Discussion 5

Chelsey Green

10/05/2018

1. Suppose the number of different computers used by students in the University computer labs last week has distribution:

Value	Probability
0	0.3
1	0.4
2	0.3

Suppose the distribution of computer use is still valid for this week and consider sampling two students at random. Let X_1 and X_2 be iid random draws from the population that represent the number of different computers used by the two students in your sample.

(a) Determine the missing elements in the table for the sampling distribution $\bar{X} = \frac{X_1+X_2}{2}$.

$\bar{X}$	Probability
0	$.3 * .3 = .09$
0.5	$2 * .4 * .3 = 0.24$
1	0.34
1.5	0.24
2.0	$.3 * .3 = .09$

Notice, total probability $.09 + .24 + .34 + .24 + .09 = 1$

(b) Calculate the expected value and standard error of $\bar{X}$ any way you want.

$\mu = E(X) = 0 * .3 + 1 * .4 + 2 * .3 = 1,$ *so* $E(\bar{X}) = 1 = 0 * .09 + 0.5 * .24 + 1 * .34 + 1.5 * .24 + 2 * .09$

and for variance: $\sigma_X^2 = (0-1)^2 * .3 + (1-1)^2 * .4 + (2-1)^2 * .3 = 0.6$ *so* $\sigma_{\bar{X}}^2 = \frac{0.6}{2} = 0.3$ *so* $SE(\bar{X}) = \sqrt{.3} = 0.5477226$

or the long way: $Var(\bar{X}) = (0-1)^2 * .09 + (0.5-1)^2 * .24 + (1-1)^2 * .34 + (1.5-1)^2 * .24 + (2-1)^2 * .09 = 0.3$ *so* $SE(\bar{X}) = \sqrt{.3} = 0.5477226$.

(c) Calculate the probability that the average number of different computers used by the two chosen students is atleast 1.5 using the pmf in part a. $P(\bar{X} \geq 1.5) = 0.24 + 0.09 = 0.33$

(d) Calculate the probability that the average number of different computers used by the two chosen students is atleast 1.5 using a Normal approximation. $P(\bar{X} \geq 1.5) \approx P(Z \geq \frac{1.5-1.0}{0.5477226}) = P(Z \geq 0.9129) = 1 - P(Z \leq 0.9129) = 1 - .8186 = 0.1814$

(e) Compare the the values in parts c and d. Explain the relationship between the two numbers. *The values are off by about 12% (with normal approximation being lower). Some reasons for the discrepency: Normal distribution is continuous (assumes that average takes on ranges of values all values 1.51, 1.501, 1.50001) while the sampling distribution for our $\bar{X}$ is very discrete (it only has 5 values). The sampling distribution of the $\bar{X}$ is not well approximated by the Normal Curve for a sample as small as 2 since the population is pretty different from a normal population.*

(f) If the sample size is increased to 50, find the probability that the average number of different computers used by the 50 chosen students is atleast 1.5. *With sample size of 50, we are more confident that CLT kicks in and* $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}}) = N(1, \sqrt{\frac{0.6}{50}}) = N(1, 0.1095)$ *so* $P(\bar{X} \geq 1.5) \approx P(Z \geq \frac{1.5-1}{0.1095}) = P(Z \geq 4.56621) \approx 0$. *With a sample size this large, we are more likely to find an average close to the true value [between* 1 ± 0.1095*].*

2. A caffeine drink company sells a drink with a label that claims a caffeine content of 86 mg. Sixteen bottles of the drink are randomly selected and analyzed for caffeine content. The resulting observations are:

83.7	88.6	83.5	88.3	83.9	84.9	85.4	85.6
89.8	86.2	83.9	86.1	84.5	87.3	85.2	86.7

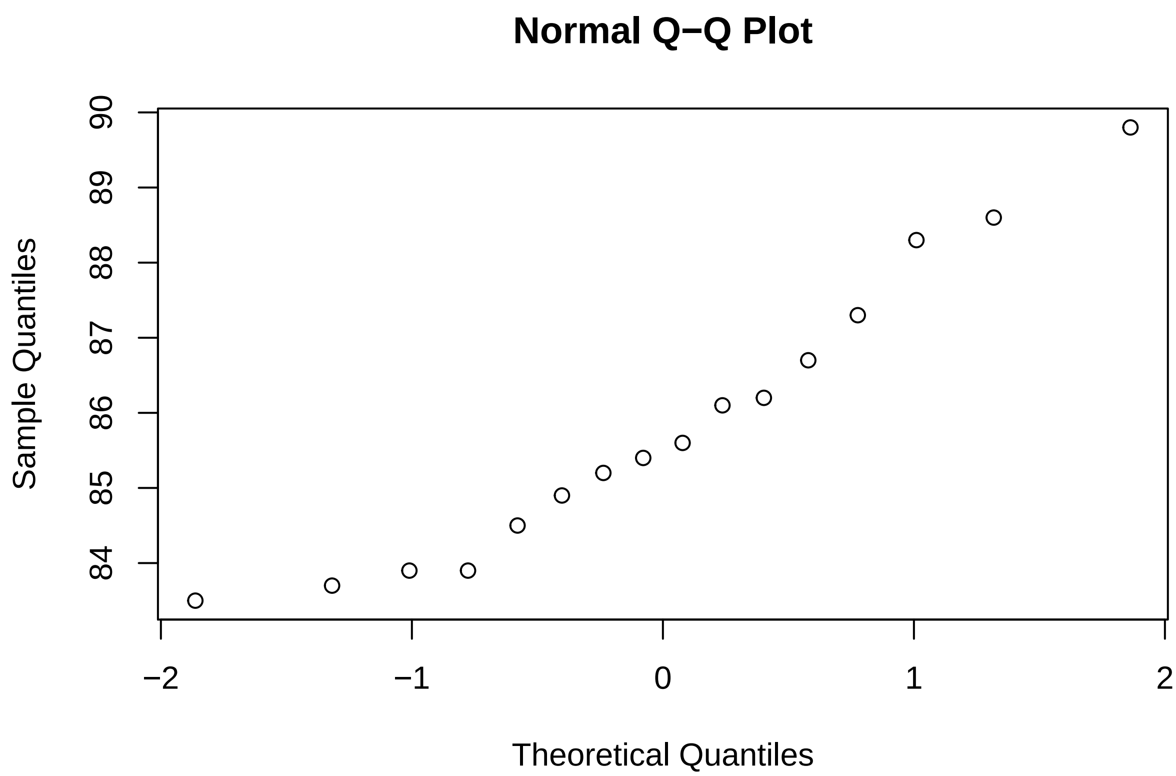
(a) Check that the assumptions for building a confidence interval are well met.

They appear to be independent observations- all from different bottles - hopefully from random selection of production batches. QQnorm plot looks pretty linear which tells me the assumption that data came from a normal population isn't unreasonable. (Or CLT would save me with sample size of 16 since data looks continuous and pretty bell-shaped)

(b) Construct an appropriate 90% confidence interval making sure to clearly show the point estimate, multiplier, and standard error estimate you used.

Since we do not know the population standard deviation, we will approximate it with the sample standard deviation $s_x = 1.88$ (Note, your calculator will also give you a value of $\sigma_x = 1.83$. This is not the value we want, we want the sample standard deviation.). We can also compute sample mean: 85.85. Then the t multiplier for $\alpha/2 = .10/2 = .05$ is $t_{15,.05} = 1.753$. So the t confidence interval for μ is $85.85 \pm 1.753 * \frac{1.88}{\sqrt{16}} = 85.85 \pm 0.82391 = (85.026, 86.67)$

```
caf_data<-c(83.7 , 88.6 , 83.5 , 88.3 , 83.9 , 84.9 , 85.4 , 85.6 , 89.8 , 86.2 , 83.9 , 86.1 , 84.5 , 8
qqnorm(caf_data)
```




```r
sd(caf_data)
```

```
## [1] 1.885736
```

```r
mean(caf_data)
```

```
## [1] 85.85
```

```r
t.test(caf_data, conf.level=.90)
```

```
##
##  One Sample t-test
##
## data:  caf_data
## t = 182.1, df = 15, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  85.02355 86.67645
## sample estimates:
## mean of x
##     85.85
```