

## 数字生物分类（校赛 B 题）：

### 基于对 Darevskia 属蜥蜴的鳞序特征和形态学特征的统计分析

#### 小组成员

侯亦驰 徐雷斌 刘阵

#### 摘要

在对许多近缘动物的物种分类工作中，常常需要通过对可测量特征的数据分析进行区分。本文通过 ANOVA、LSD、Sidak、Fisher 等方法针对一组 Darevskia 属中 8 个种 564 只蜥蜴的已知数据进行统计分析，基于对蜥蜴身体不同部位的鳞片计数（鳞序特征）和蜥蜴身体部位的线性大小（形态学特征）的数据进行处理，建立相对准确的判别种类或性别的标准，方便生物学家在野外工作条件下通过计算来判别物种。

**关键词** SPSS 判别分析 Fisher 判别函数 物种分类 特征统计

#### 1 问题提出

已知 Darevskia 属 8 个种/物种的 564 只蜥蜴的包括鳞序特征和形态学特征的测量值，欲根据已知数据建立可以较准确区分蜥蜴种类和性别的数学模型。模型应当：

- 1) 数学计算简单明了，利于生物学家野外工作时使用图形计算器做计算；
- 2) 标准建立的方法透明，以便于生物学家构建解释性理论，不建议采用包装性方案（如神经网络）；
- 3) 附有性能指标，标示出模型正确/错误分类的蜥蜴数量。

#### 2 问题分析

针对问题一，仅使用右侧的股骨孔数(FPNr)尽可能准确地将第 5 种蜥蜴与其他蜥蜴区分开来。本文发现，通过对 8 个种类的蜥蜴 FPNr 的数据做 ANOVA 分析及最小显著法(LSD)和 Sidak 法的事后比较，并绘制箱型图(Boxplots)，第 5 种蜥蜴的 FPNr 值的分布与其它种类蜥蜴的数据分布具有显著差异。接下来，本文使用两种方法建立标准。第一种方法中，本文设立关于 FPNr 的一元不等式作为判别第 5 种蜥蜴的标准。在第二种方法中，本文使用 Fisher 判别函数(Fisher Discriminant Function)作为判别标准，并指出此方法相对普遍的适用性。

针对问题二，题目要求选取两个变量尽可能准确地将第五种蜥蜴与其他蜥蜴区分开来。本文使用 ANOVA 及 LSD 事后比较展示各个特征量对第 5 种蜥蜴的区分起到的作用，并选取其中最显著的两个变量作为“最佳预测变量对”进行预测。

针对问题三、四、五，使用形态学和/或鳞序特征预计 1) 种群性别；2) 分组组内区分物种种类；3) 预测整个种群中的种类和性别。基于三个问题都属于判别分析(Discriminant Analysis)问题，本文仍然使用问题一中提及的 Fisher 判别法作为主要方法建立数学模型。首先使用 ANOVA 分析各个变量对于各组的判别是否显著，基于方程简洁的需求去掉那些相对不显著的变量。接着根据剩下的变量列出基于 Fisher 判别函数的方程组，并建立判别标准。

3 模型假设

1) 假设现有的各个特征量一定程度上由于蜥蜴物种种类的不同而有所区别。

4 符号申明

符号	含义
$f_{Fisher}$	Fisher 判别函数
$OtoP_a^b$	原#a 种蜥蜴被预测成#b 种蜥蜴的个数
$PAR_A^B$	特征 A 与 B 作为 “变量预测对” 时的准确率

5 模型建立

5.1 问题一的模型建立与求解

Step 1: 数据处理-ANOVA 分析及事后检验

本文利用 SPSS 统计软件对数据进行单因素 ANOVA 检验，因子为蜥蜴种类 (Species\_num)，因变量为右侧的股骨孔数(FPNr)，得到下图数据。可以看到，蜥蜴种类基于 FPNr 数据的  $F = 185.007$ ，远大于 1，对应  $p(Sig.) < 0.001$ ，组间具有显著差异。

ANOVA					
FPNr	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2391.854	7	341.693	185.007	<.001
Within Groups	1026.891	556	1.847		
Total	3418.745	563			

进一步对两两各组进行事后比较。本文使用最小显著法(LSD)和 Sidak 法两种方式进行检验，后者是对前者的累乘矫正。检验结果如下图（节选）所示：

Post Hoc Tests

Multiple Comparisons							
Dependent Variable: FPNr							
	(I) Species_num	(J) Species_num	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
LSD	5	1	-9.466*	.324	<.001	-10.10	-8.83
		2	-8.169*	.326	<.001	-8.81	-7.53
		3	-6.420*	.298	<.001	-7.01	-5.83
		4	-9.741*	.311	<.001	-10.35	-9.13
		6	-6.908*	.304	<.001	-7.51	-6.31
		7	-9.087*	.401	<.001	-9.88	-8.30
		8	-8.292*	.411	<.001	-9.10	-7.48
Sidak	5	1	-9.466*	.324	.000	-10.48	-8.45
		2	-8.169*	.326	.000	-9.19	-7.15
		3	-6.420*	.298	.000	-7.35	-5.49
		4	-9.741*	.311	.000	-10.71	-8.77
		6	-6.908*	.304	.000	-7.86	-5.96
		7	-9.087*	.401	.000	-10.34	-7.83
		8	-8.292*	.411	.000	-9.58	-7.00

