Applied Epidemiology in Feedlot Practices

M. D. Salman, *BVMS, PhD, Dipl ACVPM, FACE*
College of Veterinary Medicine and Biomedical Sciences
Colorado State University, Ft. Collins, CO 80523

L.L. Hungerford, *DVM, MPH, PhD*
College of Veterinary Medicine
University of Illinois, Urbana, IL 61801

Definition of epidemiology:

Epidemiology is the study of the distribution and determinants of diseases in **populations** and the development of strategies to improve **health** and **productivity** in these populations.

More informally, epidemiology has been defined to be "common sense, but with a system" (Austin & Werner, 1974), which captures the practical orientation of the discipline.

Purposes and Applications of Epidemiology:

1. Define the extent of a problem.

2. Search for causes and determinants of the occurrence and distribution of diseases or production inefficiencies.

3. Determine the mode of transmission (for infectious diseases).

4. Describe the historical trend of a disease.

5. Suggest, recommend and evaluate a control strategy and its effectiveness.

Types of Epidemiology:

1. Descriptive Epidemiology

Descriptive epidemiology includes any general observational study of the occurrence of disease or other health related event in an animal population. The major characteristics in descriptive epidemiology can be classified under the headings animal, place, and time. Types of descriptive epidemiologic studies include surveys, case reports, and case studies.

2. Analytical Epidemiology

Analytical epidemiology is a hypothesis testing method to investigate the association between a disease or health event and possible causative factors. Analytical epidemiologic studies may be either experimental or observational.

There is an unique opportunity in feedlot situations to apply one of the most powerful epidemiologic approaches to establish causation: the **cohort** study design.

What is a cohort approach?: It is a follow-up of a group of animals, which share a distinctive common character, for a definite time period to monitor a specific outcome (such as disease or a production parameter). The main purpose of a cohort approach is to establish a cause-effect relationship between a specific exposure (character) and an outcome.

Causation in Disease Processes

Ideally, the goal in epidemiology would be to find cause-effect relationships. For simple situa-tions, Koch's postulates were used to establish causal-ity. But for chronic, complex or production related diseases, Koch's postulates don't apply. Therefore, epidemiologists look first for "associations" between potential causative factors and diseases, and then use a set of criteria to assess causality. These guidelines were developed in the late 1960's, when the U. S. Surgeon General put together an expert panel to evaluate whether cigarette smoking caused lung cancer in humans. Since all of the research on the subject was observational rather than experimental (for obvious reasons since you can't randomly assign people to smoke, then follow them for 40 years to see if they develop lung cancer), the panel developed a set of criteria to judge causality using data from observational studies. These also work well for other epidemiologic studies.

1. **Temporality** - The cause must precede the effect in occurrence.

2. **Strength of association** - The greater the risk for disease in animals exposed to a causal factor relative to those that were not, the more likely that it is causal.

3. **Consistency** - Findings should be repeatable across different studies in different settings with different researchers, within reason.

4. **Specificity** - If a cause leads to one specific type of disease or a disease is associated with only one cause, this strengthens the evidence for causality. However, many environmental factors are known to lead to multiple diseases and so this guideline may not apply.

5. **Biologic Plausibility** - The proposed relationship should make biological sense.

6. **Biologic Gradient** - A dose-response relationship should be present, where increasing levels of exposure to the cause lead to greater incidence, severity or duration of disease.

7. **Coherence** - There should not be a conflict between the proposed causal relationship and what is already known about the disease.

8. **Experimental Evidence** - Available additional information from experimental studies can strengthen the evidence for causality.

Potential Causative Factors

Determinants (causes) of disease may be either necessary or sufficient causes. A necessary cause is one without which the disease cannot occur. For example, *E. coli* is a necessary cause of colibacillosis, although not of diarrhea; a sufficient cause is one that always produces the disease. Necessary or sufficient causes are rarely single factors, but are usually groups of factors known as component factors. Generally, these component factors for a disease process include an interaction between factors from the host, the environment, and the agent.

1. Host factors

Host factors play an important role in the development of the disease. Host factors are divided into intrinsic and extrinsic factors. Intrinsic host factors are those factors that are predetermined and cannot be changed. These include age, sex, genetics, etc. Extrinsic host factors are those factors that can be controlled and manipulated. These may include diet, immunization status, animal use, animal movements, contacts, behavior, and castration status.

2. Environmental factors

It has long been known that disease occurrence can be related to the environment to which the animal is exposed. These factors may include weather, housing, hygiene and sanitation, season of the year, and ubiquity of endemic agents.

3. Agent factors

Disease agents can be nutritional, chemical, physiological, genetic, parasitic (infectious agents), or commonly a combination of these.

Measures of Association

For a factor to be associated with a disease, the rate of disease in exposed animals (or animals with the factor) must be different from the rate of disease in those animals not exposed to (or without) the factor. For example, if the incidence rate of runny nose is 24% among steers in feedlot A, and the incidence rate of runny nose is 3% among steers in feedlot B, then runny nose and feedlot may be associated.

Relative Risk (RR)

The concept of risk is prominent in all health related sciences. Broadly, risk refers to the prob-ability that an individual will develop the disease over a specified period of time. Often, risk is calculated for members of specific groups (i.e. those steers in feedlot A and those in feedlot B). Relative risk is the ratio between the rate (or risk) of disease in an exposed group compared to the rate in an unex-posed group. For example, it is the risk of developing disease in feedlot A compared to feedlot B.

$$\text{Relative Risk} = \frac{\text{Incidence rate among exposed animal}}{\text{Incidence rate among unexposed animals}}$$

For the case of runny nose in feedlots A and B, the relative risk would be $\frac{24}{3}$, or expressed as a ratio, 8:1. This means that the risk of a steer developing runny nose in feedlot A is 8 times greater than the risk for a steer in feedlot B. If the incidence of disease in feedlot A was similar to that in feedlot B, the RR would be approximately equal to one, implying no association between disease and feedlot. Any deviation from RR=1 indicates some association and reflects the magnitude of that association.

Odds Ratio (OR)

The odds ratio is similar to relative risk, but is not as good a measurement of association. This is because

RR takes into account the rate of disease incidence while the OR only compares the odds of having the disease, with and without exposure. However, both can evaluate the strength or magnitude of the association. OR can be considered an estimate of the RR for studies where disease incidence cannot be directly calculated. The OR is calculated as:

$$\text{Odds Ratio} = \frac{\text{Odds of acquiring disease among exposed animals}}{\text{Odds of acquiring disease among unexposed animals}}$$

Attributable Risk (AR)

This indicates how much the exposure factor contributes to the disease.

$$AR = \frac{\text{Incidence rate among exposed} - \text{Incidence Rate among nonexposed}}{\text{Incidence Rate among those exposed}}$$

For the example case of steers with runny noses, AR would be (24-3)/24, or 87.5%. This means that feedlot A contributes 87.5 to the disease process. With attributable risk, it is understood that the disease can exist without the factor, but indicates how much the factor contributes to the disease incidence. AR is usually not 100%.

Strength of Association

Statistical tests are used to determine if apparent relationships are likely due to chance (or random variation) rather than to real effects. The stronger the relationship, the less likely it is due to chance. For RR and OR, a chi-square (χ^2) test can be used as a statistical test of strength of association. This test measures the likelihood that the observed findings could have occurred just because of chance. The larger the χ^2, the less likely the results are just due to chance variation. For a simple two-by-two table, if the value of χ^2 is greater than 3.84, then the p value is less than 0.05. This means that these results would be expected to occur just due to chance variation less than 5 times out of 100. If the χ^2 value is over 6.64, then the p-value is less than 0.01, representing even more certainty, since the observed relationship would occur by chance less than 1 of 100 times. Since individual animal variation can have large effects when small numbers of animals are compared, it is important to have as many animals as possible in the comparison groups to reduce the effects of chance.

