**Andrew Goldberg**

**Data 609 Homework 2**

**2/12/17**

**Page 69: #12**

Identify a problem worth studying and list the variables that affect the behavior you have identified. Which variables would be neglected completely? Which might be considered as constants initially? Can you identify any submodels you would want to study in detail? Identify any data you would want collected.

A company with a fleet of trucks faces increasing maintenance costs as the age and mileage of the trucks increase.

**Problem worth studying:** Predicting maintenance costs over the lifetime of a company truck

**Variables that affect behavior:** Age of truck, mileage of truck (I assume there’s more such as location/type of truck route, driver characteristics, etc)

**Neglectable variables:** I wouldn’t neglect either age of truck right away, although I assume mileage of truck is far more predictive than age.

**Constants:** presumably there would be constants moderating both age (k1) and mileage (k2) of truck

**Submodels:** My hunch is that age would be far less predictive than mileage, but it would be useful to understand the relationship between these variables

**Data collected:** Date of vehicle purchase/first usage, cost of each maintenance occurrence with date and vehicle mileage.

**Page 79, #11.**

Determine whether the data supports the stated proportionality model.

|  |  |  |  |
| --- | --- | --- | --- |
| y | x |  | K = y/ |
| 0 | 1 | 1 | 0.000 |
| 1 | 2 | 8 | 0.125 |
| 2 | 3 | 27 | 0.074 |
| 6 | 4 | 64 | 0.094 |
| 14 | 5 | 125 | 0.112 |
| 24 | 6 | 216 | 0.111 |
| 37 | 7 | 343 | 0.108 |
| 58 | 8 | 512 | 0.113 |
| 82 | 9 | 729 | 0.112 |
| 114 | 10 | 1000 | 0.114 |

Within the geometric interpretation of the data, the line does not go through the origin, so it is technically not proportional. However, the data hits quite close to the origin, and it does form a fairly straight line, so we may be strongly tempted to think of it as proportional as a simplifying assumption.

**Page 94: #4**

*Lumber Cutters—*Lumber cutters wish to use readily available measurements to estimate the number of board feet of lumber in a tree. Assume they measure the diameter of the tree in inches at waist height. Develop a model that predicts board feet as a function of diameter in inches.

1. **Consider two separate assumptions, allowing each to lead to a model. Completely analyze each model.**
2. **Assume that all trees are right-circular cylinders and are approximately the same height**

We’d have a model like board feet/10 (bf) = where h = height and d = diameter

We aren’t supplied with the height, so I assume h will meld in with k or disappear completely. I’ll keep it in for now.

Given that, we’d have:

If we took a point on the trend line at d^2 = 1000 and bf = 150:

**hk = 150/1000 = .15** leading to a model of:

Graph of the model:

1. **Assume that all trees are right-circular cylinders and that the height of the tree is proportional to the diameter.**

If height of tree is proportional to the diameter, we can assume

Leading to , which looks like:

Using points on the trendline to find k, we have k = 259/59319= .004

Giving us a model of , which looks like:

1. **Which model appears to be better? Why? Justify your conclusions.**

I would argue that the second model is better. Just by intuition, the height should be proportional to the diameter, not flat. Looking at the graphs, the data appears to be closer to an exponential d^3 fit, rather than the d^2 fit, as it starts and ends at the similar spot and appears to follow the curvature. Will wait for next lesson to do statistical analysis.

All of the data:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| x (diameter) | y (number of board feet/10) | d^2 | hk = bf/d^2 | bf = .15d^2 | d^3 | k = bf/d^3 |
| 17 | 19 | 289 | 0.065743945 | 43.35 | 4913 | 0.003867291 |
| 19 | 25 | 361 | 0.069252078 | 54.15 | 6859 | 0.003644846 |
| 20 | 32 | 400 | 0.08 | 60 | 8000 | 0.004 |
| 23 | 57 | 529 | 0.107750473 | 79.35 | 12167 | 0.004684803 |
| 25 | 71 | 625 | 0.1136 | 93.75 | 15625 | 0.004544 |
| 28 | 113 | 784 | 0.144132653 | 117.6 | 21952 | 0.005147595 |
| 32 | 123 | 1024 | 0.120117188 | 153.6 | 32768 | 0.003753662 |
| 38 | 252 | 1444 | 0.174515235 | 216.6 | 54872 | 0.004592506 |
| 39 | 259 | 1521 | 0.170282709 | 228.15 | 59319 | 0.004366223 |
| 41 | 294 | 1681 | 0.174895895 | 252.15 | 68921 | 0.004265754 |

**Page 99 #3**

Discuss several factors that were completely ignored in our analysis of the gasoline milage problem.

The explanation of the problem listed several other factors that they decided to ignore for simplification problems, such as weather, motor efficiency, etc. If I were brainstorming, I could also add factors like driver skill and efficiency. We could also add a larger factor such as the actual cost of driving inefficiency, and how that relates to the utility of driving at that speed. For instance, the benefits of getting a company’s products to market at a quicker pace may outstrip the extra costs of gas use inefficiency.

**Page 104 #2**

Tests exist to measure the percentage of body fat. Assume that such tests are accurate and that a great many carefully collected data are available. You may specify any other statistics, such as waist size and height, that you would like collected. Explain how the data could be arranged to check the assumptions underlying the submodels in this section. For example, suppose the data for males between ages 17 and 21 with constant body fat and height are examined. Explain how the assumption of constant density of the inner core could be checked.

Apologies for not being more creative myself, but if we were to use the example given at the end of this problem, we would be looking for variations in density within the equation

So if we had the weights of a large but defined population of say males between 17 and 21, and height was held constant, and let’s say you dumped each individual into a pool to measure the displacement of the water and got an accurate measure of their volume. You could then subtract out the outer layer using the aforementioned body fat test, and study the relationship between weight and volume, and see how weight density varies or doesn’t vary with changes to each other variable.