Paper No. 983128
An ASAE Meeting Presentation

Statistical Considerations for Milking Time Tests

Steven D. LeMire, Graduate Research Assistant
Douglas J. Reinemann, Ph. D., Associate Professor
Department of Biological Systems Engineering

Graeme. A. Mein, Ph.D., Visiting Professor
Department of Dairy Science and School of Veterinary Medicine

Morten D. Rasmussen, Ph.D.
Danish Institute of Animal Sciences
Visiting Professor of Biological Systems Engineering, University of Wisconsin - Madison

University of Wisconsin-Madison, Milking Research and Instruction Lab

Written for presentation at the
1998 ASAE annual International Meeting
Orlando Florida
July 12-15, 1998

Summary: Milking parlors equipped with milk meters and electronic data collection systems offer the possibility of automated assessment of short-term milking performance. Data that can be used to assess either operator or milking machine performance includes milking time, average and peak milk flow rates, and milk production for individual cows. This paper presents considerations for measurement techniques and statistical analysis of milking time data. Sample size and statistical power consideration for means tests and experimental designs are discussed. This methodology may be valuable for troubleshooting or testing before and after installation of new milking equipment.

Keywords:
Milking Time, Statistical Tests, Statistical Power, Cow Identification

Statistical Considerations for Milking Time Tests

INTRODUCTION

Computerized data collection systems are becoming more common in milking parlors. The data can be used as a monitoring and management tool to assess how equipment or management changes affect milking performance. There is considerable variability in the milking process. This variability is introduced by differences in individual cows, differences in operator technique and work routine, differences in the properties of the milking machines, changes in weather and many other factors. Statistical analysis can be used to help determine the sources of variability and more accurately determine if a real change has taken place, and if so what caused the change. This work deals with some of the statistical consideration for doing these tests. The particular emphasis in this paper is to suggest tests that can be used to most accurately determine how changes in the milking machine affect the milking process.

The statistical tests discussed in this paper are intended primarily for short-term tests performed over one or two days. The main advantage to short-term tests is that there is less of a chance to have other factors like weather or changes in herd composition influence the factors of interest. This time restriction reduces the variability from other sources and increases the ability to detect differences correlated with changes in the milking machines.

A discussion of paired and independent means tests is presented. The assumptions, sample size and statistical power for each type of test are discussed. Examples of curves are given for the mean difference in milking time, milk weight, and average milk flow rate that is detectable for different standard deviations and sample sizes. An example of how these tests might be reported is also given.

METHODS

Statistical Tests

The operator and/or manager of a milking parlor frequently asks the question, “Did a change in the milking machine (e.g. higher or lower milking vacuum, a different type of liner, different pulsation properties, a change in personnel or a change in the work routine) make a difference. Statistical analysis can be used to assess whether a change is greater than or less than the expected variability in daily milking measures. Several broad factors influence the ability to detect differences: type of statistical tests, the sample size, and the variability in the data.

The type of tests discussed here are means tests, in which we will be making claims about the difference in the average of one set of observations compared with the average of another set of observations. When we claim that the means are different, we are claiming that the data from the two groups are from different populations. Two types of errors can occur when making these claims. A false positive is referred to as a Type I error. In this case, we claim that the two means are different when they are not. The probability of a false positive is denoted by α . A false negative is referred to as a Type II error. In this case, we claim that the two means are the same when they are not. The probability of a false negative is denoted by β . One minus the probability of a false negative (β) is referred to as statistical power. Statistical power can be defined as the probability of finding a given difference if the difference exists.

Commonly used levels of error probability in scientific literature are a false positive probability (α) of 0.05 and a false negative probability (β) of 0.20. The weight of evidence, given in terms of probability, for rejecting the hypothesis of no difference is the level of significance (or p-value).Lyman [1] The smaller the p-value, the heavier the weight of the sample evidence for rejecting the hypothesis of no difference and concluding that the means are different.

