M.Sc. in ‚Transportation Systems‘

Applied Statistics in Transport

Exercises: Confidence Intervals

Population Variance known

1. In what interval will the mean of a sample of 25 observations from a N(2, σ=4) population lie in 95 percent of the time in repeated sampling?

Solution: $\bar{x} = 2, \sigma = 4, n = 25, z_{(\alpha/2)} = z_{(0.025)} = 1.96$

$\bar{x} - z_{(\alpha/2)} * \sigma/\sqrt{n} \le \mu \le \bar{x} + z_{(\alpha/2)} * \sigma/\sqrt{n}$, $2 - 1.96 * 4/\sqrt{25} \le \mu \le 2 + 1.96 * 4/\sqrt{25}$

$$0.4320 \le \mu \le 3.5680$$

Answer: The sample mean will fall in the interval $(0.4320, 3.5680)$ in 95% of the time during sampling.

2. In how many units of the unknown mean the sample mean lies in 99% of the time if the sample size is 9 and the population is N(μ, σ=1.5)?

Solution: $\bar{x}\ unknown, \sigma = 1.5, n = 9, z_{(\alpha/2)} = z_{(0.005)} = 2.576$

Scale of z: $-z_{(0.005)} \le \mu \le +z_{(0.005)}$, $\mu = 0$, $-2.576 \le 0 \le +2.576$

Scale of $\bar{\mathrm{X}}$: $\bar{\mathrm{x}} - \mathrm{z}_{(\alpha/2)} * \sigma/\sqrt{\mathrm{n}} \le \mu \le \bar{\mathrm{x}} + \mathrm{z}_{(\alpha/2)} * \sigma/\sqrt{\mathrm{n}}$,

$\mu - 2.576 * 1.5/\sqrt{9} \le \mu \le \mu + 2.576 * 1.5/\sqrt{9}\,, \mu - 1.2880 \le \mu \le \mu + 1.2880$

Answer: The mean of a sample of size 9 will fall within 1.288 units of the true mean in 99% of all random samples when sampling from N(μ,1.5).

3. Find a 90% confidence interval for the population mean in a sample of size 25 from N(μ, σ=3) using the sample mean equal to 11.

Solution: $\bar{x} = 11, \sigma = 3, n = 25, z_{(\alpha/2)} = z_{(0.05)} = 1.645$ (90%=100(1-α)% implies α=0.10, so α/2=0.05 and $z_{.05}$=1.645):

Scale of z: $-z_{(0.05)} \le \mu \le +z_{(0.05)}$, $\mu = 0$, $-1.645 \le 0 \le +1.645$

Scale of $\bar{X}$: $\bar{x} - z_{(\alpha/2)} * \sigma/\sqrt{n} \le \mu \le \bar{x} + z_{(\alpha/2)} * \sigma/\sqrt{n}$,

$11 - 1.645 * 3/\sqrt{25} \le \mu \le 11 + 1.645 * 3/\sqrt{25}\,, 10.0130 \le \mu \le 11.9870$

Answer: The population mean will fall in the interval $(10.01, 11.99)$ with a probability of 90%.

4. Speed measurements for a random sample of 50 cars on a specific road section have shown an average speed of $\bar{x} = 80km/h$. The variance of speed on this road section is known from various studies that were done in the past: $\sigma^2 = 100km^2/h^2$. What is the 95% confidence interval for the expected value of speed μ?

Solution: $\bar{x} = 80, \sigma = 10, n = 50, z_{(\alpha/2)} = z_{(0.025)} = 1.96$ (95%=100(1-α)% implies α=0.05, so α/2=0.025 and $z_{.025}$=1.96):

Scale of z: $-z_{(0.025)} \leq \mu \leq +z_{(0.025)}$, $\mu = 0$, $-1.96 \leq 0 \leq +1.96$

Scale of $\bar{X}$: $\bar{x} - z_{(\alpha/2)} * \sigma/\sqrt{n} \leq \mu \leq \bar{x} + z_{(\alpha/2)} * \sigma/\sqrt{n}$,

$80 - 1.96 * 10/\sqrt{50} \leq \mu \leq 80 + 1.96 * 10/\sqrt{50}, 77.2281 \leq \mu \leq 82.7719$

Answer: The expected value of speeds μ lies with a probability of 95% in the interval 77.2 to 82.8 km/h.

Population Variance unknown

5. Continuation of 4: Speed measurements for a random sample of 50 cars on a specific road section have shown an average speed of $\bar{x} = 80km/h$. We assume now that the population variance is unknown and has to be estimated by the sample variance. The estimated sample variance is $s^2 = 100km^2/h^2$. What is the 95% confidence interval for the expected value of speeds μ? Compare the results with the answers from exercise 4.

Solution: we assume appr. normal distribution (central limit theorem) or the t-distribution (population variance unknown): $\bar{x} = 80, s = \hat{\sigma} = 10, n = 50$

Solution normal distribution: $z_{(\alpha/2)} = z_{(0.025)} = 1.96$ (95%=100(1-α)% implies α=0.05, so α/2=0.025 and $z_{.025}$=1.96):

Scale of z: $-z_{(0.025)} \leq \mu \leq +z_{(0.025)}$, $\mu = 0$, $-1.96 \leq 0 \leq +1.96$

Scale of $\bar{X}$: $\bar{x} - z_{(\alpha/2)} * \hat{\sigma}/\sqrt{n} \leq \mu \leq \bar{x} + z_{(\alpha/2)} * \hat{\sigma}/\sqrt{n}$,

$80 - 1.96 * 10/\sqrt{50} \leq \mu \leq 80 + 1.96 * 10/\sqrt{50}, 77.2281 \leq \mu \leq 82.7719$

Solution t-distribution: 95%=100(1-α)% implies α=0.05, so α/2=0.025, n=50, $t_{.975}$=2.01 with 49=50-1 degrees of freedom:

$$80 - 2.01 * 10/\sqrt{50} \leq \mu \leq 80 + 2.01 * 10/\sqrt{50}, 77.1574 \leq \mu \leq 82.8426$$

Hence the normal distribution would underestimate the confidence interval but they are close:

Normal distribution: 82.7719 -77.2281=5.5438

t-distribution: 82.8426-77.1574=5.6852

5.5438*100/5.6852=97.51284: the confidence interval from the normal distribution is 97.5% of the confidence interval we get when we use the t-distribution, it underestimates the interval by about 2.5%.

