27

DESCRIPTION OF DATASETS

All datasets used in the examples and sample problems in this text are provided for pedagogical purposes only. They are provided so that the reader can (in conjunction with the program files listed in Chapter 28) recreate the examples included in the text, or compute solutions for sample problems provided. Contributors have made data available to the readers of this text on this understanding and consequently, this is the only use for which they are provided.

In some cases, datasets have been modified since the initial publication of results from the study which generated the data. In many cases, only a subset of the original data (*ie* a subset of variables or a subset of observations) are included. Consequently, the reader should not expect to be able to duplicate results obtained in the original publication.

In the descriptions that follow, unless otherwise specified, all variables coded 0 or 1 (0/1) have the following meaning:

0 = no, absent or negative
1 = yes, present or positive

All datasets can be downloaded from the Veterinary Epidemiologic Research website (http://www.upei.ca/ver). Datasets are directly accessible to Stata using Stata's internet commands (*eg* net from http://www.upei.ca/ver/data – see Stata documentation for details) or as zip files in a variety of statistics program formats (see website for details).

The authors extend their sincere thanks and appreciation to the contributors of these datasets.

Name **ap2**

Study type single cohort
# of records 1,114
Unit of record pig
Contributor Håkan Vigre

Reference
Vigre H, Dohoo IR, Stryhn H, Busch ME. Intra-unit correlations in seroconversion to *Actinobacillus pleuropneumonia* and *Mycoplasma hyopneumoniae* at different levels in Danish multi-site pig production facilities. Prev Vet Med 2003; accepted.

Brief description
Data were collected on 1,114 pigs from 35 batches produced on six farms that employed an 'all-in, all-out' production process. Pigs were weighed and blood sampled at the time of transfer from the weaner barn to the finisher barn (approximately 70 days of age) and again six weeks later (shortly before slaughter). Blood samples were tested for antibodies to *Actinobacillus pleuropneumonia* (Type 2), *Mycoplasma hyopnueumonia*, the influenza virus and the porcine respiratory and reproductive syndrome virus (PRRS). Two of the objectives of the study were to determine when seroconversion to the various agents occurred and at which level of the population (*eg* pig, batch or herd) most of the variation in seroconversion occurred.

(*continued on next page*)

Table of variables

Variable	Description	Codes/units
farm_id	farm identification	
batch_id	batch identification number	
litt_id	litter identification number	
pig_id	pig identification	
parity	farrowing number of sow	
vacc_mp	batch vaccinated against M. hyopneumoniae	0/1
seas_fin	season pigs in finishing unit	0 = summer 1 = winter
age_t	pig age at transfer from weaning to finishing unit	days
w_age_t	weight at age_t	kg
age_t6	age plus approx. 6 weeks	days
w_age_t6	weight at age_t6	kg
dwg_fin	daily weight gain between age_t and age_t6	gm
ap2_t	serological reac. against A. pleuropneumoniae serotype 2 at age_t	0/1
mp_t	serological reac. against M. hyopneumoniae at age_t	0/1
infl_t	serological reac. against influenza virus at age_t	0/1
prrs_t	serological reac. against PRRS virus at age_t	0/1
ap2_t6	serological reac. against A. pleuropneumoniae serotype 2 at age_t6	0/1
mp_t6	serological reac. against M. hyopneumoniae at age_t6	0/1
infl_t6	serological reac. against influenza virus at age_t6	0/1
prrs_t6	serological reac. against PRRS virus at age_t6	0/1
ap2_sc	seroconversion to ap2 during the finishing period	0/1

Name **beef_ultra**

Study type single cohort
# of records 487
Unit of record animal
Contributor Greg Keefe

Reference
Keefe GP, Dohoo IR, Valcour J, Milton RL. Assessment of ultrasonic imaging of marbling at entry into the feedlot as a predictor of carcass traits at slaughter. J Anim Sci 2003; submitted.

Brief description
Data were collected on 487 cattle at the time that they entered a feedlot for 'fattening' prior to slaughter. Data consisted of demographic information plus readings obtained from an ultrasonic evaluation of the animal. Ultrasound measurements of backfat thickness, loineye area and the percentage of intramuscular fat ('marbling') were obtained. The objective of the study was to determine if ultrasound examination of the animal at the time of entry into a feedlot was able to predict final carcass grade (AAA, AA or A). Carcass grade depends primarily on the amount of intramuscular fat in the carcass at the time of slaughter.

Table of variables

Variable	Description	Codes/units
farm	farm id	
id	animal id	
grade	carcass grade	1 = AAA 2 = AA 3 = A
breed	breed (known or estimated)	multiple
sex	gender	0 = female 1 = male
bckgrnd	animal backgrounded	0/1
implant	hormone implant used	0/1
backfat	backfat thickness	mm
ribeye	area of rib eye muscle	sq cm
imfat	intramuscular fat score	% of area
days	fattening period	days
carc_wt	carcass weight	kg

Name **bst_mast**

Study type meta-analysis
# of records 29
Unit of record group of cows
Contributor Ian Dohoo

Reference
Dohoo IR, Leslie KE, DesCôteaux L, Fredeen A, Shewfelt W, Preston A et al. A meta-analysis review of the effects of rBST. 2. Effects on animal health, reproductive performance and culling. Can J Vet Res 2003; in press.

Brief description
On request from Health Canada, the Canadian Veterinary Medical Association established an expert panel to review the production and health effect of recombinant bovine somatotropin (rBST) in dairy cattle. The panel carried out a meta-analysis of all available literature and evaluated a wide range of production and health effects. The data in this file consist of risk ratios for clinical mastitis that were associated with the use of rBST. Data from 29 distinct groups of cows, from 20 separate studies are included. The precision of the point estimate is included in the form of 95% confidence limits.

