Randomness Rules: Living with Variation in the Nutrient Composition of Concentrate Feeds

William P. Weiss
Department of Animal Sciences
Ohio Agricultural Research and Development Center
The Ohio State University, Wooster 44691

Summary

The nutrient composition of feeds, including concentrates and byproducts, varies. This is an absolute truth, and nutritionists must learn to accommodate variation when formulating rations. The variation in concentrations of most nutrients in corn grain, sorghum grain, and soybean meal is actually quite small. Feedstuffs that are a direct co-product of the production of human food or beverages (e.g., brewers grains, distillers grains, corn gluten feed) generally have moderate variation in nutrient composition. Feed ingredients that are poorly defined or have a very general definition (e.g., potato waste, grain screenings) can have extremely high variation in nutrient composition. Variation in feed composition increases risk and that has a cost. If a specific feed has a large load-to-load variation, diets either have to be over supplemented to avoid a deficiency (i.e., increased feed costs) or production may decrease because at times the diet does not provide adequate nutrients. Feeds with large variation in nutrient composition are worth less than feeds with less variation. If adequate nutrient composition data are obtained and summarized for specific ingredients from a specific source, a reliable mean and standard deviation can be calculated. Those values, rather than values from a single sample, should be used to formulate diets. The standard deviation can be used to *adjust* nutrient composition so that the risk of a deficiency of a specific nutrient can be greatly reduced. The impact of variation in nutrient composition of a feedstuff is reduced when diets contain a wide variety of different ingredients.

Introduction

A fundamental rule regarding diet formulation is that you will never know the true value of anything. Although we have reasonably accurate estimates of the **average** requirements for most nutrients, we have less certainty regarding the nutrient requirements of a specific herd or animal under specific circumstances. We have equations that accurately estimate the **average** dry matter intake for groups of cows, but we are not very good at estimating how much dry matter a specific cow is eating. We have developed several good analytical procedures to measure the concentrations of many nutrients in feeds and tables are available that contain the **average** nutrient composition of all feeds commonly fed to dairy cows. However, biological and manufacturing variation, variation caused by sampling, and variation in analytical measurements can be substantial; thus concentrations of nutrients within a specific feedstuff may be quite different from the average. Does all this uncertainty mean that we should give up on ration formulation and feed analysis? The answer to that question is obviously, no. However, the uncertainty associated with feed analysis and ration formulation must be understood and addressed. With proper sampling techniques, adequate number of samples, and appropriate data handling, you can reduce the uncertainty associated with feed analysis data. The objective of this paper is to discuss expected variation in feed composition, factors affecting that variation, and methods you can use to increase the reliability of feed analysis data. Although much of the following discussion is appropriate for all feeds, discussion will be limited to grains and byproduct feeds.

Elementary Statistics

We need to start thinking about feed composition data in terms of probabilities rather than actual, absolute concentrations. In other words, how confident are you (or should you be) that the number you have actually represents the true concentration of a nutrient in a feed? Because we are working with

probabilities, a basic understanding of some statistical principles and terminology is needed.

Populations and Samples

The ultimate goal of feed analysis is to obtain an analytical value from a sample that reflects the actual value of a *population*. Examples of a population include a truckload of distillers grains, all the distillers grains produced by a specific distillery, or perhaps all the distillers grains produced in the U.S. In statistical terms, a population is defined as a larger set from which samples are obtained. In other words, a population is defined by sampling. If distillers grains from a single distillery were sampled extensively, we would have a good estimate of the average nutrient composition of distillers grains produced at that plant. However, since other distilleries were not sampled we should be very hesitant to extrapolate the data obtained from a single distillery (i.e., a population) to the larger population of all distilleries.

