Handout 1.1: Introduction to Modern Regression Methods

Dr. Bean - Stat 5100

1 About me

- Graduated from BYU-Idaho in 2014 with a Bachelors of Science in Applied Mathematics.
- Graduated from Utah State University in 2019 with a PhD in Mathematical Sciences.
- Current interests include basketball, cross country skiing, hiking and spending time with my wife and daughter.

(Groups) What is a creative, yet appropriate, question that you have about the life/career of the instructor?

2 Why Modern Regression Methods?

Statistics, in the words of Dr. Bin Yu, is the "science that solves data problems." This science becomes more and more relevant in a world inundated with data. From the late Leo Breiman:

> The uses of statistics pervade our society. They are used and terribly misused all through the social sciences and health fields. ... It is surprising how much the world around us depends on the use of statistics. ... It's odd that even though the articles

involving statistics in the newspapers far outnumber those involving say, physics or chemistry, people in general know very little about what we do.

In this class, we will learn several of the foundational approaches for using data to make **predictions**. Perhaps more importantly, we will discuss the **cautions** we must consider when using and interpreting model output.

(Groups) Why are YOU taking this course?

3 Functional vs Statistical Modeling

We learn about functions in Math 1050 (College Algebra), a functional model takes a set of inputs X and produces a (set of) outputs Y, i.e.

$$Y = f(X)$$

Example: You write a function to model the profits from your lemonade stand. You rent the stand for \$200 a month and sell each glass of lemonade for \$1.00. If it costs you \$0.25 to make the lemonade then your monthly profits Y could be modeled as a function of the number of lemonade glasses you sell x

$$Y = 0.75x - 200.$$

The key to a functional model is that each input x produces a **unique** output Y.

In a **statistical model** we assume that the values of Y can be modeled by a function *plus* some "random noise" ϵ. The presence of the ϵ term allows for many different values of Y for the same set in inputs x.

$$Y = f(X) + \epsilon$$

Example: The relationship between ground snow and elevation in Utah (see figure 1).

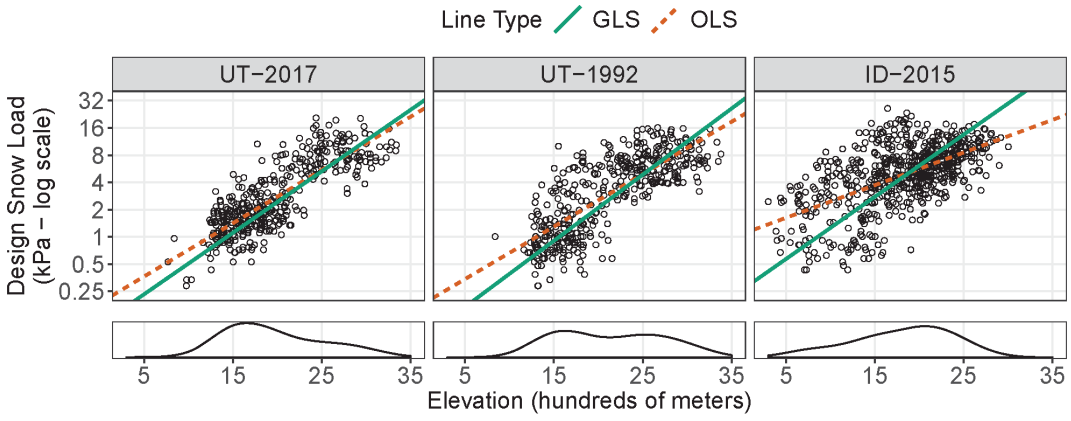


Figure 1: Plot of design ground snow loads (log-scale) vs elevation in and near Utah.

4 The Key Assumption (why this class exists)

The key assumption (and the foundation for this course) we make is that ϵ follows a probability distribution. Specifically we assume that

$$\epsilon \sim^{i.i.d} N(0, \sigma^2). \tag{1}$$

If this assumption is valid, then our **estimates** of the **model parameters** will come from well-defined probability distributions, which will allow us to determine if the linear relationships between our explanatory variables and our response variable are significant. This process is often called **statistical inference**.

(Groups) What do each of the symbols mean in (1) and why might they be important?

- **independence:** Knowing the value of one of the residuals should tell us nothing about the rest.
- **identically distributed:** Each of the residuals come from the same probability distribution.
- **zero mean:** The residuals have an average value of zero (i.e. the model is not biased).
- **constant variance:** The spread of the residuals is constant across all predictions.

Why they are important is something we will talk about for the next several weeks.

5 Why "Linear Regression"?

5.1 Why Linear?

Model are composed of:

- **coefficients:** These are *constant* values that are *estimated* to optimize the model fit.
- **variables:** These are the *observed* values, calculated from the data that we use to estimate parameters or make predictions.

A model is considered "linear" if it can be written as a sum of coefficients β multiplied by a set of variables x, i.e.

$$Y = \sum_i \beta_i X_i$$

This means that you can have nonlinear variables as long as the coefficients are linear.

(Individual) Which of the following models are linear and which are non-linear?

- $Y = \beta_0 + \beta_1 X_1 + \epsilon$
- $Y = \beta_0 + \beta_1 e^{X_1} + \epsilon$
- $Y = \beta_0 + \beta_1 X_1 X_2 + \epsilon$
- $Y = \beta_0 + X_1^{\beta_1} + \epsilon$

First three are linear, last one is not.

