

《统计学大作业报告》

——基于上海每日新冠疫情数据的线性回归分析

作者姓名：石磊

（学院：信息学院 专业：数据科学与大数据技术 学号：2052515）

摘要：上海疫情的亲历者，我对上海疫情的整个发展趋势以及各个阶段进行统计分析。我将上海疫情按时间分为了四个阶段。在每个阶段中我们分别对疫情数据进行线性回归分析，分别是：确诊病例人数随时间变化线性回归分析、无症状感染者人数随时间变化线性回归分析、确诊病例人数随无症状感染者人数变化进行线性回归分析。分析过程中，我首先计算其相关性系数 R 、回归方程；并且比较 R 方值， F 检验来判断线性相关性；接着计算回归曲线在 95% 置信水平下的置信区间以及均值的预测区间；又对相关系数进行显著性检验确定其显著性程度；最后得出确诊病例人数随时间变化、无症状感染者人数随时间变化在疫情开始发展阶段呈强正相关性，在疫情稳定结束阶段呈强负相关性；确诊病例人数随无症状感染者人数增加呈强正相关，反之亦然。

关键词：线性回归分析， R -square， F 检验， R 方值，置信区间。

引言：在这次疫情中，通过对上海官方发布的上海疫情数据，对数据进行回归分析，揭露数据底下蕴含着统计的思想，让我们利用统计的知识重新了解上海的抗疫历程，通过数据讲述疫情的整个“生命周期”。

1 收集数据

1.1 数据来源

在上海发布的从 3 月 12 日至 5 月 25 日这 75 天中的官方上海疫情数据。
数据列别包括：日期、本土确诊病例人数、本土无症状病例人数、本土确诊+无症状病例人数、时间间隔。

图 1: 上海疫情数据集展示图 (1)

date	quezhen	wu	quewu	time_code
12-Mar	1	64	65	1
13-Mar	41	128	169	2
14-Mar	9	130	139	3
15-Mar	5	197	202	4
16-Mar	8	150	158	5
17-Mar	57	203	260	6
18-Mar	8	366	374	7
19-Mar	17	492	509	8
20-Mar	24	734	758	9
21-Mar	31	865	896	10
22-Mar	4	977	981	11
23-Mar	4	979	983	12
24-Mar	29	1580	1609	13
25-Mar	38	2231	2269	14
26-Mar	45	2631	2676	15
27-Mar	50	3450	3500	16
28-Mar	96	4381	4477	17
29-Mar	326	5656	5982	18
30-Mar	355	5298	5653	19
31-Mar	358	4144	4502	20
1-Apr	260	6051	6311	21
2-Apr	438	7788	8226	22
3-Apr	425	8581	9006	23
4-Apr	268	13086	13354	24
5-Apr	311	16766	17077	25
6-Apr	322	19660	19982	26
7-Apr	824	20398	21222	27
8-Apr	1015	22609	23624	28
9-Apr	1006	23937	24943	29
10-Apr	914	25173	26087	30
11-Apr	994	22348	23342	31
12-Apr	1189	25141	26330	32
13-Apr	2573	25146	27719	33
14-Apr	3200	19872	23072	34

图 2: 上海疫情数据集展示图 (2)

15-Apr	3590	19923	23513	35
16-Apr	3238	21582	24820	36
17-Apr	2417	19831	22248	37
18-Apr	3084	17332	20416	38
19-Apr	2494	16407	18901	39
20-Apr	2634	15861	18495	40
21-Apr	1931	15698	17629	41
22-Apr	2736	20634	23370	42
23-Apr	1401	19657	21058	43
24-Apr	2472	16983	19455	44
25-Apr	1661	15319	16980	45
26-Apr	1606	11956	13562	46
27-Apr	1292	9330	10622	47
28-Apr	5487	9545	15032	48
29-Apr	1249	8932	10181	49
30-Apr	788	7084	7872	50
1-May	727	6606	7333	51
2-May	274	5395	5669	52
3-May	260	4722	4982	53
4-May	261	4390	4651	54
5-May	245	4024	4269	55
6-May	253	3961	4214	56
7-May	215	3760	3975	57
8-May	322	3625	3947	58
9-May	234	2780	3014	59
10-May	228	1259	1487	60
11-May	144	1305	1449	61
12-May	227	1869	2096	62
13-May	194	1487	1681	63
14-May	166	1203	1369	64
15-May	69	869	938	65
16-May	77	746	823	66
17-May	96	759	855	67
18-May	82	637	719	68
19-May	88	770	858	69

图 3: 上海疫情数据集展示图 (3)

20-May	84	784	868	70
21-May	52	570	622	71
22-May	55	503	558	72
23-May	58	422	480	73
24-May	44	343	387	74
25-May	48	290	338	75

2 问题二的求解与分析

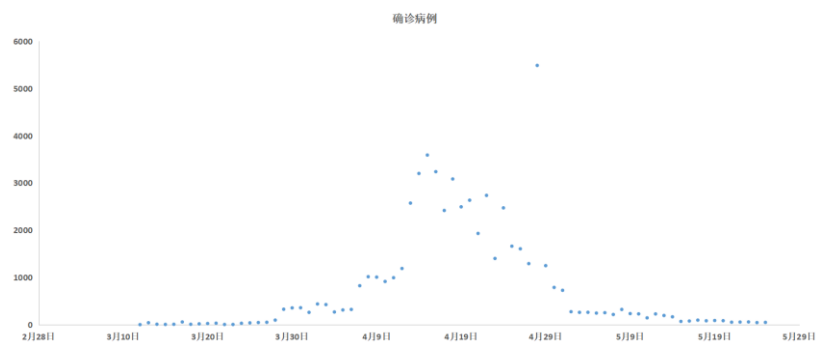
2.1 数据散点图的绘画与分析

散点图的绘画使用软件 Python，我们散点图绘画过程，部分图横坐标为日期时间，所以为了消除非数值型数据作为成图源数据带来的不方便，这里创建了新的一栏来表示持续天数。

2.1.1 散点图展示 (代码见附录：源代码 1)

