Homework 6 - STAT 231

Suggested Solution *†

Due in class: 17 Oct 2019

Problem 1-6 are from the Devore textbook.

1. **(Chapter 5: Problem 46)** Let X denote the young's modulus of the aluminum alloy sheets. From the question we have $E(X) = 70, \sigma(X) = 1.6$ with unit GPa.

   (a) $E_1(\bar{X}) = E(X) = 70, \sigma_1(\bar{X}) = \sigma(X)/\sqrt{n_1} = 1.6/\sqrt{16} = .4 \quad \square$

   (b) $E_2(\bar{X}) = E(X) = 70, \sigma_2(\bar{X}) = \sigma(X)/\sqrt{n_2} = 1.6/\sqrt{64} = .2 \quad \square$

   (c) random sample (2) is more likely to have sample mean within 1 GPa of 70 GPa due to smaller sample variance.

2. **(Chapter 5: Problem 54)** Let X denote the sediment density of the specimen. From the question we have $X \sim N(\mu, \sigma^2), \mu = E(X) = 2.65, \sigma = \sigma(X) = .85$ with unit GPa. Then $\bar{X} \sim (\mu_1, \sigma_1^2), \mu_1 = \mu = 2.65, \sigma_1 = \sigma/\sqrt{n} = .85/\sqrt{25} = .17$.

   (a) $P(\bar{X} \leq 3) = P(\frac{\bar{X}-\mu_1}{\sigma_1} \leq \frac{3-\mu_1}{\sigma_1}) = P(Z \leq \frac{3-\mu_1}{\sigma_1}) = P(Z \leq 2.06) \approx .9803 \quad \square$
   $P(2.65 \leq \bar{X} \leq 3) = P(\frac{2.65-\mu_1}{\sigma_1} \leq \frac{\bar{X}-\mu_1}{\sigma_1} \leq \frac{3-\mu_1}{\sigma_1}) = P(0 \leq Z \leq 2.06) = P(Z \leq 2.06) - P(Z \leq 0) \approx .9803 - .5 = .4803 \quad \square$

   (b) Consider a sample size n, $\bar{X} \sim (\mu_n, \sigma_n^2), \mu_n = \mu = 2.65, \sigma_n = \sigma/\sqrt{n}$, and $P(\bar{X} \leq 3) = P(\frac{\bar{X}-\mu_n}{\sigma_n} \leq \frac{3-\mu_n}{\sigma_n}) = P(Z \leq \frac{3-\mu_n}{\sigma/\sqrt{n}}) = P(Z \leq .35\sqrt{n}/.85) \geq .99 \implies .35\sqrt{n}/.85 \geq 2.33 \implies n \geq 32.02$, take $n = 33 \quad \square$

3. **(Chapter 7: Problem 3)** Let X denote the average . From the question we have $X \sim N(\mu, \sigma^2)$,

*Zhiling GU (zlgu@iastate.edu)

†Please inform me if there is any error or possible improvement

(a) Narrower. The probability of the 90% confidence interval contains $\bar{X}$ is smaller than that of the 95% confidence interval, therefore the width of confidence interval is narrower. Note the width of the confidence interval equal to $2 * \Phi^{-1}(1-\alpha/2)\sigma/\sqrt{n}$. Notice $z_{.05} = \Phi^{-1}(.05) = 1.645 < z_{.25} = \Phi^{-1}(.025) = 1.96$, the former one for the 90% confidence interval and the latter one for the 95% confidence interval.

(b) No. μ is a fixed constant, while confidence interval is random as a function of realized random variables. It is more appropriate to say there is 95% of the intervals may contain μ.

(c) No. The confidence interval is for the population mean instead of individual values.

(d) No. 95% is just the probability of confidence interval containing the true mean. It does not mean the percentage of a finite number of interval is bound to be 95%. If the process were repeated for infinite times, this statement is correct.