5.2 Why Regression?

Based on concept that things tend to "regress" to the mean:
Example: heights of fathers vs sons:

- Tall fathers tend to have tall sons, but those sons will tend to be shorter than their fathers.
- Short fathers tend to have short sons, but those sons will tend to be taller than their fathers.
- Thus, a line comparing "standard deviations" of fathers and sons heights will have a slope approximately equal to one, while the **regression** line will have a slope that is less than one.
- Because things regress to the mean, the regression line will always be flatter than the standard deviation line.

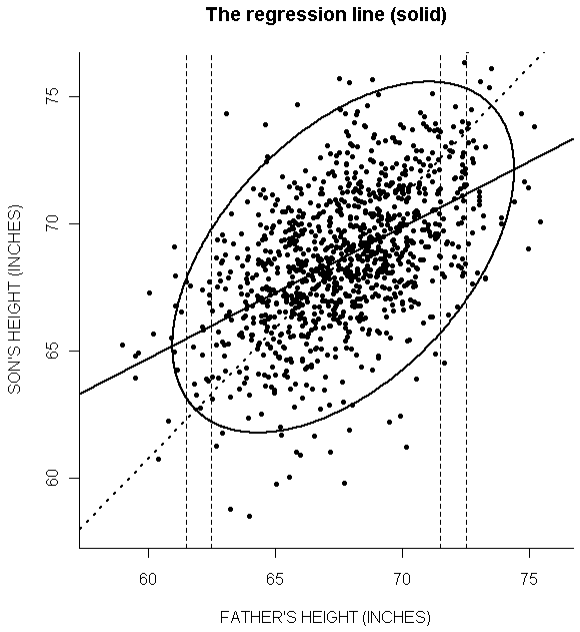


Figure 2: Plot of father vs sons' heights. The dotted line is the standard deviation line while the solid line is the regression line.

Handout 1.2: Introduction to Hypothesis Testing

Dr. Bean - Stat 5100

1 Why Hypothesis Testing?

In statistics, hypothesis tests are a way to determine if an **observed** difference is **significant** or simply due to chance. The key ingredients of a hypothesis test are:

- A null and alternative hypothesis.
- An observed statistic taken from a sample of the population.
  - Example: $t = \frac{\bar{X}-\mu_0}{SE\{X\}}$
- An assumed probability distribution for the test statistic IF the null hypothesis is true.

The climax of a hypothesis test is the determination of the **p-value**. If the p-value is small (< 0.05), we reject the null hypothesis, and if it is not small, we fail to reject the null hypothesis.
Note that we *never* accept the alternative hypothesis, we simply *fail to reject* the null hypothesis.

(Individual) What is a p-value?

The probability of obtaining our sample statistic, or one more extreme, if the null hypothesis was true.

(Groups) Determine whether or not the following statements are true or false:

- **The p-value is the probability that the null hypothesis is true. (FALSE)**
- **We reject a null hypothesis when the p-value is small. (TRUE)**
- **If the p-value is very small, it is not possible that the null hypothesis is true. (FALSE)**
- **The difference between the sample mean and the population mean is all that matters in the test statistic. (FALSE)**

(Individual) Why is the assumed distribution for the test-statistic such a big deal?

We use the assumed probability distribution is how we determine the p-value, and the p-value is how we determine the "significance" of our results. If the probability distribution is not appropriate then the p-value will be worthless.

How can we know the distribution of the test statistic?

- Through visualizations: Histograms, qqplots, boxplots.
- More often, the **Central Limit Theorem** assures us that the test statistic will follow a normal probability distribution.

2 Example:

Researchers have studied how the amount of sunlight bamboo is exposed to affects the speed of growth. One study compared the growth of 50 bamboo shoots grown under standard conditions to the growth of 49 bamboo shoots that had been exposed to 10% more sunlight. The growth was measured 40 days after planting. The observed mean and sd for bamboo under the standard growing conditions were 32.04 inches and 5.82 inches while the observed mean and sd for the more sunlight bamboo were 28.61 inches and 6.32 inches.
Hypothesis:

$$H_0 : \mu_1 = \mu_2$$
$$H_A : \mu_1 \neq \mu_2$$

We will never know the values of μ_1 and μ_2. They are **population parameters** that we could only know if we sampled every person in the population.

Rather, we **estimate** the values of these population parameters with our samples, giving us values of $\bar{X}_1 = 32.04$ and $\bar{X}_2 = 28.61$ and an observed difference of $\bar{X}_1 - \bar{X}_2 = -3.43$.

If the null hypothesis was true then the observed difference would follow a t-distribution centered at 0 with a standard deviation (assuming pooled variances) of 1.221. This also means that our observed value of $t = \frac{-3.43}{1.221} = -2.81$. The p-value associated with our observation is 0.006.

The p-value in this setting is:

> The probability of having an observed difference between the two bamboo groups as, or more, extreme than -3.43 IF the null hypothesis (no difference) was in fact true.

The p-value says that our observed difference would have been **very unlikely** (less than a 1/100 chance) if the null hypothesis was actually true. This gives us evidence to **reject the null hypothesis** and conclude that the growth rate of bamboo is different when sunlight conditions change.

To summarize:

1. We make a claim about the value of the population parameters.
2. We test the claim by obtaining statistics from a sample of the population.
3. We determine the probability of obtaining our sample statistic (or something more extreme) IF the null hypothesis was true.
4. If the probability of our observation is LOW, we reject the null hypothesis, if it is NOT LOW, then we fail to reject the null hypothesis.

