The University of Sydney
School of Mathematics and Statistics

STUVAC Statistical Snacks

MATH1062/MATH1005: Mathematics 1B/Statistical Thinking With Data — Semester 1, 2024

Lecturers: J. Baine, T. Cui, J. Spreer, M. Stewart

1. 0-1 Box (specific example) A box contains 10 tickets. 3 are $\boxed{1}$ and 7 are $\boxed{0}$. In the questions below, if necessary, round to 3 decimal places.

(a) What is the mean μ and SD σ of the box (that is, of the list of numbers represented on the tickets in the box)?

(b) Suppose $n = 100$ tickets are drawn randomly, with replacement, yielding numbers $X_1, \ldots, X_n$. Write $S = X_1 + \ldots + X_n$ for the sum of the draws and $\bar{X} = S/n$ for the average of the draws.

(i) What is $E(X_1)$?

(ii) What is $SE(X_1)$?

(iii) What is $E(X_1 + X_2)$?

(iv) What is $SE(X_1 + X_2)$?

(v) What is $E(S)$?

(vi) What is $SE(S)$?

(vii) What is $E(\bar{X})$?

(viii) What is $SE(\bar{X})$?

(c) By appealing to the Central Limit Theorem, determine a value v such that the interval $0.3 \pm v$, i.e. $[0.3 - v, 0.3 + v]$, serves as an (approximate) 98% prediction interval for $\bar{X}$. In other words, find v such that

$$P\left\{0.3 - v \le \bar{X} \le 0.3 + v\right\} \approx 0.98\,.$$

The R output below may be useful for this.

```
qnorm(0.95)

## [1] 1.644854

qnorm(0.975)

## [1] 1.959964

qnorm(0.98)

## [1] 2.053749

qnorm(0.99)

## [1] 2.326348
```

2. 0-1 Box (general case) Repeat question 1, but for a box with N tickets: pN are $\boxed{1}$ and $(1-p)N$ are $\boxed{0}$. Write out answers to the questions below in terms of general sample size n and proportion of $\boxed{1}$s p.

(a) What is the mean μ and SD σ of the box (that is, of the list of numbers represented on the tickets in the box)?

(b) Suppose n tickets are drawn randomly, with replacement, yielding numbers $X_1, \ldots, X_n$. Write $S = X_1 + \ldots + X_n$ for the sum of the draws and $\bar{X} = S/n$ for the average of the draws. You may assume that n is large enough that the Central Limit Theorem applies.

(i) What is $E(X_1)$?

(ii) What is $SE(X_1)$?

(iii) What is $E(X_1 + X_2)$?

(iv) What is $SE(X_1 + X_2)$?

(v) What is $E(S)$?

(vi) What is $SE(S)$?

(vii) What is $E(\bar{X})$?

(viii) What is $SE(\bar{X})$?

(c) By appealling to the Central Limit Theorem, determine a value v such that the interval $p \pm v$, i.e. $[p - v, p + v]$, serves as an (approximate) 98% prediction interval for $\bar{X}$. In other words, find v such that

$$P\left\{p - v \leq \bar{X} \leq p + v\right\} \approx 0.98\,.$$

The R output below part question 1 (c) may be useful for this.