

DISCLAIMER: This publication is intended for EDUCATIONAL purposes only. The information contained herein is subject to change with no notice, and while a great deal of care has been taken to provide accurate and current information, UBC, their affiliates, authors, editors and staff (collectively, the "UBC Group") makes no claims, representations, or warranties as to accuracy, completeness, usefulness or adequacy of any of the information contained herein. Under no circumstances shall the UBC Group be liable for any losses or damages whatsoever, whether in contract, tort or otherwise, from the use of, or reliance on, the information contained herein. Further, the general principles and conclusions presented in this text are subject to local, provincial, and federal laws and regulations, court cases, and any revisions of the same. This publication is sold for educational purposes only and is not intended to provide, and does not constitute, legal, accounting, or other professional advice. Professional advice should be consulted regarding every specific circumstance before acting on the information presented in these materials.

© **Copyright: 2014** by the UBC Real Estate Division, Sauder School of Business, The University of British Columbia. Printed in Canada. ALL RIGHTS RESERVED. No part of this work covered by the copyright hereon may be reproduced, transcribed, modified, distributed, republished, or used in any form or by any means – graphic, electronic, or mechanical, including photocopying, recording, taping, web distribution, or used in any information storage and retrieval system – without the prior written permission of the publisher.

LESSON 4

Market Identification and Property Characterization

Note: Selected readings can be found under "Online Readings" on your Course Resources webpage

Assigned Reading

1. UBC Real Estate Division. 2014. *BUSI 344 Course Workbook*. Vancouver: UBC Real Estate Division. Lesson 4: Market Identification and Property Characterization
2. Appraisal Institute of Canada & Appraisal Institute (US). 2010. *The Appraisal of Real Estate, Third Canadian Edition*. Vancouver: UBC Real Estate Division. pp. 13.5–11.

Recommended Reading

1. Appraisal Institute of Canada & Appraisal Institute (US). 2010. *The Appraisal of Real Estate, Third Canadian Edition*. Vancouver: UBC Real Estate Division.
Chapter 13: The Direct Comparison Approach
Chapter 14: Comparative Analysis
Chapter 15: Applications of the Direct Comparison Approach
2. UBC Real Estate Division. 2009. Supplemental Information to Accompany BUSI 344 Lesson 4.

Learning Objectives

After completing this lesson, the student will be able to:

1. identify the main steps of an appraisal assignment leading to selecting the correct market segment;
2. explain the two fundamental components (subject/market) of the process of identifying the correct market segment;
3. recognize the four "dimensional aspects" of the valuation solution, how these help and determine margins for the data frame and the most likely dataset to be used in the analysis;
4. discern the relevance and differences between the subject neighbourhood, the relevant district, and market segment;
5. distinguish a "sufficient" dataset from a data frame, and identify aspects or characteristics which may require additional research;
6. apply statistical and graphical methods to depict and quantify aspects of asset characteristics;
7. determine when circumstances dictate reconsideration of a dataset;
8. explain how experimental method principles of control, randomization, replication, and blocking apply to valuation; and
9. apply two basic analytical tools, paired group comparison and simple regression, in valuation situations requiring property characterization.

Instructor's Comments

This lesson and the next build on the exploratory data analysis in Lesson 3, but go a notch further in applying these techniques in practical valuation circumstances. In other words, rather than talking about theory and "what ifs", we observe examples of real-life applications for these principles.

This lesson focuses on market identification and property characterization. Previous lessons have emphasized using the computer to aid in market analysis, through statistical depictions and graphical visualizations. We will use these tools to identify the most relevant market (based on optimal use) for the subject and the key explanatory and predictor variables. After this market identification, we will "locate" where the subject belongs in this market. In working through property characterization, we may find we need to revisit the market segmentation decision.

We will relate the economic concepts of *neighbourhood*, *district*, and *market segment* to the data-analytic concepts of *database*, *data frame*, *dataset*, and *information set*. We will focus on the variables that characterize a particular property, locating the optimal use of the property in the right market segment. The statistical tools we will use include two simple but highly useful procedures: (1) paired group (centres) comparison, and (2) simple (two-variable) regression.

This lesson is written from the perspective of a valuation practitioner, but the concepts presented can be generalized to many aspects of real estate analysis. The techniques covered are useful for professionals who provide analytical services to a commercial client or governmental entity, as well as to consumers of such services, to support a better understanding the relationships between the appraisal problem, scope of work, variables under examination, and statistical tools applied.

Where Does Market Identification Fit in Appraisal?

Most appraisal assignments are straightforward and there is little difficulty in identifying the right market segment. However, as assignments become more complex, this choice may become increasingly difficult. A key consideration is that *the biggest appraisal mistakes are made in choosing the wrong market segment to analyze*.

Most appraisal assignments start with problem identification and defining the scope of work. You can consider asset analysis to be comprised of five parts:

1. Identify the **problem** to be solved;
2. Determine the **scope** of work necessary to solve the problem;
3. Complete the **research** necessary to **reduce** the data to a useful and informative set;
4. Complete the **analysis** necessary to produce a credible result; and
5. Interpret and **communicate** the results.

From a data-analytic perspective, the focus is economic and statistical similarity, and the data decisions which must be made. Therefore, our focus is on Steps 3 and 4 above: **research** and **analysis**. Necessarily, there is overlap between the two activities. Getting the right data may involve some analysis, and analysis may impel further research.

Another way of looking at the nature of the valuation problem is as follows:

1. Identify the right market segment (market identification); and
2. Locate the subject within that market (property characterization).

Market identification is a process of recognizing and applying key variables. This is essentially the first three parts of asset analysis listed above. To arrive at the appropriate market segment dataset, first we review and

examine the effect of the client's appraisal problem, and second, determine the scope of the solution. Third, we consider the four dimensions of similarity (transaction, time, space, and utility), specifically how they affect our reduction of a database into sets of data useful for analyzing and understanding market dynamics. That is, we apply them to our data decisions. Finally, the issue of optimal use is included in this sequence, in terms of how this may change an initial perception of the right market and right dataset.

The next section will review each of these five market identification steps in turn.

Market Identification Steps

As mentioned earlier there are five steps in identifying the right market:

1. **Problem Identification** – The initial information and question(s) provided by the client.
2. **Scope** – The solution strategy identified by the appraiser, in consultation with the client.
3. **Similarity** – A measure of relative equivalency in aspects of time, space, and utility (physical, legal, and financial qualities).
4. **Data Decisions** – Defining and refining (reducing) available data.
5. **Optimality** – Research and analysis may lead to a reconsideration of each of the above steps. This is where the process may become iterative.

Market Identification Step 1: Identify the Appraisal Problem

Identifying the right market starts with identifying the problem to be solved. This contains the following elements:

- **Identify the client and other intended users.** This helps the appraiser ascertain the type, focus, and level of work to be performed, as well as the depth, breadth, detail, and language of the report.

Examples:

1. Work for an attorney will usually differ from work for a lender.
2. Work for a developer could lead you to the conclusion that the value desired is the land for assemblage, rather than the current retail use.

Who the client is may make a strong impact on setting the rest of the problem scenario. It is a dominating factor in the scope of the report and must be clearly communicated.

- **Intended use.** A lender may need to make a loan decision or foreclosure decision. If the use is for a foreclosure decision, the client may want more emphasis on market analysis and in particular the trade-off between a quick sale and the potential costs of carrying the property for a more "normal" marketing period. The correct market might include only other quick sales, not necessarily meeting market value conditions. *Intended use will affect what data and what variables are important. It will also affect the scope of the report.*
- **Purpose.** Market value is but one type of value. The lender noted above may want a liquidation value or disposal value to help compare the carrying costs of waiting for perhaps a higher market value under normal marketing effort and time. *The type of value will affect what data and what variables are important.*

- **Date of Value.** Even the exact minute of value can make a difference while a building is burning. While a current value is the most common, there are many cases and reasons for retrospective (historical) valuations. Similarly, many construction and development valuations require a prospective (future) value, perhaps upon completion of construction or upon full absorption of rental space to tenants. Investment analysis invariably involves some forecasting of future elements. Markets exist and behave differently at different times. *Date of value clearly identifies the time dimension of the valuation solution.*
- **Property characteristics.** Physical, legal, and financial qualities of a property launch the main challenges of market identification. These substantially define property characterization. Property characteristics are treated differently because their correct identification requires competency in that type of assignment. This lesson provides methods for the recognition and control of property characteristics as analytical variables. *Property characteristics affect what data and what variables are important.*
- **Assumptions.** Appraisal assumptions can be classified into four types:
 - Limiting conditions – liability and use.
 - General – typical and ordinary.
 - Extraordinary – that which may be false.
 - Hypothetical – that which is false.

The above six elements of identifying the appraisal problem directly affect decisions regarding the data to be used, the variables to be exploited, and the type of statistical/graphical tools to be used. They are key in defining the scope of work and scope of reporting.

Market Identification Step 2: Scope

Having identified the client's problem to be solved, the appraiser now establishes the levels of property identification, inspection, data research, and analysis to be performed. The scope of the solution has a large impact on decisions for data and analysis, including the choices of variables used.

The **breadth** of the research element is itself comprised of four fundamental dimensions. Each of these requires an "extent" or framing definition. The four dimensions run parallel to the "sequence of adjustments" found throughout appraisal literature:

1. Transaction conditions – e.g., motivated sales or not;
2. Time segment researched;
3. Space – e.g., explore a similar market or not; and,
4. Characteristics range – e.g., what range of size is considered competitive?

Complexity involves a trade-off for the analyst (as well as for the client) between meticulous detail versus an easily grasped and communicated model structure. Two reporting issues arise:

1. Greater complexity may require a greater effort and competence; and,
2. Any gap between the appraiser's familiarity with a method and a client/user's familiarity (and competence) must be bridged by the report.

Detail in a report can vary. As well, the level of detail necessary in the analysis may not match that of the report, which may require only summations or conclusions with no detail. In such a case, the report will need to explain and reconcile any differences.

Emphasis is an often overlooked aspect of scope. The client may want the appraiser to pay particular analytical attention to a certain aspect, such as the sale of a particular competitor's property. Similarly, a client may want an appraiser to fully develop and report on a particular aspect (say the cost approach), while the rest of the report is delivered verbally.

The above four elements interact to determine the scope of the report.

Market Identification Step 3: Similarity

The third step in identifying the right market is specifying the similarities (and differences) between our subject and its environment. That is, we need to determine all the elements of comparison that, in essence, define our subject. These elements fall into one of four *dimensions*:

- **Transaction:** rights conveyed, financing, sale conditions (motivation);
- **Time:** market conditions;
- **Space:** adjacent and proximate influences, neighbourhood, district, market segment, and illustrative comparables; and,
- **Utility:** property characteristics: physical, legal, and financial qualities.

There is usually statistical independence between each of these dimensions, meaning that, in general, the variables from each do not have significant relationships.¹ For example, the motivation aspects of a foreclosure sale (transaction) would be unrelated to recent price changes in the sub-market (time) or the number of bedrooms in the house (utility). However, within a dimension, we often have high dependence: e.g., house living area and bedroom count are almost always highly correlated (more area usually means more bedrooms). This independence across dimensions is what allows us to use the statistical tools of *paired group comparison* and *simple regression* for market identification analyses. However, keep in mind that independence across dimensions is not absolute nor universal. It is the appraiser's task to both exploit independence when possible and to expose *dependence* when it occurs.

We will now examine each of the four dimensions in turn: transaction, time, space, and utility. In doing so, we will keep our goal in mind: relying on them to help identify the right market so we can meet our client's need.

Dimension 1: Transaction – Contract terms

The transaction dimension categorizes *non-property elements* of the contract, including *rights conveyed*, *financing*, and *sale conditions*. Each of these may be important to the identified appraisal problem, the scope of work, the dataset, and the statistical/graphical model.

Rights Conveyed

These elements entirely relate to the "bundle of rights" conveyed, not physical or income characteristics of the property or characteristics of the market. Contract terms best define transaction variables. Errors in identifying transaction elements constitute the source of the "second biggest mistake" to which an appraiser is vulnerable.

Financing

Financing may impact sale price for a number of reasons. The vendor may offer a take-back mortgage to help sell the property, the existing mortgage may be assumable to pass on a low interest rate, or a variety of tax or investment reasons may alter the financing conditions.

¹ This relative independence is the reason the aspects are called dimensions. Think of these as four perpendicular layers or planes, with values on each plane having no effect on the value of other planes.

These factors tend to relate to personal motivation, not property characteristics. Although in some situations, such as government subsidies for low-income home ownership, financing factors may in fact be connected to a certain property type or buyer type.

Sale Conditions – Motivation

Non-realty motives may arise for a number of reasons. A person may wish to buy a particular house, because of personal memories of growing up in that house. That person's motivation is non-market driven and atypical of other buyers. This factor is difficult to measure, but real.

Sale condition effects can be statistically analyzed by paired group comparison or simple regression. Because motivation is difficult to quantify, we may use a proxy variable.² We will illustrate this in Case 3 later in this lesson.

Dimension 2: Time – Market Conditions

Market conditions will vary over time due to trends or events or both.

Trends

Groups of similar properties tend to trend together. Time (market condition) adjustments can be analyzed using price indices, scatter plots, trend-line fitting, or simple regression. Paired group comparisons are discouraged because they can lead to a significantly wrong conclusion, even indicating the wrong market direction. The reason for this is that grouping of a continuous variable (time) causes loss of information. If there is a causal or random grouping of data within the groups, you are actually introducing a biasing variable for no beneficial reason – for example, if you know the day, do not round to the nearest month, as this reduces precision and may introduce bias. Similarly, if you can use a scatter plot, do not use a histogram. Be cautious about grouping, use it only when you have a specific reason for doing so.