Table of variables

Variable	Description	Codes/units
study	study number	
group	cow group number	
parity	parity group	1 = primiparous 2 = all ages combined 3 = multiparous
study_yr	year of study	
rr	risk ratio	
cilow	lower 95% confidence limit	
cihigh	upper 95% confidence limit	
dur	duration of treatment	days
dose_day	daily dosage	mg/day

Name **bst_milk**

Study type meta-analysis
# of records 28
Unit of record group of cows
Contributor Ian Dohoo

Reference
Dohoo IR, Leslie KE, DesCôteaux L, Fredeen A, Dowling P, Preston A et al. A meta-analysis review of the effects of rBST. 1. Methodology and effects on production and nutrition related parameters. Can J Vet Res 2003; in press.

Brief description
On request from Health Canada, the Canadian Veterinary Medical Association established an expert panel to review the production and health effect of recombinant bovine somatotropin (rBST) in dairy cattle. The panel carried out a meta-analysis of all available literature and evaluated a wide range of production and health effects. The data in this file consist of change in level of milk production (fat-corrected milk) that were associated with the use of rBST. Data from 28 distinct groups of cows, from 19 separate studies are included. The precision of the point estimate is included both in the form of 95% confidence limits and the SE of the point estimate.

Table of variables

Variable	Description	Codes/units
study	study number	
group	cow group number	
parity	parity group	1 = primiparous 2 = all ages combined 3 = multiparous
study_yr	year of study	
diff	difference in milk production	
cilow	lower 95% confidence limit for difference	
cihigh	upper 95% confidence limit for difference	
se	standard error of difference	
ncows	number of cows in study	
dur	duration of treatment	days
dose_day	daily dosage	mg/day

Name **bvd_test**

Study type single cohort
# of records 2,162
Unit of record cow
Contributor Ann Lindberg

Reference
Lindberg A, Groenendaal H, Alenius S, Emanuelson U. Validation of a test for dams carrying foetuses persistently infected with bovine viral-diarrhoea virus based on determination of antibody levels in late pregnancy. Prev Vet Med 2003; 51: 199-214.

Brief description
Blood or milk samples were collected from 2,162 pregnant cows at various stages of lactation. Following the birth of their calf, the status of the calf with regard to persistent infection (PI) with the bovine viral diarrhea (BVD) virus was determined. The blood and milk samples were tested using an ELISA to determine levels of BVD virus antibodies in the cow. A variety of cutpoints were then examined to determine which gave the best combination of sensitivity and specificity for detecting PI+ calves. Logistic regression was used to evaluate the effects of other factors (particularly stage of lactation) on the estimated sensitivity and specificity of the test.

(*continued on next page*)

Table of variables

Variable	Description	Codes/units
cow_id	cow identification	
breed	breed	1 = red and white 2 = black and white 3 = beef 4 = other
parity	parity group	1 = primiparous 2 = all ages combined 3 = multiparous
pregmon	pregnancy month at test	
season	calving season	1 = winter 2 = spring 3 = summer 4 = autumn
spec	type of specimen	0 = milk 1 = blood
calfst	calf status	0/1
od	optical density	
co_5	test result dichotomised at 0.5	0/1
co_6	test result dichotomised at 0.6	0/1
	. . . etc . . .	
co_15	test result dichotomised at 1.5	0/1
co_16	test result dichotomised at 1.6	0/1

Name **calf**

Study type retrospective cohort
# of records 254
Unit of record calf
Contributor Jeanne Lofstedt

Reference
Lofstedt J, Dohoo IR, Duizer G. Model to predict septicemia in diarrheic calves. J Vet Int Med 1999; 13: 81-88.

Brief description
These data come from a retrospective analysis of the medical records from all diarrheic calves which were presented to Atlantic Veterinary College, PEI, Canada between 1989 and 1993. The ultimate objective of the study was to develop a logistic model which would predict whether or not the calf was septic at the time of admission (septic calves have a much poorer prognosis than non-septic calves and are not usually worth treating, given economic considerations).

There are 254 observations (records) and 14 variables in the dataset (calf). The original dataset had far more variables (including a lot of laboratory data) but this dataset contains only a subset of the demographic data and the physical examination data collected. All observations were determined on the day of admission, except for the outcome (sepsis) which was based on all data available at the time of death or discharge.

(*continued on next page*)

Table of variables

Variable	Description	Codes/units
case	hospital case number	
age	age at admission	days
breed	breed	coded 1-9
sex	sex	0 = female 1 = male
attd	attitude of calf	0 = bright, alert 1 = depressed 2 = unresponsive, comatose
dehy	% dehydration	
eye	uveitis/hypopyon clinically evident	0/1
jnts	swollen joints clinically evident	number of joints affected
post	posture of calf	0 = standing 1 = sternal 2 = lateral
pulse	pulse rate	beats per minute
resp	respiratory rate	breaths per minute
temp	rectal temperature	^{o}C
umb	swollen umbilicus clinically evident	0/1
sepsis	sepsis (outcome)	0/1

Name **calf_pneu**

Study type cohort
# of records 24
Unit of record calf
Contributor Iver Thysen

Reference
Thysen I. Application of event time analysis to replacement, health and reproduction data in dairy cattle research. Prev Vet Med 1988; 5: 239-250.

Brief description
These published data were used in one of the early publications in the veterinary literature discussing the use of survival analysis techniques. The data consist of mortality records from 24 calves that were housed in one of two housing systems: continuous housing, or batch (*ie* all-in all-out) housing.

