ST2132 Tutorial 5 Solution 2210

1. (a) By the CLT, for large n, the distribution of

$$\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}\sigma} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

is close to N(0,1) $\sim Z$. The statement follows from $\Pr(-z_{\frac{\alpha}{2}} \le Z \le z_{\frac{\alpha}{2}}) = 1 - \alpha$, replacing Z by $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ and the equal sign by $\approx$.

(b) Do algebra on the event in (a), taking one inequality at a time.

(c) With $\mu = p$, $\sigma^2 = p(1-p)$, $\bar{X} = \hat{p}$, we get

$$\Pr\left(\hat{p} - z_{\frac{\alpha}{2}}\frac{\sqrt{p(1-p}}{\sqrt{n}} \le \mu \le \hat{p} + z_{\frac{\alpha}{2}}\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) \approx 1 - \alpha$$

2. (a) The parameter is p, the proportion of households in the town that had internet access.

(b) p is estimated as $239/500 \approx 0.48$. The SE is $\mathrm{SD}(\bar{X}) = \sqrt{p(1-p)}/\sqrt{500}$, estimated as $\sqrt{0.48 \times 0.52}/\sqrt{500} \approx 0.02$. 95%-CI is $0.48 \pm 1.96 \times 0.02 \approx (0.44, 052)$.

(c) 500 trials, success probability p.

3. (a) Something is wrong. The expectation μ and SD σ are unknown. If they are known, there is nothing to estimate. Amended: "Assume the 25 readings are realisations of IID RV's $X_1, \ldots, X_{25}$ with mean μ and SD σ. Both μ and σ are unknown."

(b) $\bar{x} = 81,411$ is an estimate of μ. The SE is estimated as $30/\sqrt{25} = 6$. The statement is true.

(c) False. The average of the 25 readings is known, so there is no need to make a CI for it.

(d) False. The parameter is fixed, not random. It is either in the range, or not in the range.

(e) This means that the data look like coming from a random experiment, indicating that it is reasonable to treat them as realisations of IID RV's, i.e., the randomness assumption is reasonable.

4. (a) Yes. The interval is a CI for μ with $z_{\frac{\alpha}{2}} = 1$. In R, pnorm(-1) gives $\alpha/2 \approx 0.16$. Thus, the confidence level is $1 - \alpha \approx 0.68$. This is roughly the proportion of intervals containing μ.

[The above uses the large-sample CI for μ. Since the X's are normal, we can also use the exact CI, i.e., $t_{\frac{\alpha}{2},99} = 1$. Then pt$(-1, 99)$ gives almost the same answer.]

(b) An interval can be constructed as follows:

```
x = rnorm(100)
xbar = mean(x)
SE = sd(x)/sqrt(100)
ci = xbar + SE*c(-1,1)
```

The interval contains 0 exactly if the two ends have opposite sign. To see if the interval contains 0:

```
ci[1]*ci[2] < 0
```

The above can be included in a for loop. But more efficiently, we can use vector operations.

```
data = matrix(rnorm(100000*100), 10000, 100) # 10000 data sets
xbar = apply(data, 1, mean) # estimates of 0
SE = apply(data, 1, sd) / 10 # estimated SE
ci_l = xbar - SE # lower endpoints
ci_u = xbar + SE # upper endpoints
v = ci_l*ci_u < 0 # contains 0?
mean(v) # proportion of intervals containing 0
```