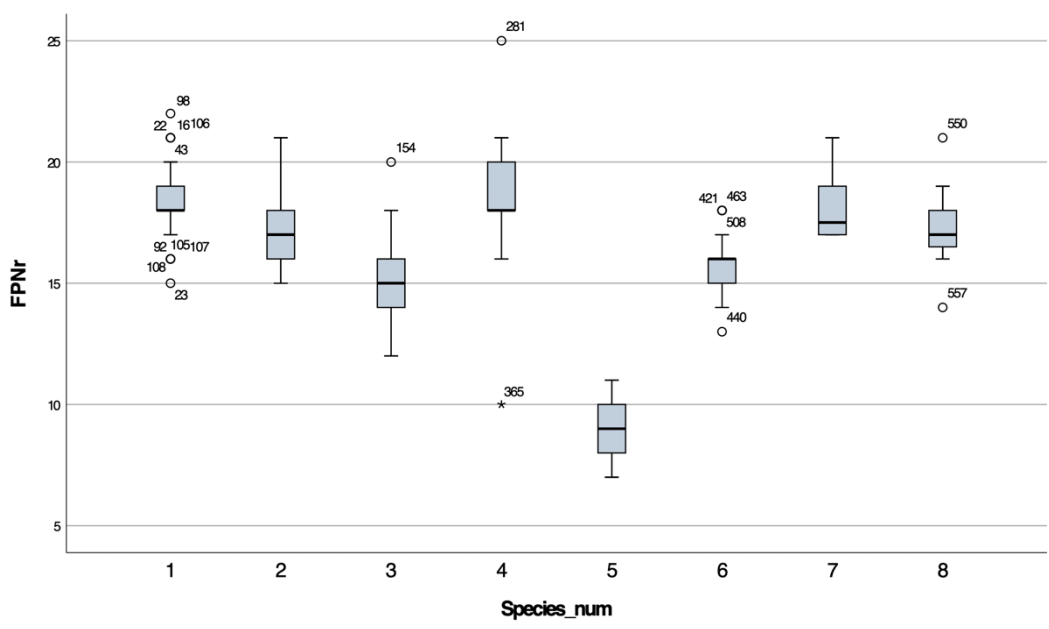
可以看到，不管是在 LSD 检验中还是在 Sidak 检验中，第 5 种蜥蜴与其它种类蜥蜴

均具有显著性差异 ( $p < 0.001$ )。

**Step 2: 建立数学模型与标准**

本文使用两种方法建立标准。

方法一：使用 SPSS 绘制箱型图，变量为右侧的股骨孔数(FPNr)，类别轴为蜥蜴种



类(Species\_num)，可以直观地观察到第五种蜥蜴 FPNr 的中位数、四分位距、最大最小值均与其它种类蜥蜴相差甚远，具有十分显著的区分性。另外，图中显示第 5 种蜥蜴的数值分布范围与其它种类蜥蜴的数值分布范围除了第 4 组中 365 号外没有相交区，且由于 365 号的数值超出第 4 组的 3 倍四分位距（图中显示为\*号），可以判断为是极端值。据此，第一种判别方法则是取第 5 组的最大值（11）与其它组的最小值（即第 3 组的最小值 12 的中数 11.5 作为界限。（FPNr 为整数）若 FPNr  $\leq 11$  则判定为第 5 种蜥蜴。使用已知数据检查，仅有 365 号 1 个数据误判，其它均有效。

方法二：使用 Fisher 判别函数判别蜥蜴种类。借助 SPSS 判别分析功能求得 Fisher 判别函数，将数值代入函数来求得结果。自变量为右侧的股骨孔数(FPNr)，分组变量为蜥蜴种类(Species\_num)，分组变量的范围为(1, 8)；函数系数选择费希尔(Fisher)。

Classification Function Coefficients								
	Species_num							
	1	2	3	4	5	6	7	8
FPNr	9.976	9.273	8.326	10.124	4.850	8.591	9.771	9.340
(Constant)	-93.976	-81.491	-66.102	-96.737	-23.805	-70.234	-90.236	-82.636
Fisher's linear discriminant functions								

在输出中得到上图并根据图表中的数据列出 Fisher 判别函数：

$$\begin{cases} y_1 = 9.976x - 93.976 \\ y_2 = 9.273x - 81.491 \\ y_3 = 8.326x - 66.102 \\ y_4 = 10.124x - 96.737 \\ y_5 = 4.850x - 23.805 \\ y_6 = 8.591x - 70.234 \\ y_7 = 9.771x - 90.236 \\ y_8 = 9.340x - 82.636 \end{cases}$$