Central Tendency and Dispersion

A population is represented by a set of observations or samples. Because of inherent variation and variation caused by sampling and analytical procedures, we know that all the sample values will not be the same. Rather than one single value, you will obtain a distribution of values. The two most important pieces of information we need to obtain from that set of samples are a measure of central tendency and a measure of dispersion. For observations that follow a normal statistical distribution, the mean (in this discussion average and mean will be used interchangeably) is the best measure of central tendency. The mean of a normal distribution is not the absolute *right* answer, but rather it is the value that has the lowest probability of being substantially wrong. The concentrations of most nutrients in plant-based feedstuffs fit an approximately normal distribution; therefore the mean concentration is the best measure of central tendency for those nutrients. With a normal distribution, approximately one-half of the samples will have values lower than the mean and one-half will have concentrations higher than the mean. The concentrations of trace minerals (and sometimes fat) in plant-based feeds often have a skewed distribution (a few observations will have very high concentrations). With this type of distribution, the mean overestimates central tendency (less than one-half of the observations will have concentrations higher than the mean). The median (the value at which half the observations are higher and half are lower) is the best measure of central tendency for this type of distribution.

Although many people are familiar with and often use measures of central tendency (i.e., the mean) in ration formulation, fewer people consider or use measures of dispersion in ration formulation. In simple terms, a measure of dispersion should be used to determine how much confidence you should have when using a mean value. When a distribution of values has a large dispersion, the probability of being substantially wrong when using the mean increases. For a normal distribution the most common measure of dispersion is the standard deviation (**SD**). In a normal distribution approximately 38% of all observations will be within $\pm$ 0.5 SD units of the mean, 68% of all observations will be within $\pm$ 1 SD of the mean and approximately 95% of the observations will be within $\pm$ 2 SD of the mean. For example, if the mean concentration of crude protein (**CP**) in a population of brewers grains was 25% and the SD was 2 we would expect that about 68% of the samples from that population would contain between 23 and 27% CP and 95% of the samples would contain between 21 and 29% CP. The smaller the SD, relative to the mean, the less likely it will be that using the mean value will cause a substantial error in diet formulation.

Sources of Variation

Understanding potential sources of variation in feed composition data will help determine which data to use and how to use it. The nutrient composition of grains and byproducts can be influenced by plant genetics (hybrid, variety, etc.) and growing conditions (drought, climate, soil fertility, etc.). In addition, the composition of byproducts is affected by manufacturing techniques. The above sources of variation are considered fixed (i.e., they can be described and replicated). Hybrid X may have been genetically selected to produce corn grain with higher than average concentrations of protein. Distillery Y might dry their distillers grains at very high temperatures, causing high concentrations of acid detergent insoluble protein. A drought may reduce kernel size; thereby increasing the concentration of fiber in corn grain. Another possible fixed source of variation is the analytical lab. Although great

progress has been made in standardizing methods, labs may use different analytical techniques to measure nutrients. If lab A measures neutral detergent fiber (**NDF**) using sulfite but another lab does not, the NDF concentrations will differ between the labs because of procedure. Other sources of variation are considered random. We do not know why the values differ, they just do. If you sample a load of brewers grains ten times and send those ten samples to a lab, you will probably get back ten slightly different concentrations of protein. The variation could be caused by variation within the load of brewers grains or it could be caused by random errors at the lab.

Ideally, random variation would be considered within population variation and fixed variation would be considered as variation between populations. For example, because of manufacturing differences, distillers grains from distillery X has consistently higher NDF concentrations than distillers grains from distillery Y. If distillers from X and Y were considered separate populations, the SD within each population would be expected to be lower than the SD if the distillers grains from both distilleries were combined. Because of blending grains and multiple sources of feedstock for manufacturing facilities, many fixed sources of variation become blurred (you probably will not know the variety of the soybeans used to make the soybean meal you purchased or whether the gluten feed you purchased was made from drought-stressed corn grain). In these situations, the fixed sources of variation become random sources, resulting in an increase in the within population variation. Nonetheless, accounting for as many fixed sources of variation as possible by defining separate populations will reduce the dispersion of the data and reduce the potential of being substantially wrong when using the mean.

Expected Variation in Nutrient Composition of Grains and Byproducts

The largest publicly-available data base of feed composition can be found in the NRC dairy publication (NRC, 2001). That database contains both means and SD for measured nutrients in most common feeds. The data used to calculate those means and SD came from a wide array of sources. Samples came from across the U.S. and over several years. For some feeds and nutrients, the number of samples used to calculate the mean and SD are quite limited and those values should be used carefully. For other feeds, the sample size is quite large and the mean and SD are probably good estimates for the broad population from which the samples were drawn. However, it is important to remember that the broad population represented in the NRC tables may not be a good estimate for a specific source of a feed. Kertz (1998) also provides data on variation in nutrient composition of a limited number of feeds.

