University of California, Los Angeles
Department of Statistics

Statistics 13 **Instructor: Nicolas Christou**

Homework 5 - Solutions

EXERCISE 1
The sample size is $n = 900$. Of these people 600 support this idea. Therefore the sample proportion is $\hat{p} = \frac{x}{n} = \frac{600}{900} \Rightarrow \hat{p} = 0.667$.

a. We want a 95% C.I. $\Rightarrow 1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025 \Rightarrow z_{0.025} = 1.96$. The confidence interval will be:

$$\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

or

$$0.667 - (1.96)\sqrt{\frac{0.667(1-0.667)}{900}} \le p \le 0.667 + (1.96)\sqrt{\frac{0.667(1-0.667)}{900}}$$

Or 0.667 ± 0.031 or $0.636 \le p \le 0.698$. Therefore with 95% confidence the true population proportion p will fall in the inerval $(0.636, 0.698)$.

b. The allowable error of estimation is $E = \pm 2\%$, also $z_{0.025} = 1.96$. Using the information from part (a) we can assume that we have some knowledge on $\hat{p}$. Therefore the sample size is:

$$n = \frac{z^2_{\frac{\alpha}{2}} \hat{p}(1-\hat{p})}{E^2} = \frac{(1.96)^2(0.667)(1-0.667)}{0.02^2} = 2133.2.$$

The City of Los Angeles will need a sample of 2134 voters to obtain a $\pm 2\%$ margin of error with 95% confidence.

c. In Westwood they do not have any prior knowledge on $\hat{p}$ so they use $\hat{p} = 0.5$. The sample size will be:

$$n = \frac{z^2_{\frac{\alpha}{2}} \hat{p}(1-\hat{p})}{E^2} = \frac{(1.96)^2(0.5)(1-0.5)}{0.02^2} = 2401.$$

We observe that the sample size is larger when we use $\hat{p} = 0.5$ (more conservative calculation).

EXERCISE 2

a. These 16 observations have $\bar{x} = 3.611$ and $s^2 = 3.4181 \Rightarrow s = 1.8488$.

b. Confidence intervals for μ:
$90\% \Rightarrow t_{0.05;15=1.753}$
$95\% \Rightarrow t_{0.025;15=2.131}$
$99\% \Rightarrow t_{0.005;15=2.947}$
The corresponding confidence intervals which based on $\bar{x} \pm t_{\frac{\alpha}{2};n-1} \frac{s}{\sqrt{n}}$ are:
90% is 3.611 ± 0.8102,
95% is 3.611 ± 0.9849,
99% is 3.611 ± 1.3621.

c. Let $E_1 = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{16}}$ be the length of the interval (on each side) when the sample size is $n = 16$, and let $E_2 = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ be the width of the interval (on each side) when the sample size is n. We want to find n such that $\frac{E_2}{E_1} = \frac{1}{2}$. Solving for n we have $\frac{\sqrt{16}}{\sqrt{n}} = \frac{1}{2} \Rightarrow n = 64$. Therefore we need to increase the sample size from 16 to 64 (48 additional observations) to halve the width of the interval.

EXERCISE 3
In this problem we do not know the population standard deviation σ. So we are going to use the student's t distribution. First we have to calculate $\bar{x}$ and s (the sample mean and the sample standard deviation), using the sample data.

X	X^2
5.2	27.04
4.4	19.36
4.8	23.04
3.6	12.96
3.1	9.61
6.2	38.44
5.8	33.64
5.1	26.01
3.2	10.24

The sample mean and sample standard deviation are calculated as follows:

$$\sum_{i=1}^{9} x_i = 41.4, \text{ and } \sum_{i=1}^{9} x_i^2 = 200.34$$

$$\bar{x} = \frac{41.4}{9} \Rightarrow \bar{x} = 4.6.$$

$$s = \sqrt{\frac{1}{n-1}[\sum_{i=1}^{9} x_i^2 - \frac{(\sum_{i=1}^{9} x_i)^2}{n}]} = \sqrt{\frac{1}{9-1}[200.34 - \frac{(41.4)^2}{9}]} \Rightarrow s = 1.11.$$

Now, we want to construct a 99% confidence interval for μ: 99% C.I. $\Rightarrow 1-\alpha = 0.99 \Rightarrow \alpha = 0.01 \Rightarrow \frac{\alpha}{2} = 0.005$. Using the t table ($n-1 = 9-1 = 8$ degrees of freedom) we find: $t_{0.005;8} = 3.355$. The confidence interval will be:

$$\bar{x} - t_{\frac{\alpha}{2};n-1}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\frac{\alpha}{2};n-1}\frac{s}{\sqrt{n}}$$

or

$$4.6 - 3.355\frac{1.11}{\sqrt{9}} \leq \mu \leq 4.6 + 3.355\frac{1.11}{\sqrt{9}}$$

Or $3.36 \leq \mu \leq 5.84$. Therefore, with 95% confidence the true population μ, will fall in the interval $(3.36, 5.84)$. Because the sample size is small we must assume that the population from where the sample was taken follows the normal distribution.