In the case of milking time tests the operator may adjust these values if there is more or less concern with false positive or false negative claims. For example, choosing a false positive probability (α) of 0.05 implies that the operator would be wrong 5 percent of the time when concluding that a change had occurred if no mean changed had occurred. With a false negative probability (β) of 0.20, the operator would conclude nothing had changed when something had changed 20 percent of the time. If the operator does not want to be bothered needlessly the false positive probability (α) and the false negative probability (β) can be reduced.

Sample Size
One form of the equation to determine the sample size for a paired means test is shown below.

$$n_{cows} = \frac{(Z_{\alpha/2} + Z_{\beta})^2 \; S_{diff}^2}{\Delta^2} \qquad \textit{Sample Size for Paired tests (Lyman [1])}$$

$$\begin{aligned} \textit{Where: } n_{cows} &: \textit{Number of Cows (Two observations per cow)} \\ Z_{\alpha/2} &: \textit{Z value for Two Tailed false postive probability } \alpha/2 \\ Z_{\beta} &: \textit{Z value for false negative probability } \beta \\ S_{diff} &: \textit{Standard deviation of the differences} \\ \Delta &: \textit{Difference in population means} \end{aligned}$$

Sample size is influenced by the levels selected for false positive probability (α) and false negative probability (β), the variability in the data, and magnitude of the difference in the means. The Z value for $Z_{0.05/2}$ is 1.96. The Z value for $Z_{0.20}$ is 0.84.

Since the number of cows is usually fixed on an individual farm, the above equation can be used to determine the minimum detectable difference for a given herd size. In the case of the variable of milk weight, this is the minimum difference in average milk production that would be considered to be "unusual" given the variability of the process. Reducing the variability in the data will increase the ability to detect changes in the means.

Two simple tests that could be programmed into the milking machine computer program for milking time data are independent sample and paired means tests. These tests can be done with a Student t-test or a Z test. The t-test is used when the sample size is less than about 30. For sample size above 30 the t-test approaches the Z test. For simplicity, the t-test will be discussed in this paper.

In the examples given in this paper, the experimental unit is an identified individual cow. The experimental unit is that to which the treatment is applied and that which is observed to collect the data. In order to take advantage of these types of statistical tests, cow identification is, therefore, highly desirable. As will be shown, the ability to collect data and draw conclusions is severely limited, if not impossible, without individual cow identification

Means Test For Independent Samples

Two assumptions of the independent sample test are that the two samples being compared are randomly drawn from independent and normal populations. Thus if we compare the average milk flow from 3 milking stalls on one side of the parlor with the average milk flow from 3 stalls on the other side of the parlor for one milking, we assume that each cow was only milked once on either side of the parlor.

An example of an independent sample test would be to compare the average milk flow rate between two stalls in a milking parlor. This might be done to detect a malfunctioning milking unit in one stall or to test the effect of changing the milking machine characteristics in one stall or group of stalls. It is assumed that any difference in allocation of cows to either stall group is independent of the responses being measured. This test would take one observation or data point for each cow and would not require that the same cow was milked in each stall.

One advantage of an independent sample test is that the test can be performed with data collected from one milking. The main disadvantage of the independent sample test is that the variability from cow to cow for milking responses can be large. This large cow to cow variability often leads to poor statistical power. One form of the independent sample test is shown below. The hypothesis is that there is no difference between the two population means. For the example above, the hypothesis tested is that there was no difference between the means in average milk flow rate between the two groups of stalls.

$$H_o: \mu_1 - \mu_2 = 0, \quad H_1: \mu_1 - \mu_2 \neq 0:, \quad t = \frac{\overline{X}_1 - \overline{X}_2}{S_{pooled}\sqrt{1/n_1 + 1/n_2}}, \quad \textit{Reject } H_o \textit{ if } |t| > t_{\alpha/2,\ (n_1+n_2-2)}$$

Where: μ_1: *Population mean for group* 1, μ_2: *Population mean for group* 2

$\overline{X}_1$: *Sample mean of group* 1, $\overline{X}_2$: *Sample mean of group* 2

n_1: *Number of cows in group* 1, n_2: *Number of cows in group* 2

S_1^2: *Sample Variance of group* 1, S_2^2: *Sample Variance of group* 2

$$S_{pooled} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

An additional assumption of the above test is that the two population variances are equal.