Based on expected variation feeds can be classified as having low, moderate, or high variability. Feeds with generally low variability include corn grain, sorghum grain, and perhaps barley (Table 1). Feeds with the largest variability in composition are byproducts that are usually not a direct end product of the production of some other product. For example, potato waste has extremely high variability because it may include cull potatoes, potato peels, waste products from the manufacturing of potato products for human consumption, rejected product, etc. Millrun, corn screenings, and cannery waste are other examples of feeds that are not well-defined and would be expected to have high variability, even when they come from the same originating source. Feeds with moderate variability include most feeds that would be considered co-products rather than byproducts. For example, distillers grains, brewers grains, and corn gluten feed are an end product of alcohol, beer, and corn sweetener production. Since production of these products is generally well-controlled, the composition of the resulting co-product is relatively constant within a production facility. To increase the accuracy of ration formulation, feeds with moderate and high variability in composition must be sampled and analyzed routinely and the data generated must be used correctly.

An accurate estimate of SD for a specific feedstuff can be extremely useful in ration formulation. The SD should be considered when deciding on ration *safety* factors. The SD in the NRC table is a function of inherent variation in composition of the grain or feedstock, lab-to-lab variation, variation among manufacturing processes, and many other sources of variation. If no other measure of dispersion is available to you, the SD in the NRC table can be used; however, you must remember that for many feeds, the actual variation you will observe could be substantially less than the SD in the NRC table (Table 2).

Table 1. Mean (0) concentrations and standard deviations (**SD**) for selected nutrients and selected feeds. Data are from NRC (2001) and represent very diverse populations.

	CP		NDF		EE		Ash	
	0	SD	0	SD	0	SD	0	SD
Grains								
Barley	12.4	2.1	20.8	8.6	2.2	0.6	2.9	0.8
Corn	9.4	1.3	9.5	2.3	4.2	1	1.5	0.5
Sorghum	11.6	1.8	10.9	5	3.1	0.8	2	0.6
Byproducts								
Brewers grains, wet	28.4	4	47.1	6.8	5.2	1.6	4.3	0.9
Corn gluten feed	23.8	5.7	35.5	6.8	3.5	1.1	6.8	1.5
Distillers grains, dry	29.7	3.3	38.8	7.8	10	3.4	5.2	1.1
Potato waste	10.5	8.4	22.1	14.3	10.8	7.8	12.8	7.4
Rice bran	15.5	2.2	26.1	6.8	15.2	4.2	10.4	1.9
Soy hulls	13.9	4.6	60.3	7.4	2.7	1.4	4.8	0.7
Soybean meal-48	53.8	2.1	9.8	5.6	1.1	0.4	6.4	0.7
Wheat midds	18.5	2.1	36.7	7.5	4.5	1.3	5	0.8

Several common feeds were sampled and analyzed over a one year period in California (DePeters et al., 2000). All analysis were conducted at a single lab and for the feeds that will be discussed, all samples within a feed came from the same production facility. A similar type study was conducted in Missouri (Belyea et al., 1989). Dried distillers were sampled in both studies. The calculated distribution of CP concentrations are shown in Figure 1 for the two studies and for NRC data. Mean concentration of CP was very similar for the three data sets (29.7, 30.6, and 31.2% of DM for NRC, MO, and CA); however dispersion differed greatly. The SD for NRC, MO, and CA were 3.3, 1.6, and 0.6. Based on the means and considering typical dietary inclusion rates for distillers grains, essentially the same concentration of dietary CP would be obtained regardless of the source of the data. However, because the SD is substantially lower when all samples were obtained from a single source, you would be much less likely to make a substantial error in formulation (i.e., diet is actually deficient in CP) when you use a mean value if the sample is from a limited, rather than a broad population. Not all feed or nutrients follow the pattern shown for distillers grains in Figure 1. Figure 2 shows the distribution of CP concentrations for rice bran from the NRC data set (broad population) and from Belyea et al. (1989; limited population). If you used the mean concentration of CP from NRC for rice bran obtained from the particular production facility sampled in the Missouri study, you would underestimate the CP concentration substantially; resulting in increased protein supplementation costs. For nutrients that are routinely measured, means obtained from a broad population (e.g., NRC) should be used only when other data, specific to a limited population, are not available.

