2,n-1}) = 1 - \alpha.$$ 

   100(1 $-$ $\alpha$)% CI for $\mu$ is

   $$\bar{x} \pm t_{\alpha/2} s \sqrt{n}.$$

   ```r
   > qt(1-0.025,9)
   [1] 2.262157
   > sd(x)
   [1] 4.13414
   ```

8. General large sample confidence interval. Suppose $\hat{\theta}$ is an unbiased estimator of some parameter $\theta$, Then 100(1 $-$ $\alpha$)% confidence interval is

   $$\hat{\theta} \pm z_{\alpha/2} \sqrt{\frac{\hat{\theta}}{n}}.$$
In many applications, $\hat{V}^\theta$ is a function of $\theta$ which makes computation of CI complicated. In this situation, we need to estimate $\hat{V}^\theta$ further.

The following is a special case. Let $X_1, \ldots, X_n$ be a random sample with mean $\mu$. For sufficiently large $n$,

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim N(0,1)$$

where $S$ is the sample standard deviation. If $n$ is sufficiently large, approximate $100(1 - \alpha)\%$ confidence interval for $\mu$ is

$$\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}},$$

where $s$ is the sample standard deviation.

**Problem**

A person tossed $n = 100$ biased coins with $P(H) = p$ and observed 38 heads. Construct a 95% CI for $p$.

```r
> X<-rbinom(100,1,0.4)
> X
[1] 1 0 0 0 0 0 0 1 0 1 1 0 0 0 0
[17] 0 0 1 1 0 1 0 1 1 0 1 0 1 0 1 0
[33] 1 1 0 0 0 1 0 1 0 0 0 0 0 1 0
[49] 0 1 0 0 0 0 1 1 0 1 0 1 1 0 0 1
[65] 0 0 1 0 0 1 0 0 0 1 0 0 1 0 0 0
[81] 1 0 1 0 0 1 1 0 0 0 1 0 0 1 1 1
[97] 0 1 1 0
> sqrt(0.38*(1-0.38)/100)*1.96
[1] 0.09513574
> 0.38+0.095
[1] 0.475
> 0.38-0.095
[1] 0.285
```

**Assigned problems.** Exercise 7.6, part (a) only, 7.20, 7.32. HW 5 due April 14th. Solve the assigned problems in Lecture 15-18.