Homework 7 - STAT 231

Suggested Solution *†

Due in class: 17 Oct 2019

Problem 1-8 are from the Devore textbook.

1. **(Chapter 7: Problem 19)**

   Recall the theorem: a confidence interval for a population proportion p with confidence level approximately $100(1-\alpha)\%$ is

   $$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}$$

   where $\tilde{p} = \frac{n\hat{p}+2}{n+4}$.

   From the question, we know $n = 356, \hat{p} = 201/356, \alpha = .05, \tilde{p} = \frac{201+2}{356+4}$. So a (two-sided) 95% confidence interval for the proportion of all dies that pass the probe is

   $$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}} = \frac{201}{356} \pm 1.96\sqrt{\frac{203}{360}\cdot\frac{157}{360}/356} = (0.513, 0.616) \quad \square$$

2. **(Chapter 7: Problem 22)**

   (a) From the question, we know $n = 143, \hat{p} = 10/143, \alpha = .05, \tilde{p} = \frac{10+2}{143+4}$. So a one-sided 95% confidence interval in the form (lower bound, 1) has the lower bound

   $$\hat{p} - z_{\alpha}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}} = \frac{10}{143} - 1.645\sqrt{\frac{12}{147}\cdot\frac{135}{147}/143} = .0323 \quad \square$$

*Zhiling GU (zlgu@iastate.edu)
†Please inform me if there is any error or possible improvement

(b) There is 95% of chance that the interval $(.0323, 1)$ will include the true population proportion of artificial hips that experience squeaking.

3. **(Chapter 7: Problem 29)** Determine the t critical value(s) that will capture the desired t-curve area in each of the following cases: Look up the t-table, we have:

   (a) Central area = .95, df = 10 $\Longrightarrow$ $\pm t_{.025,10} \approx \pm 2.228$

   (b) Central area = .95, df = 20 $\Longrightarrow$ $\pm t_{.025,20} \approx \pm 2.086$

   (c) Central area = .99, df = 20 $\Longrightarrow$ $\pm t_{.005,20} \approx \pm 2.845$

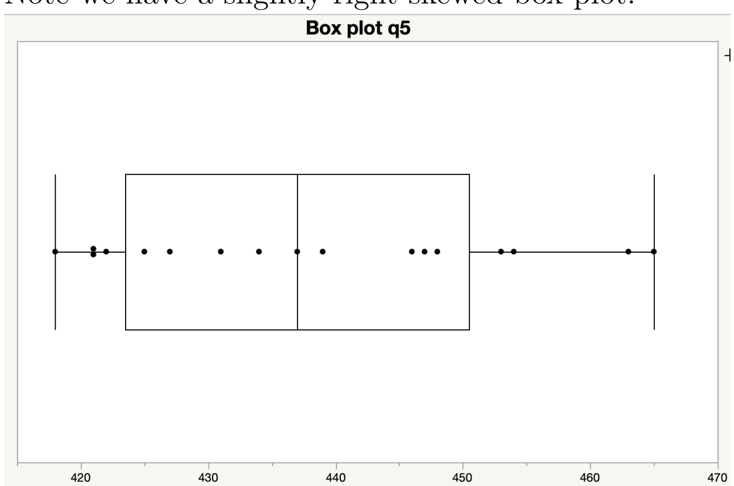
   (d) Central area = .99, df = 50 $\Longrightarrow$ $\pm t_{.005,50} \approx \pm 2.678$

   (e) Upper-tail area = .01, df = 25 $\Longrightarrow$ $t_{.01,25} \approx 2.485$

   (f) Lower-tail area = .025, df = 5 $\Longrightarrow$ $-t_{.025,5} \approx -2.571$

4. **(Chapter 7: Problem 33) Just (c)**

   (a) Note we have a slightly right-skewed box plot.



   (b) Based on a normal probability plot, it is reasonable to assume the sample observations came form a normal distribution.

   (c) $n = 17, df = n - 1 = 16, \overline{X} = 438.294, s = 15.144, t_{.025,16} = 2.120 \Longrightarrow$ $95\%CI = \overline{X} \pm \hat{t}_{.025,16} s/\sqrt{n} = 438.29 \pm 2.120 \times 15.14/\sqrt{17} = (430.51, 446.08)$ Since 440 falls in this confidence interval while 450 does not, we say 440 is a plausible value but 450 is not a plausible value for true average degree of polymerization.

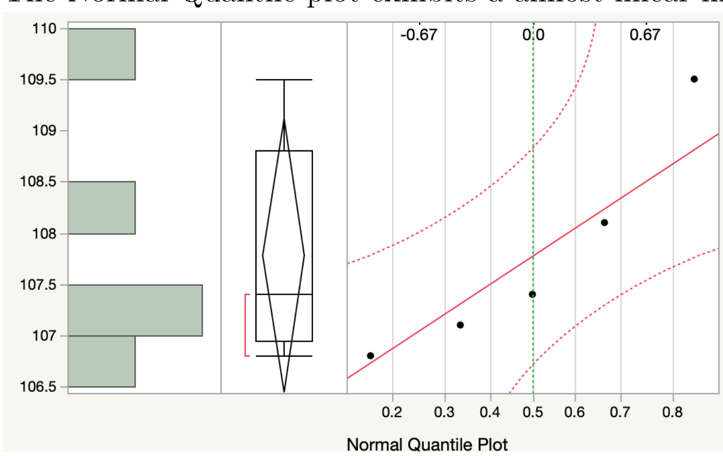
5. **(Chapter 7: Problem 35)**

   (a) $n = 15, df = 14, \overline{X} = 25, s = 3.5, t_{.025,14} = 2.1448 \implies$ 95% Confidence Interval $= \overline{X} \pm \hat{t}_{.025,14} \times s/\sqrt{n} = 25 \pm 2.1448 \times 3.5/\sqrt{15} = (23.06, 26.94)$.