How Should You Handle Variation in Feed Composition

Feeds with low expected variability do not have to be analyzed routinely and in some cases not at all.

Table 2. Mean (❶) concentrations and standard deviations (**SD**) for selected nutrients and selected feeds. The California data are from DePeters et al. (2000) and the Missouri data are from Belyea et al. (1989). Within experiment and feed, samples originated from the same production facility (i.e., limited populations). These values should be compared to those in Table 1 (a broad population).

	CP		NDF		EE	
	❶	SD	❶	SD	❶	SD
CA						
Brewers grains, wet	27	2.2	37.3	3.4	6.3	0.4
Corn gluten feed, wet	22.9	4.3	38.8	3.8	3.4	0.4
Distillers grains, dried	31.2	0.6	35.6	8.2	13	1.3
MO						
Corn gluten feed, dry	23.3	1.4	51.9	2.3	6.6	1.9
Distillers grains, dried	30.6	1.4	33	1.5	7.4	0.9
Rice bran	19.1	0.4	21.8	1.3	17.3	1.9
Soybean hulls	11.8	0.2	72.5	0.8	0.8	0.3

Sampling and analytical errors become relatively small when large numbers of samples are analyzed. For these feeds, a mean derived from a large number of samples may actually be better than a single observation or a mean from a small set of samples. For these feeds, I recommend using book values unless you have good reason to believe your particular feed is different (for example, if you grow or buy high oil corn, the mean values for regular corn would not be appropriate).

For feeds with moderate or high variability in nutrient composition, routine feed sampling and analysis is essential. Although most people realize this, it is often not done because by the time they get the report back from the lab, the load has been fed. If this is your opinion, you are not using the analytical data correctly. As stated above, we need to think in terms of probabilities, not absolute numbers. **You should be sampling and analyzing load samples to obtain estimates of mean composition and SD**; the values obtained from a single load sample are not that important. The frequency of sampling depends on the expected variation and how much error you are willing to accept. Populations with large variation require more sampling to obtain accurate estimates. I cannot give you a specific number of samples needed because it varies depending on the nutrient of interest (e.g., the number of samples needed to obtain accurate estimates of the mean and SD for CP is usually less than that needed for NDF) and the population. As a general guideline ten or so samples of a given population is reasonable. For highly variable feeds more samples would be desirable.

The approach followed by many nutritionists is to sample a load of feed, have it analyzed, and then formulate a diet based on that information. When a new analysis is obtained, the previous data is eliminated and a new diet is formulated based on the new composition. The inherent assumption of this practice is that the new data better represents the feed than did the old data. This may or may not be true. When new analytical data are obtained, the user should ask one simple question, "Is there a good reason why the composition changed?" Possible answers to that question include: "I changed suppliers", "the distillery changed production methods", or probably most commonly, "I don't know". If you cannot think of a good reason the composition changed, the change may simply be a random event. The difference could be caused by load-to-load random variation, by within load (i.e., sampling) variation, or both. In this case, the new

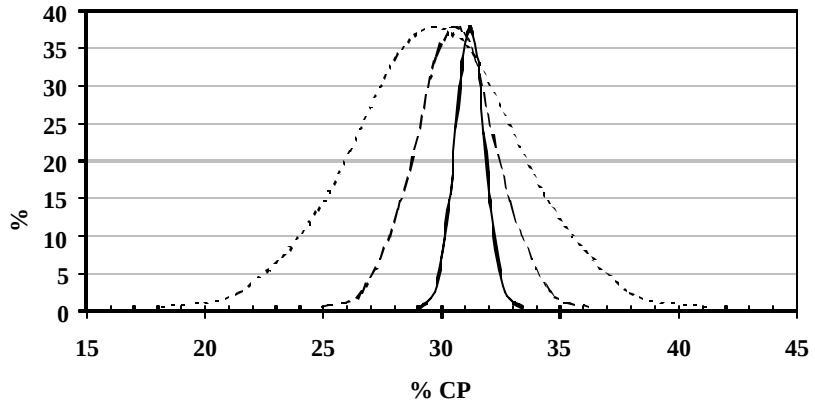

Figure 1. Distributions of crude protein (**CP**) concentrations in dried distillers grains. The small dashed line represents data from a nationwide population (NRC, 2001). The large dashed line represents samples from a single source in Missouri (Belyea et al., 1989) and the solid line represents samples from a single source in California (DePeters et al., 2000). Although mean concentrations were similar among populations, note that dispersion is substantially less for the limited populations (CA and MO) compared to the broad population (NRC). Distributions were calculated based on the mean and SD.