Contingency or two-by-two tables

These measures of association and the strength of the association can be easily calculated from a generic two-by-two frequency table. This allows the practitioner to quickly and easily make assessments of risk and strength of association.

	Disease	No Disease
Factor A	a	b
Factor B	c	d

$$RR = \frac{a/(a+b)}{c/(c+d)} \qquad OR = \frac{a/b}{c/d} \qquad AR = \frac{a/(a+b) - c/(c+d)}{a/(a+b)}$$

$$\chi^2 = \frac{(ab-dc)^2 (a+b+c+d)}{(a+b)(c+d)(a+c)(b+d)}$$

When a is very small compared to b, and c is very small compared to d, OR is approximately equal to RR. This is the case when a disease is rare. (a and c will be very small compared to the sample size).

Designing a field trial in a feedlot practice

Introduction to Field Trials

Field trials are simply the planned **comparison** of the outcome in groups receiving different levels of a treatment. For example, the treatments levels could be vaccine versus no vaccine, and the outcome might be the subsequent rate of disease. In a field trial, a treatment could be a vaccine, a therapeutic drug, or an entire program composed of a number of individual treatments such as a preconditioning program. When designing a field trial, one attempts to ensure the comparability of the animals receiving each level of the treatment, and to reduce the experimental error so that true treatment effects can be identified. In other words, we try to remove the effect of all factors that may influence the outcome except for the treatment. Thus, the true effect of the treatment can be measured. The comparability of the treatment groups is primarily dependent on the method of allocating the animals to treatment groups, and the management of the groups during the course of the trial. When the trial is complete, statistical tests are used to evaluate the likelihood that random variation produced any observed differences between treatment groups.

Steps for Designing a Field Trial

1. Develop Clear Objectives

It is important that a well defined objective for the trial be developed. The objective should identify the outcome to be measured, and the treatment to be compared. Generally, the objective is to test a specific hy-

pothesis. For example, if you hypothesize that a new treatment for a certain disease is better than the traditional treatment, the objective of the trial will be to compare the efficacy of the new and traditional methods for treating the disease in question.

2. Identify the Reference and Experimental Populations

The reference population is the population that will benefit if the treatment being tested is effective. The experimental population is the population in which the trial is conducted. If the results of the trial are to be extrapolated from the experimental population to the reference population, then the experimental population must be representative of the reference population.

3. Define the Experimental Unit

The experimental unit is the smallest independent grouping of animals that is randomized to the treatment groups. Commonly, the experimental unit is an individual animal, but in certain situations may be a group of animals such as a pen or a herd.

4. Allocate Experimental Units to Treatment Groups

It is important to use a formal randomization procedure to allocate the experimental units to treatment groups. Randomization can be achieved by flipping a coin, drawing numbers from a hat, through use of a random numbers table, or by using a random number generator available with many computer programs.

5. Define the Treatment Regimes

The different treatments, including their timing, method, route of administration, and dosages should be clearly and completely specified.

6. The Follow-Up Period

During the follow-up period, it is important to ensure that the conscious or unconscious beliefs about the value of the different treatments do not lead to differential management or assessment of the outcome. This can often be accomplished with the use of blind techniques and placebos. If the individual responsible for assessment of the outcome is not aware of the treatment status of individual animals, then personal bias can be avoided.

7. Measurement of the Outcome

When choosing the outcome to be measured, it is important to select a parameter that can be measured objectively and quantitatively. The measured outcome should be clinically relevant and meaningful rather than just easily measurable criteria, i.e. decreased level or severity of disease is more meaningful than higher titers. It is preferable to measure the outcome without knowledge of the treatment status of the animal.

8. Analysis of the Treatment Effect

The effect of the treatment is found by comparing the outcome in the treatment and control groups. The probability that any observed differences occurred by chance can be evaluated using statistical tests.

EXERCISE: Design of a Field Trial

A large biologic/pharmaceutical company has recently contacted you regarding a new "miracle" vaccine that they claim will prevent all types of respiratory disease in feedlot cattle. The vaccine must be administered 3 to 5 weeks prior to the time the cattle enter the feedlot. The vaccine is new on the market, and due to its cost, few producers are willing to try it. The company is convinced that the benefits of the vaccine will far outweigh its costs. In order to get producers interested, the company is willing to provide some free vaccine for you to use in your practice. However, they require that the free vaccine be utilized for comparative purposes, that is to compare feedlot disease rates for vaccinated and non-vaccinated calves. Therefore, you must provide them with a detailed outline of how the vaccine will be used. You decide to submit a proposal for a field trial to test the effectiveness of the vaccine.

One of your best clients is a large cow-calf producer that calves approximately 500 commercial females each spring. This operation weans calves in October, then backgrounds the calves until the first of the year. In January, the calves are shipped to a feedlot to be custom fed. Design a simple field trial to test the effectiveness of the vaccine in these calves. Begin with a hypothesis, and be sure to discuss treatment groups, assignment to treatment groups, and follow-up.

Objectives of This Exercise:

This computer exercise is designed to introduce you to Epi Info. After you have completed the exercise, you will be able to do the following procedures in Epi Info:

1. Understand how to create a questionnaire in the Epi Info format.
2. Understand how to use a questionnaire to create an Epi Info data file.
3. Analyze data stored in the data file to answer

practical questions related to the effectiveness of a new vaccine in the prevention of respiratory diseases in feedlot cattle.

INTRODUCTION TO EPI INFO: COMPUTER PROGRAMS FOR EPIDEMIOLOGY

Overview of Epi Info, Version 5.01B

Epi Info is a collection of computer programs developed to handle data that were collected using questionnaires and to perform other common epidemiologic tasks. This software was developed by epidemiologists at the Centers for Disease Control in Atlanta.

A copy of the Epi Info software is available to you for your own use for the cost of the floppy disks ($3.00 for the program). The Epi Info manual can be ordered using the form at the end of this handout. It takes 4 to 6 weeks to get the manuals. Epi Info is not copy protected, and may be copied and given to other interested parties. The manual is easy to read and provides further instructions on the capabilities of the program that are not covered in this handout. The Epi Info program also contains a series of tutorials that will help the user to learn the program. Tutorial instructions for each function of Epi Info are given at the beginning of the corresponding chapters in the manual.

Epi Info, Version 5.01B requires an IBM-compatible microcomputer running the PC-DOS or MS-DOS operating system (version 2.0 or higher); 524,288 bytes of available memory; and two floppy-disk drives or one floppy-disk drive and a hard drive. The programs in Epi Info require approximately one megabyte of hard-disk space.

Data files can consist of as many records as DOS and your disk storage can handle (up to four billion records). A questionnaire can have up to 500 lines. The number of variables is not limited, but they must fit in the 500 lines of the questionnaire.

Epi Info is a powerful program when used correctly, but does have some limitations. It does not perform all the statistical analyses one might desire. Epi Info does interface with other statistical, spreadsheet and database packages (EAS, SAS, Lotus 1-2-3, SPSS, EpiStat, dBASE II, dBASE III, Statpac and Basic) which allows further data manipulation and more advanced analyses. Epi Info data files can be converted to the format of other software programs by using the CONVERT command (see the manual for instructions on using the convert command).

Installing Epi Info:

The programs on the Epi Info disks are in a compressed format, and they must be installed before they can be used. To install the program, place Disk 1 into drive A, type **A:Install**, and press the Enter key. You will be asked a series of questions about the Source drive (in this case it is drive A where Disk 1 is located), the Destination drive (the drive where you want the program to be installed; usually drive C or D), whether you want to install the program or copy it, and the type of video board in the computer. Once these questions have been answered, the Install program will prompt you with a beep to insert a new disk as it is needed. When this happens, insert the necessary disk and press the Enter key. For further instructions on how to install Epi Info, see chapter 4 of the manual.

