Summer 2020

Identifying Probability Distributions of Key Variables in Sow Herds

Hao Tong
Iowa State University

Identifying Probability Distributions of Key Variables in Sow Herds

Hao Tong, Daniel C. L. Linhares, and Alejandro Ramirez
Iowa State University, College of Veterinary Medicine, Veterinary Diagnostic and Production Animal Medicine Department, Ames, Iowa, USA

Abstract

Figuring out what probability distribution your data fits is critical for data analysis as statistical assumptions must be met for specific tests used to compare data. However, most studies with swine populations seldom report information about the distribution of the data. In most cases, sow farm production data are treated as having a normal distribution even when they are not. We conducted this study to describe the most common probability distributions in sow herd production data to help provide guidance for future data analysis. In this study, weekly production data from January 2017 to June 2019 were included involving 47 different sow farms. We evaluated 14 variables and report descriptive statistics, including mean, median, standard deviation, range, minimum, and maximum. Variables were also analyzed to identify goodness of fit for selected distributions. Goodness of fit was identified based on the p-value of previously validated test results. A total of 15 distributions and 2 transformations were tested. Any p-value larger than the pre-set α level (0.05) was accepted which means the variable fits the candidate probability distribution. The result demonstrated that when data from all farms were combined, for variables farrowing rate (p-value=0.110) and average stillborn (p-value=0.179), the Johnson Transformation was the best approach to then fit a normal distribution. All the other variables did not fit any of the distributions or transformations tested. Then each farm's data was analyzed individually. Results showed that herd-level variables (i.e., not aggregated) often fit to multiple distributions. Very few variables were considered as normally distributed. However, most variables were fit to normal distribution after being transformed using the Johnson or Box-Cox transformation methods.

Keywords: sow herds, probability distribution, goodness-of-fit tests.

Introduction

Data analysis is very important in all kinds of fields including the swine industry. Furthermore, scorrect analysis of data strongly depends on the distribution it belongs.

The analysis of sow herd data commonly compares data using statistics that assumes that data is normal-distributed i.e., belonging to the normal (aka Gaussian) distribution. Allen and Stewart (1983) reported that data on age at puberty, weight at puberty, estrus cycle, gestation period, and litter size can be analyzed assuming these variables are normally distributed. The log-normal and exponential distributions have also been used by others (Pettigrew et al.,1986, de Roo, 1987).

According to the central limit theorem (Brown,1971), the normal distribution is always the first choice when analyzing data. But often it is easy to assume a normal distribution to analyze data by mistake. Often this is wrong because the sample size is not large enough, the variable is not independent, or the data follows some other distributions. So, identifying the data's probability distribution should always be the first step of the analysis.

Many tests could be used to conduct the goodness of fit test. The Kolmogorov-Smirnov test (Massey et al.,1951) and the Anderson-Darling test (Stephens et al.,1974) are the two most frequently used.

The Kolmogorov-Smirnov (K-S test) test could be used to compare a sample of data with a single candidate probability distribution (one-sample K-S test) or compare two samples to see if they belong to the same probability distributions (two-sample K-S test). Here we use the one-sample K-S test to serve as a goodness of fit test.

The Anderson-Darling (A-D test) test came from a modification of the Kolmogorov-Smirnov(K-S) test. Basically, the A-D test is much more sensitive to the tails of the distribution and is more powerful than the K-S test (Razali et al.,2011).

Based on the above, whether or not the variables in sow herd follows the normal distribution is questionable. Therefore, the present study was designed to do the goodness of fit test for 14 sow productivity variables using the A-D test and one-sample K-S test. Each variable was evaluated with the data as a whole as well as for each farm independently.

Variables were evaluated for goodness of fit to the following 15 distributions and 2 transformations:

Normal distribution is also called Gaussian distribution. It is the most common distribution used in many different fields. It represents many situations, showing that data near the means are more frequent then the data far from the mean. Many data in nature follows normal distribution, for example, the population height, weight, and longevity in a country, student's test scores in a school.

Lognormal and 3-parameter lognormal distribution are both continuous probability distribution. Its logarithm is normally distributed. To fit the lognormal distribution, it required your data only contain positive values.

The exponential and 2-parameter exponential distribution are the probability distribution of time events in a Poisson point process and it's a special case of the gamma distribution. It requires the data larger than zero.

Weibull distribution and 3-parameter Weibull are a continuous probability distribution. It's often used in survival analysis, reliability engineering, failure analysis, and many other different fields. It requires the data only contain positive values if you want to fit Weibull distribution.

Largest extreme value distribution and smallest extreme value distribution. Large extreme value and smallest extreme value distribution have a very strong relationship. These two distributions are often used to describes extreme phenomena such as high insurance losses (largest extreme value) or minimum temperature (smallest extreme value).

Gamma distribution and 3-parameter gamma. The gamma distribution is a maximum entropy probability distribution. The exponential distribution, Erlang distribution, and chi-squared distribution are special cases of the gamma distribution. And the model of the size of insurance claims (Boland et al.,2007) and rainfalls (Aksoy,2000) are the most common application of these distributions.

Logistic distribution is a continuous probability distribution that presents the normal distribution in shape but has heavier tails. It is a special case of Tukey lambda distribution. It has been used in sports modeling, finance, and physical sciences.

Loglogistic and 3-paramater loglogistic is continuous probability distribution requiring your data only contain non-negative random variables. It often used in a situation such as survival analysis, hydrology, economics where the parametric model whose rate increases at the beginning and decrease later.

Poisson distribution is a discrete probability distribution modeling the number of events occurring within a given time interval. It requires the data only contains integer values. This distribution is widely used in many fields such as chemistry, biology, management, etc.

Johnson transformation and Box-cox transformation distribution, those two are often used when the original data are not normal distribution, but we want to use the transformation to check the normality.

Materials and Methods

Study is based on the weekly production data from 47 different sow herds collected from January 2017 to June 2019. Multiple variables that were collected, from which14 commonly used in evaluating herd productivity were selected for this study.

Data used

The following 14 variables from 47 different sow herds were selected to be tested for best distribution fit. For weekly production data, the period is defined as one week.

- Percentage of late services: Weekly total services of sows mated after 7 days of weaning divided by the result of the number of first services minus the number of gilts in.
- Wean 1st service: Average days from wean to the first service for sows with the first service on that day divided by total first service weaned sows in that week.
- Sows aborted: Weekly number of sows aborted.
- Abortions: Percentage of sows aborted based on number of sows bred.
- Farrowing rate: Weekly number of sows farrowed divided by total services 16-17 weeks (~110-118 gestational period) prior.
- Average total born: Weekly total pigs born divided by total sows farrowed that week.
- Average stillborn: Weekly total stillborn pigs divided by weekly total sows farrowed.
- Average Mummified: Weekly total mummified pigs divided by weekly total sows farrowed.
- Average live born: Weekly total pig born alive divided by weekly total sows farrowed.
- Average gestation length: Average gestation length is the result of total gestation days for sows that farrowed in a week .
- Pre-wean mortality: From sows weaned on a week, total pigs cohort born alive minus total pigs weaned divided by total pig cohort born alive.
- Total pigs weaned: Number of pig weaned in a week.
- PWS (pigs weaned per sow): Total pigs wean in a day divided by total sows weaned

in a week.

- Average wean age: Sum (Wean Age * Pigs weaned) divided by count (Pigs weaned) for that week.

Statistical analyses

The data were analyzed by the goodness of fit test using the individual distribution identification of Minitab Statistical Software (Minitab,19.2) and one-sample Kolmogorov–Smirnov test (one-sample K–S test) of IBM SPSS Statistics (SPSS, 27.00).
The confidence level was set to 95%. Johnson Transformation p-value was set to 0.10 for the best fit. Box-Cox power transformation used optimal λ. All 14 variables were tested with data combined from all 47 sow farms, as well as individually for each sow farm.

How to interpret the results

A-D tests and one-sample K-S tests are hypothesis tests that determine whether your data followed a hypothesized probability distribution. The null hypothesis and alternative hypothesis are;

- H_0: The sample data follow the hypothesized distribution.

- H_1: The sample data do not follow the hypothesized distribution.

$P \leq \alpha$: The data do not follow the distribution (reject H_0)

If the p-value is less than or equal to the significance level, which means the null hypothesis is rejected and concluded that the data do not follow the distribution.

$P > \alpha$: Cannot conclude the data do not follow the distribution (Fail to reject H_0)

If the p-value is greater than the significant level, which means you fail to reject the null hypothesis and there is not enough evidence to conclude that the data do not follow the distribution. You can assume the data follow the distribution.

