final report

Project code: A.MFS.0158

Prepared by: Andreas Kiermeier

South Australian Research and Development Institute

Date submitted: June 2009

PUBLISHED BY
Meat & Livestock Australia Limited
Locked Bag 991
NORTH SYDNEY NSW 2059

Positive lot sampling for E. coli O157

Meat & Livestock Australia acknowledges the matching funds provided by the Australian Government and contributions from the Australian Meat Processor Corporation to support the research and development detailed in this publication.

Contents

1 Introduction

Export beef processing establishments undertake sampling and testing of beef trim for E. coli O157 on a routine basis. Each lot of beef trim produced consists of at most 700 cartons. The protocol agreed with the US Food Safety Inspection Service (FSIS) involves random selection of 12 cartons from each lot. From each carton fives samples of 5{10 g each are selected and combined, yielding a total of 60 samples with a weight of at least 375 g (Anonymous, 2008). Lots which test positive for E. coli O157 are required to be disposed of under AQIS approved arrangements.

While establishments have detected E. coli O157 in their lots, little is know about the magnitude of lot contamination. That is, little is know about how widespread and acute contamination is throughout the lot, though they are believed to be low. To obtain more information about these data gaps Meat and Livestock Australia (MLA) commissioned Food Science Australia (FSA) to undertake intensive sampling of positive lots of beef trim for E. coli O157. The lots that were sampled were identified by meat processing establishments as having tested positive.

This report details the analysis of the results obtained by FSA.

2 Methodology

The sampling and testing methodology employed is detailed in the relevant MLA reports prepared by FSA. In brief, the twelve cartons which had previously sampled and tested positive for E. coli O157 were shipped to FSA and then thawed under controlled conditions. For each carton the following pieces of information were collected prior to microbiological sampling:

- number of pieces in each carton;
- weight of pieces in each carton;
- number of pieces in each carton that included an external surface of a carcase;
- an estimate of those external surface areas.

In the remainder of this document the term external piece will be used to denote pieces with external surface area.

2.1 Microbiological Testing

From each carton, 75 samples, each weighing approximately 5 g, were taken from those external pieces. Samples were individually enriched in 50 ml of E. coli O157 MP broth and then 10 ml subsamples of the enrichment were combined into composites of five. These were further enriched and then tested for E. coli O157. On a positive composite result, the individual samples that had gone into the composite were separately tested for E. coli O157.

This approach is in principal equivalent to testing all samples individually | the advantage is a reduction in the number of samples that need to be tested and hence considerable cost savings can be achieved.

2.2 Estimation of Concentration

Since E. coli O157 was expected to be present at very low levels, the microbiological methodology relied on enrichment, followed by presence/absence testing. Subsequently, estimates of concentration are not immediately available.

3 Results

However, the following approach, which is equivalent to the MPN approach, was used to estimate the concentration of E. coli O157 in each carton.
Assume that the concentration of E. coli O157 in the carton is λ organisms per gram.
The number, X, of E. coli O157 organisms in a 1 g sample has then a Poisson distribution, Po(λ), and similarly, the number of E. coli O157 organisms in a 5 g sample has a Po(5λ) distribution.
Therefore, the probability that a sample is negative equals the probability that there are zero E. coli O157 in the 5 g sample. This is given by

$$P^- = P(X = 0) = \frac{(5\lambda)^0 e^{-(5\lambda)}}{0!} = e^{-5\lambda},$$

and hence the probability of a single positive 5 g sample is given by $P^+ = 1 - P^-$.
Here however, 75 samples are removed from each carton. We therefore assume that each of the 5 g samples has the same probability of being positive and that samples are independent[1]. The number, Y, of samples testing positive is then binomially distributed, $B(P^+, n = 75)$, and the probability of observing y positive samples is given by

$$P(Y = y) = \binom{75}{n}(P^+)^y(1 - P^+)^{75-y} = \binom{75}{n}(1 - e^{-5\lambda})^y(e^{-5\lambda})^{75-y}.$$

Solving Equation (1) for λ yields the maximum likelihood estimate $\hat{\lambda}$, which is also known as the most probable number or MPN.
The same calculations can be performed if samples are based on a per cm² instead of a per g basis - the only thing that is required is the area represented by each 5 g sample.
The results for each lot that was intensively tested by FSA are presented in this section.

3.1 Lot A

This lot consisted of 560 consecutive cartons of beef trim, derived from 226 cows and 5 other animals (downgrade male carcases and bulls).
The following is a summary of the number of piece in the 12 cartons that were sent to FSA for further testing.

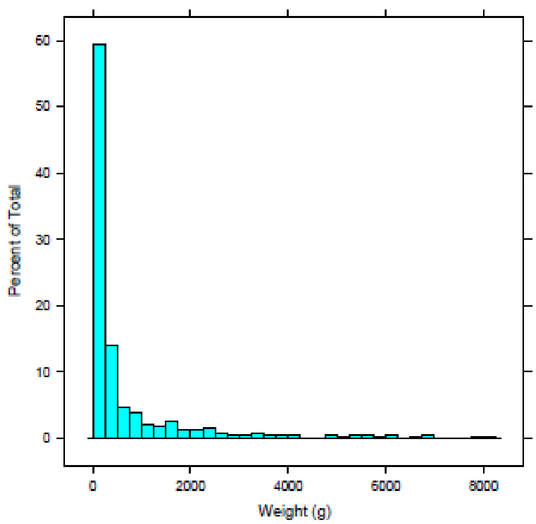
```
carton
  1   2   3   4   5   6   7   8   9  10  11  12
128  41  11   7   9  98  55   8  16   8  84  14
```

3.1.1 Weight

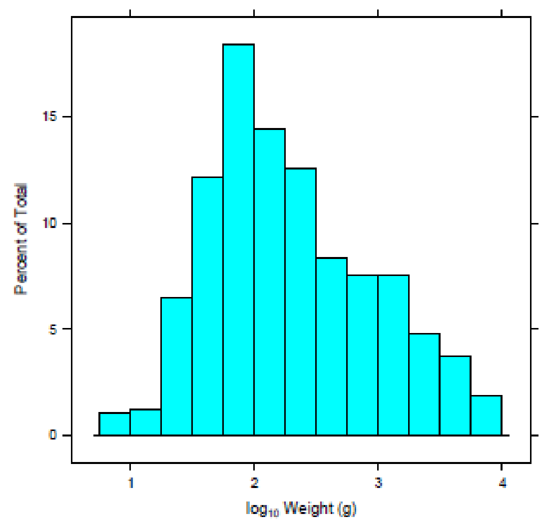
The total weight of beef trim pieces in each carton is summarised below.

```
> wcl <- with(Pieces.lot, tapply(weight, carton, FUN = sum))
> wcl
```

[1] This may not necessarily be the case if many samples are removed from large external pieces.



(a)



(b)

Figure 1: Histogram of the weight of beef trim pieces — (a) original scale and (b) $\log_{10}$ transformed.

```
    1     2     3     4     5     6     7     8     9    10    11    12
26966 26913 27079 26881 27006 26902 27097 26624 26811 26991 26814 26948
```

The following is a summary of the weight and $\log_{10}$ weight of the pieces across the sampled cartons in Lot A. Histograms of these are shown in Figure 1.

```
> with(Pieces.lot, summary(weight))

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    8.0    64.5   153.0   674.4   611.0  8215.0

> with(Pieces.lot, summary(log10(weight)))

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.9031  1.8100  2.1850  2.3100  2.7860  3.9150
```

The histogram of log10 of weight in Figure 1(b) shows a right-skewed distribution, which is not too far from a Normal distribution. This was investigated with the normal quantile-quantile plot (Figure 2) which shows departures from normality, as points do not lie closely enough to a straight line. Nevertheless, from a practical perspective, the distribution of log10 weight may reasonable be approximated by a Normal distribution with mean 2.31 log10 g and standard deviation 0.66 log10 g.

3.1.2 External Surface Area

In Lot A there were 479 pieces, of which 327 were external pieces. The total external surface area (cm2) in each carton is given below.

```
> scl <- with(Pieces.lot, tapply(surface, carton, FUN = sum))
> scl
      1       2       3       4       5       6       7       8       9      10
12492.0  8850.0  5028.0  6939.0  5837.0 10178.0 12103.0  6017.0  7959.0  4440.0
```

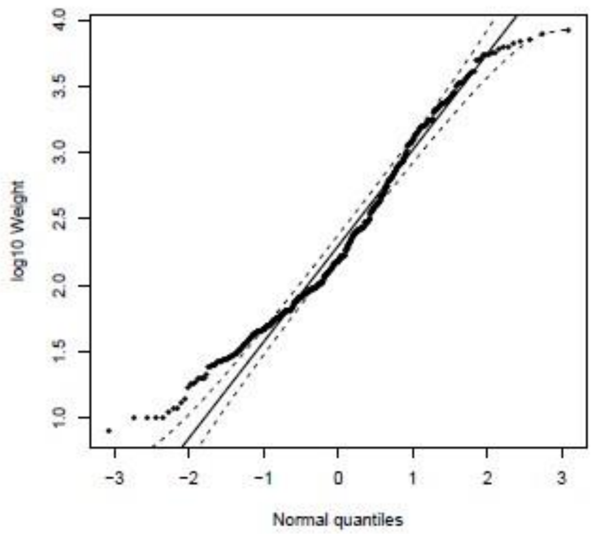


Figure 2: Normal Q-Q plot of the $\log_{10}$ weight of beef trim pieces.

```
     11      12
11057.5  8616.0
```

A summary of the external surface area of the pieces in Lot A and their log10 values is given below and histograms of the external surface area and the $\log_{10}$ of external surface area are shown in Figure 3.