EXERCISE 4
We have the following information: $n = 5$, $\bar{x} = 31.7$, and $s = 8.7$.

a. We want to construct a 90% confidence interval for the population mean. Since we are given the sample standard deviation s, we are going to use the t table. $1 - \alpha = 0.90 \Rightarrow$ from the t table $t_4 = 2.132$.

$$\bar{x} - t_{\frac{\alpha}{2};n-1}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\frac{\alpha}{2};n-1}\frac{s}{\sqrt{n}}$$
$$31.7 - 2.132\frac{8.7}{\sqrt{5}} \leq \mu \leq 31.7 + 2.132\frac{8.7}{\sqrt{5}}$$
$$\text{Or } 23.4 \leq \mu \leq 40.0$$

Therefore, with 90% confidence the true population μ will fall in the interval $(23.4, 40.0)$.

b. False. The confidence interval is an interval for the population mean μ. It is not an interval for individual values.

EXERCISE 5

a. We have the following information: $n = 40$, $\bar{x} = 128.6$, and $\sigma = 3$.

$$\bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$
$$128.6 - 1.96\frac{3}{\sqrt{40}} \leq \mu \leq 128.6 + 1.96\frac{3}{\sqrt{40}}$$
$$\text{Or } 127.67 \leq \mu \leq 129.53$$

b. The probability is 0 or 1.

EXERCISE 6
We are given the following information: $n = 50$, $\bar{x} = 115$, and $\sigma = 10$.

a. Construct a 95% confidence interval for the population μ.
$1 - \alpha = 0.95 \Rightarrow z_{\frac{\alpha}{2}} = 1.96$. The confidence interval will be:

$$\bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$
$$115 - 1.96\frac{10}{\sqrt{50}} \leq \mu \leq 115 + 1.96\frac{10}{\sqrt{50}}$$
$$\text{Or } 112.23 \leq \mu \leq 117.77$$

b. We are 95% confident that μ falls in the above interval. Also, you can say that if all possible samples of the same size (here $n = 50$), were taken, 95% of them would include the true population mean μ, and only 5% would not.

EXERCISE 7
We have the following information: $n = 100$, $\bar{x} = 15.35$, and $s = 6.10$.

a. We want to construct a 95% confidence interval for the population mean. Since we are given the sample standard deviation s, we are going to use the t table. $1 - \alpha = 0.95 \Rightarrow$ from the t table $t_{99} \approx 1.984$. The t table does not have entries for $n - 1 = 100 - 1 = 99$, so we can use degrees of freedom 100 as an approximation.

$$\bar{x} - t_{\frac{\alpha}{2};n-1} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\frac{\alpha}{2};n-1} \frac{s}{\sqrt{n}}$$
$$15.35 - 1.984 \frac{6.1}{\sqrt{100}} \leq \mu \leq 15.35 + 1.984 \frac{6.1}{\sqrt{100}}$$
$$\text{Or } 14.14 \leq \mu \leq 16.56$$

Therefore, with 95% confidence the true population μ will fall in the interval $(14.14, 16.56)$.

b. Assume now that $\sigma = 6.1$. We want to find the sample size needed using 95% confidence that the error of estimation will be at most ± 1 minute.

$$n = \left(\frac{z_{\frac{\alpha}{2}} \sigma}{E} \right)^2 = \frac{(1.96)^2 (6.1)^2}{1^2} = 142.95$$

So he must have a sample of 143 people.

EXERCISE 8
We will need to compute the sample mean and sample standard deviation from the data. Here they are:

```
> a <- read.table("http://www.stat.ucla.edu/~nchristo/statistics13/soil.txt", header=TRUE)
> mean(a$lead)
[1] 153.3613
> sd(a$lead)
[1] 111.3201
```

a. The confidence interval for the mean μ is:

$$\bar{x} \pm t_{\frac{\alpha}{2};n-1} \frac{s}{\sqrt{n}}$$

Here because $n = 155$ is large enough we can use $z_{\frac{\alpha}{2}}$ instead of $t_{\frac{\alpha}{2};n-1}$. The 95% confidence interval is:

$$153.36 \pm 1.96 \frac{111.32}{\sqrt{155}} \quad \text{or} \quad 153.36 \pm 17.53.$$

Finallly, $135.83 \leq \mu \leq 170.89$.

b. The problem with this confidence interval is that it does not show the high lead concentration in some areas around the river. There are locations that the lead concentration is well above 400 ppm.