

## Solutions Chapter 3 (Confidence Intervals) – Mathematical Statistics

### Exercise 1

- a. An estimate of the expectation ( $\mu$ ) is the sample mean:  $\bar{x} = 28.25$ .  
An estimate of the variance ( $\sigma^2$ ) is the sample variance:  $s^2 = 14.37$   
(Remark: the notations  $\mu = 28.25$  and  $\sigma^2 = 14.37$  are false is, since  $\mu$  and  $\sigma$  remain unknown!)
- b. Requested: a 95%-CI for  $\mu$  with unknown  $\sigma$  (directly from the formula sheet) has bounds:  $\bar{x} \pm c \frac{s}{\sqrt{n}}$ ,  
where  $c = 2.131$ , such that  $P(T_{15} \geq c) = \frac{\alpha}{2} = 0.025$ , from the  $t_{15}$ -distribution ( $\rightarrow$  *brief notation for the t-distribution with  $df = 15$* )

$$\text{Substitute into the formula: } 95\% - CI(\mu) = \left( 28.25 - 2.131 \cdot \frac{\sqrt{14.37}}{\sqrt{16}}, 28.25 + 2.131 \cdot \frac{\sqrt{14.37}}{\sqrt{16}} \right) \\ = (28.25 - 2.02, 28.25 + 2.02) = (26.23, 30.27)$$

(Note: in the formula we did not use the capitals  $\bar{X}$  and  $S$ , since  $\bar{x}$  and  $s$  are the numerical values in a. after observing the sample variables).

- c. Requested is a 95%-CI for  $\sigma^2$ , so we will use (formula sheet):  
 $\left( \frac{(n-1)s^2}{c_2}, \frac{(n-1)s^2}{c_1} \right)$ , with  $P(\chi_{n-1}^2 \leq c_1) = P(\chi_{n-1}^2 \geq c_2) = \frac{\alpha}{2} = 0.025$

We have  $s^2 = 14.37$  and  $n = 16$  (a.) and in the  $\chi_{16-1}^2$ -table we find  $c_1 = 6.26$  and  $c_2 = 27.49$ .

$$\text{So } 95\% - CI(\sigma^2) = \left( \frac{(n-1)s^2}{c_2}, \frac{(n-1)s^2}{c_1} \right) = \left( \frac{15 \cdot 14.37}{27.49}, \frac{15 \cdot 14.37}{6.26} \right) \approx (7.84, 34.4).$$

### Exercise 2

- a. A 99%-CI for  $\mu$  has bounds:  $\bar{x} \pm c \frac{s}{\sqrt{n}}$ , where  $c = 2.861$  can be found in the  $t$ -table with 19 degrees of freedom, such that  $P(T_{19} \geq c) = \frac{\alpha}{2} = 0.005$ , so  $c \frac{s}{\sqrt{n}} = 2.861 \cdot \frac{39.50}{\sqrt{20}} \approx 25.27$   
 $99\% - CI(\mu) = (189.74 - 25.27, 189.74 + 25.27) = (164.47, 215.01)$
- b. At a confidence level 99% the mean (expected) number of hours of sunshine in the month of July in De Bilt between 164.47 and 215.01. We are quite sure about this statement: this method will produce an interval containing  $\mu$  in 99% of the cases that we will apply the method in equivalent situations, that is, if we repeat the random samples over and over again with equally many observations.
- c. The value 164.1 of the observation  $x$  in the year 1984 should not be compared to confidence interval in a. since the interval concerns  $\mu$ , the **mean number** of hours of sunshine.

(Note that more observations result in a smaller interval and, of course, a relatively smaller proportion of observations within the interval).

If we want to check how exceptional an observation such as  $x = 164.1$  is, we can compute its  $z$ -score:

$$z = \frac{x - \bar{x}}{s} = \frac{164.1 - 189.74}{39.50} \approx -0.65.$$

This value does not at all indicate an exceptionally low observation: it is less than one standard deviation away from the mean.