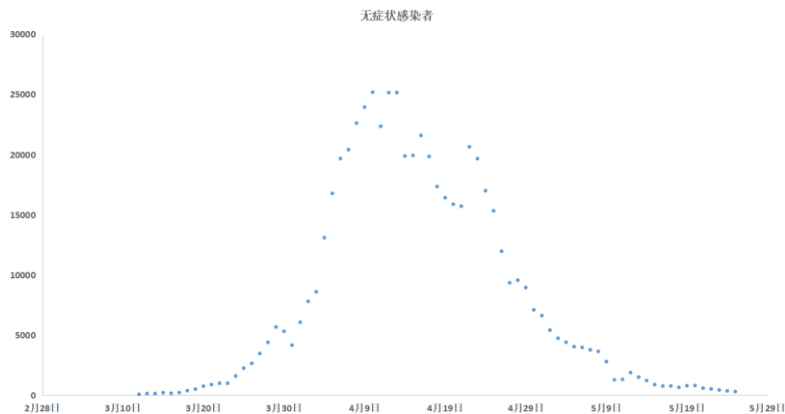
(1) 本土确诊病例人数随时间变化散点

图 4：本土确诊病例人数随时间变化散点



(2) 本土无症状病例人数随时间变化散点图

图 5：本土无症状病例人数随时间变化散点图



(3) 本土无症状与确诊病例人数随时间变化散点图

图 6：本土无症状与确诊病例人数随时间变化散点图

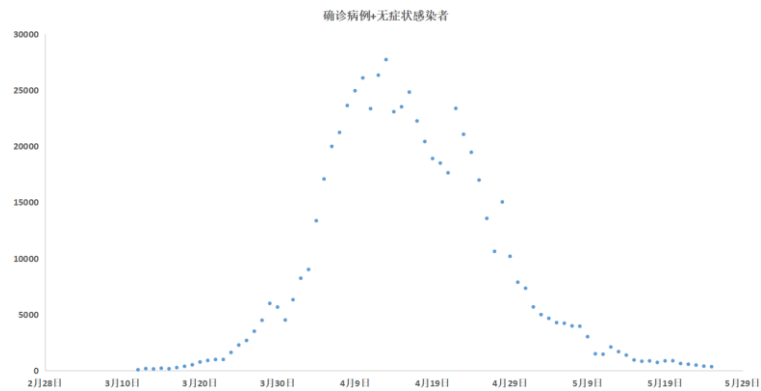


图 7：确诊病例人数随无症状人数散点图

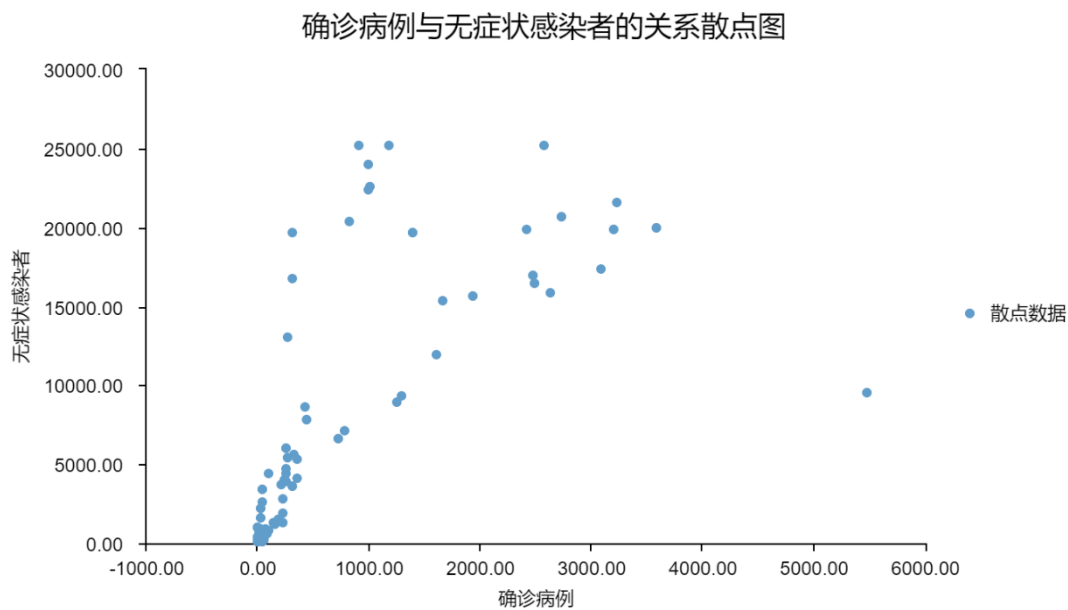
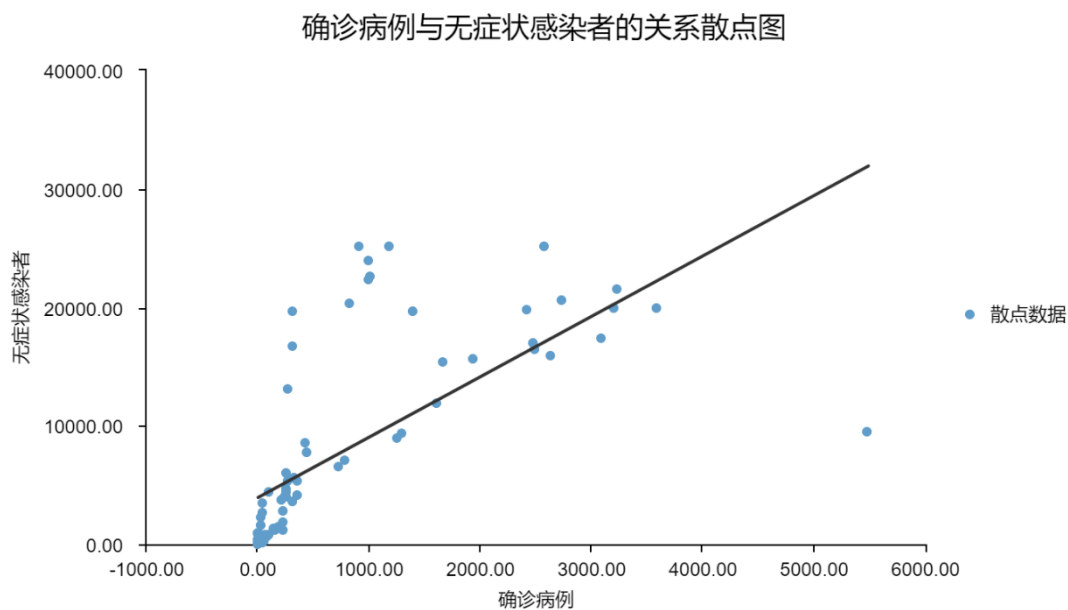


图 8：确诊病例与无症状感染者的关系散点图



散点数据线性拟合公式为：无症状感染者 = 3939.629 + 5.099*确诊病例， R 方值为 0.466。

3 问题三的求解与分析

3.1 确诊病例数据区间的划分

上海疫情数据集共 75 个样本数据，从 3 月 12 号到 5 月 25 号，这里根据问题二的散点图大致可以看出疫情趋势，为更加详细的分析，故将样本数据划分为四个区间（3 个阶段）分别进行线性回归分析。

- （1）区间 1：3 月 12 号到 3 月 29 号
- （2）区间 2：3 月 30 号到 4 月 8 号
- （3）区间 3：4 月 9 号到 4 月 27 号
- （4）区间 4：4 月 28 号到 5 月 25 号

3.2 疫情不同阶段确诊病例与时间的线性分析

由于篇幅原因，这里挑标志性的两个时间段来进行分析，剩余的见附录二

（1）3.12-3.29 阶段线性分析

图 9：确诊病例随时间拟合图

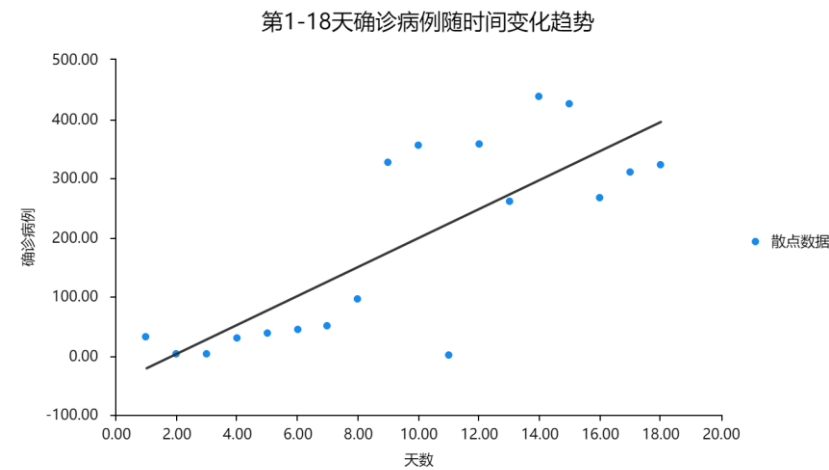
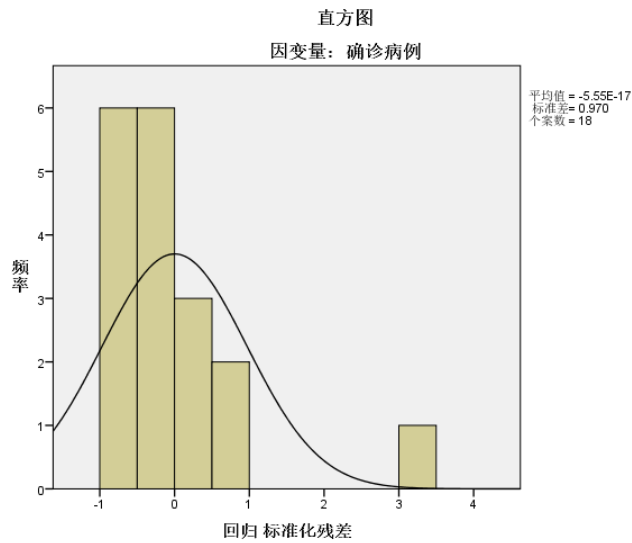


图 10：确诊病例回归标准化残差图



残差:

[0.36138415 0.91160373 0.23089338 0.04211398 -0.03074578 0.62616588
-0.26670477 -0.24339162 -0.25187374 -0.26089483 -0.81378385 -0.94473112
-0.66060243 -0.64338577 -0.6607535 -0.71436949 -0.08093388 11.24018815]

散点数据线性拟合公式为: 确诊病例 = -45.915 + 24.488*天数, R 方值为 0.629。

表 1: 线性回归分析结果-简化格式

	回归系数	95% CI	VIF
常数	-45.915 (-0.901)	-145.752 ~ 53.922	-
天数	24.488** (5.204)	15.265 ~ 33.711	1.000
样本量		18	
R^2		0.629	
调整 R^2		0.605	
F 值		$F(1,16)=27.079, p=0.000$	

因变量: 确诊病例

D-W 值: 1.153

表 1：线性回归分析结果-简化格式

回归系数	95% CI	VIF
------	--------	-----

* $p<0.05$ ** $p<0.01$ 括号里面为 t 值

从上表可知，将天数作为自变量，而将确诊病例作为因变量进行线性回归分析，从上表可以看出，模型公式为：确诊病例=-45.915 + 24.488*天数，模型 R 方值为 0.629，意味着天数可以解释确诊病例的 62.9%变化原因。

线性关系显著性检验： F 检验是回归方程总体的显著性检验， F 检验主要是检验因变量与自变量之间的线性关系是否显著，用线性模型来描述他们之间的关系是否恰当。

对模型进行 F 检验时发现模型通过 F 检验 ($F=27.079$, $p=0.000<0.05$)，也即说明天数一定会对确诊病例产生影响关系。

天数的回归系数值为 24.488 ($t=5.204$, $p=0.000<0.01$)，意味着天数会对确诊病例产生显著的正向影响关系。

同时由上表可知，95%的置信区间为[15.256, 33.711]。

总结分析可知：天数全部均会对确诊病例产生显著的正向影响关系。

(2) 4.9-4.27 阶段线性分析

图 11：第 29-47 天确诊病例随时间拟合图

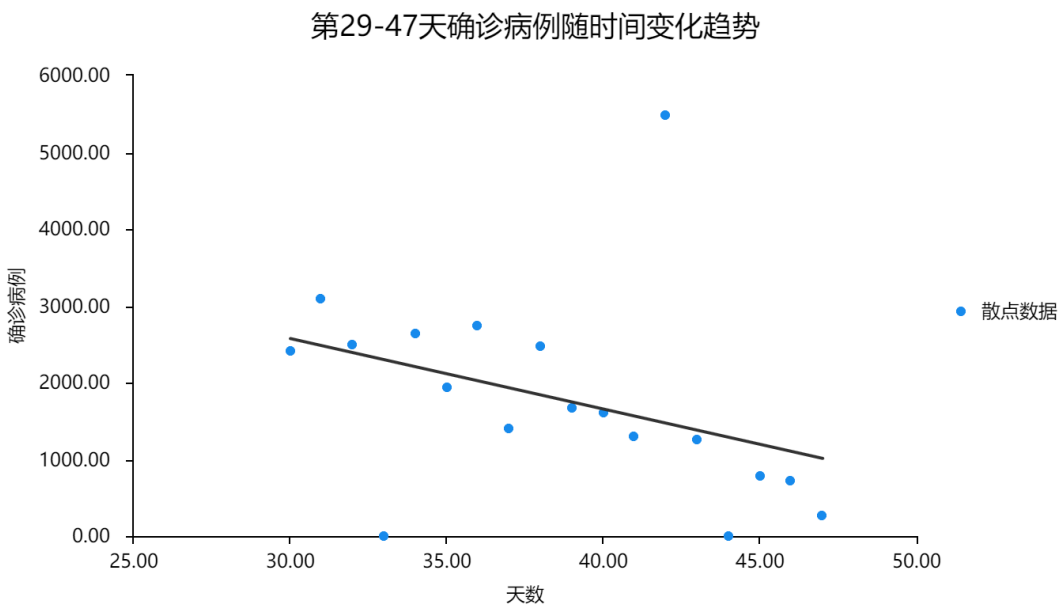
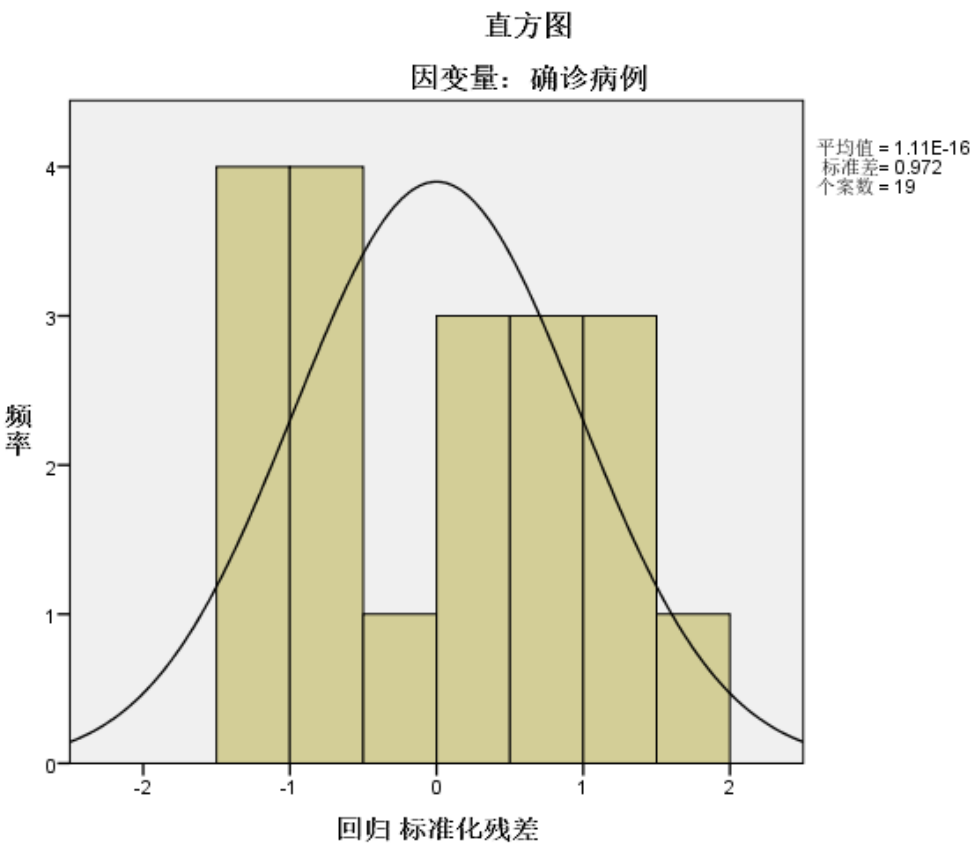


