Assignment 4

Due date : 5/20/2022

21102052 Lee Jeong-Yun

1. (Slightly modified version of Ex 7-13 on p 283). The article "Gas Cooking, Kitchen Ventilation, and Exposure to Combustion Products" (Indoor Air, 2006: 65–73) reported that for a sample of 50 kitchens with gas cooking appliances monitored during a one-week period, the sample mean CO_2 level (ppm) was 650, and the sample standard deviation was 160.

a. Calculate and interpret a 95% (two-sided) confidence interval for true average CO_2 level in the population of all homes from which the sample was selected.

b. Suppose the investigators had made a rough guess of 150 for the value of s before collecting data. What sample size would be necessary to obtain an interval width of 50 ppm for a confidence level of 95%?

a)

$z_{0.05/2} = 1.96 \quad \bar{X} = 650. \quad n = 50 \quad s = 160.$

$\left(650 - \frac{160}{\sqrt{50}},\ 650 + \frac{160}{\sqrt{50}}\right) = (605.65,\ 694.35)$

b)

$n = \left(2 \times z_{0.05/2} \times \frac{s}{w}\right)^2 = \left(2 \times 1.96 \times \frac{150}{50}\right)^2 = (11.76)^2 = 138.2967$

$n = 138.$

2. (Slightly modified version of Ex 7-20 on p 284) . The Associated Press (October 9, 2002) reported that in a survey of 4700 American youngsters aged 6 to 19, 725 of whom were seriously overweight (a body mass index of at least 30; this index is a measure of weight relative to height).

a Calculate a confidence interval using a 99% confidence level for the proportion of all American youngsters who are seriously overweight by (7.10)

b Calculate a confidence interval using a 99% confidence level for the proportion of all American youngsters who are seriously overweight by (7.11)

c Calculate a confidence interval using a 99% confidence level for the proportion of all American youngsters who are seriously overweight by by prop.test(k,n) with k=725, n=4700

3. (Slightly modified version of Ex 7-33 on p 292) The following observations are on degree of polymerization for paper specimens for which viscosity times concentration fell in a certain middle range:

428 425 418 422 425 427 431 434 437 439 446 447 448 453 454 463 465

a. Is it plausible that the given sample observations were selected from a normal distribution?

– Perform by the Shapiro-Wilk normality test

> shapiro.test(data)

– Perform by plotting normal probability plot

b. Calculate a two-sided 95% confidence interval for true average degree of polymerization

a)

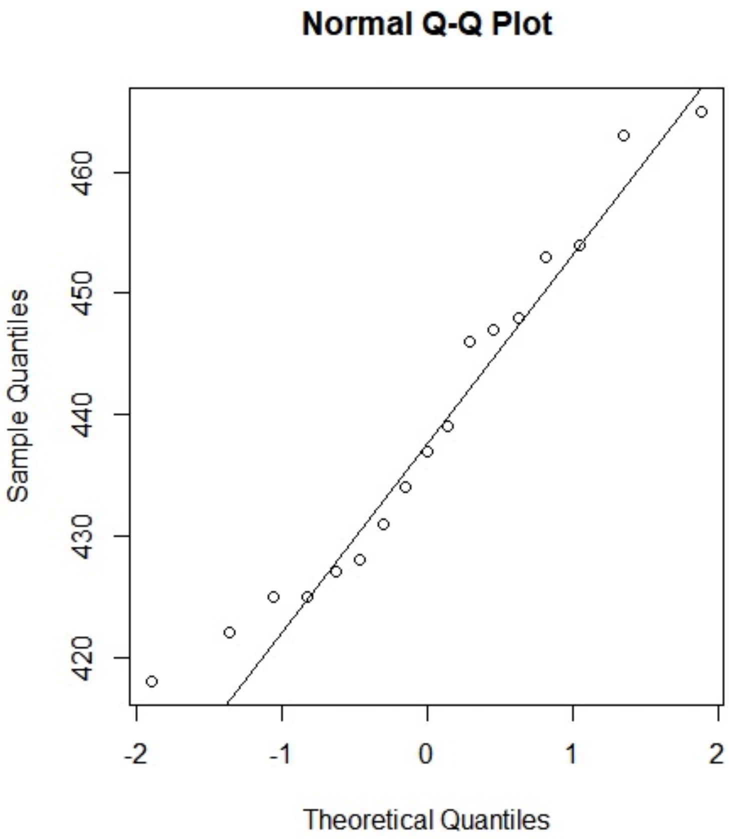
```
> x <- c(428,425,418,422,425,427,431,434,
+        437,439,446,447,448,453,454,463,465)
> shapiro.test(x)

        Shapiro-Wilk normality test

data:  x
W = 0.94629, p-value = 0.4005
```