- d. Requested is the 95%-CI for the standard deviation  $\sigma$ , so we use (formula sheet):

$$\left( \frac{(n-1)s^2}{c_2}, \frac{(n-1)s^2}{c_1} \right), \text{ where } P(\chi_{n-1}^2 \leq c_1) = P(\chi_{n-1}^2 \geq c_2) = \frac{\alpha}{2} = 0.025$$

In this case  $n = 20$  and  $s = 39.50$  are given and in the  $\chi_{20-1}^2$ -table we find  $c_1 = 8.91$  and  $c_2 = 32.85$

$$\text{So a } 95\% - CI(\sigma) = \left( \sqrt{\frac{19s^2}{c_2}}, \sqrt{\frac{19s^2}{c_1}} \right) = \left( \sqrt{\frac{19 \cdot 39.50^2}{32.85}}, \sqrt{\frac{19 \cdot 39.50^2}{8.91}} \right) \approx (30.0, 57.7)$$

### Exercise 3

- a. Using the calculator we find:  $\bar{x} = 60$  and  $s^2 = 51.25$  ( $s \approx 7.159$ ).  
( $\bar{X}$  and  $S^2$  are unbiased estimators of  $\mu$  and  $\sigma^2$ )

- b. We are interested in a confidence interval for  $\mu$ , where in this case the other parameter  $\sigma^2$  is unknown. Assuming a normal distribution for the search times we will use:  $\bar{X} \pm c \frac{s}{\sqrt{n}}$  with  $P(T_{n-1} \geq c) = \frac{1}{2}\alpha$ , where  $n = 9$ ,  $\bar{x} = 60$ ,  $s \approx 7.159$  and in the  $t_8$ -table we find  $c = 2.306$ , such that  $P(T_8 \geq c) = 0.025$ .  
 95%-CI( $\mu$ ) =  $(\bar{x} - c \frac{s}{\sqrt{n}}, \bar{x} + c \frac{s}{\sqrt{n}}) \approx (54.5, 65.5)$
- c. In this part a 95%-CI for  $\sigma^2$  is requested, so we will use (formula sheet):  
 $(\frac{(n-1)S^2}{c_2}, \frac{(n-1)S^2}{c_1})$ , with  $P(\chi_8^2 \leq c_1) = P(\chi_8^2 \geq c_2) = \frac{\alpha}{2} = 0.025$ , such that  $c_1 = 2.18$  and  $c_2 = 17.53$   
 So 95%-CI( $\sigma^2$ ) =  $(\frac{(n-1)S^2}{c_2}, \frac{(n-1)S^2}{c_1}) = (\frac{8 \cdot 51.25}{17.53}, \frac{8 \cdot 51.25}{2.18}) \approx (23.4, 188.1)$

#### Exercise 4

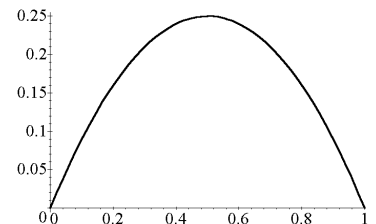
- a. 95%-CI for  $p$ :  $\hat{p} \pm c \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$  with  $\Phi(c) = 1 - \frac{1}{2}\alpha$ , (formula sheet)  
 This is a large-sample-interval since  $n = 100$  is large enough.  
 Furthermore  $\hat{p} = \frac{22}{100} = 0.22$  and  $c = 1.96$  from the  $N(0, 1)$ -table such that  $\Phi(c) = 0.975$   
*(never use the t-distribution if the binomial model applies to the observations!).*  
 Substitute:  $(0.22 - 1.96 \sqrt{\frac{0.22 \cdot 0.78}{100}}, 0.22 + 1.96 \sqrt{\frac{0.22 \cdot 0.78}{100}}) \approx (0.22 - 0.08, 0.22 + 0.08) = (0.14, 0.30)$
- b. The 95%-CI should not be wider than 0.02, so:  $2 \cdot 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq 0.02$ .

This implies:  $\sqrt{n} \geq \frac{1.96}{0.01} \sqrt{\hat{p}(1-\hat{p})}$ , so  $n \geq (\frac{1.96}{0.01})^2 \cdot \hat{p}(1-\hat{p})$ .

Using the observed value in the small sample in a., so  $\hat{p} = 0.22$ , we find  $n \geq 6593$ .