Events

While most economic forces are smooth or dampened in their impact on price levels, some events cause dramatic, even sudden changes in price levels – for example, natural disasters such as floods or earthquakes; or societal disasters, such as terrorist acts or human accidents such as the Chernobyl melt-down.

Economic demand or supply factors can also present sudden impacts. The announcement date of the opening of a large manufacturing plant is an event that would be the source of price change, impacting the volatility of the market and the nature of the demand. Similarly, the announcement of the opening of a road or transport facility will provide an event impact on price expectations and demand levels.

Paired group methods may in some instances provide insight, but should only be used in conjunction with graphical and trend-line fitting methods (simple regression). For example, consider an event-driven demand/supply change where an otherwise level market experiences a newsworthy event, such as a major plant opening or closing. Grouping data before and after the event could provide an event-time adjustment, beyond any needed trend adjustment.

Dimension 3: Space – Location

Spatial, or locational, relationships can be categorized into adjacent, proximate, near, and remote influences. Once again, we are selecting a dataset which is economically homogeneous to the subject, but now measured geographically. Locational analysis has some similar statistical characteristics to those of time-series analysis, in

² A proxy variable "stands in" for another, perhaps unmeasurable variable. These are also known as indicator, surrogate, latent, or manifest variables.

that data points close in space (or time) tend to be correlated. The appraisal model process takes advantage of this correlation as a criterion of similarity.

A *neighbourhood* is defined as a group of complementary land uses, while a *district* can be defined as a grouping or cluster of similar land uses. *Market segment* is defined as a homogeneous market, as characterized by a set of variables, for which we will use our four dimensions: transaction, time, space, and utility. We will focus on market segment in this lesson and the next when referring to issues related to location.

The typical *independence* of location from the other three analytical dimensions enables the simple tools of compare means and simple regression.

Dimension 4: Utility

Utility of a property to an owner can come in the form of physical, legal, or financial characteristics. Taken as a dataset, utility characteristics will tend to be highly correlated with each other. Thus, they form the basis of a "good comparable".

- Physical characteristics: site elements and improvement elements. They are often strongly associated with each other, but tend to have low correlation to variables of time, space, or transaction.
- Legal restrictions: zoning or other public regulation of real estate use which can narrow or enhance the utility of one property as compared to another. For example, consider two identical houses across the street from each other but in different jurisdictions. One allows basement suites, the other does not. There would be high correlation within each side of the street, but lower correlation as to comparisons across the street. Values on either side of the street would not have significant correlation to time or transaction terms. Of course, there would be spatial correlation connected with the side of the street on which the properties are located.
- Financial issues: e.g., income flow, operating expenses, quality of management, tenant mix, rent concessions, ease terms, lease expiration dates, renewal options, and lease provisions. These will normally show statistical independence to the other dimensions, but correlation among themselves.

Dimensions Independence: A Key Modeling Consideration

The four dimensions of similarity: transaction, time, space, and utility, serve as the basis for our valuation modeling pursuits. In evaluating similarity (the third step in the market identification process), the appraiser must keep in mind that the nature of available statistical tools is different for each dimension. As well, statistical independence between the four dimensions, and some sub-parts, allows the use of more easily understood and simple statistical tools.³

The rest of this lesson and the next will rely heavily on the opportunities and obstacles presented by the relative independence of various predictive variables.

³ The goal in modeling is to take advantage of the correlations *within* each of these dimensions and the relative independence (low correlation) *between* these dimensions. Transaction terms and utility elements can be easily conceptualized and lend themselves readily to "one-dimensional" analyses, such as placing the subject in the scale of relative values of comparable sales. However, a potential statistical problem with these analyses (particularly with physical elements) is collinearity or correlation between explanatory variables.

Market Identification Step 4: Data Decisions

We now apply what we know about the four dimensions of similarity to the hierarchy of analytical data groupings. Different data/information sets are useful for different modeling purposes. This is similar in concept to the fact that sometimes it is the neighbourhood, sometimes the district, and often the market segment which is important for a particular analysis.

- **Information** is data which has been sorted for a particular purpose in a manner which is understandable to a human. The entire appraisal process is designed to take data, enhance it, and reduce it so that it is meaningful to a human.

We now look at the reduction of data in a database to create a meaningful and useful set of information for a specific analysis method. We work the database into a data frame, to a (market segment) dataset, to an information set, and finally to a (reporting) illustrative comparable set. Definitions are given below.

- **Database** is a collection of records stored in a systematic, structured way most often with some way to uniquely identify those records (the mailing address of a property for example).
- **Data frame** is a selection of data chosen from a database (or collection of databases), such that any information outside the data frame is judged to have no or negligible impact on the solution desired.
- **Dataset** is the selection of data chosen from the data frame for direct analysis.
- **Information set** is the selection of data (usually comparable sales from within the market segment), which the appraiser considers directly competitive in terms of the four dimensions of similarity. The difference between a dataset and information set is editing and enhancement, which includes verification and transformation.
- **Illustrative dataset** is used for reporting or presentation purposes. It parallels the comparable sales used in traditional appraisal. The illustrative set has one primary purpose: it is a further data reduction to help simplify the methodology used for those who may have no training or experience in data analysis methods.

Data reduction strategies will be detailed in a later section in this lesson.

Market Identification Step 5: Optimality

A paradox of appraisal is that the characteristics of the subject identify the market in which it competes, and yet market forces determine which characteristics are important. In practice, which comes first? This problem can be called "analytical iteration". That is, as each step is completed it may lead to some reconsideration of a previous step. In fact, there could be many such iterations.

- **Step 1: Identify the Appraisal Problem.** Physical inspection or legal and financial discovery may require a redefinition of the appraisal problem from what your client originally believed.
- **Step 2: Scope.** Developments during the inspection, data collection, or research phases of the assignment may result in a change in the client's needs. You will have an obligation to inform the client of any such significant changes which might affect the scope of the work.
- **Step 3: Similarity.** Changes in the appraisal problem or needed scope can easily (and often do) change the relevant market. Changes in the problem or scope must be considered in terms of the four dimensions.

- Step 4: Data Decisions. Even the choice of database might change. For example, where you discover the optimal use is to develop into a retail centre, you may need to gather commercial data and put aside your residential information.
- Step 5: Optimality. Identifying the most profitable, maximally productive use may sometimes involve "alternative use" appraisals on each of the possible uses, in order to determine which use produces the best financial opportunity. Such studies must include analyses of risk for each potential use considered here.

Market Identification: Checklist

Now that we have examined the market identification steps in detail, we can summarize the market identification process in one handy reference document. The checklist in Table 4.1 will serve as a good reference tool in reading through the case studies at the end of the lesson. Beyond the course, this framework can also serve as a helpful reminder for a comprehensive approach to market identification – it will help organize your thoughts and ensure you are following a logical and consistent approach.

Controlling Elements		Explanation or Examples	
Problem identification			
Array and Breadth as requested by the client	Intended use	Loan security, litigation, tax appeal	
	Value definition	Market value, liquidation value, insured value	
	Dates	Current, prospective, retrospective	
	Characteristics	Property type: SFD, office, hotel Size, lease types, key amenities	
	Assumptions	Typical, hypothetical, extraordinary	
The scope of work: Data			
Array and Breadth, as determined by the appraiser Dimensions of Similarity: Transaction Time Space Utility	Data Decisions	Data sources	Local MLS, public records, appraiser files
		Data frame	Set to include everything believed to be of impact
		Dataset	Directly competitive market segment
		Information set (market segment)	A subset of the dataset, reduced for a particular analysis or analytical method
		Illustrative set	Comparables which "best" typify the market
The scope of work: Analysis			
Complexity, as determined by the appraiser	Unit(s) of comparison	An example of how analysis decisions (in this case unit-of-measure) interplay with data decisions	
	Statistical tools		
	Graphical tools		
The scope of work: Reporting			
Detail and Emphasis	Intended use	Oral, testimony, data stream, statement, summary, "self-contained"	
	Analytical scope		
	Client familiarity		

Property Characterization: Moving From Research to Analysis

After identifying the correct market for the subject, we then move on to characterizing the property within that market. So far we have focused on the *research* part of the valuation problem. Our focus now concentrates on *analysis* of the chosen dataset. This modeling effort will be based on the principles of experimental design, which we will explore in the following section. In addition, we will explain data reduction strategies, a key aspect of valuation modeling that was mentioned briefly in Lesson 3.

Modeling Principles and Objectives

Our goals in promoting statistical analysis are to *increase human understanding* and to *reduce variability*. To achieve these goals, we seek reliable and credible results through modeling:

- *Reliability* is defined as: freedom from random error.
- *Credibility* is defined as: worthy of belief. Credible results come from a combination of professional competence and ethical behaviour. Credibility is also dependent on our ability to effectively *communicate* results, whether verbally, in a form report, in a storyline summary, through a results-only account of findings, or through a full narrative discourse.
- *Modeling* is the mechanism by which we achieve the highest possible reliability. Modeling is both an art (a set of skills) and a science. Effective modeling requires both the right data and the right analytical procedures.

Statistical modeling requires decisions based on the practitioner's knowledge, education, and experience. Appraisal assignments often focus on an estimate of value, sometimes with an opinion of reliability (relative precision) attached. However, the objective of a valuation modeling assignment can vary widely depending on which property dimension we are dealing with.

Our focus is on applying the right analytical procedure to the right data. We first look at how the variable and measure types apply to various property types. Then we examine two of the simpler modeling tools: (1) paired group comparison (contrasting measures of central tendency, such as the mean or the median); and (2) simple (two variable) regression. These tools are examined and applied in the cases at the end of this lesson.

We will not use multiple regression analysis (MRA) in this lesson, mainly for reasons of clarity: simpler tools may be more intuitive and accessible to the analyst, as well as the client. By focusing on a combination of simpler tools here, we aim for an overall goal of lower variability and more precision, combined with ease of use. Often simple tools are the best ones to use. Multiple regression analysis will be covered in depth in Lessons 6 through 8.

Four Principles of Experimental Design

Experiment is a term more commonly used in science than in valuation. However, experimental methods are implicitly used any time we look at how one variable (dependent or response variable) is controlled or manipulated by another (independent or factor variable). In valuation, we are constantly analyzing economic experiments, viewing the cause and effect between certain variables so we can estimate value. However, we can also use experiments to measure reliability, feasibility, optimal use, or other client questions needing answers. In valuation (as in most social science problems), we must deal with observational data, rather than experimentally generated data.

In general, economics gives us the nature of the *cause* portion of our answer, while the statistics gives us the *direction* and *size* of the effect.

Recall that our statistical objective is to *increase precision* and *reduce variability*. To do so, we will apply all four of the principles of experimental design for valuation data analytics:

1. Control: limiting what data is used.
2. Randomization: relying on chance.
3. Replication: considering similar experiments.
4. Blocking: exploiting natural and logical groupings of variables.

Next, we will explain each of these terms.

1. Control

One way we can reduce variation is by constricting our research data to a narrow range of similar (or identical) conditions. In doing so we can eliminate or isolate the effect of a specific variable. This is often accomplished by removing a variable from a model to see how the relationships change without that variable. By controlling for extraneous sources of variation, we reduce the variability of the response. A key to good control is to watch for collinear or correlated variables and not carry them along into our control group.

In real estate data, there are two basic types of control:

- For *categorical data*, we can simply include or exclude the category, or we can compare the groups and derive an adjustment. An example would be industrial properties with rail spur access. We can adjust for the difference or we can eliminate the dissimilar type of feature. For example, if our subject property has no rail access, we delete all sales with adjacent rail sidings.
- For *measured data*, we typically trim the data around the parameters of the subject property. For example, if the subject is 1,450 square feet, we might restrict our research data to properties that range from 1,300 to 1,700 square feet. Alternatively, if our subject property is one of the largest properties in an area, say 2,800 square feet, we might limit our research data to homes of 2,500 square feet or more.

In applying a control, we are reducing the size of what will become our information set. The trade-off is that we may lose some beneficial randomization of other variables.

2. Randomization

This is a way of equalizing or "averaging out" the effects of unknown or uncontrollable sources of variation. Randomization protects us from effects we might have missed, such as an undiscovered outlier.

The theory underlying randomization is that troublesome variables, whether known or unknown, will tend to be evenly spread out, ideally with a random assortment on either side of the subject's parameter. This applies to categorical as well as measure variables. This is very similar to the traditional appraisal process of bracketing, wherein the appraiser attempts to have at least one comparable greater than and one comparable less than the subject.

We do go one step further here, in that our objective is not only to bracket, but to balance as much as possible, with a similar number of comparables greater than and less than in important characteristics.

A practitioner maxim says, "Control what you can, and randomize the rest".

3. Replication

Replication in experimental design means one's results should be able to be replicated by others. For example, in accounting this is related to the objectivity principle, which says that all accounts must be based on objective and verifiable facts, such that a third party (e.g., an auditor) can review the work and see the basis for the conclusions. This principle equally applies to appraisal work, where a third party (perhaps another appraiser) should be able to review the valuation results and make sense of them. It does not mean they would necessarily come to the same conclusion, but they could follow the logic from start to finish, seeing the basis for all of the decisions made along the way.

