

Data set 8:

Barnacle Death

Background

In a classic paper published in *Ecology* in 1961, Joe Connell described the competitive interaction between two barnacle species, *Chthamalus stellatus* and *Semibalanus balanoides* (then called *Balanus balanoides*). These barnacles dominate the intertidal zone of wave-exposed rocky shores of the eastern Atlantic Ocean. They co-occur throughout the British Isles, and where they do *C. stellatus* adults are restricted to the high intertidal, despite the occurrence of juveniles lower in the intertidal. This habitat restriction apparently results from *C. stellatus* being out-competed for space by the larger *S. balanoides*, which are not as desiccation tolerant as *C. stellatus* and therefore not able to survive in the high intertidal zone.

Connell's study was conducted on the Isle of Cumbrae in the Firth of Clyde, Scotland. Part of it concerned the different lethal effects of *S. balanoides* on *C. stellatus* individuals. The fates of mapped individual *C. stellatus* were observed over one year, in several patches varying in height relative to mean tide level, including some barnacle-encrusted stones which were moved to lower in the intertidal zone than *C. stellatus* normally occurred. The causes of mortality of all dead *C. stellatus* were recorded as "smothered" by *Balanus*, "undercut" by *Balanus*, "crowded" by *Balanus*, or "unknown."

This handout uses data from three sites at which *C. stellatus* settlement was reasonably high and large numbers of deaths were recorded. These were Connell's site 3 (2.2 ft above mean tide level (MTL)), site 12 (at MTL), and site 13 (1 ft below MTL).

The question:

Did the relative importance of the various causes of death differ among these sites?