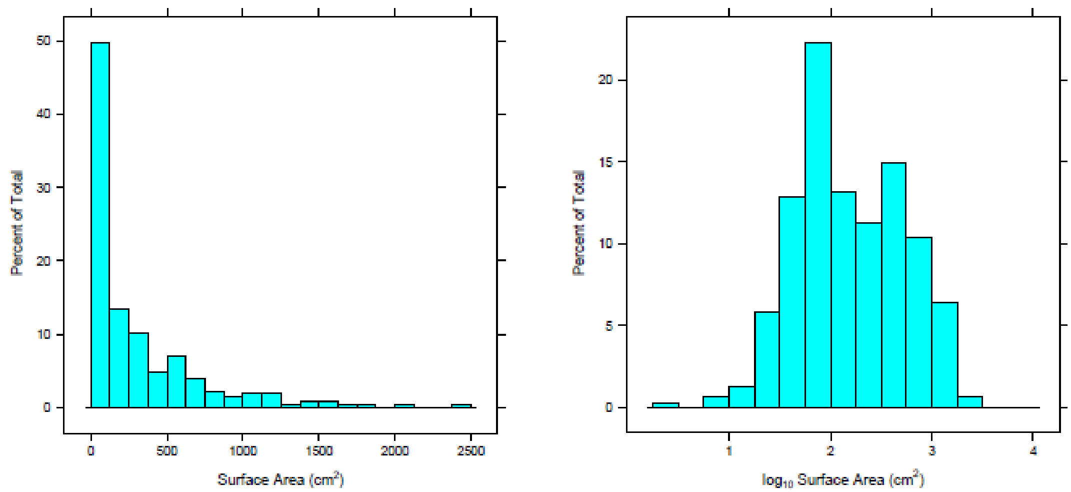
```
> with(Pieces.lot.ext, summary(surface))

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    3.0    64.0   132.0   304.3   415.0  2460.0

> with(Pieces.lot.ext, summary(log10(surface)))

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.4771  1.8060  2.1210  2.1830  2.6180  3.3910
```

Given the shape of the histogram of the $\log_{10}$ of the external surface area in Figure 3(b), a normal distribution for external surface area does not appear reasonable for this lot - there is some indication that the distribution may even be bimodal.

For external pieces, the relationship between weight and external surface area was investigated. The scatter plot of the $\log_{10}$ transformation of these two variables is shown in Figure 4.

A linear regression model appears to fit the data well and the model summary is given below.

```
Call:
lm(formula = log10(surface) ~ log10(weight), data = Pieces.lot.ext)

Residuals:
     Min       1Q   Median       3Q      Max
-0.72729 -0.10982  0.01480  0.12801  0.56583

Coefficients:
              Estimate Std. Error t value      Pr(>|t|)
(Intercept)    0.27908    0.04391   6.356 0.000000000699
log10(weight)  0.75421    0.01684  44.783        < 2e-16
```



(a) (b)

Figure 3: Histogram of external surface area of external pieces — (a) original scale and (b) $\log_{10}$ transformed.

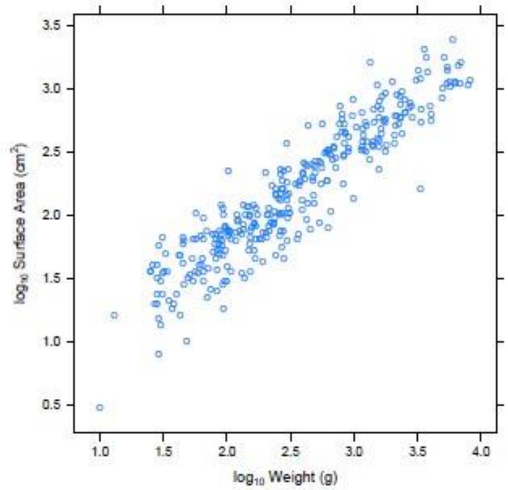


Figure 4: Scatter plot of the $\log_{10}$ external surface area versus $\log_{10}$ weight for beef trim pieces with an external surface area.

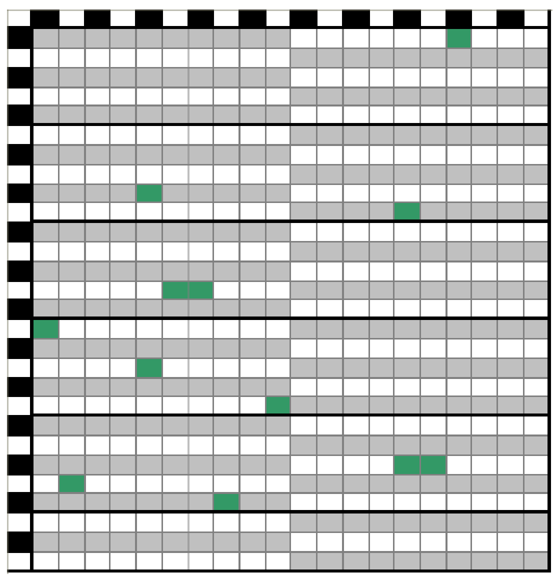


Figure 5: Diagram of location of cartons in the lot, numbered from left to right and top to bottom. Each square represents a carton in the lot — sampled cartons are presented in green and cartons that tested positive are presented as red (top row and left column are include to make counting easier).

```
Residual standard error: 0.1977 on 325 degrees of freedom
Multiple R-squared: 0.8605,       Adjusted R-squared: 0.8601
F-statistic:  2006 on 1 and 325 DF,  p-value: < 2.2e-16
```

The output from the regression model indicates that the estimated line is

$$\log_{10}(\text{External Surface Area}) = 0.28 + 0.75 \log_{10}(\text{Weight}) \,,$$

or on the original scale

$$\text{External Surface Area} = 10^{0.28} \times \text{Weight}^{0.75} \,.$$

Since the linear model was fitted on the log scale, the results can also be presented as percentage changes. This means that a 1% increase in Weight is associated with a 0.75% increase in external surface area.

In relation to the summary of total external surface area in each carton, presented at the beginning of this section, this model seems to result in reasonable estimates of the total surface area in a carton.

This can be seen from the following summary which presents the sum of the estimated surface area (from the weight) for all external pieces.

1	2	3	4	5	6	7	8	9	10	11	12
9209	8030	7047	6341	6362	8842	8104	6392	7818	6003	10218	7676

3.1.3 Contamination

This lot resulted in a total of 0 positive E. coli O157 tests. An overview of the cartons in the lot, including those that were sampled, is given in Figure 5.

Using the MPN calculations presented in Section 2.2 this indicates that the concentration in each carton is less than 0.0027 organism per g (95% CI upper bound: 0.0118). On a carton basis this would yield less than 73 organisms (95% CI upper bound: 321).

Performing the same MPN calculations on a per cm2 basis, assuming an external surface area of 10 cm2 per sample, results in concentration estimate of less than 0.0013 organism per cm2 (95% CI upper bound: 0.0059). Larger surface areas per sample would result in even lower concentration estimates.

Combined with total surface area per carton, as summarised earlier, this would indicate less than 17 organisms[2] per carton on average.

As stated earlier, there were 560 cartons in this lot. However, none of the 12 cartons tested positive during the intensive investigation, despite previously having resulted in a positive test result using the BioControl VIP 8 hour test. This subsequently leaves the following possible conclusions:

- Initial Screening test resulted in a false positive and this lot was truly negative for E. coli O157;
- The concentration (and prevalence) of E. coli O157 in the lot is so low that it was detected by chance on the initial screening test, but not on the subsequent testing.
- The concentration of E. coli O157 in the lot is very low and the carton prevalence is less than 0.265,[3] assuming that contamination and sampling occur completely randomly.

[2] Using the maximum total surface area per carton

[3] The upper bound for the 95% confidence interval, based on the Binomial distribution with n = 12

3.2 Lot B

This lot consisted of 12 cartons of veal trim, derived from 200 calves.
The following is a summary of the number of piece in the 12 cartons in Lot B.

```
carton
 1  2  3  4  5  6  7  8  9 10 11 12
28 14 18 30 11 16 44 41 24 35 25 26
```

3.2.1 Weight

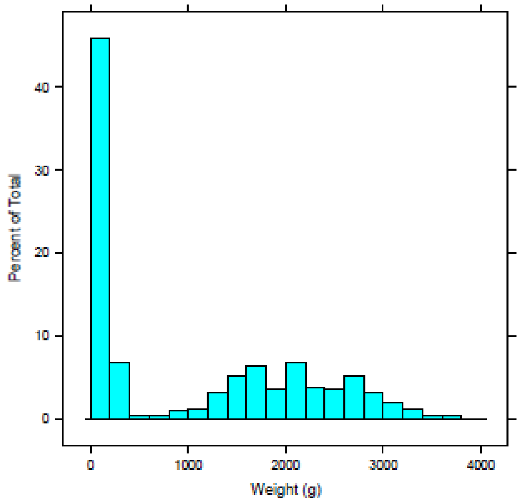
The total weight of beef trim pieces in each carton is summarised below.

```
> wcl <- with(Pieces.lot, tapply(weight, carton, FUN = sum))
> wcl

    1     2     3     4     5     6     7     8     9    10    11    12
27362 27117 26983 27331 27264 26950 27110 27109 27193 27379 27469 27226
```

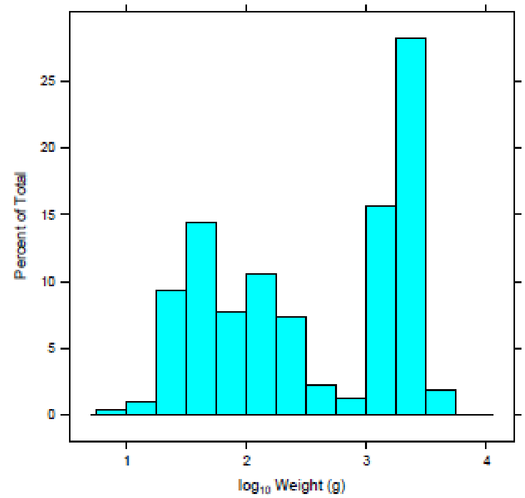
The following is a summary of the weight and log10 weight of the pieces across the sampled cartons in Lot B. Histograms of these are shown in Figure 6.

```
> with(Pieces.lot, summary(weight))

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   10.0    56.5   295.0  1046.0  2088.0  3721.0

> with(Pieces.lot, summary(log10(weight)))

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   1.752   2.469   2.533   3.320   3.571
```

Both histograms show a bimodal distribution. In particular, the distribution of weight shows a large proportion of very small pieces.



