UNIVERSITY OF Southampton

Comp6235 – Coursework for the statistics part

Junming Zhang   29299527   jz1g17@ecs.soton.ac.uk

1. Method

MeanTime	SDtime	LowerBand	UpperBand	TimeCI	MeanWeig	SDweight	LowerBand	UppoerBai	WeightCI
0.397778	0.341166	0.174882	0.620673	0.445791	2.025556	1.35737	1.138741	2.91237	1.77363
1.476923	0.260876	1.335109	1.618737	0.283628	2.102308	1.446779	1.31583	2.888786	1.572956
2.52	0.334365	2.312758	2.727242	0.414483	1.998	1.187825	1.261778	2.734222	1.472443
3.617143	0.202049	3.467463	3.766823	0.29936	1.937143	1.072749	1.142439	2.731846	1.589407
4.415	0.35248	4.23036	4.59964	0.369281	1.852143	1.126487	1.262053	2.442233	1.180179
5.523333	0.268949	5.371161	5.675505	0.304344	2.370833	1.066562	1.767369	2.974297	1.206928
6.436364	0.360646	6.223236	6.649492	0.426256	1.890909	1.175563	1.196196	2.585623	1.389427
7.428333	0.239235	7.292973	7.563693	0.27072	1.718333	1.150137	1.067582	2.369085	1.301503
8.478462	0.284659	8.323719	8.633204	0.309485	2.295385	0.756038	1.884398	2.706372	0.821974
9.675714	0.250989	9.489779	9.861649	0.37187	1.595714	0.976573	0.872259	2.31917	1.446911
10.35933	0.278066	10.21861	10.50005	0.281442	1.587333	1.238712	0.960459	2.214208	1.253749
11.50125	0.305026	11.28988	11.71262	0.422745	1.31625	1.183046	0.496441	2.136059	1.639618
12.55313	0.362321	12.37559	12.73066	0.355074	1.79125	0.899903	1.350298	2.232202	0.881905
13.38714	0.357851	13.12204	13.65224	0.530199	1.665714	0.866234	1.023999	2.30743	1.283431
14.51	0.280642	14.28544	14.73456	0.449121	1.241667	1.014247	0.4301	2.053233	1.623133
15.367	0.24281	15.2165	15.5175	0.30099	2.092	1.072617	1.427185	2.756815	1.329629
16.512	0.319484	16.23196	16.79204	0.56008	1.668	1.077831	0.723239	2.612761	1.889521
17.38778	0.287697	17.19982	17.57574	0.375924	1.8	0.97112	1.165535	2.434465	1.268931
18.48667	0.304281	18.24319	18.73014	0.48695	1.068333	0.672649	0.530102	1.606565	1.076463
19.622	0.308334	19.35173	19.89227	0.540534	1.738	0.835536	1.005621	2.470379	1.464759
20.86	NA	NA	NA	NA	1.57	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
21.2	0.27074	20.89363	21.50637	0.612742	2.373333	0.954428	1.293296	3.453371	2.160075
23.16	NA	NA	NA	0	1.34	NA	NA	NA	NA

Table 1. Processed data from fish.txt

1. Sort data based on ascending order of time and plot the distribution and linear fitting diagram.
2. Separate data to 24 independent group hour by hour.
3. Calculate the mean, standard deviation and confidence interval of separated data.
4. Plot the line chart between two variables based on processed data.
5. Plotting based on the ggplot2 library.

2. Result and Analysis

2.1 Character and generate graphs to illustrate the distribution of time and weight

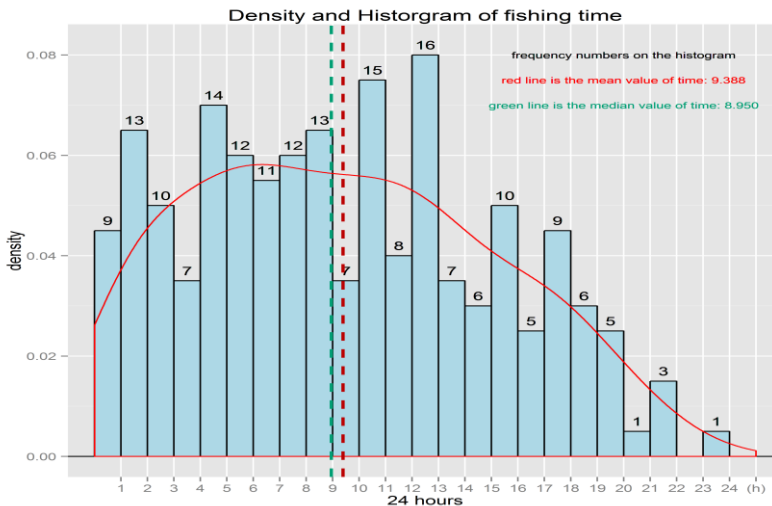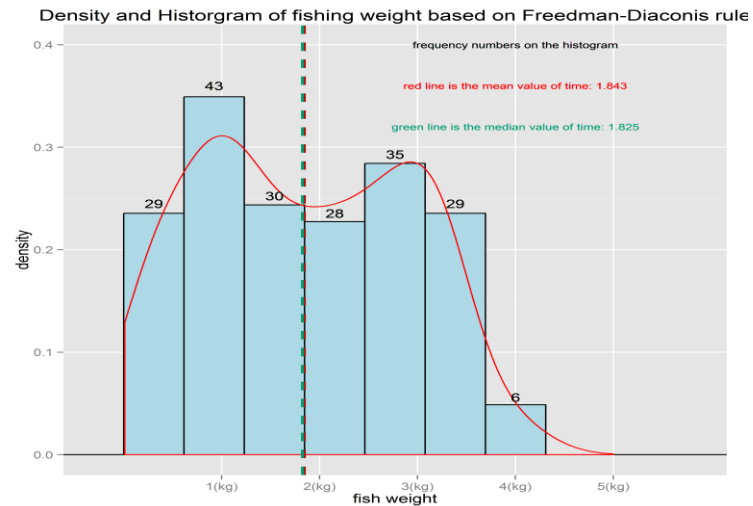*Figure 1. Distribution of catching time*