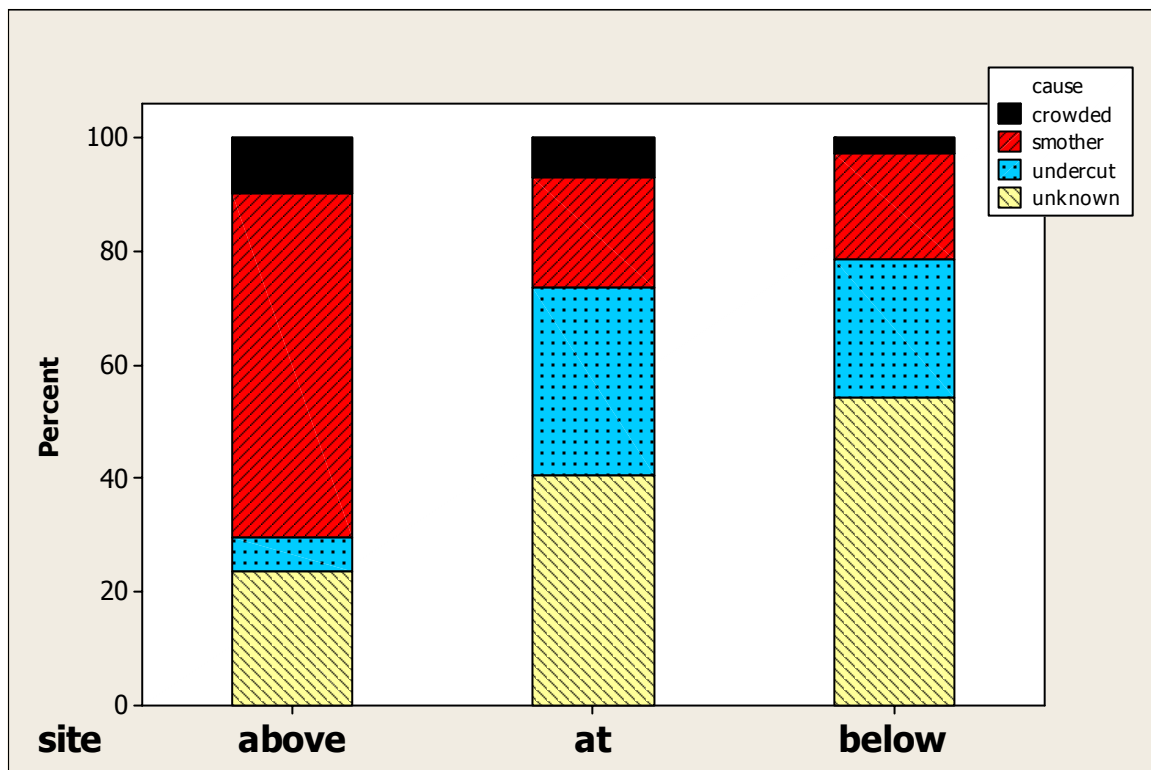


The data

	smother	undercut	crowded	unknown	total
above	49	5	8	19	81
at	11	19	4	23	57
below	13	17	2	38	70

Data exploration

Display: Stacked bar chart



Summary statistics: Percentages, per site

	smother	undercut	crowded	unknown
above	60.5	6.2	9.9	23.5
at	19.3	33.3	7.0	40.4
below	18.6	24.3	2.9	54.3

The bar chart simply shows these row percentages. Clearly smothering caused a much larger fraction of the deaths, and undercutting a much smaller fraction, at the above-tide-level site than at the other two sites. Crowding was rare at all three sites; deaths from unknown causes were fairly common, and more so the lower the site.

Inference

Sampling model and hypotheses

This is an example of sampling model 1: each site is a separate sample, and we are interested in comparing relative frequencies of the different sources of mortality across the samples.

$$\begin{aligned}
 H_0: \quad & P_{\text{smother(above)}} = P_{\text{smother(at)}} = P_{\text{smother(below)}} \\
 & P_{\text{undercut(above)}} = P_{\text{undercut(at)}} = P_{\text{undercut(below)}} \\
 & P_{\text{crowded(above)}} = P_{\text{crowded(at)}} = P_{\text{crowded(below)}} \\
 & P_{\text{unknown(above)}} = P_{\text{unknown(at)}} = P_{\text{unknown(below)}} \\
 H_a: \quad & \text{at least one of the preceding inequalities does not hold}
 \end{aligned}$$

Chi-square test of overall table

Cell entries are the observed count, expected count, and chi-square contribution. Since none of the expected frequencies are less than 1 and only 2 of 12 are less than 5 (and 1 of these only slightly so), the X^2 test should perform acceptably.

	smother	undercut	crowded	unknown	total
above	49	5	8	19	81
	28.43	15.97	5.45	31.15	
	14.89	7.53	1.19	<u>4.74</u>	
at	11	19	4	23	57
	20.00	11.24	3.84	21.92	
	<u>4.05</u>	<u>5.37</u>	0.01	0.05	
below	13	17	2	38	70
	24.57	13.80	4.71	26.92	
	<u>5.45</u>	0.74	1.56	<u>4.56</u>	

$$X^2 = 50.13 \quad df = 6 \quad P \ll 0.0005$$

The conclusion of the test is that there is strong evidence against the null hypothesis. **The probabilities of the different forms of mortality vary across the sites.**

Examination of the cell contributions to X^2 , like the data exploration above, indicates that the greatest differences are that:

- ‘smothering’ is much more frequent ‘above’ the tide line than ‘at’ or ‘below’ it
- ‘undercutting’ is less frequent ‘above’ and more frequent ‘at’ the tide line

A smaller difference is that deaths from unknown cause were increasingly frequent going from above to below the tide line; this could be due to losses resulting from (presumably) greater wave exposure and/or from the transplanted rocks the barnacles were on at the lower sites moving and crushing barnacles.

Partitioning the contingency table and overall X^2 test

Interpretation of the overall test can be examined further by partitioning the test into a set of smaller, independent tests, as shown in the following.

First partition:

This partition compares the probability that death was by **smothering**, between the **above** site and the other two sites combined.

	other	smother	Total
-----+	-----+	-----+	
above	32 53	49 28	81
-----+	-----+	-----+	
at_below	103 82	24 45	127
-----+	-----+	-----+	
Total	135	73	208

Statistic	DF	Value	Prob
-----	-----	-----	-----
Chi-Square	1	37.567	0.001
Fisher's Exact Test (2-Tail)			1.24E-09

There is very strong evidence the probability of smothering is higher at the above-tide-level site than at the other two sites.