图 12：第 29-47 天确诊病例回归标准化残差图



残差：
[-1.31476452 -1.4360097 -1.31544566 -1.05322441 0.58996213 1.35802742
1.89081118 1.36059734 0.34202405 1.12448565 0.40363368 0.55380229
-0.26552249 0.65379293 -0.93518525 0.32215547 -0.66040936 -0.75530184
-1.20729449]

散点数据线性拟合公式为：确诊病例 = 5702.484-100.889*天数，R方值为 0.182。

表 2：线性回归分析结果-简化格式

	回归系数	95% CI	VIF
常数	5702.484* (2.866)	1802.164 ~ 9602.804	-

表 2：线性回归分析结果-简化格式

	回归系数	95% CI	VIF
天数	-100.889 (-1.946)	-202.480 ~ 0.701	1.000
样本量		19	
R^2		0.182	
调整 R^2		0.134	
F 值		$F(1,17)=3.789, p=0.068$	

因变量：确诊病例

D-W 值：1.768

* $p<0.05$ ** $p<0.01$ 括号里面为 t 值

从上表可知，将天数作为自变量，而将确诊病例作为因变量进行线性回归分析，从上表可以看出，模型公式为：确诊病例=5702.484-100.889*天数，模型 R 方值为 0.182，意味着天数可以解释确诊病例的 18.2%变化原因。对模型进行 F 检验时发现模型并没有通过 F 检验 ($F=3.789$, $p=0.068>0.05$)，也即说明天数并不会对确诊病例产生影响关系，因而不能具体分析自变量对于因变量的影响关系。

由上表中的 95% CI 一栏可得到计算后 95%的置信区间
从上表可知，将天数作为自变量，而将确诊病例作为因变量进行线性回归分析，从上表可以看出，模型 R 方值为 0.182，意味着天数可以解释确诊病例的 18.2%变化原因。

4 问题四的求解与分析

4.1 无症状感染者的统计分析

此处选取两个时期进行分析，理由同上，剩余两个时间段无症状感染者统计表见附录 3
(1) 对 1-10 天无症状感染者随时间变化分析

图 13：第 1-10 天无症状感染者病例回归拟合图

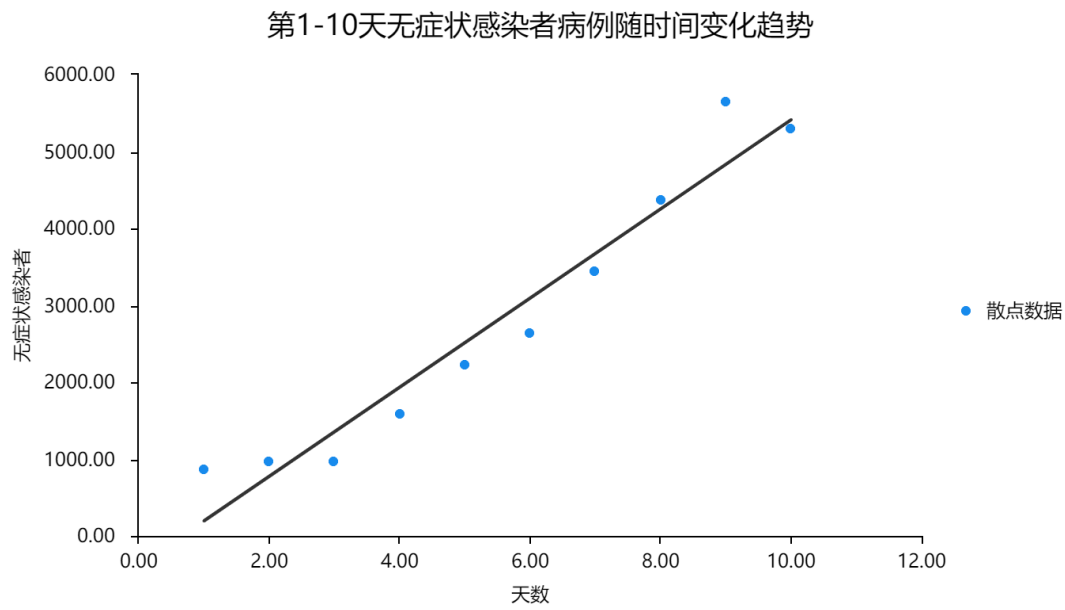
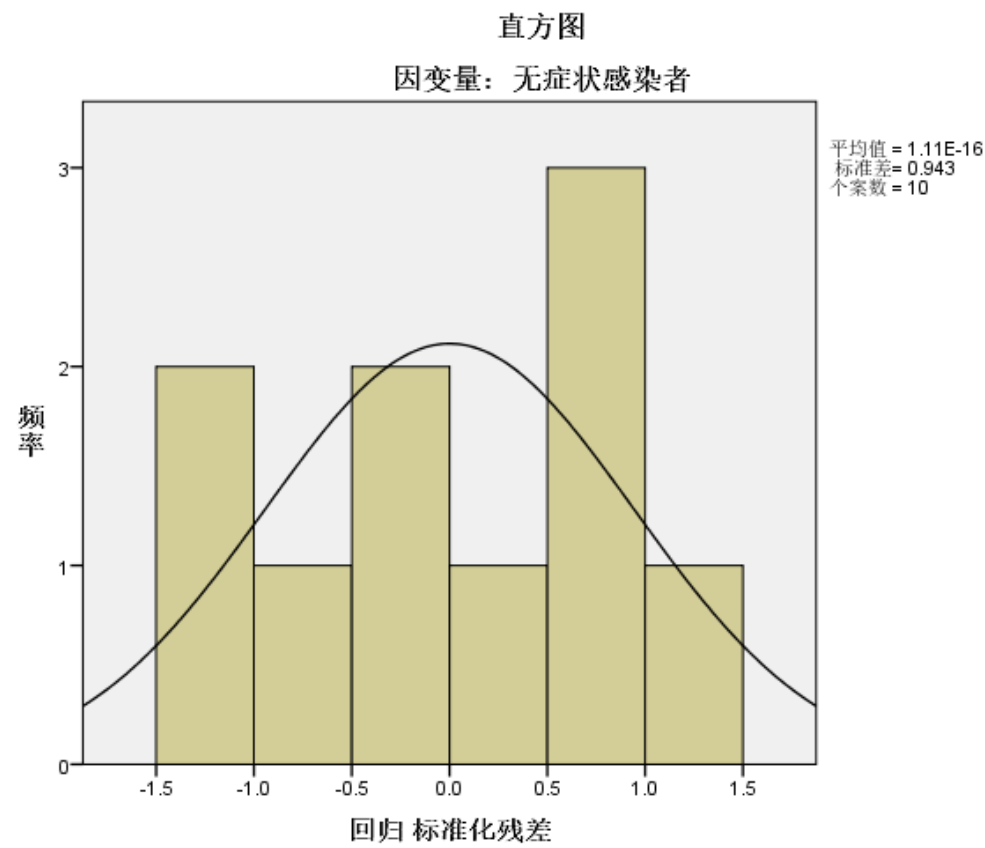


图 14：第 1-10 天无症状病例回归标准化残差图



散点数据线性拟合公式为：无症状感染者 = $-384.200 + 579.818 \times \text{天数}$ ， R 方值为 0.939。

表 3：第 1-10 天无症状感染者病例线性回归分析结果-简化格式

	回归系数	95% CI	VIF
常数	-384.200 (-1.183)	-1020.585 ~ 252.185	-
天数	579.818** (11.080)	477.255 ~ 682.381	1.000
样本量		10	
R^2		0.939	
调整 R^2		0.931	
F 值		$F(1,8)=122.772, p=0.000$	

因变量：无症状感染者

D-W 值：1.181

* $p<0.05$ ** $p<0.01$ 括号里面为 t 值

从上表可知，将天数作为自变量，而将无症状感染者作为因变量进行线性回归分析，从上表可以看出，模型公式为：无症状感染者=-384.200 + 579.818*天数，模型 R 方值为 0.939，意味着天数可以解释无症状感染者的 93.9%变化原因。对模型进行 F 检验时发现模型通过 F 检验 ($F=122.772$, $p=0.000<0.05$)，也即说明天数一定会对无症状感染者产生影响关系，最终具体分析可知：