Replication can be difficult in appraisal practice especially in cases with limited data or an unusual subject property, where the answer we get may be unique, wrong, or biased by chance. At the extreme, an assignment involving a market with only one good comparable sale may have results that are nothing more than anecdotal. It is better than *no* information, but it is anecdotal and must be considered with reservation.

Another aspect of replication in appraisal is reviewing past experience with a property type or market, and applying this to subject. This can mean digging through your files or talking to colleagues about their experience with this type of problem. Because real property is unique, you may not be able to replicate results from these past "experiments" but you may be able to parallel them in addressing the circumstances at hand.

For example, say you are appraising the sole high-rise hotel in a mid-sized city. This property will have a fundamental economic connection to its community, but may be unique compared with smaller hotels and motels in the same city. It may not compete directly with other similar "tallest" hotels in other mid-sized cities, but it might parallel many aspects. Measures such as expenses, occupancy rates, and restaurant income will probably show similarities, in terms of relationships with local variables, which can be relied on and applied to the analysis of the subject. These similarities in elements and parameters permit the "experiment" to be replicated, allowing you to build on lessons learned in other markets and property types. It also permits a form of pairing or widening the dataset.

4. Blocking

Blocking involves taking advantage of natural or logical groupings of similar variables. An example is comparing warehouses on railroad sidings with warehouses not on sidings. Rail siding is the *blocking variable*. If the warehouse properties are similar otherwise, the blocking variable provides a valid basis for paired group comparison. We have two distinct groups of similar experimental units (blocks), and we can isolate the variability from the differences between the blocks.

The four dimensions of similarity, transaction, time, space, and utility/characteristics, are all forms of natural groupings of variables. By separating the data into blocks, you can explain and reduce the variability of the overall model. Blocking is perhaps the essence of this lesson, in that we principally rely on the natural similarities *within* these four dimensions of real estate data.

As noted at the start of this section, these four principles of experimental design help contribute to achieving our goals of *increasing precision* and *reducing variability*. We now proceed to examine data reduction strategies.

Data Reduction Strategies

A key aspect of the traditional appraisal process was overcoming difficulties in gathering reliable information. Data was gathered manually with great time and expense, and jealously guarded as proprietary information. The goal in most residential appraisals was to find and confirm data for a few (usually three to eight) comparable sales. *Data reduction* took place at the start of the process, in recognition of the difficulties in data gathering.

The outcome was as much dependent on what you could get as what was the best information. Thus, a source of competitive advantage for an appraiser or other real estate professional was access to good data.

Today, data is increasingly available in electronic form. Contemporary market analysis no longer focuses on finding data, but on *exploiting* the information within it. Data-analytic methods emphasize the use of the entire relevant dataset – the identified market segment and analytical sub-sets of this, which we call information sets. We aim to squeeze out the maximum information possible from the identified dataset. We *never* eliminate data without reason.

Recall our data reduction decisions from earlier in the lesson:

- Database: Numbers and facts available or attainable
- Data frame: Excludes what is believed statistically *not* useful
- Dataset: Limited to what is believed economically relevant (market dataset)
- Information set: Data controlled for analysis and comprehension
- Illustrative set: Data chosen for its ability to exemplify the solution

Narrowing down our data to what is useful for problem solving is a process of *data reduction*. Our goal is to create an *information set* sufficiently large to minimize data variability, yet sufficiently similar to the subject so as to provide a precise answer. In pursuing this, we must respect the trade-off between data precision (similarity) and analytical (model) precision.

Data Frame to Market Dataset

The *database* is all the data attainable. In reducing to a *data frame*, we are implicitly saying: "I don't think anything beyond this will be helpful". In reducing to the *market dataset*, we then say: "This is the part which I think will be needed for one or more of my anticipated analytical methods". In each reduction, our objective is to balance understandability, analytical potential, and data completeness.

How big a dataset you want or need depends on the problem at hand. You start with *all* the data in your market. The data you rely on in each step depends on the quantity, quality, and completeness of the market dataset, the subject's conformance within the neighbourhood or market, and the nature of the variables involved. Remember, the trade-off is between having data that closely matches the subject (data precision) and creating a model that takes into account as much information as possible (model precision). When you have a bunch of great comparable sales there is no trade-off problem. When you have comparables that vary in numerous ways, the data reduction questions become quite important. Sparse data may also cause you to return to expand the dataset as necessary. As you can see, data reduction can become an iterative process.

Market Dataset to Information Set

We have limited the *market dataset* to properties competitive to our subject property. The *information set* may be the whole of the dataset or a portion, depending on what we are trying to learn, what statistical principle is in evidence, and, importantly, the level of reliability we hope to achieve given the dataset we are working with. Data are numbers which *may* be useful; information is data organized into a form that is useful for solving problems.

An information set may involve just some of the data fields or variables available from the full competitive market set. For example, the information set may use only sale prices for a price index analysis. A mean may be calculated using the full dataset or it may be a trimmed mean, with some of the extreme values deleted.

Example: Control vs. Adjust Decision

You are appraising a single-family residence, typical of the local market except it has a pool. You have dozens of recent fee simple sales from the immediate area over the past six months. However, only three of these properties had a pool, with two of the properties much larger in living area and another similar in size. The market has been fairly level for over a year. Do you control for or adjust for the (binary) pool variable? Should you reconsider (expand) your market dataset? Your thinking process might proceed as illustrated below.

Control or adjust pool variable:

- I have too few pool sales to simply eliminate all the non-pool sales. I must go ahead and use them.
- I have three sales with a pool, but I might be able to get greater reliability (through randomization) by including non-pool sales, to be sure to bracket other features of the subject. Some of these non-pool sales will be matched or implicitly balanced out with the very large homes with pools.
- I will include the non-pool sales, and I will find a way to adjust for the pool variable.

Reconsider/expand dataset:

- I might be able to go back and expand my data frame to include more pool sales in nearby neighbourhoods – but then I would need to come up with a location adjustment. The increased information might be offset by need to adjust for the additional (location) variable.
- Based on my prior experience, I have found that pools are not really worth that much in this neighbourhood, perhaps contributing less than 2% of the overall value, and I will probably end up adjusting about that amount. I just need to support the adjustment, to make sure it is not higher or lower. I will support this with paired group analysis.
- I will not expand the data frame.

Measure variables, such as site area, can be converted to categories to help make the control or adjust decision. Table 4.2 sets out the variable types and for each, the type of control to apply or the way to estimate an adjustment.

Variable Type	Control	Adjust
Categorical – binomial Categorical – 3 or more	Include or exclude Include or exclude each category separately	Paired group comparison Paired group comparison for each category
Measure Convert to categorical	Trim around subject Isolate	Simple regression or Paired group comparison

This control/adjust decision relies on the independence between the four dimensions of similarity. Paired group (or means) comparison works particularly well where there is variable independence. Similarly, simple (two-variable) regression works well when other variables do not confound the analysis – meaning independent variables cannot have dependence or collinearity with each other.

Table 4.3 summarizes the control/adjust decision for each of the four dimensions: transaction, time, space, and utility. The table provides commonly found characteristics for each dimension, the usual variable type, and the statistical principle to be applied for a solution.

Analytical Dimension	Basis of Characterization	Common Characteristic	Variable Type	Statistical Principle
Transaction	Contract terms Rights conveyed Financial Sale Conditions	Fee/Leased	Binary	Estimation
		Lease type (Gr., Net, NNN.)	Categorical/Measure	
		Subsidized govt. financing/ not	Binary	
		Seller carry different rates	Measure	
		Motivation: market/disposal/ liquidation	Categorical	
Market conditions	Time	Price index: trend – straight- line	Measure: linear	Forecasting Ex-ante (up to date of value) Ex-post (after date of value)
		Price index: polynomial fit	Measure: non-linear	
		Price index: event impact	Binary: before/after	
Spatial elements	Location	Neighbourhood: inside/outside	Binary	Estimation (A perceived need for extrapolation usually means going back to the market identification phase)
		Adjacent detriment: yes/no	Binary	
		Distance to city centre	Measure	
		Amenity proximity: next/near/far	Categorical	
Utility	Characteristics			
	Amenity	Condition: Fair/Average/ Good/Excellent	Categorical	Estimation
		Effective age	Measure	
		Rail siding	Binary	
		Highway access in kilometres	Measure	
		Highway access ordinal ranking	Categorical	
	Income	Gross income	Measure	Estimation (A forecast may be implicit, but is seldom explicit, except in yield capitalization methods)
		NOI	Measure	
		Tenant quality: (AAA, AA, A, B)	Categorical	
		Government guarantee: yes/no	Binary	

The order set out in this table coincides with the generally recommended sequence of adjustments for the direct comparison approach. The reasoning for the sequence is that transaction (contract) terms are generally unrelated to characteristics of the property. They have to do with financing, motivation, and the property rights exchanged. Market conditions (time) are seldom related to property characteristics over any relevant time period (but similar properties do tend to trend together). Similarly, location (spatial) relationships tend to keep in step with each other. While utility elements (amenities and income) often have correlation between themselves (like bedrooms and size), they are usually quite independent of the other three dimensions.

Characterization: Variable and Measure Types

We conclude the narrative portion of this lesson with a review of the variable types used to characterize the subject property, and in locating the subject within the identified market.

When characterizing a subject property for analysis, we must distinguish between *qualitative variables* (nominal and ordinal) and *quantitative variables*. Each of these is handled differently in the analysis stage. The following table provides common examples of real estate qualitative and quantitative variables, categorized by property type.

Property Type	Measure Variables (Quantitative)	Categorical Variables (Qualitative)
SFR	Living area Site area View breadth measure Basement area Garage size	Style Pool (Y/N) View (mntns, river, other) Sump (Y/N) Quality (Good, Average, Fair) Patio (Y/N) Garage (single, multi)
Condominium	Parking spaces Homeowners fee Floor level above ground	Type of parking Assessment due (Y/N) Gated (Y/N)
Multifamily	Total living area Mean living area per unit Number of units Expense ratio Rent multiplier Capitalization rate	Covered parking (Y/N) Unit breakdown by br/ba Condominium quality (Good, Average, Fair) Condition (Good, Average, Fair) Rent control (Y/N) Age restrictions (Y/N)
Development land	Acres Buildable units Density	Zoning Flood zone category
Industrial user (owner-occupant)	Building size Office size Office percentage Clear height Door height	Rail access (Y/N) Dock high (Y/N) Clean area (Y/N) Loft (Y/N) Paving type

Categorical (qualitative) variables lend themselves to the control/adjust decision. In contrast, a measure variable must first be considered for usefulness as a unit of measure, then possibly considered for simple regression (if adequate controls are possible), or potentially transformed into a grouping so that it may be treated as a categorical variable.

Case Studies: Market Identification and Property Characterization

We now bring together the fundamental components of the valuation problem: defining the dataset and applying the statistical and analytical tools of market identification and property characterization. These tools will be illustrated in the following case studies.

In reviewing the following cases, you should assume the client has provided all the information necessary for problem identification. The scope of work has been established. An initial "best use" decision has already been made, based on this scope of work. The focus of our examination will be on the techniques of market identification and property characterization, incorporating statistical analysis and illustrating how computer applications may aid in this.

In working through these case studies, you will notice that we follow the valuation problem checklist. Again, this is a handy way to ensure your approach is logical and comprehensive. You may wish to review the *Appraisal of Real Estate*, pages 13.5-11 before starting the case studies (available under "Online Readings" on your course webpage).

Case Study 1: Determining Optimal Unit of Comparison

You have been commissioned to provide an appraisal for a house in Coronado Village, an island beach community of high-priced homes (about 9,500 dwelling units) in San Diego, California. The appraisal will be used in support of a mortgage loan application. The key question in this case is to determine the best unit of measure to use to complete this assignment. In essence, you are looking for the **best unit of comparison**.

Your assistant has researched local sales data and placed this into an Excel workbook. You need to edit and analyze the database. You can find the workbook titled "Coronado" under "Online Readings" on the course webpage. We will use Excel for working through the cases as Excel is the common program in use in most valuation settings.⁴ However, you are welcome to use the SPSS files if you prefer. The Excel workbook contains the data, as well as an additional worksheet showing the data reduction (information set). The SPSS files contain only the initial sales data.

The data was downloaded from the local Multiple Listing Service, which is considered reliable in this area, and reflects 90% to 95% coverage of all sales taking place. The data includes six months of sales of single-family residences, in a relatively stable market. The area and sales in the dataset are all homogeneous, with minor differences. None are waterfront or view properties.

⁴ In Excel 2007, we are using the Data Analysis add-in as an easier way to calculate the necessary statistics (not requiring so many formulas to be written). You can find the Data Analysis selection under Excel's Data menu, as long as you have installed the Data Analysis add-in. If this has not been installed, then click on the Office Button (top left-hand corner), click Excel Options (bottom right), click Add-Ins, and from the list select Analysis ToolPak. This will install the Data Analysis menu item under Data. If you do not see Analysis ToolPak on the Add-In list, then you may need to re-install your Excel program from the Microsoft Office CD if you wish to enable this option. You may also use the Excel formulas or SPSS to calculate the necessary statistics.

The data frame appears straightforward:

1. Transaction: All reported sales are fee simple, conventional financing, and typical motivation.
2. Time: Six months data is all that is necessary, as there are plenty of sales.
3. Space: The datasets include all of the island.
4. Utility: Per your instructions, your assistant pulled only SFDs, but included all sales on the island.

The Valuation Problem Checklist below provides the parameters of the valuation problem, as well as the agreed scope of work/reporting for the appraisal assignment.