Running Epi Info:

After installation, Epi Info is located on the hard drive in a subdirectory called EPI5. To access the program, turn on the computer, monitor, and printer. At the C or D:>, change into the EPI5 subdirectory by typing **CD\EPI5** and press the Enter key (be sure to use the backslash,\, not the slash,/). Now type **EPI** at the prompt, press the Enter key and wait a few seconds for the program to load into memory and become available for use.

The Main Menu for Epi Info will appear on your monitor when the program is ready to run. There are 10 options included in the Main Menu. An option can be selected in one of two ways: 1). use the up and down arrows keys to highlight the desired program and press the Enter key or 2). type the highlighted capital letter of the command, for example, E for EPED or N for ENTER and then press the Enter key. On the right side of the screen is a brief explanation of each option.

Types of Files Used in Epi Info:

File names used in Epi Info can be eight or fewer characters in length followed by an optional period and extension name of three or fewer letters. The file name should begin with a letter not with a number. Punctuation (other than the single period before the extension), spaces, and other unusual characters are not allowed in file names.

Examples: Vaccine
Feedlot2.QES
Doc1.TXT

There are four file name extensions that have special meaning in Epi Info:

1. **QES** File names ending with the extension .QES are recognized as questionnaire files by Epi Info. All questionnaire files used in Epi Info must end with the .QES extension.

Example: Vaccine.QES (A questionnaire file to collect data on calves that were or were not vaccinated with the new vaccine.)

2. **REC** The .REC extension designates a data file. Epi Info uses the format of a previously created questionnaire file to design the structure of the data file. Epi Info also stores the questionnaire in the data file and uses the questionnaire as a data entry screen.

Example: Vaccine.REC (A data file using the structure of the questionnaire file. Vaccine.QES. The data file will store the collected information)

3. **CHK** Check files have the extension .CHK. Check files consist of commands that will reduce data entry errors. Check files are optional when using Epi Info and must be created by users.

Example: Vaccine.CHK (A program set up by the user to reduce errors while entering data into Vaccine. REC)

4. **PGM** Epi Info allows users to write programs using ANALYSIS (a program within Epi Info) commands to do desired analyses on a data set. Program file names end with the .PGM extension.

Example: Vaccine.PGM (A program to do a series of analyses on data from Vaccine.REC)

Creating a Questionnaire in the Epi Info Format

Terminology Used in Questionnaire Design:

1. Variable or Field—A variable or field (variable and field have the same meaning and will be interchanged throughout this handbook) is the location in the questionnaire where data are entered. Variables or fields contain similar information. For example, if a questionnaire has a variable for animal ID number, only animal ID numbers are entered in this variable. In Epi Info, fields have names, and these names must begin with a letter of the alphabet and be 10 or fewer characters in length. Numbers can be included in variable or field names as long as they do not occur at the beginning of the name. Punctuation marks should be avoided. In naming a field, it is best to use a name that corresponds to the kind of information that will be entered in the field. For example, a field that will contain information on birth dates of calves could be named BIRTHDATE. When Epi Info uses a questionnaire to create a data file, the field names become column names in the data file.

2. Record—All of the fields in one questionnaire make up a record. Let's assume that we are collecting data related to respiratory diseases in feedlot cattle throughout the United States. We have designed the questionnaire, and one questionnaire will be filled out for each feedlot in the study. All data collected on one questionnaire make up a record for that feedlot. In the Epi Info data file, a record corresponds to a single row of data.

3. File—A file is composed of all the completed questionnaires from our survey of feedlot operations throughout the United States. A data file has one or more rows of data with each row of data corresponding to the information from one questionnaire. Thus a file is made up of columns and rows and contains all the data collected in the survey.

Basic Rules Used to Design Questionnaires in the Epi Info Format:

The first step in using the Epi Info program is to create a questionnaire in the Epi Info format. There are three rules that must be used to design a questionnaire in the Epi Info format.

1. Creating Text—The text in an Epi Info questionnaire is only for the benefit of the people using the questionnaire (interviewer, respondent, data entry person, etc.). The computer actually ignores the text in questionnaires. The text of a questionnaire defines what information is desired. Creating the text for an Epi Info questionnaire is exactly like designing a questionnaire on a typewriter or a word processor. The layout of the text is under the discretion of the designer. The text should be neat, follow a logical order, and give enough explanation so that the user will know what information the questionnaire contains. For questionnaire responses that are recorded as coded values, it may be helpful to include the codes in the questionnaire text.

2. Naming Variables or Fields—Epi Info has two methods of assigning names to fields. The program will create these names automatically using the text that precedes the response blank on the questionnaire. The first 10 non-punctuation characters on the same line as the response blank become the field name. If you use this method, it is wise to put significant words within the first 10 characters of text preceding the blank. Words such as what, who, etc. are discarded in this process. If one line of text in the questionnaire was "PRIMARY BREED OF CATTLE IN A PEN", the field name selected by the computer would be PRIMARYBRE. Letting the computer name the fields may result in names that are difficult for the user to remember.

The second method of naming fields is to enclose the desired name between curly brackets {} to the left and on the same line as the response blank. Field names must begin with a letter of the alphabet and be 10 or

fewer characters in length. Numbers can be included in field names as long as they do not occur at the beginning of the name. Punctuation marks should be avoided. In naming a field, it is best to use a name that corresponds to the kind of information that will be entered in the field. For example, a field that will contain information on the primary breed of cattle in a pen could be named {BREED}. When Epi Info uses a questionnaire to create a data file, the field names become column names in the data file.

3. Defining Variable or Field Types and Lengths—Epi Info recognizes nine field types. Field type defines what type of data can be entered in a questionnaire response blank. Every response blank must be defined as to type of field. By pressing the Control key and the same time the Q key twice (Ctrl-Q Q) while in the EPED program, a menu will appear on the EPED screen that lists the field types. A field type can be selected from this menu by highlighting the desired type and pressing the Enter key. For Text, Upper Case, and Numeric field types, the computer will also ask you for the length of the field. The length of these fields should be long enough to hold the number of characters of the longest possible response. The field type will be placed in the questionnaire at the location of the cursor. Field types can also be typed directly into a questionnaire.

a. Text or underline fields. These fields are indicated by continuous underline characters, and the length of the field is determined by the number of underline characters. Questionnaire responses for this field type can include any printable character (letters, numbers, or punctuation marks). The maximum length of a text field is 80 characters. Text fields that are left blank during data entry are considered as missing values in data analysis, and the data file will contain blank spaces for these variables.

Example: State Where the Feedlot is Located {STATE}

_____

b. Numeric fields. Only numbers will be accepted in this field type. If nothing is entered, the result will be interpreted as a missing value in data analysis and stored as a period (.) in the data file. The length of a numeric field is indicated by the number of #. If a period is placed anywhere among the number signs, the field will be a real number, and the decimal point will remain in a fixed location. Floating decimal points are defined for any numeric field with five or more digits but without a decimal point, such as #####. A numeric field can be up to 14 characters long, counting the decimal point as one character.

Example: Number of Cattle in the feedlot {CATTLENUMB} #######

c. Upper-case fields. These are similar to text fields, but all letters entered in lower case during data entry will be converted to upper case when stored in the data file. Converting all letters to upper case makes certain processes such as sorting data more efficient. The field length is indicated by the number of spaces between the less-than and greater-than symbols, including the A.

Example: Name of the Feedlot {NAME} <A   >

d. Yes/No field. Only Y, N, and blank spaces are accepted in this field type. Spaces are interpreted as missing data in data analysis.

Example: Was This Animal Vaccinated? {VACCINATED} <Y>

e. U.S. and European data fields. Dates will be checked for validity by the computer at the time of data entry. Epi Info recognizes four different date formats.

Example: Birth Date of the Animal {BIRTHDATE}
<mm/dd/yy> or
<mm/dd> or
<dd/mm/yy> or
<dd/mm>

When entering dates, type only the numbers in the date—the program provides the separating slash marks. For example, 092353 entered in a data field will be stored in the computer as 09/23/53.

f. Local phone number fields. This type will accept local phone numbers entered as xxx-xxxx where the x's are digits.