LRT p-value (likelihood ratio test p-value) was used to determine whether adding another parameter significantly improves the fit of the distribution. An LRT p-value that is less than 0.05 suggests that a significant improvement in fit. You need to examine the LRT p-value to determine whether the 3-parameter distribution is significantly better than the 2-parameter distribution. If LRT p-value>0.05, then reject the 3-parameter distribution and stay with 2-parameter distribution.

Results and Discussion

The goodness of fit test's result of all 14 variables combining all 47 sows herd is shown in Table 1. From the results, we could figure out that only for variable farrowing rate (p-value=0.110) and average still born (p-value=0.179), the Johnson Transformation is the overall best approach to then fit a normal distribution. All the other variables did not fit to any of the distribution we tested for.

The 14 variables tested by each sow herd individually had different results. Two sow herds were censored because the sample size was too small (less than 10 weeks of production data), making the data not eligible for fitting a distribution, resulting on 45 out of 47 (95.7%) being evaluated. Results are presented in Table 2. The count indicates the number of sow herds that fits the tested probability distributions. Eligible herds indicate the total number of sow herds which were suitable to be tested by that probability distribution. Proportion indicates the proportion of sow herds which fits the candidate distribution based on total herds (n=45). Eligible proportion indicates the proportion of sow herds quantity which fits the distribution based on the number of sow herds which were actually eligible to be tested for that distribution.

The descriptive statistics, including mean, median, standard deviation, range, minimum, and maximum of 14 variables from all 47 sow herd are shown in Table 3. The histogram of all 14 variables of all 47 sow herd is shown in Table 4.

From Table 1, all the variables could not find the best fit probability except for farrowing rate and average still born. Those two needed to be transformed using the Johnson Transformation before analyzed as normally distributed. The reason most variables did not fit any distribution could be that those sow herds come from different populations, had different health status, and had different means and standard deviations affecting each other and make an irregular pattern.

From Table 2, we noticed that some sow herd did not fit a probability distribution that did fit many of the other farms. This suggests that the individual farm situation may impact the data and change the specific distribution of a variable for that particular farm at that time. This means that even it is the same variables, different sow herd may fit different distribution. So, doing a goodness of fit test before analyzing the data is necessary.

For example, 55.56% of sow herds' average stillborn fit a normal distribution. However, Figure 1 shows the histogram of average stillborn from sow herd number 13, which is not normally distributed. From the histogram we see that compared to the normal distribution, the sample around 1.0 are unnormal higher than the other tail does. This could be an improvement that reduce the stillborn number at the right tail and pushing more values to the left, thus, the arrow point area increased. This show how individual farm situations can impact the distribution of a variable.

Another example is seen with average total born where 40% sow farms data fit a normal distribution. Figure 2 shows a plot from a farm that did not fit to the normal distribution. We could see that the area indicated to the left of the arrow has a longer tail than the same area mirrored to the where the right tail is very short. This may indicate some conditions already impacted the average total born.

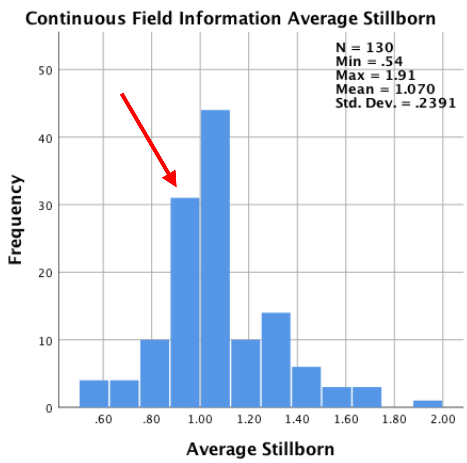


Figure 1

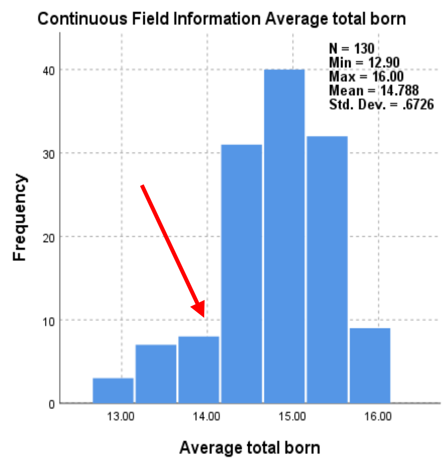


Figure 2

Conclusion

The result of data obtained form sow herds for identifying fit for probability distributions shows that the normal distribution often does not fit for the 14 variables we tested. Therefore, using a goodness of fit test to check fitness for normal distrituion before analyzing data should always be the first choice.

When variable distributions don't fit the normal distribution in the raw format, there is good chance that they still fit that distribution aftering being transformed. In this study we report that most of the 14 variables studies were successfully transformed to normal using the Johnson or Box-Cox transformation. Details from Table 2 can be very helpful as a quick guide for selecting assumptions for the type of distribution for some of the common herd production variables.

Individual farm situation may impact data and change the specific distribution of a variable for that time. Further research could focus on the question about how different herd situations could affect the data's distribution.

References

Allen, M.A. and Stewart, T.S., 1983. A simulation model for a swine breeding unit producing feeder pigs. *Agricultural Systems*, *10*(4), pp.193-211.

Pettigrew, J.E., Cornelius, S.G., Eidman, V.R. and Moser, R.L., 1986. Integration of factors affecting sow efficiency: a modeling approach. *Journal of Animal Science*, *63*(4), pp.1314-1321.

De Roo, G., 1987. A stochastic model to study breeding schemes in a small pig population. *Agricultural Systems*, *25*(1), pp.1-25.

Brown, B.M., 1971. Martingale central limit theorems. *The Annals of Mathematical Statistics*, *42*(1), pp.59-66.

Massey Jr, F.J., 1951. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*, *46*(253), pp.68-78.

Stephens, M.A., 1974. EDF statistics for goodness of fit and some comparisons. Journal of the American statistical Association, 69(347), pp.730-737.

Boland, P.J., 2007. *Statistical and probabilistic methods in actuarial science*. CRC Press.

Aksoy, H., 2000. Use of gamma distribution in hydrological analysis. Turkish Journal of Engineering and Environmental Sciences, 24(6), pp.419-428.

Razali, N.M. and Wah, Y.B., 2011. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics*, *2*(1), pp.21-33

Table 1. Goodness of fit test when combining all data from 47 farms.

whole farm distribution	Percentage of late Services	Wean 1st service	Sows aborted	Abortions	Farrowing rate	Average total born		Average Stillborn	Average Mummified
	p-value	p-value	p-value	p-value	p-value	p-value	LRT p-value	p-value	LRT p-value
Normal	<0.005	<0.005	<0.005	<0.005	<0.005	<0.005		<0.005	<0.005
Box-Cox Transformation						<0.005			
Lognormal						<0.005			
3-parameter Lognormal	*	*	*	*	*	*	0.000	*	*
Exponential						<0.003			
2-Parameter Exponential	<0.010	<0.010	<0.010	<0.010	<0.010	<0.010	0.000	<0.010	<0.010
Weibull						<0.010			
3-Parameter Weibull	<0.005	<0.005	<0.005	<0.005	<0.005	<0.005	0.000	<0.005	<0.005
Smallest Extreme Value	<0.010	<0.010	<0.010	<0.010	<0.010	<0.010		<0.010	<0.010
Largest Extreme Value	<0.010	<0.010	<0.010	<0.010	<0.010	<0.010		<0.010	<0.010
Gamma						<0.005			
3-parameter Gamma	*	*	*	*	*	*	1.000	*	*
Logistic	<0.005	<0.005	<0.005	<0.005	<0.005	<0.005		<0.005	<0.005
Loglogistic						<0.005			
3-parameter Loglogistic	*	*	*	*	*	*	0.000	*	*
Johnson Transformation					0.110			0.179	
Poisson									

LRT p-value, likelihood ratio test p-value,used to determine whether 3-parameter distribution is significant better than 2-parameter distribution(p<0.05)