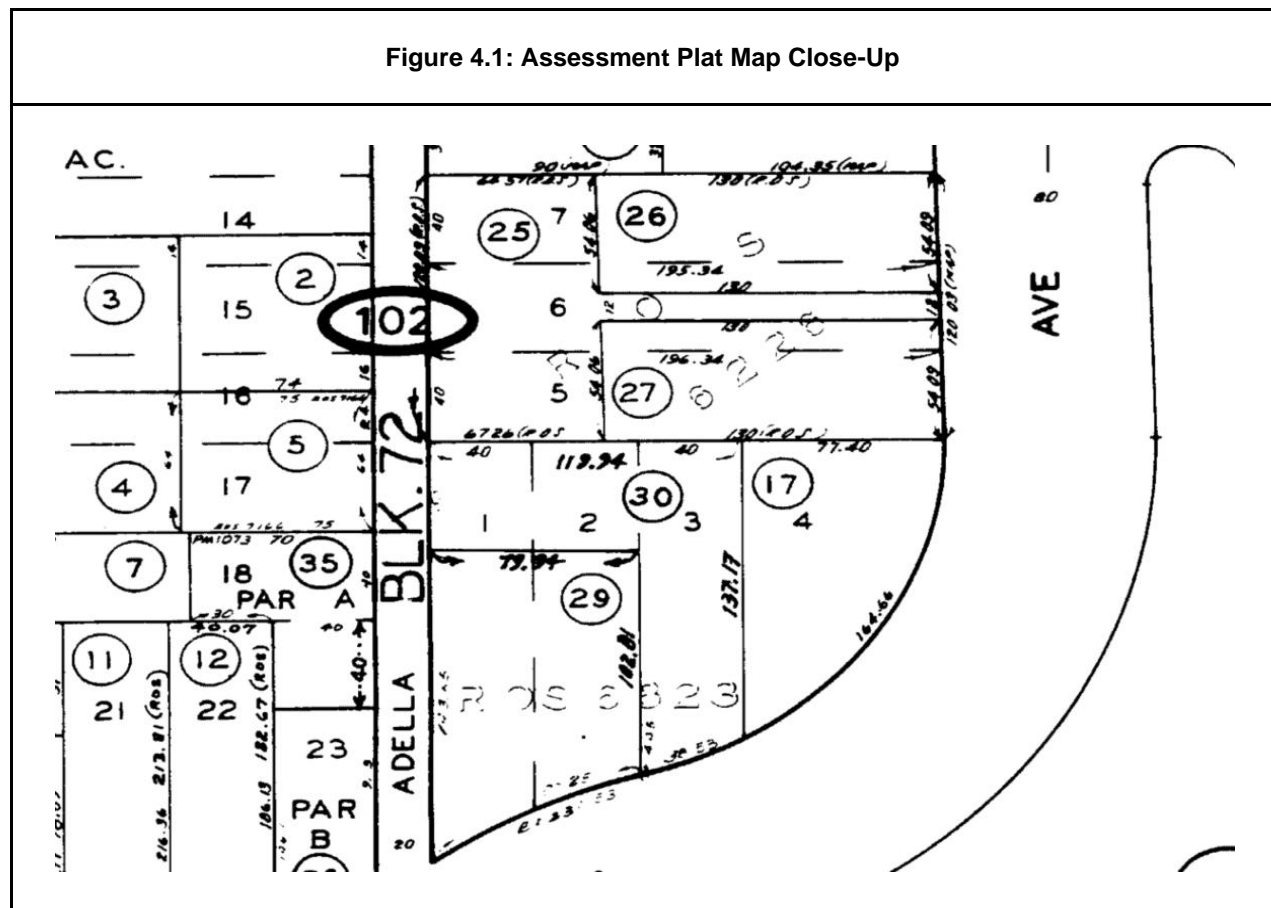
Problem identification		
Array and Breadth as requested by the client	Intended use	Loan security
	Value definition	Market value
	Dates	Current date of value: Nov 25, 2006. Date of report: Nov 25, 2006
	Characteristics	Detached residence; 1,350 SF; site is 3,200 SF; no garage; 2 BR; 1.5 Ba
	Assumptions	Typical
The scope of work: Data		
Array and Breadth, as determined by the appraiser	Data sources: market and subject	Local MLS, public records, appraiser files, interior inspection of the subject property
	Data frame	Coronado Village only, 6 months, SFDs
	Dataset	Same as the data frame
	Information set (market segment)	The focus of this case problem
	Illustrative set	Not a part of this case problem
The scope of work: Analysis		
Complexity, as determined by the appraiser	Unit(s) of comparison	Expected to be either living area or site area
	Statistical tools	Expected tools to be used are correlation comparison (sale price against site area and sale price against living area) and the coefficient of variation
	Graphical tools	
The scope of work: Reporting		
Detail and Emphasis	Intended use	Form report
	Analytical scope	
	Client familiarity	

You are confident the subject is in its highest and best use, based on your inspection of the subject property, its zoning, and surrounding uses.

Your assistant has performed some further research and found the following lot sizes from assessment plat maps:

- Sale 1: odd-shaped, but your assistant says it is about 1,320 square feet. You looked at the parcel map, and agree with your assistant's estimate.
- Sale 14: 60' by 50' Sale 27: 25' by 140'
- Sale 34: L-shaped and your assistant says it looks to be about 6,000 square feet. You decide to check his calculation and edit the dataset as necessary, using your own judgment. In reviewing Figure 1 (Lot 30), you calculate the lot size at over 9,000 square feet. You ask your assistant how he estimated only 6,000 and he explains that its odd shape means it is not fully useable, although it does allow the garage to be off the side street. In his estimation, considering the utility factor, it is about equivalent to a more rectangular lot of about 6,000 square feet. You accept his argument because he looked at the lot, and you did not.

The other lot sizes, living areas, and plat maps were not available. However, you note that the ones which your assistant looked up were unusual sizes that would have affected the analysis because they are influential points, and thus are important to the decision at hand. You compliment your assistant for understanding the importance of researching outliers and influential data points.



You first identify the steps to be followed:

1. Data editing: Review outliers and missing data.
2. Analysis: Run correlations of sale price versus living area, and sale price versus site area. Confirm results by testing the coefficient of variation of price per square foot for both.
3. Arrive at a final dataset (market segment) for this problem. In the next case, we will take this dataset down further to an information set designed for a particular analysis method.

Data Editing

In reviewing the data, you realize there are several sales that are missing measurements for either living area or site area. For these sales, you need to either complete the missing data fields or remove the sales from the analysis. In spite of the work done by your assistant you decide to remove all records with missing lot or living area information. Thus, you have reduced the dataset by 12 records, leaving 29 records. This is the information set worksheet in the Excel workbook.



Helpful Hint!

If you remove the sales, you should move them to a separate area of the datasheet or to another worksheet or file, so that you keep the data for future reference. For example, in an Excel workbook you can have numerous worksheets, so you may find it helpful to keep the original dataset on Sheet1 (perhaps renaming it to Original Data) and then create a new sheet for editing. By documenting your data decisions this way, your work becomes auditable.⁵ Alternatively, (in any software product) you could use the File → Save As function and save a series of files, one for each stage of your analysis naming them appropriately (that is so you can tell them apart and know where they fit into your series). This would allow you to go back to an earlier file if you make a misstep.

Analysis – Run Correlations and Coefficient of Deviations

To run the correlations in Excel, the columns must have the same number of cells, one continuous list and no missing data. You may wish to create a new Excel worksheet for the "information set" for each specific analysis. This makes it easier to manage the data as necessary for any specific question. It also preserves your original data in case you make a mistake.

- Use the CORREL function..., set to =CORREL(column set 1; column set 2).
- In the InformationSet worksheet, move the cursor to cell D38 and type =correl(d2:d30,c2:c30) to get the correlation between sale price (column c) and living area (column d).
- Move to cell E38 and change the d2:d30 to e2:e30 to get the correlation between sale price and site area.

Results:

- Correlation of sale price versus living area: 84.6%
- Correlation of sale price versus site area: 62.1%

Both living area and site area have moderate to strong correlations with sale price, indicating both are reasonable units of comparison. Living area has the higher correlation and appears to be a better choice for further market delineation. We will carry out further testing, comparing the coefficient of variation for the price per square foot for the two variables.

First, we add two columns to the spreadsheet and calculate the price per square foot of site area (SP/SiteSF) and living area (SP/LivSF). Then we add formulas to calculate the standard deviation and mean of each of these columns: e.g., =STDEV(F2:F30) and =AVERAGE(F2:F30). This has been done for you in row 32 and 33 on the InformationSet worksheet.

Standard deviation is not ideal for comparing results between sets of numbers because its size will vary based on the relative size of each of living area and site area. A better measure is the coefficient of variation (COV), because it expresses standard deviation as a percentage of the mean, and thus allows comparisons between variables of different sizes. More information on the COV can be found in Lesson 1.

⁵ While beyond the scope of this course, this is the key advantage of modern data analytical methods of valuation. Traditional appraisal reports are reviewable, but not auditable. Auditability enables the measurement of reliability and quality control, which can be difficult or impossible under traditional methods.

We find the following:

	SP/LivSF	SP/SiteSF
Standard Deviation	258.5	137.5
Mean	945.6	357.1
Coefficient of Variation	0.273394	0.384934

It appears that sale price per square foot of living area is much more tightly distributed, shown by the smaller coefficient of variation. Living area has a higher correlation with sale price and the price per unit variable calculated from these two variables is less dispersed as well, so we will emphasize living area as the preferred unit of comparison for this market segment.

Dataset (Market Segment)

Living area appears to be a better *unit of comparison* to use for further market delineation (i.e., going from the dataset to the information set). We will examine this visually/graphically in Case Study 2.

At this point, we have reduced the data frame down to a dataset, via controls on transaction terms, location, and time. In the next case, we continue this reduction to the information set level, specifically useful for the process of locating the subject within the market.

Case Study 2: Moving from Dataset to Information Set

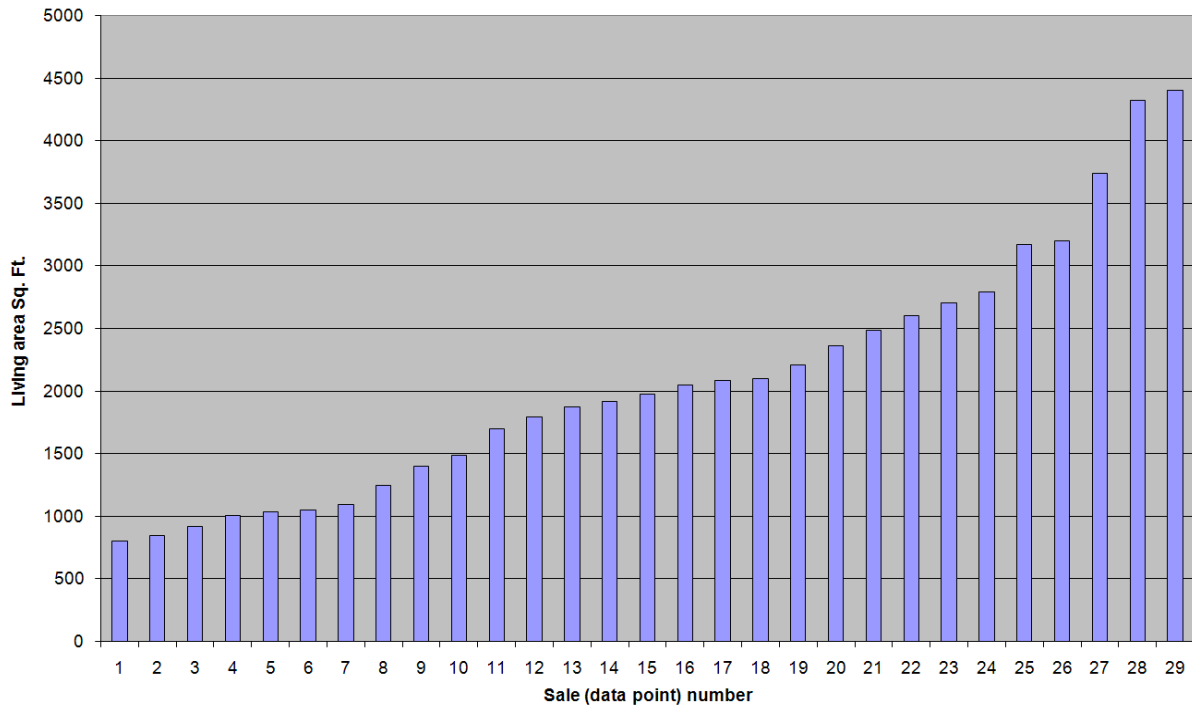
This case continues the Coronado Village study from Case 1. Previously we completed the *problem identification* portion of the Valuation Assignment Profile. We identified the square feet of living area as the best unit of measure to use as our unit of comparison. In this example, we will identify the appropriate *information set*, which is the most competitive market segment in terms of the transaction, time, space, and utility. Such an information set will produce the most reliable basis for performing analysis.

You can find the dataset in a file titled "Coronado2" under "Online Readings" on the course webpage. Again, we will use Excel to go through the case, as this is the common program in use in most valuation settings. However, you are welcome to use the SPSS files if you prefer. The Excel workbook contains the data, as well as additional worksheets showing the data reduction and output. The SPSS files contain only the initial dataset.

