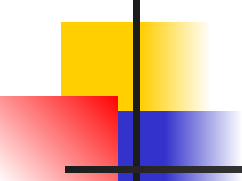


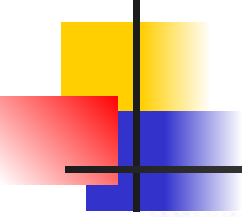


Тема лекций 10. Корреляционный анализ. Часть 11.



Бисериальный, тетракорический и поликорический показатели связи. Рассмотренные выше способы вычисления коэффициента корреляции употребляются в отношении признаков, сильно варьирующих и состояние которых можно измерить, взвесить и выразить в определенных цифровых показателях. Однако бывают случаи, касающиеся качественных признаков, когда требуется лишь констатировать наличие или отсутствие данного признака.

Поэтому многообразие связей между различными признаками, которые характеризуют генотипические, биохимические и физиологические особенности индивидуумов, приводит к необходимости установления силы связи между альтернативными (r_a), качественными (r_p) и между качественными и количественными (r_b) признаками.



Отличительной особенностью вышеперечисленных коэффициентов связей от коэффициентов корреляций между количественными признаками является то, что знак для них не имеет смыслового значения.

Выявление связи между простыми качественными признаками, имеющие четкое наследование, и полезными количественными признаками, имеющие полигенный тип наследования имеет для эколого-аналитической работы огромное значение.

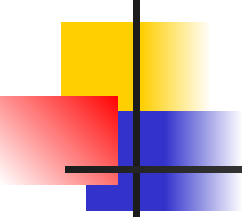
Бисериальный показатель связи. Бисериальный показатель связи (r_b) применяют при изучении связи между количественными и качественными признаками. Формула для вычисления r_b :



$$r_b = \frac{\frac{\sum f_+ \cdot a}{n_+} - \frac{\sum f \cdot a}{n}}{\sqrt{\frac{C}{n_+} - \frac{C}{n}}} \quad (50)$$

$$C = \sum f a^2 - \frac{(\sum f a)^2}{n} \quad (51)$$

где C - дисперсия количественного признака, вычисляемая по формуле 51. Подстрочный символ $+$ указывает на присутствие качественного признака.




Тетрахорический показатель связи. Коэффициент r_a применяют для определения связи между альтернативными признаками. При альтернативности признак имеет только два противоположных состояния. Поэтому признак называется альтернативным, когда он может проявиться лишь в двух состояниях (возможно двойное состояние: наличие той или иной окраски).

Для определения r_a необходимо произвести разnosку членов совокупности по клеткам четырехпольной корреляционной решетки (табл. 10.1). В такой решетке из-за возможного двойного состояния образуется только два класса.

Таблица 10.1

Четырехпольная корреляционная решетка для альтернативных признаков

| Класс – у | х | Класс х | | Итого |
|-----------|-------|-------------|-------------|-----------------------------|
| | у | 1 | 2 | |
| | 1 | f_1 | f_2 | $f_1 + f_2$ |
| | 2 | f_3 | f_4 | $f_3 + f_4$ |
| | Итого | $f_1 + f_3$ | $f_2 + f_4$ | $n = f_1 + f_2 + f_3 + f_4$ |



Формула для вычисления тетрафорического показателя связи:

$$r_a = \frac{f_1 f_4 - f_2 f_3}{\sqrt{(f_1 + f_2) \cdot (f_3 + f_4) \cdot (f_1 + f_3) \cdot (f_2 + f_4)}} \quad (52)$$

или

$$r_a = \frac{[f_1 f_4 - f_2 f_3]^{-0,5n}}{(f_1 + f_2) \cdot (f_3 + f_4) \cdot (f_1 + f_3) \cdot (f_2 + f_4)} \quad (53)$$

где f_1, f_2, f_3, f_4 - число членов совокупности по клеткам корреляционной решетки.

Формула 53 дает поправку на приближенность значений классов альтернативного признака. Квадратные скобки в числителе указывают на то, что берется абсолютная разница.

Полихорический показатель связи. Полихорический показатель связи (r_p) применяют при определении величины связи между двумя качественными признаками, имеющими несколько градаций.

Формула для вычисления r_p :

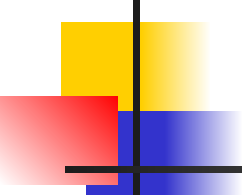
$$r_p = \frac{a-1}{\sqrt{(l_x-1) \cdot (l_y-1)}} \quad (54)$$

где a - коэффициент, который необходимо вычислить по следующей формуле:

$$a = \sum \left[\frac{\sum (f_{xy}^2 : f_y)}{f_x} \right] - \frac{(l_x-1) \cdot (l_y-1)}{n} \quad (55)$$

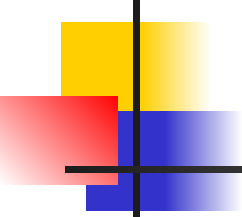
$$a = \sum \left[\frac{\sum (f_{xy}^2 : f_x)}{f_y} \right] - \frac{(l_x-1) \cdot (l_y-1)}{n} \quad (56)$$

где f_{xy}, f_{yx} -частоты в клетках корреляционной решетки.



