Statistical Analysis of Form

- ***Objectives:***

Review statistical testing / significance determination

Discuss example of real-life use of Index of Dispersion

- ***Learning Outcomes:***

Review how p values and significance are determined

Understand steps involved in testing for significant Intensity

Practice approach used in multi-scaling analysis of Intensity

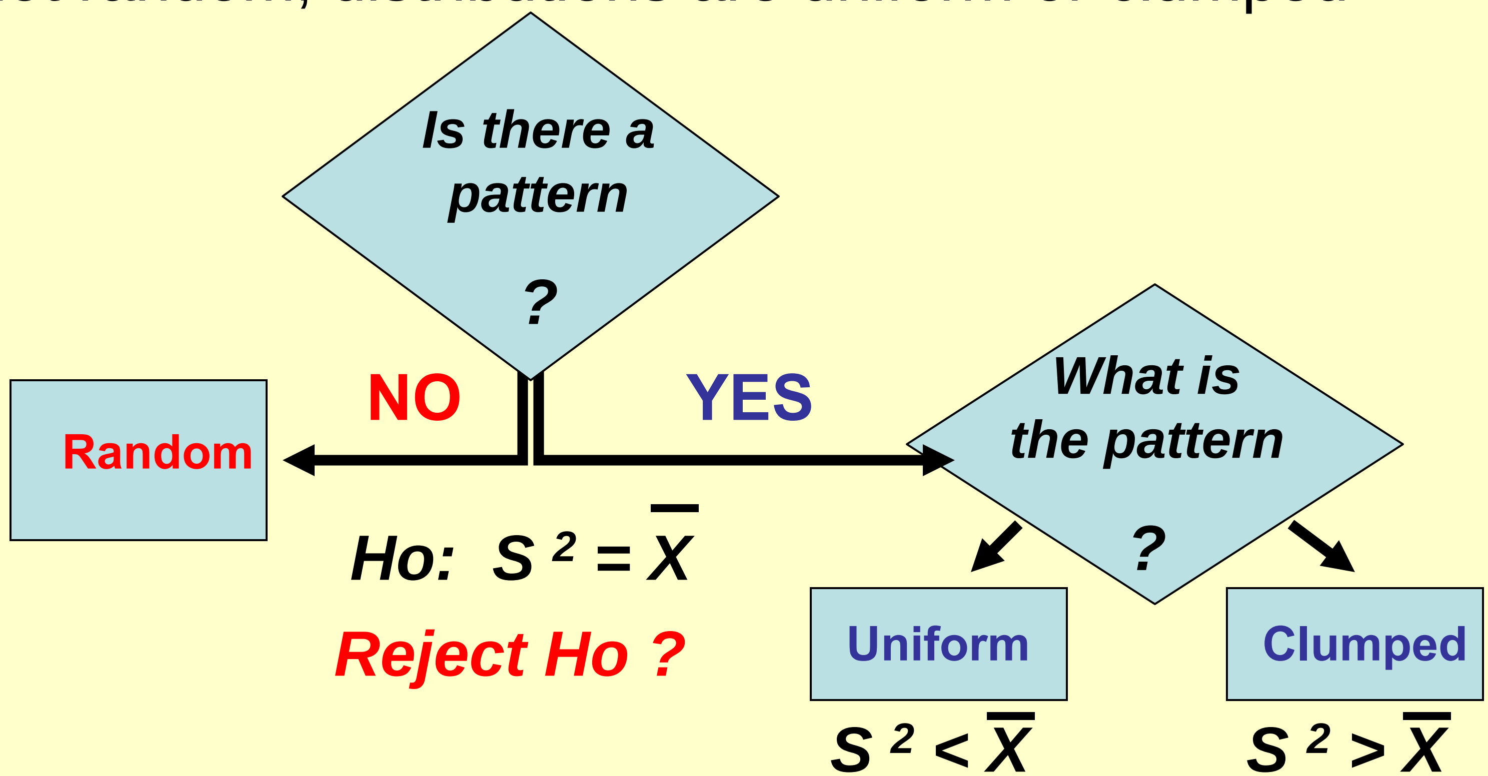
Appreciate link between analysis of *Intensity* and *Form*

