Model Project Number 1:
Trivial Pursuit of Statistics

Jessica Cottreau

Picture this, you and a few people are sitting around and decide to play a friendly game of *Trivial Pursuit*. Everything is going well, everyone is having a good time. Then you are faced with a question for a piece of pie, "Who placed the first telephone call to the moon?" Your friend looks at the back of the card, gazes at the answer, and annoyingly says, "That's easy." You really do not know the answer, so you choose an answer from the sea of names swimming in your head. Uunfortunately it was not the right answer and your friend with the card does not attempt to conceal your momentary idiocy, instead yells out "Richard Nixon" in disbelief. But did your friend really know the answer?

It is possible that your friend truly knew the answer to the question before he or she saw the answer, but it is also possible that your friend fell victim to the "I knew it all along" effect. The= "I knew it all along" effect is a psychologically researched effect that states that a person tends to overestimate his or her knowledge on a subject when the answers are presented. Spurred by my observations of this effect while playing *Trivial Pursuit*, I decided to test these observations statistically.

I began my study by handpicking five questions from Parker Brother's *Trivial Pursuit Genus IV* board game. One question was selected from the *People & Places*, *Arts & Entertainment*, *History*, *Science & Nature*, and *Sports & Leisure* categories. The *Wild Card* category was left out to avoid two questions on the same topic. Two question sheets were then designed, one with the five questions, and the other with the questions followed by the answers. After each question set the participant is asked whether the questions were easy or difficult.

Treatment 1–no answers

Answer the following questions in your head, you do not need to write the answers down.

1. Who directed Vertigo in 1958?
2. What African republic's name was inspired by its thriving elephant tusk trade?
3. Who placed the first telephone call to the moon?
4. What isn't found in liquid form on any other planet but earth?
5. Who was involved as a player or coach in three super bowls with the Cowboys, two with the Eagles and one with the Bears?

After answering these questions did you find them easy or difficult? Put a check next to one of the following statements.

- They were easy-had difficulty with 0-2 of the questions.
- They were difficult-had difficulty with 3-5 of the questions.

Treatment 2–answers

Answer the following questions in your head, the answers are recorded after each question.

1. Who directed Vertigo in 1958? *Alfred Hitchcock.*

2. What African republic's name was inspired by its thriving elephant tusk trade? *Ivory Coast.*

3. Who placed the first telephone call to the moon? *Richard Nixon.*

4. What isn't found in liquid form on any other planet but earth? *Water.*

5. Who was involved as a player or coach in three super bowls with the Cowboys, two with the Eagles and one with the Bears? *Mike Ditka.*

After answering these questions did you find them easy or difficult? Put a check next to one of the following statements.

- They were easy-had difficulty with 0-2 of the questions.

- They were difficult-had difficulty with 3-5 of the questions.

The two treatments were then ordered randomly by putting numbers 1-100 in a bucket and picking out fifty of the numbers for the first treatment and then putting the two treatments in the random order they were assigned.

Data

The alternative chosen for this experiment is the second alternative, where there will begreater successes on the treatment that provided the subject with answers.

				Row Proportions		
Treatment	Easy	Difficult	Total	Easy	Difficult	Total
1 (No answers)	12	38	50	0.24	0.76	1.00
2 (Answers)	23	27	50	0.46	0.54	1.00
Total	35	65	100			

Thus, $x = \hat{p}_1 - \hat{p}_2 = 0.24 - 0.46 = -0.22$. (Jessica chose to use the continuity correction in order to obtain a more accurate approximation.) "New" $x = -0.22 + 0.02 = -0.20$, and $\sigma = \sqrt{[35(65)]/[50(50)(99)]} = 0.0959$. Thus, $z = -0.20/0.0959 = -2.09$, and $-z = 2.09$. The approximate P-value is 0.0183.

From this P-value, you can determine that the experiment and results are statisticallysignificant.

In conclusion, this study shows that the "I knew it all along" effect has some validity in the world today. It shows that people are likely to overestimate their ability to answer a question when the answer is in front of them. It also gives me the hope that my friend really didn't know that Richard Nixon was the first person to place a telephone call to the moon.

Model Project Number 2:
Sexual Harassment Study

Erica Olsen

I performed a completely randomized design on fifty female students between the ages of eighteen and twenty-two on the University of Wisconsin, Madison campus. The experiment had a dichotomous response to investigate the influence of gender on sexual harassment. The two versions of the question that was asked are as follows.

Version 1: A 45 year-old man is interviewing a 25 year-old woman for a political position in the state capital in Madison, Wisconsin. The interview is conducted at a local restaurant. Upon the arrival of the potential employee, the interviewer insists on buying her a drink. The interviewer asked three principal questions.

1. Are you currently in or looking for a stable relationship?
2. How many people have you been romantically involved with in the past year?
3. Do you plan to have children in the future?