p-value > 0.05.

→ follow normal distribution



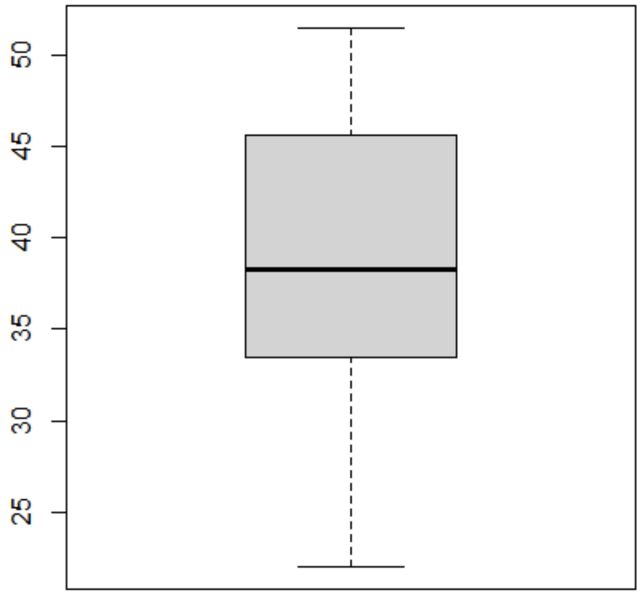
→ line & value have significant relationship.

The data follow normal distribution

4. (Slightly modified version of Ex 7-49 on p 297) For each of 18 preserved cores from oil-wet carbonate reservoirs, the amount of residual gas saturation after a solvent injection was measured at water flood-out. Observations, in percentage of pore volume, were

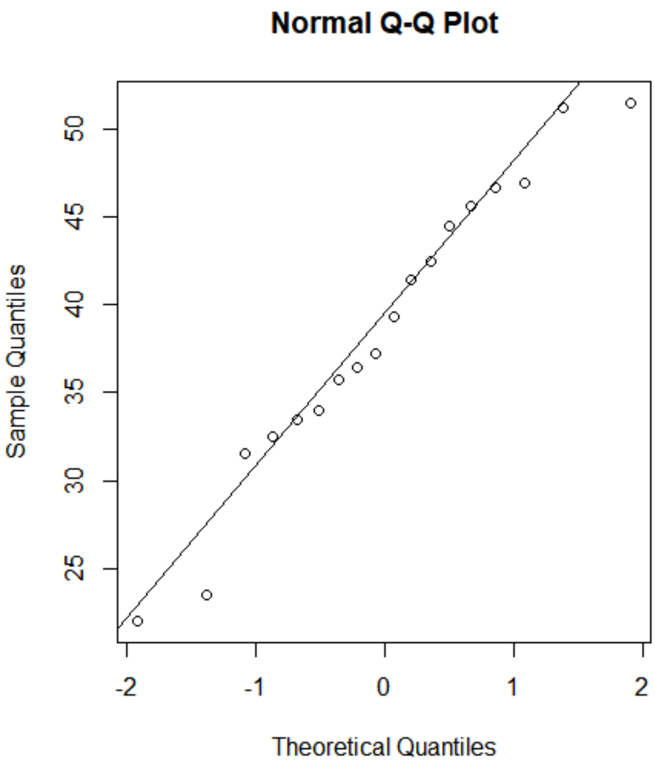
23.5	31.5	34.0	46.7	45.6	32.5	41.4	37.2	42.5
46.9	51.5	36.4	44.5	35.7	33.5	39.3	22.0	51.2

a) Construct a boxplot of this data, and comment on any interesting features.



