

# IMMC23090436

#Modelling

DDL: 2023-12-15

## 🕒 Todo

1.  $\chi^2$ 的含义

## ✓ Done

1.  $\chi^2$ 的计算

## 论文

### 问题复述

通过给出的共564组的关于8种不同蜥蜴的生理学特征，将8种蜥蜴用数学方法区分开。

注：不能使用神经网络等黑箱操作

### 3.1

#### 概要

Task 1 聚焦于使用FPNr将编号为5的蜥蜴从8种蜥蜴种区分出来。

文章通过计算不同种蜥蜴的FPNr值的平均值，发现#5的蜥蜴的FPNr和其他种类的蜥蜴有很大的区分度。

#### $\chi^2$

选择 $\chi^2$ 作为评判分类的标准。

#### $\chi^2$ 的含义

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

其中，O代表*observation*，E代表*expectation*。

在论文中的应用是在于评判分类的好与坏

Table 3: Classification result of a criterion

	Species # $i$	Not Species # $i$
Classified as Species # $i$	$a$	$b$
Classified as not Species # $i$	$c$	$d$

$$\chi^2 = \frac{(a + b + c + d)(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$$

## 3.2 and 3.3

### 概要

1. 建立一个模型去选择有用的变量
2. 基于最有效的一对变量去实现蜥蜴的识别

### 选择变量

$$r_j = \frac{j_5(\bar{j}_5 - \sigma_{j_5}, \bar{j}_5 + \sigma_{j_5})}{\sum_{i=1}^8 j_i(\bar{j}_5 - \sigma_{j_5}, \bar{j}_5 + \sigma_{j_5})}$$

含义：计算每个变量对于识别的能力(capability)

文中计算得到表格如下:

Table 6:  $r_j$  of each variable

Pholidosis		Morphometric	
Variable( $j$ )	$r_j(\times 10^{-2})$	Variable( $j$ )	$r_j(\times 10^{-2})$
MBS	59.26	SVL	5.8
VSN	5.60	TRL	3.18
CSN	21.65	HL	9.52
GSN	51.43	PL	14.95
FPNr	94.74	ESD	11.67
SDLr	5.15	HW	5.84
SCSr	5.07	HH	7.63
SCGr	10.53	MO	12.86
SMr	3.70	FFL	11.59
MTr	6.13	HFL	26.23
PA	5.36		
PTMr	5.05		
aNDSr	5.41		

通过分析各个变量的 $r_j$ 值，可以发Task 1中使用的FPNr确实是分辨能力最好的变量，侧面印证了计算 $r_j$ 的合理性。  
通过排序得到了第二有效的变量为MBS，因此文章使用MBS和FPNr这一对变量进行计算。

计算

文章使用了如下函数描述:

$$f(x,y) = x \times y$$

其中 $x,y$ 分别代表MBS和FPNr。  
文章确定了当函数值小于450时，蜥蜴被认定为#5；而当函数值大于等于450时，被认为不是#5。

🔗 Question

$M = 450$ 是如何确定的，文章中没有给出明确的方法  
为什么要选择 $f(x,y) = x \times y$

## 概要

1. 使用正态分布拟合并判断蜥蜴的性别

## 计算

原文计算公式：

$$R_{p,q} = \frac{|\bar{M}_{p,q} - \bar{F}_{p,q}|}{\sigma_{M_{p,q}} + \sigma_{M_{p,q}}}$$

### ⚡ Error

此处分母中有两个 $\sigma_{M_{p,q}}$

正确的式子应该如下：

$$R_{p,q} = \frac{|\bar{M}_{p,q} - \bar{F}_{p,q}|}{\sigma_{M_{p,q}} + \sigma_{F_{p,q}}}$$

$R_{p,q}$ 的意义：没看懂

$\bar{M}_{p,q}$	The male lizards' average value of $\frac{p}{q}$
$\bar{F}_{p,q}$	The female lizards' average value of $\frac{p}{q}$

### 🤔 Question

为什么这个式子能用来判断p,q变量的区分度

既然在3.2和3.3中已经有一个式子了为什么还要再做一个式子

### ⚡ Error

文章中没有解释为什么蜥蜴的数据呈正态分布

而是直接给出了蜥蜴的数据呈正态分布，这一点上不够合理

Table 8: The optimal  $\frac{p}{q}$  variables (linear sizes only)

Variable 1 ( $p$ )	Variable 2 ( $q$ )	$R_{p,q}$
TRL	HH	1.078
TRL	HL	1.209
<u>SVL</u>	<u>HL</u>	<u>1.296</u>
TRL	PL	1.429
TRL	HW	1.630

#### 🔗 Question

为什么要选择 $R_{p,q}$ 为1.296的SVL/HL

### 3.5

#### 🔗 Question

最主要的问题是没有说明为什么选择那些数据去进行区分

可能是因为 没有空间写了

### 4.1

#### 🔗 Question

为什么还要再写一遍sex classification