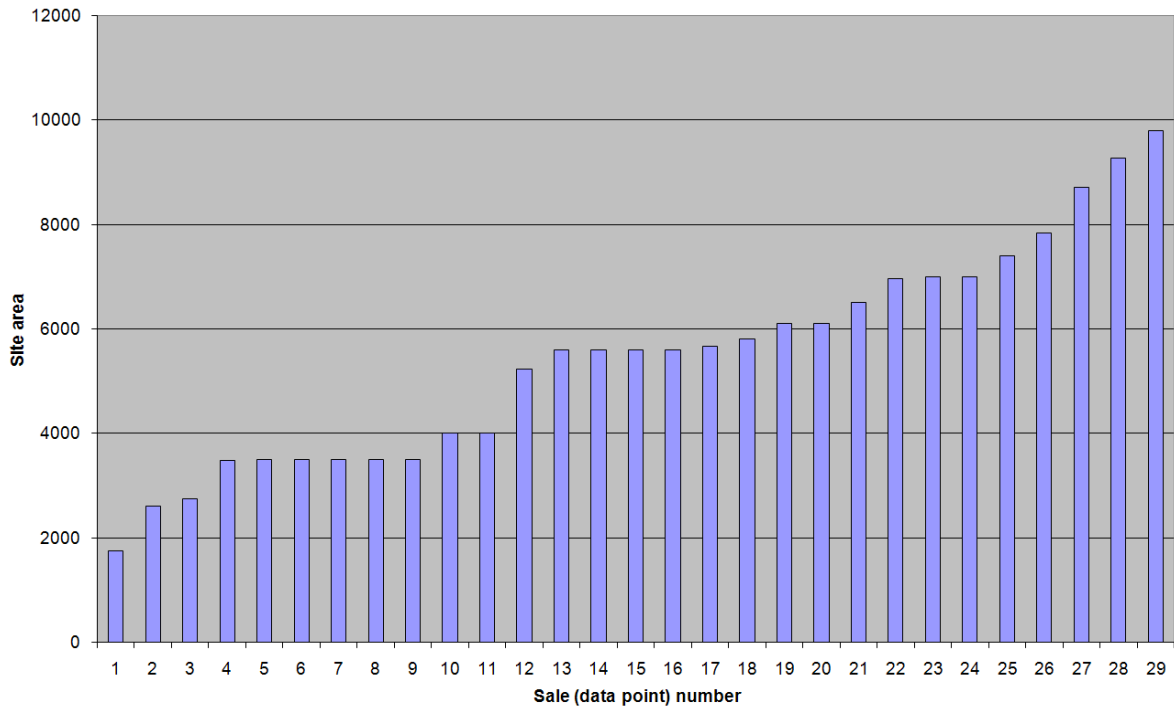
For the data scope, we previously identified our data sources and the level of subject investigation. The frame and dataset include all sales of single-family dwellings (SFDs) in Coronado Village for the past six months. While you normally prefer a longer time frame, your assistant assures you there were plenty of sales during this period.

For the analysis scope, we previously identified living area as the best unit of comparison. Below we have used simple bar graphs to observe the arrays of both living area and site area (the data was sorted prior to preparing each chart).

Living area distribution



Distribution of site area



We now proceed to identify an appropriate information set. As a part of this we will also identify statistical or graphical tools that might assist in locating the subject within this selected market. Keep in mind that such statistical or graphical tools will likely be helpful in communicating results to the client.

While a perfect information set is never achievable, the *near-ideal* information set will have the following features:

- We end up with at least, say, eight to ten sales in the information set.
- Transaction terms (rights, financing, motivation) will be identical or easily adjustable.
- Time will be concurrent or very recent.
- Space will be a single, homogeneous market area.
- Utility characteristics will be identical, similar, or:
 - Bracketed
 - Balanced
 - Randomized

Minimum number of sales. The maxim is always to use all relevant sales in your market segment. However, sometimes there are few sales, leading to a difficult trade-off. More data is always better – except when that data is overly different or otherwise unreliable. The general guideline of eight to ten sales is useful, but must be considered in light of the other desirable features of information sets: bracketing, balance, and randomization.

Bracketing. In general, we would look for at least two or three sales above and below the subject size (in terms of our unit of comparison). Two or fewer will usually mean we need to expand our dataset to include more such sales. We can do this by going back further in time, or expanding geographically.

Balance. The ideal balance would be a similar number of data points above and below the subject characteristics. If the subject is difficult to bracket, you will likely have an unbalanced information set. However, bracketing is usually more important than balance. *Sufficient* sales is another trade-off in this dance of reliability.

Randomization. If we assume for a moment that transaction, time, and space elements are identical, then any other important utility characteristics would be balanced and randomized both above and below the subject value.

In randomization, we are really not controlling at all! We are depending on features and circumstances to randomly fall on either side of the subject, thus averaging out any errors.

Poor randomization might occur by chance. We deal with this by bracketing, balance, and using as many sales as possible. Another way is by observation, looking for features (variables) which might go together (be correlated). If we have an observed variable, we can simply note any correlation by creating a correlation matrix on all variables in our dataset. You may do this in Excel by going to Data → Data Analysis → Correlation⁶ and then selecting all numerical columns in the dataset.

For *unobserved* variables (those not in our dataset), we must depend on our experience and knowledge of economic behaviour and trust that they too are balanced and randomized around our subject.

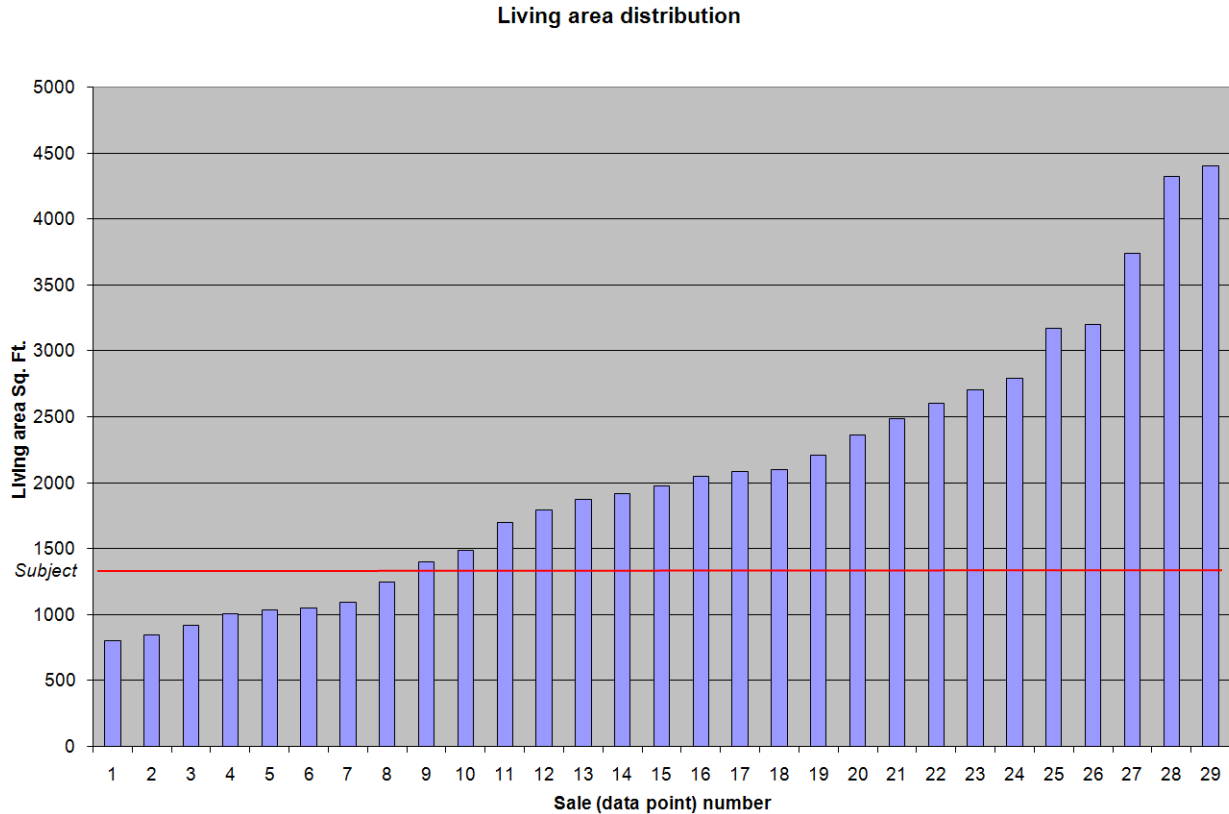
Identify Information Set

For this example, we need to identify an appropriate information set, based on the unit of comparison variable. To do so, we will:

- Locate the subject graphically on the "Living area distribution" graph, by drawing a horizontal line indicating the subject living area.
- Select an information set from the graphed dataset.

⁶ In Excel, the Data Analysis selection under the Data menu requires that you have installed the Data Analysis add-in. Please see Footnote 4 for installation instructions.

In Excel, make a copy of the "Living area distribution" graph on a new worksheet (or simply run the chart again). Click on the chart and click the Layout tab. In the Insert sub-menu select Shapes and under Lines click the single line tool on the left. Use your mouse to insert a horizontal line at the subject's living area of 1,350 square feet. You do not need to be exact. You can make the line thicker or give it a colour of your choice by right clicking on the line and selecting Format Shape....



Based on your choice of best unit of comparison (living area), which of the three choices appears to be best to reduce the *dataset* to your final *information set*? ⁷

- (a) The upper half of the dataset: say sales 15–29
- (b) The middle half of the dataset (25th percentile to 75th percentile): say sales 10–19
- (c) The lower half of the dataset: say sales 1–14

Answer: (c). The subject's living area is 1,350 SF, shown by the line on the graph. It is clear that the subject's competitive property sales lie in the lower half of the dataset. Thus we conclude that the lower half of the dataset comprises the best information set of the three choices given. The upper half is less relevant and likely might skew our results unreasonably. The subject is now nicely bracketed in terms of the unit of measure (living area) and a remaining key variable (living area).

However, reviewing this further, sales 1-18 may be a slightly better choice still based on:

1. More data.
2. The benefits of randomization might also be improved, particularly for any unobserved variables.

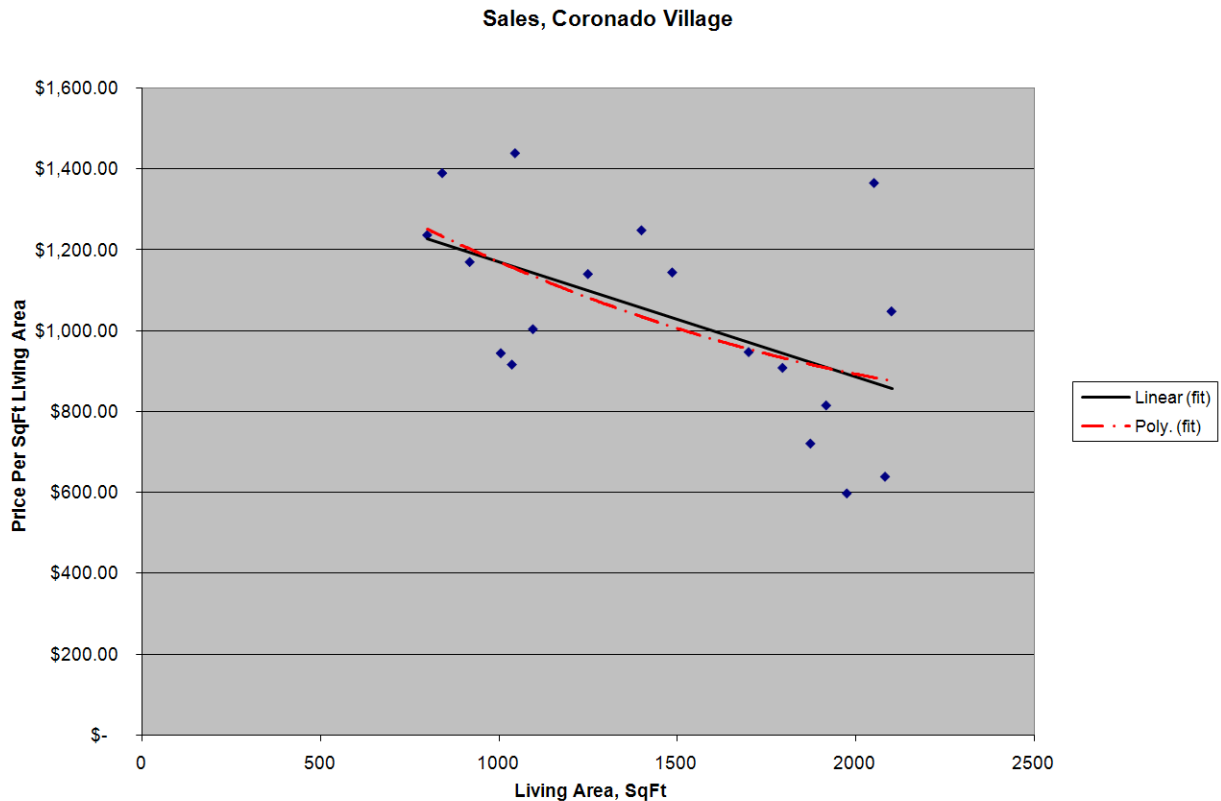
⁷ Note that in this example, you used a continuous measure variable (living area) as a control variable, to reduce the dataset to a competitive market segment of similarly sized living areas.

3. Sales 15-18 appear to all be very similar living areas (slightly above 2,000 square feet). Adding the balance of larger living areas to those below the subject size can easily be justified. Going to 18 sales seems to provide better balance, as well as increasing data quantity, without any apparent loss of reliability.

