

# Effect of Sampling Method on the Accuracy and Precision of Estimating the Mean Pig Weight of the Population<sup>1,2</sup>

*C. B. Paulk, G. L. Highland<sup>3</sup>, M. D. Tokach, J. L. Nelssen, S. S. Dritz<sup>4</sup>, R. D. Goodband, J. M. DeRouchey, and K. D. Haydon<sup>5</sup>*

## Summary

Producers have adopted marketing strategies such as topping to help reduce economic losses from weight discounts at the processing plant. Despite adopting these strategies, producers are still missing target weights and incurring discounts. One contributing factor is the error of sampling methods that producers use to estimate the mean weight of the population to determine the optimal time to top pigs. The standard sample size that has been adopted by many producers is 30 pigs. Our objective was to determine the best method for selecting 30 pigs to improve the accuracy and precision of estimating the mean pig weight of the population. Using a computer program developed in R (R Foundation for Statistical Computing, Vienna, Austria), we were able to generate 10,000 sample means for different sampling procedures on 3 different datasets. Using this program we evaluated taking: (1) a completely random sample of 30 pigs from the barn, (2) a varying number of pigs per pen to achieve a total sample size of 30 pigs, (3) selecting the heaviest and lightest pig (determined visually) from 15 pens and calculating the mean from those pigs, and (4) calculating the median of the selected pigs.

Among the 3 datasets, taking a completely random sample of 30 pigs from the barn resulted in a range between the upper and lower confidence interval as high as 23 lb. Increasing the number of pens sampled while keeping the sample size constant reduced the range between the upper and lower confidence interval; however, the confidence interval (range where 95% of weight estimates would fall) was still as high as 24 lb (241 to 265 lb) when only 30 pigs were sampled. Although the range was reduced, it was not enough to make increasing the number of pens sampled a practical means of estimating mean pig weight of the barn. Selecting the heaviest and lightest pigs in 15 pens and taking the mean of the sample resulted in a reduction of the range between the upper and lower confidence interval from 31 to 53%. Although the precision of the sample was improved, accuracy of the sampling method decreased, with the mean of the 10,000 simulations up to 8 lb lighter than the mean of the population.

Selecting the heaviest and lightest pigs can be a valuable method for improving the precision in estimating the mean of the population, but adjustments to the sampling procedure need to be developed to improve its accuracy.

---

<sup>1</sup> The authors wish to thank Elanco Animal Health, Greenfield, IN, for providing technical and financial support.

<sup>2</sup> The authors wish to thank Dr. Jason Kelly and Suidae Animal Health and Production, Algona, IA, for providing technical support and access to commercial swine facilities.

<sup>3</sup> Department of Statistics, College of Arts and Sciences, Kansas State University.

<sup>4</sup> Department of Diagnostic Medicine/Pathobiology, College of Veterinary Medicine, Kansas State University.

<sup>5</sup> Elanco Animal Health, Greenfield, IN.

Key words: finishing pig, mean weight estimation, sample size

## Introduction

Swine producers must meet processing plant requirements for specific weights of pigs as well as weight ranges to avoid economic penalties. In attempts to reduce these economic penalties, producers have adopted marketing practices such as topping or marketing the heaviest pigs several weeks before the expected barn closeout. Because pig BW typically approximates a normal distribution, subsampling methods to predict the average weight of pigs in the barn can be used to model distributions of BW within the barn. The standard sample size that has been adopted by many producers is 30 pigs. Previous data from Kansas State University reported that for a set sample size, increasing the number of pens sampled could reduce the error in estimating the mean pig weight of the population (Paulk et al., 2011<sup>6</sup>). To maximize economic return when marketing pigs, the precision of sampling pigs needs further improvement; therefore, our objective was to determine the best method of selecting 30 pigs to improve the accuracy and precision of estimating the mean pig weight of the population.

## Procedures

A total of 3 datasets (A, B, and C) were used to evaluate sampling method on the accuracy and precision of estimating the pig mean weight in the barn. The first sampling method tested was a completely random sample of 30 pigs from the barn, disregarding pen arrangements. The second sampling method tested compared the number of pigs (1 to 30 pigs) sampled from an increasing number of pens to achieve a total sample size of 30 pigs. The third and fourth sampling methods tested consisted of selecting the heaviest and lightest pig (determined visually) from 15 pens (30 pigs total) and calculating the mean and median of the selected pigs, respectively.