(Individual) Suppose the p-value for the above example had been 0.02 instead of 0.006. Would you conclusions change? What about if the p-value had been 0.13? How about 0.98?

Same conclusions for 0.02, but fail to reject the null hypothesis for p-values of 0.13 and 0.98.

(Individual) What is the different between a practical difference and a "significant" difference?

As seen before, a practical difference is not always significant, but a significant difference is not always practical.
When sample sizes are LARGE, nearly every difference is flagged as significant, even if the actual difference between groups is small.

3 Inference vs Prediction

In linear modeling, we assume that the population follows the model:

$$Y = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \cdots + \beta_{p-1} X_{i,p-1} + \epsilon_i$$

In these models, we can conduct **inference** to determine if the linear relationship between an explanatory variable X_k and the response variable Y is significant.

HOWEVER, we can also use these same models to try and make accurate **predictions** of Y.

Models that are accurate tend to have significant coefficients, but models with significant coefficients are not always accurate.

Our approach to linear modeling in this class changes slightly when our primary interest is establishing significance, vs being accurate.

(Groups) Can you think of an example when our primary motive for creating a model is to create accurate predictions? How about an example where the primary motive is determining the significance of the coefficients?

Accuracy: Predicting the market value of a house given square footage, lot size, etc.

Significance: Determining if there is a statistically significant gender bias in pay, after accounting for other demographic factors.

1.3: SAS Crash Course

Dr. Bean - Stat 5100

1 Why SAS?

SAS is a popular statistical software package used by many companies in researchers.

Advantages:

- Lots of output without having to write much code.
- Provides lots of graphical diagnostics automatically.

Disadvantages:

- Expensive (if you want a desktop version).
- Hard to customize output (particularly visualizations).

Teaching SAS in this class:

- (For non-math majors): Has the smallest learning curve to create basic linear models.
- (For math and stats students): Provides you exposure to multiple programming languages as several of our upper division courses use R.

In this class, we will focus on using SAS Studio, which is a free online version of SAS.

2 SAS Studio Online

- Navigate to `https://odamid.oda.sas.com/SASLogon/login?service=https%3A%2F%2Fodamid.oda.sas.com%2FSASODAControlCenter%2Fj_spring_cas_security_check`.
- Select the "Not registered or cannot sign in?" option.
- Follow directions for creating a SAS profile.
- Note that the email confirming your profile may take some time (5-10 minutes) to receive.
- Note also that the only tool we will use in this class is "SAS Studio".

3 Getting Started

Once you have an account, your SAS studio window will look something like this.

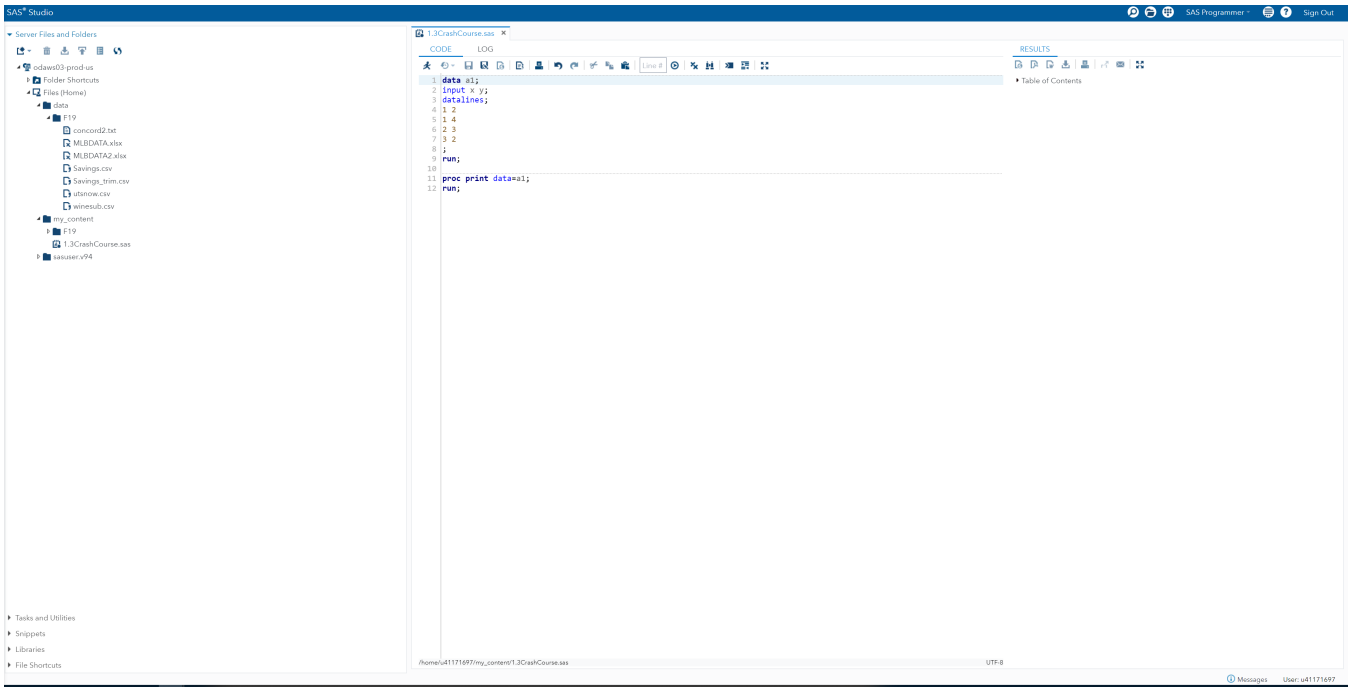


Figure 1: Sample SAS studio interface.