Paired Means Test

In the paired test discussed here, the data are paired by taking a within-cow difference between two observations from the same cow. An observation may be the average of several days of observations. For example, one could take the average flow rate from yesterday for cow 942 and the average flow for the same cow today and then take the difference. This difference then becomes one data point for the paired test. This is then done for the entire herd for the two days of milking. Taking more than one observation. from each cow reduces the variability in the data by comparing each cow to itself rather than a cow selected at random. The mean and standard deviation of the differences are used in the t-test.

The main assumption of the paired test is that the population of differences between observations is normally distributed. When a paired test takes observations from the same cow for two or more test periods, the test can be confounded with time. This means that any change caused by the milking machine cannot be separated from changes caused by other factors such as different operator or a change in weather. One way to avoid this confounding problem is to milk half the cows with one 'treatment' and the other half of the cows with the other 'treatment' in one test period. The "treatment" groups are then switched at the second test period. Cows that drop out or enter during the paired test period must be left out of the test as they do not generate 'paired' data. It must be assumed that the reason that the cows drop or enter is independent from the 'treatment'.

One form of the paired test is:

$$H_o: \mu_{diff} = 0, \quad H_o: \mu_{diff} \neq 0, \quad t = \frac{\overline{X}_{diff}}{S_{diff}/\sqrt{n}}, \quad Reject\ H_o\ if\ |t| > t_{\alpha/2,\ (n-1)}$$

Where: μ_{diff}: *Mean population difference*

$\overline{X}_{diff}$: *Sample mean difference*

S_{diff}: *Sample Standard deviation of the difference*

n: *Number of cows* (*two observations per cow*)

A paired test can be done to compare the difference of two consecutive days or between today, and an average of several previous days. The advantage of the paired test is that fewer cows are needed to find a given difference than for an independent test. This is because responses like milking time and milk weight over the short term for a given cow are usually positively correlated. This positive correlation reduces the variance of the difference of the data and improves the ability to detect mean differences.

The sample variance of the difference of two variables X and Y is:

$\mathrm{var}(X - Y) = s_x^2 + s_y^2 - 2Cov(X,Y)$ where s_x^2 *and* s_y^2 are the sample variances and the $Cov(X,Y)$ is the sample covariance for the two samples. When the correlation is zero, the sample variance of the difference reduces to: $\text{var}(X - Y) = s_x^2 + s_y^2$. The correlation coefficient (r) is related to the sample covariance as follows: $r = \dfrac{Cov(X,Y)}{\sqrt{s_x^2 s_y^2}}$.

When positive correlation between the observations is high, the variance of the difference can be lower than when they are not. The correlation coefficient may in itself be a useful measure of the uniformity of milking procedures. Herds with good milking practices can achieve correlation as high or higher than 0.9 between milk weights and milking times on consecutive days.

Figure 1 shows the effect of correlation coefficient on the standard deviation of a difference in milk yield for a paired test. Each individual milking had a variance of 80 and a standard deviation of 8.94 pounds of milk. As the correlation between cow milkings increases, the standard deviation for the differences decreases. Figure 2 shows the possible difference in mean milk weight that could be detected for a 100-cow barn with a false positive probability (α) of 0.05 and a false negative probability (β) of 0.20. The higher the cow to cow correlation, the smaller the potential effect that can be detected.