Dataset A was derived from Groesbeck et al. (2007<sup>7</sup>). Dataset A (Figure 1) comprised a total of 1,260 pigs in 48 pens with 23 to 28 pigs per pen. The mean, median, standard deviation and CV of the population were 253.0 lb, 254 lb, 32.8 lb, and 13.0%, respectively. Datasets B and C were obtained for the purposes of this experiment. Dataset B was obtained from a commercial finishing site in northern Iowa. Pigs (PIC C42 × PIC 359) weighed for Dataset B were from a single barn that was classified as healthy by the attending veterinarian. The barn was filled with pigs over a 1-wk period, and pigs were gate cut as they came off the truck to randomly place them in pens. Dataset B (Figure 2) contained a total of 1,261 pigs weighed (population mean = 213.5 lb, median = 214 lb, standard deviation = 21.5 lb, and CV = 10.1%) and housed in 19 pens with 56 to 81 pigs per pen. Dataset C was derived from a different commercial site in northern Iowa that consisted of pigs (Genetiporc F25 × G performer boar) that were weaned during a porcine reproductive and respiratory syndrome (PRRS) outbreak at the sow farm. The barn was filled with pigs over a 1-wk period, and pigs were gate cut as they came off the truck. Dataset C (Figure 3) comprised a total of 1,069 pigs weighed (population mean = 222.4 lb, median = 224 lb, standard deviation = 32.0 lb, and CV = 14.4%) from 40 pens with 20 to 35 pigs per pen.

<sup>6</sup> Paulk et al., Swine Day 2011. Report of Progress 1056, pp. 308–318.

<sup>7</sup> Groesbeck, G. N., G. Armbuster, M. D. Tokach, R. D. Goodband, J. M. DeRouchey, and J. L. Nelssen. 2007. Influence of Pulmotil, Tylan, and Paylean on pig growth performance and weight variation. American Association of Swine Veterinarians Proceedings, pp. 235–238.

A program was coded using R (R Foundation for Statistical Computing, Vienna, Austria) to demonstrate the error associated with varying sampling methods when estimating the mean weight of the population. For the first sampling method, the program was designed to take a completely random sample of the designated sample size, disregarding pen arrangements, and calculate the mean of this sample. The program conducted this sampling technique 10,000 times, generating 10,000 sample means. The 10,000 sample means were sorted from least to greatest, and a 95% confidence interval (CI) was generated by selecting the 9,751<sup>st</sup> observation (upper CI) and the 250<sup>th</sup> observation (lower CI). The distances between the upper and lower CI represent the range of the mean estimations. A similar analysis was conducted using R for the remaining sampling methods. For sampling methods 3 and 4, marketers provided by Suidae Health and Production, Algona, IA, were used to select the heaviest and lightest pigs in each pen. One marketer, marketer 1, was provided for Dataset B and two marketers, marketers 2 and 3, were provided for Dataset C. The percentages of accurately selected pigs for each dataset are presented in Table 1. Selection accuracy was incorporated into sampling methods 3 and 4 for Dataset A based on the selection accuracy of the 2 marketers from Dataset C. The probability for selecting the 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>, 4<sup>th</sup>, or 5<sup>th</sup> heaviest pig was 50, 25, 15, 5, and 5%, respectively, and the probability for selecting the 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>, 4<sup>th</sup>, or 5<sup>th</sup> lightest pig was 70, 15, 5, 5, and 5%, respectively. These were chosen because dataset A and C had similar pen arrangements. To account for selection accuracy in the simulations, a rank was assigned to the heaviest and lightest pig selected by the marketer in each pen. Next, these were combined into a list for both groups of selected pigs, the heaviest and lightest pigs. For each pen selected, a rank was randomly selected; therefore, for Dataset A, if the 1<sup>st</sup> pen randomly selected were pen 8, one pig selected from pen 8 would have a 50, 25, 15, 5, and 5% chance of being either the 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>, 4<sup>th</sup>, or 5<sup>th</sup> heaviest pig, and the other pig selected would have a 70, 15, 5, 5, and 5% chance of being either the 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>, 4<sup>th</sup>, or 5<sup>th</sup> lightest pig, respectively.

## Results and Discussion

Notably, random samples were generated using a computer program and samples taken from the barn are not truly random unless pigs are individually identified and preselected, rather than selected by the marketer.

