

罪案大数据分析

作者：CPDA 数据分析师 [广州] 陈达安

一. 背景

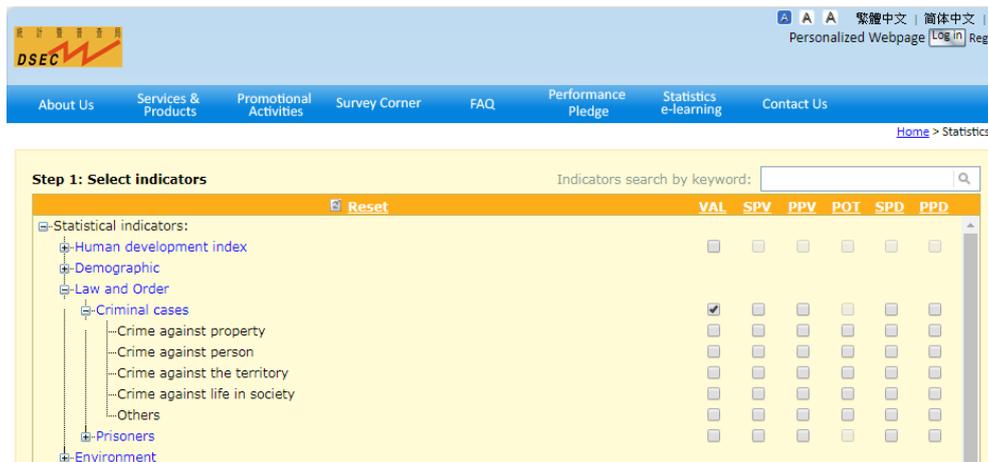
在澳门的警务部门中，不断有警员反映近年罪案不断增多，工作量及压力大增，由于办案警员要在半年内提交总结报告予检察院，每件案件需经过多名警员共同侦办，包括录取口供，侦查，取证，公文往来等多个步骤，一宗案件需花费大量的人力物力及时间调查，若案件数量及工作量继续大增，社会治安环境下降，同时办案警员工作将超出负荷，令警队士气大减，对警队形象及市民造成影响。

故此，利用 R 语言以及时间序列相关分析，结合 2000 年-2018 年 18 年的犯罪统计数字进行分析，并利用 ARIMA 寻找恰当的模式对未来两年案件数目进行预测。

二.数据的来源及解释

澳门统计局的犯罪统计料资(2000-2018 年的刑事案件数量)

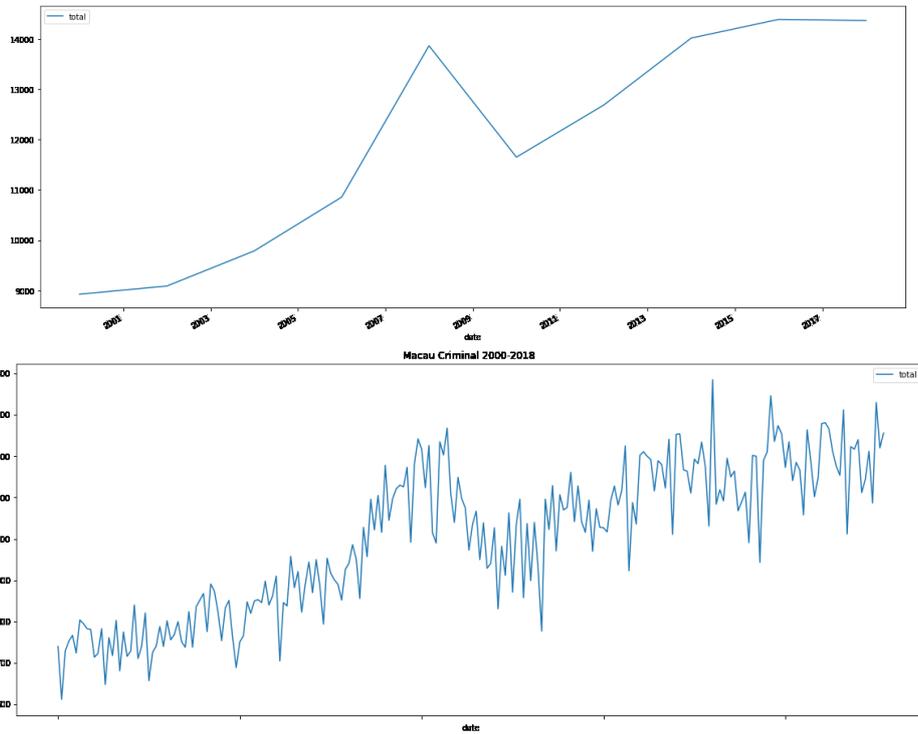
<https://www.dsec.gov.mo/TimeSeriesDatabase.aspx?lang=en-US>



