How does allocation of funding to London local authorities affect obesity cases in 2018

CASA 0005 Quantitative Methods Coursework 1

November 15, 2020

1. Introduction

Today, almost one in five Year 6 children in the UK is found to be obese, and sadly this number is not dropping during the past few years [1]. The origins of childhood obesity stem from various aspects, including lifestyle issues, genetic and environmental causes. The government has been taking considerable forms of actions to tackle this problem from its root, from sugar reduction to advertising and promotions. This study investigates how government has been allocating their funding to local authorities in London.

2. Data

The data used in this study contains population, obesity cases, total budget, and allocation of funding for local authorities across London in 2018. An illustration of data employed is shown in Table (1).

	local authorities	total obesities	total population	...	total budget	...
0	Barking and Dagenham	763	181779	...	139000	...
1	Barnet	773	355955	...	220000	...
			⋮			

Table 1: Illustration of data used, list of column names include:
Names of local authority areas; total obesity cases in each area; total population; obesity density (obesity cases divided by population); total budget allocated (in pounds); percent of budget spent on improving air quality, cleaner environment, health training, raising school awareness, media awareness and subsiding counselling.

Only one observation is considered an outlier due to its relatively small scale, and therefore dropped from the dataset, City of London data. Its population is below the average population in London boroughs by 97%, having this data in the linear regression plot of obesity cases in 2018 vs. total budget spent lowers the regression coefficient from 0.437 to 0.349.

3. Methodlogy

Three approaches were taken to investigate the criteria of funding:

i. A linear line was fitted using scipy.stats.linregress() between total budget spent and obesity density.

ii. A more outlier-robust linear approach, called Random Sample Consensus (RANSAC) was used. This method compliments the ordinary least squares methods by adding detections of outliers and accord them to have no influence on the parameters of the model (Fischler and Bolles, 1981). In `sklearn.linear_model.RANSACRegressor()`, outliers are classified as those whose residual exceed the median absolute deviation of dependent variables (Pedregosa, F. et al., 2011).

iii. Finally, a non-linear approach was taken. Polynomial regression of various degrees was fitted to the data using `numpy.polyfit()` in Python.

4. Results

i. Logically one might assume that the more obesity cases discovered per unit population, the more funding would be allocated. The linear line fitted to total budget spent vs. obesity density yields a coefficient of determination of 0.002, which means that only 0.2% of the variance of dependent variable (total budget) is explained by independent variable (obesity density). Log transformations were also applied to the variables, associated parameters are shown in Table (2).

	unchanged data	$log(x)$ and y	$log(y)$ and x	$log(x)$ and $log(y)$
slope	3×10^{-9}	0.001	3×10^{-7}	0.0574
constant	-0.002	-0.005	0.47	-0.259
R^2	0.037	0.032	0.29	0.29
Pearson correlation coefficient	0.192	0.18	0.538	0.537
p value	0.29	0.32	0.001	0.002
relationship implied	linear	exponential	exponential	power

Table 2: coefficients related to the relationship between original and log-transformed total budget spent and obesity density

As the maximum value of R^2 being under 0.3 in Table (2), the amount of funding allocated to local authorities is extremely weakly correlated with the obesity cases per unit population by the power law, thus indicating that this relationship is meaningless.

ii.

Figure 1: Best fit line for I and V

In Table (**??**), the expected I is calculated from the equation of the fitted line, which was yielded in python programme code "3.ii": $I = 6.777V + 1.998$.

iii.

Figure 2: Best fit line with uncertainties

The probability $P(22.31; 9) = 0.007953893$. As this probability is extremely close to 0, we reject null hypothesis since this implies the measurements are insignificant and unphysical. For a good fit we expect two-thirds of the data points to be within one standard error bar of the trend line. But in this graph we can only find 5 out of 11 points within one standard error bar. Therefore this is not a good fit.

This suggests that the hypothesis of linear relationship between v and i is rejected. This may be the case theoretically, but in the real scenario, linear relationship cannot be maintained since the resistance is not constant (could be affected by temperature).