Main windows:

- CODE (or Editor): The window where you type in SAS commands. Font color matters in this window.
- LOG: tells what was "done". Red notes indicate errors. This is the first thing you should check if your results are unexpected.
- RESULTS (or Results Viewer): displays results in HTML format. You can export the HTML output as a pdf or a rft (rich text file - Microsoft word) to obtain nice looking output that you can copy into your reports. Or, you can use the snipping tool to save relevant output and import them as figures in LaTeX (more on LaTeX in 1.4).

You run SAS code by clicking on the "Running Man".

Necessary Components of a SAS Program:

- a semi-colon at the end of every statement
- a data statement that either creates or imports a dataset
- at least one space between each word or statement
- (almost always) a **procedure** that performs some type of analysis with your data
- a run statement

Data

There are three ways to read data into SAS. The first way is the easiest, but also not practical for large datasets.

Read in data "by hand"

```
/* Text between the asterisks are considered comments in SAS */

data a1; /* Create a new dataset named a1 */
input x y; /* List the variable names of the data that will be included in a1 */
datalines;  /* (or 'cards') tells SAS to start reading in data */
1 2
1 4
2 3
3 2
;
run; /* Tells SAS to execute the above code */

proc print data = a2; /* Print the dataset to the output screen */
run;
```

Obs	x	y
1	1	2
2	1	4
3	2	3
4	3	2

Read in data from a file.

Assume that the same data as before our located in the SAS studio folder:

`/home/u41171697/data/mydata.txt`

Note that SAS Studio will not recognize file paths on your actual computer. All data that you wish to read into SAS must be uploaded to SAS studio first.

We could read the data directly from the file using:

```
data a1;
infile '/home/u41171697/data/mydata.txt'; /* Specify path to file */
input x y;
run;
```

Read in data from a file using a procedure.

Now suppose that the data are in the excel file mydata.xlsx with the variable names included on the first row.

```
proc import
datafile =  ''/home/u41171697/data/mydata.txt''
dbms=xlsx /* Specify the file type (what separates the variables) */
out=work.a1 /* Specify the name of the dataset */
replace; /* Overwrite any datasets in the directory with the same name */
run;
```

Altering Datasets

We can add or change variables in a dataset using the commands:

```
data a2;
set a1; /* Create a copy of the dataset a1 */
  xy = x*y; /* Multiplication */
  xsq = x**2; /* Exponentiation */
  xeq1 = 0;
  if x = 1 then xeq1=1; /* Conditional Assignment */
run;

proc print data = a2;
var x y xy xeq1; /* Print all variables to the screen except xsq */
run;
```

Obs	x	y	xy	xeq1
1	1	2	2	1
2	1	4	4	1
3	2	3	6	0
4	3	2	6	0

Filtering Datasets

We can subset datasets using conditional statements.

```
data a3; set a2;
  if y < 3.5;
    /* Default 'then' is keep */
run;
/* Same as: */
data a3; set a2;
  if y >= 3.5 then delete;
run;
proc print data=a3;
 var x y xeq1;
```

```
 title1 'Smaller Set';
run;
```

Smaller Set

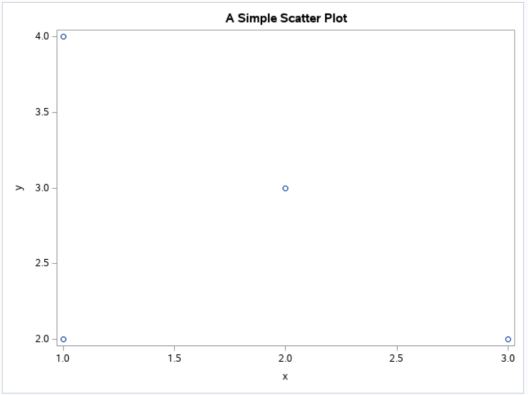
Obs	x	y	xeq1
1	1	2	1
2	2	3	0
3	3	2	0

Procedures

We prepare data in a SAS-friendly format using DATA steps. We analyze data using SAS procedures (PROC). Some PROCs that we will commonly use in this class include:

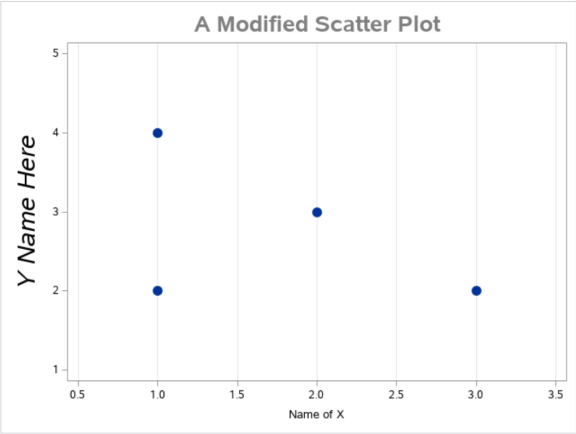
- Fitting models: PROC REG, PROC LOGISTIC, PROC ARIMA, and more
- Graphical Checks: PROC SGPLOT, PROC SGSCATTER, PROC BOXPLOT, PROC UNIVARIATE

```
proc sgplot data=a2;
  scatter x=x y=y ;
  title1 'A Simple Scatter Plot';
run;
```




```
proc sgplot data=a2;
  scatter x=x y=y /
    markerattrs=(symbol='CIRCLEFILLED'
                 size=12);
  xaxis min=.5 max=3.5 grid
    label='Name of X';
  yaxis values=(1 to 5 by 1)
    label='Y Name Here'
    labelattrs=(size=20 style=ITALIC);
```

```
  title1 height=2 color=grey
    'A Modified Scatter Plot';
run;
```




```
/ * @@ reads symbols in variable order and ignores new lines
$ indicates that variable z is a character and not numeric
. - indicates missing values */
 data a1; input x y z $ @@; cards;
   1 2 alpha   1 4 .
   2 3 gamma   3 . delta
   ;
 run;
 proc means data=a1;
   var y;
   title1 "Means Output";
 run;
 proc print data=a1;
   var y x z;
   title2 "Subtitle";
run;
```