number may be no better than the old number, but the mean of the two numbers has the lowest probability of being substantially wrong. The mean, rather than the new or old number should be used for ration formulation. Users should collate feed composition data using a spreadsheet or some other method and recalculate the mean and SD as new data are collected. If you can come up with a logical reason why composition changed (i.e., a new population), then the new number should replace the old number and you start the process of collating data again.

As mentioned several times above, the SD is an important statistic. It is an indicator of how wrong you could be. In Table 1, corn gluten feed had a mean CP concentration of 23.8 and a SD of 5.7. Assuming a normal distribution and assuming you received totally random loads of corn gluten feed (i.e., not from a single source), approximately 16% of the loads would have a CP concentration less than 18.1% and 16% of the loads would have a CP concentration greater than 29.5%. If a particular load of corn gluten feed had 18% CP and you used the mean concentration and corn gluten feed made up 10% of the diet DM, the actual CP concentration of the diet would be about 0.6% units lower than the formulated value. An error of this magnitude or larger would be expected 16 out of every 100 loads. If you are willing to accept this risk, then using the mean is the best option. However, if based on your experience, you conclude that milk production will drop 2 lbs. (or some other number) if the diet contains 0.6 percentage units less CP than formulated and you are unwilling to accept that risk (even though this will happen only 16% of the time), you need to adjust for variation. You can reduce your risk of substantially under feeding CP by *adjusting* the mean value based on its SD. Based on a normal distribution, if you use the mean minus 0.5 X SD,

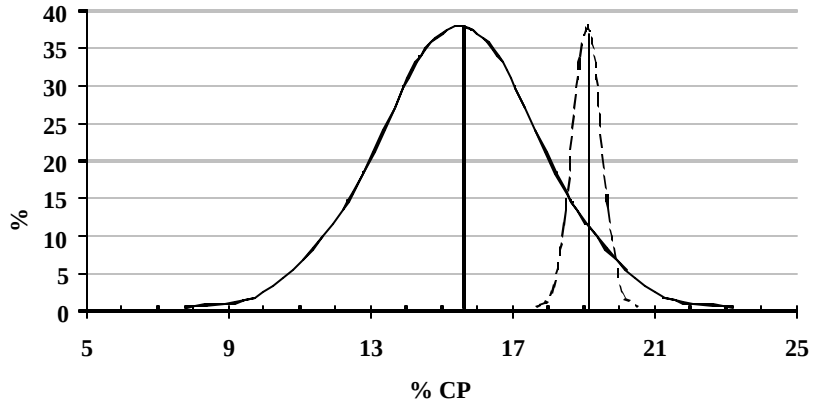

Figure 2. Distributions of crude protein (**CP**) concentrations in rice bran. The solid line represents data from a nationwide population (NRC, 2001) and the dashed line represents samples from a single source (Belyea et al., 1989). The means of the two populations are substantially different and the dispersion is much greater for the broad population. Distributions were calculated based on the mean and SD.

rather than the mean, you reduce the risk of making the error discussed above from 16% of the time to 7% of the time. If you use the mean minus 1 SD unit, you reduce the risk of making the above error to just 2% of the time. In the example above, mean CP for corn gluten feed was 23.8% (SD = 5.7). If I was willing to risk being substantially wrong 7 out of every 100 loads of corn gluten feed, I would use 23.8 - (0.5 x 5.7) or 21.0% CP for corn gluten feed when I balanced the diet. If I only wanted to be substantially wrong 2% of the time, I would use 23.8 - 5.7 = 18.1% CP. By using a lower CP concentration for corn gluten feed, I have substantially decreased the probability of being substantially deficient in CP; however, I will be over supplementing CP most of the time. You will need to determine how much risk you are willing to accept and balance that against increased feed costs.

