The Meaning of ‘Significance’

Prof. John Dover

Fancy a Gamble?

What odds would you bet on?

50% chance of being wrong?

20% chance of being wrong?

10% chance of being wrong?

5% chance of being wrong?

1% chance of being wrong?

0.1% chance of being wrong?

Would it depend on what you are betting for?

a horse winning a race?

Or your life?

Scientists are gamblers!

Generally, for most things, a 1 in 20 chance of an observation being wrong are good odds, 1 in 10 would be dodgy

Why all this stuff about odds? Because that is what we really mean when we talk about statistically significant results.

What we really mean is ‘what is the chance that our data are rubbish?’ Or perhaps more genteelly ‘ if we look at the means of (say) beetle numbers from two areas and see a difference, how likely is it that the observation reflects reality’

If we can work out the probability of our data being unreliable, then we can take that gamble (accept the observation as probably real) if the odds are in our favour. That is what Probability values from statistical analyses are all about.

If it is all about odds, why all this stuff about 0.05 then?

If we accept that 1 in 20 is good odds (I wouldn't bet my life on 1 in 20 though) then the following might help explain some of the confusion

1 in 20 is the same as 5%

5 divided by 100 (5/100) is the same as 1 divided by 20 (1/20)

And 5/100 = 5% (five hundredths)

Five hundredths is the same as 5%

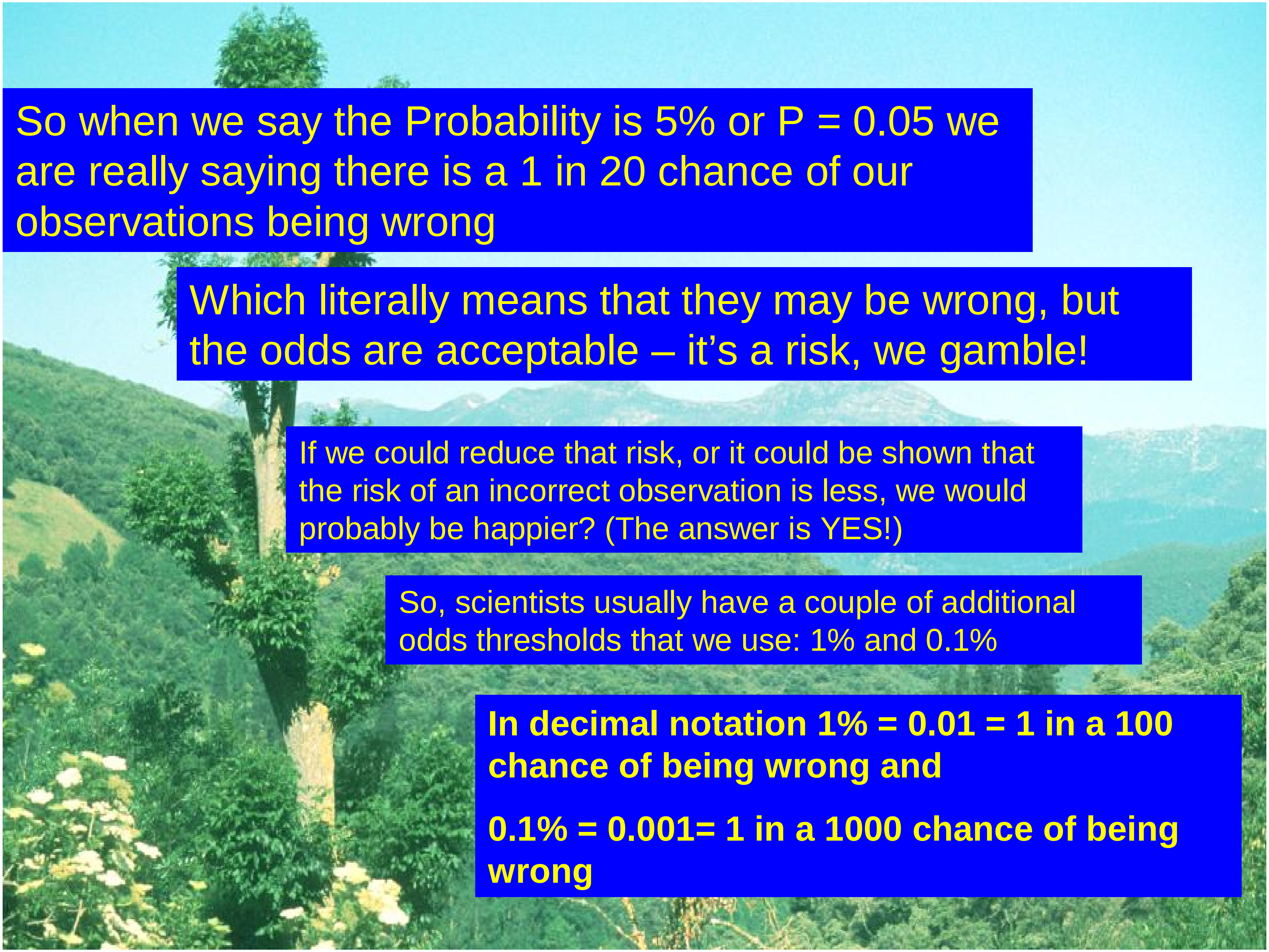
Put another way:

