Practice Exam 2 Solutions

Question 1

Part A

Given our method for constructing a range of plausible values based on a sample stastistic and variability, we expect our resulting interval to contain the true parameter 95% of the time.

Part B

Population is the collection of objects that we are interested in studying or making statements about. A value associated with the population that we are interested in is called a *parameter*. Instead of an entire population, we collect a sample which we hope to be representative of the population. The value associated with the sample that we estimate is called a statistic.

A study which involves the entire population is known as a census.

Part C

Distribution tells us what values something takes and how frequently we expect them to appear. A *sampling distribution* is a distribution on the values of a sample statistic. The sampling distribution gives me an idea of the plausible range of values a sample statistic might make. The variability of a sampling distribution comes from two things: first, the variability that we find in the population (σ) and the number of observations in our sample (n)

Part D

$$t = \frac{\overline{x} - \mu}{\hat{\sigma}/\sqrt{n}}$$

- As the difference between $\overline{x}$ and μ increases, our t statistic will get larger
- As $\hat{\sigma}$ increases, the t statistic will decrease and vice versa
- As n increases, our t-statistic will also increase

Part E

$$\overline{x} \pm C \left(\frac{\hat{\sigma}}{\sqrt{n}} \right)$$

- Changing $\overline{x}$ will shift the interval but will not change the size of the interval
- Changing calibration parameter C – making C larger will make the interval larger, reducing the type I error rate
- Increasing $\hat{\sigma}$ will increase CI
- Decreasing n will also increase CI

Part F

The CLT says that for larger enough n (typically 20-30), the sampling distribution of the sample mean will be approximately normal. Therefore, since the 68-95-99 rule holds for normal distributoins, it should hold for distribution of the sample mean.

Question 2

Part A

```r
# point estimate of CR
10528 / 6431
```

```
## [1] 1.6371
```

Part B

No, the sampling distribution is not normally distributed so the assumptions needed for 68-95-99 rule do not hold.

Part C

A 80% confidence interval would be (1.09, 1.89)

Part D

Based on the CI, it appears that $CR \neq 1$ therefore we reject null hypothesis and there is skew for the data.

Question 3

Part A

$H_0 : p = p_0 = 0.05$

Part B

```r
# proportion with VAP
63/472
```

```
## [1] 0.13347
```

$\hat{p} = 0.1333$

Part C

```r
p <- 63/472
n <- 472
numer <- p * (1-p)

se <- sqrt(numer / n)

C <- 1.96

## CI
p + c(-1, 1)*C*se
```

```
## [1] 0.10279 0.16416
```

95% CI == (0.103, 0.164)

Part D

Type I error rate = 1 - 0.95 = 0.05

Part E

Based on this, we would reject the null hypothesis

Bonus note:

If we were to find a p-value associated with this, we would have been assuming that our null distribution was centered at $p_0 = 0.05$ and, from there, constructed a range of plausible values *under the null*. Here, we see the range of plausible values under the null is (0.019, 0.08), while the value we actually observed was $\hat{p} = 0.133$. The p-value would then be asking, based on a distribution where the 95% range of plausible values was (0.019, 0.08), what is the probability that we actually observed $\hat{p} = 0.133$? In this case, our conclusion would be the same and we would reject our null hypothesis.

```
p
```

```
## [1] 0.13347
```

```
0.05 + c(-1, 1)*C*se
```

```
## [1] 0.019319 0.080681
```