Review interpretation of the correlation coefficient (r)

Quantifying Intensity

Relate the distribution of counts (e.g., organisms, sightings) across observations (e.g., samples, transects)

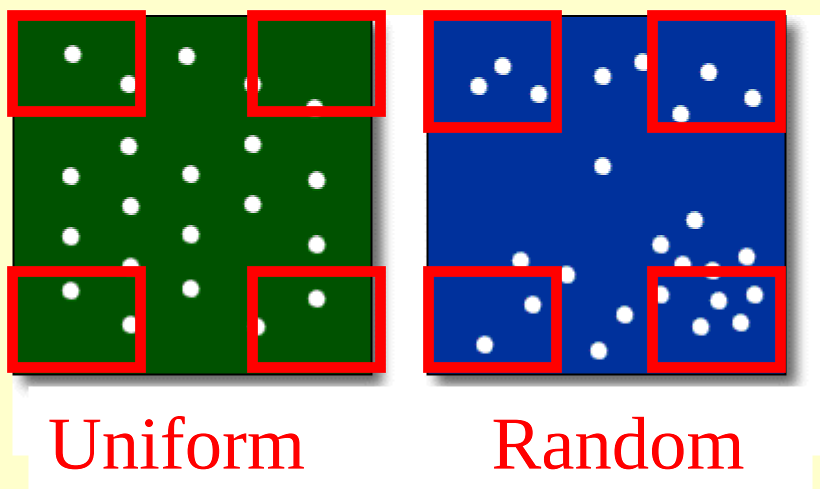
Classifies distributions into random / non-random.
If not random, distributions are uniform or clumped



Quantifying Intensity

- Index of Dispersion (I) = Variance / Mean

$$s^2 = \frac{\sum (z_i - \bar{z})^2}{(n-1)} \Bigg/ \bar{z} = \frac{\sum (z_i)}{n}$$