(a)



(b)

Figure 6: Histogram of the weight of beef trim pieces — (a) original scale and (b) $\log_{10}$ transformed.

3.2.2 External Surface Area

In Lot B there were 312 pieces, of which 209 were external pieces. The total external surface area (cm2) in each carton is given below.

```
> scl <- with(Pieces.lot, tapply(surface, carton, FUN = sum))
> scl

    1     2     3     4     5     6     7     8     9    10    11    12
11493  8100 11335  8862  8115  8928  7225  8287  7157  8040  9024  8865
```

A summary of the external surface area of the pieces in Lot B and their log10 values is given below and histograms of the external surface area and the log10 of surface area are shown in Figure 7.

```
> with(Pieces.lot.ext, summary(surface))

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  21.0   147.0   480.0   504.5   729.0  2403.0

> with(Pieces.lot.ext, summary(log10(surface)))

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.322   2.167   2.681   2.529   2.863   3.381
```

Given the shape of the histogram of the log10 of the external surface area in Figure 7(b), a normal distribution for external surface area does not appear reasonable for this lot | there is again some indication that the distributions may be bimodal, which is not surprising in light of the distribution of the weight of pieces. This is most likely due to the very small pieces having little or no external surface area.

As above, the relationship between weight and external surface area was investigated for external pieces. The scatter plot of the log10 transformation of these two variables is shown in Figure 8.
A linear regression model _ts the data well and the model summary is given below.

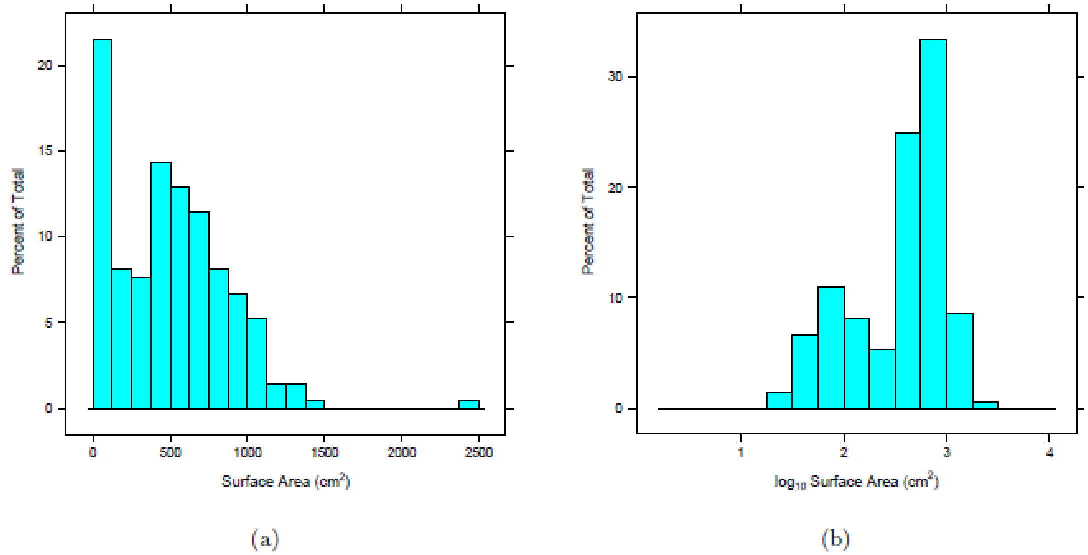


Figure 7: Histogram of external surface area of external pieces — (a) original scale and (b) $\log_{10}$ transformed.

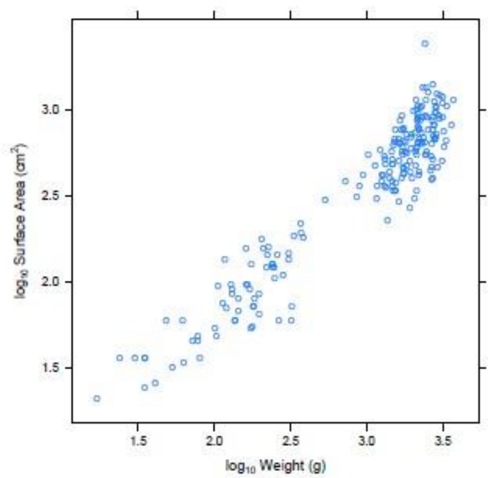


Figure 8: Scatter plot of the $\log_{10}$ external surface area versus $\log_{10}$ weight for beef trim pieces with an external surface area.

Figure 9: Diagram of location of cartons in the lot, numbered from left to right and top to bottom. Each square represents a carton in the lot — sampled cartons are presented in green and cartons that tested positive are presented as red (top row and left column are include to make counting easier).

```
Call:
lm(formula = log10(surface) ~ log10(weight), data = Pieces.lot.ext)

Residuals:
      Min        1Q    Median        3Q       Max
-0.406559 -0.089634  0.004135  0.101052  0.530348

Coefficients:
              Estimate Std. Error t value    Pr(>|t|)
(Intercept)    0.27731    0.05216   5.316 0.000000273
log10(weight)  0.76098    0.01731  43.965     < 2e-16

Residual standard error: 0.1428 on 207 degrees of freedom
Multiple R-squared: 0.9033,        Adjusted R-squared: 0.9028
F-statistic:  1933 on 1 and 207 DF,  p-value: < 2.2e-16
```

The output from the regression model indicates that the estimated linear line is

$$\log_{10}(\text{External Surface Area}) = 0.28 + 0.76\log_{10}(\text{Weight}) ,$$

or on the original scale

$$\text{External Surface Area} = 10^{0.28} \times \text{Weight}^{0.76} .$$

This model is similar to that estimated for Lots 1 and 2, but with a slightly higher intercept term. Since the linear model was fitted on the log scale, the results can also be presented as percentage changes. This means that a 1% increase in Weight is associated with a 0.76% increase in external surface area.

In relation to the summary of total external surface area in each carton, presented at the beginning of this section, this model seems to result in reasonable estimates of the total surface area in a carton. This can be seen from the following summary which presents the sum of the estimated surface area (from the weight) for all external pieces.

1	2	3	4	5	6	7	8	9	10	11	12
8401	8202	8047	8523	7921	8606	8145	8397	7934	8334	8449	8203

3.2.3 Contamination

This lot resulted in a total of 23 positive E. coli O157 tests - the number of positive samples per carton is summarised below.

1	2	3	4	5	6	7	8	9	10	11	12
0	5	0	8	0	0	7	3	0	0	0	0

An overview of the cartons in the lot, including those that were sampled, is given in Figure 9.

Using the MPN calculations presented in Section 2.2 gives the per g and per cm2 based results shown in Table 1.

Table 1: MPN calculations on per g and per cm^2 basis for the number of positive samples per carton observed in Lot B

+ve	MPN (per g)	Upper 95 (per g)	MPN (per cm^2)	Upper 95 (per cm^2)
3	0.008	0.021	0.004	0.011
5	0.014	0.030	0.007	0.015
7	0.020	0.038	0.010	0.019
8	0.023	0.042	0.011	0.021

As stated earlier, there were 12 cartons in this lot, and 4 of the 12 cartons tested positive during the intensive investigation after previously having tested positive with the BioControl VIP test with unknown incubation time. This indicates that a substantial proportion of the lot, and potentially the whole lot, was contaminated. This is also substantiated by observing that cartons early and late in this lot tested positive for E. coli O157 (Figure 9).

3.3 Lot C

This lot consisted of 24 cartons of veal trim, derived from 273 bobby calves.
The following is a summary of the number of piece in the 12 cartons sampled in Lot C.

```
carton
 1  2  3  4  5  6  7  8  9 10 11 12
15 17 16  9 10 16 15 15 14 15 16 18
```

3.3.1 Weight

The total weight of beef trim pieces in each carton is summarised below.

```
> wcl <- with(Pieces.lot, tapply(weight, carton, FUN = sum))
> wcl

    1     2     3     4     5     6     7     8     9    10    11    12
27157 27180 26947 27383 26974 27059 27258 27233 27032 27047 26616 27188
```

The following is a summary of the weight and log10 weight of the pieces across the sampled cartons in Lot C. Histograms of these are shown in Figure 10.

```
> with(Pieces.lot, summary(weight))

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   17.0    57.5  2490.0  1847.0  3245.0  5888.0

> with(Pieces.lot, summary(log10(weight)))

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.230   1.760   3.396   2.776   3.511   3.770
```

Both histograms show bimodal distributions similar to Lot B . In particular, the distribution of weight shows a large proportion of very small pieces.

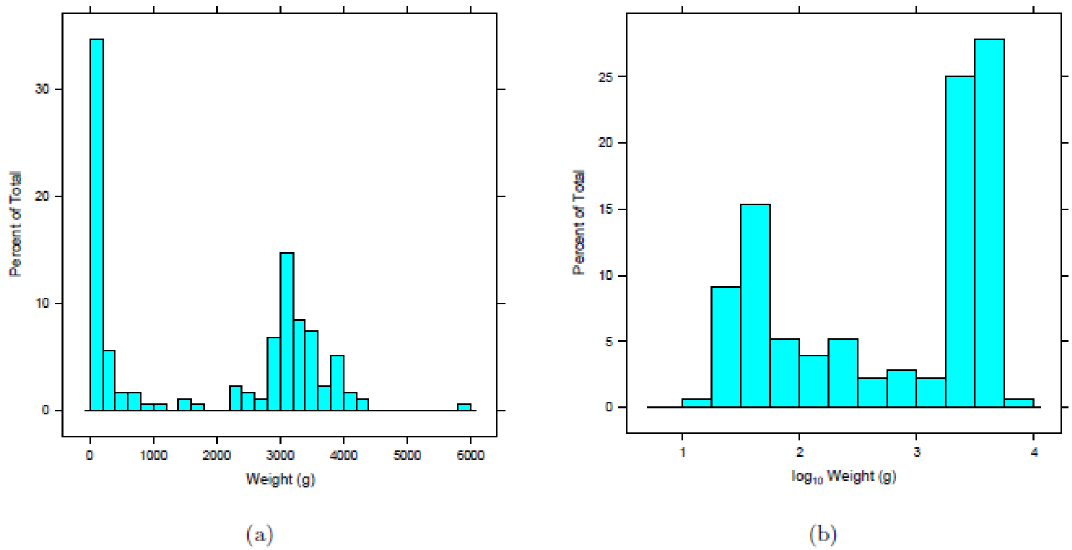


Figure 10: Histogram of the weight of beef trim pieces — (a) original scale and (b) $\log_{10}$ transformed.