其中 $x$ 为某一数据中 FPNr 数值, 将数值代入上述方程组, 求得 $y_i(1 \leq i \leq 8, i \in Z)$ ,

Classification Results <sup>a</sup>											
		Predicted Group Membership									
	Species_num	1	2	3	4	5	6	7	8	Total	
Original	Count	1	0	9	1	27	0	5	24	0	66
		2	0	23	9	9	0	9	13	0	63
		3	0	22	86	1	3	32	12	0	156
		4	0	6	0	46	1	2	38	0	93
		5	0	0	0	0	24	0	0	0	24
		6	0	25	39	0	0	52	4	0	120
		7	0	11	0	8	0	0	3	0	22
		8	0	8	1	3	0	4	4	0	20
%	1	.0	13.6	1.5	40.9	.0	7.6	36.4	.0	100.0	
	2	.0	36.5	14.3	14.3	.0	14.3	20.6	.0	100.0	
	3	.0	14.1	55.1	.6	1.9	20.5	7.7	.0	100.0	
	4	.0	6.5	.0	49.5	1.1	2.2	40.9	.0	100.0	
	5	.0	.0	.0	.0	100.0	.0	.0	.0	100.0	
	6	.0	20.8	32.5	.0	.0	43.3	3.3	.0	100.0	
	7	.0	50.0	.0	36.4	.0	.0	13.6	.0	100.0	
	8	.0	40.0	5.0	15.0	.0	20.0	20.0	.0	100.0	

若此数据中 $y_{max} = y_5$ 则可以判定为是第 5 种蜥蜴。

分类结果检验中对 24 个第 5 种蜥蜴样本进行检验, 有效率为 100%, 这说明此 Fisher 判别函数对于第 5 种蜥蜴的判别分类十分精确。

Step 3: 方法比较与总结

上述两种方法对第 5 种蜥蜴的分类均具有极高的准确性。前者分类简便有效, 适合针对离分界值较远的数据进行判定, 也适合野外工作的大致判定。后者计算量较大, 但对于数据的分类更加科学准确, 适合用于离分界值较近的数据。并且后者的判定方法适用性更加普遍, 在后几问当中有十分显著的作用。

5.2 问题二的模型建立与求解

Step 1: 使用 ANOVA 及 LSD 事后比较展示各个特征量对第 5 种蜥蜴的区分起到的作用。利用 SPSS 统计软件对数据进行单因素 ANOVA 检验, 因子为蜥蜴种类

Multiple Comparisons							
LSD							
Dependent Variable	(I) Species_num	(J) Species_num	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
MBS	5	1	-13.788*	.561	<.001	-14.89	-12.69
		2	-11.389*	.564	<.001	-12.50	-10.28
		3	-6.199*	.516	<.001	-7.21	-5.19
		4	-9.898*	.539	<.001	-10.96	-8.84
		6	-5.150*	.526	<.001	-6.18	-4.12
		7	-15.561*	.694	<.001	-16.92	-14.20
		8	-7.133*	.712	<.001	-8.53	-5.73
VSN	5	1	-1.125*	.341	.001	-1.79	-.46
		2	-.990*	.343	.004	-1.66	-.32
		3	-.965*	.313	.002	-1.58	-.35
		4	-1.636*	.327	<.001	-2.28	-.99
		6	-3.750*	.320	<.001	-4.38	-3.12
		7	-4.958*	.422	<.001	-5.79	-4.13
		8	-3.658*	.433	<.001	-4.51	-2.81
CSN	5	1	-3.420*	.235	<.001	-3.88	-2.96
		2	-2.843*	.236	<.001	-3.31	-2.38
		3	-1.625*	.216	<.001	-2.05	-1.20
		4	-2.972*	.225	<.001	-3.41	-2.53
		6	-2.925*	.220	<.001	-3.36	-2.49
		7	-4.466*	.291	<.001	-5.04	-3.90
		8	-2.775*	.298	<.001	-3.36	-2.19

(Species\_num), 因变量为各个形态学和鳞序特征, 得到下图数据 (由于原图表过长, 此图仅为原图的节选, 以供展示图表信息和格式)。可以看到, 蜥蜴种类#5 相对于其他组别对应  $p(\text{Sig.}) < 0.001$ , 组间具有显著差异的特征值为 MBS、CSN、GSN、FPNr、SCGr、MO 和 HFL。

**Step 2:** 由于通过问题一的方法已经基于 FPNr 得到有效率为 100% 的判别方程, 因此, 任何变量与 FPNr 配对的预测结果均为 100%。然而不能保证新的数据仍然适用于 FPNr 的判别, 最佳的“变量预测对”可以通过将 FPNr 与以上其它变量进行搭配, 从而可以进一步验证分类的合理性。

5.3 问题三的模型建立与求解

本文在问题一当中提到, 使用 Fisher 判别函数的方法具有更普遍的适用性, 这一点体现在 Fisher 判别法仍然可适用于问题三、四、五当中。下面就利用此方法解题。

