Statistics 541 Fall 2011

Real Estate Project

Due Monday, November 14, 3pm

Sample Solution

To: Vice President
From: Statistical Consulting
Subject: **Analysis of Rental Contract Data, Executive Summary**

We analyzed 225 commercial rentals that were on the market in the last two years. We considered scenarios for a need of 50,000 usable sqft with modern wiring and in generally good repair, with a view towards a 3 year lease. The analysis shows that the main factors affecting rental costs are 1) location (city center, old and new suburb), 2) years since renovation, 3) wiring for modern computers and telecommunications, 4) occupancy rate of the building, 5) distance from the airport, and 6) length of the lease. (These factors affect variable cost, i.e., cost per sqft; fixed rental costs exist also but are negligible for a large property of 50,000 sqft.)

We take as a baseline a 3 year lease of 50,000 sqft of newly renovated and wired space in the new suburb nearest to the airport, with 100% occupancy. This baseline has a predicted rent per sqft of $41, which puts the total annual rent in the ballpark of $2 million. The data should allow us to predict rent per sqft for 95% of properties to within about ±$5, or total rent for 50,000 sqft to within ±$0.25 million. In the following bullets we describe estimated discounts and premiums due to the factors listed above.

- City properties command an average premium of nearly $6 per sqft over the suburbs.
- Number of years since renovation is an indicator of the state of repair of a property. Compared to newly renovated properties, 1 and 6 and 15 years since renovation grant average discounts of $1.75 and $5 and $7 per sqft, respectively. (The median years since renovation are 27, 13 and 14 for the city, the new and the old suburb, respectively.)
- Missing modern wiring grants an average discount of $4 per sqft.
- Occupancy rate reflects desirability of a building. There is an average discount of $1 per sqft for every 10% decrease in occupancy. (The median occupancy rates are 90%, 87% and 78% for the city, the new and the old suburb, respectively.)
- Distance from the airport makes a difference; distances range from 10 miles on one end of the new suburb to 20 miles on the remote end of the old suburb. For the 20 miles distance there is on average a $6.4 discount per sqft over the 10 miles distance.
- Leases vary from 3 to 10 years, and there is a discount for longer term leases. For example, 6 year and 10 year leases have on average discounts of $1 and $1.5 per sqft over 3 year leases.

As an illustration, a 3-year lease on a space with 50,000 sqft, in the center of the new suburb, 12 miles from the airport, with 4 years since renovation, with 80% occupancy, and with modern wiring, is expected to cost $33.7 per square foot (95% range: $29.0…$38.3), for a total annual cost of $1.685 million (95% range: $1.450…$1.915 million).

As we see there is still some variation in rental costs which our model has been unable to capture, and it might be possible to find bargains at the lower end of the predicted ranges.

The factor with greatest impact on cost is the choice of city center versus suburbs (city premium = $6/sqft). The old suburb is less expensive than the new suburb, partly due to being older but mostly due to greater distance from the airport (old suburb discount = $3.5/sqft on average over the new suburb). Two other factors that entail discounts are high number of years since renovation and low occupancy rates, although such properties need to be looked over carefully because both factors may be proxies for potential problems.

Technical Summary

We investigated factors that drive rental cost in the target city using multiple regression. We fitted a regression model to the data on hand in order to predict rental cost from various features recorded in the data. Because of larger variability in total rental cost for larger properties, we decided not to use total rental cost as the response variable, but rent per sqft instead. In fact, we will see that this choice of response variable works well and results in a satisfactory model fit that validates statistical inferences. The predictors we included in the model are the following:

1) *Location*, a variable cost, which expresses differences between city center, new and old suburb.
2) *1/SqftLease*, a term that is needed to express overall fixed cost when modeling per unit cost.
3) The interaction of *Location* and *1/SqftLease*, to express differences in fixed costs between city center, new and old suburb.
4) The product of *Parking* and *1/SqftLease*, to include the fixed cost of a number of executive parking spaces.
5) *Occupancy*, as a measure of overall desirability of a property.
6) *Log(LeaseLength-2)*, to include an expected discount for longer leases. The logarithm expresses a diminishing return in the sense that increasing a lease from 3 to 4 years should have a proportionately larger effect than increasing it from 9 to 10 years. The minimum lease length is 3 years, so in order to obtain a zero discount for 3 years we subtracted 2 from the length of the lease (which is admittedly somewhat arbitrary, but works well on this data).
7) *Log(1+Renovation)*, to include an expected discount for dilapidation. But again a logarithm was used to express diminishing returns, as an increase from 0 to 1 year should have a greater effect than an increase from 20 to 21 years. We added 1 to achieve a zero discount for zero years, that is, newly renovated properties.
8) *Distance from the airport* turns out to be necessary to express a discount for remoteness from the airport.
9) *Wiring* for modern computing and telecommunications describes a premium for landlords' investment in modernization of their properties.

Other factors that influence rents undoubtedly exist, but the above list includes terms that have been found important and/or are economically meaningful.