*(Remark 1: if the indication of  $\hat{p}$  in part a. is not given, one could choose a "safe" value  $\frac{1}{2}$  for  $\hat{p}$  or  $\hat{p}(1-\hat{p}) = \frac{1}{4}$ : in that case we find:  $n \geq 9604$ )*

*Remark 2: The inequality  $\hat{p}(1-\hat{p}) \leq \frac{1}{4}$  can be shown by considering the function  $f(p) = p(1-p) = p - p^2$  for  $0 \leq p \leq 1$ : the graph of  $f$  is a parabola that intersects the X-axis at  $p = 0$  and  $p = 1$  and the maximum value of  $f$  is attained at  $p = \frac{1}{2}$ , for which  $f(p) = p(1-p) = \frac{1}{4}$*



#### Exercise 5

- a. This is another binomial situation: Use  $\hat{p} \pm c \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$  with  $\Phi(c) = 1 - \frac{1}{2}\alpha$ , (formula sheet)

In this case  $1 - \alpha = 0.95$  and  $\Phi(c) = 0.975$ , so  $c = 1.96$ ,  $\hat{p} = \frac{73}{400} = 0.1825$  and  $n = 400$ .

Substitution in the formula results in:

$$\begin{aligned} 95\text{-CI}(p) &= \left( 0.1825 - 1.96 \cdot \sqrt{\frac{0.1825 \cdot 0.8175}{400}}, 0.1825 + 1.96 \cdot \sqrt{\frac{0.1825 \cdot 0.8175}{400}} \right) \\ &= (0.1825 - 0.0379, 0.1825 + 0.0379) \approx (0.145, 0.220). \end{aligned}$$

Interpretation: "We are 95% confident that the real population proportion of cars that do not meet the conditions is between 14.5 and 22%."

*The more precise (frequency-)interpretation: "If we would repeat the sample many times (independently), 95% of all computed confidence intervals will contain the population proportion  $p$  of substandard cars."*

b. We want only an upper bound for the proportion  $p$ .

So we should use:  $95\text{-CI}(p) = \left(0, \hat{p} + c \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$ , where  $n = 400$  and  $\hat{p} = \frac{73}{400}$  are the same as in a., but now  $\Phi(c) = 0.95$  (one sided implies one tail probability with area  $\alpha$ ):  $c = 1.645$ .

$$95\text{-CI}(p) = \left(0, 0.1825 + 1.645 \cdot \sqrt{\frac{0.1825 \cdot 0.8175}{400}}\right) \approx (0, 0.214) \quad (\text{or, including } p = 0: [0, 0.214) )$$

Interpretation: "We are 95% confident that the proportion of cars with deficiencies is at most 21.4%."

c. **Estimation error = half of the width:**  $c \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq 0.02$ , it follows that  $n \geq \left(\frac{c}{0.02}\right)^2 \hat{p}(1-\hat{p})$ .

For a confidence level 99% is  $\Phi(c) = 0.995$ , so  $c = 2.575$  and from a. we can use  $\hat{p} = \frac{73}{400}$ .

We find:  $n \geq 2473.1$ , or:  $n = 2474$ .

## Exercise 6

Consider the sample variance  $S^2$  of a random sample  $X_1, \dots, X_n$  drawn from a normal distribution with unknown  $\mu$  and  $\sigma^2$ . The sample size is large:  $n > 25$  (sufficiently large to apply the CLT)

a. According to the CLT  $\frac{(n-1)S^2}{\sigma^2}$  is approximately normally distributed with expectation  $n-1$  and variance  $2(n-1)$ :  $\frac{(n-1)S^2}{\sigma^2} \stackrel{\text{CLT}}{\sim} N(n-1, 2(n-1))$

Hence we have approximately

$$\frac{\frac{(n-1)S^2}{\sigma^2} - (n-1)}{\sqrt{2(n-1)}} \sim N(0,1) \text{ and}$$

$$P\left(-c < \frac{\frac{(n-1)S^2}{\sigma^2} - (n-1)}{\sqrt{2(n-1)}} < c\right) \stackrel{\text{CLT}}{\approx} 1 - \alpha, \text{ where } \Phi(c) = 1 - \frac{1}{2}\alpha.$$