**Step 1:** 利用 SPSS 统计软件对数据进行单因素 ANOVA 检验, 因子为性别(Sex\_num), 因变量为各个形态学和鳞序特征, 得到下图数据。可以看到, 蜥蜴性别基于 CSN( $p = 0.064$ )、GSN( $p = 0.247$ )、SCSr( $p = 0.150$ )、PA( $p = 0.913$ )、PTMr( $p = 0.551$ )、SVL( $p =$

ANOVA					
		Sum of Squares	df	Mean Square	F
MBS	Between Groups	390.183	1	390.183	21.442
	Within Groups	10226.895	562	18.197	
	Total	10617.078	563		<.001
VSN	Between Groups	1013.501	1	1013.501	507.850
	Within Groups	1121.568	562	1.996	
	Total	2135.069	563		<.001
CSN	Between Groups	6.075	1	6.075	3.438
	Within Groups	993.194	562	1.767	
	Total	999.270	563		.064
GSN	Between Groups	15.181	1	15.181	1.341
	Within Groups	6363.110	562	11.322	
	Total	6378.291	563		.247
FPNr	Between Groups	52.916	1	52.916	8.835
	Within Groups	3365.829	562	5.989	
	Total	3418.745	563		.003
SDLr	Between Groups	442.825	1	442.825	51.240
	Within Groups	4856.897	562	8.642	
	Total	5299.722	563		<.001
SCSr	Between Groups	1.292	1	1.292	2.075
	Within Groups	350.075	562	.623	
	Total	351.367	563		.150
SCGr	Between Groups	523.145	1	523.145	55.292
	Within Groups	5317.323	562	9.461	
	Total	5840.468	563		<.001
SMr	Between Groups	2.781	1	2.781	7.170
	Within Groups	217.957	562	.388	
	Total	220.738	563		.008
MTr	Between Groups	23.304	1	23.304	24.845
	Within Groups	527.136	562	.938	
	Total	550.440	563		<.001
PA	Between Groups	.005	1	.005	.012
	Within Groups	226.695	562	.403	
	Total	226.700	563		.913

PTMr	Between Groups	.583	1	.583	.356
	Within Groups	922.303	562	1.641	
	Total	922.887	563		.551
aNDSr	Between Groups	4.573	1	4.573	41.871
	Within Groups	61.374	562	.109	
	Total	65.947	563		<.001
SVL	Between Groups	78.464	1	78.464	1.988
	Within Groups	22186.369	562	39.478	
	Total	22264.834	563		.159
TRL	Between Groups	320.474	1	320.474	19.314
	Within Groups	9325.107	562	16.593	
	Total	9645.581	563		<.001
HL	Between Groups	291.222	1	291.222	63.300
	Within Groups	2585.580	562	4.601	
	Total	2876.802	563		<.001
PL	Between Groups	194.167	1	194.167	110.778
	Within Groups	985.053	562	1.753	
	Total	1179.220	563		<.001
ESD	Between Groups	23.081	1	23.081	9.916
	Within Groups	1308.128	562	2.328	
	Total	1331.209	563		.002
HW	Between Groups	142.998	1	142.998	133.540
	Within Groups	601.801	562	1.071	
	Total	744.799	563		<.001
HH	Between Groups	42.938	1	42.938	51.527
	Within Groups	468.329	562	.833	
	Total	511.267	563		<.001
MO	Between Groups	64.223	1	64.223	32.060
	Within Groups	1125.814	562	2.003	
	Total	1190.037	563		<.001
FFL	Between Groups	201.989	1	201.989	32.358
	Within Groups	3508.167	562	6.242	
	Total	3710.156	563		<.001
HFL	Between Groups	1071.407	1	1071.407	87.745
	Within Groups	6862.291	562	12.210	
	Total	7933.697	563		<.001

0.159)数据对应的  $p$  均大于 0.05, 组间没有显著差异, 对判别蜥蜴的性别起不到显著作用。考虑到方程简洁的需求, 排除这六项数据。(通过使用 SPSS 对排除前和排除后的比较预测准确率仅相差 1%左右可以佐证)

**Step 2:** 排除上述六项数据后, 使用 Fisher 判别函数判别蜥蜴性别。借助 SPSS 判别分析功能求得 Fisher 判别函数, 将数值代入函数来求得结果。自变量为各个形态学和鳞序特征, 分组变量为蜥蜴性别(Sex\_num), 分组变量的范围为(1, 2); 函数系数选择费希尔(Fisher)。函数表如右图所示。

根据图表中的数据列出 Fisher 判别函数。考虑到数据较多, 本文使用矩阵形式展现:

$$f_{Fisher} = \begin{pmatrix} 3.056 & 15.862 & -6.117 & 4.869 & 0.048 & -2.720 & -2.329 & 29.478 & -1.673 & 0.319 & 5.469 & 0.335 & 1.998 & 0.565 & 4.189 & -0.144 & 0.923 & -359.932 \\ 3.075 & 17.207 & -5.892 & 4.944 & -0.547 & -2.813 & -2.462 & 27.753 & -1.008 & 0.345 & 4.700 & 0.489 & 0.440 & -0.910 & 3.754 & -0.338 & 0.919 & -376.083 \end{pmatrix}$$

$$X = \begin{pmatrix} x_{MBS} \\ x_{VSN} \\ x_{FPNr} \\ x_{SDLr} \\ x_{SCGr} \\ x_{SMr} \\ x_{MTr} \\ x_{aNDsr} \\ x_{TRL} \\ x_{HL} \\ x_{PL} \\ x_{ESD} \\ x_{HW} \\ x_{HH} \\ x_{MO} \\ x_{FFL} \\ x_{HFL} \\ 1 \end{pmatrix} \quad Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$$

