

## Meeting 12: Data Analysis 2

### Reading

Comprehensive and easy to understand references on this area are difficult to find, hence the large list of recommended reading below.

#### *Hypothesis testing, p-values and confidence intervals*

- Curtis CR, Salman MD and Shott S (1990) Statistics simplified – P values, Journal of the American Veterinary Medical Association, 197(3): 318-320
- Shott S (1990) Statistics simplified – Confidence intervals, Journal of the American Veterinary Medical Association, 197(5): 576-578

OR

- Thrusfield 2<sup>nd</sup> Edn, pp 199-219 or 3<sup>rd</sup> Edn, pp 247-265 (estimation and hypothesis testing)

OR

- Petrie and Watson pp 69-75 (hypothesis testing)

OR

- Dohoo, pp 131-136 (Hypothesis testing and confidence intervals)

#### *Statistical tests*

- Curtis CR, Salman MD and Shott S (1991) Statistics simplified – Comparing means, Journal of the American Veterinary Medical Association, 198(1): 62-65.
- Shott S (1990) Statistics simplified – Comparing proportions, Journal of the American Veterinary Medical Association, 197(11): 1460-1462.
- Shott S (1990) Statistics simplified – Association, Journal of the American Veterinary Medical Association, 198(3): 404-407.
- Shott S (1990) Statistics simplified – Nonparametric statistics, Journal of the American Veterinary Medical Association, 198(7): 1126-1128.

OR

- Petrie and Watson pp 78-88 (t-test and hypothesis testing), pp 90-92 (F-test for variance), pp 101-112 (Chi-squared test and hypothesis testing) pp 138-150 (non-parametric tests)

#### *Standardisation*

- Thrusfield 2<sup>nd</sup> Edn, pp 49-52 or 3<sup>rd</sup> Edn, pp 63-65 (standardisation)

#### *Stratified analysis, clustering and sources of error*

- Thrusfield 2<sup>nd</sup> Edn pp 232-234 or 3<sup>rd</sup> Edn, pp 278-280 (Mantel-Haenzel procedure)
- Dohoo, pp 29-30 and 40 (types of error), pp 459-471 (clustering and stratified analysis).
- Curtis CR, Salman MD (1990) Statistics simplified – Power and sample size, Journal of the American Veterinary Medical Association, 197(7): 838-840.
- Petrie and Watson pp 175-178

## Presentations

1. Hypothesis testing, power, type 1 and 2 errors
2. Basic statistical tests
  - Unit of analysis
  - Independence
  - Parametric and non parametric tests (including underlying assumptions, applications and limitations)
  - Tests for different data types
3. Stratified analysis and the Mantel-Haenzel procedure
4. Identifying 'tricky data' – paired and matched data, repeated measures on the same animals, clustered data and autocorrelation

## Exercises

1. Read the following papers and discuss how clustering was accounted for during the design and analysis stages of the studies.
  - Gitau GK, Perry BD, Katende JM, McDermott JJ, Morzaria SP and Young AS (1997) The prevalence of serum antibodies to tick-borne infections in cattle in small holder dairy farms in Murang'a District, Kenya; a cross-sectional study. *Prev Vet Med* 30: 95-107.
  - Haine D, Boelaert F, Pfeiffer D, Saegerman C, Lonneux J-F, Losson B, Mintiens K (2004) Herd level seroprevalence and risk mapping of bovine hypodermosis in Belgian cattle herds. *Prev Vet Med* 65: 93-104.
2. Work through the following CDC exercise on oral contraceptive use and ovarian cancer: <http://www.cdc.gov/eis/casestudies/Xocovca.student.811-703.pdf>

## Example examination questions

1. Define the null and alternative hypothesis. (2005 oral)
2. What is a p value and how is it determined? (2005 oral)
3. What are the application, advantageous and limitations of parametric and non-parametric test? Give 2 examples of each type of test. (2005 oral)
4. Briefly describe the essential features, application and limitations of
  - a. Stratified analysis (2005 written)
5. Using examples write brief notes on
  - a. p values and significance testing (2003 written)
  - b. confidence intervals (2002 written)
  - c. power in epidemiological studies (2001 written)

**Additional reading/resources**

- Epidemiological Skills in Animal Health, PGFVS Proceedings 143; pp 222 - 236 (analysing epidemiologic data)
- McDermott JJ, Schukken YH and Shoukri MM. Study design and analytic methods for data collected from clusters of animals. *Prev Vet Med* 18: 175-191.
- McDermott JJ, Schukken YH. A review of methods used to adjust for cluster effects in explanatory epidemiological studies of animal populations. *Prev Vet Med* 18:155-173.

## Meeting 13: Data analysis 3

### Reading

Comprehensive and easy to understand references on this area are difficult to find, hence the large list of recommended reading below.