(continued)

whole farm distribution	Average live born		Average gestation length		Pre-wean mortality %	Total pigs weaned	PWS	Average wean age		Average mated inventory
	p-value	LRT p-value	p-value	LRT p-value	p-value	p-value	p-value	p-value	LRT p-value	p-value
Normal	<0.005		<0.005		<0.005	<0.005	<0.005	<0.005		<0.005
Box-Cox Transformation	<0.005		<0.005					<0.005		
Lognormal	<0.005		<0.005					<0.005		
3-parameter Lognormal	*	0.000	*	0.000	*	*	*	*	0.076	*
Exponential	<0.003		<0.003					<0.003		
2-Parameter Exponential	<0.010	0.000	<0.010	0.000	<0.010	<0.010	<0.010	<0.010	0.000	<0.010
Weibull	<0.010		<0.010					<0.010		
3-Parameter Weibull	<0.005	0.000	<0.005	0.000	<0.005	<0.005	<0.005	<0.005	0.000	<0.005
Smallest Extreme Value	<0.010		<0.010		<0.010	<0.010	<0.010	<0.010		<0.010
Largest Extreme Value	<0.010		<0.010		<0.010	<0.010	<0.010	<0.010		<0.010
Gamma	<0.005		<0.005					<0.005		
3-parameter Gamma	*	1.000	*	1.000	*	*	*	*	0.000	*
Logistic	<0.005		<0.005		<0.005	<0.005	<0.005	<0.005		<0.005
Loglogistic	<0.005		<0.005					<0.005		
3-parameter Loglogistic	*	0.000	*	0.026	*	*	*	*	0.000	*
Johnson Transformation										
Poisson										

(continued)

whole farm distribution	Average gestation length		Pre-wean mortality %	Total pigs weaned	PWS	Average wean age		Average mated inventory
	p-value	LRT p-value	p-value	p-value	p-value	p-value	LRT p-value	p-value
Normal	<0.005		<0.005	<0.005	<0.005	<0.005		<0.005
Box-Cox Transformation	<0.005					<0.005		
Lognormal	<0.005					<0.005		
3-parameter Lognormal	*	0.000	*	*	*	*	0.076	*
Exponential	<0.003					<0.003		
2-Parameter Exponential	<0.010	0.000	<0.010	<0.010	<0.010	<0.010	0.000	<0.010
Weibull	<0.010					<0.010		
3-Parameter Weibull	<0.005	0.000	<0.005	<0.005	<0.005	<0.005	0.000	<0.005
Smallest Extreme Value	<0.010		<0.010	<0.010	<0.010	<0.010		<0.010
Largest Extreme Value	<0.010		<0.010	<0.010	<0.010	<0.010		<0.010
Gamma	<0.005					<0.005		
3-parameter Gamma	*	1.000	*	*	*	*	0.000	*
Logistic	<0.005		<0.005	<0.005	<0.005	<0.005		<0.005
Loglogistic	<0.005					<0.005		
3-parameter Loglogistic	*	0.026	*	*	*	*	0.000	*
Johnson Transformation								
Poisson								

Table 2 counts and proportions of the sow herds which fit to the distributions when tested individually

Distribution	Percentage of late Services				Wean 1st service			
	Count	Eligible	Proportion	Eligible Proportion	Count	Eligible	Proportion	Eligible Proportion
Normal	3	45	6.67%	6.67%	0	45	0.00%	0.00%
Lognormal	19	37	42.22%	51.35%	4	44	8.89%	9.09%
3-parameter Lognormal	0	0	0.00%	0.00%	0	0	0.00%	0.00%
Exponential	0	37	0.00%	0.00%	0	44	0.00%	0.00%
2-Parameter Exponential	0	45	0.00%	0.00%	0	45	0.00%	0.00%
Weibull	9	37	20.00%	24.32%	0	44	0.00%	0.00%
3-Parameter Weibull	9	45	20.00%	20.00%	13	45	28.89%	28.89%
Smallest Extreme Value	0	45	0.00%	0.00%	0	45	0.00%	0.00%
Largest Extreme Value	24	45	53.33%	53.33%	20	45	44.44%	44.44%
Gamma	0	37	0.00%	0.00%	2	44	4.44%	4.55%
3-parameter Gamma	0	0	0.00%	0.00%	0	0	0.00%	0.00%
Logistic	7	45	15.56%	15.56%	0	45	0.00%	0.00%
Loglogistic	23	37	51.11%	62.16%	8	44	17.78%	18.18%
3-parameter Loglogistic	0	0	0.00%	0.00%	0	0	0.00%	0.00%
Box-Cox Transformation	28	37	62.22%	75.68%	34	44	75.56%	77.27%
Johnson Transformation	41	41	91.11%	100.00%	41	41	91.11%	100.00%
Poisson	0	0	0.00%	0%	0	0	0.00%	0

(continued)

Distribution	Sows aborted				Abortions			
	Count	Eligible	Proportion	Eligible Proportion	Count	Eligible	Proportion	Eligible Proportion
Normal	0	45	0.00%	0.00%	1	45	2.22%	2.22%
Lognormal	0	4	0.00%	0.00%	2	3	4.44%	66.67%
3-parameter Lognormal	0	0	0.00%	0.00%	0	0	0.00%	0.00%
Exponential	0	4	0.00%	0.00%	0	3	0.00%	0.00%
2-Parameter Exponential	1	45	2.22%	2.22%	0	45	0.00%	0.00%
Weibull	0	4	0.00%	0.00%	1	3	2.22%	33.33%
3-Parameter Weibull	2	45	4.44%	4.44%	5	45	11.11%	11.11%
Smallest Extreme Value	0	45	0.00%	0.00%	0	45	0.00%	0.00%
Largest Extreme Value	0	45	0.00%	0.00%	1	45	2.22%	2.22%
Gamma	0	4	0.00%	0.00%	1	3	2.22%	33.33%
3-parameter Gamma	0	0	0.00%	0.00%	0	0	0.00%	0.00%
Logistic	0	45	0.00%	0.00%	0	45	0.00%	0.00%
Loglogistic	0	4	0.00%	0.00%	1	3	2.22%	33.33%
3-parameter Loglogistic	0	0	0.00%	0.00%	0	0	0.00%	0.00%
Box-Cox Transformation	1	4	2.22%	25.00%	3	3	6.67%	100.00%
Johnson Transformation	2	2	4.44%	100.00%	26	27	57.78%	96.30%
Poisson	18	0	40.00%	0.00%	0	0	0.00%	0.00%

(continued)

Distribution	Farrowing rate				Average total born			
	Count	eligible	proportion	eligible proportion	Count	eligible	proportion	eligible proportion
Normal	12	45	26.67%	26.67%	18	45	40.00%	40.00%
Lognormal	8	42	17.78%	19.05%	15	45	33.33%	33.33%
3-parameter Lognormal	0	0	0.00%	0.00%	0	0	0.00%	0.00%
Exponential	0	42	0.00%	0.00%	0	45	0.00%	0.00%
2-Parameter Exponential	0	45	0.00%	0.00%	0	45	0.00%	0.00%
Weibull	22	42	48.89%	52.38%	14	45	31.11%	31.11%
3-Parameter Weibull	13	45	28.89%	28.89%	15	45	33.33%	33.33%
Smallest Extreme Value	20	45	44.44%	44.44%	12	45	26.67%	26.67%
Largest Extreme Value	0	45	0.00%	0.00%	1	45	2.22%	2.22%
Gamma	7	42	15.56%	16.67%	13	45	28.89%	28.89%
3-parameter Gamma	0	0	0.00%		0	0	0.00%	0.00%
Logistic	2	45	4.44%	4.44%	15	45	33.33%	33.33%
Loglogistic	9	42	20.00%	21.43%	14	45	31.11%	31.11%
3-parameter Loglogistic	0	0	0.00%		0	0	0.00%	0.00%
Box-Cox Transformation	25	42	55.56%	59.52%	29	45	64.44%	64.44%
Johnson Transformation	34	35	75.56%	97.14%	29	29	64.44%	100.00%
Poisson	0	0	0.00%	0.00%	0		0.00%	0.00%

(continued)

Distribution	Average Stillborn				Average Mummified			
	Count	eligible	proportion	eligible proportion	Count	eligible	proportion	eligible proportion
Normal	25	45	55.56%	55.56%	5	45	11.11%	11.11%
Lognormal	31	44	68.89%	70.45%	16	44	35.56%	36.36%
3-parameter Lognormal	0	0	0.00%	0.00%	0	0	0.00%	0.00%
Exponential	0	44	0.00%	0.00%	0	44	0.00%	0.00%
2-Parameter Exponential	0	45	0.00%	0.00%	1	45	2.22%	2.22%
Weibull	10	44	22.22%	22.73%	4	44	8.89%	9.09%
3-Parameter Weibull	20	45	44.44%	44.44%	14	45	31.11%	31.11%
Smallest Extreme Value	1	45	2.22%	2.22%	0	45	0.00%	0.00%
Largest Extreme Value	16	45	35.56%	35.56%	14	45	31.11%	31.11%
Gamma	28	44	62.22%	63.64%	10	44	22.22%	22.73%
3-parameter Gamma	0	0	0.00%	0.00%	0	0	0.00%	0.00%
Logistic	22	45	48.89%	48.89%	4	45	8.89%	8.89%
Loglogistic	26	44	57.78%	59.09%	18	44	40.00%	40.91%
3-parameter Loglogistic	0	0	0.00%	0.00%	0	0	0.00%	0.00%
Box-Cox Transformation	38	44	84.44%	86.36%	32	44	71.11%	72.73%
Johnson Transformation	25	25	55.56%	100.00%	34	36	75.56%	94.44%
Poisson	0	0	0.00%	0.00%	0	0	0.00%	0.00%

