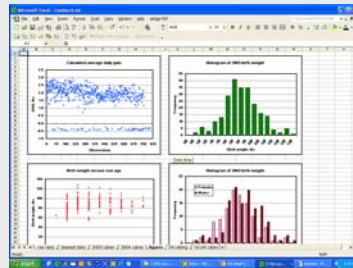
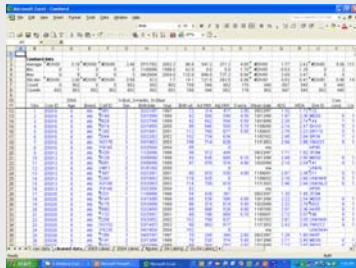




## Land O' Lakes Beef Cattle Short Course

### Data Collection/Management and SPC Applications



Kevin Dhuyvetter  
Agricultural Economist  
Kansas State University

July 19-20, 2005 – Weber Hall Room 111, Manhattan, KS



### Forces of change in the beef industry...

- Increasing consumer focus
- Industry consolidation/internal competition
- Globalization of markets
- Accelerated industrialization
- New technologies/information systems
- Competition with poultry/pork industries
- Contractual arrangements and alliances
- Increased environmental concerns



## **Increasing consumer focus...**

- Consumers demand safe, nutritious beef products that are easy to prepare in less time
- Pork and poultry industries have already responded via...
  - Integrated production
  - Superior genetics
  - Branded and case-ready product



## **Retail trends drive producer trends...**

- Increased accountability (traceability)
- Increased use of technology (bio and info)
- Decreased room for outliers
- Increased understanding of different segments
- Understanding optimal operational practices
- Beef quality assurance enhancements



## Success in light of these trends...

... will hinge upon the *rapid* exchange of *accurate* information at all levels of food animal production and marketing.

Farm — — — — —▶ — — — — —▶ — — — — —▶ — — — — —▶ Consumer



## Management strategies

### Traditional

- Batch production
- Group average info
- Average knowledge
- Manual operations
- General tactics
- Intuitive decision making

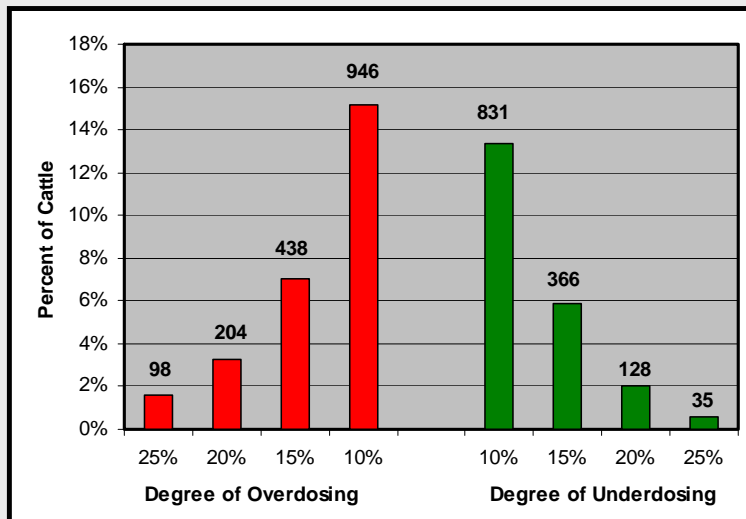
### Emerging

- Individualized production
- Individual information
- Specific knowledge
- Automatic operations
- Detailed tactics
- Informed decision making



## Example -- average weight dosing

[6,231 head (24 lots) Avg in-weight = 694 lb]



Which is worse – overdosing or underdosing?



## Why should I collect data?

- Information has value to make better decisions
  - forecasts are often based on historical data
- Potential for future use
  - cannot be collected later
- Required/regulatory
  - e.g., contractor, government, lender



## Data collection

- A process involving...
  - People
  - Materials
  - Methods
  - Machines
  - Measurements
  - Environments



## Quality of the data you collect

- What are the metrics for quality data?

***Accurate***

***Timely***



## Timeliness of data...

- When the value of data is realized can vary tremendously, which suggests timeliness may not always be that important.
- However, data seldom can be collected “after the fact” and thus timeliness in collecting data is extremely important.
- Timeliness is also important because when data are collected we are often involved with other important tasks (e.g., vaccinating).

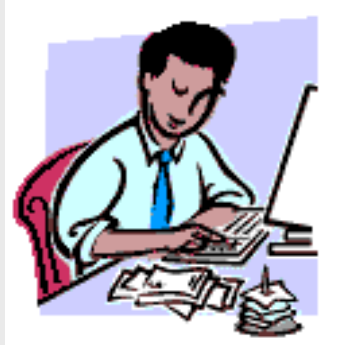


## Data are used for...

- Real-time decision making
  - Sorting cattle
  - Vaccination program (product and amount)
  - Breeding program
- Analysis of relationships for future decision-making (expectations)
  - Evaluating bulls/breeding program
  - Evaluating production programs
  - Evaluating health programs



## Problem...




...front end to the computer is a HUMAN.