Figure 2. Distribution of fish weight

For catching time(unit-h):

Mean	Median	Mode	Variance	Geometric mean	Standard deviation
9.388	8.950	12-13	31.998	6.786	5.657
Skewness	Kurtosis	IQR	Bin size	Kernel density	95% confidence interval of the mean
0.250	2.134	8.335	1	Gaussian	8.600-10.177

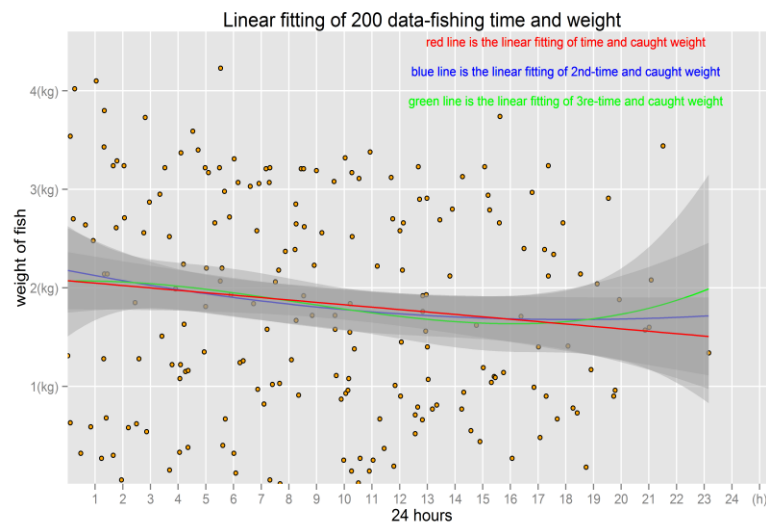The bin size of catching time is 1 because the time value is within 24 hours.

For catching weight (unit-kg):

| Mean	Median	Mode	Variance	Geometric mean	Standard deviation
1.843	1.825	0.616-1.232	1.162	1.369	1.078
Skewness	Kurtosis	IQR	Bin size	Kernel density	95% confidence interval of the mean
0.090	1.849	1.800	0.616	Gaussian	1.693-1.993

The bin size of catching weight is based on Freedman-Diaconis rule.

2.2 Analyze and generate graphs to illustrate the dependence between time and weight.

Figure 3. The linear fitting of fishing data   *Figure 4. The line chart of sampled fishing data*
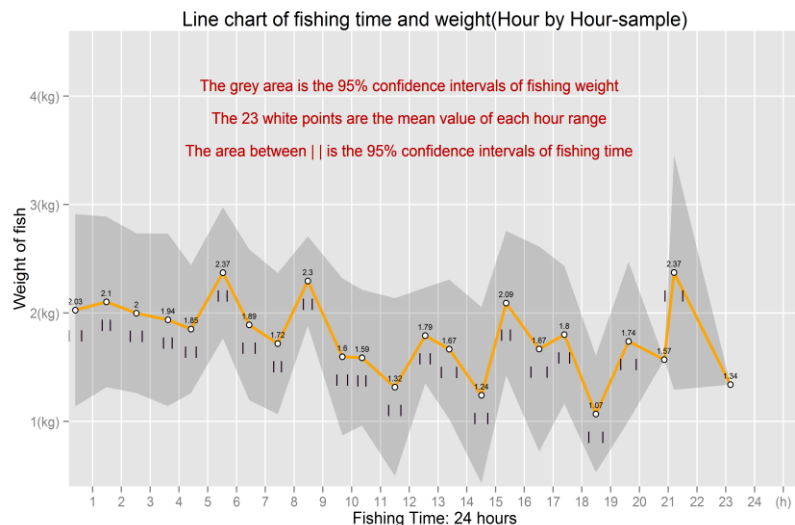
Pearson metric	P value	Kendall metric	P value	Spearman metric	P value
-0.128	0.070	-0.088	0.063	-0.126	0.073

The p-value is not small than 0.05 which means the correction test may not correct. However, in most situation, it could confirm the fishing time and caught weight are independent due to the value of each correction test is close to 0.

Then apply linear fitting to the original data:

lm1	lm2	lm3	covariance
599.329	600.893	602.529	-0.782

lm1 has the smallest value which means the red line of Figure 3 is the best regression curve of them. It could find the caught weight decreases with the increasing fishing time.

2.3 Estimate the best time to go fishing.

The regression curve of Figure 3 indicates the caught weight decrease with the time. At the meanwhile, for obtaining accurate time to go fishing, Figure 4 gives the average fish caught hour by hour, it could find time between 5:00 and 6:00 has higher average number 2.37 and highest lower band of fish weight., Therefore the estimated best time to go fishing is 5:00~6:00.

3. R code, Diagrams and Processed data

https://github.com/TROUBLEBRO/comp6235_1718_coursework-1