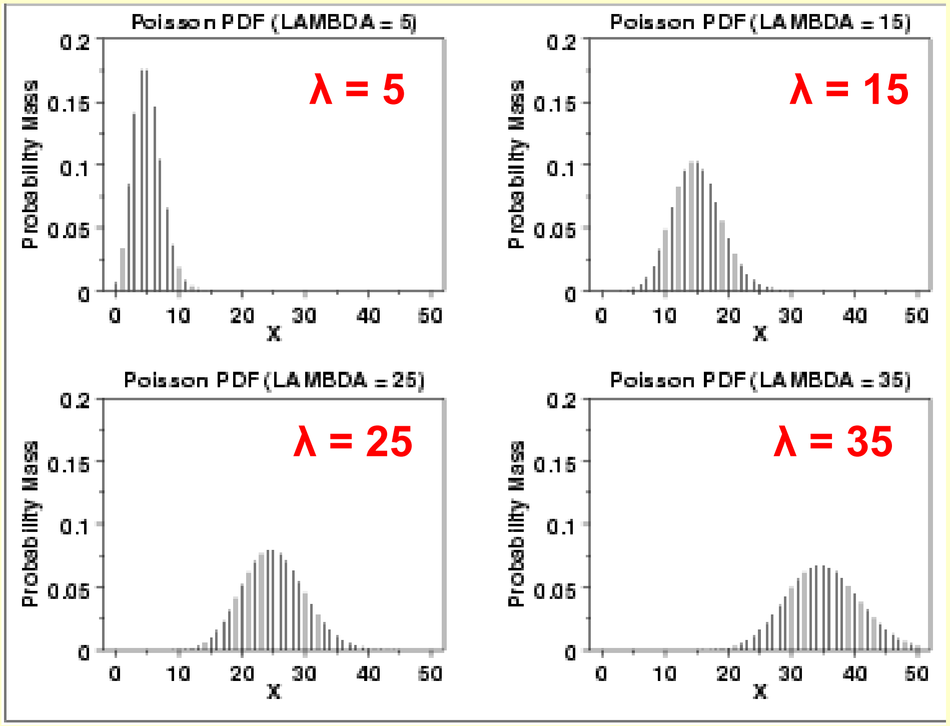
V: 0.00
M: 2.00
I: 0.00

V: 1.58
M: 3.50
I: 0.85

Poisson Distribution - Chi-Square

Spatial pattern is usually represented by the distribution of counts per unit area, volume, time, individual host or site. The simplest model of dispersion is a hypothesis of randomness, described by the Poisson probability distribution:

$$P[X=x]=\frac{e^{-\lambda}\lambda^x}{x!}, \qquad \text{for } x=0, 1, \ldots \quad (\lambda>0).$$

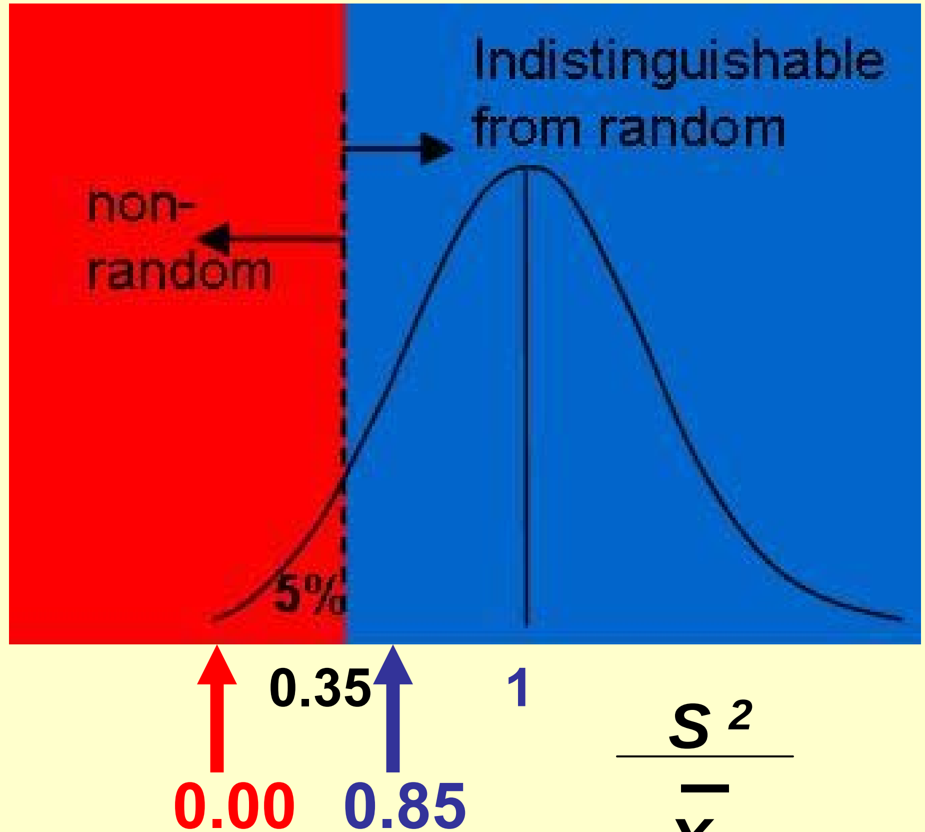


Lambda (λ) = Estimated Mean for quadrats sampled

(Johnson & Petkau 1995)

Statistical Significance - Approach

- Use Chi-Square Distribution to Assess Statistical Significance



$$\frac{S^2}{\bar{X}}$$

- Interpretation of Statistical Results

Real situation	Conclusion	
	H_0 is not rejected	H_0 is rejected
H_0 is true	good decision	type I error
H_0 is not true	type II error	good decision