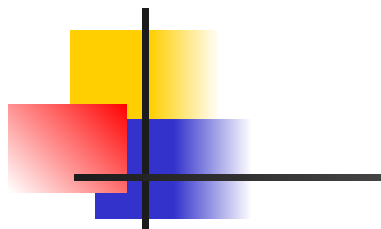
Множественная корреляция. Наряду с анализом двумерных совокупностей в биологии широкое применение находит *статистический анализ многомерных корреляционных связей*. Простейшим случаем множественной корреляции является зависимость между тремя признаками: X , Y и Z . Тесноту связи одного из них (X) с двумя другими признаками (Y и Z) измеряют с помощью *коэффициента множественной корреляции* (формула 62)

Коэффициент множественной корреляции принимает значения от нуля до единицы. Значимость этого совокупного показателя корреляции оценивают по величине критерия Стьюдента с числом степеней свободы – $k=n-3$ и принятым уровнем значимости.



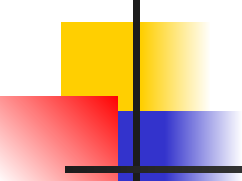
Частная корреляция. Если известна связь между признаками X , Y и Z , можно определить *частные* или *парциальные коэффициенты корреляции*, показывающие корреляционную зависимость между двумя варьирующими признаками при постоянной величине третьего признака (формулы 57-62).

Рассмотренные коэффициенты множественной и частной корреляции применяют лишь для измерения линейных связей. Анализ множественных нелинейных связей описан в специальной литературе.



| № формулы | Показатель | Формула |
|-----------|-------------------------------|---|
| 1 | 2 | 3 |
| 57 | Девятаты | $\sum (x_i - \bar{x})^2 = \sum x_i^2 - (\sum x_i)^2 / n$ |
| | | $\sum (y_i - \bar{y})^2 = \sum y_i^2 - (\sum y_i)^2 / n$ |
| | | $\sum (z_i - \bar{z})^2 = \sum z_i^2 - (\sum z_i)^2 / n$ |
| 58 | Величины сопряженной вариации | $\sum (y_i - \bar{y})(x_i - \bar{x}) = \sum yx - \sum x \sum y / n$ |
| | | $\sum (y_i - \bar{y})(z_i - \bar{z}) = \sum yz - \sum y \sum z / n$ |
| | | $\sum (x_i - \bar{x})(z_i - \bar{z}) = \sum xz - \sum x \sum z / n$ |

| 1 | 2 | 3 |
|----|--|---|
| 59 | Коэффициенты для сопряженной вариации* | $s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$ |
| | | $s_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n}}$ |
| | | $s_z = \sqrt{\frac{\sum (z_i - \bar{z})^2}{n}}$ |
| 60 | Парные коэффициенты корреляции | $r_{xy} = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{ns_x s_y}$ |
| | | $r_{yz} = \frac{\sum (y_i - \bar{y})(z_i - \bar{z})}{ns_y s_z}$ |
| | | $r_{xz} = \frac{\sum (x_i - \bar{x})(z_i - \bar{z})}{ns_x s_z}$ |
| 61 | Частные коэффициенты корреляций | $r_{xy(z)} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}$ |
| | | $r_{xz(y)} = \frac{r_{xz} - r_{xy}r_{yz}}{\sqrt{(1 - r_{xy}^2)(1 - r_{yz}^2)}}$ |
| | | $r_{yz(x)} = \frac{r_{yz} - r_{xy}r_{xz}}{\sqrt{(1 - r_{xy}^2)(1 - r_{xz}^2)}}$ |
| 62 | Множественный коэффициент корреляции | $r_{x(yz)} = \sqrt{\frac{r_{xy}^2 + r_{xz}^2 - 2r_{xy}r_{xz}r_{yz}}{1 - r_{yz}^2}}$ |



Литература:

Основная – 1 [24-66]; 2 [т.1-82-99]; 3 [163-183].

Дополнительная – 2 [34-80]; 3 [28-46]; 5 [142-184].

Контрольные вопросы:

1. Коэффициент генетической корреляции в малочисленных выборках: методика и формулы вычисления.
2. Когда применяют бисериальный показатель связи?
3. Когда применяют тетракорический показатель связи?
4. Когда применяют поликорический показатель связи?
5. Частная корреляция