天数的回归系数值为 579.818 ($t=11.080$, $p=0.000<0.01$)，意味着天数会对无症状感染者产生显著的正向影响关系。

总结分析可知：天数全部均会对无症状感染者产生显著的正向影响关系。

从上表可知，将天数作为自变量，而将无症状感染者作为因变量进行线性回归分析，从上表可以看出，模型 R 方值为 0.939，意味着天数可以解释无症状感染者的 93.9%变化原因。

(2) 对第 24-54 天无症状感染者进行分析

图 15：第 24-54 天无症状病例随时间变化回归拟合图

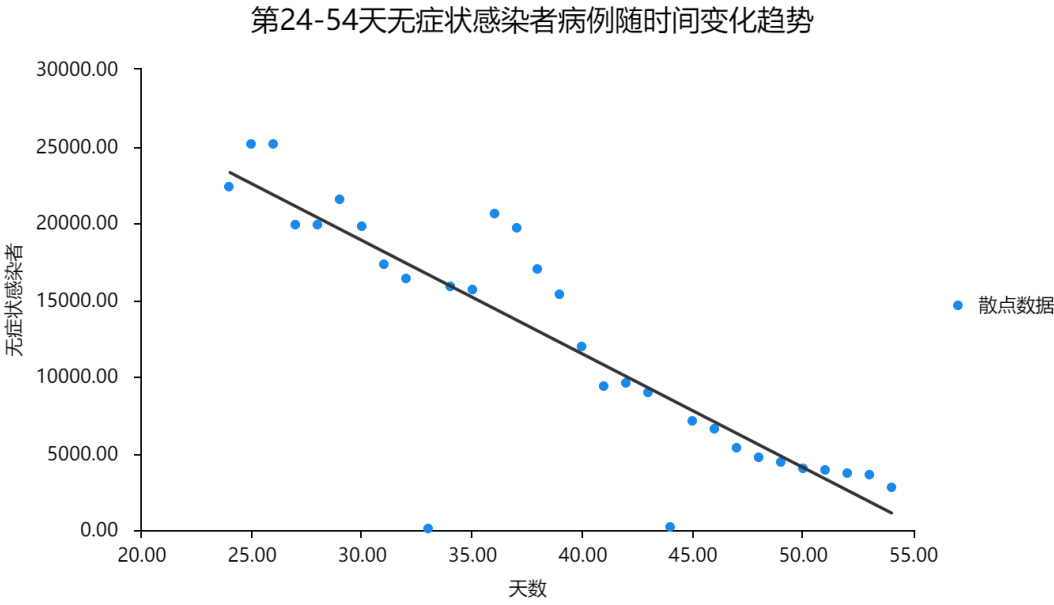
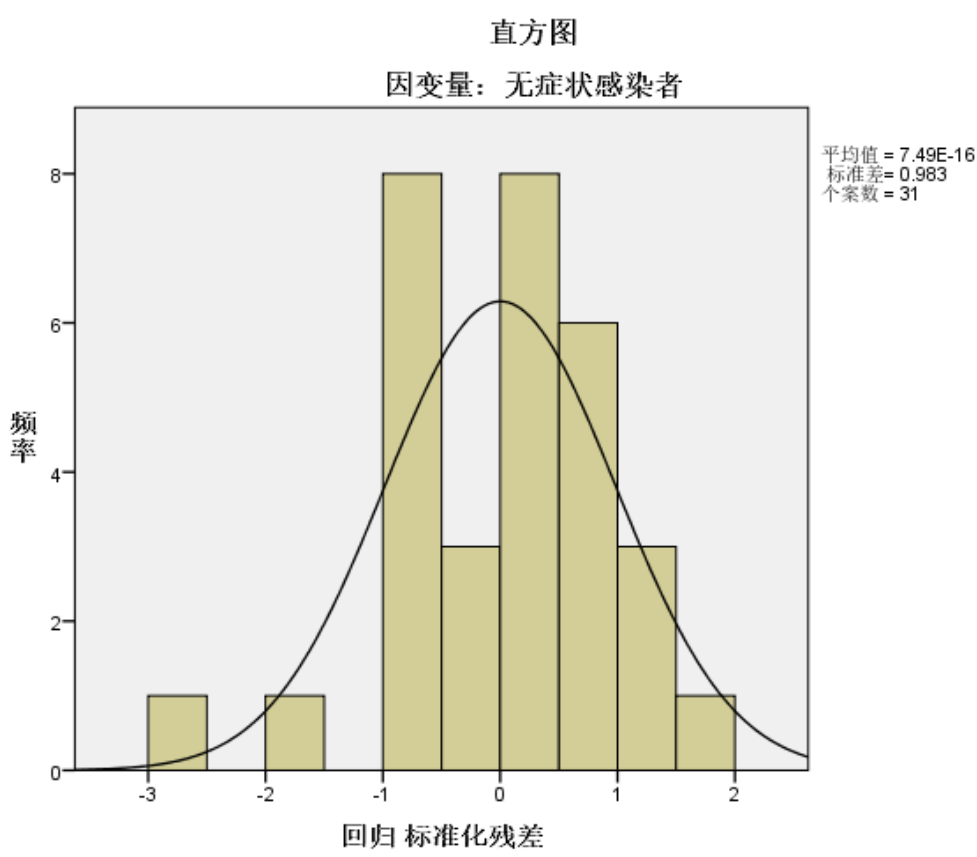


图 16：第 24-54 天无症状病例回归标准化残差图



散点数据线性拟合公式为：无症状感染者 = 41027.534-739.192*天数，R 方值为 0.738。

表 4：第 24-54 天无症状病例随时间变化线性回归分析结果-简化格式

	回归系数	95% CI	VIF
常数	41027.534** (12.532)	34611.103 ~ 47443.966	-
天数	-739.192** (-9.035)	-899.552 ~ -578.831	1.000
样本量		31	
R^2		0.738	
调整 R^2		0.729	
F 值		$F(1,29)=81.623, p=0.000$	

因变量：无症状感染者

D-W 值：0.451

* $p<0.05$ ** $p<0.01$ 括号里面为 t 值

从上表可知，将天数作为自变量，而将无症状感染者作为因变量进行线性回归分析，从上表可以看出，模型公式为：无症状感染者=41027. 534-739. 192*天数，模型 R 方值为 0. 738，意味着天数可以解释无症状感染者的 73. 8%变化原因。对模型进行 F 检验时发现模型通过 F 检验 ($F=81. 623, p=0. 000<0. 05$)，也即说明天数一定会对无症状感染者产生影响关系，最终具体分析可知：

天数的回归系数值为-739. 192 ($t=-9. 035, p=0. 000<0. 01$)，意味着天数会对无症状感染者产生显著的负向影响关系。

总结分析可知：天数全部均会对无症状感染者产生显著的负向影响关系。

从上表可知，将天数作为自变量，而将无症状感染者作为因变量进行线性回归分析，从上表可以看出，模型 R 方值为 0. 738，意味着天数可以解释无症状感染者的 73. 8%变化原因。

5 问题五的求解与分析

5.2 无症状感染者与确诊病例的统计分析

这里我将其分为确诊病例<2417 人和确诊病例>2417 人两类，（此处只列一类，还有一类图表

在附录中列出)

(1) 无症状感染者与确诊病例的关系图

图 17：第 24-54 天无症状病例与确诊病例回归标准化残差图

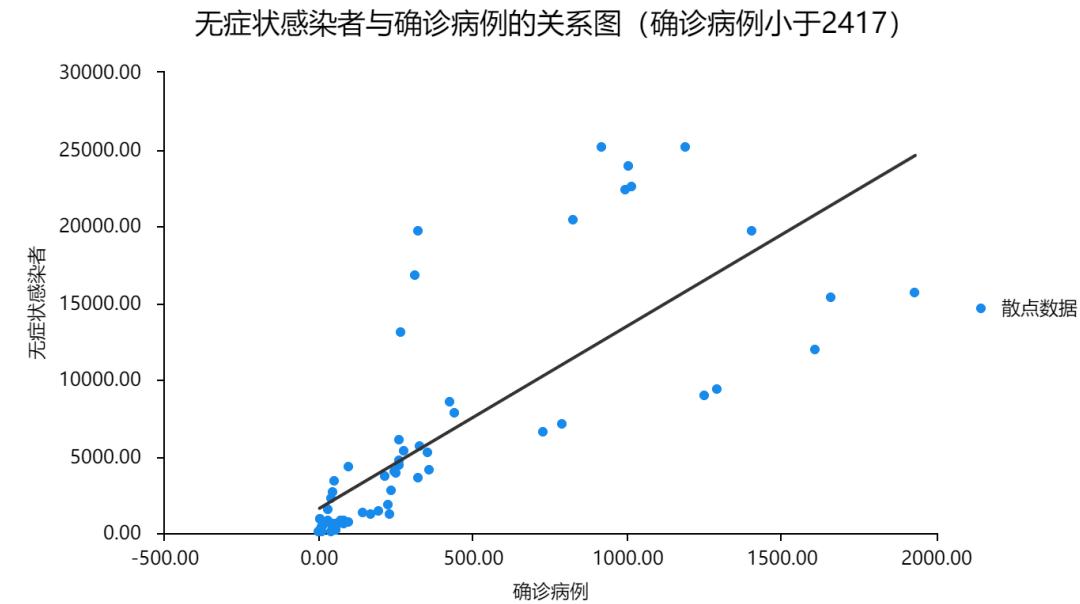
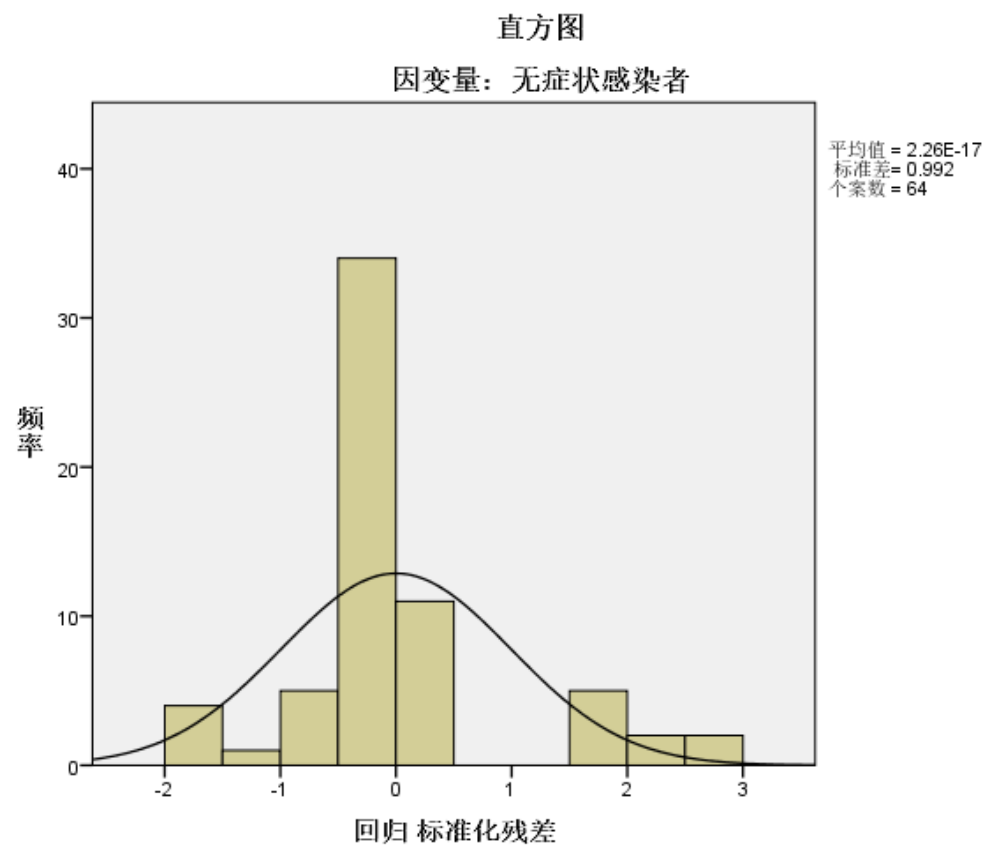