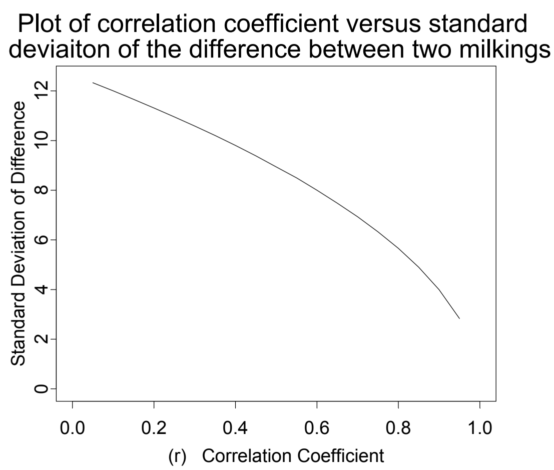

Figure 1. Plot of standard deviation of difference between two milkings and the correlation coefficient for a 100 cow farm with a variance during each milking of 80 Lbs^2.

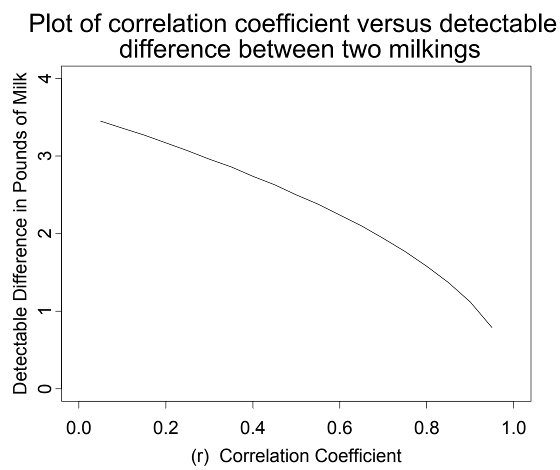

Figure 2 . Plot of detectable mean difference between two milkings and the correlation coefficient for a 100-cow farm with a variance during each milking of 80 Lbs^2. The type I error probability $\alpha = 0.05$ and type II error probability $\beta = 0.20$

Cow Identification

For the independent sample means tests, the only number that is needed is the individual milk weight for each cow. For the paired test, the cow identification number is required to be able to pair the observations. If cow identification is unreliable, the reliability of the process of trying to assess change also becomes unreliable. For example, 42 cows were paired from two days of milking. The correlation coefficient (r) was 0.86 with a standard deviation of the difference of milk weights of 7.83 pounds of milk. When 5 cows were identified incorrectly, the correlation coefficient (r) then became 0.61 with a standard deviation of the difference of 12.8 pounds of milk. In the case where the cows were all identified correctly, the detectable mean difference in milk production between the two days was 3.4 pounds. When the 5 cow numbers were identified incorrectly, the mean detectable difference increased to 5.5 pounds of milk. Both cases were calculated with a two tailed false positive probability (α) of 0.05 and a false negative probability (β) of 0.20. The loss of ability to detect changes when cows are identified incorrectly will depend on the uniformity of production responses.

Milking time
Different types of milking machines use different methods of recording the time of milking. At least one type of milk meter starts the 'milking time' measurement when the operator presses the 'attach' button before attaching the milking unit to the cow. The end of milking is taken at the time at which the decision is made to detach the milking unit from the cow. This decision is made based on some threshold value of milk flow rate.

Another type of milk meter starts the 'milking time' measurement when the milk meter first records milk flow. This is a fill and dump type meter so that the measurement chamber must be filled once to record the first milk flow. This machine ends the 'milking time' when the last dump of milk occurs and the milking unit is detached. The decision to detach is based on the time it takes to fill the measurement chamber.

These two machines thus use different methods to measure 'milking time'. The average measure of 'milking time' and the variability of this measure will differ for these two methods and they are not generally comparable. This situation illustrates the importance of understanding how the data are measured before beginning data analysis or comparison.

Milking time can be broken down into several parts as shown in Illustration 1. This additional information would allow assessment of different aspects of machine and operator performance.