When asked to identify the heaviest pig in the pen, marketers 1, 2, and 3 identified the heaviest pig in 47.4, 43.5, and 55.0% of the pens and the 2<sup>nd</sup> heaviest pig in 5.3, 35.0, and 25.0% of the pens, respectively (Figures 2, 3, and 4; Table 1). The pigs identified by marketers 1, 2, and 3 were within the actual 5 heaviest pigs in 68, 100, and 95% of the pens, respectively. When asked to select the lightest pig, marketers 1, 2, and 3 identified the lightest pig in 57.9, 75.0, and 68.4% of the pens and the 2<sup>nd</sup> lightest pig in 21.1, 17.5, and 10.5% of the pens, respectively (Figures 2, 3, and 4; Table 1). The pigs identified by marketers 1, 2, and 3 were within the actual 5 heaviest pigs in 79.5, 100, and 100% of the pens, respectively.

When taking a completely random sample of 30 pigs from datasets A, B, and C, the range between the upper and lower CI was 23.0, 15.0, and 22.5 lb, respectively. For Datasets A and C, when sampling 15 pigs from 2 pens, the estimated range between the upper and lower CI was 32.0 and 47.8 lb, respectively, but when sampling 1 pig from 30 pens the ranges between the upper and lower CI were 23.1 and 20.3 lb, respec-

tively (Table 2). For Dataset B, when sampling 30 pigs from 1 pen, the estimated range between the upper and lower CI was 38.3 lb, but when sampling 2 pigs from 15 pens, the range between the upper and lower CI was 14.8 lb; therefore, increasing the number of pens used to sample 30 pigs can improve the range between the upper and lower CI by 28, 61, and 58% in Datasets A, B, and C, respectively.

Selecting the heaviest and lightest pigs in 15 pens and taking the mean of the sample resulted in a reduction of the range between the upper and lower CI from 31 to 53%, but because specific pigs were selected, bias was introduced into the sampling procedure. This bias resulted in increased systematic error or reduced accuracy, with the mean of the 10,000 simulations being less than the actual mean of the perspective population. When pigs were selected based on the estimated selection (Dataset A), marketer 1 (Dataset B), marketer 2 (Dataset C), and marketer 3 (Dataset C), the means of the 10,000 simulations were 245.0, 207.7, 219.8, 221.8, respectively, whereas the actual means of Datasets A, B, and C were 253.0, 213.5, 222.4 lb, respectively. The deviation in accuracy of the mean can be influenced by the shape of the population distribution and the accuracy of the marketer when selecting both the heaviest and lightest pigs. Taking the median of the selected pigs did not further improve the range between the upper and lower 95% CI.

Sample size, method, variation, and distribution of pigs within a barn can substantially affect the precision of estimating the mean weight of all pigs in the barn. It is important for producers to take this into consideration when weighing pigs prior to topping to make marketing decisions. Calculating the mean of the selected heaviest and lightest pigs in each pen can improve the precision of estimating the mean; however, adjustments to the sampling method need to be determined to improve its accuracy.

**Table 1. The percentage of the selected pigs as the actual n heaviest or lightest pig<sup>1</sup>**

	Rank of pigs					
	1	2	3	4	5	>5
<b>Heaviest<sup>2</sup></b>						
Dataset B marketer 1, %	47.4	5.3	0.0	5.3	10.5	31.5
Dataset C marketer 2, %	42.5	35.0	10.0	7.5	5.0	0.0
Dataset C marketer 3, %	55.0	25.0	10.0	2.5	2.5	5.0
<b>Lightest<sup>3</sup></b>						
Dataset B marketer 1, %	57.9	21.1	10.5	0.0	0.0	10.5
Dataset C marketer 2, %	75.0	17.5	5.0	2.5	0.0	0.0
Dataset C marketer 3, %	68.4	10.5	7.9	5.3	7.9	0.0

<sup>1</sup> Marketers were asked to select the heaviest and lightest pig in each pen in the barn.

<sup>2</sup> 1 is the heaviest pig; 5 is the 5<sup>th</sup> heaviest pig.

<sup>3</sup> 1 is the lightest pig; 5 is the 5<sup>th</sup> lightest pig.