Means Output

The MEANS Procedure

Analysis Variable : y				
N	Mean	Std Dev	Minimum	Maximum
3	3.0000000	1.0000000	2.0000000	4.0000000

Means Output
Subtitle

Obs	y	x	z
1	2	1	alpha
2	4	1	
3	3	2	gamma
4	.	3	delta

4 Miscellaneous Notes

- Missing semicolons are perhaps the most common bug in SAS code.
- The best way to get started with SAS programming is to look at the course example code, then figure out how to modify that code for your particular problem.
- Export output by copying from Results Viewer (sometimes helps to use the "Download results as RTF file") and pasting into a word document, or saving the images and importing them into a LaTeX document.
- Help in SAS: The question mark symbol in the upper right hand corner of the SAS studio editor is a good place to look for function documentation. However, reading SAS documentation can be difficult if you aren't already familiar with the procedure.
- SAS code can be written across lines. Line breaks are managed with semicolons. Data can be read in continuously with "@@".
- Missing Values: SAS procedures will completely ignore an observation if one of the variables is missing; to code a value as "missing", use the period (.) character.
- "Strings": Read in character variables with $ after the name in the input line.
- Comment Lines: To comment out a line up to the next semi-colon, put an asterisk (*) before it. To comment out an entire section, start with /* and end with */
- Selective Output: SAS will usually give you more output than you want or need, so you will need to know what you want in order to do anything useful with the output. **It is not appropriate to include ALL SAS output on homework assignments and project papers.**
- **Save Code:** SAS studio will not warn you about unsaved code when you try and close your browser. **Save your code** often to avoid frustrating loss of code.
- This class requires SAS version 9.3 or later. This is automatic for anyone using the online version of SAS studio.

5 How to "win" at SAS

- The best way to learn SAS is to **use it**. Start early on the following assignments:
  - Homework 1
  - SAS Virtual Learning Course
- Take time to *understand* SAS code before you use it. This will make it much easier to modify the code on homework and projects.

1.4: Data Exploration

Dr. Bean - Stat 5100

1 Why Data Exploration

Data Modeling is a lot like:

In order to avoid disaster, you need to **look** before you **jump**.

Example: Consider four scenarios where we use to create a model that uses values of x to predict values of y. We make the assumption in each case that the data can be modeled as

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \epsilon_i \tag{1}$$

This assumption means that we assume that X and Y share a linear relationship. That is, as X increases, Y will increase proportionally. We will explore this further in Handout 2.1.
I estimate the values of β_0 and β_1 using SAS for all four scenarios. The estimated models all have identical form, with identical measures of model goodness (which we will learn about in Handouts 2.2 and beyond).

$$\hat{Y} = 3 + 0.5X$$

(Groups) Using the results of Figure 1, which models are appropriate, and which are inappropriate? Why?