3.3.2 External Surface Area

In Lot C there were 176 pieces, of which 126 were external pieces. The total external surface area (cm2) in each carton is given below.

```
> scl <- with(Pieces.lot, tapply(surface, carton, FUN = sum))
> scl

   1    2    3    4    5    6    7    8    9   10   11   12
9064 6463 8569 7138 7371 7322 7077 7048 8166 6472 8627 7980
```

A summary of the external surface area of the pieces in Lot C and their log10 values is given below and histograms of the external surface area and the log10 of surface area are shown in Figure 11.

```
> with(Pieces.lot.ext, summary(surface))

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   36.0   483.0   765.0   724.6   986.0  1450.0

> with(Pieces.lot.ext, summary(log10(surface)))

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.556   2.684   2.884   2.740   2.994   3.161
```

Given the shape of the histogram of the log10 of the external surface area in Figure 11(a), a normal distribution for external surface area does not appear reasonable for this lot | there is again some indication that the distributions may be bimodal. This is most likely due to the very small pieces having little or no external surface area.

As above, the relationship between weight and external surface area was investigated for external pieces. The scatter plot of the log10 transformation of these two variables is shown in Figure 12.

A linear regression model fits the data well and the model summary is given below.

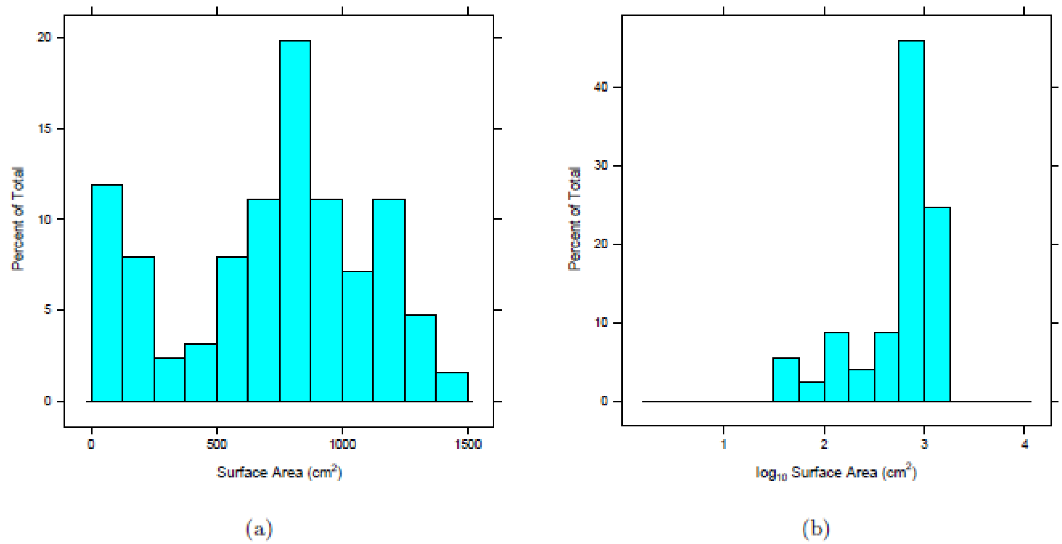


Figure 11: Histogram of external surface area of external pieces — (a) original scale and (b) $\log_{10}$ transformed.

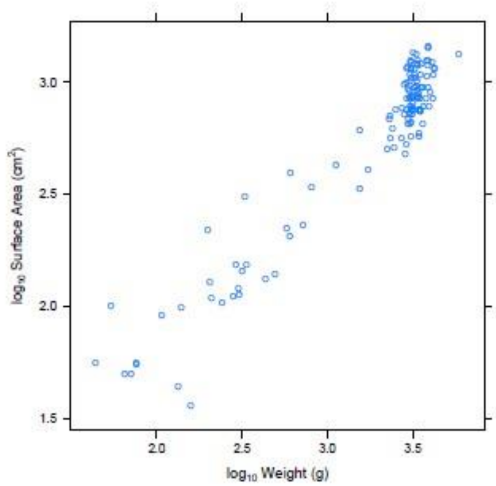


Figure 12: Scatter plot of the $\log_{10}$ external surface area versus $\log_{10}$ weight for beef trim pieces with an external surface area.

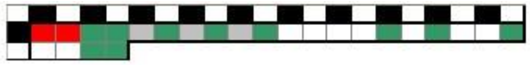


Figure 13: Diagram of location of cartons in the lot, numbered from left to right and top to bottom. Each square represents a carton in the lot — sampled cartons are presented in green and cartons that tested positive are presented as red (top row and left column are include to make counting easier).

```
Call:
lm(formula = log10(surface) ~ log10(weight), data = Pieces.lot.ext)

Residuals:
      Min         1Q    Median        3Q       Max
-0.420146 -0.080324  0.002785  0.078161  0.368491

Coefficients:
              Estimate Std. Error t value    Pr(>|t|)
(Intercept)     0.3574     0.0666   5.366 0.000000381
log10(weight)   0.7355     0.0203  36.226     < 2e-16

Residual standard error: 0.1179 on 124 degrees of freedom
Multiple R-squared: 0.9137,       Adjusted R-squared: 0.913
F-statistic:  1312 on 1 and 124 DF,  p-value: < 2.2e-16
```

The output from the regression model indicates that the estimated linear line is

$$\log_{10}(\text{External Surface Area}) = 0.36 + 0.74 \log_{10}(\text{Weight}) \,,$$

or on the original scale

$$\text{External Surface Area} = 10^{0.36} \times \text{Weight}^{0.74} \,.$$

This model is similar to that estimated for Lot A. Since the linear model was fitted on the log scale, the results can also be presented as percentage changes. This means that a 1% increase in Weight is associated with a 0.74% increase in external surface area.

In relation to the summary of total external surface area in each carton, presented at the beginning of this section, this model seems to result in reasonable estimates of the total surface area in a carton.

This can be seen from the following summary which presents the sum of the estimated surface area (from the weight) for all external pieces.

```
   1    2    3    4    5    6    7    8    9   10   11   12
7629 7537 7454 7375 7191 7353 7380 7384 7250 7374 6953 7383
```

3.3.3 Contamination

This lot resulted in a total of 3 positive E. coli O157 tests - the number of positive samples per carton is summarised below.

```
1  2  3  4  5  6  7  8  9 10 11 12
0  0  1  0  0  0  0  0  0  0  2  0
```

An overview of the cartons in the lot, including those that were sampled, is given in Figure 13.

Using the MPN calculations presented in Section 2.2 gives the per g and per cm2 based results shown in Table 2.

Table 2: MPN calculations on per g and per cm^2 basis for the number of positive samples per carton observed in Lot C

+ve	MPN (per g)	Upper 95 (per g)	MPN (per cm^2)	Upper 95 (per cm^2)
1	0.003	0.012	0.001	0.006
2	0.005	0.017	0.003	0.008

As stated earlier, there were 24 cartons in this lot, and 2 of the 12 cartons tested positive during the intensive investigation after previously having tested positive with the BioControl VIP test with unknown incubation time.
Consequently, a 95% confidence interval for the number of positive cartons in the lot is (2, 10), indicating that a substantial proportion of this lot was contaminated. Interestingly, the cartons confirmed as positive in this lot were the first two cartons in this lot (Figure 13). However, it is presently not clear whether these cartons were produced consecutively or whether other animals (not calves) were processed / boned throughout the production of this lot. This may need further clarification.

3.4 Lot D

This lot consisted of 528 consecutive cartons of beef trim, derived from 432 carcases. From these cartons, 19 cartons were included in the original sampling | some samples had been sourced across two cartons.[4]
The following is a summary of the number of piece in the 19 cartons that were sent to FSA for further testing. Where two cartons had been used for the initial sampling, those two carton were also used for intensive sampling, with 37 and 38 samples drawn from the two cartons.

```
carton
  1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19
 69  93 100  59  87  45  21  17  27 166  50  43  62 192  52  89  53 117 101
```

3.4.1 Weight

The total weight of beef trim pieces in each carton is summarised below.

```
> wcl <- with(Pieces.lot, tapply(weight, carton, FUN = sum))
> wcl

    1     2     3     4     5     6     7     8     9    10    11    12    13    14
27332 27038 26959 27708 26990 26967 26906 27027 27105 27147 27168 27171 26967 26601
   15    16    17    18    19
26882 26912 27070 26904 26818
```

[4] Random selection was based on time of production | there were several instances where 2 cartons were produced in the same time interval (Figure 18).

