Summary of Chapter 6

1 Concepts

- **Confidence Interval Estimate for μ:**

  1. σ is known:

  Two-sided: A $100(1-\alpha)\%$ confidence interval (C.I.) on μ is given by

$$\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

  One-sided: A $100(1-\alpha)\%$ upper (or lower)-confidence bound for μ is

$$\mu \le \bar{X} + Z_{\alpha}\frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \bar{X} - Z_{\alpha}\frac{\sigma}{\sqrt{n}} \le \mu$$

  2. σ is unknown:

  Two-sided: A $100(1-\alpha)\%$ C.I. on μ is given by

$$\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}$$

  One-sided: A $100(1-\alpha)\%$ upper (or lower)-confidence bound for μ is

$$\mu \le \bar{X} + t_{\alpha,n-1}\frac{S}{\sqrt{n}} \quad \text{or} \quad \bar{X} - t_{\alpha,n-1}\frac{S}{\sqrt{n}} \le \mu$$

- **Confidence Interval Estimate for p:** A $100(1-\alpha)\%$ C.I. on p is

$$p_s - Z_{\alpha/2}\sqrt{\frac{p_s(1-p_s)}{n}} \le p \le p_s + Z_{\alpha/2}\sqrt{\frac{p_s(1-p_s)}{n}}$$

where p_s is the sample proportion.

- **Confidence Interval Estimate for Population Total:** Let N be the population size, a $100(1-\alpha)\%$ C.I. for population total T is

$$N\bar{X} - Nt_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \le T \le N\bar{X} + Nt_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

- **Confidence Interval Estimate for Total Difference:** A $100(1-\alpha)\%$ C.I. for total difference D is

$$N\bar{D} - Nt_{\alpha/2,n-1}\frac{S_D}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \le D \le N\bar{D} + Nt_{\alpha/2,n-1}\frac{S_D}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

where $\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i$, $S_D = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \bar{D})^2$.

- **Determining Sample Size:** Let e denote the sampling error,

1. For μ: $n = \frac{Z^2\sigma^2}{e^2}$
2. For p: $n = \frac{Z^2 p(1-p)}{e^2}$

2 Examples

Example 1. The quality control manager at a lightbulb factory needs to estimate the mean life of a large shipment of lightbulbs. The process standard deviation is known be 100 hours. A random sample of 64 lightbulbs indicated a sample mean life of 350 hours.

a. Set up a 95% confidence interval estimate of the true population mean life of lightbulbs in this shipment.

b. Do you think the manufacturer has the right to state that the lightbulbs last an average of 400 hours?

c. Does the population of the lightbulb life have to be normally distributed here? Explain.

d. Explain why an observed value of 320 hours for a single lightbulb is not unusual, even though it is outside the confidence interval you calculated.

a. $\alpha = 0.05,\ n = 64,\ \bar{X} = 350, \sigma = 100$, and $Z_{\alpha/2} = Z_{0.025} = 1.96$.
A 95% C.I. on μ is

$$\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 350 \pm (1.96)\frac{100}{\sqrt{64}} = 350 \pm 24.5$$

$$\text{or} \qquad 325.5 \leq \mu \leq 374.5$$

We are 95% confident that the mean life of lightbulbs in this shipment is between 325.5 and 374.5 hours.

b. No. The manufacturer cannot support a claim that the bulbs last an average of 400 hours. Based on the data from the sample, a mean of 400 hours does not fall in the 95% confidence interval. In fact, a mean of 400 hours would present a distance of 4 standard deviations above the sample mean of 350 hours ($\sigma_{\bar{X}} = 100/\sqrt{64} = 12.5$).

c. No. Since σ is known and $n = 64$, from the **Central Limit Theorem**, we may assume that the sampling distribution of $\bar{X}$ is approximately normal.

d. An individual value of 320 is only 0.30 standard deviations ($0.3\sigma = 0.3 \times 100 = 30$) below the sample mean of 350. The confidence interval represents bounds on the estimate of the mean of a sample of 64, not an individual value.

Example 2. The compressive strength of concrete is being tested by a civil engineer. He tests 12 specimens and obtains the following data.

2216 2237 2249 2204 2225 2301 2281 2263 2318 2255 2275 2295

a. Using one of graphical statistical tools check the assumption that compressive strength is normally distributed. Include a graphical display in your answer.