6. 16 holes were bored to check the thickness of the road surface, they showed an average thickness of $\bar{x} = 3\ cm$; the sample standard deviation was $s = 0.5\ cm$. Does the requested value $\mu_r = 3.5\ cm$ lie in the confidence interval that includes the true mean thickness of the road surface with a 95% probability?

Solution t-distribution: 95%=100(1-α)% implies α=0.05, so α/2=0.025, n=16, $t_{.975}$=2.13 with 15=16-1 degrees of freedom:

$$3.0 - 2.13 * 0.5/\sqrt{16} \leq \mu \leq 3.0 + 2.13 * 0.5/\sqrt{16}, 2.7338 \leq \mu \leq 3.2662$$

The confidence interval calculated from the sample does not include the requested value μ_r. Hence the thickness of the road surface is not sufficient / satisfactory.

Solution normal distribution from 5.: $\bar{x} = 3, s = \hat{\sigma} = 0.5, n = 16, z_{(\alpha/2)} = z_{(0.025)} = 1.96$ (95%=100(1-α)% implies α=0.05, so α/2=0.025 and $z_{.025}$=1.96):

Scale of z: $-z_{(0.025)} \le \mu \le +z_{(0.025)}$, $\mu = 0$, $-1.96 \le 0 \le +1.96$

$$3.0 - 1.96 \ * 0.5/\sqrt{16} \le \mu \le 3.0 + 1.96 \ * 0.5/\sqrt{16}, 2.7550 \le \mu \le 3.2450$$

For n=100:

Solution t-distribution: 95%=100(1-α)% implies α=0.05, so α/2=0.025, n=50, t.975=1.98 with 99=100-1 degrees of freedom:

$$3.0 - 1.98 \ * 0.5/\sqrt{16} \le \mu \le 3.0 + 1.98 \ * 0.5/\sqrt{16}, 2.7525 \le \mu \le 3.2475$$

Confidence intervals proportion

7. Household income: A random sample of 600 households resulted in 120 households with a monthly income of less than 800 Euro. What is the 99% confidence interval for this proportion of all households in the area under investigation?

Solution: $\overline{p}_s = \frac{120}{600} = 0.2,\ n = 600, z_{(\alpha/2)} = z_{(0.005)} = 2.575$

As the condition $\overline{p}_s * \overline{q}_s * n \ge 9$ is met ($0.2 * 0.8 * 600 = 96 > 9$), we can use the normal approximation for the binomial distribution (dbinom(X,600,0.2)) for determining this interval.

The standard error for 20% gives: $\hat{\sigma}_{\%} = \sqrt{\frac{\overline{p}_s * \overline{q}_s}{n}} = \sqrt{\frac{0.2*0.8}{600}} = 0.0163,$

(99%=100(1-α)% implies α=.01, so α/2=.005 and $z_{.005}$=2.575:

Scale of z: $-z_{(0.025)} \le \mu \le +z_{(0.025)}$, $\mu = 0$, $-2.575 \le 0 \le +2.575$

$\overline{p}_s - z_{(\alpha/2)} * \hat{\sigma}_{\%} \le \overline{p} \le \overline{p}_s + z_{(\alpha/2)} * \hat{\sigma}_{\%}$

$0.2 - 2.575 * 0.0163 \le \overline{p} \le 0.2 + 2.575 * 0.0163,\ \ 0.1580 \le \overline{p} \le 0.2420$

Or in other notation: $\Delta_{crit} = \overline{p}_s \pm z_{(\alpha/2)} * \hat{\sigma}_{\%} = 0.35 \pm 2.575 * 0.0213 = 0.2 \pm 0.0420$

Answer: In the range of 15.8% to 24.2% lie 99% of all population parameters that can have "produced" the sample parameter $\overline{p}_s$=20%.

Sample size

8. The average speed on a highway in the morning peak hour should be determined. A preliminary study showed $\overline{x} = 91 \frac{km}{h}; s = 28km/h$. What is the minimal sample size if the estimated average speed should not differ more than 5km/h (5 km/h less or 5 km/h more) from the true average speed with a probability of 0.955?

Solution: E_a=5km/h

$CI = 0.955, \alpha = 0.045, \frac{\alpha}{2} = 0.0225, z_{(0.0225)} = -2.01, z_{(0.9775)} = 2.01$

$E_a = 5km/h; rci = 2 * E_a = 10$

$n \ge \frac{{z_{(\alpha/2)}}^2 * \hat{\sigma}^2}{{E_a}^2} = \frac{2.01^2 * 28^2}{5^2} = 126.6975$

The minimal sample size is 127.

Or with rci: $n = \frac{4 * {z_{(\alpha/2)}}^2 * \hat{\sigma}^2}{rci^2} = \frac{4 * 2.01^2 * 28^2}{10^2} = 126.6975$