Example: Feedlot's Phone Number {PHONE}
<phonenum>

g. Long distance phone number fields. This field type will accept long distance phone numbers entered as (xxx)xxx-xxxx where the x's are digits.

Example: Diagnostic Laboratory's Phone Number {DLABPHONE} <long distance>

h. Record identification number fields. This is a special-purpose field that maintains sequential identification numbers for the records. The first record number will be number 1, and each succeeding record will have an IDNUM one higher than the previous record. Since

IDNUM fields are filled automatically by the computer, the cursor skips them during data entry. This field type keeps track of the order in which records were entered during data entry.

Example: Record Number {RECNUMB} <Idnum>

i. Today's-date or date-of-last-change field. The computer automatically enters the system date of the computer into this field type on the date when the record was last saved. If the record is edited and saved after it is first entered into the data file, the date in this field type is updated to the last date that the record was saved.

Example: Last Date Record Was Saved {SAVEDDATE}
<today> or
<today/yy> or
<today/yyyy>

Each of the examples in this section could be a line in a questionnaire. Notice that each example has text that defines the information desired, a field name enclosed in curly brackets, and a field type that defines what type of data can be entered in the response blank. The field name must be on the same line and to the left of the field type.

When creating a questionnaire in Epi Info, the underline, number sign, less-than or greater-than characters should never be used unless a field type is being defined.

Creating a Questionnaire

The following steps were used to create the questionnaire used in this exercise:

1. Highlight EPED from the Main Menu and press the Enter key.

2. Press F2 key and a submenu will appear. Select "Open file this window" and press the Enter key. You will be asked for a name for the questionnaire file. Type **A:VACCINE.QES.** "A:" tells the computer to store this file on the floppy disk in the A drive. "VACCINE" is the name of the file. Epi Info file names can be up to 8 characters long, followed by a period and extension name of three or fewer letters. ".QES" designates this file as a questionnaire file. All questionnaire files must have the .QES extension. Press the enter key.

3. Press the F6 key to bring up the Setup menu and check to see that Word Wrap is OFF. "WW, Justify, etc" is highlighted, so press the Enter key to bring up the next

menu. If Word Wrap is ON, press the Enter key and this will turn Word Wrap OFF. When Word Wrap is ON, the computer will automatically go to the next line once you have typed to the right margin. When Word Wrap is OFF, you have to manually go to the next line of typing by pressing the Enter key. If a questionnaire is created in EPED while Word Wrap is ON, problems can occur when you use the questionnaire to generate a data file. Press the Esc key to return to the Setup menu, and highlight "Save setup" and press the Enter key. Press the Esc key twice. In general, use the Esc key to go to a previous menu. The F10 key will cause the program to return to the Main Menu or exit the program.

4. Type the questionnaire text, create field names, and designate the type and length of each field. Creating the questionnaire is similar to typing the questionnaire on a word processor. Most of the commands in EPED are similar to those in Wordstar (Wordstar is a word processing program).

The following information was used to create the questionnaire:

Information desired	Field name	Field type	Field length
Animal ID number	ANIMALID	Numeric	4
Birth date of animal	BIRTHDATE	Date	
Sex of Animal	SEX	Text	1
Breed of animal	BREED	Text	1
Was the animal vaccinated	VACCINE	Yes/No	
Date of vaccination	VACDATE	Date	
Date animal entered feedlot	IDATE	Date	
Initial weight	IWEIGHT	Numeric	5
Date at the end of feeding period	FDATE	Date	
Weight at end of feeding period	FWEIGHT	Numeric	6
Average daily gain	ADG	Numeric	3
Did animal have resp. disease	RESPDIS	Yes/No	
Date of respiratory disease	DATERESPIR	Date	

5. When a questionnaire is finished, press the F9 key to save the questionnaire. Press the F10 key to return to the Main Menu.

Retrieving a previously created questionnaire.

1. Highlight the EPED option on the Main Menu and press Enter key. Place the data disk into the A drive.

2. Press the F2 key and a submenu will appear. Select "Open file this window" and press the Enter key. You

will be asked for the name for the questionnaire file. Type **A:VACCINE.QES.** Press the Enter key, and the questionnaire will be presented on the screen. This questionnaire was previously created in the EPED program to collect data related to the occurrence of respiratory disease in feedlot cattle that had or had not been vaccinated with a new vaccine.

3. The questionnaire should look like the following:

QUESTIONNAIRE TO EVALUATE A NEW VACCINE TO PREVENT RESPIRATORY DISEASES IN FEEDLOT CATTLE

ANIMAL IDENTIFICATION NUMBER:
   {ANIMALID} ####
BIRTH DATE OF THE ANIMAL:
   {BIRTHDATE} <MM/DD/YY>
SEX OF THE ANIMAL:
   {SEX} _ M = STEER F = HEIFER
BREED OF THE ANIMAL:
   {BREED} _ A = ANGUS H = HEREFORD X
   = CROSSBRED
WAS THE ANIMAL GIVEN THE VACCINE:
   {VACCINE} <Y>
DATE OF VACCINATION:
   {VACDATE} <mm/dd/yy>
DATE THE ANIMAL ENTERED THE FEEDLOT:
   {IDATE} <mm/dd/yy>
WEIGHT OF THE ANIMAL WHEN IT ENTERED THE FEEDLOT:
   {WEIGHT} #####
DATE AFTER A 120-DAY FEEDING PERIOD:
   {FDATE} <mm/dd/yy>
WEIGHT OF THE ANIMAL AFTER A 120-DAY FEEDING PERIOD:
   {FWEIGHT} ######
AVERAGE DAILY GAIN DURING THE 120-DAY FEEDING PERIOD:
   {ADG} #.##
DID THE ANIMAL HAVE A RESPIRATORY DISEASE DURING THE 120-DAY FEEDING PERIOD:
   {RESPDIS} <Y>
DATE OF RESPIRATORY DISEASE:
   {DATERESPIR} <mm/dd/yy>

D. Print a copy of the questionnaire.

Steps to print the questionnaire:

1. Be sure that your printer is turned on and that is on line.

2. With the questionnaire on the screen in EPED, press the F5 key.

3. The print menu will appear on the screen. Highlight the option "Print file now" and press the Enter key.

4. Once the questionnaire is printed, press the F10 key to return to the Main Menu.

Creating a Data File and Entering Data

Creating a data file.

Once a questionnaire has been created in the Epi Info format, the next step in using the program is to create the structure of the data file. Epi Info uses the format of the questionnaire to create the appropriate data file structure. Field names in the questionnaire become column names in the data file. Field types and lengths in the questionnaire will define the types of data and number of characters that will be allowed in the columns of the data file.

The following steps were used to generate the data file structure used in this exercise:

1. Highlight the ENTER option on the Main Menu, and press the Enter key.

2. Name the data file the same as the questionnaire file by typing A:VACCINE when the computer asks for "Data file is:". The program will automatically give the data file the extension ".REC" indicating that this is a data file. Press the Enter key. Then press the Esc key.

3. Type in the name of the questionnaire file **A:VACCINE.** The computer automatically adds the .QES extension. Press the Enter key.

4. Epi Info automatically creates a data file corresponding to the format of the questionnaire. When the ENTER program creates a data file, it stores the entire questionnaire in the data file so that subsequent "visits" to the data file do not require the original questionnaire. The questionnaire now appears on the ENTER screen as a data entry form.
5. The data file has now been created and is ready to receive data. At this point the data file contains no data.

6. The data entry screen was used to enter records on 100 calves. Half of the calves were vaccinated with the new vaccine, the other calves were not vaccinated with the new vaccine.