## Human factors

- Humans are slow
- Humans are error prone
- The greater the complexity of the data entry, the greater the probability of making an error
- Personnel hired to enter data are often minimally skilled (and may lack knowledge of subject matter at hand)





**Ever seen characters like this?**

 **1506022L**

*Your interpretation?*

36376	31496	30457	29921
3549	35292	28515	28999
3203	30388	26467	26743
38508	30125	26425	22547
37526	25586	32227	30697
26523	32949	27	28774
30216	35224	36672	27580
36451	2	28770	30584
46704	26777	28623	36682
36399	36895	31648	28372
30285	25120	30322	26464
30917	26918	30219	28419
34092	2512	26381	36672
36410	264	30777	28965
32966	20232	20877	30630
30136	36733	28406	30660
36506	36883	36663	28874
32948	35175	36645	28914
30337	30390	36682	30339
36639	31277	28365	30667
30329	32787	30176	28381
34679	35089	26371	28781
36638	26405	36649	36668
32236	35108	30183	36658
33426	36689	26529	
35170	36633	35479	
36702	30894	35471	
36642	36377	35445	
36471			

**Ever had to interpret data like this?**

**What's the odds of an error occurring?** (even if you were the one who wrote it)

**How does this affect accuracy and timeliness of data?**



## **Automatic identification and data capture (AIDC)...**

- Systematic data capture by means other than manual notation or keyboard input or voice.



## AIDC technologies

- Bar coding
- Optical character recognition
- Magnetic stripe
- Voice recognition
- Radio frequency
- Smart card
- Biometrics

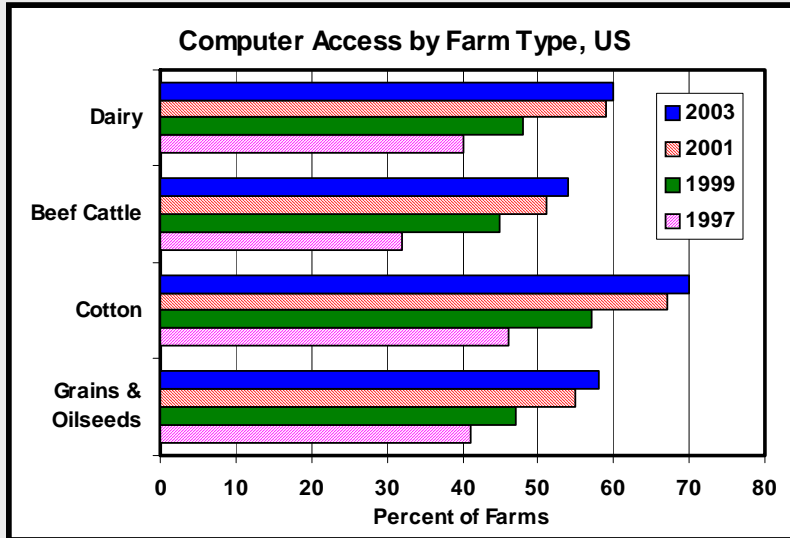


## Smart businesses...

- Recognize that good decisions need to be based on good information which requires them to collect data in an accurate and timely fashion.
- Unfortunately, many producers in the livestock industry have not always been of this mindset.



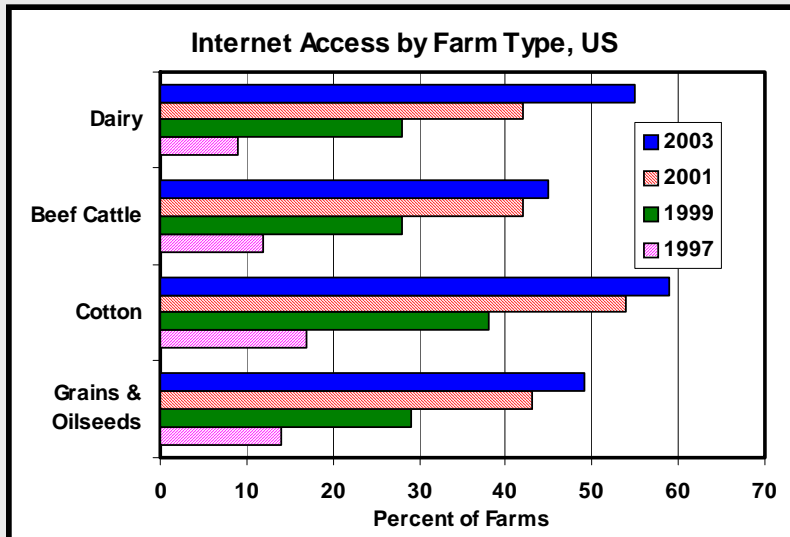
## Technology adoption...



Source: Farm Computer Usage and Ownership, July 28, 2003, NASS USDA.