The following is a summary of the weight and log10 weight of the pieces across the sampled cartons in Lot D. Histograms of these are shown in Figure 14.

```
> with(Pieces.lot, summary(weight))
```

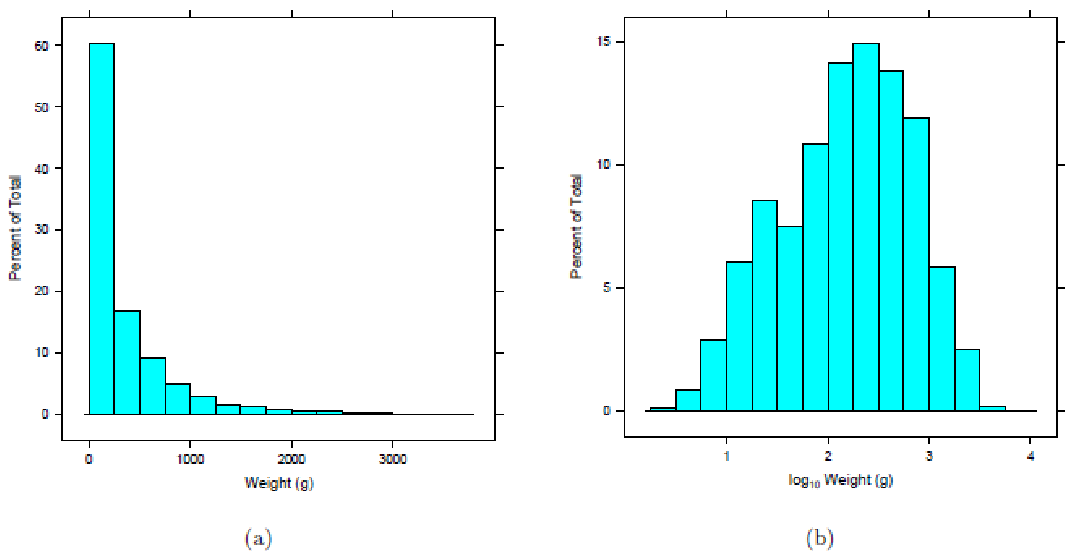


Figure 14: Histogram of the weight of beef trim pieces — (a) original scale and (b) $\log_{10}$ transformed.

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    3.0    54.5   171.0   356.0   450.0  3608.0
```

```
> with(Pieces.lot, summary(log10(weight)))
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.4771  1.7360  2.2330  2.1730  2.6530  3.5570
```

The histogram of $\log_{10}$ of weight in Figure 14(b) shows a right-skewed distribution, which is not too far from a Normal distribution. This was investigated with the normal quantile-quantile plot (Figure 15) which shows departures from normality, as points do not lie closely enough to a straight line. Nevertheless, from a practical perspective, the distribution of log10 weight may reasonable be approximated by a Normal distribution with mean 2.17 log10 g and standard deviation 0.64 log10 g.

3.4.2 External Surface Area

In Lot D there were 1443 pieces, of which 1212 were external pieces. The total external surface area (cm2) in each carton is given below.

```
> scl <- with(Pieces.lot, tapply(surface, carton, FUN = sum))
> scl
    1     2     3     4     5     6     7     8     9    10    11    12    13    14
12308 12108 10052  9904  9320  8729  6972  5626  6918  8443  6754  8985  7847 11676
   15    16    17    18    19
 9614 10707  8089  7897 10039
```

A summary of the external surface area of the pieces in Lot D and their log10 values is given below and histograms of the external surface area and the log10 of external surface area are shown in Figure 16.

```
> with(Pieces.lot.ext, summary(surface))
```

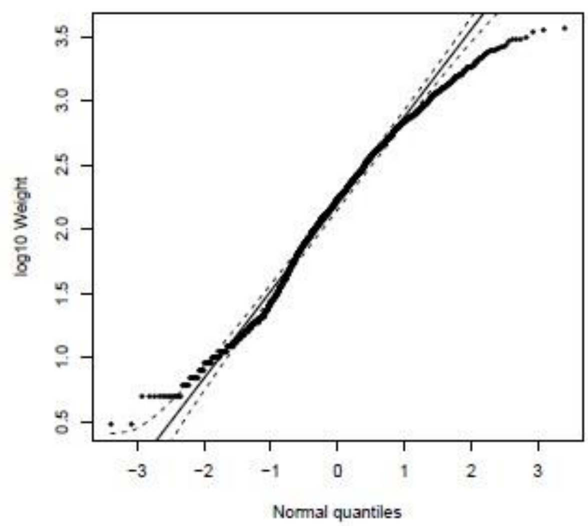


Figure 15: Normal Q-Q plot of the $\log_{10}$ weight of beef trim pieces.

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    3.0    32.0    91.0   141.9   204.0  1015.0

> with(Pieces.lot.ext, summary(log10(surface)))

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.4771  1.5050  1.9590  1.8950  2.3100  3.0060
```

Given the shape of the histogram of the log10 of the external surface area in Figure 16(b), a normal distribution for external surface area does not appear reasonable for this lot|the distribution appears left skewed.

For external pieces, the relationship between weight and external surface area was investigated. The scatter plot of the log10 transformation of these two variables is shown in Figure 17. From this graph it appears to show more variability at the low end, i.e., small pieces.
A linear regression model appears to fit the data reasonably well - however there appears to be a problem with distribution of the residuals which are not normally distributed, but instead appear to reflect the observations made for the distribution of log10 external surface area. This should however not affect the estimation of the regression parameters, but the variance estimates and therefore any inferences that are to be drawn from the model.
The model summary is given below.

```
Call:
lm(formula = log10(surface) ~ log10(weight), data = Pieces.lot.ext)

Residuals:
     Min       1Q   Median       3Q      Max
-1.03014 -0.12280  0.03386  0.14853  0.68369

Coefficients:
              Estimate Std. Error t value   Pr(>|t|)
(Intercept)    0.11928    0.02487   4.796 0.00000182
log10(weight)  0.78035    0.01056  73.892    < 2e-16
```

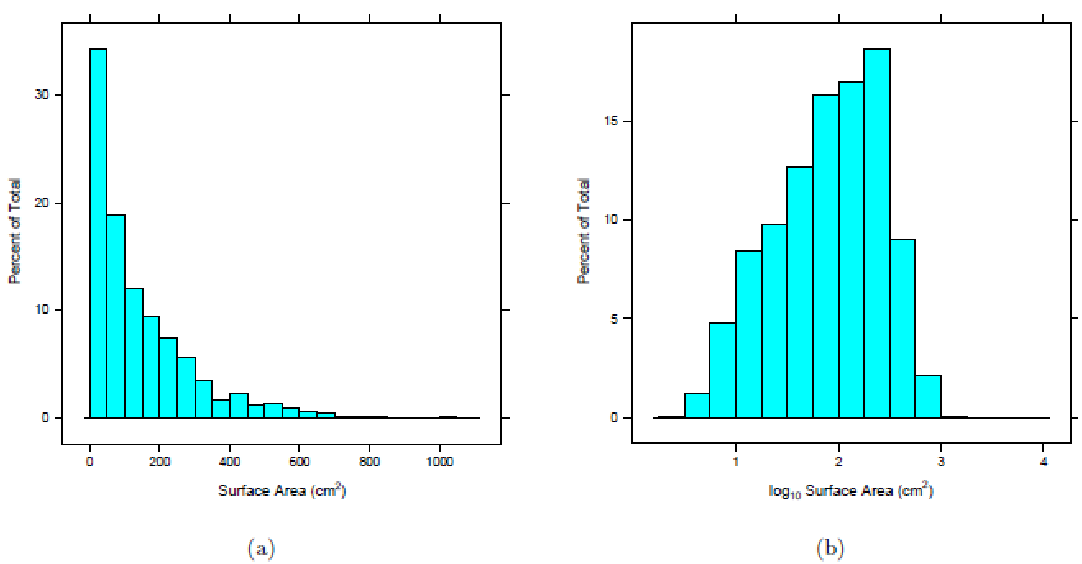


Figure 16: Histogram of external surface area of external pieces — (a) original scale and (b) $\log_{10}$ transformed.

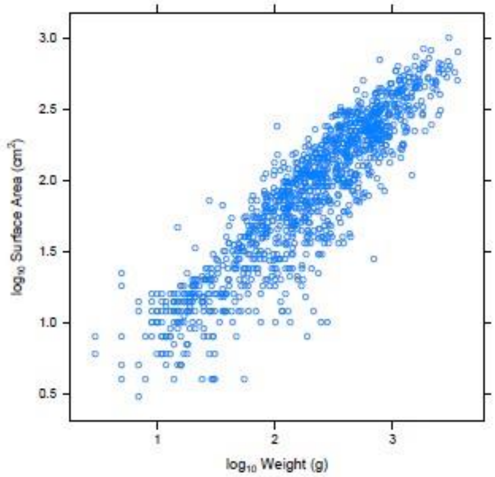


Figure 17: Scatter plot of the $\log_{10}$ external surface area versus $\log_{10}$ weight for beef trim pieces with an external surface area.

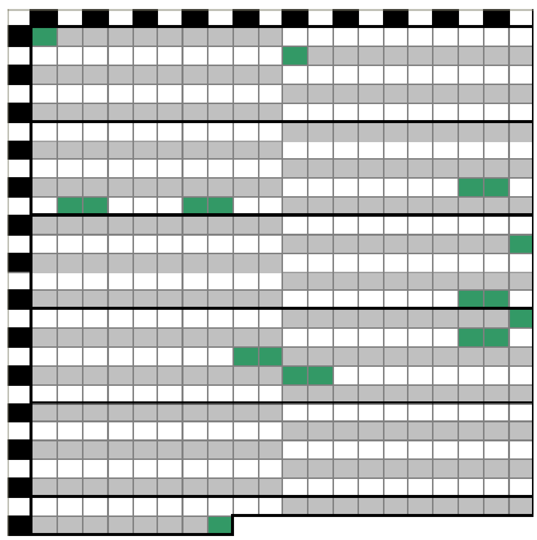


Figure 18: Diagram of location of cartons in the lot, numbered from left to right and top to bottom. Each square represents a carton in the lot — sampled cartons are presented in green and cartons that tested positive are presented as red (top row and left column are include to make counting easier).

```
Residual standard error: 0.2227 on 1210 degrees of freedom
Multiple R-squared: 0.8186,       Adjusted R-squared: 0.8184
F-statistic:  5460 on 1 and 1210 DF,  p-value: < 2.2e-16
```