For purposes of this case, we select the larger information set. We will use sales 1–18 as the information set.

The next step is to view and analyze the sales, based on the living area unit of comparison:

- Copy the "InfoWorksheet" to a new worksheet "InfoWorksheet (2)".
- Now keep only sales 1-18 (rows 2-19), **deleting** everything below (including the "DELETIONS" if you copied those over from the previous worksheet).
- Create a blank column to the right of the "Sold price" column. To do this, right click on the column to the right, then click Insert.
- Copy over the "LivingSqFt" column.
- Create a blank column to the right of this column.
- In the top blank cell, name the column "PriceSFLiving".
- In the next cell below, create the formula for this calculation:
= Sold Price/LivingSqFt [in our spreadsheet = C2/D2]
- Grab the bottom right corner of this cell, while holding down the left button, drag to the bottom of the column. The formula will be copied to all the cells.
- Now create an **XY (Scatter)** graph from those two side-by-side columns. Label the graph properly.
 - i. Select the two columns: LivingSqft and PriceSFLiving
 - ii. Click on the Insert Tab
 - iii. In the Charts sub-menu, Click Scatter, select the chart sample with no lines (top left sample)
 - iv. Right click on the chart, select Move Chart..., select New Sheet, and enter a name for the New Sheet such as Scatter Plot, Click OK. You should now be on the worksheet of the scatter plot
 - v. Click the Layout tab, and using the Labels sub-menu enter a chart title, such as "Sales, Coronado Village"; and titles for the X and Y axes (such as "Living Area Sq/Ft" and "Price per Square Foot of Living Area" respectively).
- Add two trend lines to the graph:
 - i. Click the Layout tab and in the Analysis sub-menu click Trendline, click More Trendline Options...; select Linear, click Close.
 - ii. Click the Layout tab and in the Analysis sub-menu click Trendline, click More Trendline Options...; select Polynomial, Order 2, then click Line Color, select Solid line, and select a colour (other than black), click Line Style, and make the Width 2 points and select a Dash type, click Close. This will differentiate the polynomial curve from the straight line.



The indicated range of price per square foot of living area is \$597 to \$1,438. The subject should fall in this range. Based on the subject living area of 1,350 square feet, a rough estimate appears to be \$1,050 per square foot of living area. You can verify this by drawing a vertical line and horizontal line where they intersect the linear regression of site area against price per square foot of living area. This yields an initial estimate of value for the subject of \$1,417,500.

Narrowing down the dataset based on living area has reduced the variation in sale price and you are now at the starting point for further valuation analysis. The scatterplot shows there is still considerable variation in sale price to be accounted for. You may now use the information set to examine how the subject compares to these sales and determine what adjustments may be necessary. Your goal will be to conclude the most reasonable, supportable, and credible opinion of value.

Case Study 3: Screening Available Sales Comparables

This case problem addresses sale conditions (motivation), an element of the transaction dimension of property value.

You are carrying out an appraisal of a single-family dwelling in a homogeneous area for a lender client. The subject is owner-occupied and is to be valued under normal market value conditions (no atypical motivation).

It is common in appraisal to simply eliminate sales that have atypical sale conditions, such as non-market motivation. This practice assumes you have enough data so that these sales can be eliminated and still permit a reasonable analysis. It also assumes the sales removed and the sales remaining have similar features, because if they do not, then you may be removing a significant value influence and biasing your results. You may wish to keep these in your information set because of sparse data or because one may seem to be your best comparable, identical in all respects, right next door perhaps – the only difference being that it was a motivated sale.

Examining for data problems related to atypical sale condition is difficult, because there is rarely, if ever, a variable that directly measures motivation. The best method for determining this is to interview the participants for each suspect sale, to find out if there were any unusual considerations during the negotiations. However, we will try to evaluate this effect statistically, using a proxy, or indicator variable. An ideal proxy variable has a very high correlation with the variable you wish to evaluate and a low correlation with everything else.

You can find the dataset for this case study in a file titled "Motivation" under "Online Readings" on the course webpage. The dataset has three types of sales: owner-occupied, tenant-occupied, and vacant. The sales distribution is similar otherwise; they are all from the same market area.

We will use vacant sales as a proxy for motivation, using the following logic. We know that vacant properties continue to incur expenses on the owner until eventually sold. These expenses include mortgage payments, utilities, maintenance, security, insurance, and management. Because the owner of a vacant property incurs financial loss every day, without the compensating benefit of occupancy or rent, a quicker sale is preferred. It seems reasonable to conclude that this owner may be more willing to compromise in terms of price, and thus is atypically motivated.

This case examines this *economic* hypothesis, testing it against actual data in an ideal setting, a large homogeneous market area with good data.

Data Frame to Dataset

Missing data: None to deal with.

Outliers: One outlier to deal with: 434 San Jacinto Drive, the sale price is clearly out of reasonable range. Decision is to delete, because of its negligible impact on the measures of central tendency.

Adequacy:	Quantity	Unbalanced, but adequate (more owner sales [63] than tenant sales [9]).
	Quality of the data	Good
	Completeness	Good
	Randomization	Adequate

Market Discrimination

We created an Analysis dataset, with the outlier removed (156 sales). First we sorted the dataset on occupied status: owner, tenant, or vacant. Then we calculated the mean and median sale price of the houses by "occupied" status.

	Means	Medians
Owner-occupied properties	\$446,619	\$440,000
Tenant-occupied properties	\$416,433	\$430,000
Vacant properties	\$419,097	\$417,500

Inclusion/Exclusion Decision

Using vacant properties as a proxy, it appears there is a difference in selling prices between market sales with typical motivation and sales where the purchaser is more highly motivated for disposal. If these sales were to be included in the analysis, a rough approximation of the adjustment for non-normal motivation in the sale is the difference in the mean or median above. The mean difference was \$27,522 and the median \$22,500. Because the median reduces the impact of outliers, it is given greater weight in concluding a \$23,000 adjustment for motivated (or in this case, vacant) sales.

You now have the option of including such sales in the dataset, by adjusting for the motivated seller transaction dimension. If this was the "perfect" comparable, an identical property sold next door a week ago but a motivated sale, you now have a supported adjustment that makes it useable.

This case presented a good example of econometric modeling. It required solid data, economic reasoning and statistical methods to support a mathematical result, resulting in an adjustment for motivation.

Case Study 4: Control or Adjust? Finding the Competitive Market

You recently completed an appraisal of an ocean-view home in Solana Beach for mortgage financing purposes. Your bank client reviewed the appraisal and returned it, stating your value is "way too high" and that you failed to take into account comparable sales that are very near the subject property. You are certain you used the best comparables and would like to rebut the reviewer's contention. Specifically, you wish to argue that the subject has an ocean view, while the closer comparable properties do not. You suggest that including these non-ocean view comparable sales would be inappropriate.

To support this, you decide to present a graph showing the distribution of sale prices in the market area. Your intention is that the graph will show the subject's position in the upper tier of property values, and identify the key variable that is causing the separation in home prices between the upper tier and the lower tier.

The market is level, so you have chosen to examine two years of data, with no price indexing or adjustment necessary. The subject has an ocean and lagoon view. You will compare the subject to property sales that had ocean view mentioned by the listing agent to those with no view (coded "N/K", not known).

You can find the dataset in a file titled "View" under "Online Readings" on the course webpage. In the file, first you will group all the ocean view sales and obtain measures of centre, to compare to the N/Ks. Then you will delete other sales, such as with views of mountains/hills or park-like.

Data Frame to Dataset

Missing data: You note that 16 of the 103 sales have missing living area data and you are concerned this may skew the results. You decide to compare the mean sale prices of sales with living area to sales without living area, in order to decide whether this variable might affect the results. You find a 4% difference in price for sales with no living area reported, and conclude this is not a model-significant problem; however, we will remove those 16 records leaving 87 in the dataset.

Outliers: None

Adequacy:

Quantity	Unbalanced, but adequate (more ocean sales [50] than N/K sales[19]).
Quality of the data	Good
Completeness	Good
Randomization	Adequate

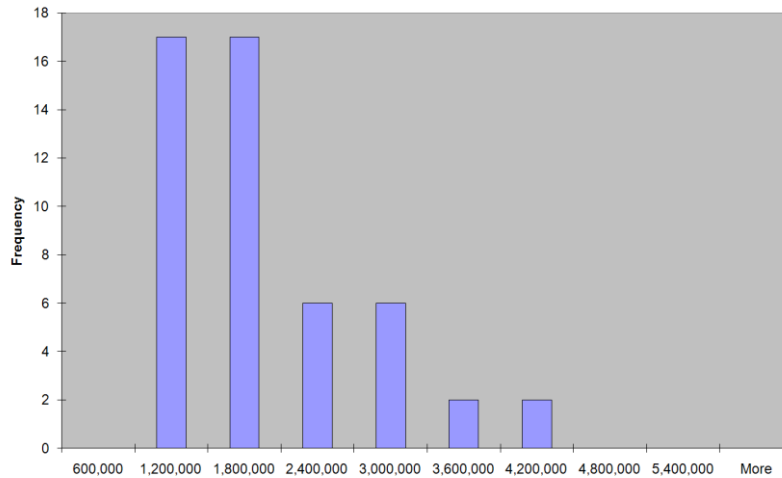
Market Discrimination

First, we sorted and grouped all sales with view ocean and N/K. Then we calculated the mean and median for both (see worksheet: 4 Analysis – ocean versus NK view).

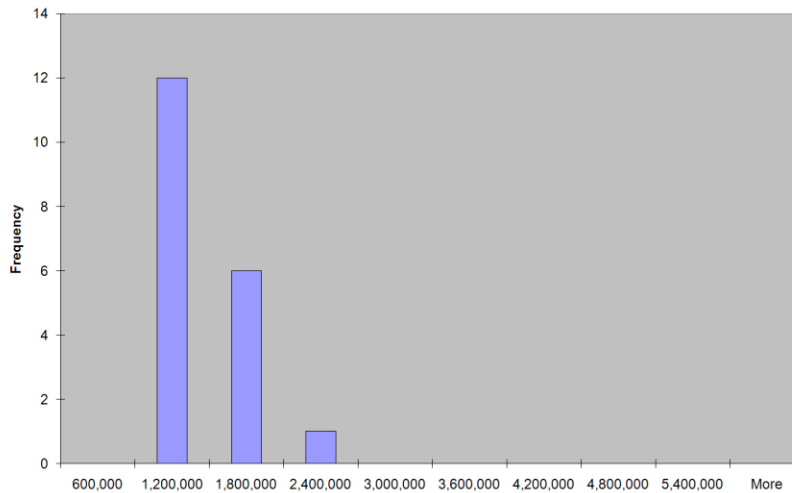
	Means	Medians
Ocean view properties	\$1,690,111	\$1,350,000
No-view properties	\$1,239,568	\$1,100,000

Second, we created a one-variable histogram of sales price groupings, for view and non-view. The distributions are both right-skewed, but the ocean view is much more spread out.

Reported Selling Prices: Ocean View



Reported Sale Prices: Unknown View



Inclusion/Exclusion Decision

There is clearly a difference in selling prices for ocean view properties compared to non-ocean view. The percentage difference between the means is 36% and the medians is 23%. This appears to support your decision to exclude non-ocean view comparables from your original appraisal.

Reconciliation and Interpretation

If you wanted to include non-ocean view sales in your analysis, then you would need to adjust those sales for view. The differences between the mean and the median of ocean-view and no-view properties is a good indicator of this adjustment:

Difference in means:	\$450,543
Difference in medians:	\$250,000

Real estate sale price distributions tend to be skewed to the right, with the mean to the right of (higher than) the median. Using the median reduces the impact of the highest priced-properties.

If the subject is close to the median in terms of living area and has an "average" ocean view, then a reasonable adjustment for view over an otherwise similar sale with no view would be \$250,000. The mean reflects the value of the highest-priced properties, which presumably have the best views, so the best view in the area might be valued at approximately \$450,000.

Case Study 5: Identifying a Similar Market Area

In this case, we will examine the possibilities for expanding the market area in an appraisal in order to increase the data available for comparison and analysis. You can find the dataset in a file titled "Boxcity" under "Online Readings" on the course webpage.

You are an appraiser valuing an office building in Box City, a distinct suburb city approximately 25 kilometres from the city centre. The subject building is 25 years old and has 30,000 square feet of net rentable area, it is located on a 2.5 acre lot.

You have applied the income approach and concluded a value estimate of \$175 per square foot of net rentable area. You want to validate this value estimate using the direct comparison approach, comparing the per square foot indications from recent similar sales.

Your research of the Box City office market has found six recent sales of comparable office buildings. However, you are concerned that these sales are somewhat dissimilar to the subject and perhaps are not optimal comparables. You believe that office buildings in Missing Valley, approximately ten kilometres to the west, may be more similar and may better bracket the subject value. You need to confirm if expanding the market area for comparables is reasonable and justified.

Data Frame to Dataset

Missing data:	All data is available and reported.	
Outliers:	Two Missing Valley sales have prices that appear substantially greater than the rest, one in particular, and two appear somewhat less than the rest. The impact of these potential outliers will be investigated.	
Adequacy:	Quantity	Good
	Quality of the data	Good
	Completeness	Good
	Randomization	Fair, but all that is available

Market Discrimination

First, we will calculate measures of centre for price per rentable square feet for each area (Price Per SF(Net) in column AM).

