

Applied Statistics

Thom Hurks	<i>0828691</i>
Puck Mulders	<i>0737709</i>
Marijn van Knippenberg	<i>0676548</i>
Rik Coenders	<i>0777004</i>

May 25, 2016

Contents

1	Introduction	1
2	Exploratory Data Analysis	2
2.1	Normality	2
2.1.1	QQ Plot and Shapiro-Wilk	2
2.1.2	Wald-Wolfowitz Runs Test	3
2.1.3	Rank Serial Correlation Test	3
3	Appendix	3

1 Introduction

Todo

2 Exploratory Data Analysis

2.1 Normality

2.1.1 QQ Plot and Shapiro-Wilk

First, let us determine if the outcomes in the dataset are normally distributed. The first step is to generate a normal QQ plot as an easy visual confirmation.

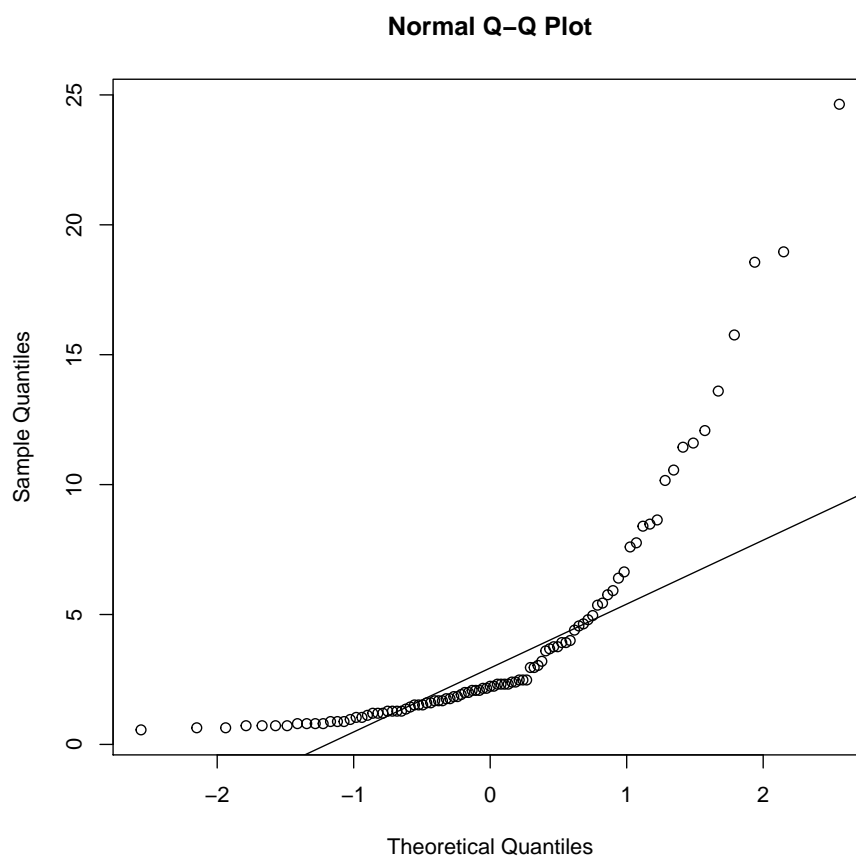


Figure 1: QQ Plot

The normal QQ plot already seems to suggest that the data is not normally distributed. When we also perform the Shapiro-Wilk normality test we can see the p-value = $1.1536883 \times 10^{-12} < 0.05$ so we can reject the null-hypothesis that the outcome is normally distributed.

2.1.2 Wald-Wolfowitz Runs Test

Next we perform the Wald-Wolfowitz runs test to test the null hypothesis that each element in the sequence is independently drawn from the same distribution. We see that the p-value = $7.1225335 \times 10^{-5} < 0.05$ so we can reject that null hypothesis. This is an interesting result, because we know the data is obtained from the same source, but part of the samples may or may not be tainted, and this result may point to that being true.

2.1.3 Rank Serial Correlation Test

Here we perform the rank serial autocorrelation test at lag 1 using the rank von Neumann ratio. This tests the null hypothesis that the lag- k autocorrelation is 0 for all values of k greater than 0 (i.e., the time series is purely random). The procedure that we use emits some warnings because the dataset contains ties, and with this specific test that can make the p-value less accurate. With that warning in mind, we present the table containing the rank serial autocorrelation test p-values of each batch:

	batch	p.value
1	B0	0.99
2	B1	0.13
3	B2	0.26
4	B3	0.36
5	B4	0.68
6	B5	0.21
7	B6	0.49
8	B7	0.37

As one can observe, the p-values are all large, which means that we cannot reject the null hypothesis that the batches are random using this test.