(continued)

Distribution	Average live born				Average gestation length			
	Count	eligible	proportion	eligible proportion	Count	eligible	proportion	eligible proportion
Normal	9	45	20.00%	20.00%	5	45	11.11%	11.11%
Lognormal	10	45	22.22%	22.22%	5	45	11.11%	11.11%
3-parameter Lognormal	0	0	0.00%	0.00%	0	0	0.00%	0.00%
Exponential	0	45	0.00%	0.00%	0	45	0.00%	0.00%
2-Parameter Exponential	0	45	0.00%	0.00%	0	45	0.00%	0.00%
Weibull	12	45	26.67%	26.67%	1	45	2.22%	2.22%
3-Parameter Weibull	13	45	28.89%	28.89%	5	45	11.11%	11.11%
Smallest Extreme Value	13	45	28.89%	28.89%	1	45	2.22%	2.22%
Largest Extreme Value	1	45	2.22%	2.22%	0	45	0.00%	0.00%
Gamma	11	45	24.44%	24.44%	5	45	11.11%	11.11%
3-parameter Gamma	0	0	0.00%	0.00%	0	0	0.00%	0.00%
Logistic	7	45	15.56%	15.56%	2	45	4.44%	4.44%
Loglogistic	7	45	15.56%	15.56%	2	45	4.44%	4.44%
3-parameter Loglogistic	0	0	0.00%	0.00%	0	0	0.00%	0.00%
Box-Cox Transformation	23	45	51.11%	51.11%	5	45	11.11%	11.11%
Johnson Transformation	28	31	62.22%	90.32%	6	7	13.33%	85.71%
Poisson	0	0	0.00%	0.00%	0	0	0.00%	0.00%

(continued)

Distribution	Pre-wean mortality %				Total pigs weaned			
	Count	eligible	proportion	eligible proportion	Count	eligible	proportion	eligible proportion
Normal	7	45	15.56%	15.56%	3	45	6.67%	6.67%
Lognormal	16	43	35.56%	37.21%	1	35	2.22%	2.86%
3-parameter Lognormal	0	0	0.00%	0.00%	0	0	0.00%	0.00%
Exponential	0	43	0.00%	0.00%	0	35	0.00%	0.00%
2-Parameter Exponential	0	45	0.00%	0.00%	0	45	0.00%	0.00%
Weibull	0	43	0.00%	0.00%	1	35	2.22%	2.86%
3-Parameter Weibull	13	45	28.89%	28.89%	2	45	4.44%	4.44%
Smallest Extreme Value	0	45	0.00%	0.00%	2	45	4.44%	4.44%
Largest Extreme Value	18	45	40.00%	40.00%	0	45	0.00%	0.00%
Gamma	12	43	26.67%	27.91%	2	35	4.44%	5.71%
3-parameter Gamma	0	0	0.00%	0.00%	0	0	0.00%	0.00%
Logistic	8	45	17.78%	17.78%	5	45	11.11%	11.11%
Loglogistic	17	43	37.78%	39.53%	3	35	6.67%	8.57%
3-parameter Loglogistic	0	0	0.00%	0.00%	0	0	0.00%	0.00%
Box-Cox Transformation	34	43	75.56%	79.07%	7	35	15.56%	20.00%
Johnson Transformation	36	37	80.00%	97.30%	16	16	35.56%	100.00%
Poisson	0	0	0.00%	0.00%	0	45	0.00%	0.00%

(continued)

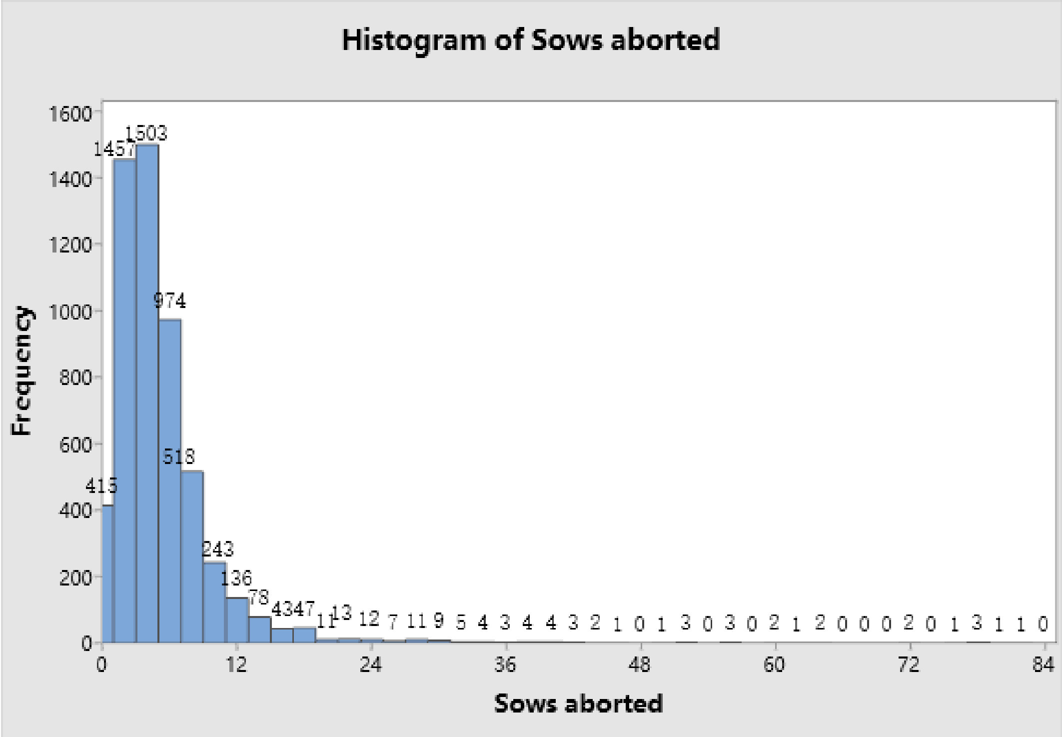
Distribution	PWS				Average wean age			
	Count	eligible	proportion	eligible proportion	Count	eligible	proportion	eligible proportion
Normal	6	45	13.33%	13.33%	20	45	44.44%	44.44%
Lognormal	5	35	11.11%	14.29%	21	45	46.67%	46.67%
3-parameter Lognormal	0	0	0.00%	0.00%	0	0	0.00%	0.00%
Exponential	0	35	0.00%	0.00%	0	45	0.00%	0.00%
2-Parameter Exponential	0	45	0.00%	0.00%	0	45	0.00%	0.00%
Weibull	5	35	11.11%	14.29%	10	45	22.22%	22.22%
3-Parameter Weibull	8	45	17.78%	17.78%	21	45	46.67%	46.67%
Smallest Extreme Value	4	45	8.89%	8.89%	4	45	8.89%	8.89%
Largest Extreme Value	0	45	0.00%	0.00%	6	45	13.33%	13.33%
Gamma	5	35	11.11%	14.29%	18	45	40.00%	40.00%
3-parameter Gamma	5	0	11.11%	0.00%	0	0	0.00%	0.00%
Logistic	5	45	11.11%	11.11%	21	45	46.67%	46.67%
Loglogistic	0	35	0.00%	0.00%	23	45	51.11%	51.11%
3-parameter Loglogistic	0	0	0.00%	0.00%	0	0	0.00%	0.00%
Box-Cox Transformation	13	35	28.89%	37.14%	33	45	73.33%	73.33%
Johnson Transformation	23	23	51.11%	100.00%	0	24	0.00%	0.00%
Poisson	0	0	0.00%	0.00%	0	0	0.00%	0.00%

