Math 321 – Summer 2019
additional practice for chapter 5, confidence intervals

1. If we flip a fair coin 1000 times, find a 95% prediction interval for the number of heads. If we do not know the coin is fair, but get 61 heads, find a 95% confidence interval for the true probability of heads.

   Solution:

   For the prediction interval, we can either use the binomial of CLT normal approximation.
   `qbinom(c(0.025,0.975),size=1000,prob=0.5)` gives $(469, 531)$ for our predicted number of heads.

   The CLT approximation that the number of heads is $X \sim N(\mu = np, \sigma^2 = np(1-p))$ gives `qnorm(c(0.025,0.975),mean=500,sd=sqrt(1000*0.5*0.5))` for the CI $(469.0102, 530.9898)$.

2. Consider that a group of 30 people is randomly chosen from of a certain demographic category. The mean heigth is found to be 5 foot 8 inches and the sample standard deviation is found to be 2 inches. Estimate the true mean height of this demographic category with a 95% CI.

   Solution:

   We can use the normal approximation from the CLT since the sample size is reasonably large. Height data is also known to be very nearly normally distributed as well. $\overline{x} \pm z_{0.975}\frac{2}{\sqrt{30}}$ substituting in our data $5.67 \pm 1.96 \cdot \frac{2}{12\sqrt{30}}$ gives a CI of 5.67 ± 0.06 ft. Thus we think the true mean height is likely between 5ft 7in and 5ft 9in.

3. An internet server has data requests arrive at a very high rate. The number of data requests in a minute is collected for 60 randomly chosen 1 minute intervals. The mean requests per minute is found to be 982. Construct a 95% CI for true mean number of requests per minute. Model the number of requests in a 1 minute period by a Poisson random variable.

   Solution:

   Let X be the number of requests in a minute. We are modeling this as Poisson, so $X \sim Pois(\lambda)$. We wish to estimate λ with a 95% CI. Our sample size is 60 with a sample mean of $\overline{x} = 982$. Each data point x_i is a random draw from a Poisson distribution whose mean and variance are both equivalent to λ, thus by the CLT $\overline{X} \sim N(\mu = \lambda, \sigma^2 = \lambda/n)$. Thus a 95% CI for μ is $\overline{x} \pm z_{0.975}\sqrt{\frac{\lambda}{n}}$. We don't know λ so we will substitute our sample mean for it in the CI formula.

   $982 \pm z_{0.975}\sqrt{\frac{982}{60}}$ gives $(972, 990)$.

   Even if we wanted to be very conservative and use a much larger standard error, say $\sqrt{\frac{2000}{60}}$ and $z_{0.975} \approx 2$ instead of $\sqrt{\frac{982}{60}}$ and $z_{0.975} \approx 1.96$ then our 95% CI would be $(970, 994)$ which is not very different.

4. Polling data of 3,000 likely voters in Florida found Trump and Biden tied at 47% each with 6% indicating other candidates. Assuming that Biden and Trump are running against each other in 2020, estimate the actual 2020 vote result for Florida with a 95% confidence interval.

   Solution:

   We can use the rule of thumb 95% CI formula for a proportion as $\hat{p} + \pm\frac{1}{\sqrt{n}}$ to get a margin of error that is $\frac{1}{\sqrt{3000}} \approx 0.018$. This is identical to the margin of error reported by the poll!

   The 95% CI's for Trump and Biden are both 0.47 ± 0.018. This could mean either or both Trump's and Biden's vote count are reduced or increased by 1.8% with other candidates vote counts changing accordingly to make up for the difference.

   Note that there are many potential sources for error in such polls, and that a significant number of voters can even change their minds multiple times about who to vote for and even whether or not to vote at all!

5. A consumer preference survey randomly sampled 50 people from county A and 75 people from county B. Thirty one people from county A indicated that they preferred product I over product II, and 52 people from county B indicated a preference for I over II. Estimate the difference in proportions between the preference for product I between the two counties with a 95% CI. Estimate the true proportion in county B who prefer product I with a 95% CI as well.

   Solution:

   $\hat{p}_1 - \hat{p}_2 \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

   $0.62 - 0.693 \pm 1.96\sqrt{\frac{0.62(0.38)}{50} + \frac{0.693(0.307)}{75}}$ gives $(-0.243, 0.097)$ so we cannot really be too certain if county B residents have a stronger preference for product I.

   For county B the 95% CI for the proportion that prefer product I is $0.693 \pm \frac{1}{\sqrt{75}} = (0.58{,}0.81)$ using the simpler rule-of-thumb formula. If we use the CI formula more precisely, we get $0.693 \pm 1.96\sqrt{\frac{(0.693)(0.307)}{75}} = (0.589, 0.797)$.