5. Best Fit by Minimizing χ^2

i. Brewster angle is given by $\theta_B = \arctan(\frac{n_2}{n_1})$, where n_1 is the refractive index of the initial medium through which the light propagates (the "incident medium"), and n_2 is the index of the other medium. Therefore, with the knowledge of Brewster angle being 57 degree, and n_1 being 1 for air, we can calculate n_2, which is 1.539865, rounded to 1.540.

The expected value of $\mathbf{R}^2$ can be calculated by

$$\mathbf{R}_\perp^2 = \left(\frac{\cos\theta_1 - n\cos\theta_2}{\cos\theta_1 + n\cos\theta_2}\right)^2 \qquad \mathbf{R}_\parallel^2 = \left(\frac{n\cos\theta_1 - \cos\theta_2}{n\cos\theta_1 + \cos\theta_2}\right)^2 \tag{1}$$

To conclude, a table with all values needed to calculate χ^2 and $P(\chi^2)$ is written below.

θ_1	θ_1 in radian	$\mathbf{R}_\parallel^2$	$\mathbf{R}_\perp^2$	θ_2 in radian	expected $\mathbf{R}_\perp^2$	expected $\mathbf{R}_\parallel^2$	$\chi_\perp^2$	$\chi_\parallel^2$
10	0.174532925	0.03	0.037	0.113008831	0.047004879	0.043388466	25.02440082	44.81275438
15	0.261799388	0.028	0.039	0.168880675	0.049393107	0.041138201	27.00417064	43.1530824
20	0.34906585	0.027	0.043	0.223978484	0.052948477	0.037974149	24.74304991	30.10798748
25	0.436332313	0.025	0.047	0.278019262	0.057909564	0.033897733	29.75464401	19.79241518
30	0.523598776	0.021	0.055	0.330698539	0.064625378	0.028937627	23.16197549	15.75148119
35	0.610865238	0.018	0.063	0.381685163	0.073594848	0.023180871	28.06269852	6.710354821
40	0.698131701	0.012	0.075	0.430616559	0.085524952	0.01682733	27.69365481	5.825778231
45	0.785398163	0.007	0.09	0.477095003	0.101415223	0.010285047	32.57682719	2.69788409
50	0.872664626	0.003	0.114	0.520685692	0.1226792	0.004338043	18.83212679	0.447589908
55	0.959931089	0.001	0.146	0.560917768	0.151316707	0.000444236	7.066843977	0.077218314
60	1.047197551	0.003	0.188	0.597289673	0.190153592	0.001271133	1.159489483	0.747245501
65	1.134464014	0.012	0.245	0.629280412	0.243166177	0.011675807	0.840726951	0.026275222
70	1.221730476	0.039	0.326	0.65636795	0.315902862	0.04053963	25.48805113	0.592615396
75	1.308996939	0.104	0.445	0.67805514	0.416000321	0.104310777	210.2453493	0.024145608
80	1.396263402	0.219	0.604	0.693901872	0.553761346	0.234129461	630.9805948	57.22514753
85	1.483529864	0.467	0.822	0.703559953	0.742712787	0.490951947	1571.615549	143.4239389

Table 3: Sets of data in order to calculate χ^2

We omit the values for 50, 55 and 60 degrees, since these are less than 3 error bars ($3 \times 0.002 = 0.006$) from zero.

The final result is listed in the following table.

	perpendicular	parallel
χ^2	2657.19	370.14
$P(\chi^2)$	0	4.451×10^{-71}

Table 4: list of results for 4.i

ii. Excel solver is used to minimise χ^2, results are listed below.

	perpendicular	parallel
χ^2	4793.01	309.32
$P(\chi^2)$	0	2.70×10^{-58}

Table 5: Minimising $\chi^2_{\parallel}$ (n=1.481)

	perpendicular	parallel
χ^2	2251.85	503.80
$P(\chi^2)$	0	2.28×10^{-99}

Table 6: Minimising $\chi^2_{\perp}$ (n=1.582)

	perpendicular	parallel
χ^2	2267.91	469.65
$P(\chi^2)$	0	4.05×10^{-92}

Table 7: Minimising both χ^2 (n=1.573)

References

[1] NHS. National Child Measurement Programme, England 2018/19 School Year., 2018.