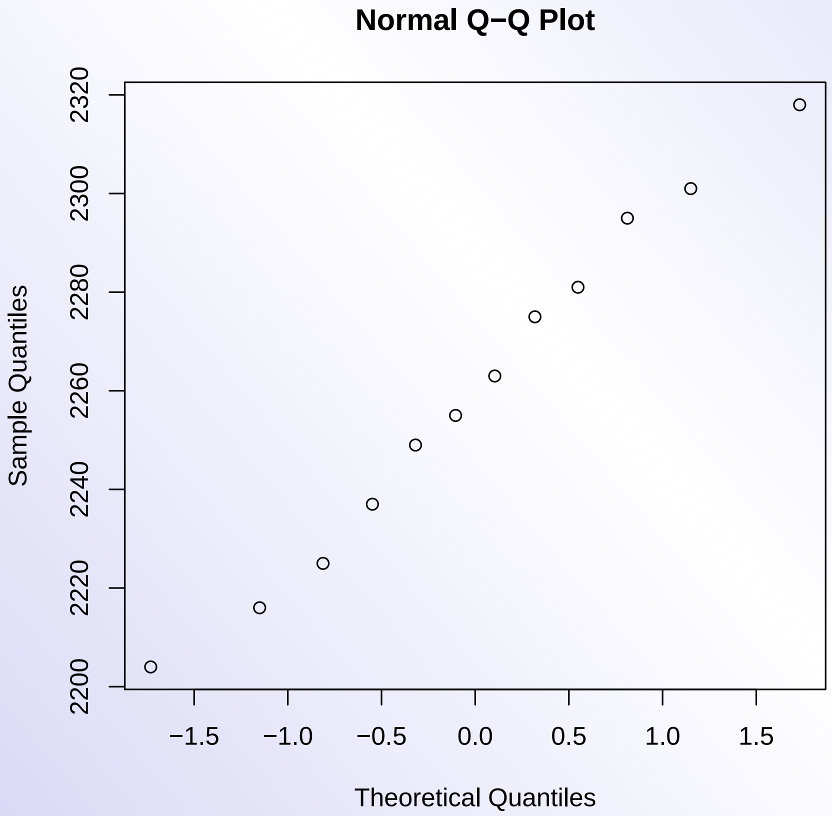
b. Construct 95% two-sided confidence interval on the mean strength.

c. Construct a 95% low-confidence bound on the mean strength. Compare this bound with the lower bound of the two-sided confidence interval and discuss why they are different.

a. The data appear to be normally distributed based on examination of the normal probability plot below.

R code:

```
> a<-c(2216,2237,2249,2204,2225,2301,2281,2263,2318,
       2255,2275,2295)
> qqnorm(a)
```



b. $n = 12, \bar{X} = 2259.9, s = 35.6,\ t_{0.025,11} = 2.201$. A 95% two-sided confidence interval on mean compressive strength is

$$\bar{X} - t_{0.025,11}\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{0.025,11}\frac{S}{\sqrt{n}}$$

$$2259.2 - 2.201\left(\frac{35.6}{\sqrt{12}}\right) \le \mu \le 2259.2 + 2.201\left(\frac{35.6}{\sqrt{12}}\right)$$

$$2237.3 \le \mu \le 2282.5$$

We are 95% confident that the true mean strength is between 2237.3 and 2282.5.

c. A 95% lower-confidence bound on mean strength is

$$\bar{X} - t_{0.05,11}\left(\frac{S}{\sqrt{n}}\right) \le \mu$$

$$2259.9 - 1.796\left(\frac{35.6}{\sqrt{12}}\right) \le \mu \qquad \text{or} \qquad 2241.4 \le \mu$$

We are 95% confident that the true mean strength is not less than 2241.4.

This lower-bound (2241.4) is different from the lower-bound of the two-sided C.I. (2237.3) because the critical values are different in the two cases (in Part b, $t_{0.025,11}$ is used, in Part c, $t_{0.05,11}$ is used).

Example 3. A study conducted by At-A-Glance communications found that 434 office workers out of a sample of 611 responded to e-mail within an hour or two (D. Haralson and S. Ward, "You Have Mail", *USA Today*, May 7, 2002, 1A).

a. Construct a 95% confidence interval for the proportion of workers who responded to e-mail within an hour or two.

b. Interpret the interval constructed in a.

c. If you were to conduct a follow-up study to estimate the population proportion of workers who responded to e-mail within an hour or two within ± 0.01 with 99% confidence interval, how many workers would you interview?

d. In Part c, if the confidence level is reduced to 90%, what is the change in the number of workers you would interview?

a. $\alpha = 0.05$, $Z_{\alpha/2} = Z_{0.025} = 1.96$, $p_s = 434/611 = 0.7103$, and $n = 611$. A 95% C.I. on p is

$$p_s \pm Z_{\alpha/2}\sqrt{\frac{p_s(1-p_s)}{n}} = 0.7103 \pm 1.96\sqrt{\frac{0.7103(1-0.7103)}{611}}$$

$$= 0.7103 \pm 0.03597, \qquad \text{or} \qquad 0.6743 \le p \le 0.7463$$

b. We are 95% confident that the true proportion of workers who responded to e-mail within an hour or two is between 0.6743 and 0.7463.

c. $\alpha = 0.01$, $Z_{\alpha/2} = Z_{0.005} = 2.5758$,

$$n = \frac{Z^2_{\alpha/2}p_s(1-p_s)}{e^2} = \frac{2.5758^2(0.7103)(1-0.7103)}{(0.01)^2} = 13652.58,$$

then we use $n = 13653$.

d $\alpha = 0.1$, $Z_{\alpha/2} = Z_{0.05} = 1.65$,

$$n = \frac{1.65^2(0.7103)(1-0.7103)}{(0.01)^2} = 5602.20,$$

then we use $n = 5603$.