Second partition:

This partition makes the analogous comparison between the **at** and **below** sites, omitting **above**.

	other	smother	Total
-----+	-----+	-----+	
at	46 46	11 11	57
-----+	-----+	-----+	
below	57 57	13 13	70
-----+	-----+	-----+	
Total	103	24	127

Statistic	DF	Value	Prob
-----	-----	-----	-----
Chi-Square	1	0.011	0.917
Fisher's Exact Test (2-Tail)			1.000

These two sites clearly do not differ much if at all in the probability that death is due to smothering: within rounding limits, the observed frequencies match the expected ones and all tests are very non-significant.

Third partition:

This partition compares the probability of death by **unknown** cause rather than by **undercutting** or **crowding** (*i.e.* the probability of an unknown cause, among deaths *not* caused by smothering), between the **above** site and the other two sites together.

	und_crwd	unknown	Total
above	13 13	19 19	32
at_below	42 42	61 61	103
Total	55	80	135

Statistic	DF	Value	Prob
Chi-Square	1	0.000	0.988
Fisher's Exact Test (2-Tail)			1.000

This test, like the preceding one, is as non-significant as it could be.

Fourth partition:

This partition compares the same sources of mortality as the previous one, but now between the **at** and **below** sites.

	und_crwd	unknown	Total
at	23 19	23 27	46
below	19 23	38 34	57
Total	42	61	103

Statistic	DF	Value	Prob
Chi-Square	1	2.928	0.087
Fisher's Exact Test (2-Tail)			0.108

In this case there is some suggestion, though certainly not clear evidence, that the probability that a cause of death will not be known (given that it is not smothering) is higher **below** than **at** tide level.

Fifth partition:

This partition compares the relative frequencies of death by **crowding** and by **undercutting**, between the **above** site and the other two sites combined.

	crowded	underct	Total
above	8 3.3	5 9.7	13
at_below	6 10.7	36 31	42
Total	14	41	55

Statistic	DF	Value	Prob
Chi-Square	1	11.681	0.001
Fisher's Exact Test (2-Tail)			0.0017

This test gives strong evidence that **crowding** is a more frequent cause of death, relative to undercutting, **above** tide level than **at** or **below** it. Due to the low expected value in the **above-crowded** cell (<5, in a 2x2 table), the X^2 test is suspect. In this case, however, the Fisher's Exact test confirms the results.

Sixth partition:

This partition makes the only remaining comparison: between relative frequencies of **crowding** and **undercutting**, at the **at** and **below** sites.

	crowded	underct	Total
at	4 3.3	19 20	23
below	2 2.7	17 16	19
Total	6	36	42

Statistic	DF	Value	Prob
Chi-Square	1	0.400	0.527
Fisher's Exact Test (2-Tail)			0.673

The Fisher's Exact test here indicates that there is no evidence the proportions of deaths by these two causes differ at these sites.

Partitioning – Conclusions and comments

This set of 6 partitions of the table shows the main differences in frequencies of causes of death between the sites:

- (1) deaths **above** tide level have a higher probability of being by **smothering** than do those **at** or **below** tide level;
- (2) of deaths caused by crowding or undercutting, **crowding** is relatively more frequent, and **undercutting** less frequent, **above** tide level than **at** or **below** it.

Note that the 6 partitions match the 6 degrees of freedom of the full table. It also can be confirmed by comparing the full table with the various partition tables, that each cell count and each marginal total in the full table appears once and only once in the subtables. In this sense the partitions are independent.

Since the hypotheses implied by these partitions were not pre-planned, and the partitions were selected after inspecting the data, the significance tests are not truly valid. Given the strong significance of the overall test, however, it is reasonable to conclude that there are some real differences in the probabilities, and that they most likely are the largest observed differences.