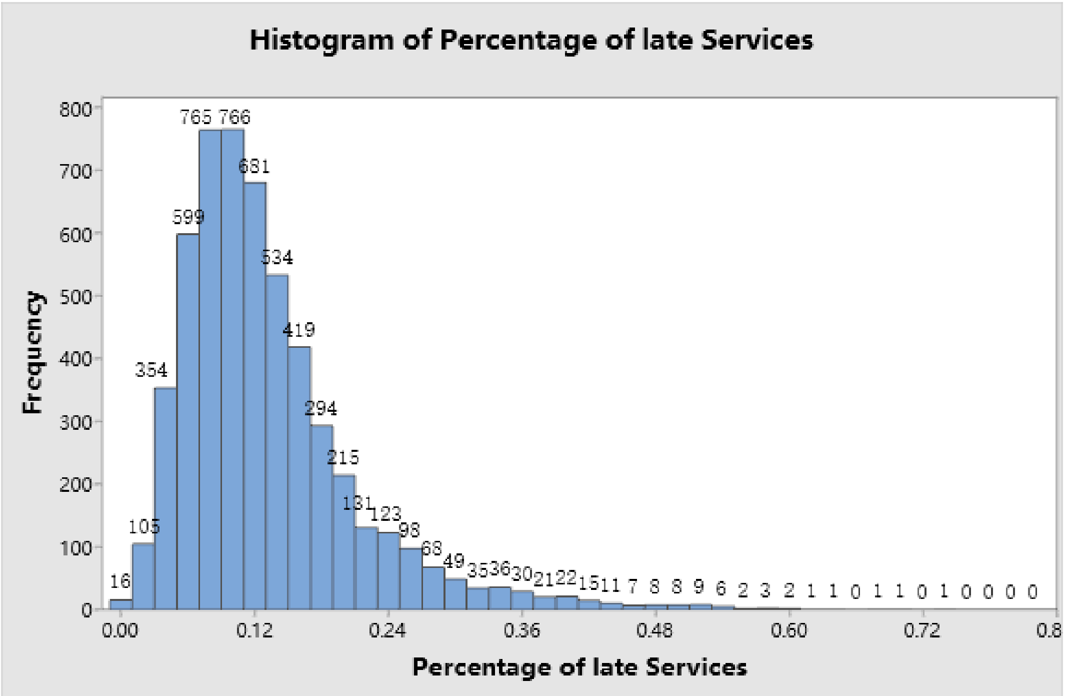
Table 3 descriptive statistics of all weekly sow herd data

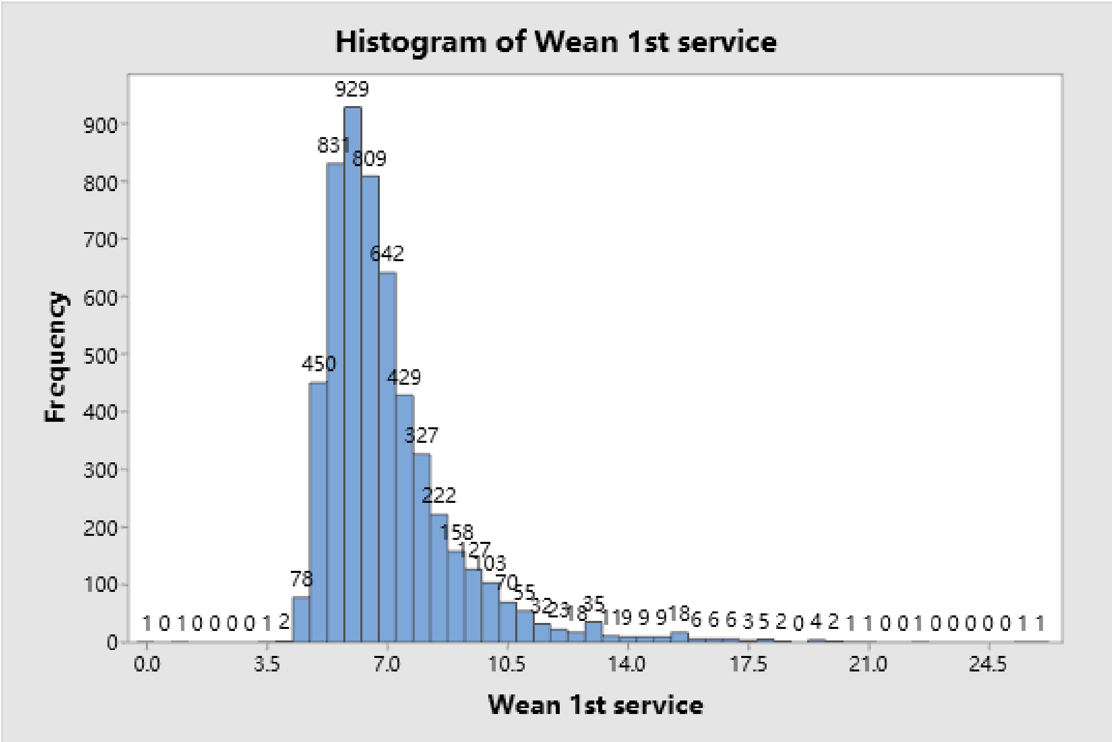
		Percentage of late Services	Wean 1st service	Sows aborted	Abortions	Farrowing rate	Average total born	Average Stillborn
Number of data points	Valid	5438	5438	5531	5576	5441	5495	5495
	Missing	169	169	76	31	166	112	112
Mean		13.188%	7.00	5.1	3%	84.426%	14.89	1.08
Median		11.359%	6.50	4.0	2%	85.500%	14.90	1.06
Std. Deviation		8.325%	2.00	9.3	7%	6.566%	0.67	0.29
Range		100.000%	26.10	366.0	293%	100.000%	7.60	4.50
Minimum		0.000%	0.00	0.0	0%	0.000%	9.30	0.00
Maximum		100.000%	26.10	366.0	293%	100.000%	16.90	4.50

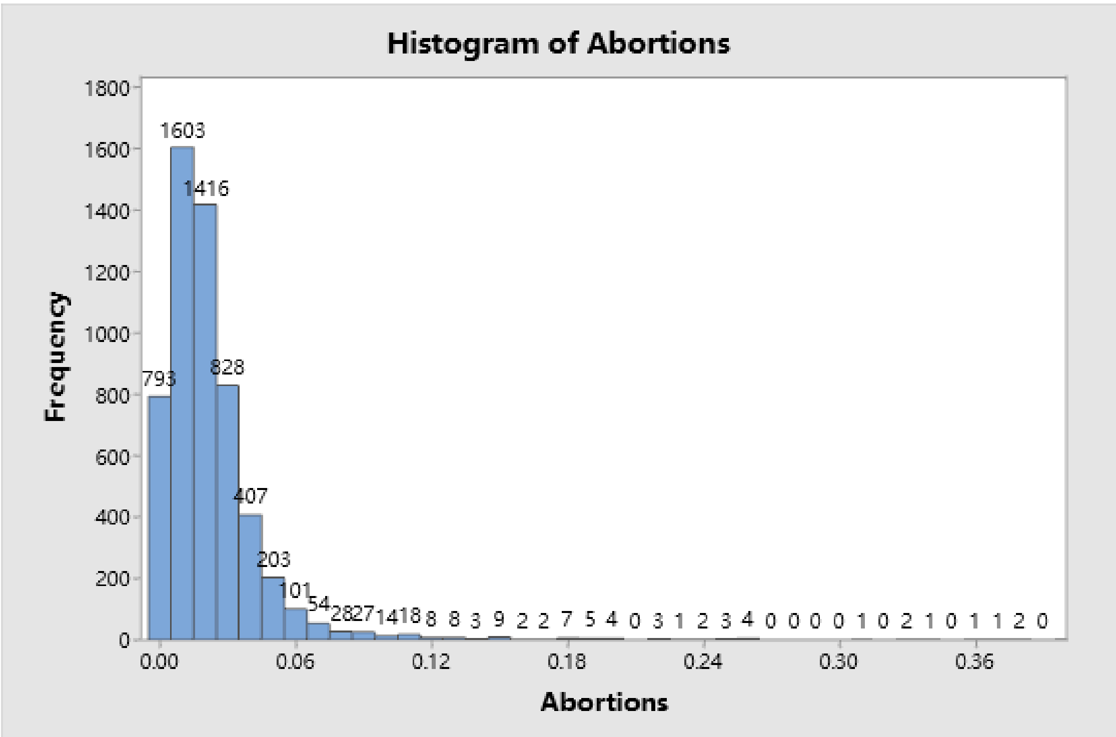
		Average Mummified	Average live born	Average gestation length	Pre-wean mortality %	Total pigs weaned	PWS	Average wean age
Number of data points	Valid	5495	5495	5495	5363	5494	5494	5458
	Missing	112	112	112	244	113	113	149
Mean		0.59	13.22	116.8	14.324%	2260.9	10.87	17.6
Median		0.50	13.30	116.8	13.799%	2188.0	11.10	17.4
Std. Deviation		0.39	0.79	0.4	4.591%	708.3	1.27	2.2
Range		4.94	8.30	7.6	100.000%	6166.0	13.00	21.3
Minimum		0.00	6.80	110.5	0.000%	0.0	0.00	7.0
Maximum		4.94	15.10	118.1	100.000%	6166.0	13.00	28.3

