ENGRD 2700: Basic Engineering Probability and Statistics
Fall 2019

Homework 7

Due **Friday Nov 22** at 11:59 pm. Submit to Gradescope by clicking the name of the assignment. See `https://people.orie.cornell.edu/yudong.chen/engrd2700_2019fa.html#homework` for detailed submission instructions.

The same stipulations from Homework 1 (e.g., independent work, computer code, etc.) still apply.

1. A sample of 100 service times at a call center has a sample mean of 9 minutes and a sample standard deviation of 6 minutes. Assume that the service times are independent and have a normal distribution.

   (a) Give a 95% confidence interval for the mean service time.
   Since we don't know the variance, the CI for this case is

   $$\bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

   Plugging in the number $\alpha = 5\%, \bar{x} = 9, s = 6, n = 100$, we obtain $t_{\alpha/2,n-1} = 1.984$ and the 95% CI

   $$9 \pm 1.19 = (7.81, 10.19).$$

   Since the sample size is large, we can use a normal distribution to approximate the t distribution. If we instead use the normal quantile $z_{\alpha/2} = 1.96$, the CI is $(7.82, 10.18)$. Using 2 instead of 1.96 is also fine, since it really doesn't change the solution much. Either the t or normal solution is fine.

   (b) Approximately how many service times we would have to collect to return a 95% confidence interval whose width is at most 15 seconds (= 1/4 minutes)?
   In this case, the number n will be very large, so the t distribution will be well approximated by a normal distribution and we can just use normal quantiles instead of the t quantile. Then the width of the CI will be $2z_{\alpha/2}\frac{s}{\sqrt{n}}$. (We use s to approximate σ.) In order to make the width at most 15 seconds or $15/60 = 1/4$ minutes, we should have

   $$2z_{\alpha/2}\frac{s}{\sqrt{n}} \leq \frac{1}{4}$$
   $$n \geq (8z_{\alpha/2}s)^2$$

   After plugging in the numbers, we obtain $n \geq 8851$. It is fine to give this answer to 2 significant figures, e.g., $n \geq 8800$ since we used s as an approximation to σ, and it is unlikely that this approximation is accurate. It is also fine to use $z_{\alpha/2} \approx 2$ in the calculation, in which case $n \geq 9216$.

2. Harry owns a bakery. The number of chocolate chips that he adds to his cookies is normally distributed with mean μ and variance $\sigma^2 = 25$, where μ is unknown. A customer buys a dozen of these cookies, and obtains the simple random sample

   31, 23, 42, 44, 28, 34, 19, 29, 30, 25, 28, 27

   (a) Compute a 95% confidence interval for μ.
   In lecture, we showed that

   $$\left[\bar{x}_n - 1.96\frac{\sigma}{\sqrt{n}},\ \bar{x}_n + 1.96\frac{\sigma}{\sqrt{n}}\right].$$

   is a 95% confidence interval for μ. Since $\bar{x}_n = 30$, we obtain the interval

   $$\left[30 - 1.96\frac{5}{\sqrt{12}},\ 30 + 1.96\frac{5}{\sqrt{12}}\right] = [27.2,\ 32.8].$$

(b) Compute 90% and 99% confidence intervals for μ.
In lecture, we showed that

$$\left[\bar{x}_n - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\ \bar{x}_n + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right].$$

is a $100(1-\alpha)\%$ confidence interval for μ, where $z_{\frac{\alpha}{2}}$ is such that

$$P\left(Z \geq z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}.$$

Since $z_{0.05} = 1.645$ and $z_{0.005} = 2.576$, our 90% confidence interval for μ is

$$\left[30 - 1.645\frac{5}{\sqrt{12}},\ 30 + 1.645\frac{5}{\sqrt{12}}\right] = [27.6,\ 32.4],$$

and our 99% confidence interval for μ is

$$\left[30 - 2.576\frac{5}{\sqrt{12}},\ 30 + 2.576\frac{5}{\sqrt{12}}\right] = [26.3,\ 33.7].$$

(c) Suppose the customer wants a 95% confidence interval that has a width of at most 2. How many cookies would he need to buy to achieve this?
For a given n, the width of the resulting interval is

$$W = \left(\bar{x}_n + 1.96\frac{5}{\sqrt{n}}\right) - \left(\bar{x}_n - 1.96\frac{5}{\sqrt{n}}\right) = 2 \cdot 1.96\frac{5}{\sqrt{n}}$$

Since the customer wants $W = 2$, solving for n yields

$$n = (5 \cdot 1.96)^2 \approx 96.04.$$

Thus, the customer would need to buy 97 cookies. (96 is also okay.)

3. Alice and Bob are running for state governor, and two polling agencies, Company X and Company Y, decide to gauge public opinion.

