Accuracy, precision & prediction of nutritive value
Precision

- The degree to which the **same result** can be obtained repeatedly on a sample.
- High precision implies less replication is required.

- Precision depends on:
  - **Sampling** error
  - **Analytical** error (equipment & procedure)
  - **Human** error

- But does high precision imply accuracy?
- Can precise methods produce inaccurate results?

See [http://www.flatsurv.com/accuprec.htm](http://www.flatsurv.com/accuprec.htm)
Accuracy

- **Accuracy** means closeness to the *true* result (in this case the reference result).

1. The degree of conformity with a standard (the "truth").

2. Depends on appropriateness of method for achieving result.

See [http://www.flatsurv.com/accuprec.htm](http://www.flatsurv.com/accuprec.htm)
Accuracy is telling the truth . . .
Precision is telling the same story over and over again.

Yiding Wang,
yiwang@mtu.edu

http://www.flatsurv.com/accuprec.htm
Which of the following shows:

1. Precision & accuracy?  
   ![Diagram A]

2. Precision with blunder?  
   ![Diagram B]

3. Accuracy with blunder?  
   ![Diagram C]

http://www.flatsurv.com/accuprec.htm
<table>
<thead>
<tr>
<th></th>
<th>Acceptable (1 s.d.)</th>
<th>Warning (2 s.d.)</th>
<th>Unacceptable (3 s.d.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DM (Moisture)</td>
<td>±0.10</td>
<td>±0.20 s.d.</td>
<td>±0.30 s.d.</td>
</tr>
<tr>
<td>Ether extract</td>
<td>±0.10</td>
<td>±0.20 s.d.</td>
<td>±0.30 s.d.</td>
</tr>
<tr>
<td>Crude protein (10%)</td>
<td>±0.10</td>
<td>±0.20 s.d.</td>
<td>±0.30 s.d.</td>
</tr>
<tr>
<td>Crude protein (20%)</td>
<td>±0.20</td>
<td>±0.40 s.d.</td>
<td>±0.60 s.d.</td>
</tr>
<tr>
<td>Ash</td>
<td>±0.10</td>
<td>±0.20 s.d.</td>
<td>±0.30 s.d.</td>
</tr>
</tbody>
</table>
## NFTA precision indices

<table>
<thead>
<tr>
<th></th>
<th>Acceptable (1 s.d.)</th>
<th>Warning (2 s.d.)</th>
<th>Unacceptable (3 s.d.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DM (Moisture)</td>
<td>±0.10</td>
<td>±0.20 s.d.</td>
<td>±0.30 s.d.</td>
</tr>
<tr>
<td>NDF (40%)</td>
<td>±0.35</td>
<td>±0.70</td>
<td>±1.05</td>
</tr>
<tr>
<td>NDF (70%)</td>
<td>±0.60</td>
<td>±1.20</td>
<td>±1.80</td>
</tr>
<tr>
<td>ADF (20%)</td>
<td>±0.20</td>
<td>±0.40</td>
<td>±0.60</td>
</tr>
<tr>
<td>ADF (40%)</td>
<td>±0.35</td>
<td>±0.70</td>
<td>±1.05</td>
</tr>
</tbody>
</table>
Predicting nutritive value

Definitions

- Predictor (Independent variable)
- Predicted term (dependent variable)

Important considerations

1. Population
   - Adequacy of no. of samples
   - Representation of typical variability

2. Relationship validity
   - E.g. should maintenance-fed cows be used to predict digestibility in lactating cows?
   - Can sheep adequately model digestion in cows?
Is forage degradability the same in cows versus sheep?

(Adesogan et al, 1995)
Criteria for choosing a predictor

- Biologically meaningful
- Rapidly & easily determined
- Inexpensive to analyze
- Routinely practicable & repeatable
- Precise & accurate
- Ethical
The relationship

- What does the \textit{scattergram} suggest
- How accurate should the relationship be \textit{i.e.}
  - What margin of error can I accept
- Has the relationship been validated
- Is it a robust relationship
- Is it a population-dependent relationship
- \textit{i.e.} Is it described best by a common line, separate lines or parallel lines
Prediction of IVDMD from Ankom bag values

\[ y = 0.89x + 5.16 \]

\[ r^2 = 0.83; \text{rsd}=3.29 \]
In vitro versus in vivo DOMD
Measures of acceptability of prediction equations