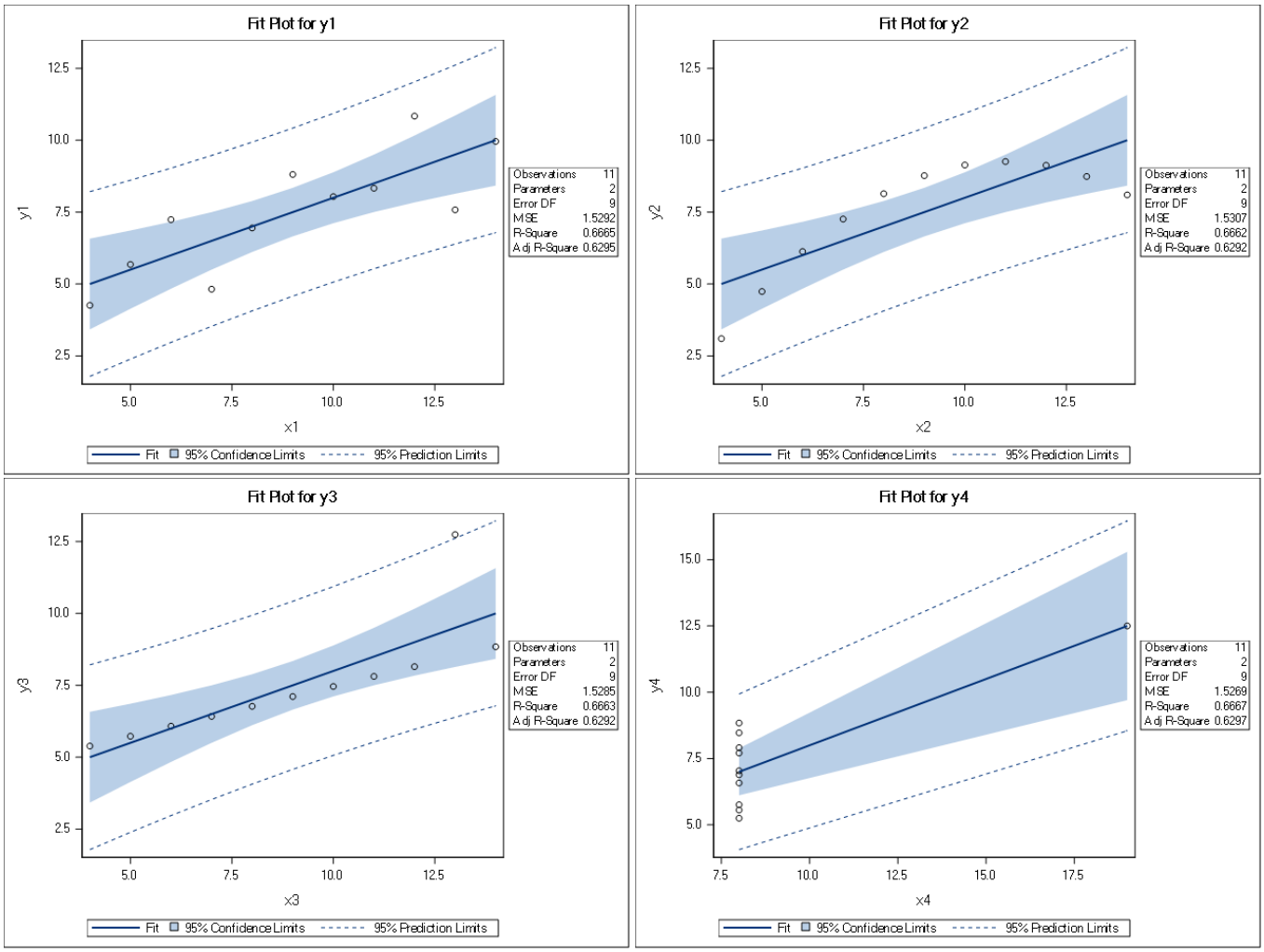


Figure 1: Plots of X vs Y, along with the estimated regression line, for Models 1-4.

Model 1 is the only appropriate model as the rest have outlier points or a non-linear relationship between X and Y.

Data Explorations BEFORE modeling will help us to detect:

- Skewed distributions
- Outlier points
- Non-linear trends

Often, we can use **variable transformations** to get data that are normal, or at least symmetric, in distribution.

Why symmetric data? Consider the "door hinge" problem.

Common Exploratory Plots

- **Boxplots:**: Show the five quartlies of the data (min, 25th percentile, median, 75th percentile, and maximum).
  - Values that are farther than 1.5*IQR (Interquartile Range, which is the 75th percentile minus the 25 percentile) above the 75th percentile or below the 25th percentile are typically plotted as "outlier" points.
  - Great way to quickly summarize the range of values.

```
proc sgplot data=concord1;
vbox Water81;
run;
```

Add option "/ datalabel =" to identify potential outlier points.

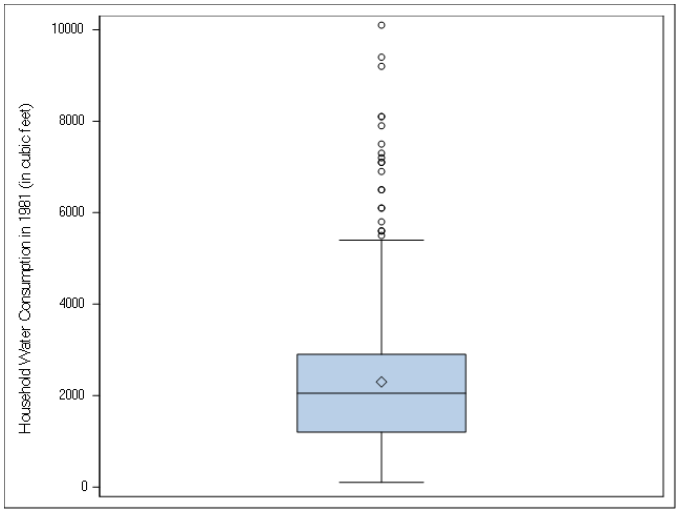


Figure 2: Sample boxplot.

- **Histograms:** Use bins to show the number of observations in a range.
  - Help us to visualize the distribution of the data by imagining a smooth curve running along the top of the bins.
  - Word of caution: the choice of bin width can drastically change the shape of a histogram.

```
PROC UNIVARIATE DATA = concord1 noprint;
HISTOGRAM Water81;
run;
```

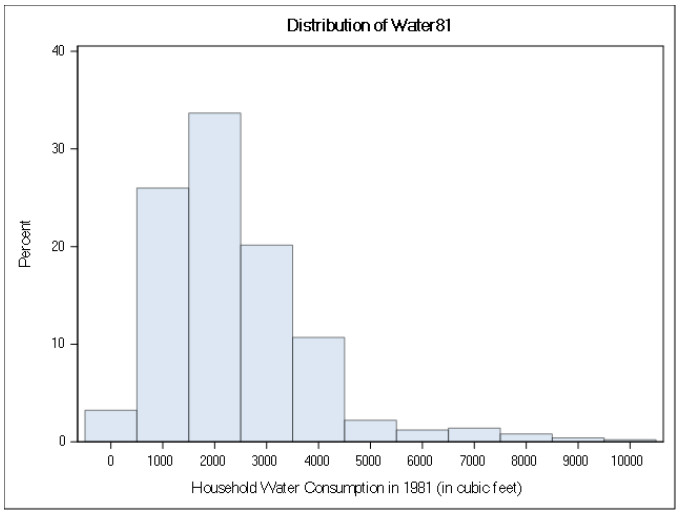


Figure 3: Sample histogram.