Ways to Reduce the Impact of Variation

The composition of all feeds vary. However, the probability that all feeds in a diet will have a lower than expected concentration of a given nutrient on a given day is low. Some feeds will have higher than expected concentrations other will have lower than expected concentrations. Therefore, the variation in nutrient composition of feedstuffs is usually greater than the variation in nutrient composition of the total mixed ration (**TMR**; assuming good, standard feeding practices are in place). The impact of variation in the composition of feedstuffs is reduced as more feeds are included in diets. Relying on a particular feedstuff that is highly variable in CP concentration to provide a large proportion of dietary CP increases the risk of being wrong. If that particular feedstuff provided only 10% of the CP in the diet, a 5 percentage unit change in its CP concentration would cause dietary CP concentration to change by only 0.5 percentage units. In Figure 3,

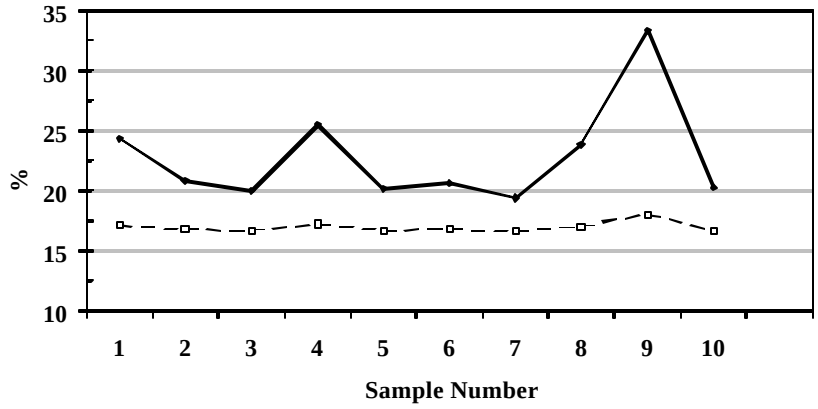

Figure 3. Crude protein concentration (**CP**; dry basis) of ten loads of corn gluten feed from a single source (solid line; DePeters et al., 2000). The dashed line represents expected concentrations of crude protein in a TMR that was balanced to contain 17% CP using the mean concentration of CP in the corn gluten feed (22.9%) and an inclusion rate of 10% of the dry matter. Note the small effect that variation in protein concentration of the ingredient has on variation in TMR protein concentration.

the concentration of CP in different loads of corn gluten feed is shown (DePeters et al., 2000). The CP concentration ranged from 19.4 to 33.4% (mean = 22.9; SD = 4.3). The load- to-load variation appears quite high. However, if the TMR was balanced for 17% CP using the mean value for corn gluten meal and the diet contained 10% corn gluten feed (DM basis) the variation in the concentration of CP is much smaller and ranged from 16.6 to 18% (Figure 3). Using a wide variety of ingredients in a TMR, and not relying too heavily on a single ingredient, is probably the best way to reduce the costs associated with variation.

Literature Cited

Belyea, R. L., B. J. Steevens, R. J. Restrepo, and A. P. Clubb. 1989. Variation in composition of by-product feeds. J. Dairy Sci. 72:2339-2345.

DePeters, E. J., J. G. Fadel, M. J. Arana, N. Ohanesian, M. A. Etchebarne, C. A. Hamilton, R. G. Hinders, M. D. Maloney, C. A. Old, T. J. Riordan, H. Perez-Monti, and J. W. Pareas. 2000. Variability in the chemical composition of seventeen selected by-product feedstuffs used by the California dairy industry. Prof. Anim. Sci. 16:69-99.

Kertz, A. F. 1998. Variability in delivery of nutrients to lactating dairy cows. J. Dairy Sci. 81:3075-3084.

National Research Council. 2001. Nutrient Requirements of Dairy Cattle. Natl. Acad. Press, Washington, DC.

