Statistik II - Exercise session 5
10.12.2014 & 17.12.2014

Info

- Classroom: SPA1 203
- Time: Wednesdays, 16:15 - 17:45
- in English
- Assignments on webpage (lvb>staff>PB)

Contact: Petra Burdejova
petra.burdejova@hu-berlin.de
Office: SPA1 R400 (upon agreement)

Schedule:

Date	Week	Exercises
15.10.14	E1	4-11, 4-18, 4-19
22.10.14	E1	4-11, 4-18, 4-19
29.10.14	E2	5-6, 5-10, 5-11, 5-14 ,5-16
5.11.14	E2	5-6, 5-10, 5-11, 5-14 ,5-16
12.11.14	E3	5-16, 5-21, 5-23, 6-1
19.11.14	E3	5-16, 5-21, 5-23, 6-1
26.11.14	E4	6-3, 6-9, 6-13, 7-1
03.12.14	E4	6-3, 6-9, 6-13, 7-1
10.12.14	E5	7-3, 7-5, 7-26, 7-45
17.12.14	E5	7-3, 7-5, 7-26, 7-45
07.01.15	E6	8-1, 8-4, 8-7, 8-11
14.01.15	E6	8-1, 8-4, 8-7, 8-11
21.01.15	E7	TBA
28.01.15	–	–
04.02.15	E8	Review for exam
11.02.15	E8	Review for exam

Review

- week 9 & week 10
- Slides:
  - Theory of sampling (05_Stichprobentheorie)
  - Estimation procedures (06_Schätztheorie)

1 Sample distribution

1.1 Sample distribution of sample mean

Distribution of $\overline{X}$				
Population	σ^2	R.v.	Distr.	Condition
$X_i \sim N(\mu;\sigma)$	known	$Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$	$N(0,1)$	
	unknown	$T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$	$t(n-1)$	for $n \leq 30$
			$\approx N(0,1)$	for $n > 30$
Unknown distribution	known	$Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$	$\approx N(0,1)$	for $n > 30$
	unknown	$T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$	$\approx N(0,1)$	for $n > 30$

1.2 Sample distribution of Sample proportion

Sample function: $\hat{\Pi} = \frac{X}{n}$

(Example: Smokers in Berlin, X - number of smokers in our sample.)

Distribution of simple random sample

$X \sim B(n;\pi)$ $\quad E(X) = n \cdot \pi \quad$ $Var(X) = n \cdot \pi \cdot (1-\pi)$

Approximation by normal distribution:

$$\hat{\Pi} \approx N\left(\pi; \sigma_{\hat{\Pi}} = \sqrt{\frac{\pi(1-\pi)}{n}}\right)$$

2 Estimation procedures

true parameter of population	θ
Estimator (function)	$\widehat{\theta} = g(X_1, \ldots, X_n)$

MSE=Mean Square Error

$$MSE = E[(\widehat{\theta}-\theta)^2] = \underbrace{E[\{\widehat{\theta}-E(\widehat{\theta})\}^2]}_{=Var(\widehat{\theta})} + \underbrace{\{E(\widehat{\theta})-\theta\}^2}_{=\text{bias}^2}$$

Example: $\theta = \mu$, $\widehat{\theta} = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$

2.1 Maximum - Likelihood (ML) Method

Likelihood-Function $L(\theta) = L(\theta|x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i|\theta)$ $\rightarrow$ maximize

LogLikelihood-Function $\log(L(\theta)) = \sum_{i=1}^{n} \log(f(x_i|\theta))$ $\rightarrow$ maximize

2.2 Least Squares (LS) Method

Quadratic Form $Q(\theta) = \sum_{i=1}^{n}(x_i - E(X_i))^2 = \sum_{i=1}^{n}(x_i - g_i(\theta))^2 \rightarrow$ minimize

2.3 Confidence interval at level $1 - \alpha$

Confidence interval for μ

		X_i normally distributed or distr. of population unknown, but $n \geq 30$
σ^2 known	Conf. interval	$P\left(\overline{X} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$ $\left[\overline{X} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}; \overline{X} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right]$ $z_{1-\frac{\alpha}{2}}$ from $N(0;1)$
	Estimator interval	$\left[\overline{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}; \overline{x} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right]$
	Length	$\ell = 2 \cdot e = 2 \cdot z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ with ℓ = length and e = error
	sample size	$n \geq \frac{\sigma^2 \cdot z^2_{1-\frac{\alpha}{2}}}{e^2}$

