Statistics II: Exercise Session 3

1 Theory of sampling

1.1 Sampling functions

Random variables X_i:	$\mathrm{E}[X_i]=\mu$, $\mathrm{Var}(X_i)=\sigma^2$, $i=1,\ldots,n$
Sampling function:	$U=U(X_1,\ldots,X_n)$
Sampling mean function:	$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$
Sample mean value:	$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$
Sampling variance function:	$S^{*2}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$
	$S'^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$
	$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$

1.2 Properties of sampling mean and variance functions

Variance of $\overline{X}$

Sample	$\frac{n}{N}$	Variance of population σ^2 known	unknown
With replacement		$\frac{\sigma^2}{n}$	$\frac{S^2}{n}$
Without replacement	<0.05	$\frac{\sigma^2}{n}$	$\frac{S^2}{n}$
	≥ 0.05	$\frac{\sigma^2}{n}\cdot\frac{N-n}{N-1}$	$\frac{S^2}{n}\cdot\frac{N-n}{N-1}$

Sampling distributions of $\overline{X}$

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

Sampling distributions of variance
Under assumption of $X_i \sim N(\mu; \sigma)$ for $i = 1, \ldots, n$

μ	Sampling function	Expected value	Distribution
Known	$S^{*2} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2$	$\mathrm{E}[S^{*2}] = \sigma^2$	$\frac{n \cdot S^{*2}}{\sigma^2} \sim \chi_n^2$
Unknown	$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$	$\mathrm{E}[S^2] = \sigma^2$	$\frac{(n-1) \cdot S^2}{\sigma^2} \sim \chi_{n-1}^2$

1.3 Central limit theorem

Assume $X_1, X_2, \ldots, X_n$ independent, identically distributed random variables with $\mathrm{E}[X_i] = \mu \neq \pm\infty$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$ (for $i = 1, \ldots, n$). Then the random variable $S_n = \Sigma_i X_i$ has the expectation value $\mathrm{E}[S_n] = n\mu$ and the variance $\mathrm{Var}(S_n) = n\sigma^2$. The distribution of the standardized random variables

$$Z_n = \frac{S_n - \mathrm{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma}$$

converges with increasing n to the standard normal distribution

$$\lim_{n \to \infty} \mathrm{P}(Z_n \leq z) = \Phi(z).$$

Exercises

6-1: *Three people*

On the population of $N = 3$ people of age 20, 22 and 24 a random sampling (with repetition) was performed with a sample size of $n = 2$.

a) Compute the mean μ and the variance σ^2 of this population.

b) How many 2-tuples are there in the sample space?

c) List all of the 2-tuples wich can be an outcome of the sampling.

d) What is the distribution of the following sampling functions?

   (i) $Y = \sum\limits_{i=1}^{2} X_i$
   Calculate $\mathrm{E}[Y]$ and $\mathrm{Var}(Y)$.

   (ii) $\overline{X} = \frac{1}{n}\sum\limits_{i=1}^{2} X_i$
   Calculate $\mathrm{E}[\overline{X}]$ and $\mathrm{Var}(\overline{X})$.

   (iii) $S'^2 = \frac{1}{n}\sum\limits_{i=1}^{2}(X_i - \overline{X})^2$
   Calculate $\mathrm{E}[S'^2]$.

   (iv) $S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{2}(X_i - \overline{X})^2$
   Calculate $\mathrm{E}[S^2]$.

e) Using the results from d) prove the following statements:

   (i) $\mathrm{E}(Y) = n \cdot \mu$ and $\mathrm{Var}(Y) = n \cdot \sigma^2$
   (ii) $\mathrm{E}(\overline{X}) = \mu$ and $\mathrm{Var}(\overline{X}) = \sigma^2/n$
   (iii) $\mathrm{E}(S'^2) = \sigma^2 \cdot (n-1)/n$
   (iv) $\mathrm{E}(S^2) = \sigma^2$

6-3: *Headache pills*

The amount of an active ingredient in headache pills is normally distributed. Since the low amount of the active ingredient does not work and high amount causes side effects, the production has to be monitored. With the help of random sampling the average amount of the active ingredient μ (in mg) is estimated. What is the probability that the sample mean $\overline{X}$ takes on values which are more than 0.5 mg greater than the mean μ, if

a) $\sigma = 1$ mg and $n = 16$?

b) $\sigma = 1$ mg and $n = 64$?

c) $\sigma = 2$ mg and $n = 64$?

6-9: *Tennis instructor*

A tennis instructor offers 8 hours of training every day in a month (30 days). Over the years he found out that during a lecture his student hits the ball over the fence of the court (without returning) 2 or 7 times each with probability of 0.1 and 1 or 6 times each with probability of 30 %. None of his students hits the ball over the fence 3, 4, less than 1 or more than 7 times.

a) What is the approximate probability that 900 - 1000 balls are hit over the fence in one month?

b) What is the approximate probability that more than 1050 balls are hit over the fence in one month?

c) Calculate the (symmetric) interval for expected value so that the number of balls hit over the fence in one month lies between its boundaries with the probability of 99 %.

d) Validate the assumptions used in this task.