date	total	2017/10/1	1212
2000/1/1	740	2017/11/1	1176
2000/2/1	612	2017/12/1	1154
2000/3/1	729	2018/1/1	1312
2000/4/1	753	2018/2/1	1012
2000/5/1	767	2018/3/1	1223
2000/6/1	724	2018/4/1	1217
2000/7/1	804	2018/5/1	1240
2000/8/1	795	2018/6/1	1112
2000/9/1	783	2018/7/1	1144
2000/10/1	781	2018/8/1	1212
2000/11/1	714	2018/9/1	1087
2000/12/1	723	2018/10/1	1330
2001/1/1	783	2018/11/1	1220
2001/2/1	648	2018/12/1	1256

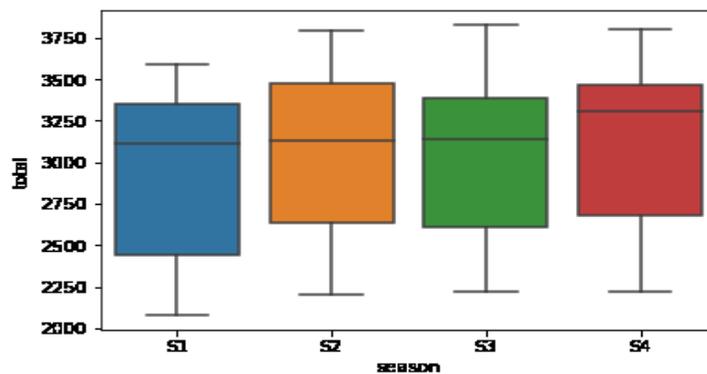
三.数据探索性分析

1. 折线图-查看趋势



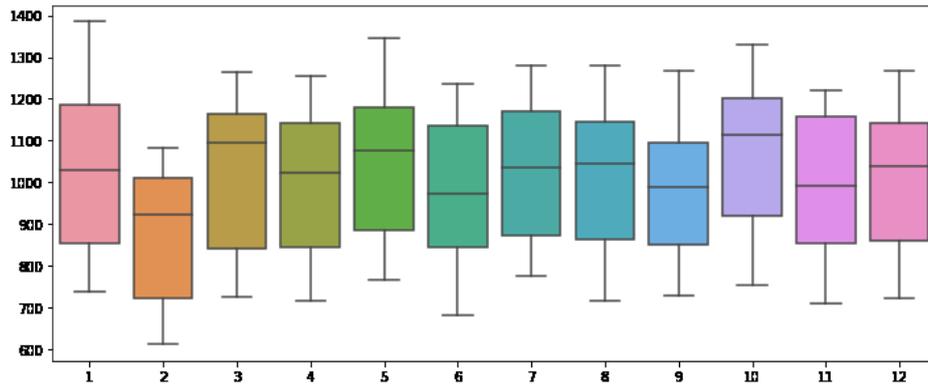
2. 箱型图-查看数据分布

结果:从折线图中可看出犯案数字有上升趋势, 其中可能含有季节性因素, 但仍需进一步分析。

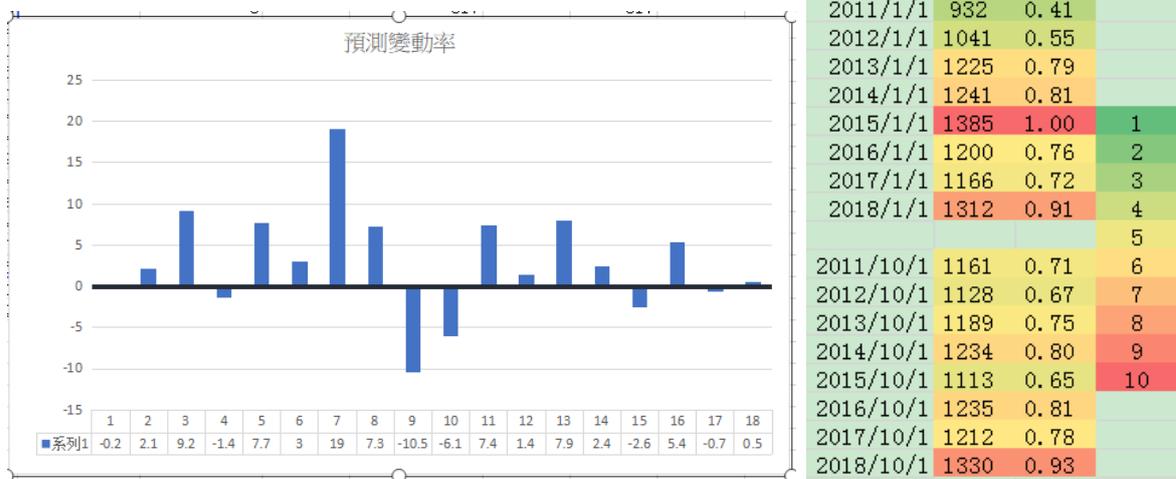


从季度箱型图中可得出: 第1季中位数及第4季中位数偏高, 数据分布集中在顶部, 应作进一步分析。

3. 月度箱型图



4. 棒型图-查看变动率及极值



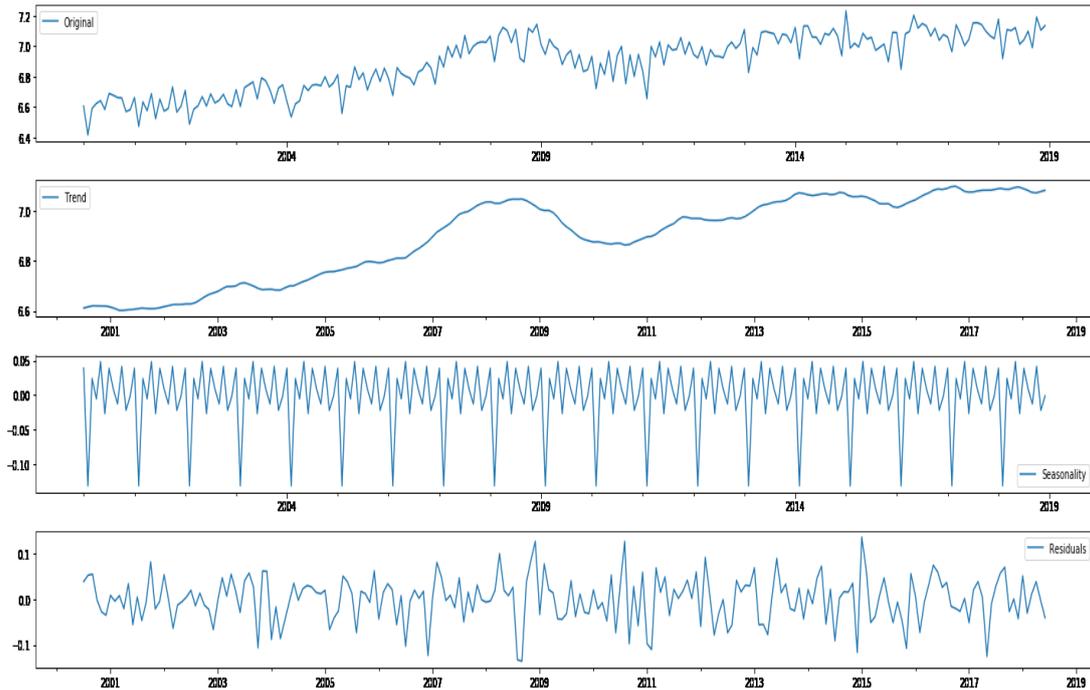
从月箱型图中可看出: 2 月份罪案数字整体比其余月份低, 推断由于 2 月为农历新年, 全城休息庆祝, 澳门市民回乡过年, 故 2 月罪案较低为正常现象。

