

Preliminary findings in analysis of  
Kathariou's msmnts of bacterial colonization  
Consultant: Jason A. Osborne

Clients: Sophia Kathariou and Roger Thompson

To investigate whether colonization rates vary by farm company, flocks were taken as sampling units in the initial analysis of the experiment. If any of the 5 (or 6) birds in flock  $i$  from company  $j$  was colonized by any of the species of bacteria and the colonization was detected by any of the 2 (or 3) samples taken on the bird, the response is coded  $y_{i(j)} = 1$  and is coded  $y_{i(j)} = 0$  otherwise:

$$y_{i(j)} = \begin{cases} 1 & \text{flock } i \text{ in co. } j \text{ colonized} \\ 0 & \text{no colonization detected in flock } i \text{ in co. } j \end{cases}$$

Birds were counted as having bacteria of interest if they were not coded as species=4.

The probability of colonization for observation  $y_{ij}$  is denoted  $p_{ij}$ . There were 30 flocks from each of two companies. The number and proportion of flocks colonized within each company are given in the table below:

The SAS System  
The FREQ Procedure

1

Table of colonies by FarmCo

colonies	FarmCo		
Frequency	1	2	Total
Col Pct			
0	4	5	9
	13.33	16.67	
1	26	25	51
	86.67	83.33	
Total	30	30	60

These two flock colonization rates, 87% for company 1 and 83% for company 2, are not significantly different.

A more naive (and less defensible) approach is to ignore flock effects, so that birds, rather than flocks, are taken as the experimental units. Under this approach, the individual bird colonizations are modelled as a Bernoulli sequence. The table below breaks down infected birds by company.

The SAS System  
The FREQ Procedure

1

Table of infected by FarmCo

infected	FarmCo		
Frequency			
Col Pct	1	2	Total
0	60	41	101
	40.00	27.15	
1	90	110	200
	60.00	72.85	
Total	150	151	301

Statistics for Table of infected by FarmCo

Statistic	DF	Value	Prob
Chi-Square	1	5.5710	0.0183
Likelihood Ratio Chi-Square	1	5.5957	0.0180
Continuity Adj. Chi-Square	1	5.0097	0.0252
Mantel-Haenszel Chi-Square	1	5.5525	0.0185

Fisher's Exact Test

Cell (1,1) Frequency (F)	60
Left-sided Pr <= F	0.9936
Right-sided Pr >= F	0.0125
Table Probability (P)	0.0061
Two-sided Pr <= P	0.0206

The infection rates for individual birds among the two companies differ significantly ( $p < 0.05$ ). The rate for company 1 is 60% and the rate for company 2 is 73% for a difference of 13%. A 95% confidence interval (assuming the appropriateness of the model) for the difference is (0.016, 0.241), so that very small differences between the companies infection rates are plausible, even if 0 isn't.

Rather than assignment of 0 or 1 to each flock for detection/non-detection of bacterial infection, scores could be assigned based on the proportion of infected birds within the flock. Such scores would be highly discrete, but ordinal, so that non-parametric methods are appropriate. One such method is the Mann-Whitney, or Wilcoxon procedure for comparing means from two independent samples. The mean score among the flocks within each farm company are given below, as is an approximate  $p$ -value from the Wilcoxon procedure. The mean scores for the two groups were about the same as the proportions of infected flocks among the two companies, 60% for Farm Co #1 and 73% for Farm Co #2. The difference was not statistically significant under the usual  $t$ -test ( $p = 0.23$ ) or under the Wilcoxon procedure  $p = 0.26$ .)

The SAS System

The NPAR1WAY Procedure

Analysis of Variance for Variable inf\_rate  
Classified by Variable FarmCo

Farm Co	N	Mean
1	30	0.600000
2	30	0.726667

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Among	1	0.240667	0.240667	1.4852	0.2279
Within	58	9.398667	0.162046		

Average scores were used for ties.

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable inf\_rate  
Classified by Variable FarmCo

Farm Co	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
1	30	843.50	915.0	62.556754	28.116667
2	30	986.50	915.0	62.556754	32.883333

Average scores were used for ties.

### Wilcoxon Two-Sample Test

Statistic 843.5000

#### Normal Approximation

Z -1.1350

One-Sided Pr < Z 0.1282

Two-Sided Pr > |Z| 0.2564

#### t Approximation

One-Sided Pr < Z 0.1305

Two-Sided Pr > |Z| 0.2610

Z includes a continuity correction of 0.5.

### Infection scores