Data Analysis Using Epi Info

The ANALYSIS program within Epi Info allows the user to perform many operations on the data stored in a data file in order to answer questions from the data. ANALYSIS produces lists of data, frequency distribution tables, cross tabulations, graphs, etc. Appropriate epidemiologic statistics, such as odds ratios, relative risk, confidence limits, the Fisher exact test, and chi square tests accompany the results from the Tables

command. Stratified analyses with the Mantel-Haenszel technique, one-way analysis of variance (ANOVA), Kruskal-Wallis tests, linear regression, and analysis of matched case-control studies are also offered. Records can be sorted using one or more variables. Certain records can be selected for analysis using the Select command, "if, then" statements, or by using mathematical and logical operations. New variables can also be defined in the ANALYSIS program. During data analysis, it is often desired to perform calculations on variables already in the data file, to group values of a variable into categories, or to change the existing values of a variable. These new values are stored temporarily in new variables. These new variables can be made a permanent part of a data file by using the Route and Write commands of ANALYSIS. The results from ANALYSIS commands can be routed to the screen, a printer, or a disk file.

In the ANALYSIS exercises, we will use data stored in the data file named VACCINE.REC. VACCINE.REC contains information related to the incidence of respiratory diseases and performance of 100 calves in a feedlot. These calves came from the same ranch, and half of the calves were vaccinated with a new vaccine approximately four weeks prior to entering the feedlot. The other half of the calves were not vaccinated. Calves in each treatment group were selected using a random numbers table. Upon entering the feedlot, all calves were weighed, processed using the feedlot's normal procedures, and placed in a single pen. Feedlot personnel did not know which calves were in each treatment group. All calves were fed for a 120-day period and were observed daily for the occurrence of respiratory diseases. At the end of the feeding period, calves were weighed and average daily gains were calculated.

In this exercise, Epi Info will be used to evaluate the effectiveness of the new vaccine in preventing respiratory diseases and to determine if the vaccine had an effect on the performance of the calves in the feedlot.

Steps to use the ANALYSIS program:

1. Highlight ANALYSIS on the Main Menu and press the Enter key.

2. The ANALYSIS screen is divided into two parts: a larger upper window where the results of ANALYSIS commands are displayed and a lower window where the ANALYSIS commands are entered at the EPI>. Results that have disappeared from the screen can be displayed by pressing the PgUp or the PgDn key. To scroll through the results one line at a time, press the Ctrl key at the same time as the PgUp or PgDn key.

3. ANALYSIS commands can either be typed at the EPI> or selected using the F2 and F3 keys.

Special Keys in the ANALYSIS program:

F1 or the Help Key—Pressing the F1 key will display a list of ANALYSIS commands and subject areas related to the ANALYSIS program. Use the up and down arrow keys to highlight one of the options and press the Enter key. An explanation of what the command does, the proper syntax or structure of the command, and examples of the command will be displayed on the screen. Press the Esc key to get out of the help menu.

F2 or the Command Key—Pressing the F2 key will list all of the commands that are available in the ANALYSIS program. If you do not like to type, you can use the F2 key to select a command name. To select a command name, press the F2 key, highlight the command that you want to use, and press the Enter key. The command name will be placed at the EPI>. All commands in ANALYSIS begin with a command name followed by one or more field names. Only one ANALYSIS command can be executed at a time. If you place two command names in the same command line, the computer will give you an error message.

F3 or the Variable Key—After a data file has been read into the active memory of the computer using the Read command, a list of field names from the data file can be obtained by using the F3 key. As with the Command key, variable names can be selected and placed on the command line by highlighting the variable name and pressing the Enter key. In commands that include two or more variables, you can tag the variable names by highlighting the first variable and pressing the Plus (+) key instead of the Enter key. Highlight the next variable name and tag it with a +. When the last variable name has been tagged, press the Enter key, and all the tagged variable names will appear on the command line after the command name.

F4 or the Browse Key—The F4 key will display the data in an open data file in a column and row format. The user can scroll through the data using the arrow keys. The data in a data file cannot be edited while using the Browse command. When the user is in Browse, pressing the F4 key will display the data in a questionnaire format. Pressing the F4 key again will return the data to a column-row format. If the F5 key is pressed while browsing, the data will be displayed in a column-row format on the top of the screen and in a questionnaire format on the bottom of the screen.

F5 or the Printer Control Key—The F5 key routes all output from the ANALYSIS commands to the printer.

Output will continue to be sent to the printer until the F5 key is pressed a second time. This routes the output to the screen only.

F9 or the DOS Key—If the F9 key is pressed from ANALYSIS, the computer gets out of the Epi Info program and goes to DOS. At the DOS prompt, the user can execute any DOS command. To return to ANALYSIS type EXIT.

Arrow keys—After commands are executed in ANALYSIS, they are stored in temporary memory in a Command stack for possible reuse. When the cursor is at the EPI>, pressing the Up Arrow key will bring back the previously executed command, pressing the Up Arrow key again will bring up the command before that, and so on. The last 20 commands are stored by ANALYSIS. Once a command has been recalled, it can be executed again by pressing the Enter key, or it can be edited and then executed. The Down Arrow key scrolls back through the commands to return to the EPI>.

ANALYSIS Commands

1. Read—the Read command must be executed before any analysis can be performed on a data file. The Read command places the data in a data file into the active memory of the computer. If the user wants to work with data from another data file, another Read command must be executed, and the data from this data file replaces the data from the first file. More than one data file can be read into the active memory of the computer at the same time, but these files must all have exactly the same format. This means that these data files were all developed using the same questionnaire. The name and the number of records in the active data file are displayed in the upper left hand corner of the ANALYSIS screen.

Execute each of the following example ANALYSIS commands:

Example: EPI>**READ A:VACCINE.REC**

2. Variables—the Variables command displays a list of all the field names in an open data file as well as the type and length of each field. A list of the field names can also be obtained by pressing the F3 key. Use the F3 key to select field names to be placed on the command line when developing ANALYSIS commands.

Example: EPI>**VARIABLES**
Example: EPI>**F3** (notice the differences between these two commands)

3. Browse or the F4 key—the Browse command displays the current data file in a spread sheet format. You can scroll through the data by using the arrow keys. Browse will not allow you to make changes in the data.

Example: EPI>**BROWSE** or **F4**

While in Browse, experiment with the F4 and F5 keys. Although it is confusing at first, the Function keys (the F keys) have different functions in each Epi Info program. The functions of these keys are displayed either at the top or the bottom of the screen.

4. Route—the Route command specifies where the user wants the output of the ANALYSIS commands to go. Output will always be displayed in the upper window of the ANALYSIS screen. Output can also be directed to the printer to produce a hard copy or to a file on a floppy disk or the hard drive for storage. To cancel a previous Route to a file or printer, use ROUTE SCREEN.

Example: EPI>**ROUTE PRINTER or F5**

5. List—the List command produces a line listing of the data in an open data set. If you want to see data that are already stored in the data file, use the List command. If no field names are specified in the List command, the computer will list as many fields as will fit across the monitor screen. If field names are given, List will only display these variables.

Example: EPI>**LIST** (will list as many variables as will fit across one screen)
EPI>**LIST ANIMALID BIRTHDATE SEX** (will only list these three variables)
EPI>**ROUTE SCREEN** (this cancels the Route Printer command)
EPI>**LIST * NOT VACDATE IDATE FDATE** (will list all variables in the data file except VACDATE, IDATE, and FDATE)

6. Freq—The Freq command produces a frequency distribution of the variable or variables designated in the command. If two or more variables are included in a Freq command, the program produces separate frequency distribution for each variable. Data from the variables are not combined. This command tells you the number of observations of each possible value of a variable in the data file.

From now on you can decide if you want the output to be sent to the printer or only to appear on the screen. It may be helpful to have a printed copy of the exercises.