1 月份犯罪数字出现最大值, 但 Q1,Q3,中位数与其余月份分布大致相同, 1 月偏高并有极值, 2 月减至最低, 说明情况异常, 值得进一步探索, 而 1 月的极值说明被强行拉高, 配合棒型图, 说明 2007 年时罪案数字变化最大, 突发增多, 出现失控的情况。

按照经验中得出, 普遍的犯罪为侵犯财产的罪行, 由于警队宣传与市民防盗防骗意识相关, 故警队的宣传仍有改善空间, 必须在未来的计划在农历新年期间加大宣传, 以免再次出现失控的现象。

10 月份中位数及整体有偏高及整体上升趋势, 10 月为国庆及十一黄金周, 内地旅客会大批涌到澳门购物, 由于人多济迫, 容易制造罪犯犯案的机会, 而针对 10 月的行动, 必须在旅游区, 巴士上加大宣传, 提醒旅客, 以免财物被盗。

5. 分解序列-分解周期性因素及趋势



从图可得知序列有明显的趋势性及季节性因素，噪声有波动，由于序列中呈现线性趋势，故使用较广泛的 ARIMA 时间序列模型进行预测。

四.建模并撰择模型

模型 1--使用 R-AUTO.ARIMA 自动建立模型，RMSE=76

模型 2--使用 DATAHOOP WINTER 指数平滑建立模型，RMSE=80

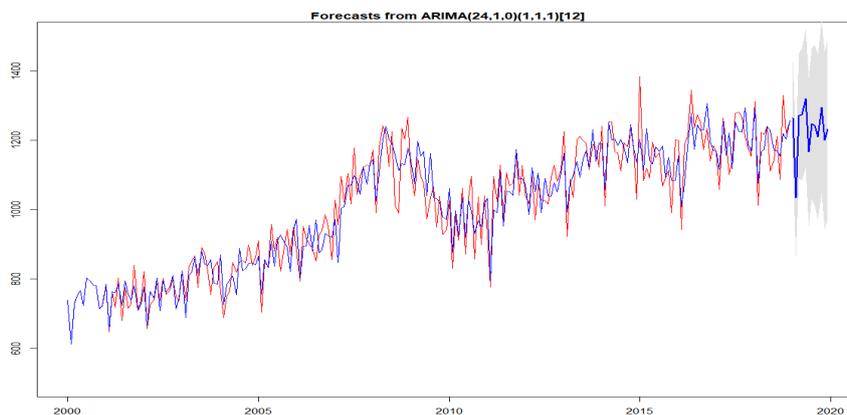
模型 3--通过分析后自建立模型，各评价指标最好，参数为

`arima(x = tsdata, order = c(24, 1, 0), seasonal = list(order = c(1, 1, 1), period = 12))`

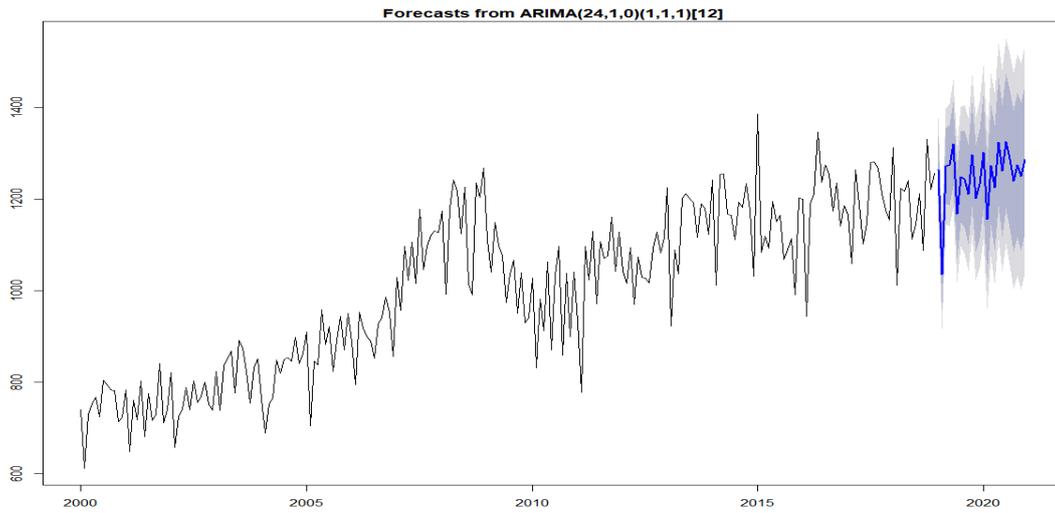
置信水平为 0.05，RMSE=53（撰择自设参数建立模型）

五.预测结果

1. 拟合结果如下:



2. 预测结果如下:



3. 预测年度与月度的数值结果如下:

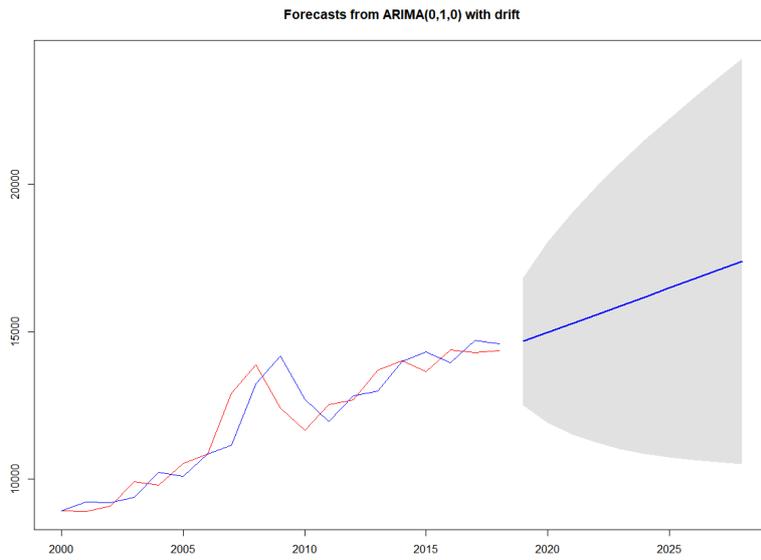
```
> x.fore
Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
Jan 2019 1263.986 1190.8340 1337.137 1152.1099 1375.861
Feb 2019 1035.085 957.5968 1112.573 916.5770 1153.593
Mar 2019 1271.191 1188.6865 1353.696 1145.0111 1397.371
Apr 2019 1274.330 1187.0690 1361.591 1140.8759 1407.784
May 2019 1320.476 1229.6891 1411.262 1181.6297 1459.322
Jun 2019 1167.500 1069.8224 1265.177 1018.1151 1316.885
Jul 2019 1248.203 1148.7488 1347.657 1096.1009 1400.305
Aug 2019 1242.964 1136.9983 1348.929 1080.9035 1405.024
Sep 2019 1210.039 1101.0846 1318.994 1043.4074 1376.672
Oct 2019 1295.828 1182.0277 1409.629 1121.7854 1469.871
Nov 2019 1201.038 1084.5776 1317.499 1022.9269 1379.150
Dec 2019 1232.128 1112.9648 1351.292 1049.8835 1414.373
Jan 2020 1301.754 1177.9436 1425.565 1112.4022 1491.106
Feb 2020 1154.714 1028.5674 1280.860 961.7896 1347.638
Mar 2020 1273.263 1142.9297 1403.597 1073.9352 1472.591
Apr 2020 1224.621 1089.1065 1360.135 1017.3694 1431.873
May 2020 1324.259 1184.3316 1464.187 1110.2583 1538.260
Jun 2020 1261.211 1117.2960 1405.126 1041.1119 1481.311
Jul 2020 1325.455 1178.7828 1472.127 1101.1393 1549.771
Aug 2020 1286.791 1136.1460 1437.436 1056.3994 1517.183
Sep 2020 1239.083 1085.0416 1393.124 1003.4972 1474.668
Oct 2020 1274.228 1116.0449 1432.411 1032.3078 1516.148
Nov 2020 1248.774 1087.2768 1410.272 1001.7852 1495.764
Dec 2020 1285.491 1123.2813 1447.701 1037.4125 1533.570
```

