1 Sampling

Example 1.1. *We have population with an unknown population mean and population variance. We want to estimate μ and σ^2 using a random sample $y_1, \ldots, y_n$. The basic estimates are:*

- *To estimate the population mean μ, we use the sample mean $\bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j$.*
- *If we know μ, to estimate the population variance σ^2, we use $\frac{1}{n}\sum_{j=1}^{n}(y_j - \mu)^2$.*
- *If we do not know μ, to estimate the population variance σ^2, we use the sample variance $s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(y_j - \bar{y})^2$.*

Example 1.2. *A biologist is interested in studying a type of rodents. The weight of newborns of this type of rodents has a population mean of μ and a population variance σ^2. μ and σ^2 are obtained weighting all the rodents in the world. However, the biologist cannot weight all the rodents of this type in the world. The biologist takes a random sample of the weights of 10 newborn rodents with the following results*

$$6.51, 9.88, 9.97, 6.91, 8.11, 6.13, 11.68, 5.66, 8.90, 3.45$$

The weights are given in grams. From this sample, the biologist estimates the μ by

$$\bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j = \frac{6.51 + 9.88 + 9.97 + 6.91 + 8.11 + 6.13 + 11.68 + 5.66 + 8.90 + 3.45}{10} = 7.72.$$

7.72 *is the best estimate for the population mean which the biologist can give looking to the data she has. The biologist estimates the population variance σ^2 by the sample variance*

$$\begin{aligned}
&s^2 = \tfrac{1}{n-1}\textstyle\sum_{j=1}^{n}(y_j - \bar{y})^2 \\
&= \tfrac{1}{10}\left((6.51 - 7.72)^2 + (9.88 - 7.72)^2 + (9.97 - 7.72)^2 + (6.91 - 7.72)^2 + (8.11 - 7.72)^2\right. \\
&\left.+(6.13 - 7.72)^2 + (11.68 - 7.72)^2 + (5.66 - 7.72)^2 + (8.90 - 7.72)^2 + (3.45 - 7.72)^2\right) \\
&= 6.008778
\end{aligned}$$

Theorem 1.1. *Suppose that $Y_1, \ldots, Y_n$ are independent identically distributed random variables from a normal distribution with mean μ and variance σ^2. Then,*

- *$\bar{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j$ has a normal distribution with mean μ and variance σ^2.*
- *$\frac{1}{\sigma^2}\sum_{j=1}^{n}(Y_j - \mu)^2$ has a chi-square distribution with n degrees of freedom.*
- *$\frac{1}{\sigma^2}\sum_{j=1}^{n}(Y_j - \bar{Y})^2$ has a chi-square distribution with $n - 1$ degrees of freedom.*

Theorem 1.2. *We have the following basic equalities:*

$$\mathrm{Var}(Y) = E[Y^2] - (E[Y])^2$$

$$\frac{1}{n}\sum_{j=1}^{n}(y_j - \bar{y})^2 = \frac{1}{n}\sum_{j=1}^{n} y_j^2 - (\bar{y})^2$$

$$\frac{1}{n}\sum_{j=1}^{n}(y_j - \bar{y})^2 = \frac{1}{n}\sum_{j=1}^{n}(y_j - \mu)^2 - (\bar{y} - \mu)^2$$

The previous inequality gives the following for r.v.'s $Y_1, \ldots, Y_n$,

$$\frac{1}{\sigma^2}\sum_{j=1}^{n}(Y_j - \mu)^2 = \frac{1}{\sigma^2}\sum_{j=1}^{n}(Y_j - \bar{Y})^2 + \frac{n(\bar{Y} - \mu)^2}{\sigma^2}$$

If $Y_1, \ldots, Y_n$ are independent identically distributed random variables from a normal distribution with mean μ and variance σ^2, then

- $\frac{1}{\sigma^2}\sum_{j=1}^{n}(Y_j - \mu)^2$ *has a chi-square distribution with n degrees of freedom.*
- $\frac{1}{\sigma^2}\sum_{j=1}^{n}(Y_j - \bar{Y})^2$ *a has chi-square distribution with $n-1$ degrees of freedom.*
- $\frac{n(\bar{Y}-\mu)^2}{\sigma^2}$ *has a chi-square distribution with 1 degree of freedom.*