Table of variables

Variable	Description	Codes/units
calf	calf id	
stock	stocking method	
days	time to death or censoring	days
died	died	

Name **colostrum**

Study type single cohort
# of records 180
Unit of record calf
Contributors Gilles Fecteau

Reference
None

Brief description
Data on the colostrum fed to 180 calves were collected from several dairy herds in Quebec. Herd identification was recoded to be the single large herd in the study compared to an amalgamation of small herds. The bacterial load in the colostrum was determined and the quantity of colostrum fed to the calf recorded. Calves were followed for three weeks and their health status over that period recorded as healthy (no illness), mild illness or serious illness (including death). The objective of the study was to determine if bacterial load in colostrum affected calf health.

Table of variables

Variable	Description	Codes/units
herd	herd of origin	0 = collection of small herds 1 = large herd
calf_id	calf identification	
health	health score	0 = healthy 1 = mild illness 2 = serious illness
qty	quantity of colostrum	litres
log_tot	natural log of total bacterial load	

Name **dairy_dis**

Study type survey (cross-sectional)
# of records 2454
Unit of record cow
Contributors John VanLeeuwen, Greg Keefe

Reference

VanLeeuwen J, Keefe GP, Tremblay R, Power C, Wichtel JJ. Seroprevalence of infection with *Mycobacterium avium* subspecies *paratuberculosis*, bovine leukemia virus, and bovine viral diarrhea virus in Maritime Canada dairy cattle. Can Vet J Res 2001; 42: 193-198.

Brief description

These data were collected as part of a prevalence survey of four infectious diseases of dairy cattle in eastern Canada. 30 herds in each of three provinces (Prince Edward Island, Nova Scotia and New Brunswick) were randomly selected from lists of all dairy herds participating on a milk-production monitoring program. Within each herd, approximately 30 animals were randomly selected and blood samples collected. These samples were tested for antibodies to: *Neospora caninum*, *Mycobacterium avium (*subsp. *paratuberculosis)* and enzootic bovine leukosis virus. In addition, a group of non-vaccinated heifers were bled and tested for bovine virus diarrhea virus, but these test results are not included in this dataset. Sampling weights were computed as the inverse of the product of the probability of a herd being selected and the probability of a cow being selected within a herd.

(*continued on next page*)

Table of variables

Variable	Description	Codes/units
prov	province	
herd	herd identification number	
cow	cow identification number	
lact	lactation number	
dim	days in milk	days
johnes	Johne's test result	0/1
leukosis	leukosis test result	0/1
neospora	neospora test results	0/1
tot_hrd	total herds in province	
prob_hrd	probability of herd being selected	
tot_cow	total cows in herd	
tot_smpl	total cows sampled in herd	
prob_cow	probability of cow being selected	
prob_smp	overall probability of a cow being selected	
weight	sampling weight	

Name **daisy**

Study type single cohort
# of records 162
Unit of record cow
Contributor Wayne Martin

Reference
None

Brief description
These data are based on real cow-reproduction data but have been modified in order to demonstrate a number of points related to linear regression. Consequently, they are now fictional data. They consist of data about the occurrence of a number of disease conditions which occur in the early post-partum period, along with measures of reproductive performance such as the interval from calving to first estrus, interval to first breeding and the calving to conception interval (all measured in days). The objective of the studies based on these data is to evaluate the effect of various diseases on reproductive performance.

Table of variables

Variable	Description	Codes/units
farmnum	farm identifier	
cownum	cow identifier	
firstest	first observed estrus	
firstbrd	postpartum to first breeding	days
calvcon	postpartum to conception	days
age	age	yrs
culled	cow removed from the herd	0/1
dayscull	postpartum to cow removal	days
endomet	endometritis	0/1
mastitis	mastitis	0/1
metritis	metritis	0/1
milkfev	milk fever	0/1
ovar	cystic ovarian disease	0/1
pyomet	pyometritis	0/1
retpla	retained fetal membranes (placenta)	0/1

Name **elisa_repeat**

Study type experimental
# of records 40
Unit of record milk sample
Contributor Javier Sanchez

Reference
Sanchez J, Dohoo IR, Markham RJF, Leslie KE, Conboy G. Evaluation of the repeatability of a crude adult indirect *Ostertagia ostertagi* ELISA and methods of expressing test results. Vet Parasitol 2002; 109: 75-90.

Brief description
Forty individual cow milk samples were repeatedly tested (six times) using a crude *Ostertagia* antigen indirect ELISA. Results were recorded both as raw optical density (OD) values and values adjusted based on the readings for the positive and negative controls in the plate.

Table of variables

Variable	Description	Codes/units
id	sample identification	
raw1	raw OD - sample #1	
raw2	raw OD - sample #2	
	. . .etc . . .	
raw6	raw OD - sample #6	
adj1	adjusted OD - sample #1	
adj2	adjusted OD - sample #2	
	. . .etc . . .	
adj6	adjusted OD - sample #6	

Name **fec**

Study type single cohort
# of records 2,250
Unit of record monthly fecal egg count
Contributors Ane Nødtvedt, Javier Sanchez, Ian Dohoo

Reference

Nødtvedt A, Dohoo IR, Sanchez J, Conboy G, DesCôteaux L, Keefe GP et al. The use of negative binomial modelling in a longitudinal study of gastrointestinal parasite burdens in Canadian dairy cows. Can J Vet Res 2002; 66: 249-257.

Brief description

Monthly (in some herds less frequently) fecal egg samples were collected from lactating age dairy cows (n=313) in 38 herds over a period of 1 year. The data were collected as part of a multifaceted study into parasitism in lactating dairy cows which included a longitudinal epidemiologic investigation and a controlled trial of the effects of deworming at calving with eprinomectin. The effects of factors at the sampling-day, cow and herd levels on fecal egg counts were evaluated.