**Table 2. The resulting mean, upper 95% confidence interval (CI), lower 95% CI, and range for the various sampling methods to give a total sample size of 30 pigs**

Sampling method	Mean of 10,000 simulations	Upper CI	Lower CI	Range
Dataset A <sup>1</sup>				
Method 1, 30 random pigs <sup>2</sup>	253.0	264.2	241.2	23.0
Method 2 <sup>3</sup>				
15 pigs from 2 pens	253.2	268.6	236.6	32.0
10 pigs from 3 pens	253.1	267.1	238.4	28.8
6 pigs from 5 pens	253.1	266.0	239.4	26.6
5 pigs from 6 pens	253.0	265.6	239.7	26.0
3 pigs from 10 pens	253.1	265.2	240.7	24.6
2 pigs from 15 pens	253.1	264.7	241.2	23.5
1 pig from 30 pens	253.0	264.3	241.2	23.1
Method 3, mean <sup>4</sup>	245.0	252.4	237.7	14.7
Method 4, median <sup>5</sup>	251.6	263.5	240.0	23.5
Dataset B <sup>6</sup>				
Method 1, 30 random pigs <sup>2</sup>	213.4	220.8	205.8	15.0
Method 2 <sup>3</sup>				
15 pigs from 2 pens	213.5	224.7	186.4	38.3
10 pigs from 3 pens	213.5	223.9	197.6	26.3
6 pigs from 5 pens	213.6	223.3	201.3	22
5 pigs from 6 pens	213.6	222.5	203.9	18.6
3 pigs from 10 pens	213.5	222	204.6	17.4
2 pigs from 15 pens	213.6	221.3	205.6	15.7
1 pig from 30 pens	213.5	220.9	206.1	14.8
Method 3, mean <sup>4</sup>	207.7	211.3	204.4	6.9
Method 4, median <sup>5</sup>	207.9	218.0	194.5	23.5

*continued*

**Table 2. The resulting mean, upper 95% confidence interval (CI), lower 95% CI, and range for the various sampling methods to give a total sample size of 30 pigs**

Sampling method	Mean of 10,000 simulations	Upper CI	Lower CI	Range
Dataset C <sup>7</sup>				
Method 1, 30 random pigs <sup>2</sup>	222.3	233.3	210.8	22.5
Method 2 <sup>3</sup>				
15 pigs from 2 pens	223.0	244.6	196.8	47.8
10 pigs from 3 pens	223.1	242.0	201.5	40.4
6 pigs from 5 pens	223.0	238.6	205.9	32.7
5 pigs from 6 pens	223.0	237.5	207.6	30.0
3 pigs from 10 pens	223.0	235.5	209.8	25.7
2 pigs from 15 pens	223.0	234.3	211.3	22.9
1 pig from 30 pens	223.0	233.1	212.8	20.3
Method 3, mean <sup>4</sup>				
Marketer 2	219.8	227.4	212.0	15.5
Marketer 3	221.8	229.6	213.8	15.8
Method 4, median <sup>5</sup>				
Marketer 2	221.0	231.0	210.5	20.5
Marketer 3	222.4	234.5	208.0	26.5

<sup>1</sup>A total of 1,260 pigs were used (mean = 253.0 lb, median = 254 lb, standard deviation = 32.8 lb, and CV = 12.98%) with 23 to 28 pigs per pen and a total of 48 pens.

<sup>2</sup>30 pigs were randomly selected from the barn.

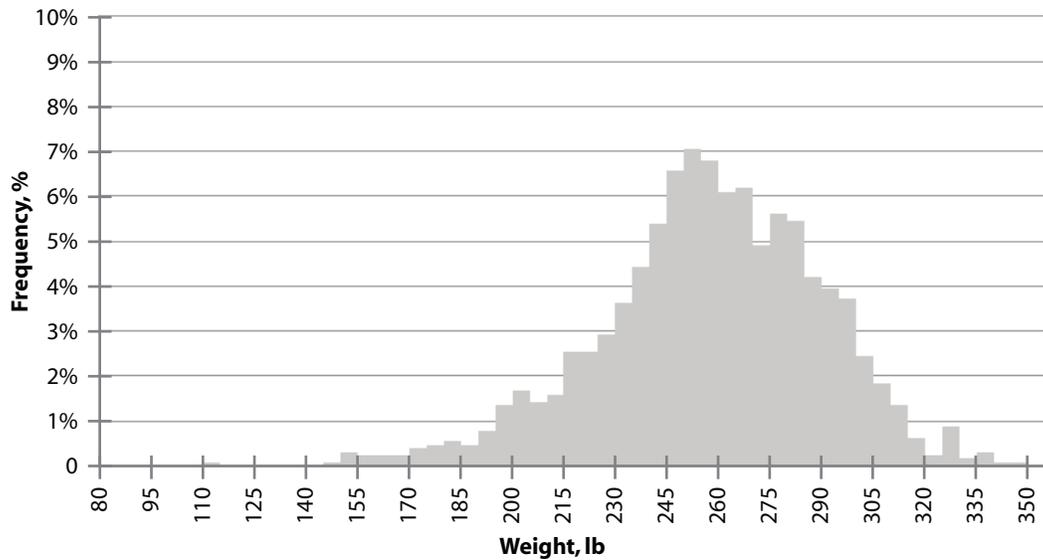
<sup>3</sup>The number of random pigs selected from the number of randomly selected pens.