Table 4 histograms of all 14 variables of all 47 sow herd shows

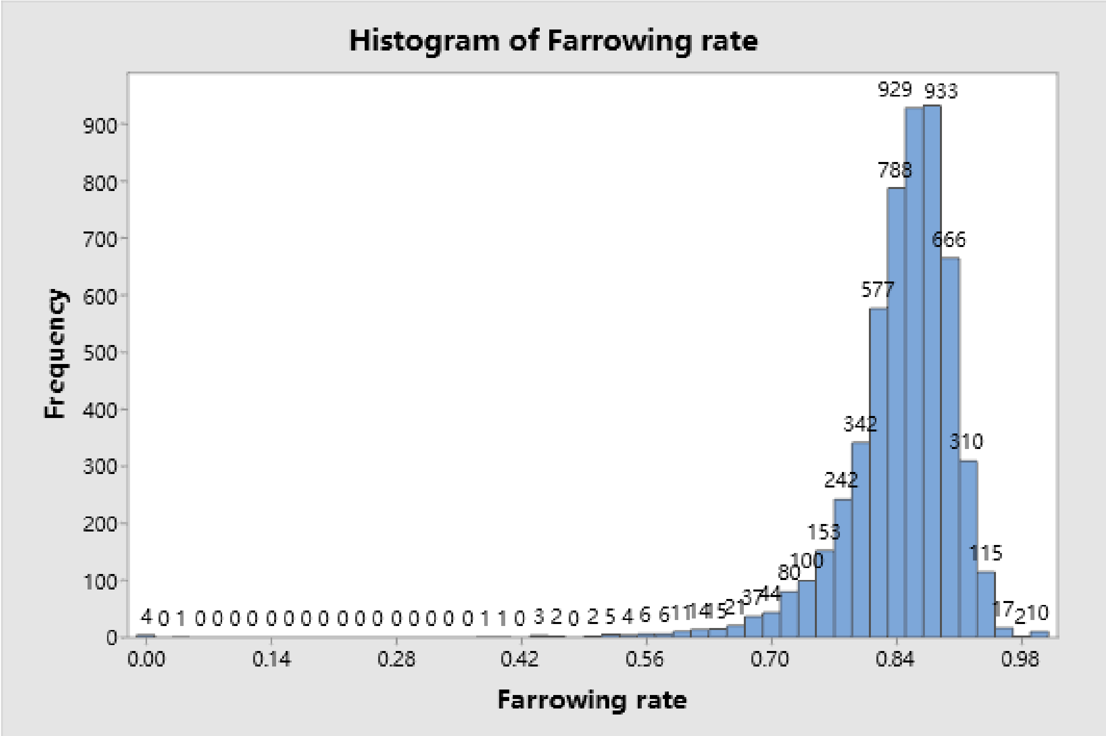


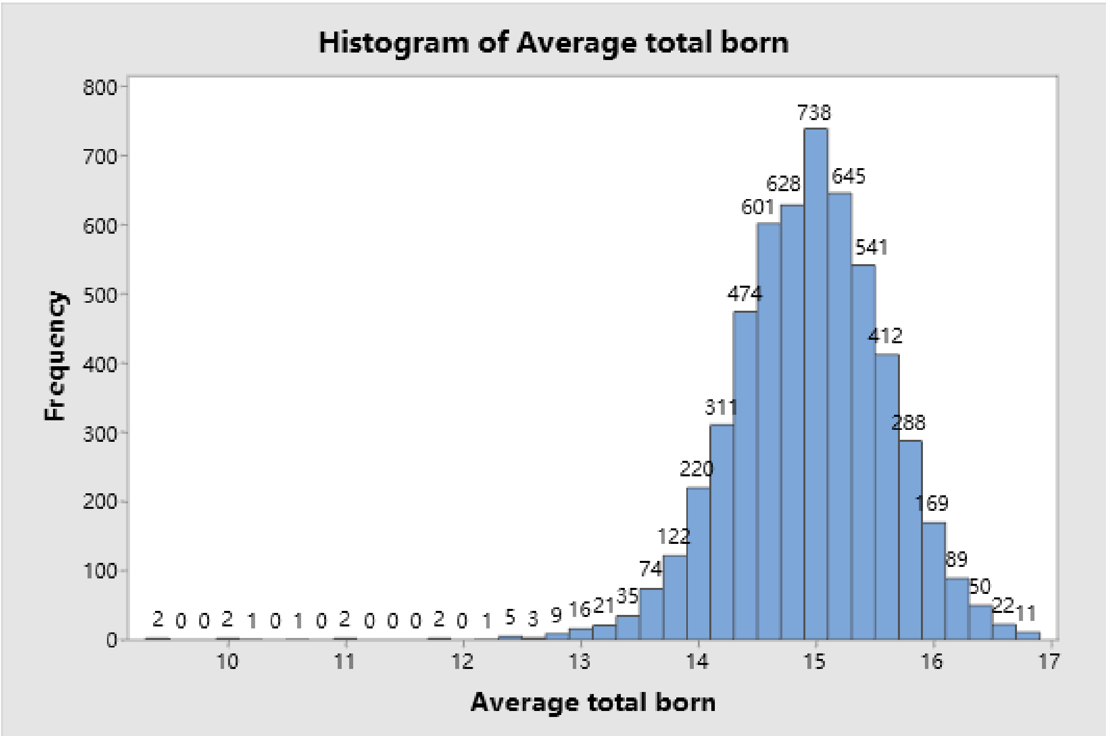


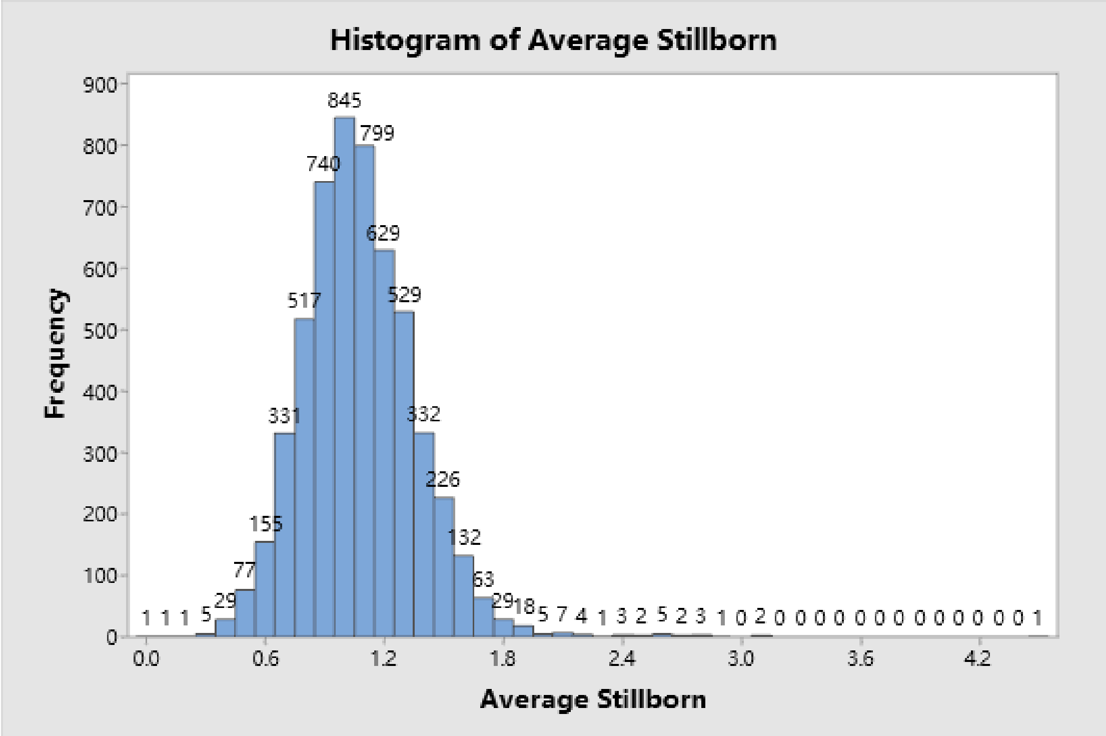