Min	1Q	Median	Mean	3Q	Max	IQR
22.00	33.62	38.25	38.66	45.33	51.50	11.71

b) Is it plausible that the sample was selected from a normal population distribution?

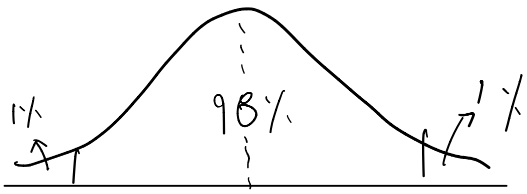


Excluding the extreme value, since the data are distributed close to the approximate line, it can be said that it follows the normal distribution.

c) Calculate a 98% CI for the true average amount of residual gas saturation.

$\Sigma x = 695.9$ $\quad \Sigma x^2 = 28124.79$ $\quad \bar{x} = \frac{695.9}{18} = 38.66$

$$s = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1}} \fallingdotseq 8.47$$



98% CI $\rightarrow 38.66 \pm \boxed{5.12}$ $\rightarrow t_{0.01,\,17} \times \frac{s}{\sqrt{n}} = 2.567 \times \frac{8.47}{\sqrt{18}}$

$\therefore [33.54,\ 43.78]$

5. (Slightly modified version of Ex 8-33 on p 323) The article "Uncertainty Estimation in Railway Track Life- Cycle Cost" (J. of Rail and Rapid Transit, 2009) presented the following data on time to repair (min) a rail break in the high rail on a curved track of a certain railway line.

   159 120 480 149 270 547 340 43 228 202 240 218

   A normal probability plot of the data shows a reasonably linear pattern, so it is plausible that the population distribution of repair time is at least approximately normal. The sample mean and standard deviation are 249.7 and 145.1, respectively.

a) Is there compelling evidence for concluding that true average repair time exceeds 190 min? Carry out a test of hypotheses using a significance level of .05.

b) Using σ =150, what is the type II error probability of the test used in (a) when true average repair time is actually 290 min? That is, what is $\beta(290)$?

a)

$H_0 = \mu = 190$

$H_1 = \mu > 190$

$\alpha = 0.05$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{249.7 - 190}{145.1/\sqrt{12}} = 1.425$$

p-value = 0.077

$\rightarrow$ accept H_0

6. (Slightly modified version of Ex 8-45 on p 328) A random sample of 160 recent donations at a certain blood bank reveals that 85 were type A blood. Does this suggest that the actual percentage of type A donations differs from 40%, the percentage of the population having type A blood? Carry out a test of the appropriate hypotheses using a significance level of .01. Would your conclusion have been different if a significance level of .05 had been used?

$n = 160$

$\hat{p} = \frac{85}{160} \approx 0.53 \qquad p_0 = 0.4 \qquad \rightarrow \qquad z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.53 - 0.4}{\sqrt{\frac{0.4 \times 0.6}{160}}} = 3.33$

$\begin{cases} H_0 : p = 0.4 \\ H_1 : p \neq 0.4 \end{cases}$

$\alpha = 0.01$

$z_{\frac{\alpha}{2}} = 2.58 < z \rightarrow$ reject H_0, $p = 0.4$

$\alpha = 0.05$

$z_{\frac{\alpha}{2}} = 1.96 < z \rightarrow$ reject H_0, $p = 0.4$

7. (Ex 9-32 on p 364) The degenerative disease osteoarthritis most frequently affects weight-bearing joints such as the knee. The article "Evidence of Mechanical Load Redistribution at the Knee Joint in the Elderly when Ascending Stairs and Ramps" (Annals of Biomed. Engr., 2008: 467–476) presented the fol- lowing summary data on stance duration (ms) for samples of both older and younger adults.