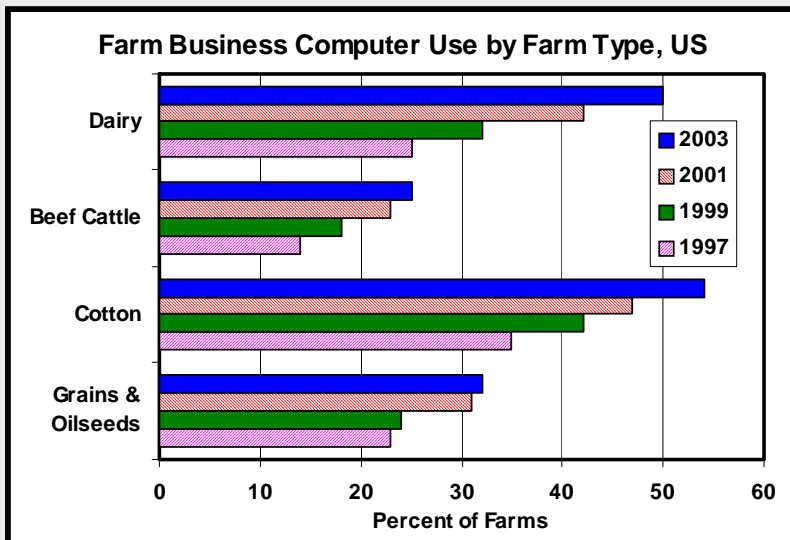
## Technology adoption...



Source: Farm Computer Usage and Ownership, July 28, 2003, NASS USDA.



## Technology adoption...



Source: Farm Computer Usage and Ownership, July 28, 2003, NASS USDA.



## Existing technologies...





## Current technologies ...

- fast, reliable computers
- amazing network capability
- excellent software

**... give us the ability to collect data accurately and timely better today than ever in history.**



## Collecting data summary

- Market is changing rapidly based on consumer's demand (safe, convenient product)
- Market is changing rapidly based on production sector (consolidation, industrialization, etc)
- New technologies are being adopted that allow these changes
- Changes will require livestock operation managers to think differently than in the past



The goal for data collection is to have useful data (*accurate* and *timely*), such that you can **turn data into knowledge and information** that can be used to make good business decisions.



### **Statistical Process Control (SPC) ...**

... the process of analyzing data so that it can be transformed into knowledge and information that is useful in the decision-making process.



## What is SPC?

- Collection of tools, mostly statistical which help us understand what is going on in any process that generates data, products, or services.
- Helps us attain insight into the inherent behavior of those processes which enables us to exercise control over that process.
- Serves to assist in the evaluation, redesign and improvement of the process (planning & control).



## Why statistically analyze data?

### TO MAKE INFERENCES!

- Inferences: relationships expected to hold in the future
  - Optimal management: make the most profitable decisions
  - Management becomes less random



## Some tools of SPC ...

- Flow charts
- Run charts
- Histograms
- Control charts
- Scatter graphs
- Summary statistics
- Sorts/queries
- Statistical analyses (regression, ANOVA)

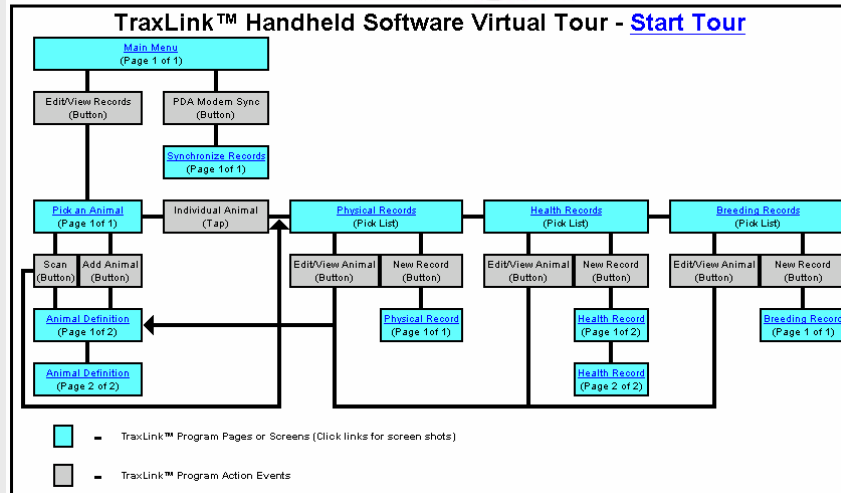


## Flow charts

- A flow chart is a diagram that shows the progress of work or the flow of materials or information through a sequence of operations.
- Not considered a statistical tool, but FC are often useful because they provide an overall view of the entire process.