4. **(Chapter 7: Problem 4)** We compute $\bar{x} \pm \Phi^{-1}(1-\alpha/2)\sigma/\sqrt{n}$

(a) $95\%CI = 58.3 \pm 1.96 \times \frac{3}{\sqrt{25}} = (57.12, 59.48)$

(b) $95\%CI = 58.3 \pm 1.96 \times \frac{3}{\sqrt{100}} = (57.71, 58.88)$

(c) $99\%CI = 58.3 \pm 2.58 \times \frac{3}{\sqrt{100}} = (57.53, 59.07)$

(d) $82\%CI = 58.3 \pm 1.34 \times \frac{3}{\sqrt{100}} = (57.90, 58.70)$

(e) $\text{width} = 2 \times 2.58 \times \frac{3}{\sqrt{n}} = 1 \implies n = 239.6 \approx 240$

5. **(Chapter 7: Problem 12)**

(a) Yes. The central limit behavior of the mean does not depend on the original distribution of random variables.

(b) From the data we have $n = 43, \hat{\mu} = 1191.627907, s = \hat{\sigma} = 506.63 \implies CI = \hat{x} \pm 2.58 * s/\sqrt{n} = (992.29, 1390.96)$.

6. **(Chapter 7: Problem 14)** $\bar{x} = 1427, s = 325, n = 514$

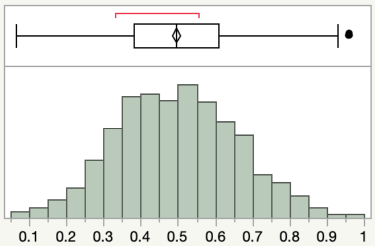
(a) $95\%CI = 1427 \pm 1.96 \times \frac{325}{\sqrt{514}} = (1398.90, 1455.09)$

(b) $\text{width} = 2 \times 1.96 \times \frac{320}{\sqrt{n}} = 50 \implies n = 629.4 \approx 630$

7. 1000×3 Uniform(0,1) simulations

(a) See below

(X1+X2+X3)/3



Quantiles

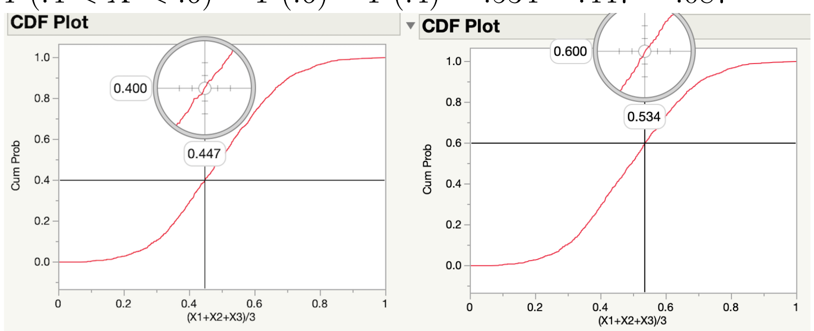
100.0%	maximum	0.9605894898
99.5%		0.9224973321
97.5%		0.8190891077
90.0%		0.703037872
75.0%	quartile	0.6075804684
50.0%	median	0.4935566164
25.0%	quartile	0.3821108107
10.0%		0.2913627915
2.5%		0.1804139798
0.5%		0.0985577493
0.0%	minimum	0.0636469581

Summary Statistics

Mean	0.4957354
Std Dev	0.160646
Std Err Mean	0.0050801
Upper 95% Mean	0.5057043
Lower 95% Mean	0.4857666
N	1000

i. The histogram is symmetric and bell shaped, looks similar to the theoretical normal histribution. Note from the results the simulated mean is .4957 and the simulated standard deviation is .1606.

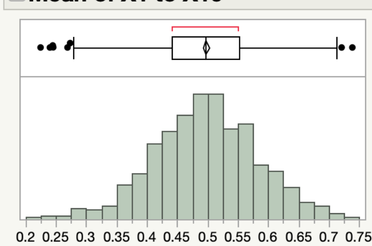
ii. $\hat{P}(.4 < \bar{X} < .6) = \hat{F}(.6) - \hat{F}(.4) = .534 - .447 = .087$



iii. $n = 3, \mu = .5, \sigma = \sqrt{1/12} = .2887$, therefore the theoretical standard deviation for $\bar{x}$ is $\sigma/\sqrt{3} = 1/6 = .1667$, theoretical mean for $\bar{x}$ is .5. They are close to the simulated results. Percentage absolute errors of the mean and standard deviation are resp. .86% and 3.66%.

