

Math 2311

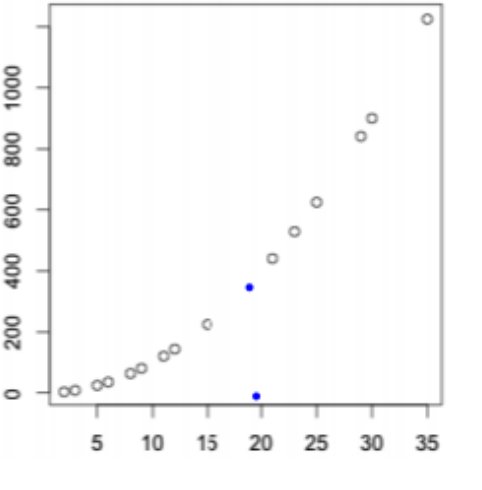
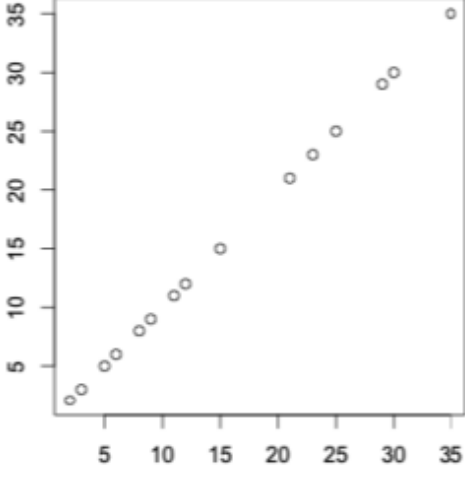
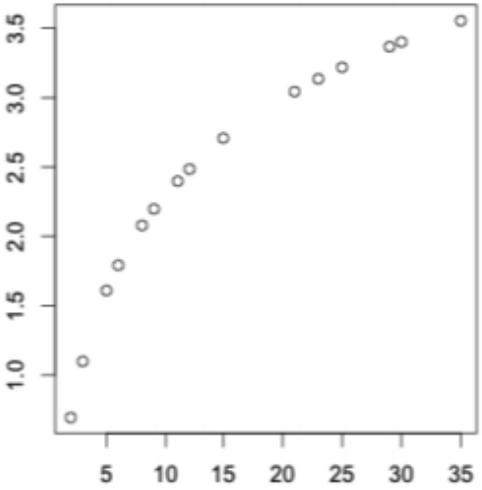
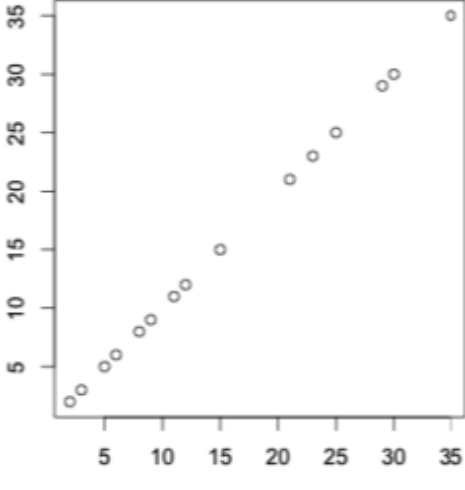
Bekki George – bekki@math.uh.edu

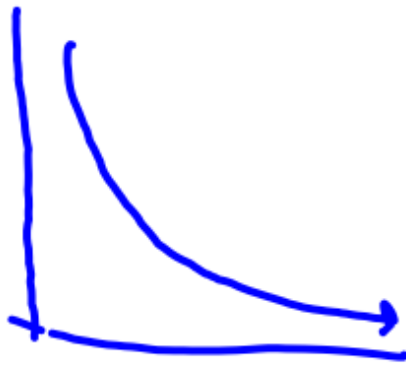
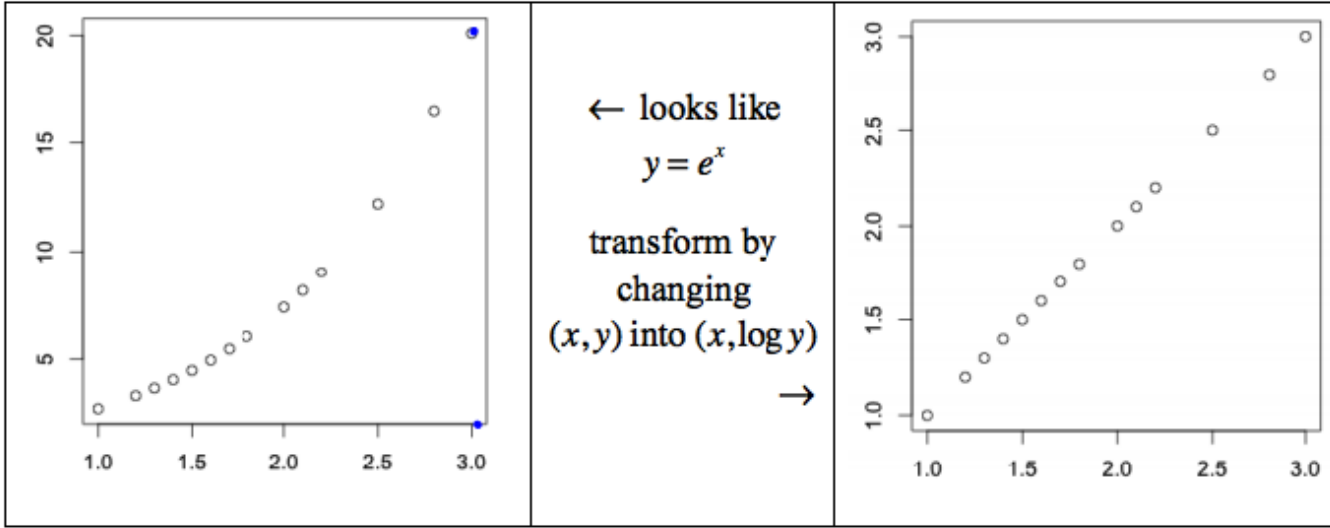
Office Hours: MW 11am to 12:45pm in 639 PGH