## Flow charts

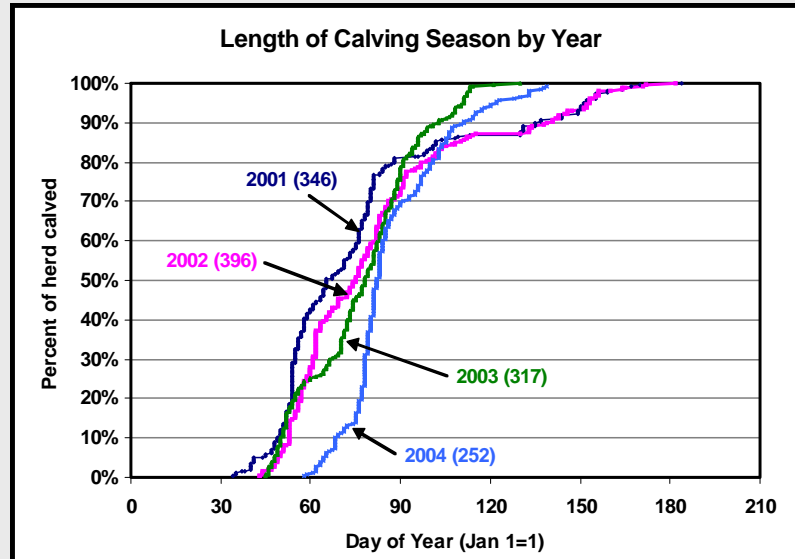


## Run charts

- A run chart is a charting of sample results (data), usually one sample reading at a time, in chronological sequence



## Run charts

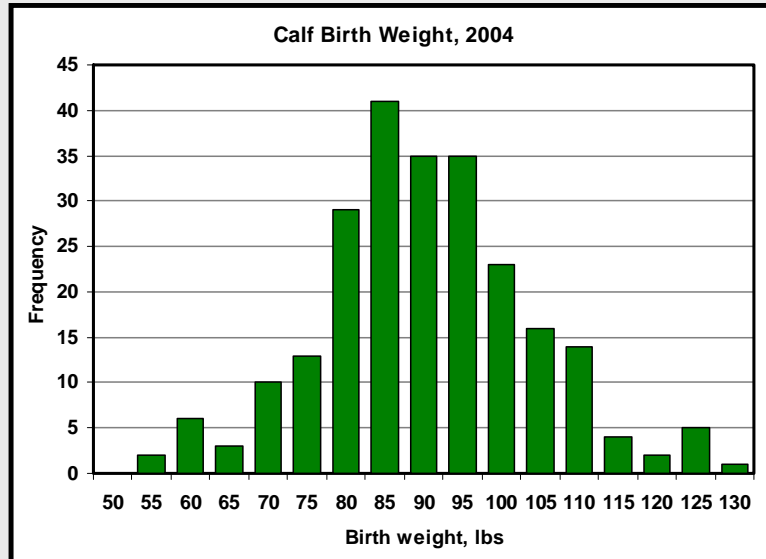


## Histogram (frequency distribution)

- Fundamental statistical tool of SPC
- Histograms are an effective way of summarizing data graphically that, without calculations, can illustrate three characteristics of the data.
  - A sense of the magnitude of the mean (average)
  - A sense of the variability (distribution) of the data
  - Some idea of the possible pattern of variation



## Histogram

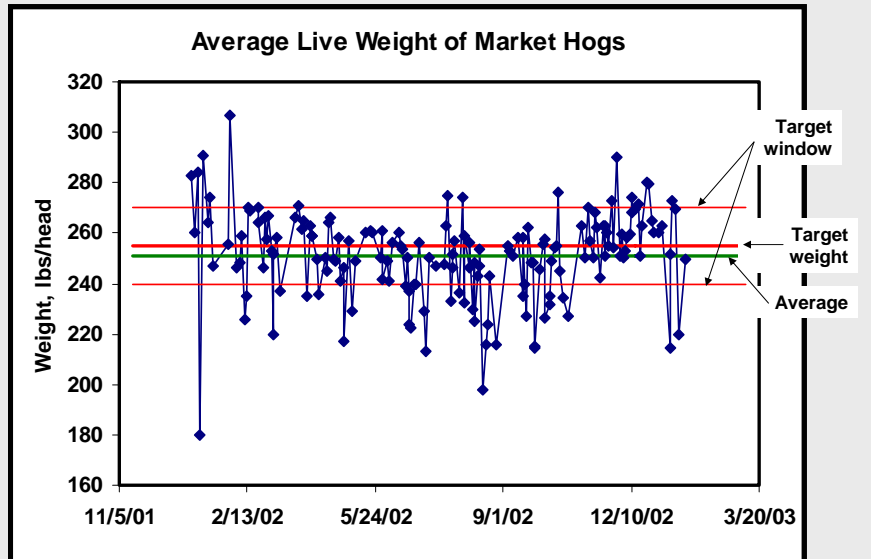


## Control charts

- Control charts are completely identified with SPC
- Control charts give insight into the statistical behavior of processes
- Process (def) – Everything that works together to produce a product, service or other output
  - People
  - Facilities and machinery
  - Genetics
  - Production inputs (e.g., feed, vaccines)
  - Environment