	Lo80	预测值	Hi80
2019	13583	14762	15929
2020	13462	15199	16928

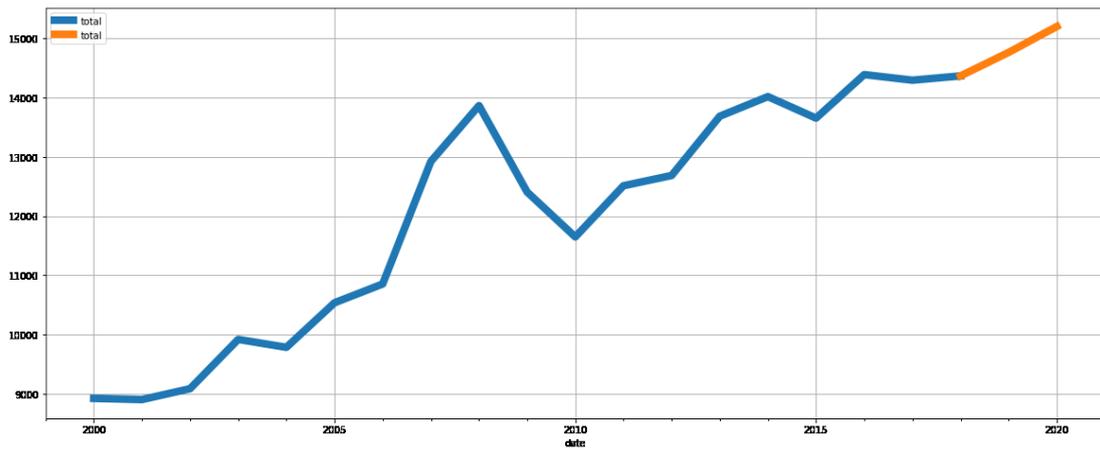
4. 检验预测结果及对比

		预测值	真实值	差值	Lo 80	真实值	Hi 80
Jan	2019	1264	1320	-56	1191	1320	1337
Feb	2019	1035	893	142	958	893	1113
Mar	2019	1271	1151	120	1189	1151	1354
Apr	2019	1274	1173	101	1187	1173	1362
May	2019	1320	1215	105	1230	1215	1411
Jun	2019	1168	1168	-1	1070	1168	1265
Jul	2019	1248	1283	-35	1149	1283	1348

5. 年度趋势预测

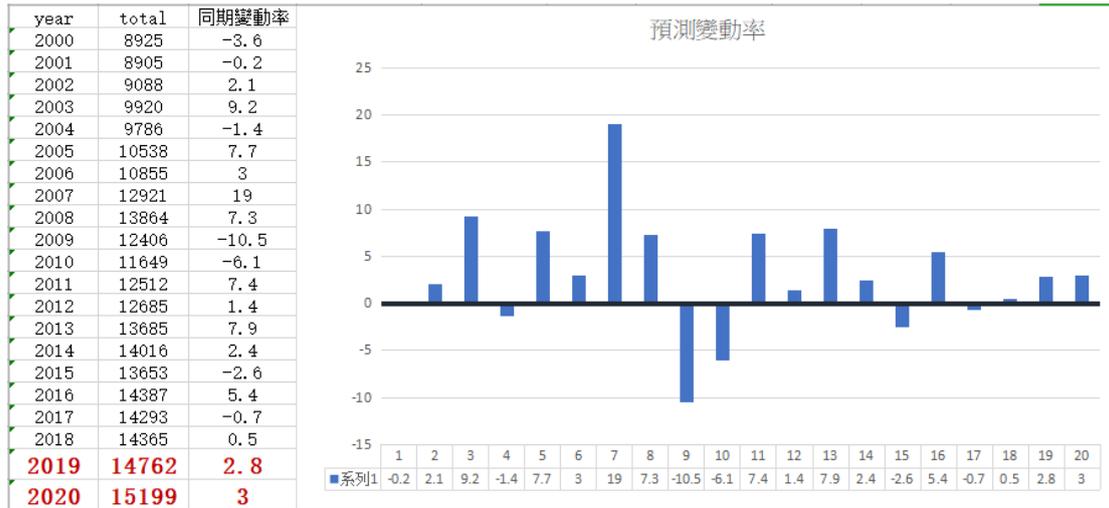


6. 月度趋势预测



7. 预测犯罪案件宗数及同期变动率

		司法與罪案	
		罪案	
		總數	
		宗	同期變動率
2019	1月	1,320	0.6
	2月	893	-11.8
	3月	1,151	-5.9
	4月	1,173	-3.6
	5月	1,215	-2.0
	6月	1,168	5.0
	7月	1,283	12.2