The output from the regression model indicates that the estimated line is

$$\log_{10}(\text{External Surface Area}) = 0.12 + 0.78 \log_{10}(\text{Weight}) ,$$

or on the original scale

$$\text{External Surface Area} = 10^{0.12} \times \text{Weight}^{0.78} .$$

The slope of the $\log_{10}$-linear model is similar to those estimated for the previous lots while the intercept is considerably lower.
Since the linear model was fitted on the log scale, the results can also be presented as percentage changes. This means that a 1% increase in Weight is associated with a 0.78% increase in external surface area.
In relation to the summary of total external surface area in each carton, presented at the beginning of this section, this model seems to result in reasonable estimates of the total surface area in a carton.
This can be seen from the following summary which presents the sum of the estimated surface area (from the weight) for all external pieces.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
9135	9199	8915	8364	9007	8028	6626	6815	7044	8393	7516	8069	7568	9446	8101	9059	7995

18	19
8326	8748

3.4.3 Contamination

This lot resulted in a total of 0 positive E. coli O157 tests. An overview of the lot is given in Figure 18.
Using the MPN calculations presented in Section 2.2[5] this indicates that the concentration in each carton is less than 0.0055 organism per g (95% CI upper bound: 0.0241). On a carton basis this would yield less than 149 organisms (95% CI upper bound: 655).
Performing the same MPN calculations on a per cm2 basis, assuming an external surface area of 10 cm2 per sample, results in concentration estimate of less than 0.0027 organism per cm2 (95% CI upper bound: 0.012). Larger surface areas per sample would result in even lower concentration estimates.
Combined with total surface area per carton, as summarised earlier, this would indicate less than 34 organisms[6] per carton on average.
As stated earlier, there were 528 cartons in this lot. However, none of the 19 cartons tested positive during the intensive investigation, despite previously having resulted in a positive test result using 15-22h PCR based screening test using DuPont BAX MP. This subsequently leaves the following possible conclusions:

- Initial Screening test resulted in a false positive and this lot was truly negative for E. coli O157- given that the screening test was undertaken by Symbio Alliance, who have experience in this type of testing, this outcome appears unlikely;
- The concentration (and prevalence) of E. coli O157 in the lot is so low that it was detected by chance on the initial screening test, but not on the subsequent testing.
- The concentration of E. coli O157 in the lot is very low and the carton prevalence is less than 0.176[7] assuming that contamination and sampling occur completely randomly.

3.5 Lot E

This lot consisted of 399 cartons of bull trimming, derived from 113 bulls.

[5] The assumption here is that only 37 samples were drawn from each carton | this is true for 7 cartons. The remaining cartons had more samples drawn | 38 or 75 | which will result in a lower MPN. Consequently, this assumption results in a conservative approach
[6] Using the maximum total surface area per carton.
[7] The upper bound for the 95% con_dence interval, based on the Binomial distribution with n = 19.

The following is a summary of the number of piece in the 12 cartons in Lot E. Compared to some of the previous lots it is immediately obvious that there are fewer pieces per carton.

```
carton
 1  2  3  4  5  6  7  8  9 10 11 12
15 26 20 45 13 37 23 12 12 25 25 25
```

3.5.1 Weight

The total weight of beef trim pieces in each carton is summarised below.

```
> wcl <- with(Pieces.lot, tapply(weight, carton, FUN = sum))
> wcl

    1     2     3     4     5     6     7     8     9    10    11    12
27303 27016 27160 27041 27118 26854 26956 26872 26981 26890 26902 27247
```

The following is a summary of the weight and log10 weight of the pieces across the sampled cartons in Lot E. Histograms of these are shown in Figure 19.

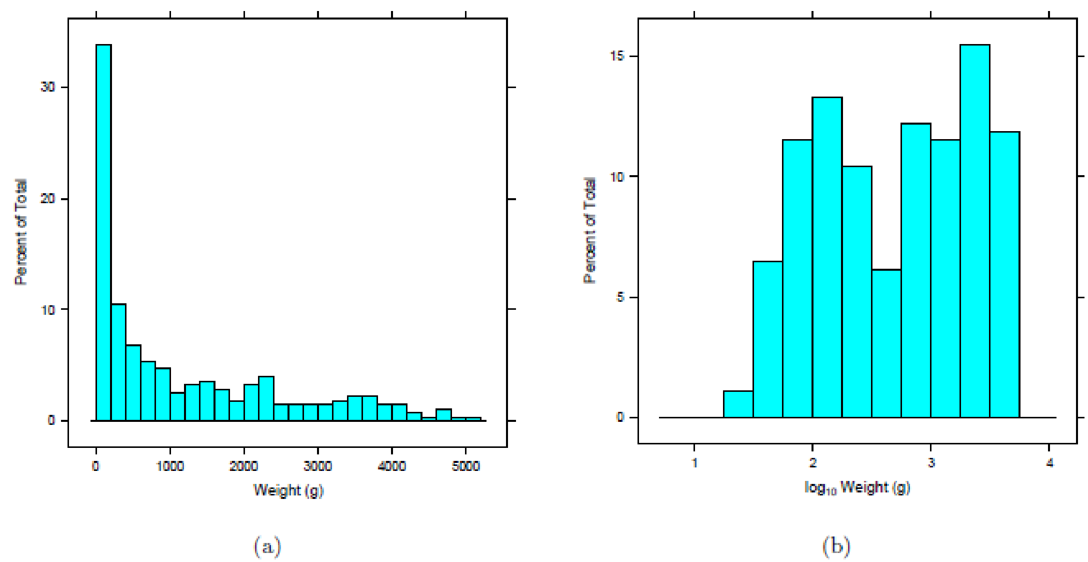


Figure 19: Histogram of the weight of beef trim pieces — (a) original scale and (b) $\log_{10}$ transformed.

```
> with(Pieces.lot, summary(weight))

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   18.0   126.0   584.5  1167.0  1949.0  5062.0

> with(Pieces.lot, summary(log10(weight)))

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.255   2.100   2.767   2.689   3.290   3.704
```

The histogram of weight is highly right skewed while the histogram of log10 weight indicates a distribution which could be bimodal. A normal approximation is clearly not applicable to either.

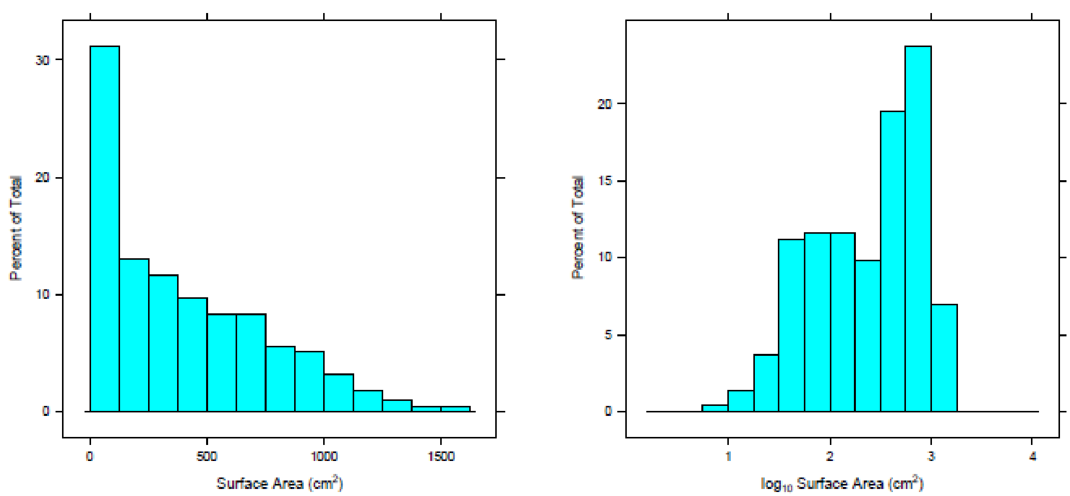
3.5.2 External Surface Area

In Lot E there were 278 pieces, of which 215 were external pieces. The total external surface area (cm2) in each carton is given below.

```
> scl <- with(Pieces.lot, tapply(surface, carton, FUN = sum))
> scl

   1    2    3    4    5    6    7    8    9   10   11   12
5230 5419 4365 7562 5634 8354 7136 8353 8462 7894 9761 8263
```

A summary of the external surface area of the pieces in Lot E and their log10 values is given below and histograms of the external surface area and the log10 of surface area are shown in Figure 20.

```
> with(Pieces.lot.ext, summary(surface))

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
     10      84     319     402     656    1596

> with(Pieces.lot.ext, summary(log10(surface)))
```



(a) (b)

Figure 20: Histogram of external surface area of external pieces — (a) original scale and (b) $\log_{10}$ transformed.

```
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.000   1.924   2.504   2.367   2.817   3.203
```

Neither histogram in Figure 20 shows a distribution which could reasonably be approximated by a Normal distribution.
The relationship between weight and external surface area was again investigated for external pieces.
The scatter plot of the log10 transformation of these two variables is shown in Figure 21.
A linear regression model _ts the data reasonably well | the model summary is given below.

```
Call:
lm(formula = log10(surface) ~ log10(weight), data = Pieces.lot.ext)

Residuals:
     Min       1Q   Median       3Q      Max
-0.82513 -0.10958  0.02539  0.15225  0.51297

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.09820    0.07957  -1.234    0.219
log10(weight)  0.85057    0.02697  31.534   <2e-16

Residual standard error: 0.2188 on 213 degrees of freedom
Multiple R-squared: 0.8236,       Adjusted R-squared: 0.8228
F-statistic: 994.4 on 1 and 213 DF,  p-value: < 2.2e-16
```

The output from the regression model indicates that the estimated linear line is

$$\log_{10}(\text{External Surface Area}) = -0.1 + 0.85 \log_{10}(\text{Weight}) ,$$

or on the original scale

$$\text{External Surface Area} = 10^{-0.1} \times \text{Weight}^{0.85} .$$

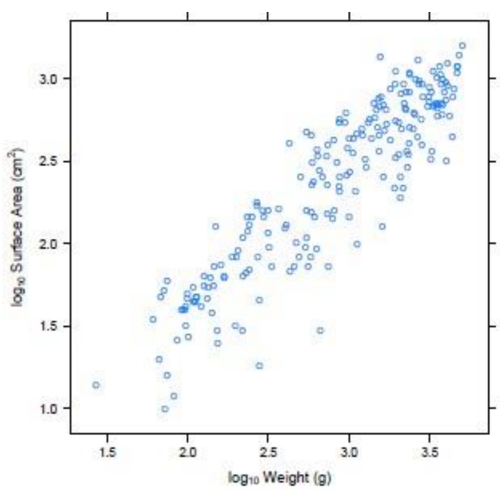


Figure 21: Scatter plot of the $\log_{10}$ external surface area versus $\log_{10}$ weight for beef trim pieces with an external surface area.