- **QQ Plot:** "Quantile Comparison" plots help to easily compare the observed distribution of points to a theoretical (typically normal) distribution.
  - Plots the data quantiles against the theoretical quantiles of similar observations that are normal in distribution.
  - Points that closely follow the diagonal line indicate that the observed data follow the theoretical distribution.
  - While they don't help to visualize shape, qqplots are superior to histograms as a visual check for normality.

```
PROC UNIVARIATE DATA = concord1 noprint;
qqplot Water81 / NORMAL(mu=est sigma=est);
run;
```

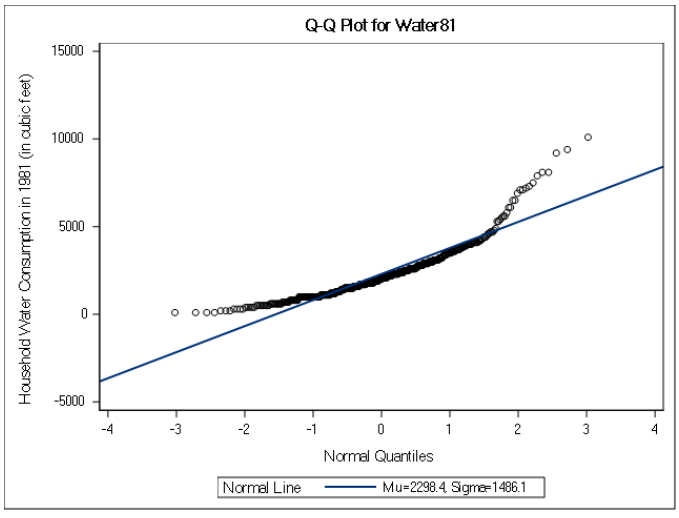


Figure 4: Sample quantile comparison plot for a normal probability distribution.

- **Scatterplots:** Plots paired observations from two variables as points on a two-dimensional plot.

- Excellent way to determine if two variables share a relationship.
- Can combine in a **scatterplot matrix** when looking at relationships between more than two variables.
- Subject to **overplotting** when you have thousands of observations that you are trying to plot at the same time.

```
proc sgscatter data=concord1;
matrix Water81 Water80 Water79;
run;
```

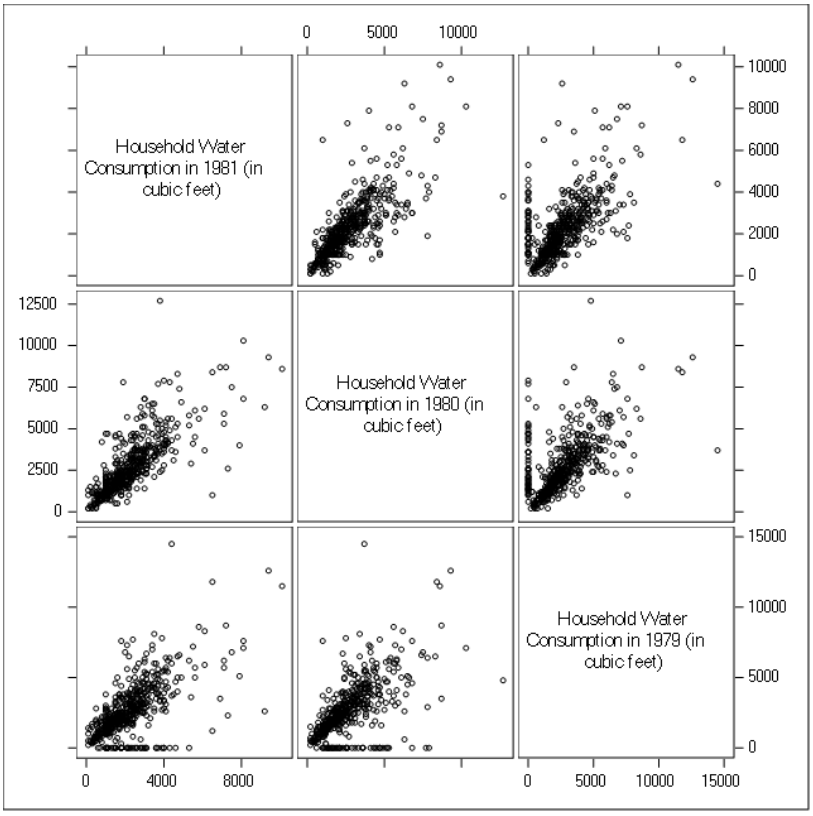


Figure 5: Sample scatterplot matrix.

See **Handout 1.4.2** for an extended example in SAS of data explorations.