<sup>4</sup>Selecting the heaviest and lightest pig (determined visually) from 15 pens and calculating the mean from those two pigs. The 15 pen means were averaged to obtain an estimated weight of the barn.

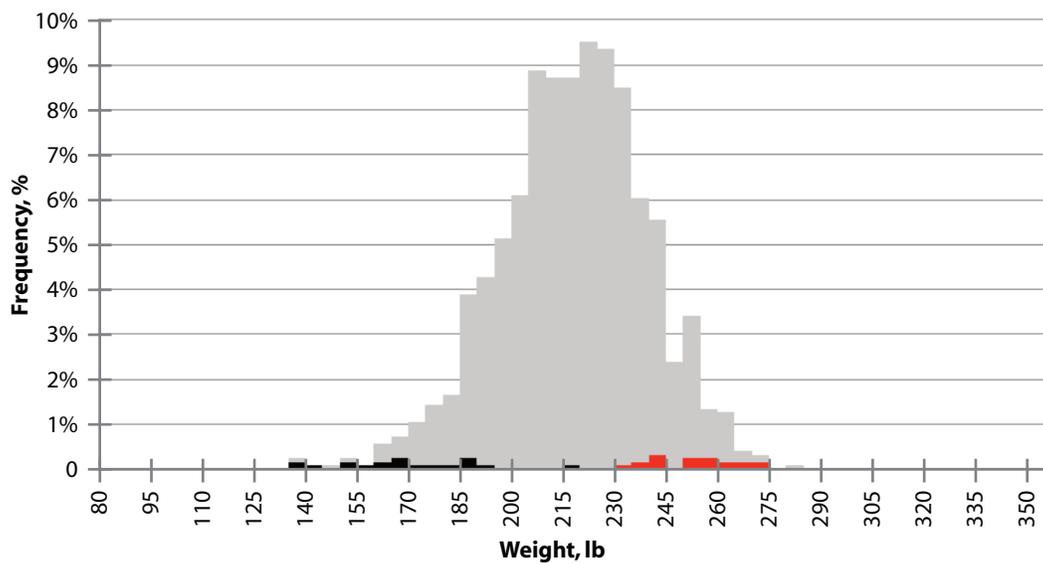
<sup>5</sup>Selecting the heaviest and lightest pig (determined visually) from 15 pens and calculating the median from those pigs

<sup>6</sup>A total of 1,261 pigs were used (population mean = 213.5 lb, median = 214 lb, standard deviation = 21.5 lb, and CV = 10.1%) with 56 to 81 pigs per pen and a total of 19 pens.

<sup>7</sup>A total of 1,069 pigs weighed (population mean = 222.4 lb, median = 224 lb, standard deviation = 32.0 lb, and CV = 14.4%) with 20 to 35 pigs per pen and a total of 40 pens.



**Figure 1. Histogram of Dataset A, a total of 1,260 pigs (mean = 253.0 lb, median = 254 lb, standard deviation = 32.8 lb, and CV = 13.0%) with 23 to 28 pigs per pen and a total of 48 pens.**



**Figure 2. Histogram of Dataset B and marketer 1's selections. A total of 1,261 pigs were weighed (population mean = 213.5 lb, median = 214 lb, standard deviation = 21.5 lb, and CV = 10.1%) with 19 pens and 56 to 81 pigs per pen. The marketer selected the heaviest and lightest pig in each pen. The 2 histograms of the marketer's selections are imposed on top of the population histogram.**