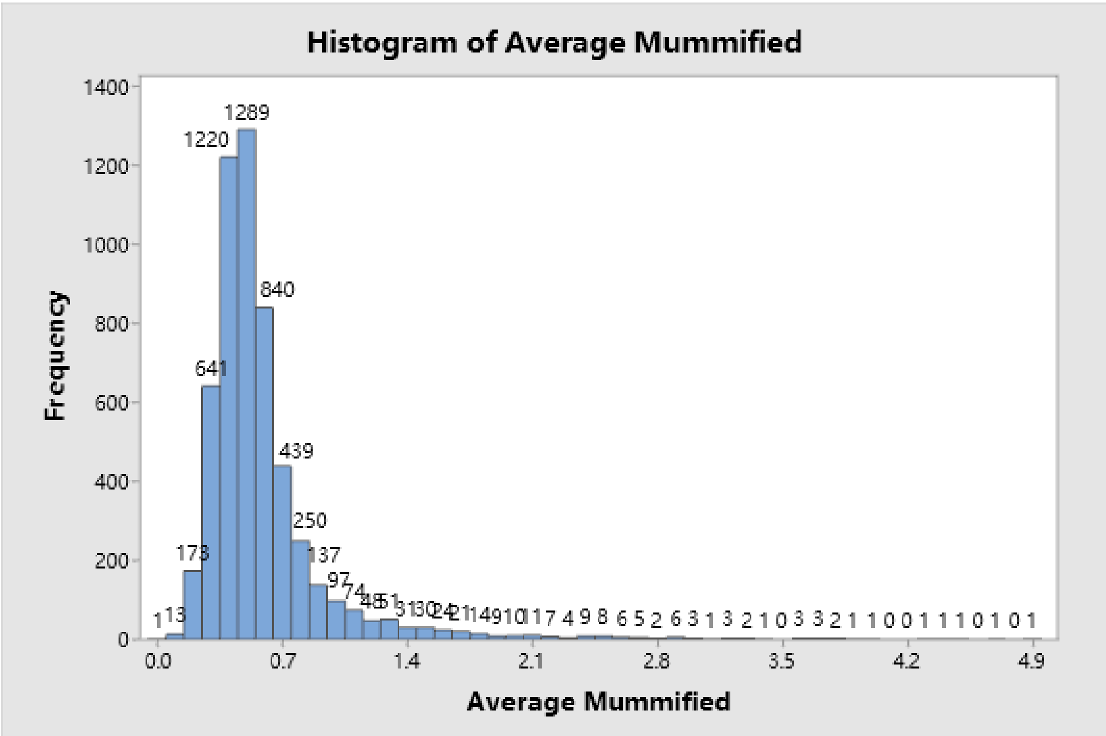


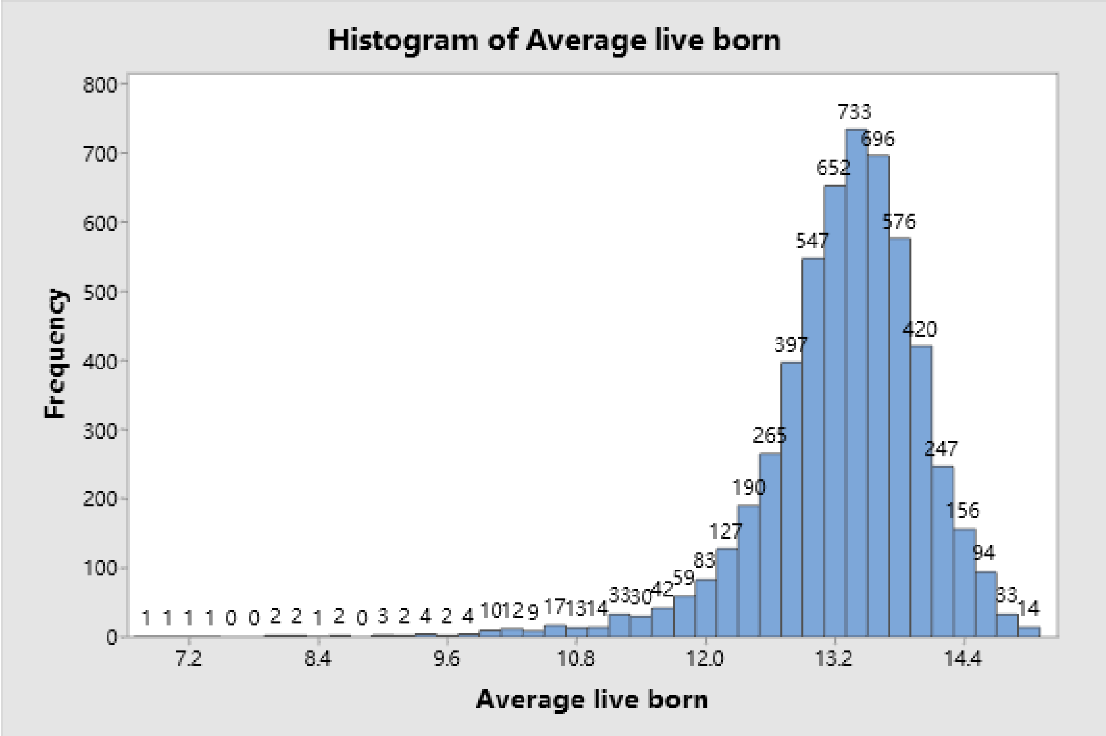
(The data of abortion which larger than 0.4 had been censored to make the graph more reasonable)