1.5: Introduction to Statistical Writing

Dr. Bean - Stat 5100

1 Writing Fellows

There will be three writing-intensive assignments this semester. Two will be completed individually and one will be completed in groups. For each one of these assignments, you will be required to make two submissions of your paper. The first will be a rough draft submission that must be complete in the content, but might need some refinements in the presentation of the content. The second will be a final, polished version of the papers. Between the two submissions, each of you will be **required** to meet with one of our writing fellows. The contact information for the writing fellows for this class are:

- Tate Shepherd
- Hala Louviere
- Tess Tureson
- Emma Kristensen

Important points on writing fellows:

- You are **required** to meet with a writing fellow as part of your grade for each paper you write. However, the writing fellows do not grade any aspect of your paper.
  - Note that points will be deducted if you are late to your appointment with your writing fellow.
- Writing fellows are equipped to help you improve the clarity and the formatting of your writing. They will not comment on the technical content of your paper.
- You do not have to take every suggestion the writing fellow gives you, but it is in your best interest to incorporate nearly all of their suggestions.

2 Why so much writing?

For the undergraduate students:

- The university requires that you take a communications intensive course beyond ENGL 2010.
- The university has determined that this course contains enough opportunities to communicate (through writing and presentations) to qualify as such a course.
  - If this course was not a CI course, some of our undergraduate statistics majors would fall short of the requirements of graduation.

For the graduates:

- Most, if not all, of you will be required to communicate your research, often in the form of a thesis, dissertation, or journal article.

- The ability to effectively communicate your efforts in these venues is absolutely vital to the success of your research.
- Despite this importance, we offer very little training in how to write effectively.

For all:

- **The ability to effectively communicate quantitative information will give you a competitive advantage in graduate school or in the workforce.**

3 General Writing Tips

Use concise, straightforward language.

- Intelligent writing doesn't require using fancy words.
- Concise writing will help retain interest in what you are saying.

(start)

> We first employed forward variable selection on our model. This told us that we should remove the variables X_2 and X_7. We acknowledge that the aforementioned method is suboptimal for use in variable selection. We then tried backwards variable selection, which told us to remove X_1, X_2, X_4, and X_{11}. Finally, we tried the stepwise variable selection approach which told us to remove X_2 and X_{11}. Because the backwards and stepwise regression both suggested the removal of X_2 and X_{11}, we decided that we would eliminate these variables from our model.

(improved)

> We tried several variable selection techniques, including forward, backwards, and stepwise selection. Each method suggested we remove different variables, but backwards and stepwise selection both recommended the removal of X_2 and X_{11}. This agreement across selection methods prompted us to remove these variables in our final model.

Use active voice whenever possible

- Active voice is more concise and gives you ownership over your results.
- Personal pronouns are OK, but use "we" instead of "I", even if you are the only author.

(start)

> It was determined that the variable X_2 should be removed from the model.

(better)

> We removed the variable X_2 from our model.

Make sure you provide meaning to the numerical results in the introduction and conclusion of your paper.

- Your ultimate goal is to persuade people that their is valuable information contained in the data.
- Simply presenting a table of results fails to persuade people as to why the results are important.
- Providing a "why" in your writing make readers more likely to pay attention to your analysis.

4 Using LaTeX

- LaTeX is a markup language intended for scientific writing.
- It is particularly good for including **references** and **mathematical equations**.

Writing equations:

%	Add a comment to your document (ignored when compiling the document).
$ ... $	Add an equation to the current line of text.
$$ ... $$	Put an equation on its own line.
\[... \]	Put an equation on its own line.

Referencing Equations

You can also number and label equations to include them in the text.

```
\begin{equation}
E = mc^2
\label{eq1}
\end{equation}
\begin{equation}
Y = \beta_0X_{i,1} + \epsilon
\label{regression}
\end{equation}
Reference Equation \ref{eq1} and \ref{regression} in the text.
```

$$E = mc^2 \tag{1}$$

$$Y = \beta_0 X_{i,1} + \epsilon \tag{2}$$

Reference Equation 1 and 2 in the text.

The same goes for referencing figures and tables.

```
\begin{figure}[H] % H command requires 'float' package
\centering
\includegraphics[width = 0.25\textwidth]{figures/module1/usu.png}
\caption{This is the Utah State Logo}
\label{fig1}

The USU logo is included in Figure \ref{fig1}.
\end{figure}
```

Figure 1: This is the Utah State Logo

The USU logo is included in Figure 1.
The equation environment includes commands for all the greek letters as well. Check out:

`https://www.overleaf.com/learn/latex/List_of_Greek_letters_and_math_symbols`

Document Headers

LaTeX Document include document headers that allow you to customize the overall format of the final document. These headers can seem like a pain at first but they are extremely valuable in helping you quickly change the format of your paper depending on the context. A typical document header looks something like this:

```
\documentclass[11pt]{article}

% Add necessary packages
\usepackage{amsmath}

% Begin Document
\begin{document}

%% ADD ALL PAPER CONTENT HERE

\end{document}
```

Getting Started with LaTeX

There are several free document editors that allow you to start using LaTeX, most notably:

- Windows: MiKTeX (`https://miktex.org/download`)
- Mac: MacTeX (`http://www.tug.org/mactex/`)
- Linux: TeXLive (`http://www.tug.org/texlive/`)

You can also get started with an online LaTeX editor called Overleaf. (`https://www.overleaf.com/`)

5 Using Microsoft Word

Microsoft Word also has an equation editor that can be quickly accessed using the "alt + equal" keystrokes. Many latex commands, particularly those for greek variables, subscripts, and superscripts, are recognized in Microsoft's equation editor.

Remember: All equations included in your homework or papers must use a professional equation editor environment. Formatting points will be taken away for equations written outside of such an environment such as Y = b0 + b1x. vs $Y = b_0 + b_1 x$.