(a) Company X interviews 453 people, and finds that 55% of individuals want to vote for Alice. Construct a 95% confidence interval for p, the proportion of all voters in the state supporting Alice.
Given $\hat{p} = 0.55, n = 453$, and $\alpha = 0.05$, a $100(1-\alpha)\%$ confidence interval of p is

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = [0.504, 0.596].$$

(b) Company Y conducts its own independent study, and obtains the interval $[0.492, 0.568]$ from a sample size of 378. What confidence level did Company Y use?
Since

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = [0.492, 0.568],$$

and

$$\hat{p} = \frac{0.492 + 0.568}{2} = 0.53,$$

we have

$$z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.53 - 0.492 = 0.038.$$

Plugging in $\hat{p} = 0.53$ and $n = 378$, we get

$$z_{\alpha/2} = \frac{0.038}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} = 1.48.$$

Using `pnorm(1.48)-pnorm(-1.48)`, we find the corresponding confidence level is 86%.

(c) Suppose, contrary to the information specified in part (b), that Company Y obtained the 90% confidence interval [0.509, 0.591] instead. How many individuals did Company Y interview?

Since

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = [0.509, 0.591],$$

and

$$\hat{p} = \frac{0.509 + 0.591}{2} = 0.55, z_{\alpha/2} = 1.645,$$

we have

$$z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.55 - 0.509 = 0.041.$$

Solving for n yields

$$n = \left(\frac{z_{\alpha/2}}{0.041}\right)^2 \cdot \hat{p}(1-\hat{p}) \approx 398. \qquad \text{(399 is also okay.)}$$

4. The files `ithaca.csv` and `syracuse.csv` contain daily temperature data in Ithaca and Syracuse during the month of March. Import these datasets into R or RStudio.

(a) Let μ_i and μ_s denote the mean temperatures in both cities during the month of March. Construct a 97% confidence interval for $\mu_i - \mu_s$.

Let X represent the dataset `ithaca.csv` and Y represent the dataset `syracuse.csv`. Create new sample D from the differences

$$D_j = X_j - Y_j, \qquad j = 1, ..., 31.$$

Then a 97% confidence interval for $\mu_i - \mu_s$ is

$$\left[\bar{D} \pm z_{\alpha/2}\frac{s_D}{\sqrt{n}}\right] = [-0.540, 1.959].$$

We compute the confidence interval using the following `R` code:

```
syracuse = read.csv("syracuse.csv");
ithaca = read.csv("ithaca.csv");
diff = ithaca$maxtemp - syracuse$maxtemp;
n = length(diff);
left = mean(diff)-qnorm(0.985)*sd(diff)/sqrt(n);
right = mean(diff)+qnorm(0.985)*sd(diff)/sqrt(n);
```

(b) Repeat part (a), assuming the temperature in Ithaca is independent of the temperature in Syracuse (which is not true in reality). How do your intervals compare?

Dataset `ithaca.csv` has $\bar{x}_i = 51.452, n_i = 31, S_i = 12.662$. Dataset `syracuse.csv` has $\bar{x}_s = 50.742, n_s = 31, S_s = 12.596$. Thus $\bar{x}_i - \bar{x}_s = 0.71$ and a 97% confidence interval for $\mu_i - \mu_s$ is

$$\left[(\bar{x}_i - \bar{x}_s) \pm z_{\alpha/2}\sqrt{\frac{S_i^2}{n_i} + \frac{S_s^2}{n_s}}\right] = [-6.252, 7.671].$$

The interval is much larger when the two datasets are assumed to be independent.

5. Luigi is known far and wide for his meatball subs. Although he advertises that the subs weigh 400 g, Vinny, a regular, suspects that the subs weigh less. Vinny buys a sub each day for 81 consecutive days. He obtains the 95% confidence interval [393.08, 400.92] for the mean weight μ of a sub.

(a) Find $\bar{x}$ and s, the mean and standard deviation of his sample.

The midpoint of the confidence interval is 397, implying that $\bar{x} = 397$. It follows that

$$400.92 = \bar{x} + 1.96\frac{s}{\sqrt{n}} = 397 + 1.96\frac{s}{\sqrt{81}},$$

and so $s = 18$.

(b) Approximately how many subs would Vinny need to buy in order to halve the width of his confidence interval?

Since the width of the confidence interval is inversely proportional to $\sqrt{n}$, Vinny would need to quadruple his sample size to $4 \cdot 81 = 324$ subs in order to halve the width of his confidence interval.

(c) Compute 90% and 99% confidence intervals for μ.

Since $z_{0.05} = 1.645$ and $z_{0.005} = 2.576$, a 90% confidence interval for μ is

$$397 \pm 1.645\frac{18}{\sqrt{81}} = [393.71,\ 400.29],$$

and a 99% confidence interval for μ is

$$397 \pm 2.576\frac{18}{\sqrt{81}} = [391.85,\ 402.15].$$