Online Thursdays 4-5:30pm

And by appointment

Class webpage: <http://www.math.uh.edu/~bekki/Math2311.html>

Original Scatter-plot	Type of plot and Transformation	Modified Scatter-plot
	<p>← looks like $y = x^2$</p> <p>transform by changing (x, y) into (x, \sqrt{y}) →</p>	
	<p>← looks like $y = \log x$</p> <p>transform by changing (x, y) into (x, e^y) →</p>	

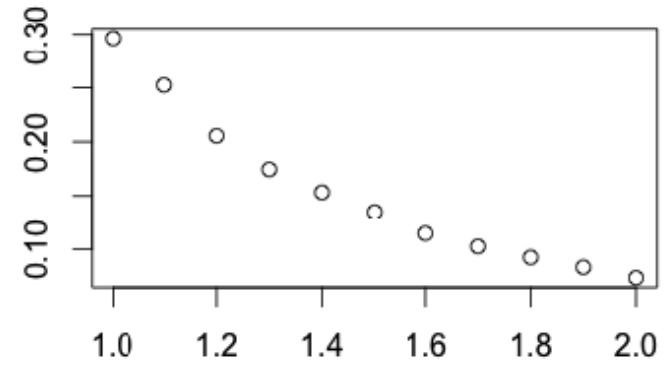


$y = \frac{1}{x}$ → $(x, \frac{1}{y})$
or $y = \frac{1}{x^2}$ → $(x, \sqrt{\frac{1}{y}})$
 $\frac{1}{y} = x^2$

3. In a physics class, the intensity of a 100-watt bulb was measured by a sensing device at various distances from the light source, and the following data were collected. Note that a candela (cd) is an international unit of luminous intensity.

Distance (meters)	Intensity (candelas)
1.0	.2965
1.1	.2522
1.2	.2055
1.3	.1746
1.4	.1534
1.5	.1352
1.6	.1145
1.7	.1024
1.8	.0923
1.9	.0832
2.0	.0734

- Give a scatter-plot and residual plot of the data.
- Based on the graphs in part a, propose a model for the data. Show me evidence to support your conclusion. Go through all necessary steps to construct a model of the type you chose.



$$d = c(1, 1.1, 1.2, \dots)$$

$$i = c(.2965, .2522, \dots)$$

$$r = -.959$$

$$\hat{y} = .4686 - .21x$$

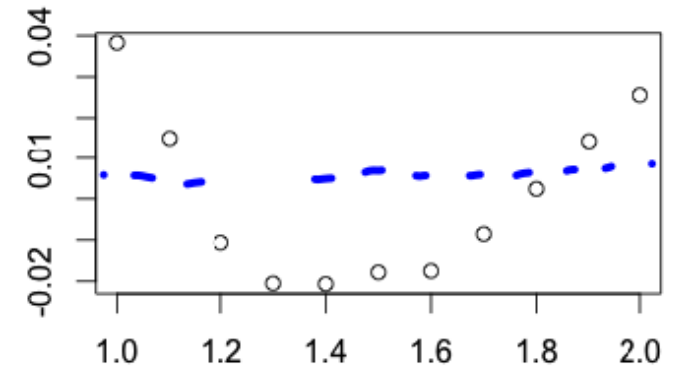
```
> modp3=lm(i~d)
> modp3
```

```
Call:
lm(formula = i ~ d)
```

```
Coefficients:
(Intercept)          d
  0.4686         -0.2104
```

```
plot(d, resid(modp3))
```