图 18：第 24-54 天无症状病例回归标准化残差图



散点数据线性拟合公式为：无症状感染者 = 1601.727 + 11.912*确诊病例， R 方值为 0.584。

表 5：无症状感染者与确诊病例的线性回归分析结果-简化格式

	回归系数	95% CI	VIF
常数	1601.727* (2.079)	92.040 ~ 3111.414	-
确诊病例	11.912** (9.321)	9.407 ~ 14.417	1.000
样本量		64	
R^2		0.584	

表 5：无症状感染者与确诊病例的线性回归分析结果-简化格式

回归系数	95% CI	VIF
调整 R^2	0.577	
F 值	$F(1,62)=86.873, p=0.000$	

因变量：无症状感染者

D-W 值：0.459

* $p<0.05$ ** $p<0.01$ 括号里面为 t 值

从上表可知，将确诊病例作为自变量，而将无症状感染者作为因变量进行线性回归分析，从上表可以看出，模型公式为：无症状感染者=1601.727 + 11.912*确诊病例，模型 R 方值为 0.584，意味着确诊病例可以解释无症状感染者的 58.4%变化原因。对模型进行 F 检验时发现模型通过 F 检验($F=86.873, p=0.000<0.05$)，也即说明确诊病例一定会对无症状感染者产生影响关系，最终具体分析可知：

确诊病例的回归系数值为 11.912($t=9.321, p=0.000<0.01$)，意味着确诊病例会对无症状感染者产生显著的正向影响关系。

总结分析可知：确诊病例全部均会对无症状感染者产生显著的正向影响关系。

从上表可知，将确诊病例作为自变量，而将无症状感染者作为因变量进行线性回归分析，从上表可以看出，模型 R 方值为 0.584，意味着确诊病例可以解释无症状感染者的 58.4%变化原因。

6 问题六的求解与分析

6.1 时间与下列三点（确诊病例，无症状感染者，确诊+无症状人数）的相关性

时间与确诊病例相关

表 6：全时间段与确诊病例 Pearson 相关-详细格式

		确诊病例
日期	相关系数	-0.082

表 6：全时间段与确诊病例 Pearson 相关-详细格式

确诊病例	
p 值	0.487

* $p<0.05$ ** $p<0.01$

从上表可知，利用相关分析去研究确诊病例分别和日期共 1 项之间的相关关系，使用 Pearson 相关系数去表示相关关系的强弱情况。具体分析可知：
确诊病例和日期之间的相关系数值为-0.082，接近于 0，并且 p 值为 0.487>0.05，因而说明确诊病例和日期之间并没有相关关系。

时间与无症状感染者相关

表 7：全时间段与无症状感染者的 Pearson 相关-详细格式

无症状感染者		
日期	相关系数	-0.072
	p 值	0.538

* $p<0.05$ ** $p<0.01$

从上表可知，利用相关分析去研究无症状感染者分别和日期共 1 项之间的相关关系，使用 Pearson 相关系数去表示相关关系的强弱情况。具体分析可知：
无症状感染者和日期之间的相关系数值为-0.072，接近于 0，并且 p 值为 0.538>0.05，因而说明无症状感染者和日期之间并没有相关关系。

时间与确诊病例+无症状感染者相关

表 8：全时间段与确诊病例+无症状感染者的 Pearson 相关-详细格式

确诊病例+无症状感染者	
相关系数	-0.076
日期	
<i>p</i> 值	0.518

* $p<0.05$ ** $p<0.01$

从上表可知，利用相关分析去研究确诊病例+无症状感染者分别和日期共 1 项之间的相关关系，使用 Pearson 相关系数去表示相关关系的强弱情况。具体分析可知：确诊病例+无症状感染者和日期之间的相关系数值为-0.076，接近于 0，并且 p 值为 $0.518>0.05$ ，因而说明确诊病例+无症状感染者和日期之间并没有相关关系。

6.2 整个疫情的走势：

确诊病例在第 1 到第 26 天阶段呈现缓慢增长趋势，（从日 1 例到达日 322 例），接下来在第 27-第 35 天阶段急剧上升（从日 824 例到达日 3590 例），在之后的第 36 天到第 52 天的阶段开始急速下降（从日 3238 例到达日 274 例），除了第 48 天异常上升达到顶峰数值 5487 例，在最后的第 53 天到第 75 天这个阶段缓慢下降（从日 260 例到达日 48 例）。

无症状感染者在第 1 到第 12 天阶段呈现缓慢增长趋势，（从日 64 例到达日 977 例），接下来在第 13-第 30 天阶段急剧上升达到顶峰（从日 979 例到达日 25173 例，），在之后的第 31 天到第 60 天的阶段开始急速下降（从日 22348 例到达日 1259 例），在最后的第 61 天到第 75 天这个阶段缓慢下降（从日 1305 例到达日 290 例）。

6.3 上海应对措施的有效性

由疫情稳定与结束阶段，在问题三、四、五的分析中可知，此阶段拟合效果优胜说明在上海市政府的措施实行下，疫情在稳定消散，感染者人数日渐下降，说明上海应对疫情采取的措施起到了很好的效果，线性拟合的优胜代表了疫情可预测性的下降，按照一定的速度逐渐消失，体现了上海措施的有效。

6.4 上海疫情可能的变化预测

根据疫情稳定与发展阶段线性回归分析可知，确诊人数和无症状感染者人数随时间增加而减少，线性回归模型拟合优胜。所及由回归分析可预测，在接下来时间，上海现存的确诊及无症状感染者将完全治愈，感染者人数会逐步清零，疫情已经得到有效控制，感染者人数

大幅下降，未来一月上海将完全恢复正常，摆脱疫情的影响，人民学习生活逐渐走向正轨。

7 问题七的求解与分析

7.1 讨论线性分析工具对上海疫情变化的分析

线性回归工具将上海疫情的发展阶段分析的较为透彻，一系列疫情阶段性的特征直接反应到线性回归模型上，我们可以通过模型的正确性与相关系数的显著性检验，证明疫情数据之间的相关性。使用线性分析工具对上海疫情的各个阶段进行描述、分析；使得上海疫情各阶段的趋势十分清晰地展现在了我们眼前，让我们看到了上海市的经历，上海政府抗疫的效率，大家齐心协力跟随政府共克时艰，打赢了这场“战争”。这一个个模型之下就是上海的历程，经历疫情的无声扩散、毫无源头到疫情全面爆发再到有效措施的实施最后疫情结束，线性分析结果显著。

7.2 更优的模型

对于模型的选择而言，由于该变化情况可能有时用线性回归模型来拟合并不一定准确，所以可以考虑深度学习中的神经网络模型来进行后续的预测。

9、参考文献：

[1] 贾俊平，何晓群. 统计学基础（第 8 版）. 中国人民大学出版社有限公司, 2021

10、附录部分

附录一（源代码和数据）

源代码：

#加载模型库

```
import pandas as pd
```

```
import numpy as np
```

```
import scipy
```

```
import statsmodels.api as sm
```

```
import matplotlib.pyplot as plt
```

解决中文显示问题

```
plt.rcParams['font.sans-serif'] = ['SimHei'] # 指定默认字体
```

```
plt.rcParams['axes.unicode_minus'] = False # 解决保存图像是负号'-'显示为方块的问题
```

#读取数据

```
def get_data(path):
```

```
    return pd.read_excel(path)
```

#数据分段

```
def split(data, start, end):
```

```
    return data[start:end]
```

#线性回归，返回预测值和残差

```

def lr(x, y):
    x1 = sp.add_constant(x)
    model = sp.OLS(y, x1)
    results=model.fit()
    print(results.summary())
    y_pred=model.predict(results.params, x1)
    outliers = results.get_influence()
    resids1 = outliers.resid_studentized_external
    plt.figure(figsize=(8,8))
    print('残差: ')
    print(resids1)
    plt.subplot(2,1,2)
    plt.hist(resids1)
    plt.subplot(2,1,1)
    plt.plot(x, y, '.')
    plt.plot([min(x), max(x)], [min(x)*(results.params[1])+results.params[0],
                                max(x)*(results.params[1])+results.params[0]])

    # return y_pred, resids1