Table of variables

Variable	Description	Codes/units
province	Canadian province	1 = PEI 2 = Quebec 3 = Ontario 4 = Saskatchewan
herd	herd identifier	
cow	unique cow identifier	
visit	visit number	
tx	eprinomectin treatment at calving	0/1
fec	fecal egg count	eggs/5 gm
lact	lactation	0 = primiparous 1 = multiparous
season	season	1 = oct-dec 99 2 = jan-mar 00 3 = apr-jun 00 4 = jul-sep 00
past_lact	lactating cows have access to pasture	0/1
man_heif	manure spread on heifer pasture	0/1
man_lact	manure spread on cow pasture	0/1

Name **feedlot**

Study type case-control
# of records 588
Unit of record animal
Contributor Wayne Martin

References

1. Martin, SW, Harland, RJ, Bateman, KG, Nagy, É. The association of titres to *Haemophilus somnus* and other putative pathogens, with the occurrence of bovine respiratory disease and weight gain in feedlot calves. Can J Vet Res 1998; 62: 262-267.
2. Martin, SW, Nagy, É, Shewen, PE, Harland, RJ. The association of titres to bovine coronavirus with treatment for bovine respiratory disease and weight gain in feedlot calves. Can J Vet Res 1998; 62: 257-261.

Brief description

This dataset represents the combined data from a number of studies on the role of specific micro-organisms as causes of bovine respiratory disease (BRD). Typically these beef cattle enter feedlots in the fall of the year and approximately 30% will develop BRD. The general strategy for the studies was to bleed all of the animals on arrival at the feedlot and again 28 days later (since most of the occurrence of BRD occurs in that time period). For analyses, we used all of the samples from cases and an approximately equal number from controls. In some of the smaller groups we used all samples and hence in these the study design was essentially a single cohort. The studies were conducted at essentially the same feedlots in different years, but depending on their size, one feedlot could have numerous groups of calves on the study in any given year.

The titres were recorded in a quantitative manner but have been dichotomised in this dataset.

Note At the time these data were collected, one of the important bacteria was called Pasteurella hemolytica and it is referred to as such in this dataset. Elsewhere in the text it is referred to by its newer name Mannheima hemolytica.

(*continued on next page*)

Table of variables

Variable	Description	Codes/units
group	group identification	
tag	eartag number	
province	province of feedlot	1 = Alberta 2 = Ontario
brd	clinical bovine respiratory disease (case-control)	0/1
brsvpos	arrival titre to brsv	0/1
brsvsc	seroconversion to brsv during study	0/1
bvdpos	arrival titre to bvd virus	0/1
bvdsc	seroconversion to bvd during study	0/1
ibrpos	arrival titre to ibr virus	0/1
ibrsc	seroconversion to ibr virus during study	0/1
pipos	arrival titre to pi3 virus	0/1
pisc	seroconversion to pi3 virus during study	0/1
phcypos	arrival titre to Ph cytotoxin	0/1
phcysc	seroconversion to Ph cytotoxin during study	0/1
phaggpos	arrival titre to Ph agglutinins	0/1
phaggsc	seroconversion to Ph during study	0/1
hspos	arrival titre to Hs	0/1
hssc	seroconversion to Hs during study	0/1
wt0	arrival weight	kg
wt28	28-day weight	kg

Name **fish_morts**

Study type cross-sectional
# of records 236
Unit of record sea-cage
Contributors Larry Hammell, Ian Dohoo

References

1. Hammell KL, Dohoo IR The epidemiology of hemorrhagic kidney syndrome - infectious salmon anemia in Atlantic salmon in Atlantic Canada. Bristol, England: Society for Veterinary Preventive Epidemiology and Preventive Medicine, 1999.
2. Hammell KL, Dohoo IR Challenges of investigating mortality patterns and management factors associated with ISAV outbreaks in eastern Canada. Paris, France: O.I.E., Risk Analysis in Aquatic Animal Health, 2000.
3. Hammell KL, Dohoo IR. Mortality patterns in Infectious Salmon Anemia virus outbreaks in New Brunswick, Canada. Journal of Fish Diseases 2003; accepted.

Brief description

Following the introduction of infectious salmon anemia virus to the Bay of Fundy (Canada), an epidemiological investigation of risk factors for the disease was initiated. At the time the study was started, the etiology of the mortalities was not known and cages were designated as 'outbreaks' or not, based on the pattern of mortalities observed in the cage. One of the risk factors identified as being associated with an increased risk of elevated mortalities was the feeding of dry (as opposed to wet- or moist-) feed. These data are a very small subset of the original data collected. They describe mortalities in 236 cages from 16 different sites over a period of just a few days (depending on the number of days between dives for collection of dead fish).

Table of variables

Variable	Description	Codes/units
cage_id	cage id	
days	time since last dive	days
morts	number of dead fish retrieved	
fish	estimated number of fish in cage	
feed	dry feed (compared to wet)	0 = wet 1 = dry

Name **isa_risk**

Study type cross-sectional
# of records 182
Unit of record sea cage
Contributors Larry Hammell, Ian Dohoo

References

1. Hammell KL, Dohoo IR The epidemiology of hemorrhagic kidney syndrome - infectious salmon anemia in Atlantic salmon in Atlantic Canada. Bristol, England: Society for Veterinary Preventive Epidemiology and Preventive Medicine, 1999.
2. Hammell KL, Dohoo IR Challenges of investigating mortality patterns and management factors associated with ISAV outbreaks in eastern Canada. France: O.I.E., Risk Analysis in Aquatic Animal Health, 2000.
3. Hammell KL, Dohoo IR. Mortality patterns in infectious salmon anemia virus outbreaks in New Brunswick, Canada. Journal of Fish Diseases 2003; accepted.