列出方程如下：

$$Y = f_{Fisher} X$$

带入具体某一数据的X，解上述方程得 $Y_1$ ， $Y_2$ ；  
比较 $Y_1$ 和 $Y_2$ ，若 $Y_1$ 大，则该蜥蜴性别为 Sex\_num = 1；若 $Y_2$ 大，则该蜥蜴性别为 Sex\_num = 2。

SPSS 的判别分析测试结果图如下。图表显示该模型预测准确率较高，总准确率为 93.3%，其中对 Sex\_num = 1 的蜥蜴预测的准确率为 95.6%，对 Sex\_num = 2 的蜥蜴预测的准确率为 91.0%。

	Classification Function Coefficients	
	Sex_num	
	1	2
MBS	3.056	3.075
VSN	15.862	17.207
FPNr	-6.117	-5.892
SDLr	4.869	4.944
SCGr	.048	-.547
SMr	-2.720	-2.813
MTr	-2.329	-2.462
aNDsr	29.478	27.753
TRL	-1.673	-1.008
HL	.319	.435
PL	5.469	4.700
ESD	.335	.489
HW	1.998	.440
HH	.565	-.910
MO	4.189	3.754
FFL	-.144	-.338
HFL	.923	.919
(Constant)	-359.932	-376.083
Fisher's linear discriminant functions		

Classification Results <sup>a</sup>					
		Predicted Group Membership			
		Sex_num	1	2	Total
Original	Count	1	263	12	275
		2	26	263	289
	%	1	95.6	4.4	100.0
		2	9.0	91.0	100.0

a. 93.3% of original grouped cases correctly classified.

## 5.4 问题四的模型建立与求解

问题四采用相同思路，用 Fisher 判别函数判别蜥蜴种类。

**Step 1:** 利用 SPSS 统计软件对数据进行单因素 ANOVA 检验，因子为性别 (Species\_num)，因变量为各个形态学和鳞片特征，得到下图数据。可以看到，所有变量对判别蜥蜴的种类都起到显著作用 ( $p < 0.001$ )。

ANOVA					
		Sum of Squares	df	Mean Square	F
MBS	Between Groups	7539.009	7	1077.001	194.542
	Within Groups	3078.069	556	5.536	
	Total	10617.078	563		
VSN	Between Groups	999.700	7	142.814	69.937
	Within Groups	1135.369	556	2.042	
	Total	2135.069	563		
CSN	Between Groups	460.147	7	65.735	67.793
	Within Groups	539.122	556	.970	
	Total	999.270	563		
CSN	Between Groups	4613.909	7	659.130	207.708
	Within Groups	1764.382	556	3.173	
	Total	6378.291	563		
FPNr	Between Groups	2391.854	7	341.693	185.007
	Within Groups	1026.891	556	1.847	
	Total	3418.745	563		
SDLr	Between Groups	2645.561	7	377.937	79.171
	Within Groups	2654.161	556	4.774	
	Total	5299.722	563		
SCGr	Between Groups	52.033	7	7.433	13.807
	Within Groups	299.334	556	.538	
	Total	351.367	563		
SCGr	Between Groups	3996.893	7	570.985	172.202
	Within Groups	1843.575	556	3.316	
	Total	5840.468	563		
SMr	Between Groups	41.299	7	5.900	18.281
	Within Groups	179.439	556	.323	
	Total	220.738	563		
MTr	Between Groups	277.099	7	39.586	80.520
	Within Groups	273.341	556	.492	
	Total	550.440	563		
PA	Between Groups	86.379	7	12.340	48.895
	Within Groups	140.321	556	.252	
	Total	226.700	563		

PTMr	Between Groups	468.190	7	66.884	81.786	<.001
	Within Groups	454.697	556	.818		
	Total	922.887	563			
aNDsr	Between Groups	28.169	7	4.024	59.224	<.001
	Within Groups	37.778	556	.068		
	Total	65.947	563			
SVL	Between Groups	8589.136	7	1227.019	49.886	<.001
	Within Groups	13675.698	556	24.597		
	Total	22264.834	563			
TRL	Between Groups	2259.376	7	322.768	24.297	<.001
	Within Groups	7386.205	556	13.285		
	Total	9645.581	563			
HL	Between Groups	880.954	7	125.851	35.059	<.001
	Within Groups	1995.848	556	3.590		
	Total	2876.802	563			
PL	Between Groups	445.457	7	63.637	48.220	<.001
	Within Groups	733.763	556	1.320		
	Total	1179.220	563			
ESD	Between Groups	515.714	7	73.673	50.230	<.001
	Within Groups	815.495	556	1.467		
	Total	1331.209	563			
HW	Between Groups	341.278	7	48.754	67.177	<.001
	Within Groups	403.522	556	.726		
	Total	744.799	563			
HH	Between Groups	232.176	7	33.168	66.077	<.001
	Within Groups	279.091	556	.502		
	Total	511.267	563			
MO	Between Groups	277.065	7	39.581	24.105	<.001
	Within Groups	912.972	556	1.642		
	Total	1190.037	563			
FFL	Between Groups	1854.966	7	264.995	79.419	<.001
	Within Groups	1855.190	556	3.337		
	Total	3710.156	563			
HFL	Between Groups	3881.653	7	554.522	76.089	<.001
	Within Groups	4052.044	556	7.288		
	Total	7933.697	563			