```

```

path = '../新冠/xinguan.xlsx'
data = get_data(path)

```

```

### 日期、确诊、无症状、确诊+无症状、日期编号
data_date = data['date'].values
data_quezhen = data['quezhen'].values
data_wu = data['wu'].values
data_quewu = data['quewu'].values
data_time_code = data['time_code'].values

```

各个散点图

#各种需要的散点图

```

plt.figure(figsize=(10,12))
plt.subplot(3,1,1)
plt.plot(data_date, data_quezhen, '.')
plt.title('确诊病例')
#plt.xticks(data_time_code, data_date,)
plt.subplot(3,1,2)
plt.plot(data_date, data_wu, '.')
plt.title('无症状感染者')
plt.subplot(3,1,3)
plt.plot(data_date, data_quewu, '.')
plt.title('确诊病例和无症状感染者')

```

确诊和无症状的关系

```
#无症状感染者和确诊病例的关系，x 是确诊，y 是无症状
lr(data_quezhen, data_wu)
```

确诊区间

```
#确诊病例分时间区间进行验证
```

```
qz_1_18 = split(data_quezhen, 0, 18)
time_1_18 = split(data_time_code, 0, 18)
```

```
qz_19_28 = split(data_quezhen, 18, 28)
time_19_28 = split(data_time_code, 18, 28)
```

```
qz_29_47 = split(data_quezhen, 28, 47)
time_29_47 = split(data_time_code, 28, 47)
```

```
qz_48_75 = split(data_quezhen, 47, 75)
time_48_75 = split(data_time_code, 47, 75)
```

```
lr(time_19_28, qz_19_28)
lr(time_29_47, qz_29_47)
lr(time_48_75, qz_48_75)
```

无症状区间

```
wu_1_10 = split(data_wu, 0, 10)
time_1_10 = split(data_time_code, 0, 10)
```

```
wu_11_23 = split(data_wu, 10, 23)
time_11_23 = split(data_time_code, 10, 23)
```

```
wu_24_54 = split(data_quezhen, 23, 54)
time_24_54 = split(data_time_code, 23, 54)
```

```
wu_55_75 = split(data_quezhen, 54, 75)
time_55_75 = split(data_time_code, 54, 75)
lr(time_1_10, wu_1_10)
lr(time_11_23, wu_11_23)
lr(time_24_54, wu_24_54)
lr(time_55_75, wu_55_75)
```

附录二：确诊病例各时期统计表

第 19-28 天

图 19：第 19-28 天确诊病例随时间病例回归图

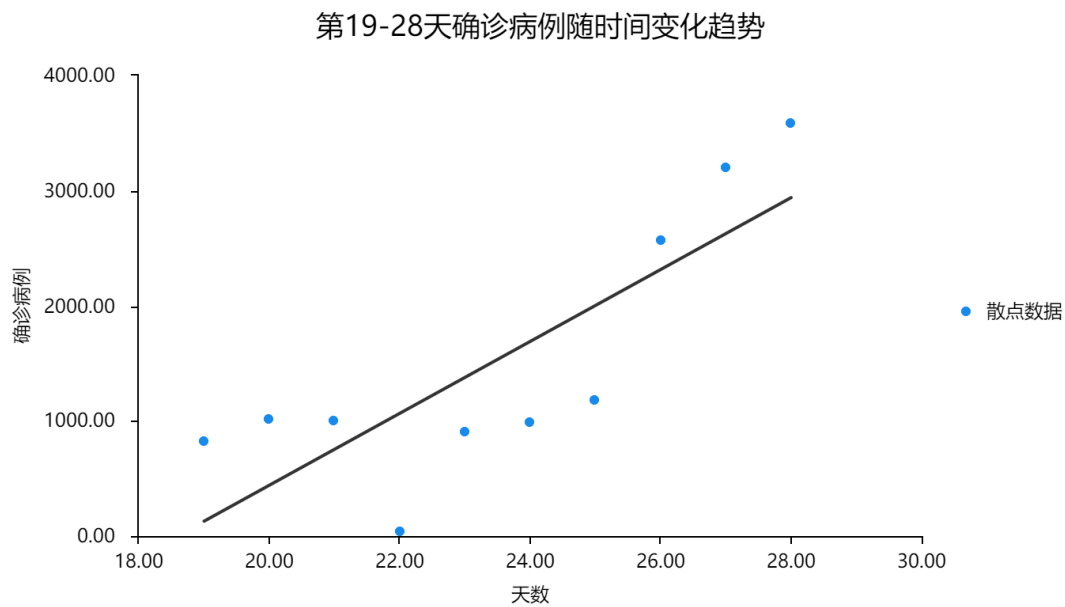
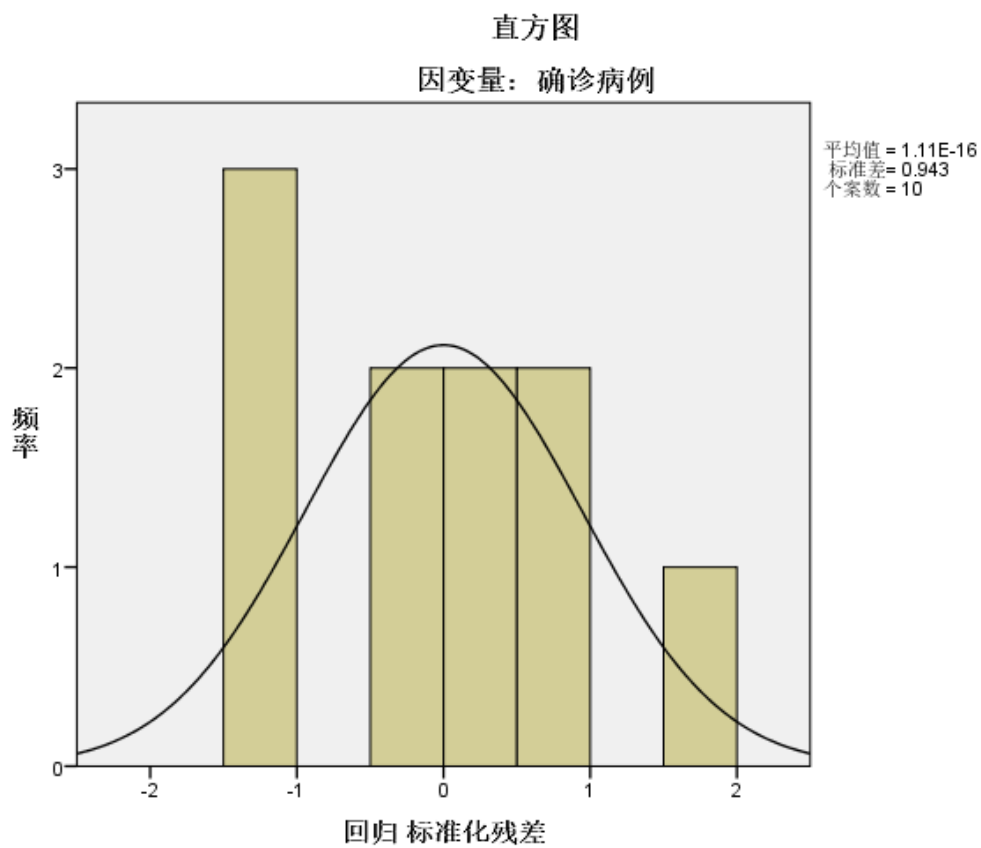


图 20：第 19-28 天确诊病例回归标准化残差图



散点数据线性拟合公式为：确诊病例 = -5807.085 + 312.412*天数， R 方值为 0.662。

表 9：第 19-28 线性回归分析结果-简化格式

	回归系数	95% CI	VIF
常数	-5807.085* (-3.110)	-9467.207 ~ -2146.963	-
天数	312.412** (3.961)	157.813 ~ 467.011	1.000
样本量		10	
R^2		0.662	
调整 R^2		0.620	
F 值		$F(1,8)=15.687, p=0.004$	