(b) When $n = 10$, see below

Mean of X1 to X10



Quantiles

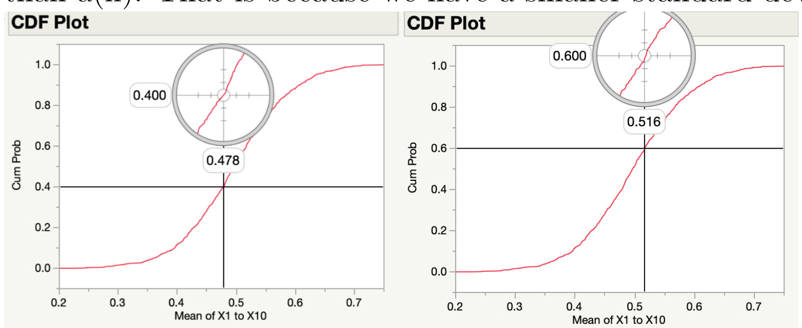
100.0%	maximum	0.7391164249
99.5%		0.7119540956
97.5%		0.6651148721
90.0%		0.6058281966
75.0%	quartile	0.5526917056
50.0%	median	0.4962527191
25.0%	quartile	0.4419239505
10.0%		0.3952916392
2.5%		0.3218144712
0.5%		0.269353766
0.0%	minimum	0.2248933434

Summary Statistics

Mean	0.4971918
Std Dev	0.0839639
Std Err Mean	0.0026552
Upper 95% Mean	0.5024021
Lower 95% Mean	0.4919814
N	1000

i. The histogram is symmetric and bell shaped, while it is less spread. The mean of $\bar{X}_{10}$ is .4972, the standard deviation of $\bar{X}_{10}$ is .084. While the mean gets closer to .5 , the standard deviation gets smaller. Note that the theoretical value of mean and standard deviation become .5 and $\sigma/\sqrt{10} = 1/\sqrt{120} = .091$. Percentage absolute errors of the mean and standard deviation are respectively .056% and 7.7%.

ii. $\hat{P}(.4 < \bar{X} < .6) = \hat{F}(.6) - \hat{F}(.4) = .516 - .478 = .038$ which is narrower than a(ii). That is because we have a smaller standard deviation.



8. Note the number of confidence intervals should be around $1000 * 90\% = 900$. In this question, we assume the population standard deviation $\sigma = 2$ is known. See program in the 'q8.jmp' file.

9. (a) For the mpg, denoted by X, we have $\bar{X} = 29.36, \hat{\sigma} = 5.54 \implies 95\%\,CI = \bar{X} \pm 1.96 \times \hat{\sigma}/\sqrt{110} = (28.32, 30.39)$, which is close to the jmp output.

Summary Statistics

Mean	29.363636
Std Dev	5.5367451
Std Err Mean	0.5279079
Upper 95% Mean	30.409933
Lower 95% Mean	28.31734
N	110

Confidence Intervals

Parameter	Estimate	Lower CI	Upper CI	1-Alpha
Mean	29.36364	28.31734	30.40993	0.950
Std Dev	5.536745	4.889216	6.383542	0.950

(b) For the size, denoted by Y, let $Y = 1$ if midsized, $Y = 0$ otherwise. Then we have $Y \sim Bernoulli(p)$. From the data, we have $\hat{p} = \bar{Y} = .31, \hat{\sigma} = \sqrt{\hat{p}(1-\hat{p})} = \sqrt{.31 \times .69} \approx .462 \implies 95\%\,CI_p = \bar{Y} \pm 1.96 \times \hat{\sigma}/\sqrt{110} = (.2236, .396)$, which is close to the jmp output.

Frequencies

Level	Count	Prob
Large	29	0.26364
Midsized	34	0.30909
Small	47	0.42727
Total	110	1.00000
N Missing	0	

Confidence Intervals

Level	Count	Prob	Lower CI	Upper CI	1-Alpha
Large	29	0.26364	0.190283	0.352941	0.950
Midsized	34	0.30909	0.2304	0.400666	0.950
Small	47	0.42727	0.338823	0.520631	0.950
Total	110				

Note: Computed using score confidence intervals.