Brief description

Following the introduction of infectious salmon anemia virus to the Bay of Fundy (Canada), an epidemiological investigation of risk factors for the disease was initiated. At the time the study was started, the etiology of the mortalities was not known and cages were designated as 'outbreaks' or not, based on the pattern of mortalities observed in the cage. A large number of risk factors were evaluated and this dataset consists of the records for 182 cages which had complete data on a subset of those factors (see list below). While the factors listed below were all fixed factors (*ie* didn't change during the study period), the data were used to compute a time-varying factor: whether or not there had been another positive cage (net-pen) at the site. This was used in survival models of the time to the occurrence of an outbreak.

Table of variables

Variable	Description	Codes/units
sitepen	(1000*site)+cage identifier	
site	site identifier	
net-pen	cage identifier	
datestrt	date fish first put in cage	
apr01_97	April 1 1997	
date	date of outbreak OR censoring	
case	case (outbreak)	0/1
cummrt96	cum. mort. during 1996	
size	cage size	0 = <10,000 1 = >10,000
par	initial population at risk in cage (number of fish)	
numcage	number of cages at site	

Name **isa_test**

Study type cross-sectional
# of records 1,071
Unit of record fish
Contributors Carol McClure, Larry Hammell

Reference
McClure C, Hammell KL, Stryhn H, Dohoo IR, Hawkins LJ. Application of surveillance data in the evaluation of infectious salmon anemia diagnostic tests. Dis of Aquatic Org 2003; submitted.

Brief description
Following the identification of the infectious salmon anemia virus in the Bay of Fundy (Canada), a lot of fish were tested using a variety of diagnostic tests. It was realised that tests often gave conflicting results and the available data were used to provide a preliminary evaluation of the operating characteristics of each test. Fish that were derived from sea cages (net-pens) that had a confirmed outbreak of ISA were considered to be 'gold standard positive.' Fish sampled from sites which did not have any outbreaks of ISA (in any cages) during the study period were considered 'gold standard negative.' Other fish sampled were not included in this study. Test results from a total of 1071 fish that had multiple tests performed and which could be classified as positive (n=264) or negative (n=807) were included in the dataset.

Table of variables

Variable	Description	Codes/units
id	case identification	
date	submission date	
site	site identification	
cage	cage identification	
subm	submission identification	
fish	fish number for each case	
dz	disease status (clinical)	
histo	histology	0 = negative 1 = suspicious 2 = positive
histo_np	histo neg/pos (pos=susp+pos)	0/1
ifat1	IFAT laboratory 1	0-4
ifat1_np	IFAT-lab1 neg/pos (pos is ≥ 1)	0/1
ifat2	IFAT - laboratory 2	0-4
ifat2_np	IFAT-lab2 neg/pos (pos is ≥ 2)	0/1
pcr	PCR	0/1
vi	virus isolation	0/1

Name **lympho**

Study type clinical trial (fictional)
# of records 300
Unit of record dog
Contributor Ian Dohoo

Reference
None

Brief description
These data are from a fictional clinical trial of two treatments for lymphosarcoma in dogs. The study was (hypothetically) conducted as a multicentre (n=7 clinics) controlled trial. Dogs meeting the eligibility criteria for entry into the trial (n=300) had the tumour surgically removed (only dogs with tumours which could be surgically removed were eligible) and then were randomly assigned to one of four treatment groups: no treatment, radiation only, chemotherapy only and both radiation and chemotherapy. Dogs were randomly assigned within each centre, so the total number of dogs on each treatment group are not exactly equal for all treatments. Each dog was followed from the time of treatment until it died from a relapse of the lymphosarcoma or was lost to follow-up (*eg* died of other causes, owner moved away from the study site) and the time to the occurrence of either of those was recorded.

Table of variables

Variable	Description	Codes/units
dogid	the dog's study identification number	
age	age of dog in years when it was diagnosed with lymphosarcoma	yrs
rad	whether or not the dog received radiation therapy	0/1
chemo	whether or not the dog received chemotherapy	0/1
died	whether the dog died or was lost to follow-up	0/1
months	the number of months after the start of therapy before the dog died or was lost to follow-up	mo

Name **nocardia**

Study type case-control
# of records 108
Unit of record herd
Contributors Lynn Ferns, Ian Dohoo

Reference
Ferns L, Dohoo IR, Donald A. A case-control study of Nocardia mastitis in Nova Scotia dairy herds. Can Vet J Res 1991; 32: 673-677.

Brief description
This dataset contains a subset of the data obtained from a case-control study of Nova Scotia dairy herds with and without Nocardia mastitis. There had been a dramatic increase in the incidence of Nocardia mastitis in Canada since 1987 and this study was carried out to identify risk factors associated with the occurrence this disease. A total of 54 case herds and 54 control herds were visited for data-collection purposes during the summer of 1989.