因变量：确诊病例

D-W 值：1.196

* $p<0.05$ ** $p<0.01$ 括号里面为 t 值

第 48-75 天

图 21：第 48-75 天确诊病例随时间病例回归图

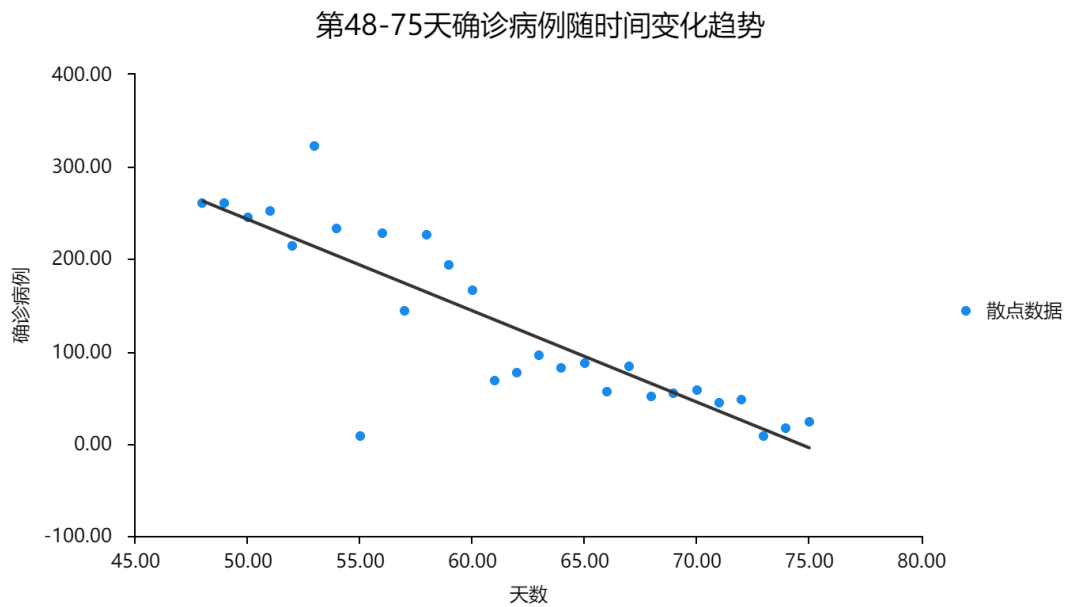
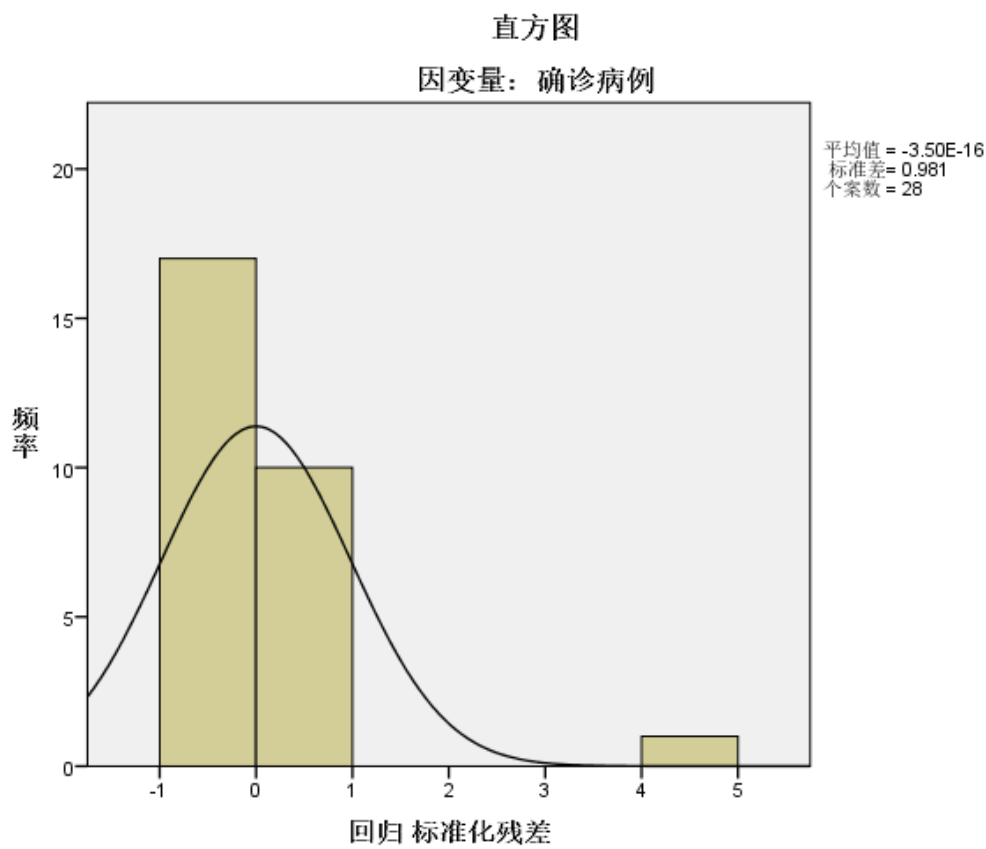


图 22：第 48-75 天确诊病例标准化残差图



散点数据线性拟合公式为：确诊病例 = $738.084 - 9.901 \times \text{天数}$ ， R 方值为 0.728。

表 10：第 48-75 天确诊病例随时间线性回归分析结果-简化格式

	回归系数	95% CI	VIF
常数	738.084** (10.037)	593.950 ~ 882.217	-
天数	-9.901** (-8.352)	-12.225 ~ -7.578	1.000
样本量		28	
R^2		0.728	
调整 R^2		0.718	
F 值		$F(1,26)=69.750, p=0.000$	

因变量：确诊病例

D-W 值：1.100

* $p<0.05$ ** $p<0.01$ 括号里面为 t 值

附录 3:

第 11-23 天

图 23：第 11-23 天无症状病例随时间病例回归图

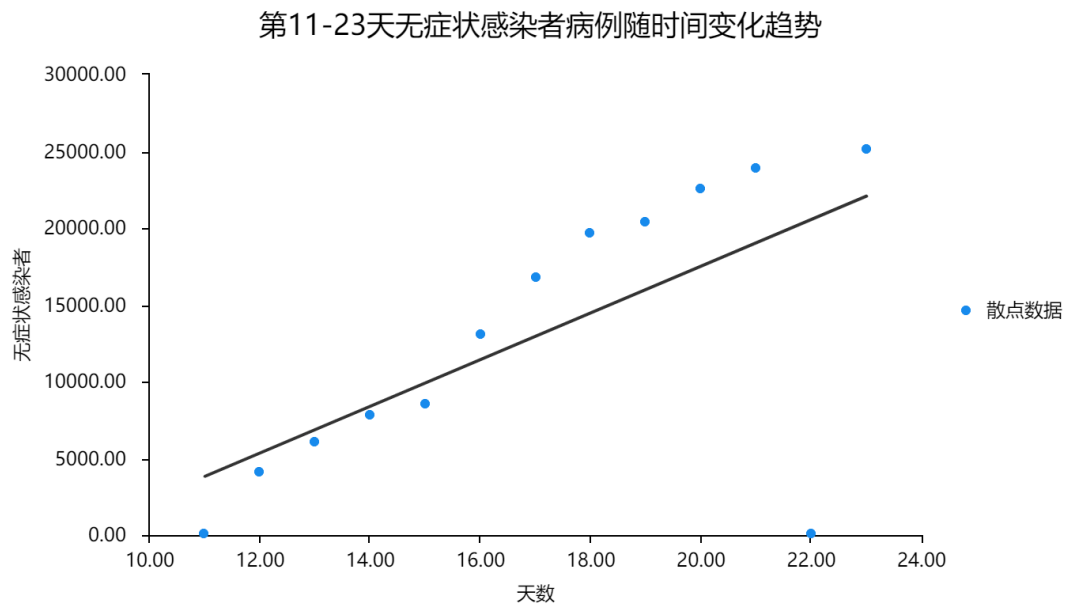
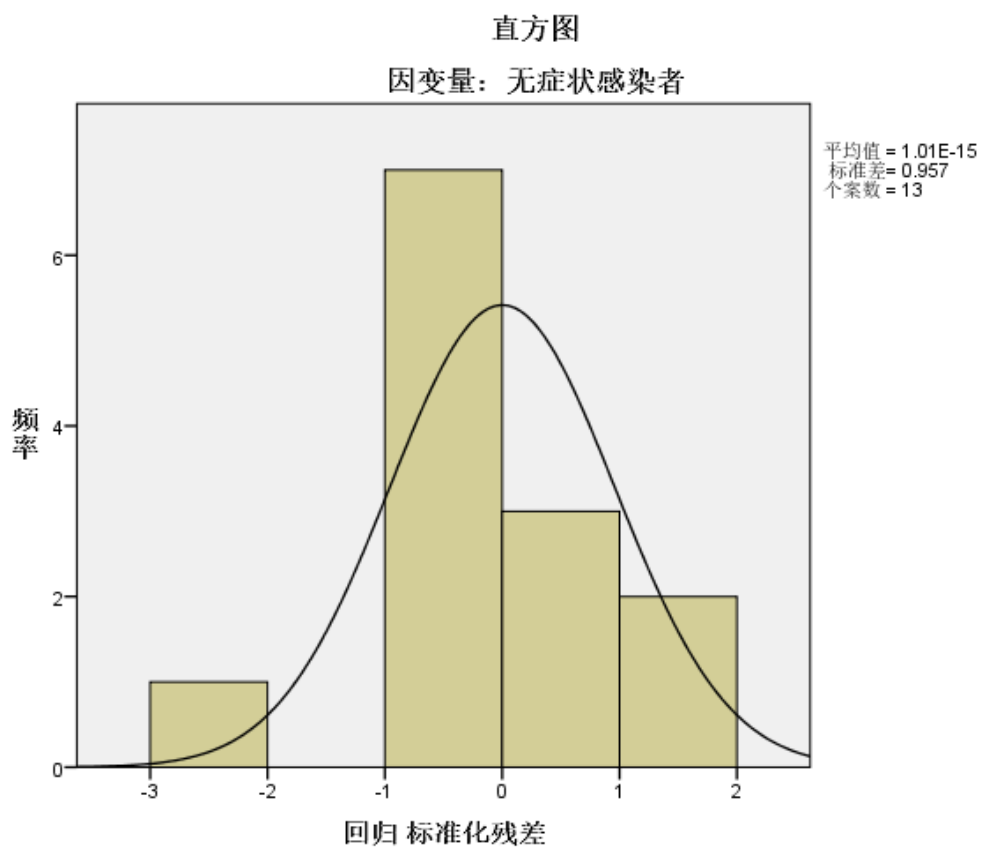


图 24：第 11-23 天无症状病例标准化残差图



散点数据线性拟合公式为 无症状感染者 = $-12901.225 + 1520.819 \times \text{天数}$, R 方值为 0.430。

表 11：第 11-23 天无症状病例与时间的线性回归分析结果-简化格式

	回归系数	95% CI	VIF
常数	-12901.225 (-1.403)	-30926.988 ~ 5124.538	-
天数	1520.819* (2.878)	485.266 ~ 2556.372	1.000
样本量		13	
R^2		0.430	
调整 R^2		0.378	
F 值		$F(1,11)=8.285, p=0.015$	