This equivalent to  $P\left((n-1) - c\sqrt{2(n-1)} < \frac{(n-1)S^2}{\sigma^2} < (n-1) + c\sqrt{2(n-1)}\right) \approx 1 - \alpha$

or:  $P\left(\frac{(n-1)S^2}{(n-1) + c\sqrt{2(n-1)}} < \sigma^2 < \frac{(n-1)S^2}{(n-1) - c\sqrt{2(n-1)}}\right) \approx 1 - \alpha$  (assuming that  $(n-1) - c\sqrt{2(n-1)} > 0$ )

b. Substituting  $n = 101$ ,  $s^2 = 50$  and  $c = 1.96$  ( $\Phi(c) = 0.975$ ) in the formula of a. we find:

$$95\text{-CI}(\sigma^2) \approx \left(\frac{100 \cdot 50}{100 + 1.96\sqrt{200}}, \frac{100 \cdot 50}{100 - 1.96\sqrt{200}}\right) \approx (39.15, 69.17).$$

c.  $95\text{-CI}(\sigma^2) = \left(\frac{(n-1)S^2}{c_2}, \frac{(n-1)S^2}{c_1}\right)$ , where  $n = 101$ ,  $s^2 = 50$  and

$$P(\chi_{100}^2 \leq c_1) = P(\chi_{100}^2 \geq c_2) = \frac{\alpha}{2} = 0.025, \text{ so that } c_1 = 74.22 \text{ and } c_2 = 129.56:$$

$$95\text{-CI}(\sigma^2) = \left(\frac{100 \cdot 50}{129.56}, \frac{100 \cdot 50}{74.22}\right) \approx (38.59, 67.37)$$

(Note: considering the differences of the results in b. and c. it seems to be better to use the  $\chi^2$ -tables, and linear interpolation, if  $df \leq 100$ )

### Exercise 7

a.  $\bar{x} = 12.6$  is an estimate of  $E(X) = \frac{1}{\lambda}$ , so  $\frac{1}{\bar{x}} = \frac{1}{12.6} \approx 0.0794$  is an estimate of  $\lambda$ .

*Note: In chapter 2 we showed that  $1/\bar{X}$  is the maximum likelihood estimator (mle).*

b. Construction of a confidence interval of  $\lambda$ :

$\bar{X}$  is according to the CLT approximately  $N\left(\mu, \frac{\sigma^2}{50}\right)$  with  $E(X) = \frac{1}{\lambda}$ , then approximately  $\frac{\bar{X} - \frac{1}{\lambda}}{\sqrt{\frac{\sigma^2}{50}}} \sim N(0,1)$ .

$\sigma^2$  is unknown, but we can approximate  $\sigma^2$  with  $S^2$ . (In this case we have  $\sigma^2 = \frac{1}{\lambda^2}$ , therefore an alternative estimator would be  $\bar{X}^2$ ! Check that  $\bar{x}^2 \approx s^2$ .)

We can use the  $N(0,1)$ -distribution to choose the value of  $c$  such that  $P(-c < Z < c) = 1 - \alpha$

$$P\left(-c < \frac{\bar{X} - \frac{1}{\lambda}}{\sqrt{\frac{S^2}{50}}} < c\right) \approx 1 - \alpha$$

$$P\left(-c \sqrt{\frac{S^2}{50}} < \bar{X} - \frac{1}{\lambda} < c \sqrt{\frac{S^2}{50}}\right) \approx 1 - \alpha$$

$$P\left(\bar{X} - c \sqrt{\frac{S^2}{50}} < \frac{1}{\lambda} < \bar{X} + c \sqrt{\frac{S^2}{50}}\right) \approx 1 - \alpha$$

$$P\left(\bar{X} - c \sqrt{\frac{S^2}{50}} < \frac{1}{\lambda} < \bar{X} + c \sqrt{\frac{S^2}{50}}\right) \approx 1 - \alpha$$

$$P\left(\frac{1}{\bar{X} + c \sqrt{\frac{S^2}{50}}} < \lambda < \frac{1}{\bar{X} - c \sqrt{\frac{S^2}{50}}}\right) \approx 1 - \alpha$$

(If  $\bar{X} - c \sqrt{\frac{S^2}{50}} < 0$ , the upper bound should be  $\infty$ .)

An alternative (better?) approach is to use that  $\sigma^2 = \frac{1}{\lambda^2}$ , so that  $\bar{X}$  is approximately  $N\left(\frac{1}{\lambda}, \frac{1}{\lambda^2 n}\right)$ .