(*continued on next page*)

Table of variables

Variable	Description	Codes/units
id	herd identification number	
casecont	case/control status of herd	0 = control 1 = case
numcow	number of cows milked	
prod	average milk production for the herd	kg/cow/day
bscc	average bulk-tank SCC over the first 6 months of 1988	‘000s of cells/ml
dbarn	type of barn dry cows kept in	1 = freestall 2 = tiestall 3 = other
dout	type of outdoor area used for dry cows	1 = pasture 2 = yard/drylot 3 = none 4 = other
dcprep	method of teat end preparation prior to dry cow therapy administration	1 = no prep. 2 = washed only 3 = washed and disinfected 4 = dry cow therapy not used
dcpct	percent of dry cows treated with dry-cow therapy	%
dneo	dry-cow product containing neomycin used on farm in last year	0/1
dclox	dry cow product containing cloxacillin used on farm in last year	0/1
doth	Other dry cow products used (eg penicillin or novobiocin based) used on farm in last year	0/1

Name **pgtrial**

Study type clinical trial
# of records 319
Unit of record cow
Contributor Jeff Wichtel

Reference
None

Brief description
A clinical trial of the effect of prostaglandin administration at the start of the breeding period was carried out in three North Carolina dairy herds. On each of the three farms, the producer determined when he was ready to start breeding cows in his herd and at that time, cows were randomly assigned to receive a single injection of prostaglandin or a placebo. These cows were then followed (up to a maximum of 346 days) until they conceived (confirmed by rectal examination) or were culled. In addition to evaluating the effect of treatment on reproductive performance, three other factors were considered (parity, body condition score and herd).

Table of variables

Variable	Description	Codes/units
herd	herd identification number	
cow	cow identification number	
tx	treatment	0/1
lact	lactation number	
thin	body condition	0 = normal 1 = thin
dar	days at risk	days
preg	pregnant or censored	0 = censored 1 = pregnant

Name **pig_adg**

Study type cross-sectional
# of records 341
Unit of record pig
Contributor Theresa Bernardo

References

1. Bernardo TM, Dohoo IR, Donald A, Ogilvie T, Cawthorne R. Ascariasis, respiratory disease and production indices in selected PEI swine herds. Can J Vet Res 1990; 54: 267-273.
2. Bernardo TM, Dohoo IR, Ogilvie T. A critical assessment of abattoir surveillance as a screening test for swine ascariasis. Can J Vet Res 1990; 54: 274-277.
3. Bernardo TM, Dohoo IR, Donald. Effect of ascariasis and respiratory disease on growth rates in swine. Can J Vet Res 1990; 54: 278-284.

Brief description

These are data on the growth performance and abattoir findings of pigs from a selection of Prince Edward Island, Canada farms. The data were collected to study the inter-relationships among respiratory diseases (atrophic rhinitis and enzootic pneumonia), ascarid levels and daily weight gain. Atrophic rhinitis score was determined by splitting the snout and measuring the space ventral to the turbinates. An adjustment to the score was made if the nasal septum was deviated. Lung scores were recorded on a scale of 0 to 3 (negative to severe pneumonia) and then converted to either the presence or absence of pneumonia. Parasite burdens were evaluated using fecal egg counts, counts of adult worms in the intestine and visual assessment of the liver for ascarid tracks. Production data were recorded by monitoring the pigs on the farms of origin from birth through to slaughter.

(*continued on next page*)

Table of variables

Variable	Description	Codes/units
farm	farm identification number	
pig	pig identification number	
sex	sex of the pig	0 = female 1 = castrate
dtm	days to market (ie from birth to slaughter)	days
adg	average daily weight gain	gm
mm	measurement of snout space	mm
ar	atrophic rhinitis score	0-5
lu	lung score for enzootic pneumonia	0 = negative 1 = mild 2 = moderate 3 = severe
pn	pneumonia (lu>0)	0/1
epg5	fecal gastrointestinal nematode egg count at time of slaughter	eggs/5 gm
worms	count of nematodes in small intestine at time of slaughter	
li	liver score (based on number of parasite induced 'white spots')	0 = negative 1 = mild 2 = severe
ar2	severe atrophic rhinitis (ar>4)	0/1

Name **pig_farm**

Study type cross-sectional
# of records 69
Unit of record farm
Contributor Dan Hurnik

Reference

1. Hurnik D, Dohoo IR, Donald AW, Robinson NP. Factor analysis of swine farm management practices on Prince Edward Island. Prev Vet Med 1994; 20: 135-146
2. Hurnik D, Dohoo IR, Bate LA. Types of farm management as risk factors for swine respiratory disease. Prev Vet Med 1994; 20: 147-157.

Brief description

A cross-sectional study of pig farms in Prince Edward Island (Canada) was carried out to investigate risk factors for respiratory diseases (enzootic pneumonia and pleuritis). The prevalence of each disease was determined at slaughter from routine evaluations of thoracic viscera. Data on risk factors were collected by the investigator during visits to each farm. Data on a wide variety of factors were collected and the challenge was to sort out relationships among these factors and between them and the respiratory diseases given a very limited sample size.

(*continued on next page*)

Table of variables

Variable	Description	Codes/units
farm_id	farm identification	
pneu	pneumonia prevalence	
pncode	pneumonia - categorical (3 levels)	0 < 10% 1 = 10-40% 2 > 40%
pleur	pleuritis prevalence	
plcode	pleuritis - categorical (3 levels)	0 = 0% 1 = 0-8% 2 > 8%
num	number of pigs examined at slaughter	
size	herd size	
growth	average daily gain	gm/day
cmpfd	pigs fed complete mixed feed	0/1
suppl	supplement added to feed	0/1
prmx	premix fed	0/1
strmed	starter ration medicated	0/1
selenium	selenium added to feed	0/1
dryfd	feed fed dry (vs wet)	0 = wet 1 = dry
flrfd	pigs fed on floor	0/1
rooms	number of separate rooms in barn	
m3pig	air volume per pig	m^3
shipm2	density (pigs shipped per m^2)	pigs/m^2
exhaust	exhaust fan capacity (proportion of recommendation)	
inlet	air inlet size (proportion of recommendation)	
maninlt	manual adjustment of air inlets	0/1
mixmnr	manure mixed between pens	0/1
straw	straw bedding used	0/1
washpns	frequency of pen washings (per yr)	
strdnst	floor space - starter hogs (sq m)	m^2
grwdnst	floor space - grower hogs (sq m)	m^2
fnrdnst	floor space - finishing hogs (sq m)	m^2

