Using Template 2 – the Normalized Data and Model Spreadsheet

You probably noticed that the linear regression ("best-fit line") you just generated using the raw rent data would be useless in forecasting occupancy rates as a function of nightly rents. The standard deviation of the occupancy rates in the data is 16.3%, and the standard error of forecasts using our linear regression is 16.1% (making our usual assumptions, which you may remember from the Linear Regression discussion in the Mastering Data Analysis with Excel course, that the occupancy rates and residuals are drawn from Gaussian distributions). So using our model is only 0.2% more accurate – a trivial amount – than predicting, every time, no matter what the rent is, that the occupancy rate is the average of the data, or 45.6%.

So we need to modify the x-axis data to make its association with occupancy rates apparent and to allow us to create a model that is more predictive. This process, which is often necessary in data-analysis, is called "normalization."

Normalization can take many forms. In our case, we are going to consider each comparable property nightly rent in terms of how it compares to the rents of other properties of the same type in the same location. The data we have to help us do that are the 10^{th} percentile (low) and 90^{th} percentile (high) rents for each of the 244 combinations of comparable property type and location.

Rents differ in two main ways between property types and locations. First, the 50^{th} percentile may be higher or lower with the same dispersion – so an inexpensive location might have a dispersion from \$50 at the 10^{th} percentile to \$150 at the 90th, with a 50^{th} percentile at \$100, and an expensive location might have the same dispersion over \$150 to \$250, with a 50^{th} percentile rent of \$200. Second, the dispersion of rents may vary. So two locations may both have a 50^{th} percentile-rent of \$125 for a one-bedroom house, but one market may have prices from \$75 to \$175 while another location has rents from \$40 to \$210.

We also assume that rents are distributed uniformly between the 10^{th} and 90^{th} percentiles, and that we can build a linear model over the 10^{th} to 90^{th} percentile range. Note that all 244 sample comparable properties fortunately have rents that fall within the 10^{th} to 90^{th} range for that type and location. Later, when trying to find the optimum (revenue-maximizing) rent for a particular Watershed property type and location, we will also assume that no rent can be set lower than 10^{th} percentile or higher than 90^{th}.

A simple example should make this clearer. Let's say the 10^{th} percentile rent was \$20 and the 90^{th} percentile rent was \$180. Between the 10^{th} and 90^{th} percentile, our assumption that rents are uniformly distributed means that each percentile change corresponds to an increase in \$2 in the rent. So a rent of \$22 would be 11^{th} percentile. A rent of \$40 would be 20^{th} percentile, and so on. A rent of \$100 – half way between \$20 and \$180 – would be 50^{th} percentile.

Let me show you the arithmetic behind an example.

Under the reasonable simplifying assumption that properties are uniformly distributed by price between the 10^{th} and 90^{th} percentile,

Sample property Percentile rent
= .1 + .8*[((sample rent) – (10^{th} percentile [low] rent))
/ ((90^{th} percentile [high] rent) – (10^{th} percentile [low] rent))].

For example:
10^{th} percentile = $100
90^{th} percentile = $260
example property with occupancy data has nightly rent of $220

percentile for example property
= .1 + .8[(220 – 100) / (260-100)]
= .1 + .8[120/160]
= .1 + .8(.75)
= .7 = 70^{th} percentile.

Following the guidelines given in the *Best Practices for Setting up an Excel Spreadsheet* document, you will want to program this formula into each row of Template Spreadsheet 2, the *Normalized Data and Model* Spreadsheet, so that you can quickly calculate the percentile "example rent" for each of the 244 combinations of property type and location.

Once you have the correct percentiles for each, and understand how to normalize rents from dollars to percentiles using the 10^{th} and 90^{th} percentile values for a given property, proceed to the *Normalization* quiz to test your understanding.

Note that to help you better understand normalization, we also provide a Guide Spreadsheet, Dollars to Percentile Conversion, which contains the above formulas, and three related explanatory videos:
Normalizing Rents to Improve Occupancy Forecasting
Using the Dollars to Percentile Conversion Guide Spreadsheet,
Optimizing rents to Maximize Revenues.