Statistical Significance – p values

Uniform	Random
V: 0.00	V: 1.58
M: 2.00	M: 3.50
I: 0.00	I: 0.85

- **How many degrees of freedom do we have ?**

(Hint: Number of samples (4) minus 1) **3**

Degrees of Freedom	Probability, *p*				
	0.99	**0.95**	**0.05**	**0.01**	**0.001**
3	0.115	0.352	7.82	11.35	16.27

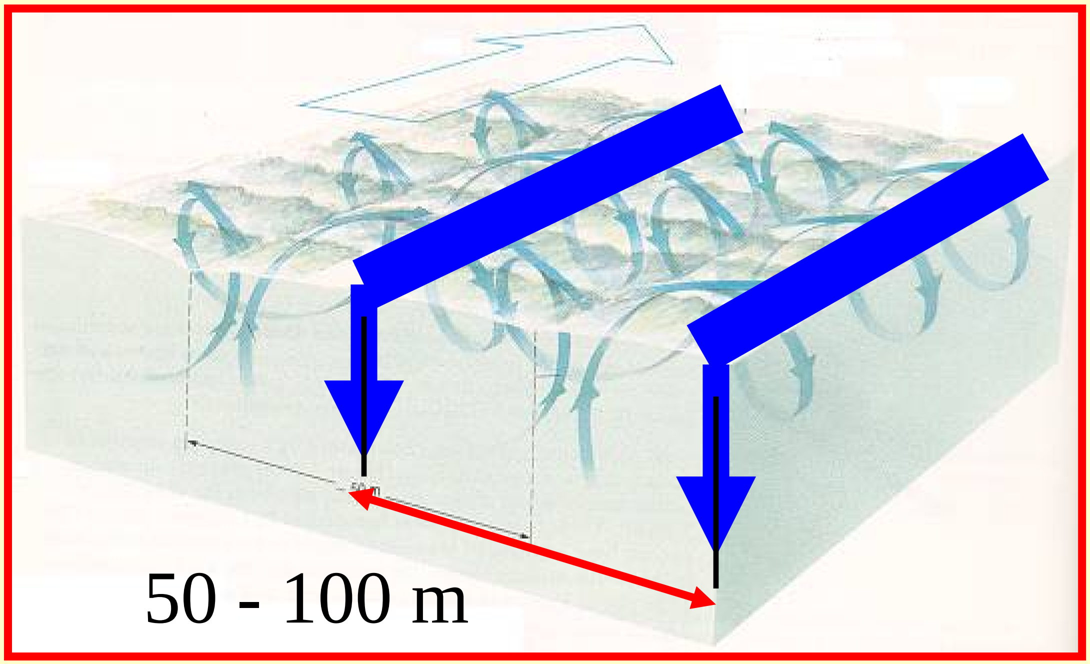
0.00 **0.85**

(Fisher et al. 1922)

Dispersion Index - Real Life Example

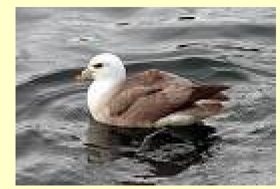
Top View

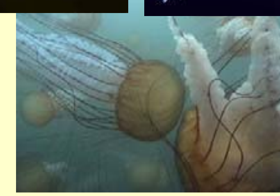
Cross Section



- Jellies aggregated by Langmuir cells

- Dense food source for northern fulmars





Up to:

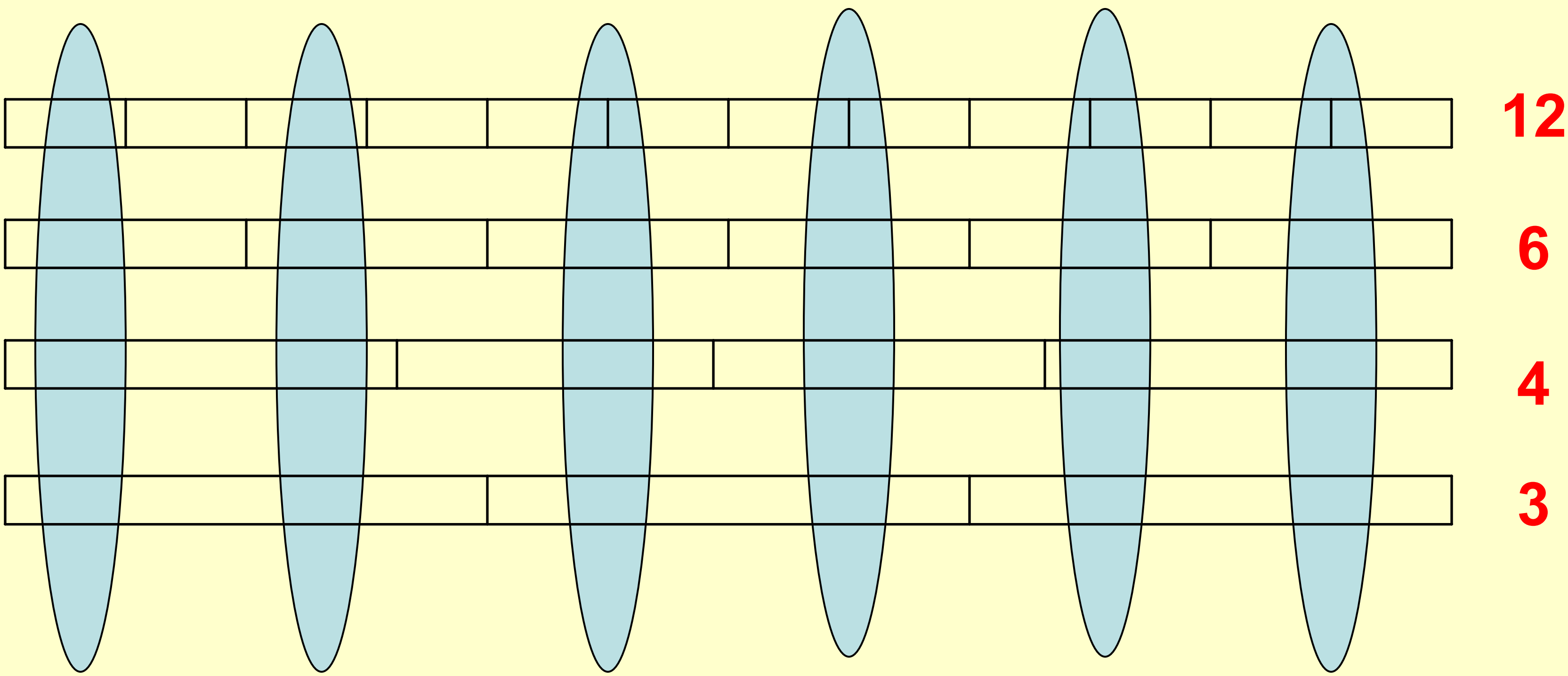
1000 / m ^{3}

(Hamner & Schneider 1986)

Variance / Mean – Multi-scale Analysis

Rationale:

A relatively low coefficient of dispersion (<1) was expected at interval lengths equal to any regular spacing of jellyfish. The results of this analysis were checked against a time-series analysis (Box and Jenkins 1976).



