

# Practise Exam (Section 1)

BIOB11 2025

## Instructions

The exam has two sections. Each section can award a maximum of 15 point (30 points total). The pass mark (godkänt) is 60%, the high pass mark (väl godkänt) is 80%. You do not need to pass both sections to pass the exam.

## This is Section 1.

In this section, you are *not* permitted to access any materials. Under no circumstances are you allowed to communicate with anyone during the exam. Write your answers in either Swedish or English. For multiple choice questions, your chosen answer must be clearly indicated.

Note that in the real exam, there will also be a four point question from Kaj's lecture and subsequent discussion. Hence this practise exam has only 11 points.

Good luck!

---

## Question 1 (1 point)

A group of researchers at Lund University want to estimate the GC content (the proportion of DNA bases that are guanine G or cytosine C rather than adenine A or thymine T) in immune function related genes in the genome of fältsorken (*Microtus arvalis*). They select 250 known immune function related genes spread across the genome and calculate the GC content. Using the meaning of the following words in statistics, what is the population and what is the sample in this scenario?

**Suggested Answer:** The population is all the immune function related genes in the genome of *Microtus arvalis*. The sample is the 250 selected genes from which GC content was measured.

### Question 2 (1 point)

When testing a hypothesis, researchers often calculate and report a p-value. What information does the p-value provide?

**Suggested Answer:** The p-value provides information as to how likely the result (or a result more extreme) would be if the null hypothesis was correct.

### Question 3 (1 point)

Allometry is the study of how the characteristics of organisms change with size. You have measured the carapace width (mm) and claw width (mm) of 60 krabbtaskor (*Cancer pagurus*) caught off the coast of Norway. What statistic could you use to describe the relationship between body mass and claw width?

Choose one:

- A) The slope of the line fit with linear regression
- B) F-statistic
- C) Median
- D) Chi-squared

**Correct Answer:** A - The question is interested in the relationship between two continuous variables. The slope of a line fit with linear regression provides a way of describing how change in one variable (carapace width) results in change in another (claw width). A correlation coefficient would also be an option, but it was not among the options.

### Question 4 (1 point)

You have measured the mass (grams) of blåmesar (*Cyanistes caeruleus*) chicks from nest boxes in four different environments: high density urban (Malmö), low density urban (Dalby), forested (Skrylle) and agricultural (outskirts of Genarp). What would be an appropriate test statistic to use to investigate differences in nestling mass between these environments?

Choose one:

- A) The standard deviation
- B) F-statistic
- C) Chi-squared
- D) The inter-quartile range

**Correct Answer:** B - As we have four groups, a simple difference in mean is not possible. The F statistic (the ratio of between-group variance to within-group variance), if more extreme than the null distribution, would tell us if at least one of the groups differs from the overall mean. Notably, this assumes the variances within groups are not extremely different.

### Question 5 (1 point)

In the same study as described in question 4, you also recorded if each chick successfully fledges (leaves) the nest. What would be an appropriate test statistic to use to test if there is a difference in the proportion of chicks that successfully fledge between the environments?

Choose one:

- A) The standard deviation
- B) F-statistic
- C) The slope of the line fit with linear regression
- D) Chi-squared

**Correct Answer:** D - We want to test for an association between two categorical variables, environment (high density urban, low density urban, forested, and agricultural) and fledge status (successful, unsuccessful). Since we have more than two groups in one of the categories, Chi-squared is our only option.

### Question 6 (2 points)

You read in a research paper that the mean length of strömmingar (*Clupea harengus membras*) caught in 2024 was 152 mm, with a 95% confidence interval of 139 mm to 163 mm. What is a 95% confidence interval, and what information does it provide?

**Suggested Answer:** The 95% confidence interval strict interpretation is if we repeated the experiment many times and calculated a 95% CI each time, the 95% CI's would include the "true" value 95% of the time. In general terms, the 95% confidence interval provides information about how certain we are in our estimate of (in this case) the mean. Also in more general terms, we might expect that if we were to take another sample, that its mean would fall within our 95% CI. Any answer which can express the idea of uncertainty, the sampling distribution, etc would suffice. An answer that only provided code examples to calculate it would not.

### Question 7 (1 point)

A group of researchers have measured the body lengths (mm) of 120 adult hussysors (*Acheta domesticus*). The distribution of body masses has a single peak, with a mean of 18.2 mm and a median of also 18.2 mm. The standard deviation is 1.4 mm. What does this suggest about the shape of the distribution of body lengths?

Choose one:

- A) The distribution is right-skewed (tail extends to the right)
- B) The distribution is left-skewed (tail extends to the left)
- C) The distribution is symmetrical (equal sized tails on either side)
- D) The distribution is leptokurtotic (wide tails and positive kurtosis)

**Correct Answer:** C - Since the mean and median are the same, and we know that the distribution has a single peak, this suggests that it is likely symmetrical.

### Question 8 (2 points)

What is the purpose of a null distribution in a hypothesis test? How is it used to decide whether to reject or fail to reject the null hypothesis?

**Suggested Answer:** The null distribution is a sampling distribution derived under a model of the null hypothesis. It represents the range of possible values of the test statistic that could be observed assuming the null hypothesis is correct. It is compared with the observed statistic to decide if the observed statistic is either likely compatible with the null hypothesis (fail to reject), or too extreme such that observing it under the null hypothesis is very unlikely (reject).

### Question 9 (1 point)

From a sample (sample size =  $N$ ), a single bootstrap sample can be generated by which of the following processes?

Choose one:

- A) Randomly resampling  $N/2$  values from the original sample without replacement  $N$  times
- B) Randomly resampling  $N/2$  values from the original sample with replacement  $N$  times
- C) Randomly resampling  $N$  values from the original sample without replacement
- D) Randomly resampling  $N$  values from the original sample with replacement

**Correct Answer:** D - To generate a single (emphasis on single) bootstrap sample from our sample, we need to resample it *with* replacement. If we sample without replacement (C), we will just get our original sample. A bootstrap sample should also be the same size as the original sample, which rules out (A) and (B).