This model is quite different to the previous models | the intercept is essentially zero and the slope is larger than those previously observed. These observations are consistent with this lot having larger pieces with little surface area, compared to previous lots. This may be due to the nature of the carcases coming from bulls rather than cows.
Since the linear model was _tted on the log scale, the results can also be presented as percentage changes. This means that a 1% increase in Weight is associated with a 0.85% increase in external surface area.
In relation to the summary of total external surface area in each carton, presented at the beginning of this section, this model seems to result in reasonable estimates of the total surface area in a carton.
This can be seen from the following summary which presents the sum of the estimated surface area (from the weight) for all external pieces.

1	2	3	4	5	6	7	8	9	10	11	12
6730	6459	5878	6638	6505	7013	6687	6474	6515	6760	6935	7098

3.5.3 Contamination

This lot resulted in a total of 20 positive E. coli O157 tests | the number of positive samples per carton is summarised below.

1	2	3	4	5	6	7	8	9	10	11	12
0	8	0	4	4	0	0	1	0	1	1	1

An overview of the lot is given in Figure 22. Currently still awaiting information about the location of cartons in the lot assumed at present that cartons were consecutive, starting with 7601. From the location of the cartons within this lot it is fairly clear that cartons were not randomly sampled. It appears that, at best, a random starting point was selected and that consecutive cartons (except were the operator couldn't keep up with production) were sampled as they came of the production line.

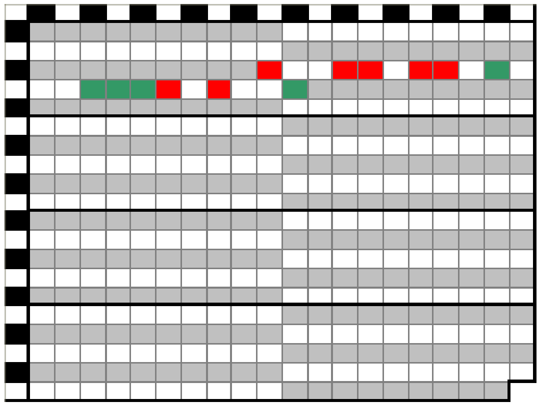


Figure 22: Diagram of location of cartons in the lot, numbered from left to right and top to bottom. Each square represents a carton in the lot — sampled cartons are presented in green and cartons that tested positive are presented as red (top row and left column are include to make counting easier).

Table 3: MPN calculations on per g and per cm^2 basis for the number of positive samples per carton observed in Lot E

+ve	MPN (per g)	Upper 95 (per g)	MPN (per cm^2)	Upper 95 (per cm^2)
1	0.003	0.012	0.001	0.006
4	0.011	0.025	0.005	0.013
8	0.023	0.042	0.011	0.021

Using the MPN calculations presented in Section 2.2 gives the per g and per cm2 based results shown in Table 3.
As stated earlier, there were 12 cartons in this lot, and 7 of the 12 cartons tested positive during the intensive investigation after previously having tested positive with the an undisclosed screening test.
Since the 12 cartons were produced over a 10 minute period, it is unlikely that they represent a random sample of the whole lot. This make it difficult to determine how widespread the contamination was in this lot.

3.6 All Lots Combined

In this section the information from all lots (excluding the two lots of veal trim) is combined to get an overall picture of lots that are positive for E. coli O157. The results in this section are of particular interest in relation to the process model that has been developed for E. coli O157 in the production beef trim.
A summary of the pieces of beef trim in each carton is given below. From the Quantile-Quantile Plot in Figure 23 it can be seen that the number of pieces per carton could reasonably be model by a Poisson distribution with mean λ = 51 pieces per carton[8].

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
7.00   16.50   41.00   51.16   76.50  192.00
```

[8] Most points fall within the `confidence envelopes' which are indicated by the dashed lines.

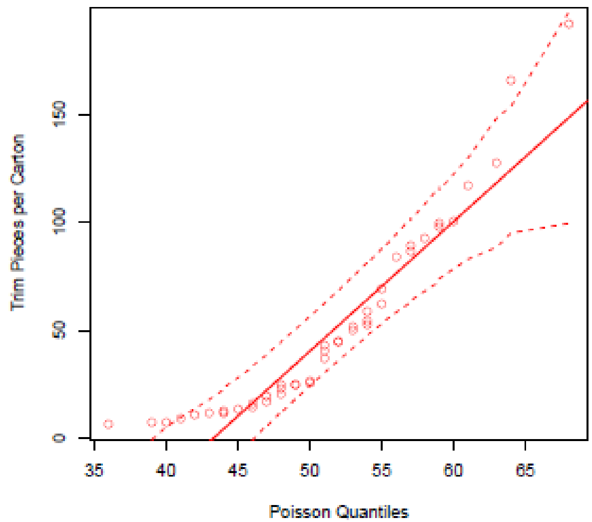


Figure 23: Poisson Q-Q plot of the number of beef trim pieces in a carton.

3.6.1 Weight

A summary of the total weight of beef trim pieces in a carton is given below. This indicates that in general cartons contain close to the 27.2 kg of beef trim.

```
> wcl <- as.vector(with(Pieces.lot, tapply(weight, list(lot, plant.carton),
+     FUN = sum)))
> summary(wcl[!is.na(wcl)])

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  26600   26900   26970   27000   27090   27710
```

The following is a summary of the weight and $\log_{10}$ weight of the pieces across all sampled cartons. Histograms of these are shown in Figure 24.

```
> with(Pieces.lot, summary(weight))

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    3.0    65.0   183.0   527.7   582.0  8215.0

> with(Pieces.lot, summary(log10(weight)))

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.4771  1.8130  2.2620  2.2680  2.7650  3.9150
```

While the histogram of weight in Figure 24(a) shows a right-skewed distribution, the histogram of $\log_{10}$ weight in Figure 24(b) shows a symmetrical distribution, which is not too far from a Normal distribution. This was investigated with the normal quantile-quantile plot (Figure 25) which shows departures from normality near the extremes (high and low). This indicates that extreme log10 weights (high and low) are observed less often than expected from normally distribution data.

Nevertheless, from a practical perspective, the distribution of log10 weight may reasonable be approximated by a Normal distribution with mean 2.27 $\log_{10}$ g and standard deviation 0.66 $\log_{10}$ g.

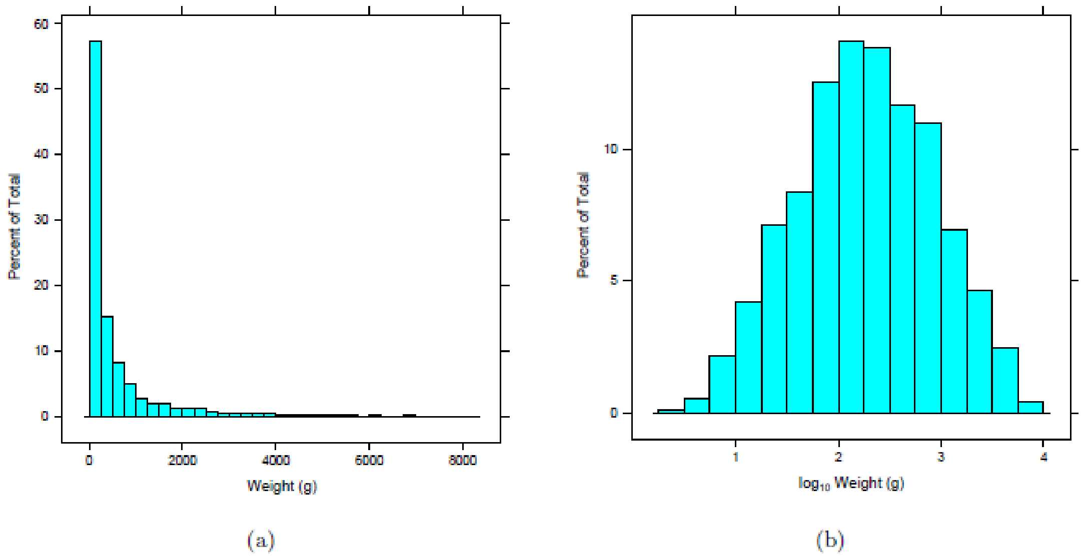


Figure 24: Histogram of the weight of beef trim pieces — (a) original scale and (b) $\log_{10}$ transformed.

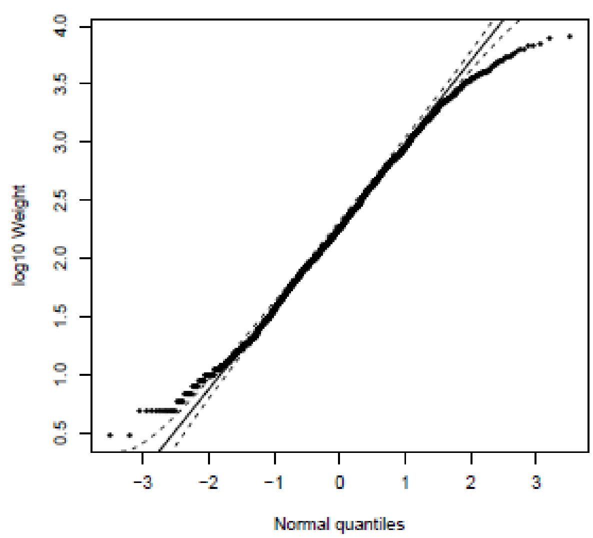
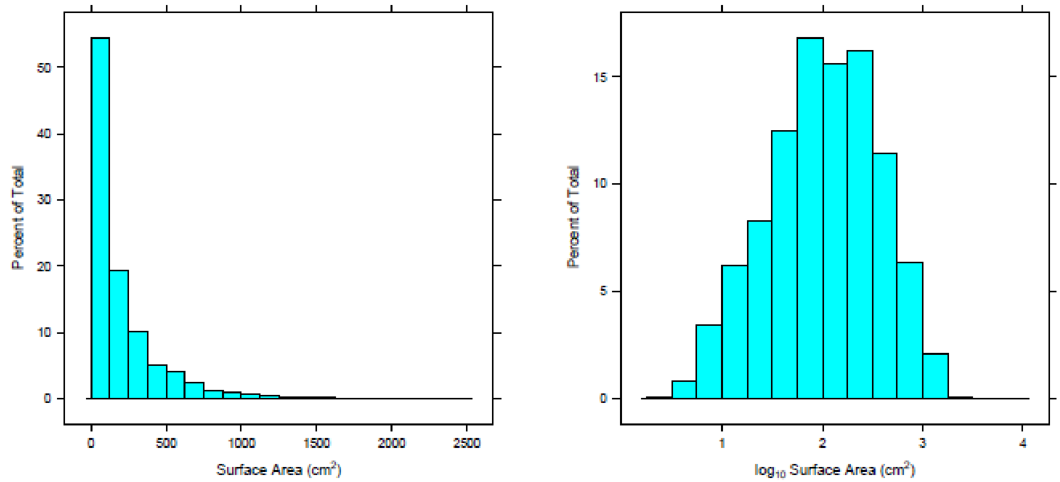


Figure 25: Normal Q-Q plot of the $\log_{10}$ weight of beef trim pieces.



(a) (b)

Figure 26: Histogram of external surface area of external pieces — (a) original scale and (b) $\log_{10}$ transformed.

3.6.2 External Surface Area

Across all lots there were 2200 pieces, of which 1754 were external pieces. A summary of the total external surface area (cm2) per carton is given below.

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
4365    6928    8353    8324    9832   12490
```