residual plot →



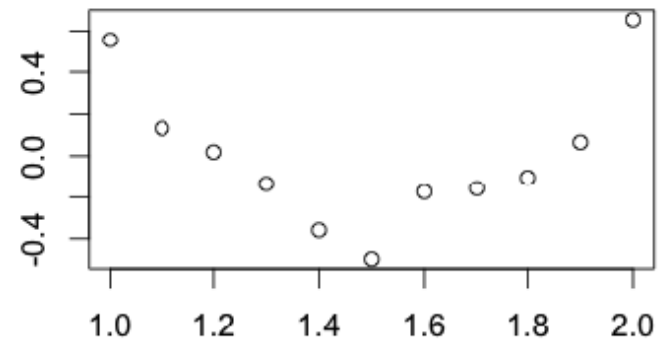
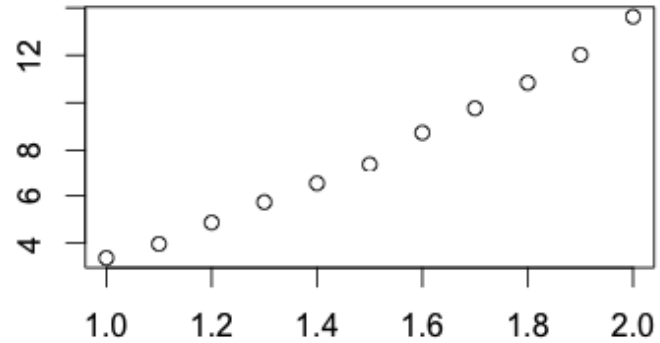
lets transform

```
> newy1=1/i
> newy1
 [1] 3.372681 3.965107 4.866180 5.727377 6.518905 7.396450
 [7] 8.733624 9.765625 10.834236 12.019231 13.623978
> plot(d,newy1)
> modp3_2=lm(newy1~d)
> modp3_2
```

```
Call:
lm(formula = newy1 ~ d)
```

```
Coefficients:
(Intercept)          d
   -7.334         10.152
```

```
plot(d,resid(modp3_2))
```

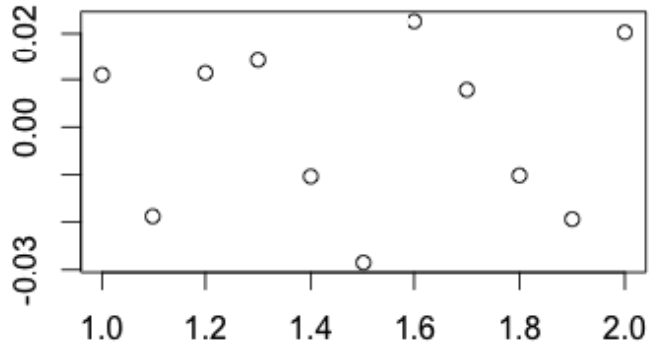


trying $y = \frac{1}{x}$ model

(let newy's = 1/old y's)

new $r = .99466$

for $y = \frac{1}{x^2}$ model (let $y^{\text{new}} = \sqrt[3]{\text{old } y}$)



← residuals
(nice)

```
> mod3_3=lm(newy2~d)
> mod3_3
```

```
Call:
lm(formula = newy2 ~ d)
```

```
Coefficients:
(Intercept)          d
-0.01975         1.84526
```

$$\hat{\text{newy}} = -0.01975 + 1.845x$$

$$\sqrt[3]{\hat{y}} = -0.01975 + 1.845x$$

$$\sqrt[3]{y} = \frac{1}{-0.01975 + 1.845x} \rightarrow$$

$$\text{So } \hat{\sigma}_0 = \left(\frac{1}{-.01975 + 1.845x} \right)^2$$

Popper 13

1. What is the difference between a population and a sample?
 - a. A sample is a subset of the population
 - b. A population is a subset of the sample
 - c. They are the same thing

6.1 – Sampling

Terms:

Population – each element (or person) from the set of observations that can be made

Sample – a subset of the population

Census – systematically getting information about an entire population

Sampling – studying a part (a sample) in order to gain information about an entire group

Sampling Frame – the list of individuals from which a sample is actually selected

Types of Sampling –

- Voluntary Response sample – people who choose themselves by responding to a general appeal (over represents people with strong opinions)
- ⚡ A simple random sample (SRS) consists of individuals from the population chosen in such a way that every set of individuals has an equal chance to be the sample actually selected.

A probability sample gives each member of the population a known chance to be selected.

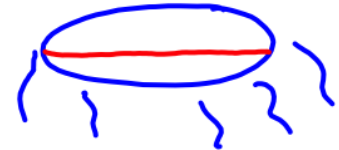
A stratified sample divides the population into groups of similar individuals, called strata, and chooses a SRS in each stratum and combines these to form the full sample.