Then we can solve  $\lambda$  from the inequality in the following (approximate) probability statement:

$$P\left(-c < \frac{\bar{X} - \frac{1}{\lambda}}{\sqrt{\frac{1}{\lambda^2 n}}} < c\right) \approx 1 - \alpha, \quad \text{where } \Phi(c) = 1 - \frac{1}{2}\alpha$$

$$\Leftrightarrow P(-c < (\lambda \bar{X} - 1)\sqrt{n} < c) \approx 1 - \alpha$$

$$\Leftrightarrow P(-c/\sqrt{n} < (\lambda \bar{X} - 1) < c/\sqrt{n}) \approx 1 - \alpha,$$

$$\Leftrightarrow P\left(1 - \frac{c}{\sqrt{n}} < \lambda \bar{X} < 1 + \frac{c}{\sqrt{n}}\right) \approx 1 - \alpha,$$

$$\Leftrightarrow P\left(\frac{1}{\bar{X}}\left(1 - \frac{c}{\sqrt{n}}\right) < \lambda < \frac{1}{\bar{X}}\left(1 + \frac{c}{\sqrt{n}}\right)\right) \approx 1 - \alpha$$

c.  $n = 50, \bar{x} = 12.6, s^2 = 150.3$  and  $c = 1.96$  if  $P(-c < Z < c) = 0.95$

$$95\% - BI(\lambda) = \left( \frac{1}{\bar{X} + c\sqrt{S^2/50}}, \frac{1}{\bar{X} - c\sqrt{S^2/50}} \right) \approx (0.0625, 0.1087)$$

(the alternative formula results in the (slightly wider) interval (0.057, 1.014))

**Exercise 8:** Analysis of the problem to find a correct approach:

a. We want to know what proportion ( $p$ ) of the 2250 visitors are buyers. Firstly, on the basis of the sample of  $n = 75$  visitors, we have to determine an interval estimate of  $p$  (do not use  $n = 2250$ ), and then we can estimate  $2250 \times p$ . Note that from the construction of the  $CI(p)$  it follows:

$$P(L < p < U) = 1 - \alpha \iff P(a \cdot L < a \cdot p < a \cdot U) = 1 - \alpha \quad (a = 2250)$$

b. The turnover = the number of buyers (1350) times  $\mu$ :  $\mu =$  "the expected turnover per buyer".

So first we determine a  $CI$  of  $\mu$  on the basis of a sample of  $n = 45$  buyers (60% of the 75 visitors).

Solutions:

a. Binomial Model:  $X =$  "the number of buyers in a random sample of  $n = 75$  visitors."

$X$  is  $B(75, p)$ -distributed with unknown  $p =$  "proportion of buyers". (Formula sheet:)

$$95\%-CI(p) = \left( \hat{p} - c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right) = \left( 0.60 - 1.96\sqrt{\frac{0.6 \times 0.4}{75}}, 0.60 + 1.96\sqrt{\frac{0.6 \times 0.4}{75}} \right)$$

$$\approx (0.4891, 0.7109), \text{ where } n = 75 \text{ (visitors), } \hat{p} = 0.60 \text{ (so } x = 45) \text{ and } c = 1.96.$$

Then for the expected number of buyers on a day with 2250 visitors we have:

$$95\%-CI(2250p) \approx (2250 \times 0.4891, 2250 \times 0.7109) \approx (1100, 1600)$$

("We are 95% confident that the number of buyers on a day with 2250 visitors is between 1100 and 1600")

b. As stated in a, we assume that  $X_1, \dots, X_{45}$  are the turnovers of the 45 buyers in the sample:

$X_1, \dots, X_{45}$  are independent and (approximately) normally distributed with unknown  $\mu$  and  $\sigma^2$ .

$\mu =$  the "mean" (expected) turnover of all (potential) buyers.

$$\text{Formula sheet: } 95\%-CI(\mu) = \left( \bar{X} - c \cdot \frac{s}{\sqrt{n}}, \bar{X} + c \cdot \frac{s}{\sqrt{n}} \right)$$

Where  $n = 45, \bar{X} = 40, s = 10$  and  $c = 2.021$  from the  $t_{44}$ -table (use the  $t_{40}$ -table):  $P(T_{40} \geq c) = 0.025$

$$95\%-CI(\mu) \approx (36.99, 43.01)$$

An interval estimate of the total turnover  $1350\mu$  is:

$$95\%-CI(1350\mu) = (1350 \times 36.99, 1350 \times 43.01) = (49937, 58064)$$

("At a 95% level of confidence the total turnover of the shop on a day with 1350 buyers is between 50 and 58 thousand Euro").

c. As stated in a, we assume that  $X_1, \dots, X_{45}$  are the observed turnovers of the 45 buyers in the sample: the (new) buyer spends  $X_{46}$ .