5/100 = 0.05 (try it on your calculator)

0.05

Tenths

Hundredths



So when we say the Probability is 5% or P = 0.05 we are really saying there is a 1 in 20 chance of our observations being wrong

Which literally means that they may be wrong, but the odds are acceptable – it's a risk, we gamble!

If we could reduce that risk, or it could be shown that the risk of an incorrect observation is less, we would probably be happier? (The answer is YES!)

So, scientists usually have a couple of additional odds thresholds that we use: 1% and 0.1%

In decimal notation 1% = 0.01 = 1 in a 100 chance of being wrong and

0.1% = 0.001= 1 in a 1000 chance of being wrong

How does this help you?

If you have a very low P-value e.g. 0.0001 then you can have a lot of confidence in your data.

1/10000 = 0.01% or 1 in 10,000 chance of your data being unreliable – great odds!!!

If you have a large P-value e.g. 0.276 – then your data are pretty unreliable – worse than a 1 in 4 chance of your data being unreliable!!!

276/1000 = 27.6% or 1 in 3.6 – rotten odds!!!

And actually that is all your P-values tell you

P-values simply tell you how much confidence you can have in your data

Getting a P-value is not the end of your job: you still have to explain what your data actually means!!!

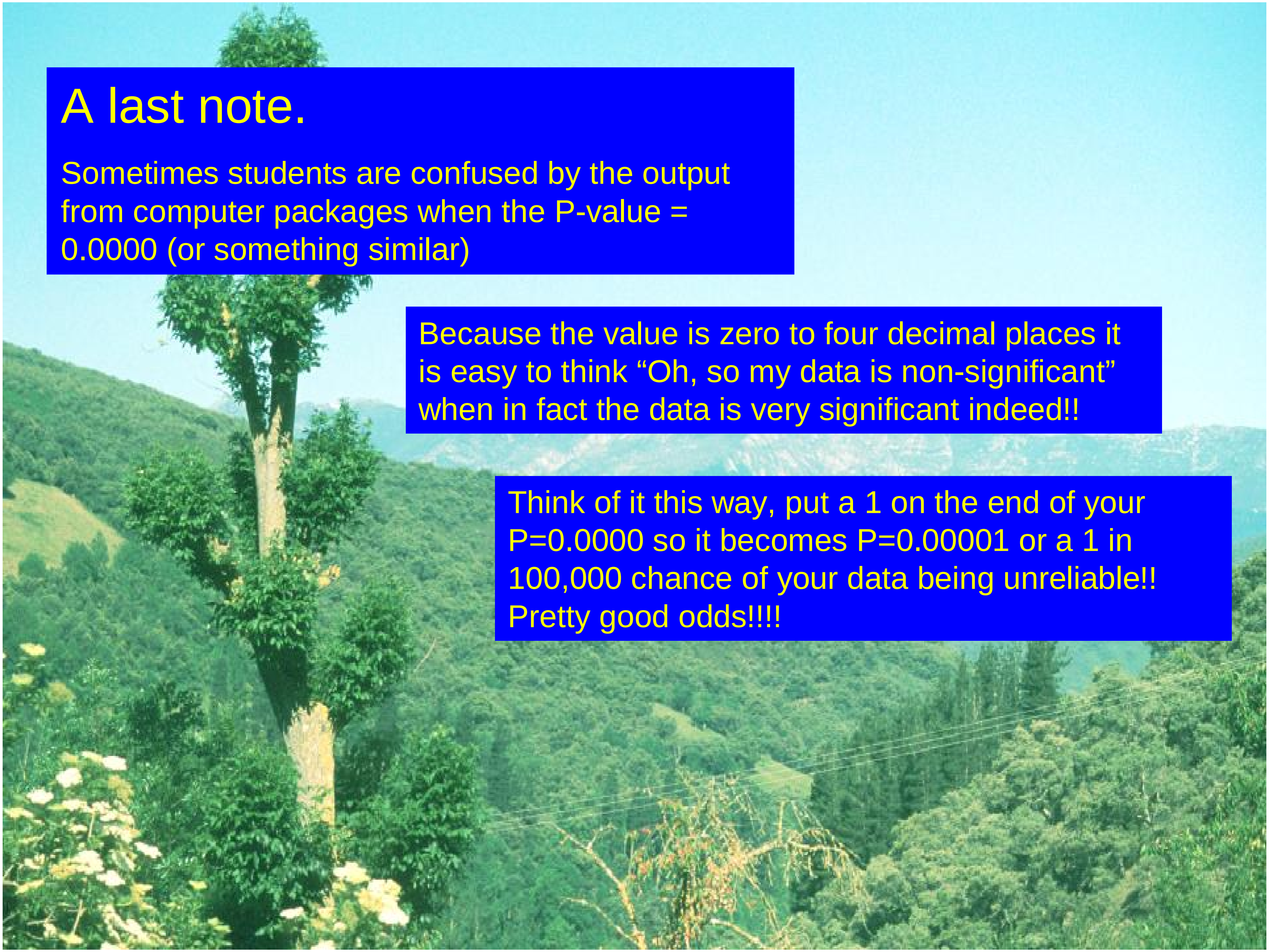
The significance thresholds we use are:

If the P-value is greater than 5% or P>0.05 (> means greater than) the differences between two (or more) data sets are considered non-significant so, for example, P=0.06 is non-significant or unreliable

If P=0.05 the differences are said to be ‘significant’

If P=0.01 the differences are ‘highly significant’

If P=0.001 the disfferences are ‘very highly significant’



A last note.

Sometimes students are confused by the output from computer packages when the P-value = 0.0000 (or something similar)

Because the value is zero to four decimal places it is easy to think “Oh, so my data is non-significant” when in fact the data is very significant indeed!!

Think of it this way, put a 1 on the end of your P=0.0000 so it becomes P=0.00001 or a 1 in 100,000 chance of your data being unreliable!! Pretty good odds!!!!

Thresholds are ancient history?

The thresholds were developed before electronic calculators, let alone computers, had been developed – so tables of probability were printed one page for 5%, one for 1% and another for 0.1%. It simply was too much effort (and would have cost a fortune) to work out all the probabilities in between those three ranges!

Now, with computers, exact probabilities (e.g. P=0.0045) can be almost instantly be calculated. So are the ranges redundant? Yes and no: they can be a useful reminder of relative value.

