Mathematical Statistics
Lab 5

1. We want to measure average expenditures of 4th quarter of 2019 for second year students in natural science in Tartu. Describe population and sample.

2. We are interested in average alcohol consumption among Tartu's female students. We go to the bar in the end of September on Thursday night and ask randomly selected 100 customers how many drinks they had. Can this sample help us answer our research question? If not, then what would be a better sample?

3. We are interested in a population's age mean. To find it, every investigator takes a sample with a size of n=20 and finds the sample's mean value.

   The sample:

   The mean of the sample:

   Is the sample mean between $\mu \pm \frac{\sigma}{\sqrt{n}}$, if we know that the age mean in the population is 40 and the variance 26.48?

   Ages in the population: 45 44 46 37 38 45 43 43 38 35 40 37 34 44 35 33 45 46 40 42 45 37 33 40 47 43 41 35 34 47 47 30 40 45 32 47 37 43 36 54 43 38 46 40 41 40 45 41 38 44 34 46 51 34 40 44 44 36 46 43 41 35 43 39 29 39 40 38 36 47 41 34 34 41 41 34 39 37 47 34 43 39 38 35 38 44 46 40 36 48 30 41 39 33 29 31 48 41 43 32

4. Show that if we assume the random sample to have a distribution $N(\mu, \sigma^2)$, where μ is known and σ is unknown, then the estimate

   $$\hat{\sigma}_1^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

   is an unbiased estimate for the parameter σ^2, and if the estimate

   $$\hat{\sigma}_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2$$

   is used, we obtain a biased estimate for the parameter σ^2. Find the bias.

Mathematical Statistics Lab 5

***Exercise 5** The sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is an unbiased estimate of σ^2.

***Exercise 6** For any constant c, d,

$$\sum_{i=1}^{n}(x_i - c)(y_i - d) = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) + n(\bar{x} - c)(\bar{y} - d).$$

***Exercise 7** A sample $x_1, \ldots, x_n$ is from a population with normal distribution $N(\mu, \sigma^2)$.We consider the following estimates for σ^2:

$$\hat{\sigma_1^2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2, \hat{\sigma_2^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2, \hat{\sigma_3^2} = \frac{1}{n+1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Which of these estimates is unbiased?

***Exercise 8**

1. A sample $x_1, \ldots, x_n$ is from a population with uniform distribution $U(\theta, \theta+1)$, unknown parameter $\theta > 0$. $\bar{x}$ is the sample mean. Show that $\hat{\theta}_1 = \bar{x} - \frac{1}{2}$ is unbiased estimate of θ.
2. A sample $x_1, \ldots, x_n$ is from a population with uniform distribution $U(\theta, 2\theta)$. Show that $\hat{\theta} = \frac{2}{3}\bar{x}$ is an unbiased and consistent estimate of θ.

***Exercise 9** Suppose that sample $x_1, \ldots, x_3$ is a sample from the population X, $E(X) = \mu$, $varX$ exists. Prove that the following statistic are all unbiased for μ.

(1) $\hat{\mu}_1 = \frac{1}{2}X_1 + \frac{1}{3}X_2 + \frac{1}{6}X_3$;

(2) $\hat{\mu}_2 = \frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{3}X_3$;

(3) $\hat{\mu}_3 = \frac{1}{6}X_1 + \frac{1}{6}X_2 + \frac{2}{3}X_3$;

Compare the efficiency of these estimates.

$\boxed{3}$ $X \sim N(40, 26{,}48)$

41, 0

$\bar{X} \sim N\left(40, \frac{26{,}48}{n}\right)$

$\bar{X}$ $n=20$

$E\bar{x} = E\left[\frac{1}{n}\sum^{n} x_i\right] =$

38.5

38,7

$= \frac{1}{n} E\left[\sum^{n} x_i\right] = \frac{\cancel{n} \cdot 40}{\cancel{n}} = 40$

40,5

41.0

$V\bar{X} = V\left[\frac{1}{n}\sum^{n} x_i\right] =$

$= \frac{1}{n^2}\sum_{i}^{n} Var\, x_i =$

$= \frac{\cancel{n}}{n^{\cancel{2}}} 26{,}48 =$

$40 \pm 1{,}15$

$(38{,}85 ; 41{,}15)$

quantile $\mathcal{L}$ $\Rightarrow$

$\sim 60\%$

$\boxed{4}$

$$E[\hat{\sigma}^2] = \sigma^2$$

$$B = E[\hat{\sigma}^2] - \sigma^2$$

$$E[\hat{\sigma}_1^2] = E\left[\frac{1}{n}\sum^{n}(x_i-\mu)^2\right] =$$

$$= \frac{1}{n}E\left[\sum^{n}(x_i-\mu)^2\right] = \frac{1}{n}\sum^{n}E[(x_i-\mu)^2] =$$

$$= \frac{1}{n}\sum^{n} V X_i = \frac{n\sigma^2}{n} = \sigma^2$$

unbiased

$$E[\hat{\sigma}_2^2] = \frac{1}{n-1}\sum^{n}E(x_i-\mu)^2 =$$

Biased

$$= \frac{n\sigma^2}{n-1}$$

$$B = \frac{n\sigma^2}{n-1} - \sigma^2 = \frac{n\sigma^2 - (n-1)\sigma^2}{n-1} = \frac{\sigma^2}{n-1}$$

$\boxed{5}$

$$E[S^2] = \frac{1}{n-1} \sum_{i=1}^{n} E\left[(x_i - \bar{x})^2\right] = \frac{nS^2}{n-1}$$

$$\sum^{n} (x_i - \bar{x})^2 = \sum^{n} (x_i^2 - 2x_i\bar{x} + \bar{x}^2) =$$

$$= \sum x_i^2 - 2\bar{x}\underbrace{\sum x_i}_{n\bar{x}} + n\bar{x}^2 = \sum^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2$$

$$= \sum^{n} x_i^2 - n\bar{x}^2$$

$$V[X] = EX^2 - (EX)^2$$

$$E[x_i^2] = Var[X_i] + (EX_i)^2 = \sigma^2 + \mu^2$$

$$E(\bar{X}^2] = Var\bar{X} + [E\bar{X}]^2 = \frac{\sigma^2}{n} + \mu^2$$

$$\parallel$$

$$Var\left[\frac{1}{n}\sum^{n} x_i\right] = \frac{\sigma^2}{n}$$

$$E[S^2] = \frac{1}{n-1} E\left[\sum^n X_i^2 - n\overline{X}^2\right] =$$

$$= \frac{1}{n-1}\left(\sum E X_i^2 - E\,\overline{X}^2\right) =$$

$$= \frac{1}{n-1}\left(n(\sigma^2+\mu^2) - n\left[\frac{\sigma^2}{n} + \mu^2\right]\right) =$$

$$= \frac{1}{n-1}\left(n\sigma^2 \cancel{+ n\mu^2} - \sigma^2 - \cancel{n\mu^2}\right) =$$

$$= \frac{1}{n-1}\,\sigma^2(n-1) = \sigma^2$$

$\boxed{7}$