(*continued on next page*)

Table of variables

Variable	Description	Codes/units
lqdmnr	manure handled as a liquid	0/1
floor	floor slatted	0/1
sldprtn	solid partitions between some pens	0/1
hlfsld	half-solid partitions between some pens	0/1
pigwtr	pigs per water nipple	
numpen	number of pens	
mixgrp	pigs from multiple groups mixed	0/1
hldbck	slow growing pigs held back from slaughter	0/1
dstfrm	distance (km) to nearest hog farm	km
hmrsd	all pigs home raised	0/1
nmbsrc	number of sources of pigs	
mnlds	only minimal disease pigs raised	0/1
vet	veterinary visits per year	
feedsls	feed salesman visits per year	
neighbr	neighbour visits per year	
pigprdc	pig producer visits per year	
trucker	trucker visits per year	
you	owner works in barn	0/1
family	family members work in barn	0/1
hrdhlp	hired help works in barn	0/1
exprnce	years of experience	yrs

Name **prew_mort**

Study type cross-sectional
# of records 6552
Unit of record litter
Contributor Jette Christensen

Reference
Christensen J, Svensmark B. Evaluation of producer-recorded causes of preweaning mortality in Danish sow herds. Prev Vet Med 1997; 32: 155-164.

Brief description
These data are a subset of 16 herds from a dataset collected by Jette Christensen in Denmark to study factors affecting preweaning mortality in pigs. These data have three levels in the hierarchy (litters (*n*=6552) within sows (*n*=3162) within farms (*n*=16)):

The key outcome of interest is preweaning mortality with a litter classified as having preweaning mortality or not if one or more piglets died before weaning.

Table of variables

Variable	Description	Codes/units
herd	unique herd id	
sowid	unique sow id	
litter	unique litter id	
lmort	prewmort in litter	0/1
herdtype	herd type	0 = production l = breeding herd
year		
month	month	jan = l dec = 12
quarter	quarter of year	l = jan-mar 2 = apr-jun 3 = jul-sept 4 = oct-dec
sow_parity	parity of sow	
sow_tx	sow required treatment (2d before to 7d after farrowing)	0/1
dead	number of dead piglets in litter	
lsize	litter size	
n	number at risk in litter	
stillb	number stillborn	

Name **reu_cc**

Study type single cohort
# of records 2509
Unit of record lactation
Contributors Emmanuel Tillard, Ian Dohoo

Reference
Dohoo IR, Tillard E, Stryhn H, Faye B. The use of multilevel models to evaluate sources of variation in reproductive performance in dairy cattle in Reunion Island. Prev Vet Med 2001; 50: 127-144.

Brief description
These data were collected as part of an ongoing research programme into dairy cattle fertility being carried out on Reunion Island (a French overseas department located in the Indian ocean) by researchers with CIRAD (Centre de Coopération Internationale en Recherche Agronomique pour le Développement). Two separate datasets have been compiled. This one contains data about the calving to conception interval, while the second had data on the interval from calving to first service. The data have a 4-level hierarchy (lactations (*n*=2509) within cows (*n*=1345) within herds (*n*=50) within geographic regions (*n*=5)).

Table of variables

Variable	Description	Codes/units
region	geographic region	
herd	herd number	
cow	unique cow number	
obs	unique observation number	
lact	lactation number	
cc	calving to conception interval	days
lncc	calving to conception interval - log transformed	
lncfs_ct	calving to first service interval - log transformed and centred	
heifer	age	0 = multiparous 1 = primiparous
ai	type of insemination at first service	0 = natural 1 = ai

Name **reu_cfs**

Study type single cohort
# of records 3027
Unit of record lactation
Contributors Emmanuel Tillard, Ian Dohoo

Reference
Dohoo IR, Tillard E, Stryhn H, Faye B. The use of multilevel models to evaluate sources of variation in reproductive performance in dairy cattle in Reunion Island. Prev Vet Med 2001; 50: 127-144.

Brief description
These data were collected as part of an ongoing research program into dairy cattle fertility being carried out on Reunion Island (a French overseas department located in the Indian ocean) by researchers with CIRAD (Centre de Coopération Internationale en Recherche Agronomique pour le Développement). Two separate datasets have been compiled. This one contains data on the interval from calving to first service. The data have a 4-level hierarchy lactations (*n*=3027) within cows (*n*=1575) within herds (*n*=50) within geographic regions (*n*=5)].

A second dataset containing only the first recorded lactation within each cow was saved as reu_cfs_1lact.

Table of variables

Variable	Description	Codes/units
region	geographic region	
herd	herd number	
cow	unique cow number	
obs	unique observation number	
lact	lactation number	
cfs	calving to first service interval	days
lncfs	calving to first service interval - log transformed	
fscr	first service conception	0/1
heifer	age	0 = multiparous 1 = primiparous
ai	type of insemination at first service	0 = natural 1 = ai

Name **sal_outbrk**

Study type matched case-control
# of records 112
Unit of record individual (person)
Contributor Tine Hald

References

1. Molbak K, Hald DT. An outbreak of Salmonella typhimurium in the county of Funen during late summer. A case-controlled study. Ugeskr Laeger 1997; 159(36): 5372-5377.
2. Hald DT. *Salmonella* in pork: Epidemiology, control and the public health impact. Copenhagen: Royal Veterinary & Agric. Univ. 2001.

Brief description

The data are from an investigation of an outbreak of Salmonella in Funen County of Denmark in 1996. The data consisted of 39 cases of Salmonella typhimurium phage type 12 and 73 controls matched for age, sex and municipality of residence. Data on numerous food exposures were recorded and a small subset of those data are included in the dataset -sal_outbrk-.