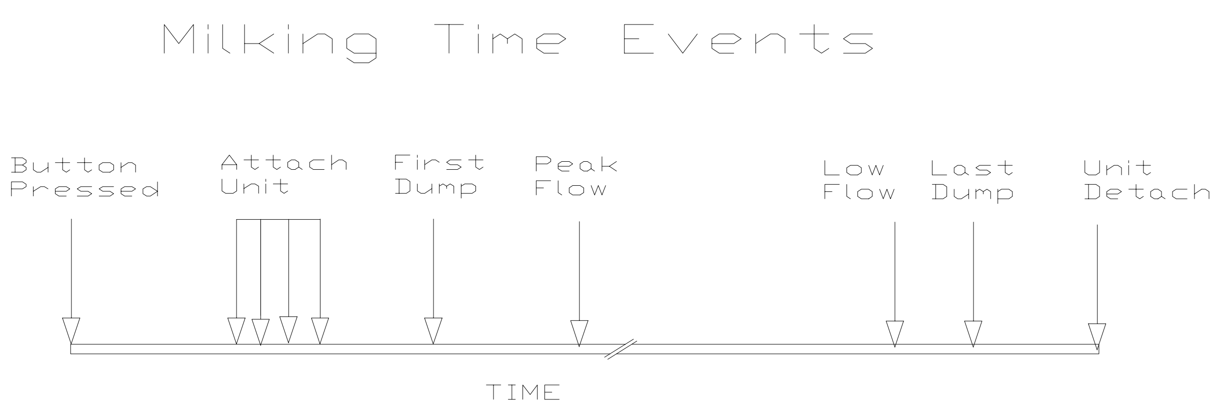

Illustration 1. Milking time events.

The first measurement interval, from the time that the 'attach' button is pressed to the time that the milking unit is attached, is a measure of the work routine of the operator. The four arrows under Attach Unit represent the attachment of each teat cup. The time from when the milking unit is attached until the first milk is recorded and/or when the peak milk flow rate is reached is a measure of the effectiveness of the preparation and stimulation procedures used by the operator. The time of low flow can be used to adjust detacher settings. Most milking machines use only one measure of milking time. This can confound operator effects with machine effects. Time intervals that should be simple to obtain on a milking machine computer are when the 'attach' button is pressed, first dump, and last dump. The time from when the 'attach' button is pressed until the first dump could be used to assess milker routine while the time from first dump to last dump is a better measure of machine efficiency.

Since milking time is used in the calculation of the average milk flow rate as well, the same difficulties occur for this measure. Separating operator from machine effects when collecting the data will reduce its variability and improve the ability to detect differences in both operator and machine effects. If these two are not separated, it will be much harder to fine-tune the milking machine for maximum efficiency.

Distribution Assumptions For Milking Time

One of the assumptions of the independent sample test is that the samples are from normal populations. For the paired test, it is assumed that the difference of the observations is normally distributed. In the case of milking time, there is evidence that the data are not normal. A number of researchers have modeled milking times as a gamma distribution. Bickert [2], Micke [3] ,Thomas [4]

A histogram of sample milking times is shown in Figure 3. A random sample from a normal distribution with the same mean and variance as the sample milking times from Figure 3 is shown in Figure 4. The distribution of sample milking times shown in Figure 3 was assumed to be distributed as a gamma distribution. The mean of this sample was 6.2. The mean of the squared observations was 42. If this distribution is assumed to be a gamma distribution, its shape parameter would then be 10.8 and its scale parameter would then be 0.57 for milking time. A simulation of 1,000,000 independent tests with a sample size of 100 per group was performed to determine the cost of assuming a normal distribution for milking time for two independent groups with these parameters. It was found that the test would be slightly less conservative with an effective probability of a Type I error of 0.052 instead of 0.050. For this setup, the cost of assuming a normal distribution for milking time is therefore not large.