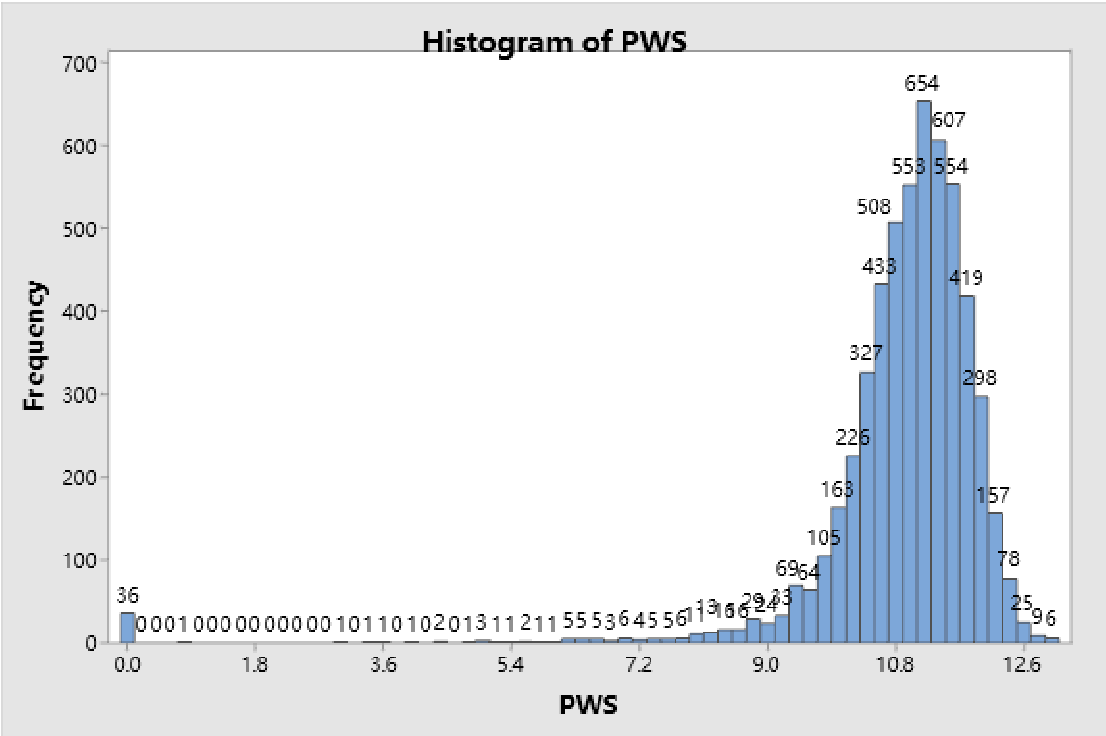

Histogram of Farrowing rate
Frequency
Farrowing rate

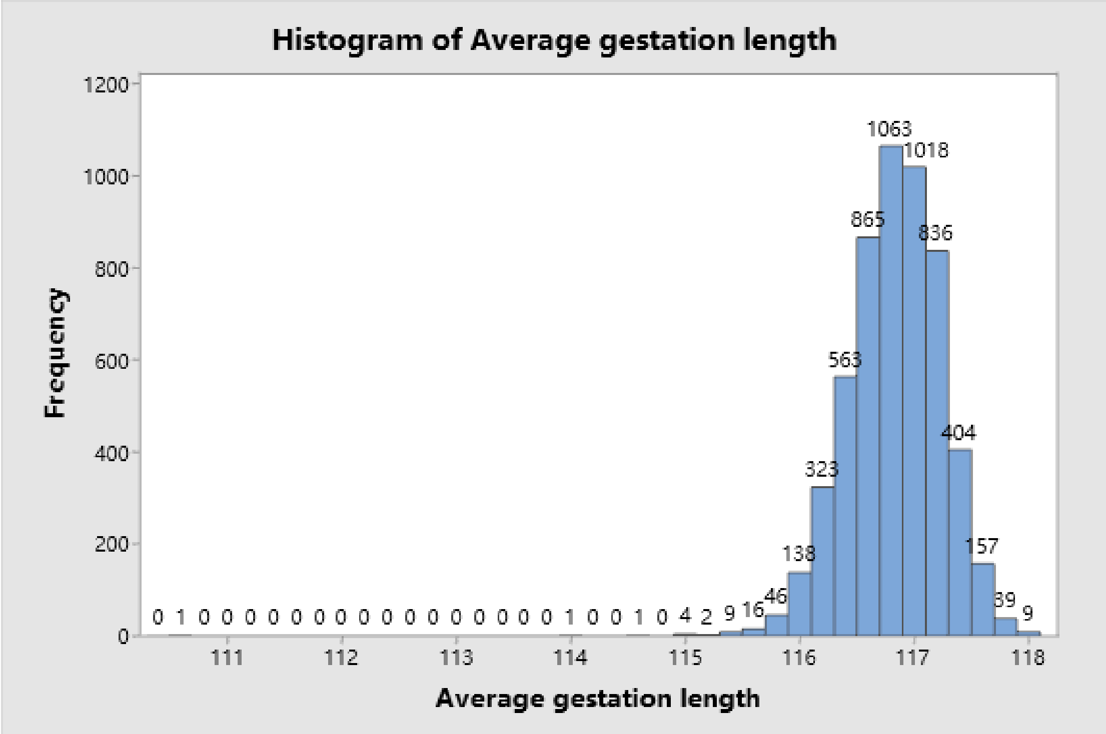

Histogram of Average total born
Frequency
Average total born

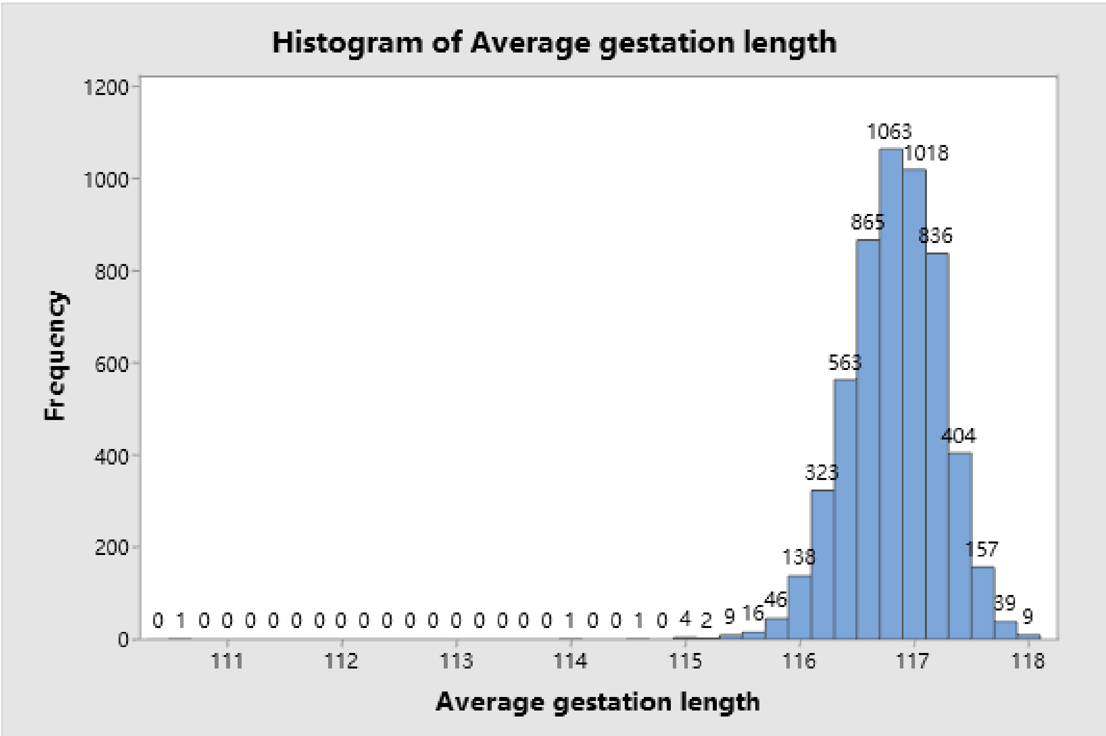

Histogram of Average Stillborn
Frequency
Average Stillborn
0
100
200
300
400
500
600
700
800
900
0.0
0.6
1.2
1.8
2.4
3.0
3.6
4.2

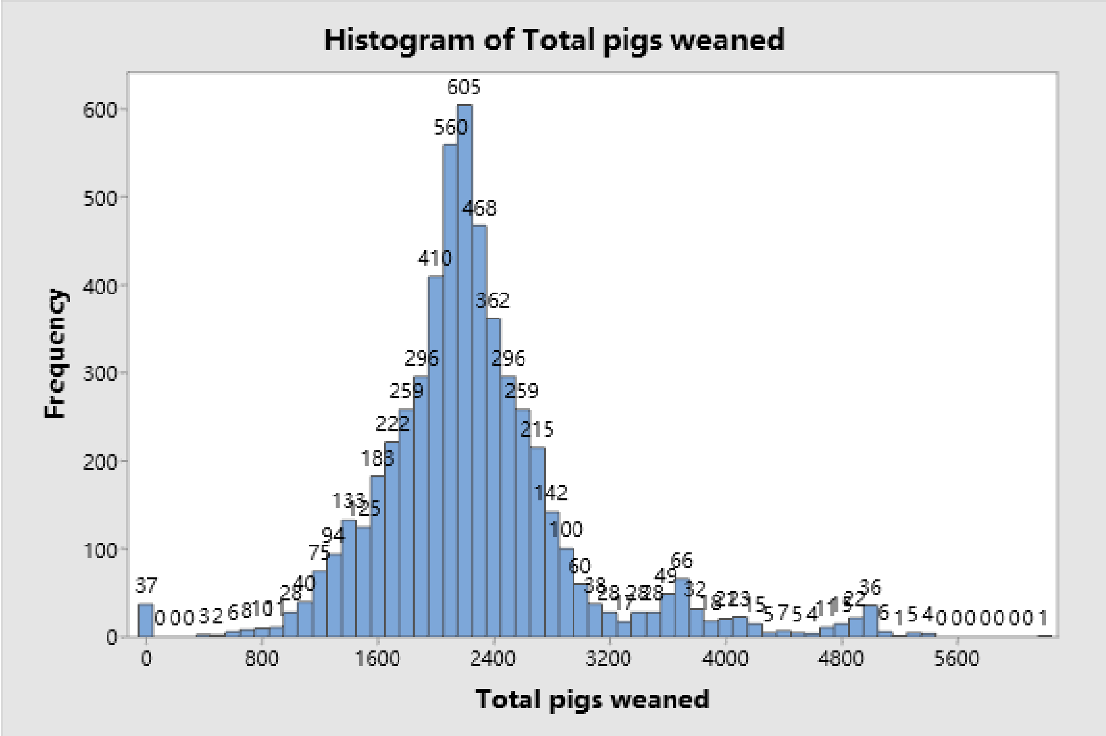

Histogram of Average Mummified
Frequency
Average Mummified
0
200
400
600
800
1000
1200
1400
0.0
0.7
1.4
2.1
2.8
3.5
4.2
4.9


Histogram of Average live born
Frequency
Average live born


Histogram of PWS
Frequency
PWS


Histogram of Average gestation length
Frequency
Average gestation length
0 200 400 600 800 1000 1200
111 112 113 114 115 116 117 118
0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 4 2 9 16 46 138 323 563 865 1063 1018 836 404 157 39 9


Histogram of Average gestation length
Frequency
Average gestation length
0 200 400 600 800 1000 1200
111 112 113 114 115 116 117 118
0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 4 2 9 16 46 138 323 563 865 1063 1018 836 404 157 39 9


Histogram of Total pigs weaned
Frequency
Total pigs weaned
0
800
1600
2400
3200
4000
4800
5600
100
200
300
400
500
600
605
560
468
410
362
296
259
222
183
142
100
60
66
37