Age	Sample Size	Sample Mean	Sample SD
Older	30	800	115
Younger	18	780	70

Assume that both stance duration distributions are normal.

a) Calculate and interpret a 99% CI for true average stance duration among elderly individuals.

b) Carry out a test of hypotheses at significance level .05 to decide whether true average stance duration is larger among elderly individuals than among younger individuals.

a) $n=30$ T-distribution

$\bar{x}=800$

$s_1=115$

$$\left(\bar{x}-t_{0.005,29}\cdot\frac{s}{\sqrt{n}},\ \bar{x}+t_{0.005,29}\frac{s}{\sqrt{n}}\right)$$

$$=\left(800-2.462\cdot\frac{115}{\sqrt{30}},\ 800+2.462\cdot\frac{115}{\sqrt{30}}\right)$$

$$= 748.31 \le \mu \le 851.69$$

b) $H_0: \mu_1-\mu_2=0$

$H_1: \mu_1-\mu_2>0$

$$t=\frac{\bar{x}-\bar{y}-\Delta_0}{\sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{n}}}=\frac{20}{\sqrt{440.83+272.22}}=0.7489$$

$$v \Rightarrow \frac{\left(\frac{s_1^2}{m}+\frac{s_2^2}{n}\right)^2}{\frac{\left(\frac{s_1^2}{m}\right)^2}{m-1}+\frac{\left(\frac{s_2^2}{n}\right)^2}{n-1}}=45.97 \rightarrow v=45$$

$t_{0.05,45}=1.6794$

$t<t_{0.05,45}$

$\Rightarrow$ not reject H_0

8. (Ex 9-36 on p 371) Consider the accompanying data on breaking load (kg/25 mm width) for various fabrics in both an unabraded condition and an abraded condition ("The Effect of Wet Abrasive Wear on the Tensile Properties of Cotton and Polyester-Cotton Fabrics," J. Testing and Evaluation, 1993: 84–93). Use the paired t test, as did the authors of the cited article, to test H_0: $\mu_D = 0$ versus H_a: $\mu_D > 0$ at significance level .01.

	Fabric							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
U	36.4	55.0	51.5	38.7	43.2	48.8	25.6	49.8
A	28.5	20.0	46.0	34.5	36.5	52.5	26.5	46.5

$H_0 : \mu_D = 0$ $\alpha = 0.01$

$H_1 : \mu_D > 1$

$U - A = d.$

$\sum d = 58$, $\sum d^2 = 1405.98$

$\bar{d} = 7.25$

$S_d = 11.853.$

$t = 1.73$

$t < t_{0.1,7} = 2.9979$

$\rightarrow$ cannot reject H_0

9. (Ex 9-49 on p 380) Is someone who switches brands because of a financial inducement less likely to remain loyal than someone who switches without inducement? Let p_1 and p_2 denote the true proportions of switchers to a certain brand with and without inducement, respectively, who subsequently make a repeat purchase. Test H_0: $p_1 - p_2 = 0$ versus H_a: $p_1 - p_2 > 0$ using $\alpha = 0.01$ and the following data

m = 200 number of success = 30

n = 600 number of success = 180

$\hat{p}_1 = 30/200 = 0.15 \qquad \hat{p}_2 = \frac{180}{600} = 0.3$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{m} + \frac{1}{n}\right)}}$$

$$\hat{p} = \frac{m \cdot \hat{p}_1}{n+m} + \frac{n \times \hat{p}_2}{m+n} = \frac{200 \times 0.15}{800} + \frac{600 \times 0.3}{800} = 0.0375 + 0.225 = 0.2625$$

$$z = \frac{0.15 - 0.3}{\sqrt{0.2625 \times 0.7375\left(\frac{1}{200} + \frac{1}{600}\right)}} \approx -4.18$$

$z_\alpha = P(Z \geq z_\alpha) = 0.01$

$1 - P(Z \leq z_\alpha) = 0.01$

$P(Z \leq z_\alpha) = 0.99 \qquad z_\alpha = 2.33$

$-z_\alpha = -2.33$

$z = -4.18 < -2.33 = -z_\alpha$

cannot reject H_0