Example: EPI>**FREQ BREED**
EPI>**FREQ SEX RESPDIS**
EPI>**FREQ *** (does a frequency distribution on all variables)
EPI>**FREQ * NOT BIRTHDATE VACDATE** (does a frequency distribution of all variables except BIRTHDATE and VACDATE)

7. Charts and Graphs—the ANALYSIS program can create histograms, scatter plots, pie charts, bar graphs and line graphs using data from a data file. Histograms, pie charts, and bar and line graphs use one variable per command. Scatter plots require two variables, and they are most useful when both variables contain continuous numeric data. The least squares linear regression line can be included in a scatter plot by adding "/R" at the end of the command. Bar graphs differ from histograms in that their bars are separated by spaces and they do not include values whose counts are 0.

Example: EPI>**HISTOGRAM BREED** (press the F10or Esc key to return to the EPI>)
EPI>**PIE BREED**

8. Title—the Title command will add titles to frequency distributions, cross tabulation tables, graphs, etc. Up to five titles can be used. The Title command must precede the command for which you want the title to appear. A Title 1 command without any text will cancel previous Title commands.

Example: EPI>**TITLE 1 DISTRIBUTION OF BREEDS OF CALVES**
EPI>**PIE BREED**
EPI>**TITLE 1** (this will delete Title1)

9. Sort—the Sort command allows you to specify the order that records will be written when using the List and Write commands. The data in a data file can be sorted by one or more variable names. If more than one variable is included in a Sort command, the records will first be sorted in ascending order by the first variable, then within a common value of the first variable the records will be sorted by the second variable, and so forth for each variable listed in the command. A Sort command without any field name will cancel the previous Sort command.

Example: EPI>**SORT SEX BIRTHDATE**
EPI>**LIST ANIMALID SEX BIRTHDATE** (notice how the data are sorted)
EPI>**SORT** (this cancels the previous Sort command)

10. Select—the Select command allows you to perform ANALYSIS operations on a subset of the original data file. Select enables you to define conditions that must be true before a record is used in subsequent ANALYSIS commands. Several Select commands can be executed in succession, and their effects are additive. Select used without an expression will cancel all previous Select commands. The active selection criteria is shown in the top left hand corner of the ANALYSIS screen.

Example: EPI>**SELECT BREED = "A"** (specific values of Text, Upper Case, Yes/No, and Date fields must be enclosed in double quotes when used in ANALYSIS commands)
EPI>**LIST** (notice what data are listed, look at selection criteria at the top of the ANALYSIS screen)
EPI>**SELECT** (this cancels the previous Select command)

11. Tables—the Tables command does a cross tabulation of two categorical variables and tests the association between these variables by using appropriate statistical analyses. At least two variable names must be included in this command. Values of the first variable name will appear on the left margin of the cross tabulation table, and those of the second variable will be across the top of the table. All variable names after the first two listed in a Tables command are used for stratification.

Example: EPI>**TABLES SEX RESPDIS**

12. Means—the Means command followed by one variable name that contains continuous data will calculate the mean value of that variable. A Means command followed by a continuous variable name (the continuous variable name must come first) and a categorical variable name (the second variable divides the first variable into groups or samples) will produce a frequency distribution table and perform appropriate statistical tests to determine the association between the two variables. The frequency distribution table can be eliminated from the output by typing /N after the second variable name. Depending on the size of the table and whether the data are normally distributed, the Means command does the following analyses:

For normally distributed data
 1. ANOVA (for two or more samples)
 2. Student's T test (for two samples)

For data that is not normally distributed
 1. Kruskal-Wallis one-way ANOVA (for two or more samples)

2. Mann-Whitney U test=Wilcoxon Rank Sum Test (for two samples)

You are not expected to understand all the statistical tests performed by Epi Info. If you want to know more about these analyses, consult your statistical references.

Example: EPI>**MEANS ADG**
            EPI>**MEANS ADG SEX**
            EPI>**MEANS ADG SEX/N** (/N will eliminate the frequency distribution table)

EVALUATION OF THE FIELD TRIAL

1. Graphs present data in a visual form and can give valuable information in evaluating a field trial. Create an epidemic curve (a histogram of the date of onset of the disease) for calves in each treatment group that had a respiratory disease.

EPI>**SELECT RESPDIS = "Y"**
EPI>**SELECT VACCINE = "Y"**
EPI>**TITLE 1 EPIDEMIC CURVE OF VACCI-NATED CALVES**
EPI>**HISTOGRAM DATERESPIR**
EPI>**SELECT**
EPI>**SELECT RESPDIS = "Y"**
EPI>**SELECT VACCINE = "N"**
EPI>**TITLE 1 EPIDEMIC CURVE OF NONVACCINATED CALVES**
EPI>**HISTOGRAM DATERESPIR**
EPI>**SELECT**
EPI>**TITLE 1**

What conclusions can be made from these graphs?

2. Calculate the incidence rate of respiratory diseases during the 120-day feeding period for all calves and for each treatment group. Assume that all cases were new cases during the feeding period and that there was no mortality.

EPI> **FREQ RESPDIS**
EPI> **SELECT VACCINE = "Y"**
EPI> **FREQ RESPDIS**
EPI> **SELECT**
EPI> **SELECT VACCINE = "N"**
EPI> **FREQ RESPDIS**
EPI> **SELECT**

Does there appear to be a difference in the incidence rates of respiratory diseases between the vaccinated and unvaccinated calves?

3. Test the association between vaccination treatment and the occurrence of respiratory diseases during the 120-day feeding period.

EPI> **TABLES VACCINE RESPDIS**

Was there a significant association between vaccination treatment and the occurrence of respiratory diseases? Interpret the odds ratio and the relative risk. Was the vaccine effective in reducing the incidence of respiratory diseases.

4. Were other factors associated with the occurrence of respiratory diseases? Test the associations of sex and breed on the occurrence of respiratory diseases.

EPI> **TABLES SEX RESPDIS**
EPI> **TABLES BREED RESPDIS**

5. What was the average daily gain of all the calves during the 120-day feeding period?

EPI> **MEANS ADG**

6. Did receiving the new vaccination have any effect on the performance of the calves during the 120-day feeding period?

EPI> **MEANS ADG VACCINE/N**

7. Test the effects of sex, breed, and having a respiratory disease on the performance of the calves during the feeding period.

EPI> **MEANS ADG SEX/N**
EPI> **MEANS ADG BREED/N**
EPI> **MEANS ADG RESPDIS/N**

8. Make a simple report of the animal numbers, sexes, and average daily gains for calves that were not vaccinated and had a respiratory disease. Sort the data first by animal sex and within sex, sort by average daily gain.

EPI> **SELECT (VACCINE = "N") AND (RESPDIS = "Y")**
EPI> **SORT SEX ADG**
EPI> **LIST ANIMALID SEX ADG**
EPI> **SELECT**

1. Graphs present data in a visual form and can give valuable information in evaluating a field trial. Create an epidemic curve (a histogram of the date of onset of the disease) for calves in each treatment group that had a respiratory disease.