## Control charts

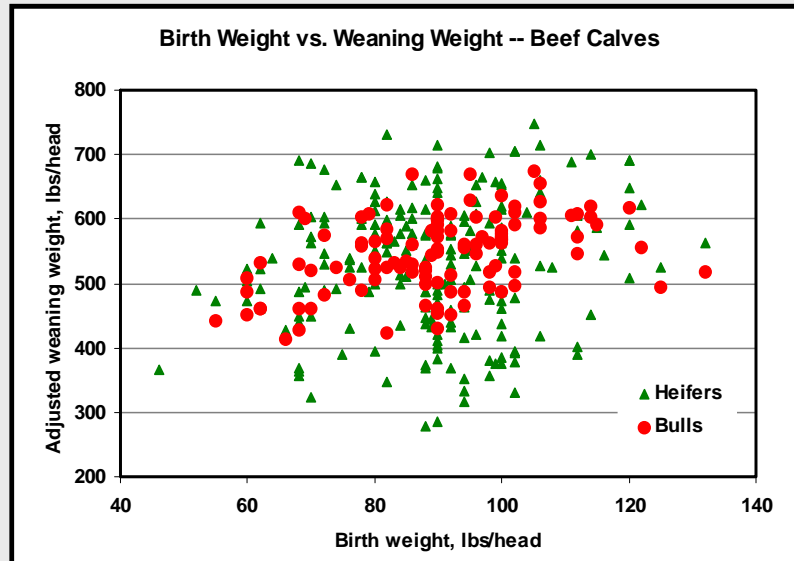


## Scatter graphs

- Scatter diagrams help to discover associations or relationships (correlation) between two variables.
- Usually does not go beyond the graphic representation of the data (can show trend and mathematical relationship fairly easily).



## Scatter graphs



## Data analysis -- visual vs. numerical ...

- Visual data analysis (figures) stimulates ideas that never might emerge otherwise
- Figures get too confusing with more than one or two factors – must move to numerical analysis



## Some tools of SPC ...

- Flow charts
- Run charts
- Histograms
- Control charts
- Scatter graphs
- Summary statistics
- Sorts/queries
- Statistical analyses

Graphical tools are extremely helpful in getting a feel for the data (i.e., first-cut analysis).

Numerical analyses are used to quantify relationships that exist for making decisions.

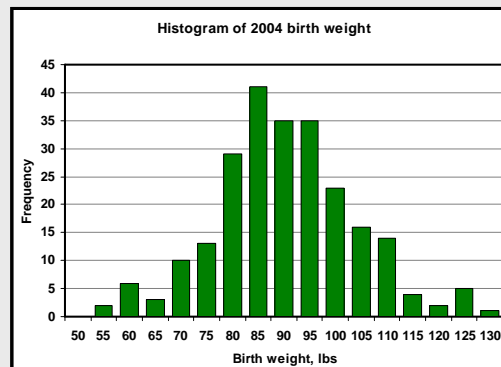


## 2004 calf birth weight ...

- What is the average, how variable is it, and what kind of distribution does it follow?

2004 Birth wt	
Mean	88.95
Standard Error	0.87
Median	89
Mode	80
Standard Deviation	13.45
Sample Variance	180.89
Kurtosis	0.36
Skewness	0.16
Range	74
Minimum	54
Maximum	128
Sum	21260
Count	239

**Descriptive statistics**



**Histogram**



## Measures of central tendency ...

- Mean – simple average (i.e., each observation is given equal weight)
- Median – middle data point in a set of values ranked from low to high
  - Median > mean (left skewed)
  - Median < mean (right skewed)
- Mode – value that occurs most frequently in a data set



## The mean is often our best forecast

	farm A	farm B
Year 1	-\$80	-\$5
Year 2	\$200	\$30
Year 3	-\$50	\$20
Year 4	-\$270	\$25
Year 5	\$300	\$30
Average	\$20	\$20

**How much confidence do you have in the \$20 estimate?**



## When analyzing data ...

- The mean is a powerful measure/concept
- However, the mean does not convey all important and relevant information.
- We often also want to consider the variability in the data.



## Understanding variation...

### Causes of variation

- Materials / inputs
  - Raw materials exhibit variation (e.g., feed)
- Machines
- Methods
  - Ways that materials, machines and people are configured to produce goods and services
- Genetics (applies to both people and animals)
- Environment



## Measures of variability ...

- Range -- the difference between the largest reading and the smallest reading.
- Standard deviation -- a measurement of the total variability of the data. It is an average of deviations from the mean.
- Coefficient of variation (CV) -- normalized measure of variability equal to standard deviation divided by the mean.

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$



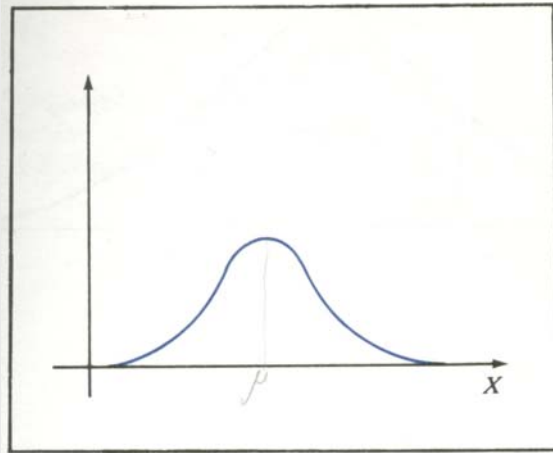
## Standard deviation ...

