16

Data analysis in cytopathology

Andrew Evered

Summary

- For visualizing cytology data, use pie charts, bar charts, or stacked bar charts for nominal data, box and whisker charts for non-Gaussian ordinal data, and bar charts with error bars for normally distributed continuous data. Use a scatterplot to display the relationship between two continuous variables and use a line graph to display trends.
- Cytologists interpret cells correctly most of the time, but occasionally make mistakes. The number of correct (true negatives and true positives) and incorrect (false negatives and false positives) interpretations are combined using mathematical formulae to produce the performance indicators sensitivity, specificity, negative predictive value, and positive predictive value. Receiver operating characteristic (ROC) analysis is an alternative approach to performance assessment.
- The diagnostic performance of cytologists and cytology laboratories is managed through a robust system of training, internal quality control, external quality assurance, and audit.
- Cytology research aims to produce generalizable new knowledge that can be used to make important technical, clinical, and managerial decisions.
- The data generated by cytology research are diverse. Methods for analysing them are equally varied.
- Test for associations between nominal variables using the Chi-squared test.
- Before applying inferential statistical tests to ordinal or continuous data, determine whether your data have a normal (Gaussian) or a non-normal (non-Gaussian) distribution.
- Use nonparametric methods of analysis (e.g. Wilcoxon's signed-rank test, Mann-Whitney U test, Kruskal-Wallis test) to test for differences between groups of non-Gaussian data.
- Use parametric methods of analysis (paired t-test, independent t-test or analysis of variance) to test for differences between groups of Gaussian data.
- Use linear regression analysis to look for correlations between variables.
- To predict an outcome from several independent variables, use logistic regression analysis when there are just two possible outcomes or discriminant analysis when there are multiple possible outcomes.