EPIDEMIC CURVE OF VACCINATED CALVES
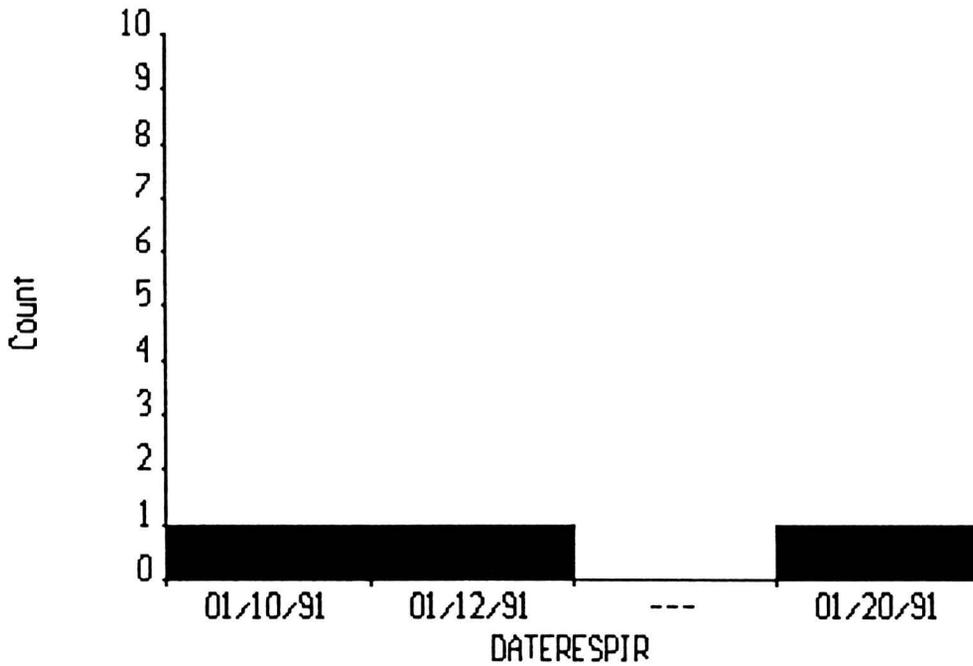
EPIDEMIC CURVE OF NONVACCINATED CALVES
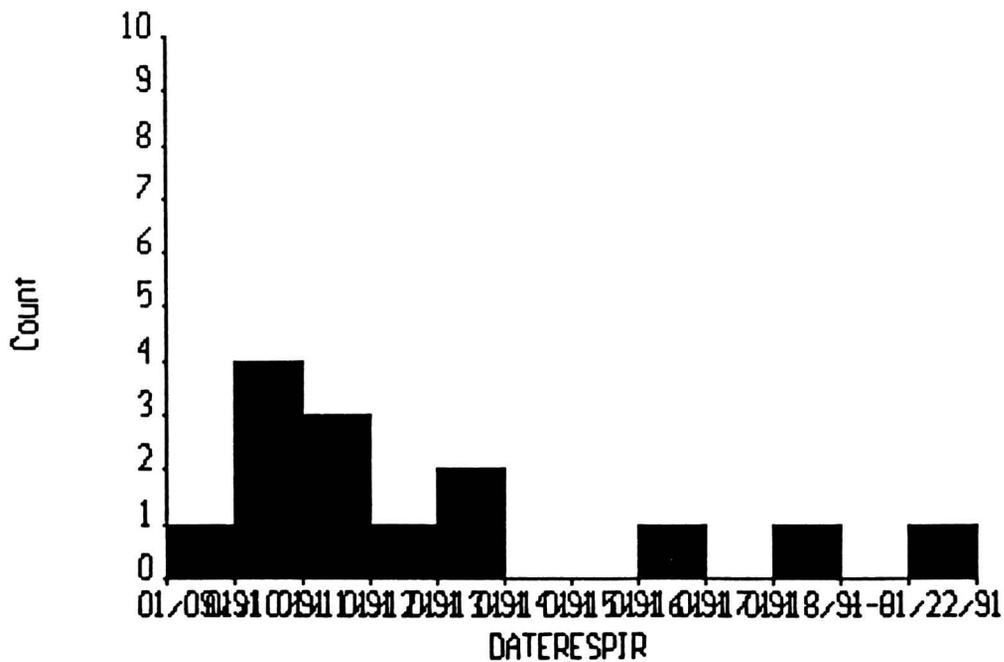
2. Calculate the incidence rate of respiratory diseases during the 120-day feeding period for all calves and for each treatment groups. Assume that all cases were new cases during the feeding period and that there was no mortality.

```
RESPDIS |  Freq  Percent   Cum.
--------+-----------------------
+       |    17   17.0%    17.0%      Overall incidence rate = 17%.
-       |    83   83.0%   100.0%
--------+-----------------------
  Total |   100  100.0%
```

Current selection: vaccine = "y"

```
RESPDIS |  Freq  Percent   Cum.
--------+-----------------------
+       |     3    6.0%     6.0%      Incidence rate in vaccinated calves = 6%.
-       |    47   94.0%   100.0%
--------+-----------------------
  Total |    50  100.0%
```

Current selection: vaccine = "n"

```
RESPDIS |  Freq  Percent   Cum.
--------+-----------------------
+       |    14   28.0%    28.0%      Incidence rate in unvaccinated calves = 28%.
-       |    36   72.0%   100.0%
--------+-----------------------
  Total |    50  100.0%
```

The incidence rate of respiratory disease is numerically higher in the unvaccinated calves, 28 vs 6%. However, to determine if the rates are significantly different, appropriate statistical analyses must be performed.

3. Test the association between vaccination treatment and the occurrence of respiratory diseases during the 120-day feeding period.

```
                    RESPDIS
VACCINE    |     +      -  | Total
-----------+--------------+------
        + |     3     47 |    50
        - |    14     36 |    50
-----------+--------------+------
    Total |    17     83 |   100
```

                     Single Table Analysis
Odds ratio   0.16
Cornfield 95% confidence limits for OR   0.03 < OR < 0.69

Relative risk of (RESPDIS=+) for (VACCINE=+)   0.21
Greenland, Robins 95% conf. limits for RR   0.07 < RR < 0.70
            (Biometrics 1985;41:55-68)
        Ignore relative risk if case control study.

```
                 Chi-Squares   P-values
                 -----------   --------
    Uncorrected:      8.58     0.00340720 <---
    Mantel-Haenszel:  8.49     0.00357158 <---
    Yates corrected:  7.09     0.00776377 <---
```

There was a significant association between vaccination treatment and the occurrence of respiratory diseases as indicated by the small P-value of the Chi-squares analyses. The odds ratio tells us that the odds of vaccinated calves having a respiratory disease is 0.16 times that of unvaccinated calves. The relative risk tells us that the risk of vaccinated calves having a respiratory disease is 0.21 time that of unvaccinated calves. Based on these results, the vaccine was effective in reducing the incidence rate of respiratory diseases in feedlot cattle.

4. Were other factors associated with the occurrence of respiratory diseases? Test the associations of sex and breed on the occurrence of respiratory diseases.

```
                             RESPDIS
SEX          |     +      -  | Total
-----------+-------------+------
         F |     9     45  |    54
         M |     8     38  |    46
-----------+-------------+------
     Total |    17     83  |   100
```

Single Table Analysis

Odds ratio   0.95
Cornfield 95% confidence limits for OR   0.29 < OR < 3.07

Relative risk of (RESPDIS=+) for (SEX=F)   0.96
Greenland, Robins 95% conf. limits for RR   0.40 < RR < 2.28
                (Biometrics 1985;41:55-68)
         Ignore relative risk if case control study.

	Chi-Squares	P-values
Uncorrected:	0.01	0.92340427
Mantel-Haenszel:	0.01	0.92378704
Yates corrected:	0.03	0.86428116

```
                            RESPDIS
BREED        |     +      -  | Total
-----------+-------------+------
         A |     7     24  |    31
         H |     2     28  |    30
         X |     8     31  |    39
-----------+-------------+------
     Total |    17     83  |   100
```

Chi square =   3.30
Degrees of freedom =   2
p value = 0.19247631

Neither sex nor breed of the animal had a significant association with the occurrence of respiratory diseases.