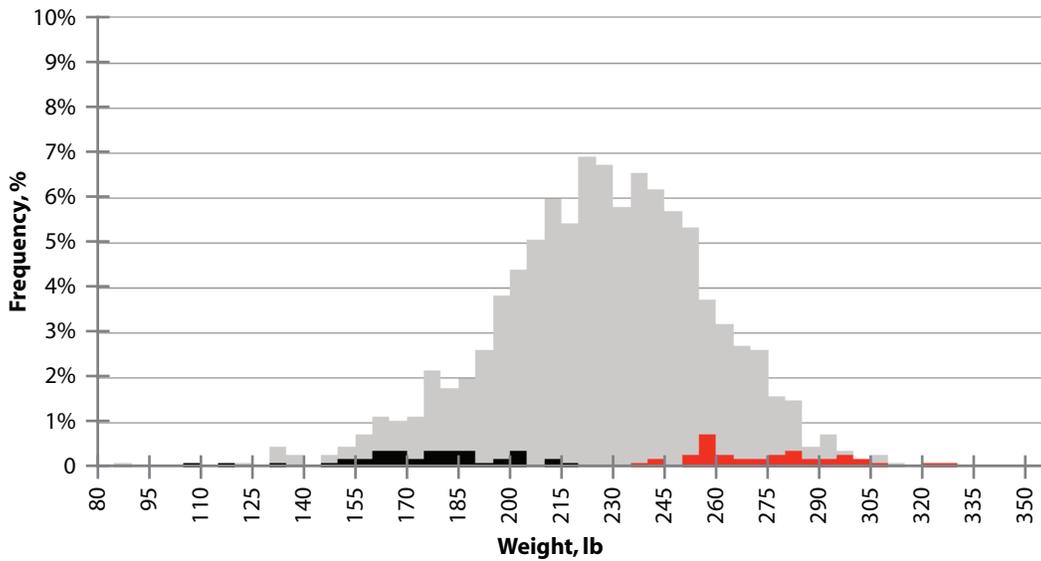


Figure 3. Histogram of Dataset C and marketer 2's selections. A total of 1,069 pigs were weighed (population mean = 222.4 lb, median = 224 lb, standard deviation = 32.0 lb, and CV = 14.4%) with 40 pens and 20 to 35 pigs per pen. The marketer selected the heaviest and lightest pig in each pen. The 2 histograms of the marketer's selections are imposed on top of the population histogram.

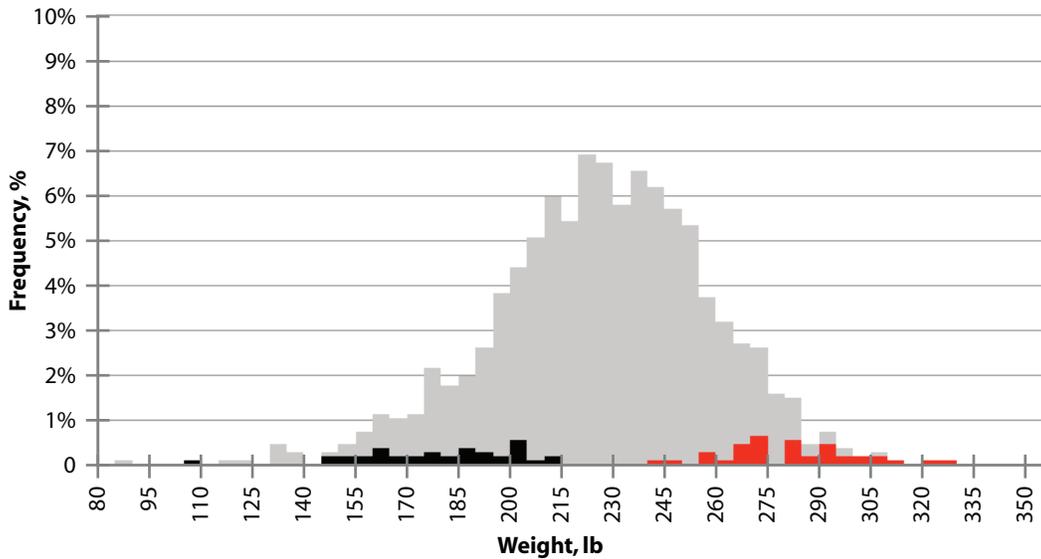


Figure 4. Histogram of Dataset C and marketer 3's selections. A total of 1,069 pigs were weighed (population mean = 222.4 lb, median = 224 lb, standard deviation = 32.0 lb, and CV = 14.4%) with 40 pens and 20 to 35 pigs per pen. The marketer selected the heaviest and lightest pig in each pen. The 2 histograms of the marketer's selections are imposed on top of the population histogram.