	Mean	Median
Missing Valley	\$163.96	\$169.64
Box City	\$125.72	\$120.21

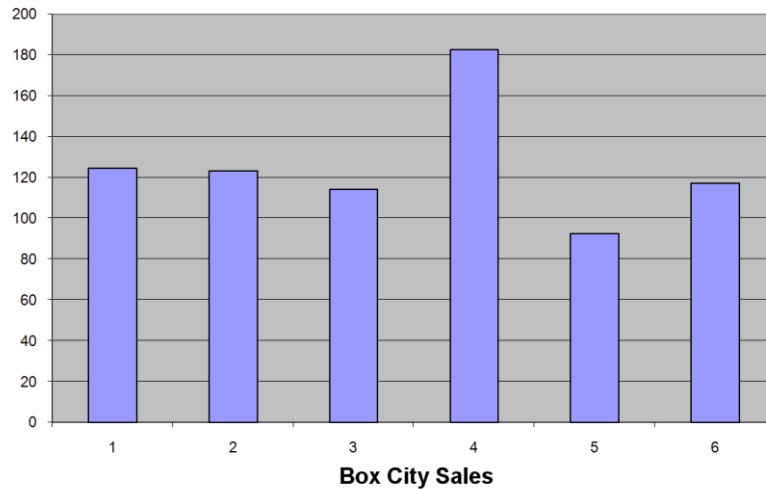
We tested the impact of the outliers by excluding them and recalculating the Missing Valley statistics. They had little effect on the mean and median and thus do not need to be removed.

Second, we will create two graphs, one for Missing Valley sales and one for Box City sales. The dependent (Y) variable is price per rentable square feet for each sale.

Price per Square Foot



Price per Square Foot



These results indicate the Missing Valley sales have a significantly different price per square foot of rentable area than most of the sales in Box City. The subject's \$175 per square foot estimated by the income approach is considerably greater than all but one of the values in Box City. If the analysis is limited to the Box City market, the subject's price per rentable square foot must be extrapolated beyond the market indications. However, if the dataset can be expanded to include Missing Valley, the subject will be bracketed, and the indicated price per square foot supported.

Further exploration of the database indicates some key differences in attributes between Box City and Missing Valley properties, with the subject better represented by those in Missing Valley. In comparison to the subject, the Box City properties on average have smaller lots, newer buildings, smaller net rentable areas, and slightly poorer building quality and condition. The Missing Valley properties appear to better match the subject's key attributes.

		Acres	Age	Sq.Ft.
MISSING VALLEY	Mean	3.0	25.8	47,633
	Median	2.3	25.0	20,988
BOX CITY	Mean	0.7	22.3	11,696
	Median	0.5	19.0	11,485
SUBJECT		2.5	25.0	30,000

Inclusion/Exclusion Decision

Widening the market area appears justified here. Because the subject is dissimilar from its most localized comparables, its indicated value is not bracketed. By expanding the market to a more regional dataset, we better match the subject's competitive set. This is a common issue in commercial property valuation, where the competition for investment funds often extends beyond neighbourhoods into wider regional, provincial, or even national markets. In cases such as these, we are justified in relaxing our requirement for locational similarity in favour of introducing properties with more similar attributes.

Case Study 6: Price-Indexing a Market Segment

You have been hired to appraise a house in a large homogeneous market area with plenty of sales. As a part of this assignment, you want to perform a trend analysis on the market segment. You will use the price index to establish time adjustments for the direct comparison approach.

The database for this case can be downloaded from the course website under "Online Readings". The database is titled "priceindex". The table below provides a summary of the assignment.

Valuation Assignment Profile Price Index		
Problem identification		
Array and Breadth, as requested by the client	Intended use	Loan security
	Value definition	Market value
	Dates	Current
	Characteristics	Property type: SFD, 3,042 sq. ft., otherwise typical for this homogenous area
	Assumptions	Typical
The scope of work: Data		
Array and Breadth, as determined by the appraiser	Data sources: market and subject	Local multiple, public records, appraiser files
	Data frame	2 postal zones, 18 months of sales, living area from 2,000 to 4,330 sq. ft.
	Dataset	Same as the data frame
	Information set (market segment)	Same as the data frame
	Illustrative set	Not part of this case example
The scope of work: Analysis		
Complexity, as determined by the appraiser	Unit(s) of comparison	Sq. ft. living area
	Statistical tools	Simple (2-variable) regression Scatterplot, linear and polynomial trend lines
	Graphical tools	
The scope of work: Reporting		
Detail and Emphasis	Intended use	Not part of this case example
	Analytical scope	
	Client familiarity	

Price-Indexing Steps

Our goal in this assignment will be to support a time adjustment using linear regression, highlighting market trends. We will carry out the following five steps:

1. Obtain and prepare the data
2. Graph and visualize the data
3. Estimate the trend
4. Reconcile and conclude
5. Document and report

Step 1: Obtain and prepare the data

- a. Identify the market segment: area, time, and key characteristics
- b. Download and/or enter data
- c. Control on the key variable (only review data in the range of 2,000–4,330 sq. ft. living area)
- d. Arrange columns (in Excel) for ease of use (ensure ClosedDate and SalePrice are side-by-side)

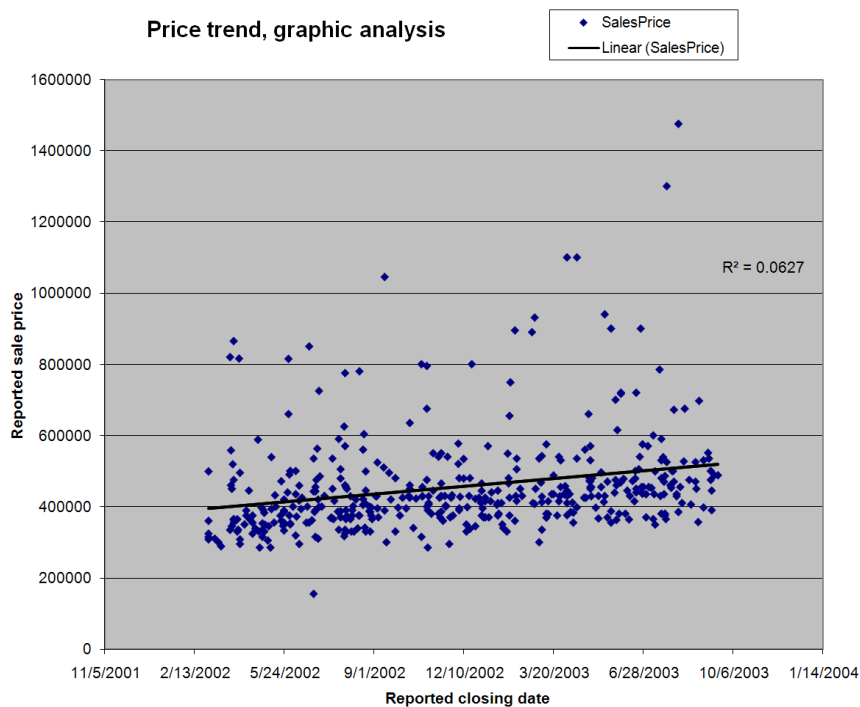
For this problem, your assistant has already prepared the information set, based on your instructions. Because you are familiar with SFDs in this neighbourhood, you know that the best unit of comparison is living area (in square feet). Your assistant has controlled around the subjects' living area of 3,042 sq. ft. by only reviewing data in the range of 2,000–4,330 sq. ft. living area.

Step 2: Graph and visualize the data ("let the data speak to you")

- a. Create a scatterplot. Excel 2007 instructions are as follows:
 - i. Select the two columns: SalesPrice and ClosedDate⁸
 - ii. Click on the Insert Tab
 - iii. In the Charts sub-menu, Click Scatter, select the chart sample with no lines (top left sample)
 - iv. Right click on the chart, select Move Chart..., select New Sheet, and enter a name for the New Sheet such as Scatter Plot, Click OK. You should now be on the worksheet of the scatter plot
 - v. Click the Layout tab, and using the Labels sub-menu enter a chart title, such as "Price Index and Market Analysis"; and titles for the X and Y axes (such as "Reported Closing Date" and "Reported Sale Price" respectively. *Note:* You are using the word "Reported" because you have not individually confirmed these sales.
- b. "See" the market, revise the dataset if necessary, document any edits, enhancements, or deletions of data (there should be none necessary in this problem). Take a moment to consider what you see here.
- c. Add linear and polynomial trendlines:
 - i. Click the Layout tab and under the Analysis sub-menu click Trendline, click More Trendline Options...
 - ii. Here you can add a Linear trendline and its R-squared value, as well as Polynomial trendlines. Try each level of order allowed, 2 through 6. If you wish, you can give each a different colour by right-clicking on the line, selecting Format Trendline... and selecting a colour (using the Solid Line choice under Line Color).

⁸ You will notice there are two closed date variables in this dataset, ClosedDate and ClosedDate(code). The date format for ClosedDate is only useful in Excel's Data Analysis module; for regular Excel commands or for SPSS, this date format is not understood. To address this, we have also included the ClosedDate(code) variable, which expresses ClosedDate in a numerical code understood by all of these programs (we created this variable in Excel by selecting ClosedDate and then Format → Cells → Number → General). You may substitute the ClosedDate(code) variable for these procedures when using any program other than Excel Data Analysis.

When you change the polynomial order from 2 to 6, this revises the shape and ends of each curve. However, you will notice that all the polynomial lines tend to stay close to the linear one. This is because of the large amount of data in this problem. The end-points tend to "flop" because there is no dampening effect of data beyond the end. For higher polynomial orders, the end points are almost totally controlled by the last two or three data points.



You will also note there may be two or three possible outliers. We will examine the impact of these in the next section.

Step 3: Estimate the trend (bivariate regression)

- a. Bivariate regression, \$/day indication of trend
- b. Polynomial best fit, for changes in the trend

In this problem, the polynomial lines tend to lie on top of the linear trend line. This is a good sign that the trend is linear over the last 12 months of the analysis and appears to be continuing at the same rate.

Open the Information Set worksheet (full dataset provided in this example). On the Excel menu, select Data → Data Analysis, and select Regression from the Data Analysis window.⁹ For Input Y Range, select the numbers in the SalesPrice column (d2:d430). To do this quickly, you may select the top number in the column, then Ctrl-Shift-Downarrow. The selection should automatically select the numbers down to the last, in row 430 (d2:d430). Repeat this selection method for ClosedDate for the Input X Range. Click OK to run the regression.¹⁰

⁹ In Excel, the Data Analysis selection under the Data menu requires that you have installed the Data Analysis add-in. Please see Footnote 4 for installation instructions.

¹⁰ To run this regression in SPSS, first open princeindex.sav in SPSS, then use the following steps, Click Analyze → Regression → Linear..., put SalesPrice in the Dependent Variable box, put Closed Date (code) in the Independent(s) box; Click OK. The regression output will be appear in the SPSS Output file.

In the standard Excel output for the regression function, the only number which we really care about is the coefficient for the X Variable, shown below as 220, or \$220 dollars per day. This is our market conditions adjustment for this subject property, at this time, for this market area. To convert this into percentage terms, we would divide this by the mean sale price of \$456,234, resulting in a time adjustment of 0.048% per day or approximately 1.45% per month (non-compounded).

Summary Output

<i>Regression Statistics</i>		Time-series regression						
Multiple R	0.250495015	Zip 91910, 91911						
R Square	0.062747753	March 1, 2002 - current						
Adjusted R Square	0.060552782	Age 1990 or before						
Standard Error	138580.5688	Living area 2000 - 5000 SqFt						
Observations	429							
ANOVA								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	1	5.49E+11	5.49E+11	28.58706	1.46E-07			
Residual	427	8.2E+12	1.92E+10					
Total	428	8.75E+12						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	-7820546.24	1548035	-5.05192	6.5E-07	-1.1E+07	-4777828	-1.1E+07	-4777828
X Variable 1	\$220	41.17971	5.346687	1.46E-07	139.2348	301.1152	139.2348	301.1152

Now we will examine the impact of removing the two high outliers, sale prices \$1,300,000 and \$1,475,000. The slope of the regression line drops to \$182, indicating a revised adjustment of 0.040% per day or approximately 1.21% per month (non-compounded). These outliers appear to have had a significant effect on the results. If they had been in the middle of the data array (earlier in time), their effect on the slope would have been reduced. But because they were among the most recent sales, they were very influential on the regression results, effectively pulling the regression line upward. The results with these outliers removed are more reflective of the effect of market movement on typical properties in this market.

Step 4: Reconcile and conclude (market conditions adjustment)

You have plentiful data in a homogeneous market and the trend appears quite linear and unchanging near the end of the relevant time period. Therefore, your reconciliation is simple, reflecting these two facts.

Step 5: Document and report

You might include the graph and regression results in your written report. The time adjustment of +1.2% per month should be applied to each comparable for the change in market conditions.

You have successfully price-indexed the market segment with its own price level adjustment. Given the data available, this is an optimal market conditions adjustment.

In closing, remember that the depth of scope and detail you go into in this part of your analysis depends on:

- the challenges of the particular assignment;
- any agreement you may have with your client; and/or
- your desire to produce a superior appraisal product.

Summary: Market Identification and Property Characterization

This lesson built on the approach to exploratory data analysis introduced in Lesson 3, attempting to foster a better understanding of the complex interplay of property characteristics and market identification. We provided a practical checklist approach for integration of appraisal procedures with simple statistical tools.

