Problem: 2, 7, 8, 9, 10, 11, 12.

Problem 2.

20% of the population is "1".

(the true value of the population proportion is known)

Then, the standard error of $\hat{p}$ is

$$se(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}.$$

(this is the formula for p known).

$\hat{p}$: the estimator of p.

p: the population proportion (known, $p = 0.2$)

n: sample size. ($n = 60$)

- Note: if p is unknown, we use

$$\hat{se}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

In this problem, we have $se(\hat{p}) = \sqrt{\frac{0.2 \times (1-0.2)}{60}}$

$$= 0.0516$$

This formula does not depend on N.

$\Rightarrow$ "No".

Comment: "independent" means we should put back the ticket after each draw.

Problem 7.

Unbiased estimator ?

- An estimator $\hat{p}$ is unbiased, iff $E[\hat{p}] = p$.

In this problem, we denote the (population) proportion of chocolate chips in the jar by p.

$\Rightarrow$ We know that, the estimator

$$\hat{p} = \begin{cases} 1, & \text{if the draw is a chocolate chip;} \\ 0, & \text{otherwise.} \end{cases}$$

$$\begin{aligned} &\mathbb{P}(\hat{p} = 1) \\ &= \mathbb{P}(\text{the draw is a chocolate chip}) \\ &= p. \end{aligned}$$

$$\begin{aligned} E[\hat{p}] &= 1 \cdot (p) + 0 \cdot (1-p) \\ &= p. \end{aligned}$$

So $\hat{p}$ is unbiased.

Problem 8.

- Consistency: when sample size goes to ∞, the estimator converges to the population value (true value)

In this problem, the sample size is always 1, which is fixed. We could not use the term "consistent" on this estimator.

Modification of this problem:

Suppose we draw n cookies from the jar. If the first draw gives a chocolate chip, our estimate is 1; otherwise 0.

- In this case, we could claim that the estimator is inconsistent. Because it could only take value of 0 or 1, it could not converge to the true value.

Problem 9.

Suppose that the true proportion is p.

We need

$$\mathbb{P}\left(|\hat{p}-p| \le 2.5 \times \hat{se}(\hat{p})\right)$$

Note: the true proportion is a fixed number,

So, "... within $\underbrace{\text{2.5 SE}}_{\text{the SE / estimated SE of } \hat{p}.}$ $\underbrace{\text{of the true proportion}}_{p \pm \text{the SE of } \hat{p}.}$"

- $\hat{p}$ is approximately normal (when the sample size is large). $p = E[\hat{p}]$.

$$\mathbb{P}\left(|\hat{p}-p| \le 2.5 \times \hat{se}(\hat{p})\right)$$

$$= \mathbb{P}\left(\left|\frac{\hat{p}-p}{\hat{se}(\hat{p})}\right| \le 2.5\right)$$

$\frac{\hat{p}-p}{\hat{se}(\hat{p})} \overset{\text{approx.}}{\sim} N(0,1)$.

$$= \mathbb{P}(|Z| \le 2.5), \quad Z \sim N(0,1).$$

$$= 0.9876.$$

(R code: pnorm(2.5) − pnorm(−2.5))

Problem 10.

We need $\mathbb{P}\left(|p-\hat{p}| \le 2.5 \times \hat{se}(\hat{p})\right)$,

which is identical to the result in problem 9.

Problem 11.

CI: (for large same size) for confidence level $(1-\alpha)$,

$$CI = \left[\hat{p} - z_{1-\alpha/2} \cdot \hat{se}(\hat{p}),\ \hat{p} + z_{1-\alpha/2} \cdot \hat{se}(\hat{p})\right]$$

$$= \hat{p} \pm z_{1-\alpha/2} \cdot \hat{se}(\hat{p})$$

- $\hat{p} = m/n$.
- $\hat{se}(\hat{p}) = \sqrt{\hat{p}(1-\hat{p})/n}$
- $z_{1-\alpha/2}$ is the z-score (i.e. quantile of standard normal), which could be calculated by R using qnorm$(1-\alpha/2)$.

$\hat{p} = 13/40 = 0.325$

$\hat{se}(\hat{p}) = \sqrt{0.325 \cdot (1-0.325)/40}$

$= 0.074$

α: 5%, if we want 95% CI.

10%, if we want 90% CI.

$z_{0.975} = 1.96$

$z_{0.95} = 1.645$.

Plug in and we get the final result.

Problem 12.

- False
- The true proportion is fixed.
  We can not simply talk about the distribution of the "true value".
- When we do the sampling process many times, the prob. that the CI contains the true value is 90%. (v)
  - CI is a random set
    But the true p is fixed!!