Biostats VHM 801 Course
Winter 2026, Atlantic Veterinary College, PEI
Henrik Stryhn

Questions for oral exam selected from quiz sessions (final version)

Note: **All statements are to be identified as True/False, and answers should be supplemented by an explanation (if applicable) or additional information related to the statement. At the exam, the statements will be presented without numbers.**

3.1) The sum of all the probabilities in a probability distribution for a discrete random variable must equal 1.

3.2) The standard deviation of a random variable is always nonnegative.

3.3) The graph of a probability density function/curve may extend below the x-axis.

3.4) For a continuous random variable there is no (positive) probability associated with a single value.

4.1) The density curve for any normal distribution is symmetrical and bell-shaped.

4.2) The two parameters of normal distributions are the mean and the median.

4.3) Normal distributions have positive probability for intervals below zero.

4.4) If a P-value for a normality test is low, the data do not match a normal distribution well.

4.5) The two most crucial assumptions behind a binomial distribution are independence and equal probabilities in all trials.

4.6) A binomial distribution can be used also for counts of three or more different outcomes in each trial.

4.7) The binomial distribution has two parameters, of which one is (almost) always known.

5.1) A confidence interval for a population mean (μ) is guaranteed to contain the true value (μ).

5.2) A confidence interval for a population mean always includes the sample mean.

5.3) As the confidence level increases, the confidence interval becomes longer.

5.4) A statistical hypothesis is never stated in terms of a sample estimate.

5.5) **The P-value is the probability that the null hypothesis is true.**

5.6) **A low P-value means that the data are incompatible with the null hypothesis (or other assumptions).**

5.7) **In a test of a statistical hypothesis, we attempt to find evidence in favour of the null hypothesis.**

6.1) **Inference about a population mean, when the standard deviation is unknown, is based on a suitable t-distribution.**

6.2) **The degrees of freedom of t-distributions are calculated from sample statistics, such as the sample mean.**

6.3) **The $t^\star$ (t-star) value used for a confidence interval for the population mean is the value of t-test statistic.**

6.4) **If we have 95% confidence intervals for two population means (from independent samples), we can get a 95% confidence interval for their difference by subtracting the confidence intervals from each other.**

6.5) **It is only valid to form differences between the observations in two samples if the samples are paired.**

7.1) **An approximate 95% confidence interval for an unknown proportion p is the sample mean plus or minus its standard error.**

7.2) **The classical (large-sample) significance test for a population proportion is based on a t-distribution.**

7.3) **If two sample proportions are equal, then the sample counts must be equal.**

7.4) **A confidence interval for $p_1 - p_2$ can be used to draw a (statistical) conclusion about the equality of two population proportions (p_1 and p_2).**

7.5) **Rank-based test do not use the full information in the data.**

7.6) **Non-parametric tests are more sensitive to outliers than tests based on the normal distribution.**

8.1) **If we reject the null hypothesis in the test for homogeneity of populations, there is evidence to suggest the category proportions are all the same.**

8.2) **If we reject the null hypothesis in the test for independence of two categorical variables, there is evidence to suggest the two variables are dependent.**

8.3) **If a two-way table has an observed count of zero, the conditions for use of the chi-square distribution for the Pearson X^2 can never be met.**

8.4) A χ^2 (chi-square) distribution is similar to a standard normal distribution, in particular when the degrees of freedom is large.

8.5) As long as the statistical analysis is correct, it does not matter how it is reported.

8.6) It is better to report actual P-values instead of ranges such as $P < 0.05$.

8.7) In a journal not focused on statistics, the description of statistical methods should always be kept to a few sentences.

8.8) The number of observations used for any specific analysis should always be retrievable from the information provided.

9.1) In a one-way ANOVA, all group sizes must be the same in order for the F-statistic to be valid.

9.2) A balanced design means that all groups have the same number of observations.

9.3) If we reject the null hypothesis in a one-way ANOVA, there is evidence to suggest all the population means of the groups are different.

9.4) The Bonferroni multiple comparisons procedure can be used (also) if we are only interested in some of the comparisons.

9.5) For the ANOVA F-statistic, we reject the null hypothesis only for large values of the test statistic (no matter the alternative hypothesis).

10.1) In a simple linear regression, all the points lie on the estimated regression line.

10.2) In a simple linear regression, the slope being zero corresponds to no linear association between y and x.

10.3) The correlation is always between 0 and 1.

10.4) When predicting y from x in a simple linear regression, the standard error and prediction error don't depend on x.

10.5) The hypothesis of no linear association (against a two-sided alternative) can be tested both by a t-test and an F-test, with the same P-value.

10.6) A simple linear regression model assumes the dispersion (spread) about the line to be constant.

11.1) Parallel lines in an interaction plot for two factors means that the relationship between the factors is additive.

11.2) **In a two-way ANOVA, if there is evidence of a significant interaction, the analysis stops.**

11.3) **If there is no evidence of a significant interaction, the two-way ANOVA needs to be followed up by one-way ANOVAs for each of the factors.**

11.4) **When the data have replication, the two-way ANOVA model (with interaction) is equivalent to a one-way ANOVA model for the factor formed by combining the two factors.**

11.5) **The normality assumption in ANOVA and regression is assessed by a normal plot and normality test for the outcome.**

12.1) **"Statistically significant at the 0.05 level" can be explained as: There is only probability 0.05 that the null hypothesis is true.**

12.2) **If the *P*-value is larger than 0.05, the null hypothesis is true.**

12.3) **If you find an interesting pattern in a set of data, it is appropriate to then use a significance test to determine its significance.**

12.4) **If given the choice between presenting a confidence interval and a *P*-value, one should pick the confidence interval.**

12.5) **A *P*-value does not give information about the size of the effect.**

12.6) **Sample size calculations based on precision and power require different types of information about the data (to be obtained in the study).**

12.7) **The sample size obtained from a statistical sample size calculation accounts for missing values.**