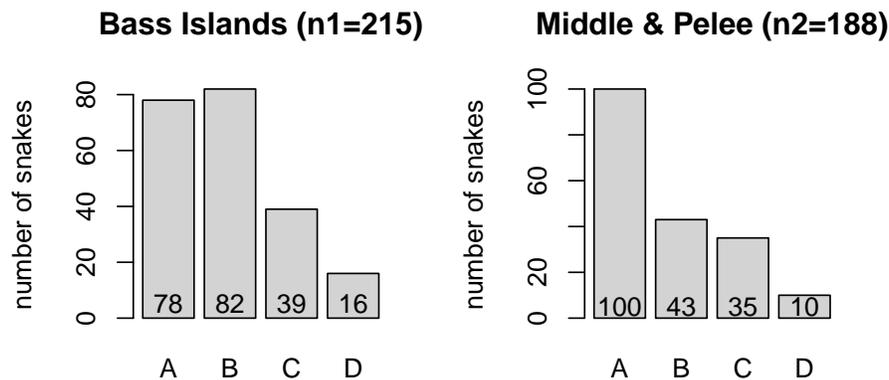


## STATISTICS FOR EES — EXERCISE SHEET 9

- From a cohort of high-school graduates 20 female students and 11 male students study biology, 6 female students and 12 male students study computer science, 2 female and 8 male students study physics, 5 female and 10 male students study mathematics, 5 female students and 1 male student study chemistry, and 10 students of each gender study statistics. Is the choice of the subject significantly gender-biased?
- The file `qqnorm_exerc_14.pdf` shows normal-qqplots for data sampled according to six different distributions. Draw density polygons or histograms to show for each of the six distributions how it differs from a normal distribution with the same mean and the same variance.
- The markings of the water snake *Nerodia sipedon* in Lake Erie can be grouped into four classes. On the mainland all snakes have strong markings (A). On the islands many snakes have no markings (A) or only weak markings (B,C). Frequencies of the four markings in samples from the Bass islands and from Middle Island and Pelee Island are as follows.



- Let  $H_0$  be the hypothesis that the distribution of marking classes were the same for all islands. What is the expected number of snakes of class A sampled from the Bass Islands if we assume that  $H_0$  is true?
  - We apply a Chi-square test with null hypothesis  $H_0$ . How much does the table entry (Bass Islands : class A) contribute to the test statistic  $X^2$ ?
  - If we sum  $(O_i - E_i)^2/E_i$  for all table entries we find  $X^2 = 14.74$ . How many entries does the table have? How many degrees of freedom does  $X^2$  have?
  - Using the quantile tables on the website of the course, what can you say about the  $p$ -value in the snake example?
- In a study<sup>1</sup> about the alkaline-phosphatase gene, three distinguishable alleles “S”, “T” and “F” were found. The following numbers of genotypes were observed in 332 persons: SS: 141, SF: 111, FF: 28, SI: 32, FI: 15, II: 5.

<sup>1</sup>Harris (1966) Enzyme polymorphism in Man. *Proc. Roy. Soc. B* **164**:1153-64

- (a) Compute the relative allele frequencies for S, I and F.
- (b) Use the allele frequencies to compute the expectation values for the genotypes in a sample of 332 persons, assuming a Hardy-Weinberg equilibrium for this gene.
- (c) Is the observed deviation from Hardy-Weinberg equilibrium significant?