The following tables show the results (after setting aside two extreme rental properties, see below) of fitting the regression equation described above:

Summary of Fit

RSquare	0.814613
RSquare Adj	0.804949
Root Mean Square Error	2.321446
Mean of Response	29.57139
Observations (or Sum Wgts)	223

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Location	2	2	917.79321	85.1526	<.0001
1/Sqft	1	1	148.38583	27.5344	<.0001
Location*1/Sqft	2	2	7.60681	0.7058	0.4949
Parking*1/Sqft	1	1	2.35226	0.4365	0.5095
Occupancy	1	1	307.72671	57.1016	<.0001
log(LLen-2)	1	1	68.24317	12.6632	0.0005
log(1+Renov)	1	1	722.46720	134.0606	<.0001
DistAirp	1	1	107.95992	20.0330	<.0001
Wiring	1	1	565.56366	104.9456	<.0001

Expanded Estimates

Nominal factors expanded to all levels

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	36.344279	2.469441	14.72	<.0001	31.476343	41.212215
Location[CITY]	3.8260977	0.29901	12.80	<.0001	3.2366684	4.4155269
Location[SUBNEW]	-1.731821	0.43469	-3.98	<.0001	-2.588713	-0.87493
Location[SUBOLD]	-2.094276	0.499168	-4.20	<.0001	-3.078272	-1.110281
1/Sqft	5026.6316	957.9424	5.25	<.0001	3138.2679	6914.9953
Location[CITY]*1/Sqft	1591.0456	1363.69	1.17	0.2446	-1097.156	4279.2473
Location[SUBNEW]*1/Sqft	-928.4294	1278.845	-0.73	0.4686	-3449.379	1592.5198
Location[SUBOLD]*1/Sqft	-662.6162	1386.774	-0.48	0.6333	-3396.323	2071.0906
Parking*1/Sqft	1059.9079	1604.294	0.66	0.5095	-2102.59	4222.4056
Occupancy	10.843677	1.435002	7.56	<.0001	8.0148995	13.672454
log(LLen-2)	-0.739977	0.207944	-3.56	0.0005	-1.149891	-0.330062
log(1+Renov)	-2.520812	0.217716	-11.58	<.0001	-2.949989	-2.091635
DistAirp	-0.642201	0.143482	-4.48	<.0001	-0.925044	-0.359359
Wiring[NO]	-2.0759	0.20264	-10.24	<.0001	-2.475358	-1.676443
Wiring[YES]	2.0759004	0.20264	10.24	<.0001	1.6764429	2.4753579

The overall measure of fit, R-square, has an astonishingly high value of 0.81, which indicates overall closeness of fitted and observed rents per sqft in relation to their overall variability.

The RMSE is $2.32; it is an overall measure of closeness between fitted and observed rents per sqft in absolute magnitude (dollars). We rounded the RMSE up to $2.50, which is a conservative thing to do. In the executive summary we mentioned ±$5 as a margin of error for predicted rents per sqft, which equals approximately two RMSEs. For properties with 50,000 sqft, the $5 margin of error for predicted rents per sqft translates to a margin of error of ±$250,000 for predictions of total rent. These figures give a rough idea of the precision that can be achieved with the reported regression model.

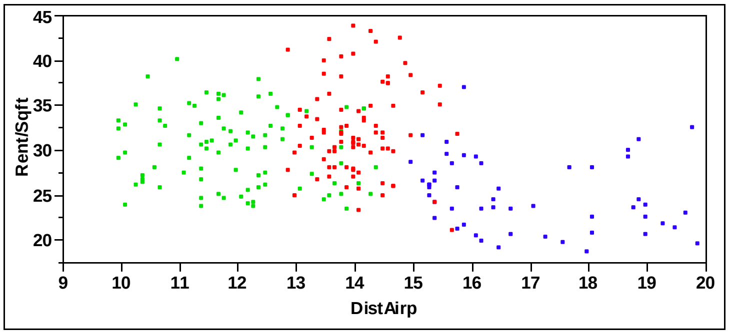
However, the prediction range quoted for the second scenario in the executive summary was not based on approximate RMSE considerations but on technically exact methods. Here are "exact" predictions for both scenarios, with their respective predictor values as inputs:

a) Baseline scenario: *Location*=new suburb, *Parking*=0, *Occupancy*=1.00, *Length of Lease*=3 years, *Renovation*=0 years, *Distance to Airport*=10 miles (= the approximate minimum for the metropolitan area due to remote location of the airport), *Wiring*=yes. The resulting prediction of rent per sqft is $41.18, and the 95% prediction interval is $36.39…$45.99.
b) Final scenario: *Location*=new suburb, *Parking*=0, *Occupancy*=0.80, *Length of Lease*=3 years, *Renovation*=4 years, *Distance to Airport*=12 miles. This latter value is the average distance of properties in the new suburb from the airport.
The resulting prediction of rent per sqft is $33.68, and the 95% prediction interval is $29.03…$38.36.

