Probability & Statistics: Assignment 7

Due on 20 December, 2018 at 8:30am

Prof. D. Eynard

A. Romanelli / A. Vicini

Problem 1

1. We know from the nature of politics that the distribution that we are dealing with will be a binomial one. As seen in class we can get an estimation of $\hat{p}$ by computing:

$$\hat{p} = \left(\sum_{i=1}^{n} x_i\right) \cdot \frac{1}{n}$$

Which yields the following proportion for candidate A:

$$\frac{185}{351} \approx 0.527 = 52.7\% \text{ of the voters}$$

Against candidate B:

$$\frac{351 - 185}{351} \approx 0.473 = 47.3\% \text{ of the voters}$$

We can see that candidate A will be slightly favoured with a 2.7% margin of vantage.

2. We want to find a 95% confidence interval for the $\hat{p}$ that we just computed:

$$\text{se} = \sqrt{\hat{p}(1-\hat{p})/n} = 0.027$$

We can thus attempt to find an interval so that our confidence results:

$$P(\hat{p} - \delta < \hat{P} < \hat{p} + \delta) = 0.95$$

The estimation of the proportion can be treated as a random variable binomially distributed, since each sample would either satisfy our confidence interval or not. We can then treat this distribution as a normal one according to the **Central Limit Theorem**:

$$\begin{aligned} P(\hat{p} - \delta < \hat{P} < \hat{p} + \delta) &= P\left(-\tfrac{\delta}{\text{se}} < \tfrac{\hat{P}-\hat{p}}{\text{se}} < \tfrac{\delta}{\text{se}}\right) \\ &= 2 \cdot \Phi\left(\tfrac{\delta}{\text{se}}\right) - 1 \end{aligned}$$

We now need to find $2\Phi\left(\frac{\delta}{\text{se}}\right) - 1 = 0.95$, thus:

$$\Phi\left(\frac{\delta}{\text{se}}\right) = \frac{1+0.95}{2}$$

$$\frac{\delta}{\text{se}} = \Phi^{-1}\left(\frac{1+0.95}{2}\right)$$

$$\delta = 1.96 \cdot \text{se} = 0.053$$

We have now found the bounds of our interval to be: $[\hat{p} - \delta, \hat{p} + \delta] = [0.474, 0.58]$
Since our $\hat{p}$ lies within this interval we can be pretty confident that the estimated proportion is favourable to A.

3. We want to have a margin of error $\delta = 0.02$, thus we'd need to increase our sample population n such that:

$$\delta = 1.96 \cdot \text{se} = 0.02$$

$$\text{se} = \frac{0.02}{1.96}$$

$$\sqrt{\hat{p}(1-\hat{p})/n} = 0.0102$$

$$\frac{0.249271}{0.00010404} = n$$

$$n = 2396$$

Problem 2

1. Given the following data: (13.0, 18.5, 16.4, 14.8, 20.8, 19.3, 18.8, 23.1, 25.1, 16.8, 20.4, 17.4, 16.0, 21.7, 15.2, 21.3, 19.4, 17.3, 23.2, 24.9, 15.2, 19.9, 19.1, 18.1, 25.2, 23.1, 15.3, 19.4, 21.5, 16.8, 15.6, 17.6) we computed the sample mean $\overline{x}$ and the standard deviation σ with the following code:

```
samples <- c(13.0, 18.5, ..., 15.6, 17.6)
mu <- mean(samples)
print(mu)
sigma <- sqrt(var(samples))
print(sigma)
```

   Which gives us $\overline{x} = 19.068, \sigma = 3.255$

2. Under the assumption of an underlaying normal distribution and variance $\sigma^2 = 3.12$, we can attempt to find a 95% confidence interval in the following way:

$$[\overline{x} - \delta, \overline{x} + \delta]$$

$$\delta = z^* \frac{\sigma}{\sqrt{n}}$$

   Where the z value for our 95% confidence interval will be $z^* = 1.96$:

$$\delta_{0.95} = 1.96 \cdot \frac{\sqrt{3.12}}{\sqrt{32}} = 1.96 \cdot \frac{1.766}{5.657} = 0.612$$

   Thus we get the resulting :

$$\text{CI}_{0.95} = [\overline{x} - \delta_{0.95}, \overline{x} + \delta_{0.95}] = [19.068 - 0.612, 19.068 + 0.612] = [18.46, 19.68]$$

3. In case of a 99% confidence interval our z^* would be 2.58 instead, which would result in the following margin of error:

$$\delta_{0.99} = 2.58 \cdot \frac{\sqrt{3.12}}{\sqrt{32}} = 0.806$$

   And it's clearly visible that $\delta_{0.95} < \delta_{0.99}$ and thus we'd have a larger confidence interval than before. This comes from the fact that a wider interval grants us more confidence that our sample lies within the interval than not.

   Increasing the sample size n will reduce the margin of error and also tighten the interval.

Problem 3

a) Under the assumption of a biased coin with $\hat{p} = 0.6$, we can compute the probability as the one of a binomial distribution:

$$P(S_n = n \cdot \hat{p}) = \binom{n}{n \cdot \hat{p}} \hat{p}^{n\hat{p}} (1 - \hat{p})^{n(1-\hat{p})}$$

$$P\left(S_n = \frac{3}{5}n\right) = \binom{n}{\frac{3}{5}n} \left(\frac{3}{5}\right)^{\frac{3}{5}n} \left(\frac{2}{5}\right)^{\frac{2}{5}n}$$

For $n = 10$:

$$P(S_{10} = 6) = \binom{10}{6} \left(\frac{3}{5}\right)^{6} \left(\frac{2}{5}\right)^{4} \approx 210 \cdot 0.0466 \cdot 0.0256 = 0.25$$

For $n = 100$:

$$P(S_{100} = 60) = \binom{100}{60} \left(\frac{3}{5}\right)^{60} \left(\frac{2}{5}\right)^{40} \approx 0.081$$

For $n = 1000$:

$$P(S_{1000} = 600) = \binom{1000}{600} \left(\frac{3}{5}\right)^{600} \left(\frac{2}{5}\right)^{400} \approx 0.026$$

Problem 4

Problem 5

To accept or reject H0 we need to compute the z-value:

$$z = \frac{x - \mu}{\sigma/\sqrt{n}} \Rightarrow z = \frac{113.5 - 110}{10/4} = 1.4$$

If the resulting z is greater than what we just computed, then we'll accept H0, else we reject it.

a) A 5% significance level corresponds to $\alpha = 0.05$: $z_{5\%} = \Phi^{-1}(0.95) \overset{?}{>} 1.4$

$$z_{5\%} = 1.65 > 1.4, \text{ thus we accept H0}$$

b) Given $\alpha = 0.10$: $z_{10\%} = \Phi^{-1}(0.90) \overset{?}{>} 1.4$

$$z_{10\%} = 1.28 < 1.4, \text{ thus we reject H0 and accept H1 instead}$$

c) The p-value corresponds to the right-tail-end probability of the null hypothesis H0:

$$p = 1 - \Phi(z) = 1 - \Phi(1.4) \approx 1 - 0.919 = 0.081$$

Problem 6