- Same unit of measure as the original data
- Affected by extreme values, which tend to enlarge the standard deviation.
- Larger values for standard deviation indicate the data are more widely dispersed around the mean (i.e., more variable).
- Generally has the most meaning when data are normally distributed.



## The Normal “Bell-Shaped” Curve

- Symmetry
- A bell shape
- Smoothness



## Distribution variations ...



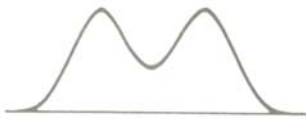
Normal distribution.



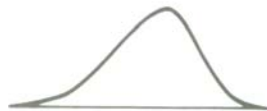
Measurements tail off to the right.



Measurements from two different groups



Measurements from two different groups that overlap.



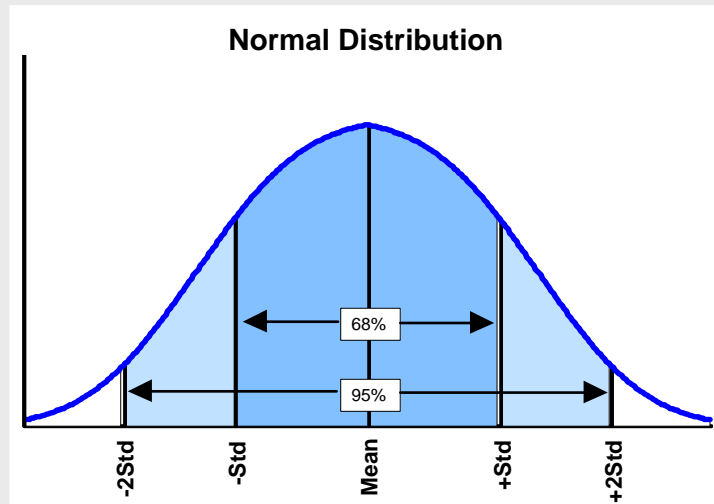
Measurements tail off to the left.



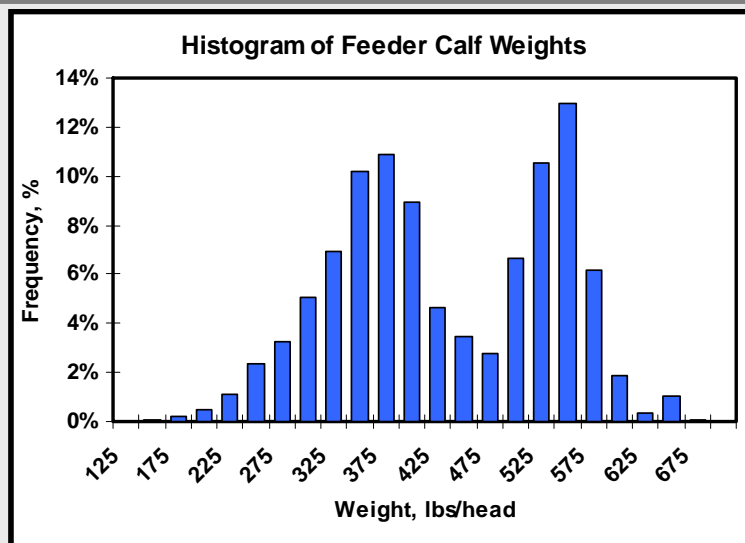
Tails of the distribution are chopped off. Parts were probably sorted.



- Normal distribution -- Many statistical analysis measures assume a normal distribution



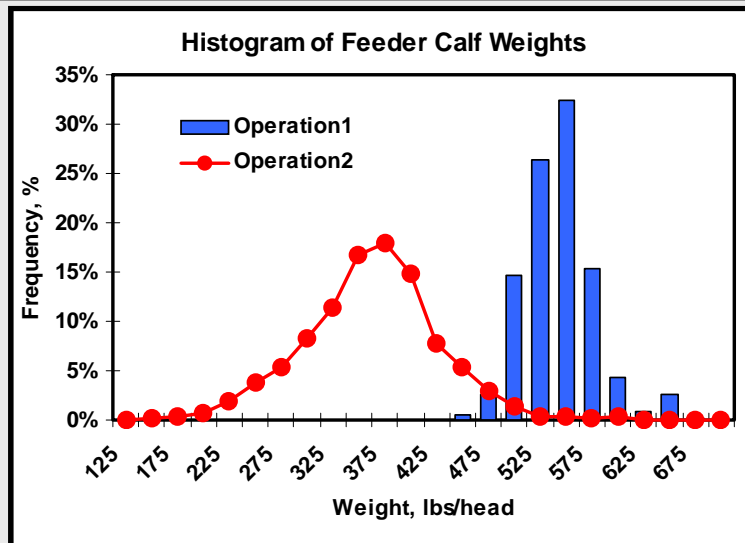
## Histogram



**Bi-modal distribution, what might be going on?**



## Histogram



Given more information, distributions appears normal.