Model:  $X_1, \dots, X_{45}, X_{46}$  are independent and (approximately) normally distributed with unknown  $\mu$  and  $\sigma^2$ .

a 90%-PI for a new observation  $X_{46}$ , based on the observed  $X_1, \dots, X_{45}, X_{17}$  has bounds:  $\bar{x} \pm cs\sqrt{1 + \frac{1}{n}}$ ,

where  $n = 45, \bar{x} = 40, s^2 = 100$  and  $c = 1.681$  (interpolation between the values at  $df = 40$  and  $df = 50$ , such that  $P(T_{44} \geq c) = \frac{\alpha}{2} = 0.05$ , from the table of the  $t_{44}$ -distribution.

$$90\%-PI(X_{46}) \approx (23.00, 57.00)$$

("At a 90%-level of confidence the turnover of buyer 1000 is between 23 and 57 Euro").

**Exercise 9**

a. For a  $(1-\alpha)100\%$ -confidence interval of  $p$  the probability statement is (for sufficiently large  $n$ ):

$$P\left(-c < \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} < c\right) \stackrel{\text{CLT}}{\approx} 1 - \alpha \quad (c \text{ such that } \Phi(c) = 1 - \frac{1}{2}\alpha.)$$

Solving  $(\hat{p} - p)^2 < c^2 \frac{p(1-p)}{n}$  with respect to  $p$  (e.g. first solving the equality):

$$\text{Equality: } \hat{p}^2 - 2\hat{p} \cdot p + p^2 = \frac{c^2}{n} \cdot p - \frac{c^2}{n} \cdot p^2$$

$$\Leftrightarrow \left[1 + \frac{c^2}{n}\right] \cdot p^2 - \left(2\hat{p} + \frac{c^2}{n}\right) \cdot p + \hat{p}^2 = 0 \quad (\text{the left hand side is parabola that opens upward:}$$

*solutions of  $p$  for the inequality lie between the two zero's)*

$$\Leftrightarrow p_{1,2} = \frac{\left(2\hat{p} + \frac{c^2}{n}\right) \pm \sqrt{\left(2\hat{p} + \frac{c^2}{n}\right)^2 - 4\left[1 + \frac{c^2}{n}\right] \cdot \hat{p}^2}}{2\left[1 + \frac{c^2}{n}\right]} = \frac{\left(2\hat{p} + \frac{c^2}{n}\right) \pm \sqrt{\left(2\hat{p} + \frac{c^2}{n}\right)^2 - 4\left[1 + \frac{c^2}{n}\right] \cdot \hat{p}^2}}{2\left[1 + \frac{c^2}{n}\right]}$$

$$\Leftrightarrow p = \frac{\left(2\hat{p} + \frac{c^2}{n}\right) \pm \sqrt{\left(2\hat{p} + \frac{c^2}{n}\right)^2 - 4\left[1 + \frac{c^2}{n}\right] \cdot \hat{p}^2}}{2\left[1 + \frac{c^2}{n}\right]} = \frac{\left(2\hat{p} + \frac{c^2}{n}\right) \pm \sqrt{\left(2\hat{p} + \frac{c^2}{n}\right)^2 - 4\left[1 + \frac{c^2}{n}\right] \cdot \hat{p}^2}}{2\left[1 + \frac{c^2}{n}\right]}$$

$$\Leftrightarrow p = \frac{\left(2\hat{p} + \frac{c^2}{n}\right) \pm \sqrt{4\hat{p}^2 + \frac{4c^2}{n}\hat{p} + \frac{c^4}{n^2} - 4\hat{p}^2 - \frac{4c^2}{n}\hat{p}^2}}{2\left[1 + \frac{c^2}{n}\right]} = \frac{\left(2\hat{p} + \frac{c^2}{n}\right) \pm \sqrt{\frac{4c^2}{n} \cdot \left[\hat{p}(1-\hat{p}) + \frac{c^2}{4n}\right]}}{2\left[1 + \frac{c^2}{n}\right]}$$

$$\Leftrightarrow p = \frac{n}{n+c^2}\hat{p} + \frac{c^2}{2(n+c^2)} \pm \frac{n}{n+c^2}c\sqrt{\frac{\hat{p}(1-\hat{p}) + \frac{c^2}{4n}}{n}}$$

$$= \frac{n}{n+c^2} \left[ \left(\hat{p} + \frac{c^2}{2n}\right) \pm c\sqrt{\frac{\hat{p}(1-\hat{p}) + \frac{c^2}{4n}}{n}} \right]$$