5. What was the average daily gain of all the calves during the 120-day feeding period?

ADG	Freq	Percent	Cum.
2.67	1	1.0%	1.0%
2.70	1	1.0%	2.0%
2.76	1	1.0%	3.0%
2.83	1	1.0%	4.0%
2.85	2	2.0%	6.0%
2.89	1	1.0%	7.0%
2.92	3	3.0%	10.0%
2.97	1	1.0%	11.0%
2.98	1	1.0%	12.0%
2.99	1	1.0%	13.0%
3.00	2	2.0%	15.0%
3.04	2	2.0%	17.0%
3.05	1	1.0%	18.0%
3.08	2	2.0%	20.0%
3.12	1	1.0%	21.0%
3.13	2	2.0%	23.0%
3.15	2	2.0%	25.0%
3.17	1	1.0%	26.0%
3.18	1	1.0%	27.0%
3.19	1	1.0%	28.0%
3.21	2	2.0%	30.0%
3.22	1	1.0%	31.0%
3.23	4	4.0%	35.0%
3.25	1	1.0%	36.0%
3.26	1	1.0%	37.0%
3.27	2	2.0%	39.0%

```
ADG  |  Freq  Percent   Cum.    - continued
-----+----------------------
3.28 |    4    4.0%     43.0%
3.29 |    2    2.0%     45.0%
3.31 |    1    1.0%     46.0%
3.32 |    2    2.0%     48.0%
3.33 |    4    4.0%     52.0%
3.34 |    2    2.0%     54.0%
3.35 |    2    2.0%     56.0%
3.37 |    1    1.0%     57.0%
3.40 |    1    1.0%     58.0%
3.41 |    1    1.0%     59.0%
3.42 |    2    2.0%     61.0%
3.43 |    2    2.0%     63.0%
3.44 |    1    1.0%     64.0%
3.47 |    2    2.0%     66.0%
3.50 |    3    3.0%     69.0%
3.52 |    2    2.0%     71.0%
3.53 |    2    2.0%     73.0%
3.54 |    1    1.0%     74.0%
3.57 |    2    2.0%     76.0%
3.58 |    2    2.0%     78.0%
3.59 |    1    1.0%     79.0%
3.60 |    1    1.0%     80.0%
3.63 |    3    3.0%     83.0%
3.66 |    2    2.0%     85.0%
3.67 |    1    1.0%     86.0%
3.68 |    2    2.0%     88.0%
3.69 |    1    1.0%     89.0%
3.71 |    1    1.0%     90.0%
3.72 |    1    1.0%     91.0%
3.75 |    1    1.0%     92.0%
3.76 |    3    3.0%     95.0%
3.81 |    1    1.0%     96.0%
3.84 |    2    2.0%     98.0%
3.87 |    1    1.0%     99.0%
3.92 |    1    1.0%    100.0%
-----+----------------------
Total |  100  100.0%

Sum                  =    334.28
Mean                 =      3.34
Standard deviation   =      0.28
```

6. Did receiving the new vaccine have any effect on the performance of the calves during the 120-day feeding period?

MEANS of ADG for each category of VACCINE

VACCINE	Obs	Total	Mean	Variance	Std Dev
+	50	170	3.404	0.066	0.258
-	50	164	3.282	0.090	0.300
Difference			0.122		

VACCINE	Minimum	25%ile	Median	75%ile	Maximum	Mode
+	2.850	3.210	3.420	3.580	3.920	3.040
-	2.670	3.080	3.285	3.500	3.810	2.920

ANOVA
(For normally distributed data only)
The p value is equivalent to that for the Student's T Test,
since there are only 2 samples.

Variation	SS	df	MS	F statistic	p-value
Between	0.370	1	0.370	4.724	0.030183
Within	7.668	98	0.078		
Total	8.038	99			

Bartlett's test for homogeneity of variance
Bartlett's chi square = 1.114   deg freedom = 1   p-value = 0.291287

The variances are homogeneous with 95% confidence.
If samples are also normally distributed, ANOVA results can be used.

Mann-Whitney or Wilcoxon Two-Sample Test (Kruskal-Wallis test for two groups)

```
Kruskal-Wallis H (equivalent to Chi square) =       3.377
                          Degrees of freedom =           1
                                     p value =    0.066125
```

The average daily gain of the unvaccinated calves was significantly lower than that of the vaccinated calves.

7. Test the effects of sex, breed, and having a respiratory disease on the performance of the calves during the feeding period.

MEANS of ADG for each category of SEX

SEX	Obs	Total	Mean	Variance	Std Dev
F	54	174	3.226	0.054	0.232
M	46	160	3.479	0.080	0.283
Difference			-0.253		

SEX	Minimum	25%ile	Median	75%ile	Maximum	Mode
F	2.670	3.080	3.230	3.330	3.760	3.230
M	2.830	3.340	3.525	3.680	3.920	3.500

ANOVA
(For normally distributed data only)
The p value is equivalent to that for the Student's T Test, since there are only 2 samples.

Variation	SS	df	MS	F statistic	p-value
Between	1.588	1	1.588	24.135	0.000040
Within	6.449	98	0.066		
Total	8.038	99			

Bartlett's test for homogeneity of variance
Bartlett's chi square =   1.868 deg freedom = 1   p-value = 0.171668

The variances are homogeneous with 95% confidence.
If samples are also normally distributed, ANOVA results can be used.

Mann-Whitney or Wilcoxon Two-Sample Test (Kruskal-Wallis test for two groups)

```
Kruskal-Wallis H (equivalent to Chi square) =      22.980
                          Degrees of freedom =           1
                                     p value =    0.000002
```


MEANS of ADG for each category of BREED

BREED	Obs	Total	Mean	Variance	Std Dev
A	31	104	3.365	0.122	0.350
H	30	99	3.315	0.044	0.210
X	39	130	3.346	0.080	0.283

BREED	Minimum	25%ile	Median	75%ile	Maximum	Mode
A	2.670	3.080	3.440	3.630	3.870	2.850
H	2.920	3.180	3.315	3.430	3.720	3.130
X	2.830	3.150	3.330	3.530	3.920	3.500

ANOVA
(For normally distributed data only)

Variation	SS	df	MS	F statistic	p-value
Between	0.039	2	0.019	0.236	0.792896
Within	7.999	97	0.082		
Total	8.038	99			

Bartlett's test for homogeneity of variance
Bartlett's chi square =   6.975 deg freedom = 2   p-value = 0.030580

Bartlett's Test shows the variances in the samples to differ.
Use non-parametric results below rather than ANOVA.

Kruskal-Wallis One Way Analysis of Variance

```
Kruskal-Wallis H (equivalent to Chi square) =        1.216
                      Degrees of freedom =            2
                                 p value =       0.544557
```

MEANS of ADG for each category of RESPDIS

RESPDIS	Obs	Total	Mean	Variance	Std Dev
+	17	50	2.958	0.033	0.181
-	83	284	3.422	0.055	0.234
Difference			-0.464		

RESPDIS	Minimum	25%ile	Median	75%ile	Maximum	Mode
+	2.670	2.850	2.920	3.080	3.340	2.850
-	2.920	3.260	3.410	3.600	3.920	3.280

ANOVA
(For normally distributed data only)
The p value is equivalent to that for the Student's T Test,
since there are only 2 samples.

Variation	SS	df	MS	F statistic	p-value
Between	3.038	1	3.038	59.560	0.000000
Within	4.999	98	0.051		
Total	8.038	99			

Bartlett's test for homogeneity of variance
Bartlett's chi square = 1.532 deg freedom = 1   p-value = 0.215820

The variances are homogeneous with 95% confidence.
If samples are also normally distributed, ANOVA results can be used.

Mann-Whitney or Wilcoxon Two-Sample Test (Kruskal-Wallis test for two groups)

```
Kruskal-Wallis H (equivalent to Chi square) =       32.381
                      Degrees of freedom =            1
                                 p value =       0.000000
```

The average daily gain of heifers was significantly lower than that of steers. Calves that had a respiratory disease gained significantly slower that those that did not have a respiratory disease. Breed had no effect on average daily gain.

8. Make a simple report of the animal numbers, sexes, and average daily gains for calves that were not vaccinated and had a respiratory disease. Sort the data first by animal sex and within sex, sort by average daily gain.

Current selection: (vaccine = "n") and (respdis = "y")

REC	ANIMALID	SEX	ADG
15	9031	F	2.67
12	9027	F	2.70
47	9095	F	2.76
20	9043	F	2.92
50	9036	F	3.00
31	9063	F	3.05
24	9050	F	3.08
3	9004	M	2.83
5	9009	M	2.85
17	9037	M	2.89
19	9040	M	2.92
27	9055	M	3.08
41	9084	M	3.23
26	9052	M	3.34