## Statistical measures

- Correlation – degree to which two series tend to rise and fall together.
- *R*-square – percent variability in the dependent variable that is explained by the independent variable(s) [equal to the squared correlation between predicted and actual series].
- *P*-value – probability of rejecting an hypothesis when it is actually true and  $(1 - p\text{-value})$  is an indication of the statistical confidence we can place in our results.



## Statistical significance

	<u>farm A</u>	<u>farm B</u>
Year 1	-\$80	-\$5
Year 2	\$200	\$30
Year 3	-\$50	\$20
Year 4	\$300	\$25
Year 5	-\$270	\$30
Average	\$20	\$20

How much confidence do you have in the \$20 estimate?

Expected value vs. confidence in expectations.

$H_0$ : Mean = 0;

$H_A$ : Mean > or < 0;

$p$ -value = 0.85 (A)

$p$ -value = 0.02 (B)



## Data queries...

- Conditional means, i.e., means of subgroups
  - Weaning weight (WW) for heifer calves
  - WW for heifers from cows 3-5 years old
  - WW for heifers from cows 3-5 years old on pasture A
- Requires data from many observations
- A big thing for data warehouses
- May not lead to optimal decisions



## Statistical models

- Linear regression – statistical equation where variation in independent variable(s) explain variation in a dependent variable
- Generalizing “causal models”  
 $y = A + Bx + e$   
 $y = A + B_1x + B_2x^2 + B_3z + B_4xz + e$



## Regression analysis

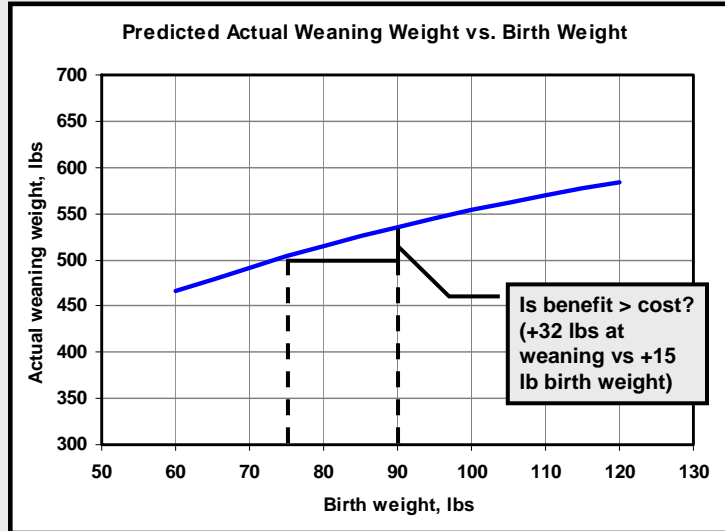
$$\text{Actual WW} = b_0 + b_1(\text{BW}) + b_2(\text{BW})^2 + b_3(\text{A}) + b_4(\text{S}) + b_5(\text{BD}) + b_6(\text{AxBD}) + e$$

where,

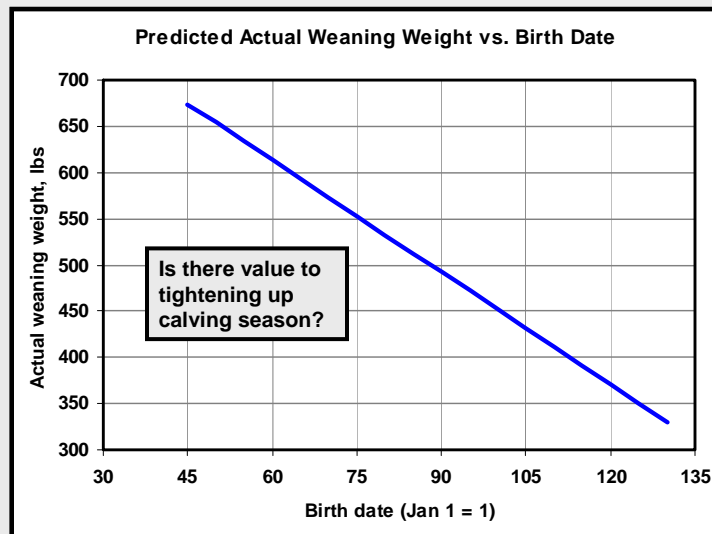
- BW = birth weight (pounds)
- A = age of cow at calving (years)
- S = sex of calf (steer=1, heifer=0)
- BD = birth date of calf
- $b_0 - b_6$  are coefficients to estimate and  $e$  depicts an estimation error ( $b$ 's are “picked” to minimize sum of squared errors)