**Step 2:** 使用 Fisher 判别函数判别蜥蜴种类。借助 SPSS 判别分析功能求得 Fisher 判别函数，将数值代入函数来求得结果。自变量为各个形态学和鳞序特征，分组变量为蜥蜴种类(Species\_num)，分组变量的范围为(1, 8)；函数系数选择费希尔(Fisher)。函数表如下图所示。

Classification Function Coefficients								
	Species_num							
	1	2	3	4	5	6	7	8
MBS	8.049	7.644	6.620	6.713	6.128	6.602	8.370	6.696
VSN	20.477	20.094	20.036	19.941	18.863	21.737	22.527	22.015
FPNr	4.819	4.143	3.111	4.717	.336	3.961	4.605	4.485
SDLr	6.436	6.313	5.857	5.515	4.870	4.950	5.359	4.858
SCGr	2.110	2.449	1.790	-.099	-.075	.500	2.194	.853
SMr	-2.322	-1.108	-1.167	-.659	-2.435	-1.405	-1.570	-2.211
MTr	4.622	1.283	.189	1.187	1.101	-.644	1.321	1.558
aNDSr	26.500	27.197	34.829	29.932	28.849	27.257	27.050	28.949
TRL	-2.400	-2.257	-2.243	-1.923	-2.094	-2.998	-3.185	-2.590
HL	2.230	1.699	2.059	1.752	1.988	1.879	2.247	2.431
PL	9.631	8.745	9.640	9.898	8.583	9.184	9.873	8.074
ESD	-3.541	-3.384	-1.339	-1.609	-1.021	-1.307	-2.054	-3.182
HW	2.360	3.666	1.158	1.407	1.895	.712	-.533	5.343
HH	-8.254	-7.840	-4.625	-4.879	-2.508	-5.301	-7.201	-5.392
MO	2.590	3.149	.432	-.604	.863	1.606	1.257	1.204
FFL	-2.217	-1.913	.011	.516	.297	-.053	-.687	.277
HFL	3.751	3.041	3.178	3.721	2.816	3.004	3.044	3.016
CSN	8.529	8.226	7.187	8.351	5.291	8.404	10.061	8.302
GSN	.906	1.134	1.224	3.495	.681	1.148	1.336	1.455
SCSr	4.749	3.411	4.599	5.704	5.700	4.752	3.299	4.543
PA	7.310	4.076	8.525	6.484	9.374	5.864	6.935	5.215
PTMr	3.729	5.895	4.191	1.155	2.336	4.012	5.366	3.864
SVL	1.148	.988	.604	.769	.303	1.220	1.570	.817
(Constant)	-824.887	-753.307	-708.684	-805.304	-551.192	-720.904	-875.720	-763.198
Fisher's linear discriminant functions								