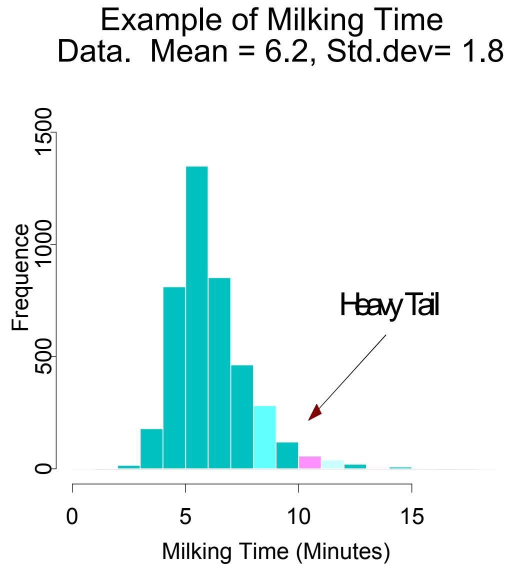

Figure3. Histogram of milking time data with mean of 6.2 and std. dev. of 1.8

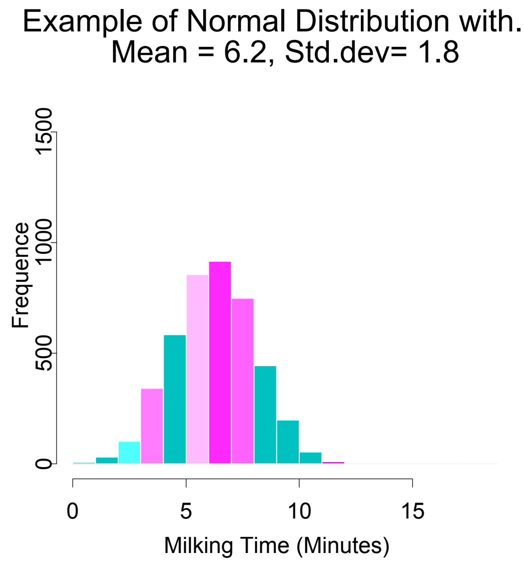

Figure 4. Histogram of a random sample from a normal distribution with mean of 6.2 and std. dev of 1.8.

For the paired test, the distribution problems for the milking time lessen. Comparisons of means from distributions with similar skewness tend to behave adequately, as the skewness cancels out. Yandell [5] Because milking times that are paired tend to be correlated, the difference between them tends to be more normally distributed so there is even less distributional concern for the paired test.

Figures 5 and 6 show the detectable difference in average milking time for paired and independent sample tests for given sample sizes and variablity in the data. A false positive probability (α) of 0.05 and false negative probability (β) of 0.20 have been assumed.

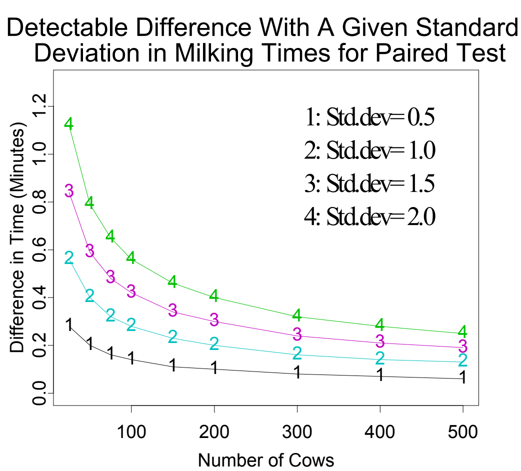

Figure 5. Detectable difference curves for paired t.test for milking time in minutes

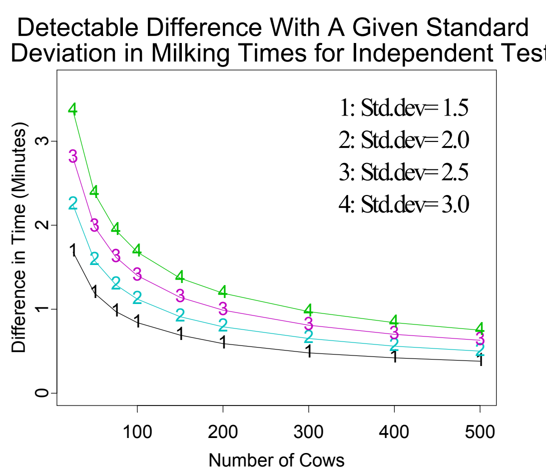

Figure 6. Detectable difference curves for independent t.test for milking time in minutes

The horizontal axis represents the total number of cows tested. In the paired test, two observations (one on each of two test periods) would be taken for each cow. In the independent test, the number of cows is divided into two equal groups and one observation is taken for each cow. For example, an independent sample test with a 400-cow farm could have 200 cows in each group receiving a different treatment. If the pooled standard deviation was 3 minutes, one could expect to find a minimum of 1 minute difference if the difference existed.