In this lesson, we:

- Identified the main steps in defining an appraisal problem, scope of work, and selecting the correct market segment.
- Described the four dimensions of similarity and their role in market identification.
- Identified the best variable for use as a unit of comparison.
- Applied visual/statistical methods to discriminate between markets.
- Related property characterization to the subject's relevant market.
- Identified three asset analysis components: problem identification, data collection, and estimation.
- Integrated the role of statistics in the asset analysis problem.
- Explored the four principles of experimental design: control, randomization, replication, and blocking.
- Examined methods of data reduction for market datasets and analytical information sets.
- Summarized analytical methods by asset dimensionality and variable types.
- Examined several property types and commonly used quantitative and qualitative variables.
- Applied the two statistical tools, paired group comparison and simple regression, in practical scenarios.

The six case studies covered in this lesson are summarized below:

- Case Study 1: Using correlation and the coefficient of variation to determine the best unit of comparison.
- Case Study 2: Using bracketing and trend line analysis to establish a baseline value.
- Case Study 3: Using logic and a proxy variable to measure the adjustment needed for sales which were motivated (minimized time on the market).
- Case Study 4: Using paired group analysis to determine the impact of ocean view over no view.
- Case Study 5: Using paired group analysis to support expanding the comparable market set.
- Case Study 6: Using simple regression and outlier elimination to determine a time adjustment factor.

Review and Discussion Questions

1. Consider the four dimensions of similarity. How might you classify each of the elements of the checklist relative to the home or apartment building in which you live.
2. In Case Study 3, we examined the potential motivation related to vacant property sales. In our analysis, we also calculated the mean and median sale prices for tenant-occupied property sales. Do the reported means and medians for tenant-occupied properties make economic sense, as they relate to the other two values?
3. Two reasons were discussed for why an analyst might need to go back to expand an original dataset or even widen the data frame. One reason was simply that more data was needed. What was the other major reason and why is it important?
4.
 - (a) What are the four dimensions of comparability (similarity) and why are they important to identifying the right market?
 - (b) The transaction dimension has three elements. How do these differ substantially from the other three dimensions?
 - (c) The utility dimension encompasses the two fundamental types of property benefits, amenities and income. What are the three forms of utility?
5. What are the five stages of data reduction for our purposes?
6. While neighbourhoods and districts are important to real estate economics, what economic division is the one we primarily rely on for defining our datasets? How is it defined?
7. Name two major types of changes in market conditions (time)?
8.
 - (a) How is information "better" than data?
 - (b) How is an information set different from the market dataset?
9. What statistic is most useful in identifying the best unit of measure?
10. Why do you save sales deleted from your dataset in a separate file?
11. Appraisal has historically used less intensive data analysis methods. Why? How has this changed today?
12. What is the difference between "reliability" and "credibility"?
13. Multiple regression analysis (MRA) is a powerful tool. Why don't we just use MRA to solve all valuation problems?
14. How is similarity identified and measured?
15. You have started an internship with an older well-respected appraiser. Everyone tells you that you are lucky to get a position with someone of his stature. You are ready to begin applying your modern tools. You get your first assignment, and enthusiastically download what you believe is a good market dataset. Being your first assignment, you have your new mentor check your search parameters. He laughs, and says: "All that statistical stuff is OK for school, but a good appraiser never needs to use more than three 'comps'! Just pick the best three and use those". What can you say?
16. What are the four appraisal principles of experimental design? Provide a one-sentence explanation of each.

17.
 - (a) List the sequence of five data group subsets we have learned regarding data reduction.
 - (b) Which of the above sets is most similar to the traditional three comparable sales?

18.
 - (a) Which is better as a unit of comparison, continuous variables or categorical variables?
 - (b) Which is better for control and blocking, continuous variables or categorical variables?

19. You have applied graphs and simple regression in an appraisal. You receive a scathing, angry review, highly critical of your methods, stating: "You can prove anything with statistics. These are not approved methods and were not taught in the appraisal courses I took. I want real support for your adjustments, like actual pairs". What can you say?

ASSIGNMENT 4

LESSON 4: Market Identification and Property Characterization

Marks: 1 mark per question.

1. The four dimensions of property similarity:
 - (1) enable determination of scope.
 - (2) often have high correlation within each dimension.
 - (3) generally have high correlation across dimensions.
 - (4) are all variants on locational advantage.

2. Which one of the following is described in this lesson as "the biggest appraisal mistake"?
 - (1) Not measuring the property correctly.
 - (2) Errors in identifying transaction elements.
 - (3) Identifying the wrong market segment.
 - (4) Applying an adjustment the wrong way.

3. Which one of the following is described in this lesson as the "second biggest mistake" in appraisal?
 - (1) Not measuring the property correctly.
 - (2) Errors in identifying transaction elements.
 - (3) Identifying the wrong market segment.
 - (4) Applying an adjustment the wrong way.

4. What must be considered in identifying similarity?
 - (1) The four dimensions of comparison.
 - (2) Location and subject characteristics.
 - (3) Transaction elements and subject characteristics.
 - (4) Utility and age of the property.

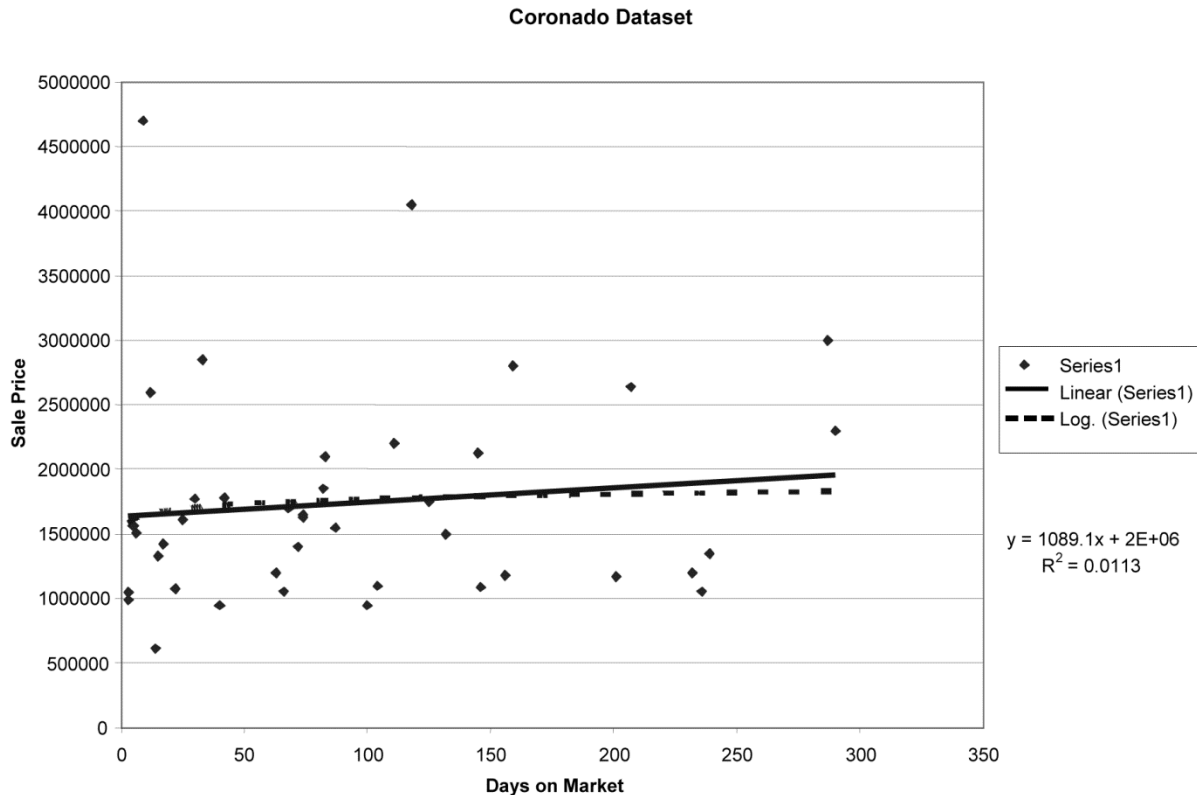
5. In any given project, the depth of scope and detail of your analysis will depend on:
 - (1) the challenges of the particular assignment.
 - (2) any agreement you may have with your client.
 - (3) your desire to produce a superior work product.
 - (4) all of the above.

6. The appraisal process is designed to:
- (1) validate sale prices in a fair market.
 - (2) make data meaningful for decision-making.
 - (3) clarify scope of work to redirect client problems.
 - (4) turn good information into data explanations.
7. Property characterization:
- (1) is the start of the appraisal process.
 - (2) provides an alternative to finding the right market.
 - (3) may cause you to return to your market identification decision.
 - (4) can only be valid with full and complete hypothesis testing on all variables of interest.
8. Credibility:
- (1) means having a high reliability and low correlation among dimensions.
 - (2) involves getting the model right, as well as the right scope of work.
 - (3) depends on competence and good communication.
 - (4) means freedom from random error.
9. Which of the following is correct regarding evaluation of market conditions?
- (1) Trends can be evaluated using a price index or scatter graph.
 - (2) Events are best analyzed by paired group methods alone.
 - (3) The closure of a major employer in a market area is an example necessitates a trend impact analysis.
 - (4) A steady market increase necessitates an event impact analysis.
10. The scope of the research element is composed of four fundamental dimensions. Which of the following is NOT one of these four dimensions?
- (1) Transaction conditions
 - (2) Range of characteristics
 - (3) Space
 - (4) Complexity
11. The four principles of valuation experimental design are:
- (1) control, randomization, replication, and blocking.
 - (2) balance, bracketing, replication, and control.
 - (3) control, randomization, balance, and blocking.
 - (4) dimensions, bracketing, estimation, and control.

12. Replication means:
- (1) other similar experiments should produce similar results.
 - (2) "averaging out" the effects of uncontrollable sources of variation.
 - (3) you can reply to challenges with reproducible data.
 - (4) taking advantage of natural grouping of similar variables.
13. The difference between a market dataset and an information set is:
- (1) control, randomization, replication, and blocking.
 - (2) that some data must be added due to presence of outliers.
 - (3) the degree of iteration necessary.
 - (4) that an information set is designed for use with a specific analysis method.
14. In Case Study 3, the indicated adjustment to apply to comparables to account for atypical motivation was:
- (1) decrease comparable sale price by \$417,500.
 - (2) increase comparable sale price by \$23,000.
 - (3) increase comparable sale price by 17%.
 - (4) decrease comparable sale price by 27%.
15. In Case Study 4, the indicated adjustment to apply to comparables with no view was:
- (1) increase comparable sale price by \$250,000.
 - (2) decrease comparable sale price by \$450,000.
 - (3) increase comparable sale price by 8%.
 - (4) decrease comparable sale price by 5.2%.
16. Which of the following is/are TRUE regarding Case Study 5?
- (1) Missing Valley sales are similar enough to the subject's characteristics that they can effectively be considered as a similar market segment, and used in further valuation analysis.
 - (2) The subject's immediate geographic area does not offer sufficient good comparables, given the subject's uniqueness in age and size.
 - (3) Comparison of means and graphs for prices between the two markets verifies the subject is more similar to Missing Valley properties, in spite of location differences.
 - (4) All of the statements are true.

17. Consider the following graph from the "Coronado" dataset, of "Mkt Time" by "Sold Price". Which of the following statements is TRUE?

- (1) Days on market and sale price appear to be strongly correlated.
- (2) The dotted logarithmic trendline (order = 2) indicates a very non-linear relationship.
- (3) The outliers appear to have skewed the analysis, rendering conclusions invalid.
- (4) None of the above is true; all of the statements are false.



18. You are interested in knowing more about the relationship between sale prices and listing prices in the "Coronado2" Dataset (the entire unedited dataset). Which of the following statements is TRUE?

- (1) The correlation of 55% shows these variables are at best moderately related.
- (2) The X-Y scatterplot with a polynomial trendline (order = 2) is curved, indicating a probable non-linear relationship.
- (3) The difference in means of approximately \$370,000 indicates an average reduction of 20% between list price and sale price.
- (4) The correlation of 99% shows these variables are strongly related.

THE FOLLOWING TWO QUESTIONS RELATE TO THE BUSI 344 PROJECT 1 DATABASE CALLED "CONDOSALES". PLEASE DOWNLOAD THE PROJECT 1 FILES FROM THE ONLINE READINGS ON THE COURSE RESOURCES WEBPAGE.

19. Run a correlation matrix for the following variables: sale price, sale price per square foot, sale price per bedroom, size, bedrooms, bathrooms, competitive set rank, and location quality rank. Which of the following statements is FALSE?
- (1) The correlations are greater than 0.7 between each pair of the size, bedrooms, and bathrooms variables
 - (2) The competitive set rank is more highly correlated with the sale price per square foot than it is with the sale price.
 - (3) The location quality rank is most correlated with the sale price.
 - (4) The sale price per bedroom is most correlated with the competitive set rank.
20. Run a paired group analysis using a boxplot and/or compare means table of Sale Price versus Secure Parking. Which of the following statements is TRUE?
- (1) The existence of secure parking has an impact on the selling price of the condominiums in the database.
 - (2) To use the sales that do not have secure parking in the analysis an adjustment would have to be made to those sales.
 - (3) The sales with secure parking and the sales without secure parking represent different markets in this data.
 - (4) All statements are TRUE.

20 Total Marks



Planning Ahead

Project 1 is based on the information from Lessons 1-5, in particular Lessons 4 and 5. Project 1 is due in three weeks (with Assignment 7). You should begin working on this well in advance, to ensure you have sufficient time to complete it satisfactorily.