Variance / Mean – Multi-scale Analysis

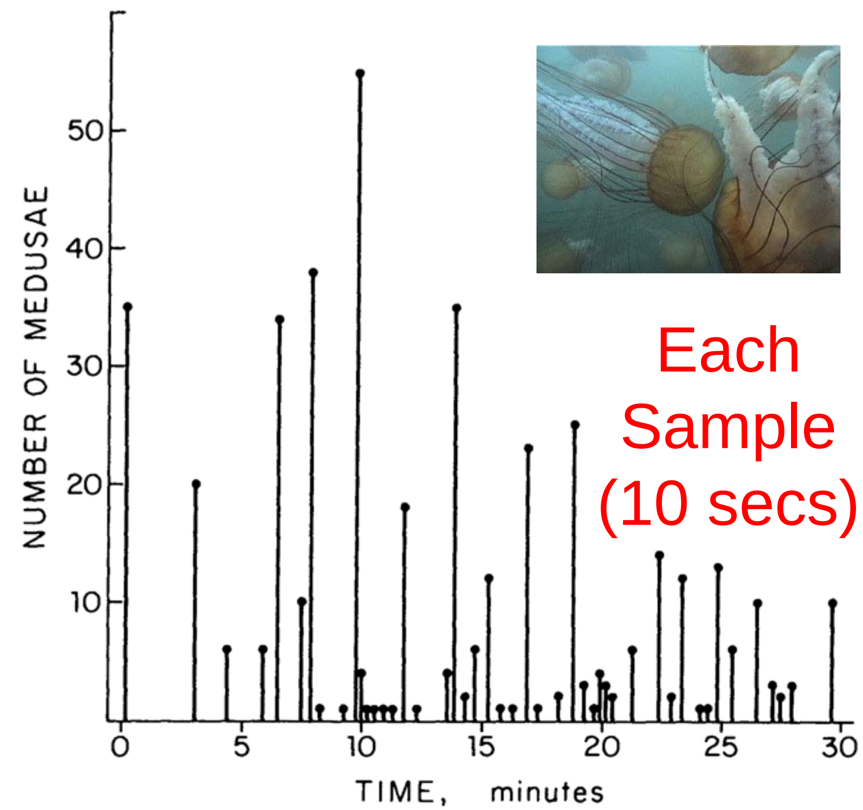


Fig. 1. Number of medusae that drifted past the ship within 2 m of the hull over a 30-min period on 2 August 1982 at PROBES station 6.