Milk Weight

The detectable difference in herd average milk weights for paired and independent sample tests are shown in Figures 7 and 8. A range of herd size and variablity in the data typical of commercial herds is given. A false positive probability (α) of 0.05 and a false negative probability (β) of 0.20 have been assumed.

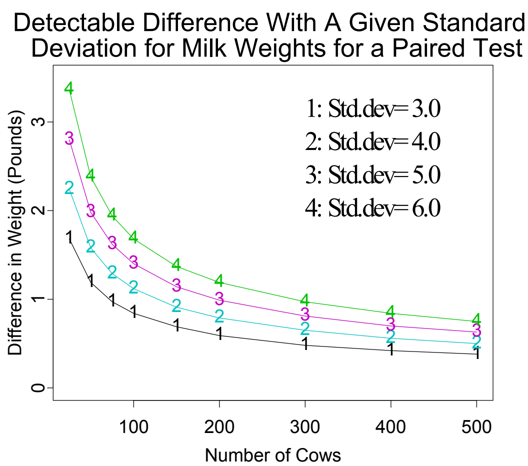

Figure 7. Detectable difference curves for paired t.test for milk weight in pounds.

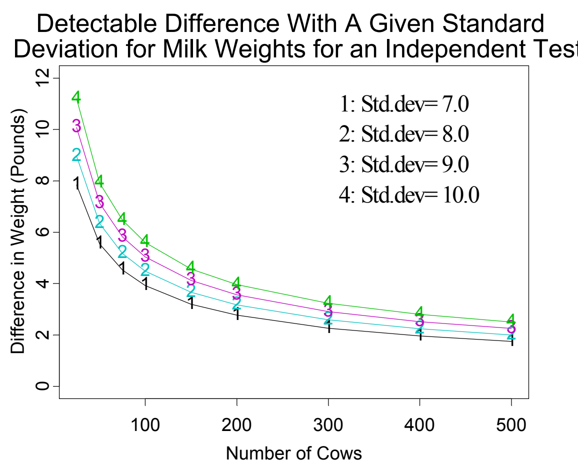

Figure 8. Detectable difference curves for independent t.test for milk weight in pounds.

Average Milk Flow Rate

The average milk flow rate may be a variable of interest to assess operator and/or milking machine performance. The detectable difference in herd average milk flow rate for paired and independent tests is shown in Figures 9 and 10. A range of herd size and variablity in the data typical of commercial herds is given. A false positive probability (α) of 0.05 and a false negative probability (β) of 0.20 have been assumed. The increased varibility associated with measuring milking time, as discussed previously, will create increased variabiltiy in measuring milk flow rate if the milking time is used to calculate milk flow rate.

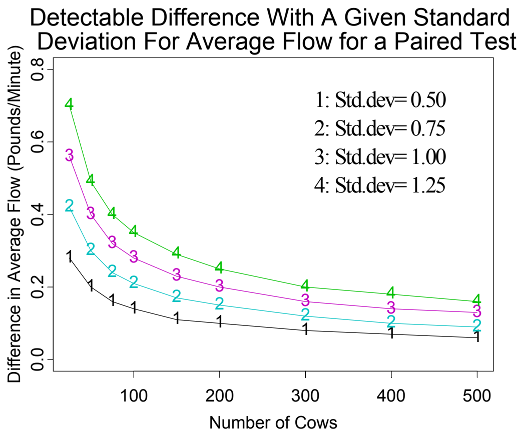

Figure 9. Detectable difference curves for paired t.test for average flow in pounds per minute.