In multistage sample design samples are taken from various subsets of the population until a manageable number of samples to interview are arrived upon.

Convenience sampling is a non-probability type of sample where the sample is chose based on their convenient accessibility and proximity.

--- Activity ---

Jelly Blubbers!



To open Jelly Blubber Activity: <http://www.math.uh.edu/~bekki/2311/notes/jellyblubbers%20activity.pdf>

Picking : 2.2 cm , 2.37cm , 3.15cm , 4.13 , 2.8

Random : 2.37 , 1.92 , 0.8 , 2.32 , 2.04 , 2.34

1.94 cm

Popper 13

2. What is a census?
 - a. When information is gathered from a subject
 - b. When information is gathered from the entire population

3. A voluntary response sample is always a good representation of the population
 - a. True
 - b. False

✦ A table of random digits is a long string of the digits 0 – 9 where each entry in the table is equally likely to be any of the 10 digits and the entries are independent of each other.

→ Experiment – actively impose some treatment in order to observe the response

→ Observational study – investigators observe subjects and measure variables of interest without assigning treatments to the subjects.

Two elements are confounded when their effects on a response variable cannot be distinguished from one another.

Statistical inference provides ways to answer specific questions from data with some guarantee that the answers are good ones. In inference we must think about how to produce data as well as analyze data.

The design of sample refers to the method used to choose the sample from the whole population.

* Voluntary response and convenience sampling are examples of bad sample design.

The design of a study is biased if it systematically favors certain outcomes.

Under-coverage occurs when some groups in the population are left out of the process of choosing the sample.

Line								
101	98360	26534	47384	94612	88666	14170	10847	05567
102	55556	59863	86607	00094	77213	35711	52851	42108
→ 103	<u>31634</u>	<u>15399</u>	<u>73476</u>	<u>77412</u>	06186	16636	54307	14947
104	13785	11509	54891	98375	68377	50572	08453	80376
→ 105	80376	73842	98465	59746	38078	25727	78502	95324
106	21198	99781	53374	25595	12153	54228	13068	71757

37, 38, 42, 46, 07

31, 63, 41, 53, 99, 73, 47, 67,
74, ...

1..100 → 00..99

Nonresponse occurs when an individual chosen for the sample can't be contacted or refuses to cooperate.

Things to watch out for in interviewing technique:

❖ Response Bias – when an interviewer's attitude suggests that some answers are more desirable than others gives the interviewer specific answers more often

Wording of Questions – confusing or misleading questions can introduce strong bias

Examples:

1. Identify the population and the sample then describe the sampling method that was used.

To conduct a pre-election opinion poll on a proposed city ordinance, a random sample of telephone numbers from the city phone book were chosen and called. (Assume all who were called answered).

Population - people in that city

Sample - people who were called

SRS of people who are in phone book

2. Determine if the study is an experiment or an observational study. Give a reason for your answer.

a. A personnel director at a large company studied the eating habits of employees by watching the movements of a selected group of employees at lunchtime. The purpose of the study was to determine the proportion of employees who buy lunch in the cafeteria, bring their own lunches, or go out to lunch.

study

b. A pharmacy student would like to know if there is a difference in results from a specific brand of drug and its equivalent generic prescription. She randomly selects 50 people who take the drug and has them complete a questionnaire regarding their symptoms and improvements after taking the specific brand or generic equivalent.

exp.

3. It is believed that 75% of all apartment dwellers in a large city deadbolt their doors in addition to locking them as an added precaution against burglary. Describe how you would select an SRS of 20 apartment dwellers to survey if there are 50 complexes in the city and each complex has 250 residents. (Use the random digit table)

⇒ # the complexes 00..49 (or 01..50)

Choose 5 #'s

pick 10 #'s for each complex

4. Bias is present in each of the following sample designs in the situations below. In each case, describe the type of bias involved and state whether you think the sampling frequency obtained is lower or higher than the actual population parameter.

a. A political pollster seeks information about the proportion of American adults that oppose gun control. He asks a SRS of 1000 American adults, "Do you agree or disagree with the following statement: Americans should preserve their constitutional right to keep and bear arms." A total of 910 or 91% agreed.

b. A restaurant chain wants to know what percentage of American families go out to eat for dinner at least 3 nights per week. They call a SRS of 1000 households between the hours of 6:00 p.m. and 8:00 p.m. and talk to 400 people. 300 of those 400 said they do not eat out 3 or more times a week.

Popper 13

4. Under-coverage occurs when some groups in the population are left out of the process of choosing the sample.
 - a. True
 - b. False

5. Two elements are _____ when their effects on a response variable cannot be distinguished from one another.
 - a. Interconnected
 - b. Confounded
 - c. Confused
 - d. Explanatory