因变量：无症状感染者

D-W 值：1.195

* $p<0.05$ ** $p<0.01$ 括号里面为 t 值

第 55-75 天

图 25：第 55-75 天无症状病例标准化残差图

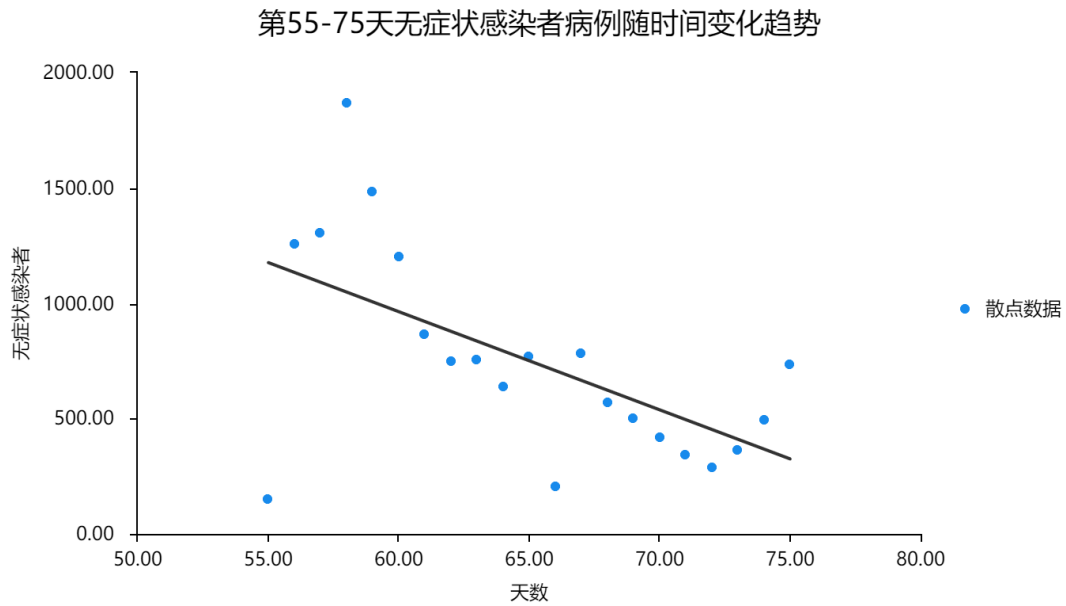
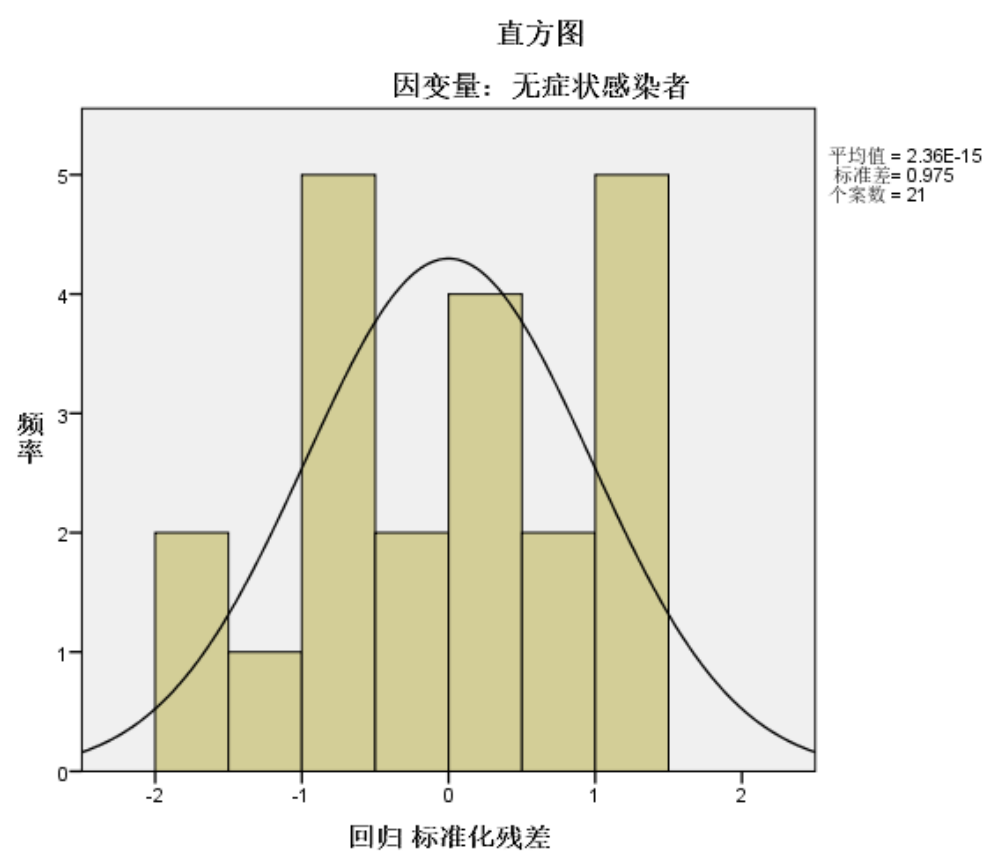


图 26：第 55-75 天无症状病例标准化残差图



散点数据线性拟合公式为：无症状感染者 = $3517.076 - 42.562 \times \text{天数}$ ， R 方值为 0.345。

表 12：第 11-23 天无症状病例与时间的线性回归分析结果-简化格式

	回归系数	95% CI	VIF
常数	3517.076** (4.001)	1794.062 ~ 5240.089	-
天数	-42.562** (-3.161)	-68.956 ~ -16.169	1.000
样本量		21	
R^2		0.345	
调整 R^2		0.310	
F 值		$F(1,19)=9.990, p=0.005$	

因变量：无症状感染者

D-W 值：0.527

* $p<0.05$ ** $p<0.01$ 括号里面为 t 值

附录 4：

图 27：无症状病例与确诊病例（确诊病例数>2416）的回归关系图

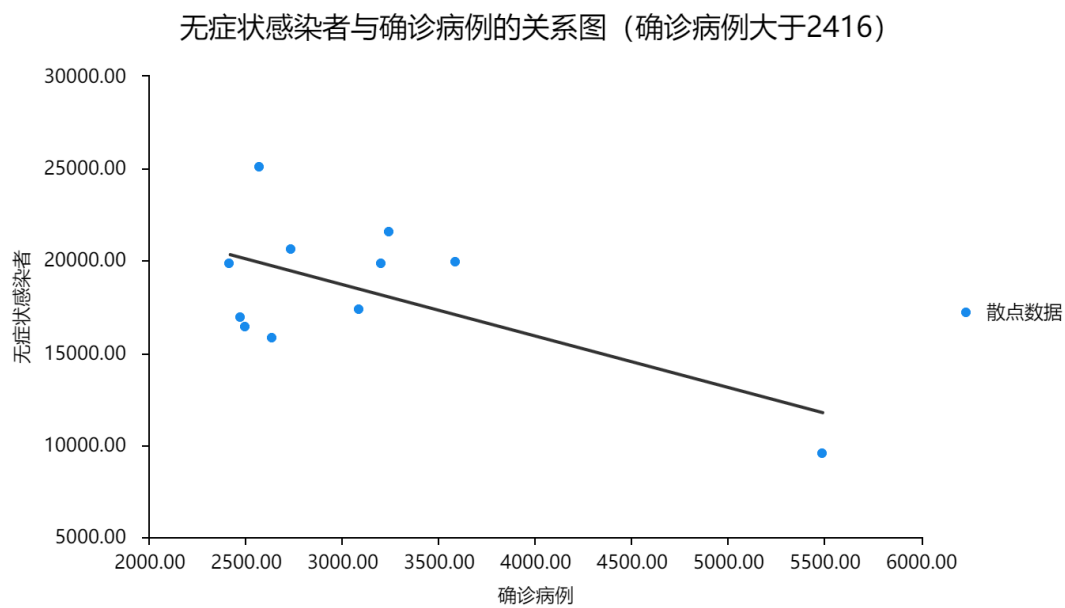
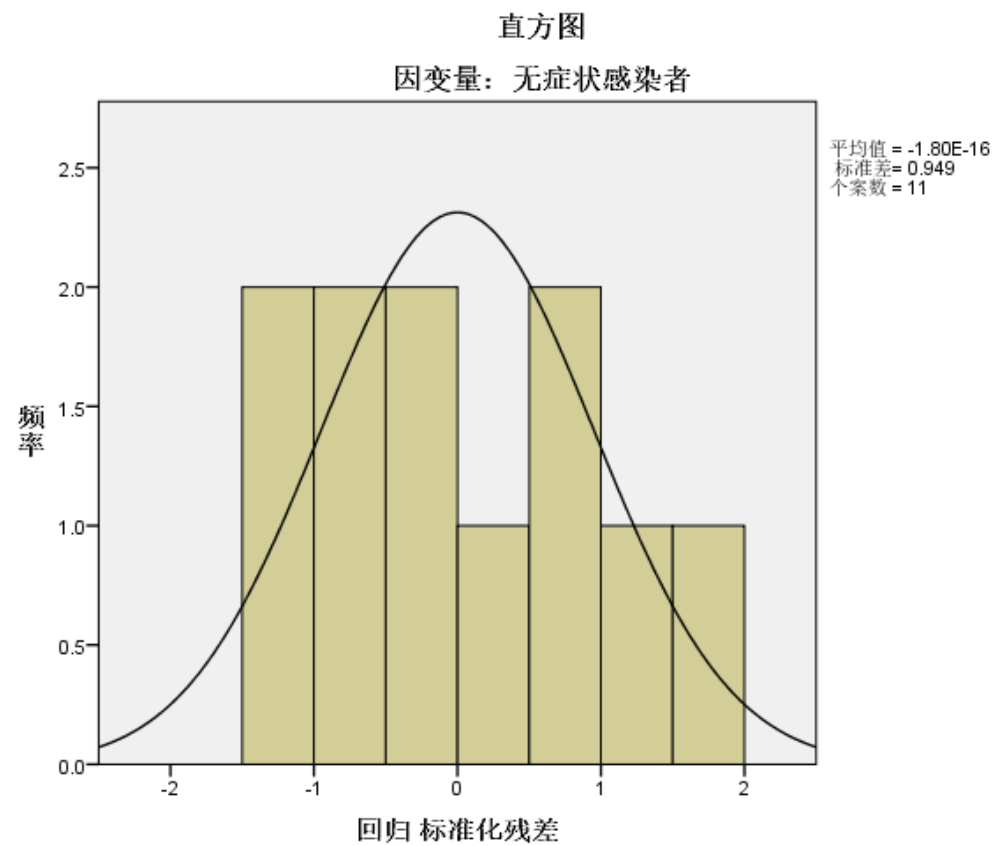


图 27：无症状病例与确诊病例（确诊病例数>2416）的无症状感染者标准化残差图



散点数据线性拟合公式为：无症状感染者 = 27078.875-2.793*确诊病例， R 方值为

0.383。

表 13：无症状病例与确诊病例（确诊病例数>2416）的线性回归分析结果-简化格式

	回归系数	95% CI	VIF
常数	27078.875** (7.174)	19680.830 ~ 34476.920	-
确诊病例	-2.793* (-2.366)	-5.107 ~ -0.479	1.000
样本量		11	
R^2		0.383	
调整 R^2		0.315	
F 值		$F(1,9)=5.597, p=0.042$	

因变量：无症状感染者

D-W 值：0.850

* $p<0.05$ ** $p<0.01$ 括号里面为 t 值

无症状感染者与确诊人数线性回归分析结果-简化格式