A summary of the proportion of pieces per carton with external surface area, irrespective of the amount, is presented below.

```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.5078  0.7700  0.8605  0.8272  0.9287  1.0000
```

Summaries of the external surface area of the pieces across all lots and their log10 values are given below and histograms of the external surface area and the log10 of external surface area are shown in Figure 26.

```
> with(Pieces.lot.ext, summary(surface))

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   3.00   42.25  108.00  204.10  257.50 2460.00

> with(Pieces.lot.ext, summary(log10(surface)))

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.4771  1.6260  2.0330  2.0070  2.4110  3.3910
```

Given the shape of the histogram of the log10 of the external surface area in Figure 26(b), a normal distribution for external surface area could be a reasonable approximation. The Normal Q-Q plot is shown in Figure 27, which, as for weight, indicates that very small and very large surface areas

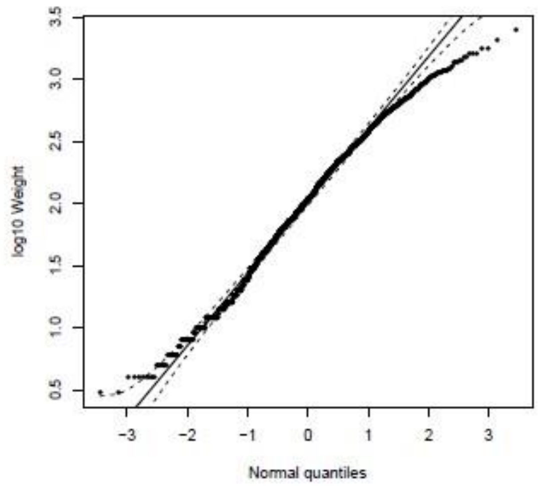


Figure 27: Normal Q-Q plot of the $\log_{10}$ external surface are of external beef trim pieces.

occur less frequently than would be expected from normally distributed data. Nevertheless, from practical perspective, the distribution of log10 external surface area may reasonable be approximated by a Normal distribution with mean 2.01 log10 g and standard deviation 0.55 log10 g. For external pieces, the relationship between weight and external surface area was again investigated.

The scatter plot of the log10 transformation of these two variables is shown in Figure 28.
In general, the linear regression model appears to fit the data reasonably well | the same residual problems are observed as for Lot 4, which obviously influenced the fit. Again, it can be expected that this would not affect the fit too much, but the inference that might be drawn. The model summary is given below.

```
Call:
lm(formula = log10(surface) ~ log10(weight), data = Pieces.lot.ext)
```

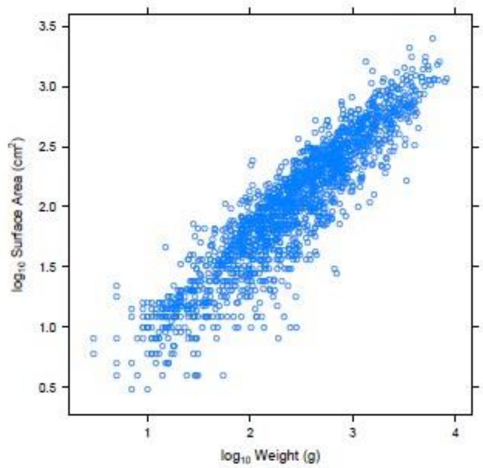


Figure 28: Scatter plot of the $\log_{10}$ external surface area versus $\log_{10}$ weight for beef trim pieces with an external surface area.

```
Residuals:
     Min       1Q   Median       3Q      Max
-1.04614 -0.11627  0.02756  0.14821  0.66987

Coefficients:
              Estimate Std. Error t value      Pr(>|t|)
(Intercept)   0.123841   0.020404   6.069 0.00000000157
log10(weight) 0.785026   0.008217  95.542       < 2e-16

Residual standard error: 0.2213 on 1752 degrees of freedom
Multiple R-squared: 0.839,       Adjusted R-squared: 0.8389
F-statistic:  9128 on 1 and 1752 DF,  p-value: < 2.2e-16
```

The output from the regression model indicates that the estimated line is

$$\log_{10}(\text{External Surface Area}) = 0.12 + 0.79\log_{10}(\text{Weight}) ,$$

or on the original scale

$$\text{External Surface Area} = 10^{0.12} \times \text{Weight}^{0.79} .$$

Since the linear model was fitted on the log scale, the results can also be presented as percentage changes. This means that a 1% increase in Weight is associated with a 0.79% increase in external surface area.

In relation to the summary of total external surface area in each carton, presented at the beginning of this section, this model seems to result in reasonable estimates of the total surface area in a carton.

This can be seen from the following summary which presents the sum of the estimated surface area (from the weight) for all external pieces (each row represents the lot in the leftmost column and cartons are presented by columns).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A	7825	6990	6310	5757	5763	7528	6974	5787	6892	5473	8688	6803	NA	NA	NA	NA	NA	NA	NA
D	9502	9562	9262	8710	9363	8368	6933	7133	7364	8712	7839	8413	7890	9803	8440	9418	8331	8656	9099
E	6733	6521	5858	6899	6426	7333	6728	6405	6451	6974	7117	7294	NA	NA	NA	NA	NA	NA	NA

3.6.3 Contamination

This lot resulted in a total of 20 positive E. coli O157 tests - the number of positive samples per carton and lot is summarised below.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A	0	0	0	0	0	0	0	0	0	0	0	0	NA	NA	NA	NA	NA	NA	NA
D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	8	0	4	4	0	0	1	0	1	1	1	NA	NA	NA	NA	NA	NA	NA

Using the MPN calculations presented in Section 2.2 gives the per g and per cm2 based results shown in Table 4.
There were 43 cartons tested across all lots, and 7 of these tested positive during the intensive investigation. It should be noted that all positive cartons originated from a Lot E - all cartons sampled

Table 4: MPN calculations on per g and per cm^2 basis for the number of positive samples per carton observed in all Lots.

+ve	MPN (per g)	Upper 95 (per g)	MPN (per cm^2)	Upper 95 (per cm^2)
1	0.003	0.012	0.001	0.006
4	0.011	0.025	0.005	0.013
8	0.023	0.042	0.011	0.021

were produced over a 10 minute period. Consequently, it could be considered appropriate to ignoring this lot, which would result in a very conservative prevalence estimate of zero positive cartons out of 12 tested, i.e., 26.5%,[9] while a more aggressive estimate would be 1 positive out of 528 cartons produced (size of the smaller lot), i.e., 1.1%.[10]

In the past, the approach by Habraken et al. (1986) has been used to estimate the probability of accepting / rejecting lots under the current sampling scheme. This was done without specific

[9] The upper bound for the 95% con_dence interval, based on the Binomial distribution with n = 31.
[10] The upper bound for the 95% con_dence interval, based on the Binomial distribution with n = 528 and having observed 1 positive cartons | one from the smaller lot. That is after all why the lot was included in this intensive testing

information on the proportion of a lot that is contaminated nor the concentration of contamination in that part of the lot. However, the results from this project now allow us to undertake those calculations in a more \educated" way.

As a worst case scenario, assume that the concentration (MPN per g) is given by the upper 95% confidence bound for 8 positives samples per carton (Table 4) and that the prevalence is as estimated above. Then under the two prevalence scenarios, we could expect that 0.017% and 0.416% of lots would be rejected, when five 6.5 g samples are removed from each of 12 cartons (60 samples of 375 g total weight).

4 References

Anonymous. 2008. AQIS Meat Notice 2008/05: Escherichia coli O157:H7 testing of raw ground beef components destined for export to the US and US Territories.

Habraken, C. J. M., D. A. A. Mossel, and S. van der Reek. 1986. Management of Salmonella risks in the production of powered milk products, Netherlands Milk and Dairy Journal 40, 99{116.