For  $n = 1200$ ,  $c = 1.96$  (95% confidence level) and the observed  $\hat{p} = \frac{300}{1200} = 0.25$  we find:  
 $0.25080 \pm 0.02447$ , so 95%-BI( $p$ )  $\approx$  **(0.2263, 0.2753)**

Comparing this interval to the “standard” interval on the formula sheet with bounds  $\hat{p} \pm c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$  we find:  
 95%-CI( $p$ )  $\approx$   $(0.25 - 0.0245, 0.25 + 0.0245) =$  **(0.2255, 0.2745)** (rounded at 3 decimals the same result).

### Exercise 10

- a. Requested: a 90%-PI for a new observation  $X_{17}$  (assuming now that  $X_1, \dots, X_{16}, X_{17}$  is a random sample from a normal distribution with unknown  $\mu$  and  $\sigma^2$  (directly from the formula sheet) has bounds:  $\bar{x} \pm cs\sqrt{1 + \frac{1}{n}}$ , where  $n = 16$ ,  $\bar{x} = 189.74$ ,  $s^2 = 14.37$  and  $c = 1.753$ , such that  $P(T_{15} \geq c) = \frac{\alpha}{2} = 0.05$ , from the table of the  $t_{15}$ -distribution.

Substitution into the formula:

$$90\%-PI(X_{17}) = \left( 28.25 - 1.753 \cdot \sqrt{14.37} \sqrt{1 + \frac{1}{16}}, 26.25 + 1.753 \cdot \sqrt{14.37} \sqrt{1 + \frac{1}{16}} \right) \\ = (28.25 - 6.85, 28.25 + 6.85) = (21.40, 35.10)$$

- b. 90% of the 16 observations are expected to be in the interval: 10% (1.6 observations) outside. In this case none of the observations is outside the interval, which is not strange: if we have 16 trials and the success rate is 10%, the probability of no success is  $0.9^{16} = 18.5\%$ .

- c. Requested: a 95%-CI with only a lower bound for  $\mu$  with unknown  $\sigma$  (using the formula sheet)

$$\left( \bar{x} - c \frac{s}{\sqrt{n}}, \infty \right),$$

where  $c = 1.753$ , such that  $P(T_{15} \geq c) = \alpha = 0.05$ , from the  $t_{15}$ -distribution.

$$\text{Substitute into the formula: } 95\% - CI(\mu) = \left( 28.25 - 1.753 \cdot \frac{\sqrt{14.37}}{\sqrt{16}}, \infty \right)$$

$$\approx (26.6, \infty)$$

(“The expected pressure resistance is at least 26.6, at a 95%-confidence level”).

### Exercise 11

- a. Requested: a 90%-PI for a new observation  $X_{201}$  (assuming that  $X_1, \dots, X_{200}, X_{201}$  is a random sample from a normal distribution with unknown  $\mu$  and  $\sigma^2$ ) has bounds:  $\bar{x} \pm cs \sqrt{1 + \frac{1}{n}}$ ,

where  $n = 200$ ,  $\bar{x} = 47.3$ ,  $s = 4.0$  and  $c = 1.645$ , such that  $P(T_{199} \geq c) = \frac{\alpha}{2} = 0.05$ , from the table of the  $t_{\infty}$ -distribution, or use the approximate standard normal distribution:  $\Phi(c) = 0.95$

Then the bounds are  $47.3 \pm 1.645 \cdot 4.0 \sqrt{1 + \frac{1}{200}}$ , so  $90\%-PI(X_{17}) \approx (40.7, 53.9)$

(We are 90% confident that John works between 40.7 and 53.9 hours per week).

- b. Using the result of exercise 6a:  $CI(\sigma^2) \stackrel{CLT}{\approx} \left( \frac{(n-1)s^2}{(n-1)+c\sqrt{2(n-1)}}, \frac{(n-1)s^2}{(n-1)-c\sqrt{2(n-1)}} \right)$ , where  $\Phi(c) = 1 - \frac{1}{2}\alpha$ .