(*continued on next page*)

Table of variables

Variable	Description	Codes/units
match-grp	case-control pair identifier	
date	interview date	
age	age	yrs
gender	gender	0 = male 1 = female
casecontrol	case-control status	0/1
eatbeef	ate beef in previous 72 hours	0/1
eatpork	ate pork in previous 72 hours	0/1
eatveal	ate veal in previous 72 hours	0/1
eatlamb	ate lamb in previous 72 hours	0/1
eatpoul	ate poultry in previous 72 hours	0/1
eatcold	ate cold sliced meats in previous 72 hours	0/1
eatveg	ate vegetables in previous 72 hours	0/1
eatfruit	ate fruit in previous 72 hours	0/1
eateggs	ate eggs in previous 72 hours	0/1
slt_a	ate pork processed at slaughterhouse A	0/1
dlr_a	ate pork marketed by wholesaler A	0/1
dlr_b	ate pork marketed by wholesaler B	0/1

Name **scc_40**

Study type longitudinal
# of records 14,357
Unit of record test-day observations
Contributors Jens Agger and Danish Cattle Organization, Paul Bartlett, Henrik Stryhn

Reference
Stryhn H, Andersen JS, Bartlett PC, Agger JFA. Milk production in cows studies by linear mixed models. Proc. of symposium in applied statistics, Copenhagen, January 2001. Proceedings (ed. Jensen NE. Linde P): 1-10.

Brief description
These data are a very small subset of a large mastitis dataset collected by Jens Agger and the Danish Cattle Organization. This dataset contains records from 14,357 test-day observations in 2,178 cows from 40 herds. Milk weights (production records) were collected approximately monthly, and only records from a single lactation for each cow were included in this dataset. Factors that may have affected the somatic cell count (SCC) were also recorded. The major objective of this study was to determine if the relationship between the somatic cell count and milk production varies for cows with different characteristics (age, breed, grazing or not *etc*).

A subset of these data called -scc40_2level- was created by only taking the first observation for each cow, thereby reducing the dataset to two levels (herds and cows).

Table of variables

Variable	Description	Codes/units
herdid	herd id	
cowid	cow id	
test	approximate month of lactation	0 to 10
h_size	average herdsize	
c_heifer	parity of the cow	1 = heifer 0 = multiparous
t_season	season of test day	l = jan-mar 2 = apr-jun 3 = jul-sep 4 = oct-dec
t_dim	days in milk on test-day	days
t_lnscc	log somatic cell count on test day	

Name **smpltype**

Study type longitudinal
# of records 1114
Unit of record pig
Contributor Håkan Vigre

Reference
Vigre H, Dohoo IR, Stryhn H, Busch ME. Intra-unit correlations in seroconversion to *Actinobacillus pleuropneumonia* and *Mycoplasma hyopneumoniae* at different levels in Danish multisite pig production facilities. Prev Vet Med 2003; submitted.

Brief description
These data are derived from the -ap2- dataset. In addition to the original data, this dataset contains indicator variables (made up) that identify those pigs that were part of a: simple random sample, systematic random sample, stratified random sample, cluster sample and multistage sample.

Table of variables

Variable	Description	Codes/units
farm_id	farm identification	
pig_id	pig identification	
barn_ord	order of pigs in barn (as they were walked down an alley)	
parity	farrowing number of the sow (categorised)	1 = 1 2 = 2 3 = 3-4 4 = 5+
dwg_t	daily weight gain to approx 65 days	gm/day
smp_srs	pig in the simple random sample	0/1
smp_syst	pig in the systematic random sample	0/1
smp_strat	pig in the stratified (by parity) sample	0/1
smp_clust	pig in the cluster (by herd) sample	0/1
smp_ms	pig in the multistage sample (psu=herd)	0/1

Name **tb_real**

Study type retrospective cohort
# of records 134
Unit of record animal groups
Contributors Ian Dohoo, Fonda Munroe

References

1. Munroe FA, Dohoo IR, Mcnab WB. Estimates of within-herd incidence rates of *Mycobacterium bovis* in Canadian cattle. Prev Vet Med 2000; 45: 247-256.
2. Munroe FA, Dohoo IR, Mcnab WB, Spangler L. Risk factors for the between-herd spread of *Mycobacterium bovis* in Canadian cattle and cervids between 1985 and 1994. Prev Vet Med 1999; 41: 119-133.

Brief description

A retrospective evaluation of all (*n*=9) outbreaks of tuberculosis in domestic animals (dairy and beef cattle, cervids and bison) in Canada between the years of 1985 and 1994 was carried out to investigate risk factors for the spread of tuberculosis within and between herds. Detailed records from the epidemiologic investigation of all outbreaks (including records on all contact herds) were reviewed and a summary of each outbreak prepared. This dataset contains data only from herds in which tuberculosis was observed. In each herd, the most probably date on which the infection entered the herd was determined and the number of new cases arising within the herd determined from the herd testing results. The number of animals in each age, sex and type group was determined and the number of animal days at risk was computed. The effects of age (three groups), sex (two groups), and animal type (five groups) on the incidence rate of new infections was investigated. **Note** To meet confidentiality and regulatory concerns, these data have deliberately been falsified.

Table of variables

Variable	Description	Codes/units
obs	observation number	
farm_id	farm identification	
type	type of animal	1 = dairy cattle 2 = beef cattle 3 = cervid 4 = other
sex	sex	1 = female 2 = male
age	age category	1 = 0-12 mo 2 = 12-24 mo 3 = >24 mo
reactors	number of pos/reactors in the group	
par	animal days at risk in the group	