一、根据图一撰择最优模型的拟合结果显示，模型拟合结果好，没有较大偏差。

二、根据图三，得出预测的月度与年度的预测值及预测区间。

三、根据图四，在给置信度 0.05 的前提下，真实值落在我们估计区间之内的概率为 99.5，而真实值落在比正常预测的估区间窄(80%置信水平)的范围附近，说明模型的预测结果较为准确。

四、根据图五，图六，进行年度与月度的预测，均显示犯罪数字在未来两年呈上升趋势，并创新高。

五、而图七的显示今年 2019 总犯案数字约 14700 宗，而至 2020 年将达 15100 宗，预测未来两年的犯罪变动率每年上升 3%，创历年新高。

六.总结

为了使数据具有可比性，取 2010-2018 的数据正规化，将罪案数字缩放到[0-1]之间作，并分 10 级作对比。

	2000		2001		2002		2003		2004		2005		2006		2007		2008		2009		2010	
1月	740	0.17	783	0.22	821	0.27	824	0.27	763	0.20	910	0.39	887	0.36	1028	0.54	1173	0.73	1109	0.64	1027	0.54
2月	612	0.00	648	0.05	657	0.06	738	0.16	689	0.10	705	0.12	794	0.24	957	0.45	992	0.49	1040	0.55	831	0.28
3月	729	0.15	761	0.19	725	0.15	836	0.29	751	0.18	846	0.30	953	0.44	1096	0.63	1180	0.73	1149	0.69	982	0.48
4月	753	0.18	718	0.14	741	0.17	852	0.31	766	0.20	838	0.29	917	0.39	1022	0.53	1242	0.82	1098	0.63	912	0.39
5月	767	0.20	803	0.25	788	0.23	868	0.33	848	0.31	958	0.45	901	0.37	1105	0.64	1217	0.78	1075	0.60	1063	0.58
6月	724	0.14	681	0.09	740	0.17	776	0.21	820	0.27	882	0.35	890	0.36	1016	0.52	1124	0.66	973	0.47	871	0.34
7月	804	0.25	775	0.21	802	0.25	891	0.36	850	0.31	921	0.40	852	0.31	1178	0.73	1226	0.79	1036	0.55	1033	0.54
8月	795	0.24	716	0.13	756	0.19	874	0.34	853	0.31	823	0.27	926	0.41	1045	0.56	1014	0.52	1067	0.59	1096	0.63
9月	783	0.22	729	0.15	769	0.20	823	0.27	846	0.30	891	0.36	941	0.43	1097	0.63	990	0.49	950	0.44	858	0.32
10月	781	0.22	840	0.29	800	0.24	754	0.18	898	0.37	944	0.43	986	0.48	1121	0.66	1235	0.81	1039	0.55	1037	0.55
11月	714	0.13	711	0.13	751	0.18	833	0.29	840	0.29	870	0.33	952	0.44	1130	0.67	1203	0.76	929	0.41	899	0.37
12月	723	0.14	740	0.17	738	0.16	851	0.31	862	0.32	950	0.44	856	0.32	1126	0.66	1268	0.85	941	0.43	1040	0.55
	2011		2012		2013		2014		2015		2016		2017		2018							
1月	932	0.41	1041	0.55	1225	0.79	1241	0.81	1385	1.00	1200	0.76	1166	0.72	1312	0.91			1			
2月	777	0.21	1016	0.52	923	0.40	1011	0.52	1084	0.61	943	0.43	1058	0.58	1012	0.52			2			
3月	1096	0.63	1094	0.62	1088	0.62	1253	0.83	1119	0.66	1190	0.75	1264	0.84	1223	0.79			3			
4月	1023	0.53	970	0.46	1036	0.55	1254	0.83	1092	0.62	1210	0.77	1188	0.75	1217	0.78			4			
5月	1129	0.67	1073	0.60	1202	0.76	1167	0.72	1195	0.75	1346	0.95	1102	0.63	1240	0.81			5			
6月	971	0.46	1028	0.54	1211	0.77	1164	0.71	1150	0.70	1236	0.81	1147	0.69	1112	0.65			6			
7月	1107	0.64	1027	0.54	1200	0.76	1111	0.65	1164	0.71	1274	0.86	1279	0.86	1144	0.69			7			
8月	1070	0.59	1017	0.52	1192	0.75	1193	0.75	1068	0.59	1254	0.83	1281	0.87	1212	0.78			8			
9月	1076	0.60	1093	0.62	1116	0.65	1182	0.74	1090	0.62	1173	0.73	1266	0.85	1087	0.61			9			
10月	1161	0.71	1128	0.67	1189	0.75	1234	0.80	1113	0.65	1235	0.81	1212	0.78	1330	0.93			10			
11月	1042	0.56	1082	0.61	1180	0.73	1175	0.73	991	0.49	1141	0.68	1176	0.73	1220	0.79						
12月	1128	0.67	1116	0.65	1123	0.66	1031	0.54	1202	0.76	1185	0.74	1154	0.70	1256	0.83						