3 Appendix

The dataset used for assignment 1:

	BATCH	RUN	OUTCOME	UNIT	TIME
1	B0	1.00	0.72	5.00	-25.00
2	B0	1.00	0.72	7.00	-25.00
3	B0	1.00	6.64	9.00	-25.00
4	B0	1.00	0.64	10.00	-25.00
5	B0	1.00	0.80	18.00	-25.00

Continued on next page

	BATCH	RUN	OUTCOME	UNIT	TIME
6	B0	1.00	2.00	23.00	-25.00
7	B0	2.00	4.80	1.00	-25.00
8	B0	2.00	0.88	4.00	-25.00
9	B0	2.00	1.28	15.00	-25.00
10	B0	2.00	2.40	19.00	-25.00
11	B0	2.00	10.16	20.00	-25.00
12	B0	2.00	0.80	24.00	-25.00
13	B1	1.00	1.04	7.00	31.00
14	B1	1.00	3.20	12.00	31.00
15	B1	1.00	1.28	13.00	31.00
16	B1	1.00	1.60	16.00	31.00
17	B1	1.00	1.68	21.00	31.00
18	B1	1.00	1.44	24.00	31.00
19	B1	2.00	1.76	4.00	31.00
20	B1	2.00	1.68	9.00	31.00
21	B1	2.00	1.92	11.00	31.00
22	B1	2.00	5.92	17.00	31.00
23	B1	2.00	4.96	19.00	31.00
24	B1	2.00	3.92	20.00	31.00
25	B2	1.00	11.44	5.00	94.00
26	B2	1.00	1.52	7.00	94.00
27	B2	1.00	8.40	16.00	94.00
28	B2	1.00	2.40	19.00	94.00
29	B2	1.00	0.72	25.00	94.00
30	B2	1.00	0.88	27.00	94.00
31	B2	2.00	3.76	4.00	94.00
32	B2	2.00	2.32	8.00	94.00
33	B2	2.00	24.64	10.00	94.00
34	B2	2.00	18.96	12.00	94.00
35	B2	2.00	5.76	22.00	94.00
36	B2	2.00	15.76	26.00	94.00
37	B3	1.00	1.52	5.00	101.00
38	B3	1.00	13.60	10.00	101.00
39	B3	1.00	18.56	13.00	101.00
40	B3	1.00	4.40	20.00	101.00
41	B3	1.00	2.96	21.00	101.00
42	B3	1.00	3.04	26.00	101.00
43	B3	2.00	1.68	2.00	101.00
44	B3	2.00	8.48	6.00	101.00
45	B3	2.00	12.08	14.00	101.00
46	B3	2.00	10.56	18.00	101.00
47	B3	2.00	3.76	22.00	101.00
48	B3	2.00	5.36	24.00	101.00
49	B4	1.00	0.64	6.00	150.00

Continued on next page

	BATCH	RUN	OUTCOME	UNIT	TIME
50	B4	1.00	0.80	13.00	150.00
51	B4	1.00	0.88	15.00	150.00
52	B4	1.00	1.28	17.00	150.00
53	B4	1.00	0.72	21.00	150.00
54	B4	1.00	1.20	23.00	150.00
55	B4	2.00	1.20	2.00	150.00
56	B4	2.00	0.56	3.00	150.00
57	B4	2.00	1.04	8.00	150.00
58	B4	2.00	2.24	11.00	150.00
59	B4	2.00	0.80	14.00	150.00
60	B4	2.00	1.52	20.00	150.00
61	B5	1.00	2.16	6.00	157.00
62	B5	1.00	1.60	8.00	157.00
63	B5	1.00	2.32	10.00	157.00
64	B5	1.00	2.48	12.00	157.00
65	B5	1.00	3.68	21.00	157.00
66	B5	1.00	2.96	22.00	157.00
67	B5	2.00	2.24	1.00	157.00
68	B5	2.00	1.76	2.00	157.00
69	B5	2.00	2.32	4.00	157.00
70	B5	2.00	2.08	16.00	157.00
71	B5	2.00	7.60	23.00	157.00
72	B5	2.00	5.44	24.00	157.00
73	B6	1.00	8.64	6.00	234.00
74	B6	1.00	2.08	9.00	234.00
75	B6	1.00	0.96	17.00	234.00
76	B6	1.00	1.84	19.00	234.00
77	B6	1.00	1.20	20.00	234.00
78	B6	1.00	1.12	22.00	234.00
79	B6	2.00	2.48	3.00	234.00
80	B6	2.00	2.00	5.00	234.00
81	B6	2.00	2.48	12.00	234.00
82	B6	2.00	2.16	13.00	234.00
83	B6	2.00	1.36	14.00	234.00
84	B6	2.00	4.64	16.00	234.00
85	B7	1.00	1.28	1.00	241.00
86	B7	1.00	2.32	4.00	241.00
87	B7	1.00	11.60	10.00	241.00
88	B7	1.00	4.56	11.00	241.00
89	B7	1.00	7.76	15.00	241.00
90	B7	1.00	3.92	23.00	241.00
91	B7	2.00		12.00	241.00
92	B7	2.00	6.40	14.00	241.00
93	B7	2.00	3.60	17.00	241.00

Continued on next page

	BATCH	RUN	OUTCOME	UNIT	TIME
94	B7	2.00	1.84	18.00	241.00
95	B7	2.00	4.00	19.00	241.00
96	B7	2.00	2.08	21.00	241.00