#### *Linear regression*

- Shott, S (1991) Statistics simplified: Regression. JAVMA 198: 798 - 801.  
OR
- Dohoo Chapter 14, pp 273-313 (for sections 14.5, 14.6,. 14.9 and 14.10, identify major points only)

#### *Logistic regression*

- Shott, S (1991) Statistics simplified: Logistic regression and discriminant analysis. JAVMA 198: 1902 - 1905.
- Bennett, D (2001) Logistic regression analysis: an epidemiological perspective. Australasian Epidemiologist 8: 38 - 44.  
OR
- Dohoo Chapter 16, pp 335-370 (for sections 16.6, 16.7, 16.10-16.14, identify major points only)

#### *Survival analysis*

- Shott, S (1991) Survival analysis. JAVMA 198: 1513 - 1515.
- Dohoo Chapter 19, sections 19.1-19.7.5, 19.8.2 pp 409-424, p 428  
OR
- Stevenson MA (2005). Introduction to Survival Analysis: 195.721 Study Guide. Massey University, Palmerston North, New Zealand. pp 1 – 11.

#### *Meta-analysis*

- Egger M, Davey Smith G (1997) Meta-analysis: Potentials and promise. British Medical Journal 315: 1371 - 1374.  
OR
- Thrusfield 3<sup>rd</sup> edn, pp 300-304 (meta analysis)
- Dohoo Chapter 24; pp 543-558  
AND
- Cochrane Collaboration Open Learning Material for Reviewers. URL: <http://www.cochrane-net.org/openlearning/>. Chapter 3.

## Presentations

Choose one of the statistical techniques listed above (linear regression, logistic regression, survival analysis, meta-analysis). Imagine you are to give a 20 minute presentation about this technique to a group of colleagues who know very little about epidemiology (or statistics). Your talk should include:

- A description of what the technique is used for.
- A [brief] description of the principles and assumptions that underpin the technique.
- Examples of where the technique has been used to provide insight into an animal health problem.
- What to look out for when evaluating studies that have used the technique.

## Exercises

Critically appraise three or more of the following four studies, describe the following:

- a) objectives of the data analyses
  - b) the unit(s) of analysis
  - c) types of data (outcome and exposure variables)
  - d) the type of analysis used (linear regression, logistic regression, survival analysis, meta-analysis) and why that approach was appropriate
  - e) any potential sources of clustering
- Arunvipas P, Dohoo I, VanLeeuwen J, Keefe G (2003) The effect of non-nutritional factors on milk urea nitrogen levels in dairy cows in Prince Edward Island, Canada. *Preventive Veterinary Medicine* 59, 83 - 93.
  - Borges V, Bernardi M, Bortolozzo F, Wentz I, (2005) Risk factors for stillbirth and foetal mummification in four Brazilian swine herds. *Preventive Veterinary Medicine* 70, 165 – 176 [logistic regression].
  - Matsuda R, Morizane T, (2005) *Helicobacter pylori* infection in dental professionals: A 6-year prospective study. *Helicobacter* 10, 307 – 311 [logistic regression].
  - Proudman C, Smith J, Edwards G, French N (2002) Long-term survival of equine surgical colic cases. Part 1: Patterns of mortality and morbidity. *Equine Veterinary Journal* 34, 432 – 437 [survival analysis].
  - Rabiee AR, Lean IJ, Stevenson MA (2004) A Bayesian meta-analysis of the effects of administering an intra-vaginal (CIDR) device in combination with other hormones on the reproductive performance of cycling, anoestrous and inseminated cows. *New Zealand Veterinary Journal* 52 (6): 384-393
  - Wilesmith JW, Ryan J BM, Stevenson MA, Morris RS, Pfeiffer DU, Lin D, Jackson R, Sanson RL (2000) Temporal aspects of the epidemic of bovine spongiform encephalopathy in Great Britain: holding-associated risk factors for the disease. *Veterinary Record* 147 (12): 319-325

- Singer ER, Saxby F, French NP (2003) A retrospective case-control study of horse falls in the sport of horse trials and three-day eventing. *Equine Veterinary Journal* 35 (2): 139-145
- Charlier J, Claerebout E, Muelenaere E. de, Vercruyse J (2005) Associations between dairy herd management factors and bulk tank milk antibody levels against *Ostertagia ostertagi*. *Veterinary Parasitology* 133 (1): 91-100
- Sanchez J, Dohoo I, Carrier J, DesCoteaux L (2004) A meta-analysis of the milk-production response after anthelmintic treatment in naturally infected adult dairy cows. *Preventive Veterinary Medicine* 63, 237 – 256 [meta-analysis].

### **Example examination questions**

1. Describe the application, advantages and limitations of linear regression in veterinary epidemiology.
2. Describe the application, advantages and limitations of logistic regression in veterinary epidemiology.
3. Briefly describe the essential features and application of survival analysis (2003 written).
4. Describe the application, advantages and limitations of meta-analysis (2005 oral).

### **Additional reading/resources**

- Noordhuizen J, Frankena K, van der Hoofd C, Graat E (1997) Application of Quantitative Methods in Veterinary Epidemiology. Wageningen Pers, Wageningen pp 135 – 178 [logistic regression].
- Ottenbacher K, Ottenbacher H, Tooth L, Ostir G (2004) A review of two journals found that articles using multivariable logistic regression frequently did not report commonly recommended assumptions. *Journal of Clinical Epidemiology* 57, 1147 – 1152 [logistic regression].
- Noordhuizen J, Frankena K, van der Hoofd C, Graat E (1997), Application of Quantitative Methods in Veterinary Epidemiology. Wageningen Pers, Wageningen pp 179 – 214 [survival analysis].