1. Correlation coefficient (r)
2. Coefficient of determination ($r^2$)
3. Root mean square error (RMSE) or residual standard deviation (rsd),
4. Standard error of the estimate (sy.x),
5. Standard error of prediction (SEP),
7. Coefficient of variation (CV)

Dependence on population size

- High
- Low
Coefficient of variation

- Can be used to predict the number of animals required for an experiment

- See Dr. Mike Galyean’s website

- http://www.asft.ttu.edu/home/mgalyean/
The correlation coefficient

\[ r = \text{degree of association between two variables} \]
Correlation coefficient

Measures only linear association
- Curvilinear relationships regarded as uncorrelated

Does not reflect causation
- E.g. correlation b/w of vets & cow deaths in Florida could be high but meaningless

Significant r values ≠ accurate prediction equations (values > 0.8 required for accuracy).
Coefficient of determination \((r^2)\)

- Measures the variability due to the predictor in the predicted term.

- However:
  - Does not imply a causal relationship.
  - Measures the prediction strength of a population, not an individual value.
  - Is population-size & range dependent.

- \(r^2\) adjusted. Adjusts the \(r^2\) value for sample size, giving a more realistic value for small samples.
Measures of accuracy

Neither ‘r’ nor ‘r²’ indicates the error associated with the prediction.

Yet predictions in any biological system involves an error of prediction which estimates the limits within which the actual predicted value will be found (Givens, 1986).
Residual standard deviation

- Indicates the variance of observations about the relationship
  - Provided the relationship is linear.
  - Is the square root of the ANOVA RMS
  - Normally numerically equivalent to the SEC
  - Unaffected by population size
  - Also called the residual mean square error (RMSE)

- Is the minimum error that must be applied to the predicted term i.e. used to build confidence intervals for the mean

- rsd of $\leq 2\%$ of the mean often indicates accuracy
Accuracy of predicting in vivo digestibility

<table>
<thead>
<tr>
<th>Technique</th>
<th>r</th>
<th>RSD (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP</td>
<td>+0.44 to +0.79</td>
<td>2.0 - 6.5</td>
</tr>
<tr>
<td>NDF</td>
<td>-0.45 to -0.80</td>
<td>2.4 - 5.1</td>
</tr>
<tr>
<td>ADF</td>
<td>-0.75 to -0.88</td>
<td>3.6 - 9.0</td>
</tr>
<tr>
<td>Lignin</td>
<td>-0.61 to -0.83</td>
<td>4.3</td>
</tr>
<tr>
<td>IVOMD</td>
<td>+0.94</td>
<td>1.6</td>
</tr>
<tr>
<td>Enzymatic OMD</td>
<td>+0.94 to +0.99</td>
<td>1.8 - 3.2</td>
</tr>
<tr>
<td><em>In situ</em> degradability</td>
<td>+ 0.98</td>
<td>1.0 - 3.0</td>
</tr>
<tr>
<td>NIRS</td>
<td>+0.95</td>
<td>0.6 – 2.7</td>
</tr>
</tbody>
</table>

(Lopez, 2000)
Multivariate relationships

- Use $R^2$, not $r^2$
- May not be more accurate
- Additional predictor use improves prediction
- Additional predictors must be
  - Biologically justifiable
  - Not autocorrelated with the first predictor
  - Significantly improve the fit of the regression
Before considering correlation coefficients and regressions, the ranges and variations of chemical composition and energy value in the population used must be appreciated to prevent the use of regressions beyond such ranges (Morgan, 1973).

Extrapolation beyond the limits of the population used to develop an equation is ill-advised and populations must be well defined to avoid the need for extrapolation (Weiss, 1993).
Empirical vs Mechanistic equations

- **Empirical models**
  - Define the requirement for each nutrient in isolation for given target production levels
  - E.g. $\text{OMD} = 0.566 + 0.00081\text{CP}$
  - May produce accurate, but biologically meaningless results
  - Poorly account for other mitigating factors
  - V. widely used
Summative equations

- Include empirical estimates but also some biological meaning.
  - E.g. Van Soest's summative equation
  - DMD = 0.98 \times NDS + NDF_d - 12.9

- May attempt to integrate underlying mechanisms for nutrient absorption with genetic potential
Mechanistic models

- Often more complex, may require computer simulations
- Explain the underlying mechanisms
- E.g. Digy prediction from
  - sample processing, particle size, intake, interactions among feeds, rates of passage and diet composition
- Do benefits (biological meaningfulness & improved accuracy) justify the cost
- Currently many are less accurate than empirical models but are useful for highlighting areas requiring research
- Future models may be more accurate
References