We have  $n = 200$ ,  $s = 4.0$  and  $c = 1.645$ , since  $\Phi(c) = 1 - \frac{1}{2}\alpha = 0.95$ .

$$90\% - CI(\sigma) = \left( \sqrt{\frac{199 \cdot 4.0^2}{199 + 1.645\sqrt{398}}}, \sqrt{\frac{199 \cdot 4.0^2}{199 - 1.645\sqrt{398}}} \right) \approx (3.71, 4.38)$$

### Exercise 12

Copying the approach in section 3.5:

- We assume that all observations are independent and all  $N(\mu, \sigma^2)$ . In particular,  $\bar{X}$  and  $\bar{Y}$  are independent.

- $\bar{X}$  serves as a prediction of  $\bar{Y}$ :  $\bar{Y} - \bar{X}$  is the prediction error, with

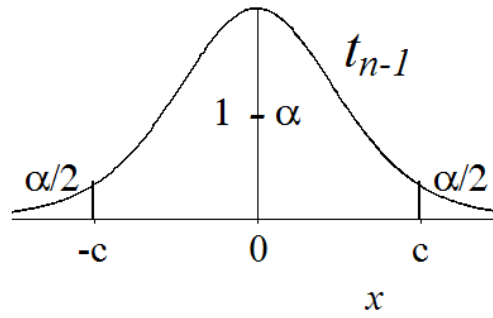
Expectation  $E(\bar{Y} - \bar{X}) = E(\bar{Y}) - E(\bar{X}) = \mu - \mu = 0$  and

and variance  $var(\bar{Y} - \bar{X}) \stackrel{\text{ind.}}{=} var(\bar{Y}) + var(\bar{X}) = \frac{\sigma^2}{m} + \frac{\sigma^2}{n} = \sigma^2 \left( \frac{1}{m} + \frac{1}{n} \right)$

$$\text{So: } X_{n+1} - \bar{X} \sim N\left(0, \sigma^2 \left( \frac{1}{m} + \frac{1}{n} \right)\right) \quad \text{and} \quad \frac{\bar{Y} - \bar{X}}{\sqrt{\sigma^2 \left( \frac{1}{m} + \frac{1}{n} \right)}} \sim N(0, 1)$$

Substituting  $S^2$  for  $\sigma^2$  results in a  $t$ -distributed “pivot” with  $n - 1$  degrees of freedom.

(Verify:  $T = \frac{Z}{\sqrt{W/n}} = \frac{\bar{Y} - \bar{X}}{\sqrt{S^2 \left( \frac{1}{m} + \frac{1}{n} \right)}} \sim t_{n-1}$  if we choose  $Z = \frac{\bar{Y} - \bar{X}}{\sqrt{\sigma^2 \left( \frac{1}{m} + \frac{1}{n} \right)}} \sim N(0, 1)$  and  $W = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$ )



$$P\left(-c < \frac{\bar{Y} - \bar{X}}{\sqrt{S^2\left(\frac{1}{m} + \frac{1}{n}\right)}} < c\right) = 1 - \alpha$$

$$\Leftrightarrow P\left(-c\sqrt{S^2\left(\frac{1}{m} + \frac{1}{n}\right)} < \bar{Y} - \bar{X} < c\sqrt{S^2\left(\frac{1}{m} + \frac{1}{n}\right)}\right) = 1 - \alpha$$

$$\Leftrightarrow P\left(\bar{X} - c\sqrt{S^2\left(\frac{1}{m} + \frac{1}{n}\right)} < \bar{Y} < \bar{X} + c\sqrt{S^2\left(\frac{1}{m} + \frac{1}{n}\right)}\right) = 1 - \alpha$$

$$(1 - \alpha)100 - PI(\bar{Y}) = \left(\bar{X} - c\sqrt{S^2\left(\frac{1}{m} + \frac{1}{n}\right)}, \bar{X} + c\sqrt{S^2\left(\frac{1}{m} + \frac{1}{n}\right)}\right)$$

Note that for  $m = 1$  this formula gives the prediction interval for one observation and for  $m \rightarrow \infty$  we find the confidence interval of  $\mu$  ( $\bar{Y}$  converges in probability to  $\mu$ ).