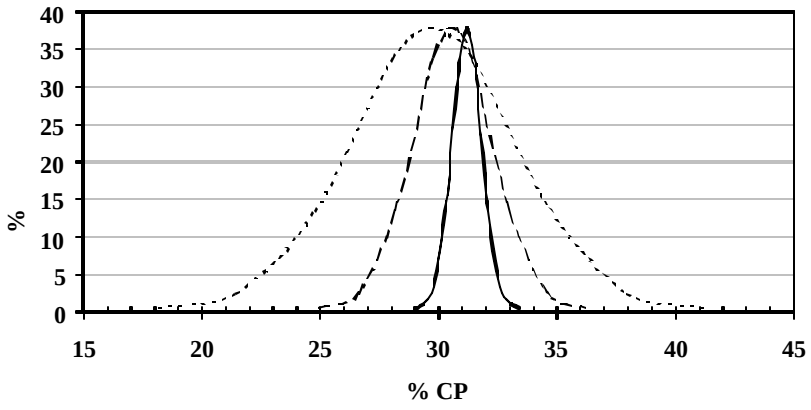

Figure 1. Distributions of crude protein (**CP**) concentrations in dried distillers grains. The small dashed line represents data from a nationwide population (NRC, 2001). The large dashed line represents samples from a single source in Missouri (Belyea et al., 1989) and the solid line represents samples from a single source in California (DePeters et al., 2000). Although mean concentrations were similar among populations, note that dispersion is substantially less for the limited populations (CA and MO) compared to the broad population (NRC). Distributions were calculated based on the mean and SD.

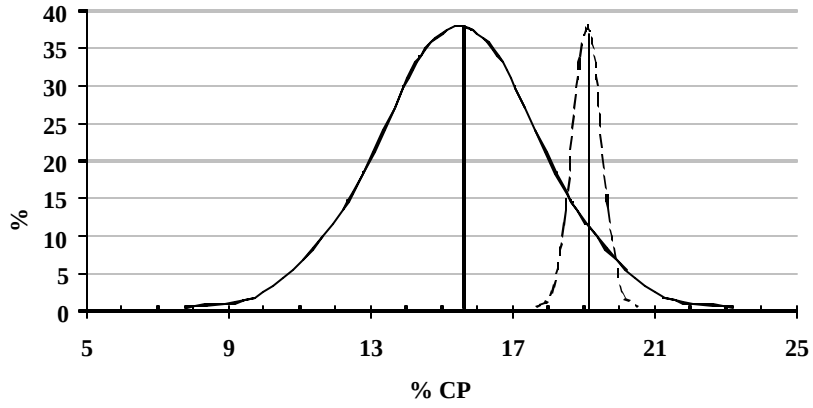

Figure 2. Distributions of crude protein (**CP**) concentrations in rice bran. The solid line represents data from a nationwide population (NRC, 2001) and the dashed line represents samples from a single source (Belyea et al., 1989). The means of the two populations are substantially different and the dispersion is much greater for the broad population. Distributions were calculated based on the mean and SD.

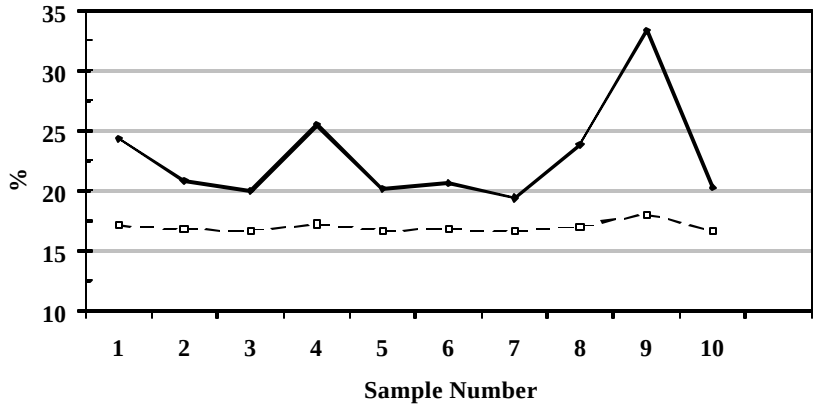

Figure 3. Crude protein concentration (**CP**; dry basis) of ten loads of corn gluten feed from a single source (solid line; DePeters et al., 2000). The dashed line represents expected concentrations of crude protein in a TMR that was balanced to contain 17% CP using the mean concentration of CP in the corn gluten feed (22.9%) and an inclusion rate of 10% of the dry matter. Note the small effect that variation in protein concentration of the ingredient has on variation in TMR protein concentration.

CP