Confidence interval for proportion π for Normal approximation

	$X \sim B(n;\pi)$ and $\hat{\Pi} = X/n$ is approximately normally distributed
Approximative Confidence interval	$P\left(\frac{X}{n} - z_{1-\frac{\alpha}{2}} \cdot \sigma_{\hat{\Pi}} \leq \pi \leq \frac{X}{n} + z_{1-\frac{\alpha}{2}} \cdot \sigma_{\hat{\Pi}}\right) = 1 - \alpha$ $\left[\frac{X}{n} - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\frac{X}{n} \cdot \left(1 - \frac{X}{n}\right)}{n}}; \frac{X}{n} + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\frac{X}{n} \cdot \left(1 - \frac{X}{n}\right)}{n}}\right]$ $z_{1-\frac{\alpha}{2}}$ from $N(0,1)$
Estimator interval	$\left[\frac{x}{n} - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\frac{x}{n} \cdot \left(1 - \frac{x}{n}\right)}{n}}; \frac{x}{n} + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\frac{x}{n} \cdot \left(1 - \frac{x}{n}\right)}{n}}\right]$
Sample size	$n \geq \frac{z^2_{1-\alpha/2}}{4 \cdot e^2}$

Exercises

Exercise 7-3 - Lamps

A supply of $N = 2000$ lamps will be investigated by means of a simple random sample of size $n = 20$. With a help of the random variable X: "number of defective bulbs in the sample of size n=20" anumber d of defective lamps in the supply is estimated.

a) Give an unbiased estimator $\theta = f(X)$ for d and show that $\mathrm{E}(\theta) = d$.

b) In a given sample, the number of defective bulbs is equal to 3.
How many defective bulbs do you estimate in the delivery?

Exercise 7-5 - Gambling machine

A gambling machine has the following probability distribution for the win X per game (in EUR):

x	-1	0	1
$\mathrm{P}(X = x)$	p	p	1- 2p

The producer of these machines hired a statistician to perform an estimate for p to know whether the value of p has changed since the gambling machines starts up.

a) The statistician draws a sample of size $n = 6$, i.e. plays with the machine 6 times and writes down the win. The sample $(X_1, X_2, X_3, X_4, X_5, X_6)$ had realization as follows: (-1,1,-1,0,1,1).
Verbalize this sample result.

b) Calculate the following probabilities: $\mathrm{P}(X = 0)$,$\mathrm{P}(X = 1)$,$\mathrm{P}(X = -1)$.

c) How would you determine the probabilities of win X per game according to sample mentioned above, if you have no information about the probability distribution of X?

d) What is the probability $\mathrm{P}\left\{(X_1, X_2, X_3, X_4, X_5, X_6) = (-1, 1, -1, 0, 1, 1)\right\}$ based on the above probability distribution?

e) Determine the maximum likelihood estimator for p in this problem.

f) Estimate p by this sample result through the maximum likelihood method.

g) Estimate p by this sample result through the least squared method.

Exercise 7-26 - Dioxin emmisions

It is believed that the dioxin emissions of a cosmetic factory pre minute are normally distributed with the mean 5 and st. deviation 1 $\sim \mathcal{N}(5kg; 1kg)$.

a) What is the probability that the average of a sample of size $n = 9$ is between 4 and 6 kg.

b) What is the area, where the average value will be with probability of 95% ?

c) How large should be the sample, so that the average dioxin emissions are exactly estimated with probability 95% and est.error for $e = 0.5$ kg/min ?

d) Compute the confidence intervals for the average dioxin emissions at the confidence level $1 - \alpha$.

e) They measured 9 times of the dioxin emissions randomly (kg/min):
7; 4; 5; 10; 9; 6; 8 ; 6.5; 7.5 .
Calculate the estimation interval at a confidence level $1 - \alpha = 0.98$.

Exercise 7-45 - Kilometrage

A) For a test 49 randomly drawn car of the same type were fuelled with the same amount of fuel. With this amount of fuel the cars went on average 50 km. Assume that st. deviation is known 7km.

   a) Give an explicit confidence interval $[V_L, V_U]$ for average kilometrage μ for this type of car at the confidence level $1 - \alpha$.

   b) Determine the interval for μ when $1 - \alpha = 95\%$.

   c) What sample size n is needed, if the estimated interval for μ at the same level shall have a width 2 km?

B) Some visitors of this test event were randomly chosen by journalist and asked about their membership in the ADAC (German automobile club). Among 200 people 40 were ADAC members. Determine the interval fro π when $1 - \alpha$=99 %.

C) Coffee machine was installed at the tribune for this event. It fills 0.2l cup with coffee. Assume that the quantity is normally distributed. Random sample of size $n = 5$ has following values: 0.18, 0.25, 0.12, 0.20, 0.25.

   a) Give an explicit confidence interval $[V_L, V_U]$ for average quantity μ for this machine at the confidence level $1 - \alpha$.

   b) Determine the interval for μ when $1 - \alpha = 95\%$.