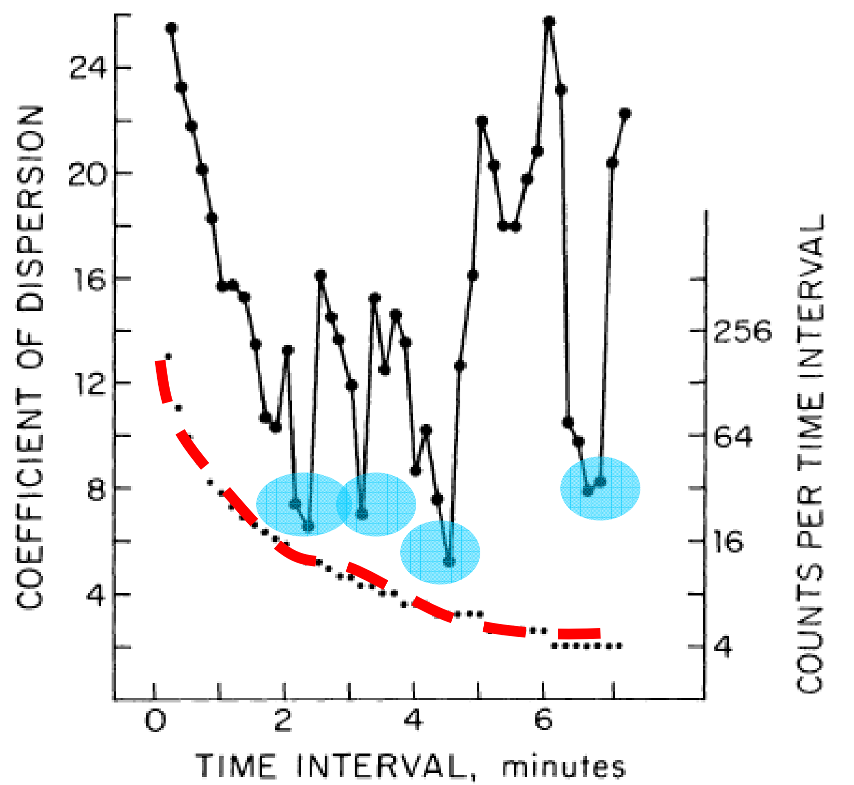


Fig. 2. Poisson analysis for regularity in the spacing of large numbers in the data from Fig. 1. Low coefficients of dispersion (variance-to-mean ratio) at 2 min 20 s, 4 min 30 s, and 6 min 40 s indicate highly regular spacing of large counts of medusae at distances about 129 m apart (*see text*).

Variance / Mean – Multi-scale Analysis

Results: Regularly spaced rows of jellies

Spacing of closest rows ~ 130 m

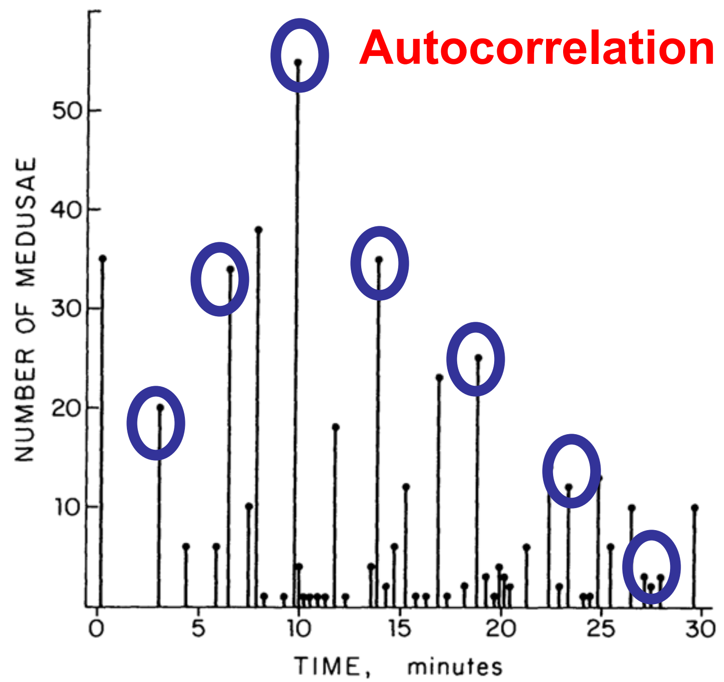


Fig. 1. Number of medusae that drifted past the ship within 2 m of the hull over a 30-min period on 2 August 1982 at PROBES station 6.

Results – Intensity:

Minimum values of the coefficient of dispersion (Fig. 2) occurred at time intervals of 2 min 20 s, 4 min 30 s, and 6 min 40 s. The spacing interval was slightly <2 min 20 s (with low variance-to-mean ratios recurring at even multiples of this interval).