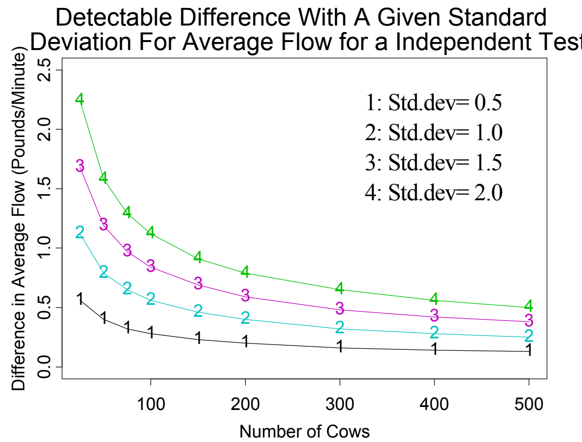

Figure 10. Detectable difference curves for independent t.test for average flow in pounds per minute.

RESULTS AND DISCUSSION

These simple statistical tests can be used periodically, e.g., to determine if a different inflation will improve milking performance. Alternately, statistical tests could be performed on a daily basis to determine if a change has occurred in the milking process. Additional tests could then be performed to help determine the cause for the change.

An example of how statistical tests might be incorporated into a milking report is shown in Table 1. The production for today as the average pounds harvested per cow is reported in Row 4. The average production per cow for the two days prior to today is reported in Row 5. The results of a two tailed paired t-test with false positive probability (α =0.05) on the difference in production are reported in Row 6. A summary of the statistical test is given in Row 7. A two-tailed, paired t-test was performed; e.g., the difference is computed for each cow and the average of the differences reported. The two-tailed test assumes that the difference may be either positive or negative. As was shown above, the paired test is able to detect smaller differences than an independent test. The mean of the difference between the two periods is –0.8 Lbs./cow/day. The standard deviation for the differences for each cow for the two periods is 5.1 Lbs./cow/day. The detectable difference is given as 1.5 Lbs./cow/day. This detectable difference is calculated from the standard deviation of the difference and selected probability error levels as discussed previously. The p-value is the probabilistic evidence for concluding that there was a difference in the means. In this case if one concluded that the mean milk production was statistically different, they would be wrong 33 percent of the time.

Therefore, from the Example Production Summary in Table 1, the dairy manager would have evidence to conclude that the mean change in production from the average of the previous two days versus today would be within the normal variation. Such an analysis could help the dairy manager determine if a real change had occurred or if the change was within the normal variation in the milking process.

Table 1. Example Production Summary

1	No. of Cows Milked Today	87
2	Total Production Today (Lbs)	6107
3	Total Production, Previous 2 days average, (Lbs)	6177
4	Average Production Today (Lbs./cow/day)	70.2
5	Average Production, Previous 2 days average, (Lbs./cow/day)	71.0
6	Is the difference significant? (α =0.05)	**NO**
7	Statistical Summary	
	Difference (Lbs./cow/day)	-0.8
	Standard Deviation of Difference.	5.1
	Detectable Mean Difference	1.5
	p value	0.33
8	Average Milking Time	5.9
9	Average Days In Milk	167.8

REFERENCES

[1] Lyman, Ott. 1988. An Introduction to Statistical Methods and Data Analysis. PWS-Kent publishing Company. Boston.

[2] Bickert, W.G.,J.B. Gerrish, and D.V. Armstrong. 1972. Semi-automatic milking in a polygon parlor: a simulation. Trans. Am. Soc. Agric. Eng. 15(2):2355.

[3] Micke, C. F., and R.D. Appleman. 1973. Simulating herringbone and side-opening milking parlor operations. J. Dairy Sci. 56:1063.

[4] C.V. Thomas, M. A. Delorenzo, and D.R. Bray. 1993. Prediction of Individual Cow Milking Time for Milking Parlor Simulation Models. J Dairy Sci. 76:2184-2194.

[5] Yandell, Brian S. 1997. Practical Data Analysis for Designed Experiments. Chapman Hall, New York