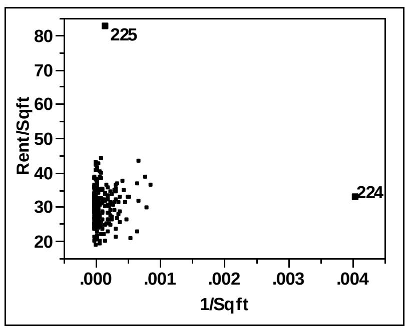
The impact of the individual predictors on rent as estimated by the above equation is as follows:

1) *Location* has considerable impact in that there is a premium on rent in the city center. The zero-centered coefficients for the three locations are approximately $3.83 (stderr=$.30) for the city center, -$1.73 (stderr=$.43) for the new suburb, and -$2.09 (stderr=$.50) for the old suburb. This amounts to an estimated average premium of $5.56 for the city center over the new suburb, and $5.92 over the old suburb.
2) *1/SqftLease* estimates the fixed cost for renting at all. The estimate of $5,027 (stderr=$958) is statistically significant (t=5.25), but for the size of property of interest (50,000 sqft) with total rental cost in excess of $1.5 million, any contribution below $10,000 can be safely neglected because it amounts to less than 1% of the total.
3) The interaction of *Location* and *1/SqftLease* can be neglected also because it describes differential fixed costs among locations, which are amounts even smaller than overall fixed cost. Also, the coefficients are statistically insignificant (F=.7, see Effect Tests).
4) The product of *Parking* and *1/SqftLease* can be neglected also. The estimate of about $1,060 per parking space is statistically insignificant (t=.66). In addition, over 100 properties have no executive parking, and over 200 have less than 5 spaces; the maximum is 12 spaces.
5) *Occupancy* is an important predictor with a coefficient of $10.8 (stderr=$1.44). This value expresses the effect of a change of occupancy from 0.00 to 1.00, which is unrealistic. More meaningful is a change of occupancy of 0.1 (10%), which entails an estimated average change in rent per sqft of $1.08. In the executive summary we rounded this value to $1.
6) *Log(LeaseLength-2)* has a coefficient of -$0.74 (stderr=$.21), expressing a discount for longer leases as expected. Because the diminishing return effect of this transformation is somewhat difficult to describe, we presented in the executive summary estimated average discounts for 6 year and 10 year leases.
7) *Log(1+Renovation)* has a coefficient of -$2.52 (stderr=$.22), expressing a discount for more years since renovation. Again, the diminishing return effect is opaque, so we presented in the executive summary estimated discounts for 1, 6, and 15 years since renovation.
8) *Distance from the airport* plays a surprisingly strong role: the coefficient of -$.64 (stderr=$.14) expresses an average estimated discount of 64 cents per additional mile from the airport. This may not seem much, but in view of the fact that the metropolitan area stretches from 10 to 20 miles from the airport, this amounts to a differential discount of up to $6.4 between the extremes of distances. The plot of *Rent/Sqft* versus *DistAirp* below reflects the geographic situation well; the color coding is green for the new suburb, red for the city center, blue for the old suburb. Recall that we didn't see a strong difference in the estimates of rent per sqft between new and old suburb; now we see that there exists a difference, but it is reflected in the equation largely by differences in distance from the airport.

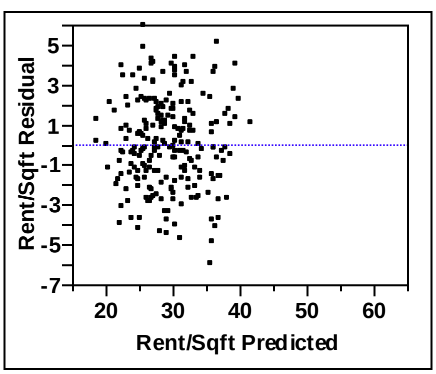
9) *Wiring* is the final factor in the equation, with a premium of $4.2 per sqft (stderr=$.40) if present. (Note that the coefficients are centered at zero, hence the full difference is obtained by doubling the numbers, including the standard error.)

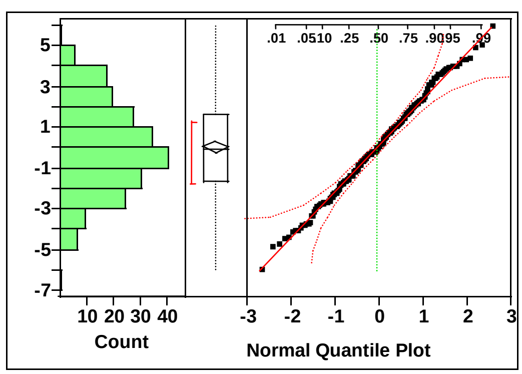


Finally, we address model adequacy and outliers. One vertical (case 225) and one horizontal outlier (case 224) were apparent in the plot of *Rent/Sqft* versus *1/SqftLease*. The horizontal outlier is a tiny property with less than 250 sqft. The vertical outlier might be a mis-recording because there are simply no properties with a rent per sqft as high as $80; all other rents are below $45 per sqft. We removed both properties for the subsequent regression analysis.



A residual analysis did not reveal any problems. The plot of residuals versus predicted has no discernable pattern, and the normal quantile plot of the residuals looks good:





Leverage plots (not shown) did not reveal any particularly influential points.