## Relationship between BW & Act WW ...



## Relationship between birth date & Act WW ...





## Regression analysis

$$ADG = b_0 + b_1(TRT1) + b_2(WT) + b_3(WT)^2 + b_4(CT) + b_5(DH) + b_6(RES) + e$$

where,

- ADG = average daily gain (pounds)
- TRT1 = respiratory treatment number 1 (yes/no)
- WT = initial weight of calf (pounds)
- CT = castrated (yes/no)
- DH = dehorn (yes/no)
- RES = days treated for respiratory problems
- $b_0 - b_6$  are coefficients to estimate and  $e$  depicts an estimation error ( $b$ 's are "picked" to minimize sum of squared errors)



## Relationship between health factors & ADG

	OwnerM	OwnerP	Average
Starting weight	531	351	427
Total days on feed	84.1	93.8	89.7
<b>Health factor</b>			
Castrate	-0.388	-0.212	-0.284
Respiratory days	-0.105	-0.087	-0.092
Avg days treated for res	2.4	3.0	2.7
Max days treated for res	17.0	19.0	19.0
<b>ADG impact</b>			
Castrate	-32.6	-19.8	-25.5
Treat for respiratory (avg days)	-20.7	-24.0	-22.4
Treat for respiratory (max days)	-149.8	-154.2	-157.2
<b>Purchase price discount, \$/cwt</b>			
Castrate	-\$3.36	-\$3.08	-\$3.27
Treat for respiratory (avg days)	-\$2.13	-\$3.74	-\$2.87
Treat for respiratory (max days)	-\$15.41	-\$23.98	-\$20.13
Value of gain, \$/cwt	\$54.62	(1994-2003 avg for 200# gain from Oct-Jan)	



## Why should we analyze data?


It is through the analysis of data, that “collected data” are transformed into knowledge, which then gives managers the information they need to make sound business decisions.

***It's all about improving the decision-making process!***

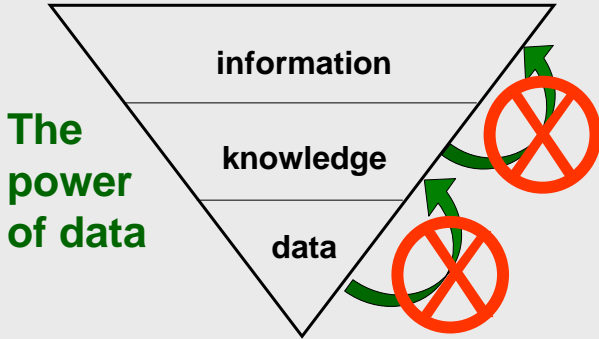


## Analyzing your data ...

- Many ways of analyzing data
  - Visual (e.g., charts)
  - Numerical (e.g., summary statistics, regression)
- No one method is correct or “best” in every situation.
- The key to working with your data (regardless of the method used) is to generate knowledge that helps give information to improve the decisions made by the manager.



## What happens if data or method of analyses are wrong?



**The power of data**

information

knowledge

data

... worse yet, you don't know data are wrong and you base decisions on bad information.

AgManager: Info on Crops, Livestock, Farm Management and Policy - Microsoft Internet Explorer

Address <http://www.agmanager.info/>

about contributors useful links site map feedback Search

Contact Us: Department of Agricultural Economics 342, Waters Hall, Manhattan, KS 66506 (785) 532-5823

**AG MANAGER INFO**  
Kansas State University & Extension

- Agribusiness
- Crops
- Farm Management
- Human Resources
- Income Tax & Law
- Livestock & Meat
- Policy
- Ag Econ News
- Contributors
- Events
- Programs
- Sponsors

**A Website Providing Information and Tools For The Competitive Business**

[www.agmanager.info](http://www.agmanager.info)

**Questions ???**

**Site Updates**

- Canadian Beef Ban Lifted**  
KKSU AgToday July 18, 2005
- In The Cattle Markets**  
July 18, 2005 by LMIC
- Updated Crop Basis Tool**  
July 16, 2005 by Kevin Dhuyvetter
- Updated Cattle Databases**  
July 15, 2005 by Jim Mintert
- Livestock and Hay Charts**  
July 15, 2005 by Jim Mintert
- Crop Questions**  
July 14, 2005 by Mike Woolverton
- Labor Law Updates**  
July 14, 2005 by Sarah Fogelman
- Planning For Farm Business Transition/Succession**  
July 11, 2005 by Rodney Jones
- Diesel Fuel and NIB Price Forecasts**  
July 1, 2005 by Kevin Dhuyvetter
- Updated Crop Databases**  
June 27, 2005 by Kevin Dhuyvetter
- Crop Basis Maps**  
June 15, 2005 by Dhuyvetter
- Quarterly Livestock Round-Up**  
June 7, 2005 by Mintert and LMIC

Kevin C. Dhuyvetter  
785-532-3527  
kcd@ksu.edu

Economics K-State Research & Extension College of Agriculture Kansas State University

start Inbo... 2 Win... Windo... Microso... AgMan... 4:53 PM