根据图表中的数据列出 Fisher 判别函数。考虑到数据较多，本文使用矩阵形式展现：

$$f_{Fisher} = \begin{pmatrix} 8.049 & 20.477 & 4.819 & 6.436 & 2.110 & -2.322 & 4.622 & 26.500 & -2.400 & 2.230 & 9.631 & -3.541 & 2.360 & -8.254 & 2.590 & -2.217 & 3.751 & 8.529 & 0.906 & 4.749 & 7.310 & 3.729 & 1.148 & -824.887 \\ 7.644 & 20.094 & 4.143 & 6.313 & 2.449 & -1.108 & 1.283 & 27.197 & -2.257 & 1.699 & 8.745 & -3.384 & 3.666 & -7.840 & 3.149 & -1.913 & 3.041 & 8.226 & 1.134 & 3.411 & 4.076 & 5.895 & 0.988 & -753.307 \\ 6.620 & 20.036 & 3.111 & 5.857 & 1.790 & -1.167 & 0.189 & 34.829 & -2.243 & 2.059 & 9.640 & -1.339 & 1.158 & -4.625 & 0.432 & 0.011 & 3.178 & 7.187 & 1.224 & 4.599 & 8.525 & 4.191 & 0.604 & -708.684 \\ 6.713 & 19.941 & 4.717 & 5.515 & -0.099 & -0.659 & 1.187 & 29.932 & -1.923 & 1.752 & 9.898 & -1.609 & 1.407 & -4.879 & -0.604 & 0.516 & 3.721 & 8.351 & 3.495 & 5.704 & 6.484 & 1.155 & 0.769 & -805.304 \\ 6.128 & 18.863 & 0.336 & 4.870 & -0.075 & -2.435 & 1.101 & 28.849 & -2.094 & 1.988 & 8.583 & -1.021 & 1.895 & -2.508 & 0.863 & 0.297 & 2.816 & 5.291 & 0.681 & 5.700 & 9.374 & 2.336 & 0.303 & -551.192 \\ 6.602 & 21.737 & 3.961 & 4.950 & 0.500 & -1.405 & -0.644 & 27.257 & -2.998 & 1.879 & 9.184 & -1.307 & 0.712 & -5.301 & 1.606 & -0.053 & 3.004 & 8.404 & 1.148 & 4.752 & 5.864 & 4.012 & 1.220 & -720.904 \\ 8.370 & 22.527 & 4.605 & 5.359 & 2.194 & -1.570 & 1.321 & 27.050 & -3.185 & 2.247 & 9.873 & -2.054 & -0.533 & -7.201 & 1.257 & -0.687 & 4.044 & 10.061 & 1.336 & 3.299 & 6.935 & 5.366 & 1.570 & -875.720 \\ 6.696 & 22.015 & 4.485 & 4.858 & 0.853 & -2.211 & 1.558 & 28.949 & -2.590 & 2.431 & 8.074 & -3.182 & 5.343 & -5.392 & 1.204 & 0.277 & 3.016 & 8.302 & 1.455 & 4.543 & 5.215 & 3.864 & 0.817 & -763.198 \end{pmatrix}$$

$$X = \begin{pmatrix} x_{MBS} \\ x_{VSN} \\ x_{FPNr} \\ x_{SDLr} \\ x_{SCGr} \\ x_{SMr} \\ x_{MTr} \\ x_{aNDSr} \\ x_{TRL} \\ x_{HL} \\ x_{PL} \\ x_{ESD} \\ x_{HW} \\ x_{HH} \\ x_{MO} \\ x_{FFL} \\ x_{HFL} \\ x_{CSN} \\ x_{GSN} \\ x_{SCSr} \\ x_{PA} \\ x_{PTMr} \\ x_{SVL} \\ 1 \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{pmatrix}$$

列出方程如下：

$$Y = f_{Fisher}X$$

带入具体某一数据的 $X$ ，解上述方程得 $Y$ ；比较 $y_i(1 \leq i \leq 8, i \in Z)$ ，得 $y_{MAX} = y_m$ ，则该数据相应的蜥蜴种类应当属于 Species\_num=m。

SPSS 的判别分析测试结果图如下。图表显示该模型预测准确率较高，总准确率为 97.5%，说明此方程模型可以很好地区分 8 个品种的蜥蜴。

Classification Results <sup>a</sup>											
		Predicted Group Membership									
	Species_num	1	2	3	4	5	6	7	8	Total	
Original	Count	1	2	3	4	5	6	7	8		
	1	63	2	0	0	0	0	0	1	66	
	2	2	61	0	0	0	0	0	0	63	
	3	0	0	152	2	1	1	0	0	156	
	4	0	0	0	91	0	1	0	1	93	
	5	0	0	0	0	24	0	0	0	24	
	6	0	0	1	0	0	118	0	1	120	
	7	0	0	0	0	0	0	22	0	22	
	8	0	0	0	0	0	1	0	19	20	
%	1	95.5	3.0	.0	.0	.0	.0	.0	1.5	100.0	
	2	3.2	96.8	.0	.0	.0	.0	.0	.0	100.0	
	3	.0	.0	97.4	1.3	.6	.6	.0	.0	100.0	
	4	.0	.0	.0	97.8	.0	1.1	.0	1.1	100.0	
	5	.0	.0	.0	.0	100.0	.0	.0	.0	100.0	
	6	.0	.0	.8	.0	.0	98.3	.0	.8	100.0	
	7	.0	.0	.0	.0	.0	.0	100.0	.0	100.0	
	8	.0	.0	.0	.0	.0	5.0	.0	95.0	100.0	

a. 97.5% of original grouped cases correctly classified.

下文使用 $OtoP_a^b$ 表示原#a 种蜥蜴被预测成#b 种蜥蜴的个数。

问题 a 比较物种#6 与#7。从上图数据可以得到总共对 120 个#6 种蜥蜴进行预测，其中 $OtoP_6^6 = 118, OtoP_6^7 = 0$ ；总共对 22 个#7 种蜥蜴进行预测，其中 $OtoP_7^7 = 22, OtoP_7^6 = 0$ 。#6 与#7 间的预测准确率(预测准确率指实际某种蜥蜴数占预测成该种蜥蜴的数量的比率)分别为 100%，100%。因此，此模型可以极好地分辨 #6 #7 两种蜥蜴。