(一) 观看 2000-2018 的案件有不断上升趋势，特别在 2016-2019 年，警示级数处于第

8,9 级, 而根据预测的情况, 今年与下年的情况将会再创新高, 2019 年为接近 14700 宗[上下限为 13583-15929], 2020 年为 15100 宗[上下限为 13462-16928], 每年上升 3%, 显示社会将面临罪犯高企的情况, 而前线警员将处于工作量继续增多的困境, 这样大大对社会及警队造成负面影响。

(二) 2000-2018 年间的数据显示 1 月份为犯罪高频月份, 案发率及增长率偏高, 其中于 2015 年 1 月出现最大值 1385 宗, 2013 年-2018 年平均案件为 1245 宗, 偏向高值, 2018 年 1 月的案件宗数为 1312, 级数为第 9 级, 接近极值, 由于案件发生在农历新年前夕, 建议在 12 月至 1 月期间加强向市民作宣传, 以防犯罪率再持续上升, 在新年前夕重新部署警力安排, 加大宣传及警力巡逻, 最后检查部署成果及绩效。

(三) 数据中显示 10 月的罪发率偏高, 相信为国庆十一黄金周影响, 综观过去每年 10 月份的案发率逐渐上升, 过去整个 2011-2018 年 10 月份的案发率逐渐上升, 2018 年 10 月的案件宗数为 1330, 级数为第 9 级, 接近极值, 这一规律可作为警示信号, 表示今年将有可能再创新高, 鉴于会有大量内地旅客到澳门旅游区购物, 建议在口岸加强宣传及增大告示内容, 并派便衣及巡逻警员到旅游及购物区巡逻, 防止小偷有机可乘, 以产生阻吓作用, 在适当部署后, 查看日后的部署效果及绩效。

(四) 减在少变动率及犯案率直接降低工作量的前提下, 建议首要任务在 1 月及 10 月加强宣传及打击犯罪, (1)10 月的黄金周应在主要购物区加强宣传及告示, 告示牌应放置足够大幅以便每位旅客都能清晰看到, 内容表达直接明确, 主要警力应部署在店铺的营业及旅客的出门时段, 时间为早上 11 时至晚上 11 时。(2)在 1 月前应提前准备对社区和旅客作宣传, 并在 1 月期间针对社区及旅游区部署适当警力, 对应社区则撰择在晚上及人少的地方加强巡逻, 以防发生入屋盗窃。

(五) 由于罪案直接影响警员的工作量及压力, 建议先鼓励, 表扬及赞赏前线警员及对他们的工作作出认同及肯定, 以提升警队士气, 往后将案件分类重新调配, 适当分配到次紧急的侦查部门, 以缓解某部门工作量过多的情况。

(六) 未来两年的犯罪率将会持续上升, 并以每年约 3%的增长增幅, 对于警员, 每增加一宗案件则增加一担压力, 减少罪犯发生及缓解前线警员压力的任务, 重要且紧急。

在此本人希望借此分析报告对市民及警务部门作出警示, 市民自身做好防护措施, 小心保管财物能直接断绝罪案发生, 除保障自身的财产利益外, 更能维持社区稳定安全; 而警务部门在市民的配合下, 工作成效事半功倍, 双得益彰。

在未来让市民与警务部门积极配合, 令澳门各社区安定繁荣, 长治久安。