In many research papers you might see asterisks or letters used in data tables. If you look at the foot of these tables you will see something like: ns = P>0.05; * = P<0.05; ** = P<0.01; ***=P<0.005. In practice the statistics programmes will have calculated exact P-values, but the researchers will have simply reported the statistic e.g. a F, t or chi^2 and given the P-value threshold. So for example P=0.026 would be * because it was less than P=0.05, but more than P=0.01; similarly P=0.006 would be ** because it was less than P=0.01 but more than P=0.001

Table 4
Mean occurrence ± SE of the 18 most abundant bird species per 50 m of green lanes and 100 m of single hedgerows

Species	Lanes	Hedges	*P*
	Means ± SE	Means ± SE	
Blackbird	0.84 ± 0.11	0.56 ± 0.10	0.018
Song Thrush	0.08 ± 0.04	0.01 ± 0.01	0.004
Robin	0.65 ± 0.11	0.36 ± 0.07	0.074
Great tit	0.52 ± 0.07	0.29 ± 0.06	0.108
Blue tit	0.69 ± 0.12	0.28 ± 0.09	0.018
Long-tailed tit	0.15 ± 0.01	0.04 ± 0.02	0.132
Linnet	0.06 ± 0.04	0.05 ± 0.02	0.705
Chaffinch	0.66 ± 0.10	0.50 ± 0.10	0.074
Greenfinch	0.12 ± 0.04	0.05 ± 0.03	0.002
Goldfinch	0.03 ± 0.02	0.01 ± 0.01	1.000
Yellowhammer	0.07 ± 0.05	0.06 ± 0.04	0.257
Wren	0.76 ± 0.11	0.33 ± 0.07	0.000
Dunnock	0.10 ± 0.03	0.16 ± 0.05	0.796
Chiffchaff	0.41 ± 0.12	0.02 ± 0.02	0.001
Whitethroat	0.10 ± 0.03	0.04 ± 0.04	0.008
Blackcap	0.10 ± 0.03	0.01 ± 0.01	0.007
Magpie	0.09 ± 0.02	0.04 ± 0.02	0.033
Wood pigeon	0.50 ± 0.13	0.21 ± 0.06	0.157

Differences between lanes and hedgerows tested using Friedman's ANOVA.

Here exact P-values are being reported (*Biological Conservation 126 (2005) 540-547)*

5.1. Individual ha[illegible] factors

[I]n this study, the mean volume of the hedges was found to influence bird abundance, number of bird territories, and bird species richness in green lanes. Hedge size has been [illegible] bird abundan[illegible] Lack, 1992; [illegible] 2000). Again t[illegible] in habitat vol[illegible] have a lower [illegible] thinner coun[illegible] et al., 1995), a[illegible] species.

The presen[illegible] positive effect[illegible]

Here n, a, b and c are being used to indicate threshold values of singificance (*Agriculture, Ecosystems and Environment* 80 (2000) 227–242)

Table 3
Factors affecting the number of species of butterfly in the fi[illegible] margins and green lanes of the Warburton study area 1997[a,b]

All species				Open pop[illegible]tion species				Closed population species			
Variable name	Regression coefficient	r^2	F-value[c] at end	Variab[illegible] nam[illegible]	Regression coefficient	r^2	F-value[c] at end	Variable name	Regression coefficient	r^2	F-value[c] at end
Constant	−0.8230			[illegible]nstant	3.7016			Constant	−1.3913		
Inside GL	3.5226	25.25	33.25 c	Ragwort stems	4.6767	17.23	6.28 a	Inside GL	1.2873	31.83	12.84 c
Outside GL	1.9430	49.23	14.54 c	Grass bank	−1.6339	29.57	18.57 c	Rough grass	2.5679	48.24	18.89 c
Ragwort stems	6.8031	57.07	5.27 a	Single hedge	−1.2059	42.97	11.51 b	Ragwort stems	3.8841	59.05	6.35 a
Rough grass	2.8669	61.62	6.17 a					Grass bank	−0.6167	63.92	4.97 a
Log length	1.6975	64.88	4.37 a					Log length	0.9772	67.13	5.27 a
								Bramble	0.8464	69.72	3.93 n

[a] Length of transect section transformed by $\log(x+1)$ prior to analysis.
[b] n—P>0.05; a—P<0.05, b—P<0.01, c—P<0.001. [illegible]
[c] Residual ms from final model used as denominator.

Explanation of codes here