问题 b 比较物种#1 与#2。从上图数据可以得到总共对 66 个#1 种蜥蜴进行预测，其中 $OtoP_1^1 = 63, OtoP_1^2 = 2$ ；总共对 63 个#2 种蜥蜴进行预测，其中 $OtoP_2^2 = 61, OtoP_2^1 = 2$ 。#1 与#2 间的预测准确率分别为 96.92%，96.82%。因此，此模型可以较好地分辨 #1 #2 两种蜥蜴。

问题 c 比较物种#3#4 与#5。从上图数据可以得到总共对 156 个#3 种蜥蜴进行预测，其中 $OtoP_3^3 = 152, OtoP_3^4 = 2, OtoP_3^5 = 1$ ；总共对 93 个#4 种蜥蜴进行预测，其中 $OtoP_4^4 = 91, OtoP_4^3 = 0, OtoP_4^5 = 0$ ；总共对 24 个#5 种蜥蜴进行预测，其中 $OtoP_5^5 = 24, OtoP_5^3 = 0, OtoP_5^4 = 0$ 。#3, #4 与#5 间的预测准确率分别为 100%，97.84%，96%。因此，此模型可以较好地分辨 #6 #7 两种蜥蜴。

### 5.5 问题五的模型建立与求解

问题五实际上是问题三、四的结合，由于本文在解答问题三、四时已经给出蜥蜴在整个种群中的性别、种类的数学模型，因此，解答问题五仅需要将数据分别代入两个数学模型，即可预测。

## 6 灵敏度分析

## 7 模型的评价与推广

### 7.1 模型的优/缺点

### 7.3 模型的推广

## 8 参考文献



## 附录

### 1) 问题一 SPSS 语法：

```
ONEWAY FPNr BY Species_num  
  /ES=OVERALL  
  /STATISTICS DESCRIPTIVES HOMOGENEITY  
  /MISSING ANALYSIS  
  /CRITERIA=CILEVEL(0.95)  
  /POSTHOC=LSD SIDAK ALPHA(0.05).
```

```
EXAMINE VARIABLES=FPNr BY Species_num  
  /PLOT=BOXPLOT  
  /STATISTICS=NONE  
  /NOTOTAL.
```

```
DISCRIMINANT  
  /GROUPS=Species_num(1 8)  
  /VARIABLES=FPNr  
  /ANALYSIS ALL  
  /SAVE=CLASS PROBS  
  /PRIORS EQUAL  
  /STATISTICS=MEAN STDDEV UNIVF BOXM COEFF RAW CORR TABLE  
  /PLOT=COMBINED SEPARATE MAP  
  /PLOT=CASES  
  /CLASSIFY=NONMISSING POOLED.
```

### 2) 问题二 SPSS 语法：

```
ONEWAY MBS VSN CSN GSN FPNr SDLr SCSr SCGr SMr MTr PA PTMr aNDSr SVL TRL HL  
PL ESD HW HH MO FFL HFL  
  BY Species_num  
  /STATISTICS DESCRIPTIVES HOMOGENEITY  
  /MISSING ANALYSIS  
  /CRITERIA=CILEVEL(0.95)  
  /POSTHOC=LSD ALPHA(0.05).
```

### 3) 问题三 SPSS 语法：

```
ONEWAY MBS VSN CSN GSN FPNr SDLr SCSr SCGr SMr MTr PA PTMr aNDSr SVL TRL HL  
PL ESD HW HH MO FFL HFL  
  BY Sex_num  
  /STATISTICS DESCRIPTIVES HOMOGENEITY  
  /MISSING ANALYSIS  
  /CRITERIA=CILEVEL(0.95).
```

```
DISCRIMINANT  
  /GROUPS=Sex_num(1 2)
```

```
/VARIABLES=MBS VSN FPNr SDLr SCGr SMr MTr aNDSr TRL HL PL ESD HW HH MO FFL  
HFL  
/ANALYSIS ALL  
/SAVE=CLASS  
/PRIORS EQUAL  
/STATISTICS=MEAN STDDEV COEFF TABLE  
/PLOT=COMBINED SEPARATE MAP  
/CLASSIFY=NONMISSING POOLED.
```

4) 问题四 SPSS 语法

```
ONEWAY MBS VSN CSN GSN FPNr SDLr SCSr SCGr SMr MTr PA PTMr aNDSr SVL TRL HL  
PL ESD HW HH MO FFL HFL  
BY Species_num  
/STATISTICS DESCRIPTIVES HOMOGENEITY  
/MISSING ANALYSIS  
/CRITERIA=CILEVEL(0.95).
```

```
DISCRIMINANT  
/GROUPS=Species_num(1 8)  
/VARIABLES=MBS VSN FPNr SDLr SCGr SMr MTr aNDSr TRL HL PL ESD HW HH MO FFL  
HFL CSN GSN SCSr PA  
PTMr SVL  
/ANALYSIS ALL  
/SAVE=CLASS  
/PRIORS EQUAL  
/STATISTICS=MEAN STDDEV COEFF TABLE  
/PLOT=COMBINED SEPARATE MAP  
/CLASSIFY=NONMISSING POOLED.
```