   (b) 95% Prediction Interval $= \overline{X} \pm \hat{t}_{.025,14} \times s \times \sqrt{1 + 1/n} = 25 \pm 2.1448 \times 3.5 \times \sqrt{1 + 1/15} = (17.25, 32.75)$. Prediction interval is wider than confidence interval.

6. **(Chapter 7: Problem 38)**

   (a) The Normal Quantile plot exhibits a almost linear line:



   (b) $n = 5, df = 4, \overline{X} = 107.78, s = 1.075, t_{.025,4} = 2.776 \implies$ 95% Confidence Interval $= \overline{X} \pm t_{.025,44} \times s/\sqrt{n} = 107.78 \pm 2.776 \times 1.075/\sqrt{5} = (106.445, 109.115)$. Since 107 falls in the confidence interval, it is a plausible value. Since 110 does not fall in the confidence interval, it is not plausible.

   (c) 95% Predict Interval $= \overline{X} \pm t_{.025,14} \times s \times \sqrt{1 + 1/n} = 107.78 \pm 2.776 \times 1.075 \times \sqrt{1 + 1/5} = (104.51, 111.05)$. The width is wider than confidence interval.

   (d) 95% Tolerance Interval $= \overline{X} \pm \tau_{n,95\%} s = 107.78 \pm \tau_{5,95\%} = 107.78 \pm 5.079 \times 1.075 = (102.32, 113.24)$.

7. **(Chapter 7: Problem 42)**

   (a) $\chi^2_{.1,15} = 22.307$

   (b) $\chi^2_{.1,25} = 46.925$

   (c) $\chi^2_{.01,25} = 34.381$

(d) $\chi^2_{.005,25} = 11.523$

(e) $\chi^2_{.99,25} = 44.313$

(f) $\chi^2_{.995,25} = 10.519$

8. **(Chapter 7: Problem 44)**

   $n = 9, df = 8, \bar{s} = 2.81, \chi^2_{.975,8} = 2.180, \chi^2_{.025,8} = 17.534 \implies$ 95% Confidence Interval for $\sigma^2 = ((n-1)s^2/\chi^2_{.025,8}, (n-1)s^2/\chi^2_{.975,8}) = (\frac{8\times 2.81^2}{17.534}, \frac{8\times 2.81^2}{2.180}) = (3.602646, 28.97651)$. Then 95% Confidence Interval for σ is $(\sqrt{3.602646}, \sqrt{28.97651}) \approx (1.898, 5.383)$.

9. (a) See q9.jmp

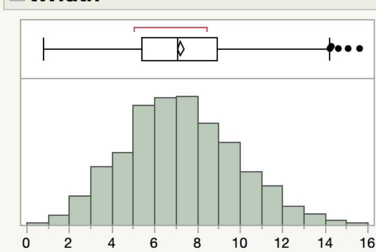
   (b) See q9.jmp

   (c) The interval in (b) exhibits 80.1% coverage, significantly lower than 90% which is due to the noisy sample standard deviation we applied without adjusting for the noise in them. The standard deviations, as we can see from the table, vary a lot and can be far away from the true standard deviation 4 due to the small sample size.

   (d) See q9.jmp

   (e) The widths of the t intervals are on average (mean 7.20) much wider than the z^* interval (constant 5.885). See the width of the simulated t^* intervels as following.

tWidth



Quantiles

100.0%	maximum	15.630226489
99.5%		14.238477454
97.5%		12.780521736
90.0%		10.749041017
75.0%	quartile	8.9429267955
50.0%	median	7.0683768212
25.0%	quartile	5.4181102242
10.0%		3.8567183266
2.5%		2.4163636716
0.5%		1.6797280657
0.0%	minimum	0.8121207814

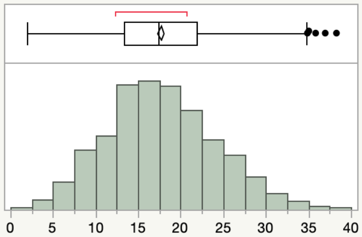
Summary Statistics

Mean	7.2052408
Std Dev	2.6004303
Std Err Mean	0.0822328
Upper 95% Mean	7.3666096
Lower 95% Mean	7.0438719
N	1000

   (f) See q9.jmp

   (g) The widths of the prediction intervals are on average (mean 7.21) wider than the confidence interval (constant 5.885). See the width of the simulated prediction intervels as following.

pWidth



Quantiles

100.0%	maximum	38.286079463
99.5%		34.877004477
97.5%		31.305756899
90.0%		26.329665715
75.0%	quartile	21.905607456
50.0%	median	17.313916522
25.0%	quartile	13.271605419
10.0%		9.4469919817
2.5%		5.9188580283
0.5%		4.1144766675
0.0%	minimum	1.989281524

Summary Statistics

Mean	17.649163
Std Dev	6.3697274
Std Err Mean	0.2014285
Upper 95% Mean	18.044435
Lower 95% Mean	17.253892
N	1000

(h) See q9.jmp

(i) See q9.jmp

(j) The proportion of the prediction intervals containing the new value is 90.1% which is expected to be close to 90%. Your answers may differ since simulations are random, but the proportion should be close to 90% anyway.