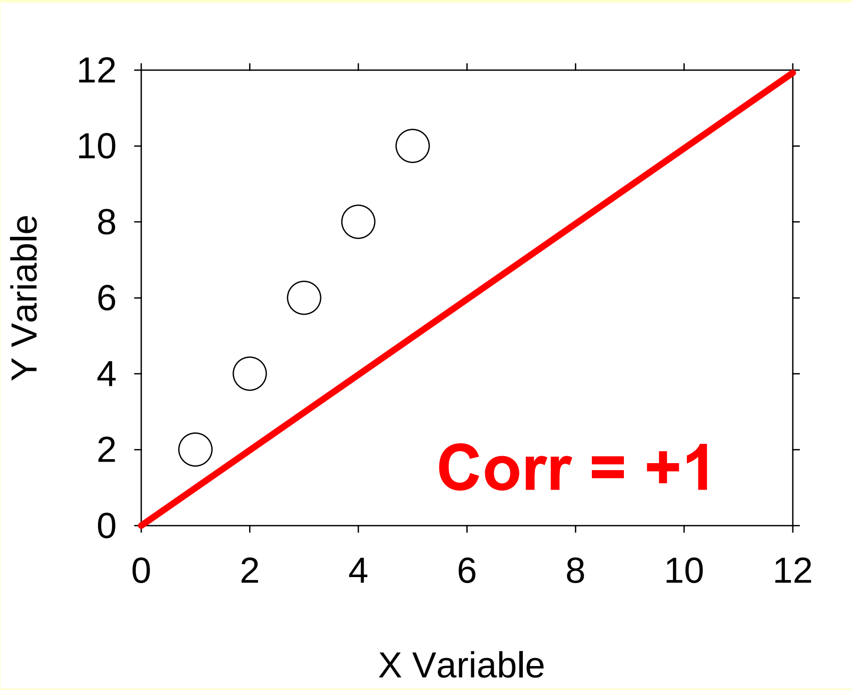
Results – Form:

A time-series analysis of the data gave similar results. Significant autocorrelation ($P = 0.05$) occurred at lags of 80 s, 130 s, and 240 s. The largest coefficient occurred at a lag of 130 s.

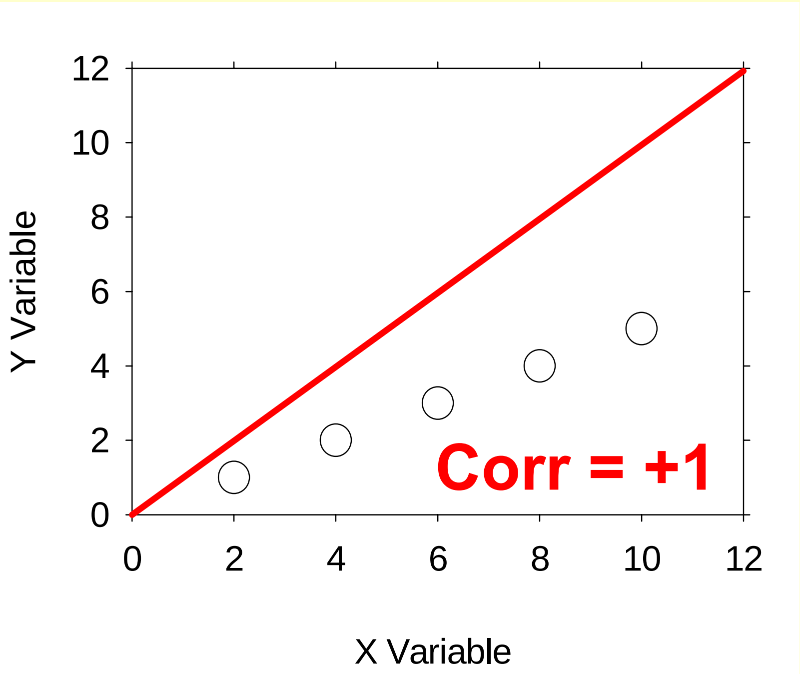
Autocorrelation – Background

The correlation coefficient (-1 ≤ r ≤ 1) indicates strength and direction of a linear relationship between two variables

X = 1,2,3,4,5   Y = 2,4,6,8,10



Y = 1,2,3,4,5   x = 2,4,6,8,10



How about these examples:

X = 5 4,3,2,1   Y = 2,4,6,8,10

Y = 1,2,3,4,5   x = 10,8,6,4,2

Autocorrelation – Quantifying Form

Informally: "Everything is related to everything else, but closer things are more related than distant things …"

Formally: "Systematic spatial variation in a mapped variable"

(Cliff & Ord 1981)

Defining Autocorrelation:

Property of random variables taking values, at pairs of locations a given distance apart, that are more similar (+) or dissimilar (-) than expected from randomly associated observations

Basically…

values closer together are more predictable than farther apart