Sample exam questions

These are sample exam questions. Their purpose is to give you the possibility to check your knowledge and understanding. This is not a comprehensive list. The problems on the exam will be similar but not exactly the same. You are strongly advised to concentrate on the problem solving approaches and methods rather than the specific solutions to these problems.

• Lecture 5

1. Definitions, meanings and ability to use *very comfortably* the notions of mean, median, mode, percentile, variance, range, standard deviation, etc.

2. Consider the following set of microarray measurements: \( x_1 = 9, x_2 = 10, x_3 = 8, x_4 = 13, x_5 = 11, x_6 = 10 \). Calculate the following statistics (it may be useful if you do not use a calculator). Pay attention to the precision you use:

   (a) \( \sum x_i \)
   
   (b) \( \sum_{i=2}^{5} x_i \)
   
   (c) \( \bar{x} \)
   
   (d) \( \sum(x_i - \bar{x}) \)
   
   (e) \( \sum(x_i - \bar{x})^2 \)
   
   (f) \( \sum x_i^2 \)
   
   (g) \( (\sum x_i)^2 \)
   
   (h) \( \sum x_i^2 - \left(\frac{\sum x_i}{n}\right)^2 \)
   
   (i) \( \sum x_i - 10 \)
(j) $\sum_{i=1}^{3}(x_i + 2)$

(k) $\sum_{i=1}^{n} 5$

(l) $\sum 2x_i$

(m) $2 \sum x_i$

(n) Sample standard error

(o) Sample mean

(p) Sample mode

(q) Sample median

(r) Sample variance

(s) Standard deviation

(t) Range

(u) $s^2$

(v) $s$

(w) Sample mean and variance for the variable $y_i = x_i + 2$ (Hint: do not use a calculator; calculate the sample mean and variance of the variable $y_i$ as a function of the sample mean and variance of the variable $x_i$).

(x) Sample mean and variance for the variable $z_i = 3x_i$ (Hint: do not use a calculator; calculate the sample mean and variance of the variable $y_i$ as a function of the sample mean and variance of the variable $x_i$).

(y) The expression levels of two different genes are measured several times as follows:
<table>
<thead>
<tr>
<th>Gene 1</th>
<th>Gene 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.9</td>
<td>3.1</td>
</tr>
<tr>
<td>9.6</td>
<td>10.0</td>
</tr>
<tr>
<td>11.2</td>
<td>9.9</td>
</tr>
<tr>
<td>9.4</td>
<td>5.1</td>
</tr>
<tr>
<td>9.9</td>
<td>21.2</td>
</tr>
<tr>
<td>8.9</td>
<td>3.8</td>
</tr>
<tr>
<td>10.4</td>
<td>17.0</td>
</tr>
<tr>
<td>9.3</td>
<td>2.9</td>
</tr>
<tr>
<td>9.7</td>
<td>18.0</td>
</tr>
<tr>
<td>11.0</td>
<td>9.3</td>
</tr>
</tbody>
</table>

i. Calculate the mean, median and mode for the two genes. What can you say about the two genes?

ii. Calculate the variance, standard deviation and range for the two genes. What can you say about the two genes now?

iii. What can you conclude about the mean, median and mode vs. variance, standard deviation and range?

- Lecture 6

1. Calculate the covariance and correlation between the two genes in the previous exercise.

2. Definitions, meanings and ability to use *very comfortably* the notions of probability (empirical and classical, axioms, properties, etc.), conditional probabilities, independent events, etc.

3. We decide to declare a microarray spot as having high confidence if it is either perfectly circular or perfectly uniform (the standard deviation of the signal intensities is less than a chosen threshold) or both. There are 10000 spots of which 6500 are circular and 7000 spots are uniform. 5000 spots are both circular and uniform. What is the probability that a randomly chosen spot is good (has high confidence).
4. Assume that the probability that a certain mRNA fragment incorporates the dye is 0.9, the probability that the fragment hybridizes on a given spot is 0.95 and the probability that we observe it using a microarray experiment is 0.8. Assume that we labeled the mRNA and we can now select only the stained mRNA. What is the probability to obtain a spot with a non-zero signal?

5. Assume that the probability that a certain mRNA fragment incorporates the dye is 0.9, the probability that the fragment hybridizes on a given spot is 0.95 and the probability that we observe it using a microarray experiment is 0.8. Are hybridization and labeling independent?

6. Assume two unrelated genes have are expressed 0.85% and 0.75% of the time in the condition under study. What is the probability that a certain sample has both genes expressed?

7. Assume that the probability that a certain mRNA fragment incorporates the dye is 0.9 and the probability that the labeled fragment hybridizes on a given spot is 0.888. What is the probability to obtain a spot with a non-zero signal?

8. Definitions, meanings and ability to use very comfortably the notions of pdf, cdf, expected values, etc.

9. You are using a microarray with 100 spots distributed evenly across the array surface. The total surface of the spots is $100 \text{ mm}^2$ and the total surface of the array is $1000 \text{ mm}^2$. Consider that a spot is contaminated if anything falls within the signal area. What is the probability that a single dust particle contaminates a given spot? Now consider that a spot is contaminated if anything falls anywhere within the signal area or in its local neighbourhood (assume the array is divided in equal squares with a spot in each square). What is the probability that a single dust particle contaminates a spot in these conditions?