Is this interview a form of sexual harassment? YES NO

Version 2: A 45 year-old woman is interviewing a 25 year-old man for a political position in the state capital in Madison, Wisconsin. The interview is conducted at a local restaurant. Upon the arrival of the potential employee, the interviewer insists on buying him a drink. The interviewer asked three principal questions.

1. Are you currently in or looking for a stable relationship?
2. How many people have you been romantically involved with in the past year?
3. Do you plan to have children in the future?

Is this interview a form of sexual harassment? YES NO

I randomized the treatments the subjects were assigned by using a stack of fifty well-shuffled cards, lying face down. Twenty-five of the cards were marked with a '1,' and twenty-five of the cards were marked with a '2.' Subjects took the top card from the stack. Subjects drawing a '1' read version one, while subjects drawing a '2' read version two. Subjects placed their cards in a box after answering either yes or no.

During data collection, I was pleasantly surprised to find that subjects were very willing to participate in my study. The collection process went very smoothly, and the randomization turned out to be less complicated than I had originally imagined.

I am a member of the Kappa Alpha Theta sorority at the University of Wisconsin, Madison. I attended one of our weekly chapter meetings and set up the randomization technique mentioned above in order to assign treatments. These subjects are of interest to me because I fit into the same gender category, age range, and am a student at the same University as the subjects. I felt that accumulating data from subjects with similar interests and experiences as myself would generate interesting results. The issue of sexual harassment is of interest to me because it has become a big issue in today's world. There are a good deal of misconceptions as to what constitutes sexual harassment and who can be a victim of it. I wanted to study the changes in results based on which gender is experiencing sexual harassment and which gender is committing harassment. The theme of sexual harassment is especially pertinent to students on college campuses, so I utilized my student status to get feedback from other students.

Before beginning data collection, I performed Fisher's test for the appropriate alternative. I assumed the skeptic was correct, and therefore that the null hypothesis $H_0 : p_1 = p_2$ was also correct. I chose an alternative hypothesis, $H_1 : p_1 > p_2$, or that there would be more successes on treatment one than on treatment two. I chose this hypothesis because of the way I feel the majority of society sees sexual harassment, as the victimization of a woman by a man, and not the other way around. I felt it was inconceivable that more people would feel that the woman was harassing the man, than would feel that the man was harassing the woman. After collecting the data, I found that there were indeed more successes on treatment one than treatment two. Treatment one had 23 successes out of 25 possible answers while treatment two had 15 successes out of 25. The observed value of the test statistic is $x = 0.32$. In addition, $s = 0.122$, and $z = 2.62$. Using these data and the standard normal curve, I obtained an approximate P-value of 0.0044. The data are presented in the table below.

	Harassment?			Row Proportions		
Treatment	Yes	No	Total	Yes	No	Total
1 (Male interviewer)	23	2	25	0.92	0.08	1.00
2 (Female interviewer))	15	10	25	0.60	0.40	1.00
Total	38	12	50			

The data collected are highly statistically significant according to the interpretation of the P-value. This means that the data support the alternative. While these results tell a lot about my subjects, they are not representative of an entire population or even the student population at UW–Madison because my subjects were all female and within a relatively small age range. I have learned, within my subject group, that more subjects feel that sexual harassment occurred when the woman was interviewed by the man than when the man was interviewed by the woman. This correlates with my alternative hypothesis and shows that gender is a factor in drawing the line between sexual harassment and friendly conversation.

Model Project Number 3:
Double Shot Basketball
Lisa Mattiacci

In my house, my roommates and I have an electronic basketball shooting game called *Double Shot.* We love playing *Double Shot* and we are constantly competing against one another to gain the high score. I performed a balanced CRD to determine the effect that the use of different basketballs have on my scores. I shot for 30 seconds (the time allotted in the game). In the first 20 seconds, each basket is worth 2 points. In the last 10 seconds, each basket is worth 3 points. The response was the total number of points I accumulated in the 30 seconds. The first treatment was shooting with the official *Double Shot* basketball. The second treatment was shooting with a *Milwaukee Bucks* basketball that was slightly heavier and flat. I performed 40 trials, randomly drawing from a shuffled stack of 40 cards, 20 red and 20 black. (The red cards represented Treatment 1.) Before collecting the data, I selected the first alternative, $\mu_1 > \mu_2$; that is, I believed that my scores would be higher on Treatment 1 than on Treatment 2.

Below are the sorted scores with the official *Double Shot* basketball (Treatment 1):

21	22	24	25	25	25	26	26	27	27
28	29	29	30	31	32	32	32	33	36

Below are the sorted scores with the *Milwaukee Bucks* basketball (Treatment 2):

17	19	19	20	21	22	23	24	24	24
24	25	25	25	26	26	27	29	29	31

Below are stem-plots of the data, by treatment.

Double Shot		*Milwaukee Bucks*	
1		1	7
1		1	99
2	1	2	01
2	2	2	23
2	4555	2	4444555
2	6677	2	667
2	899	2	99
3	01	3	1
3	2223	3	
3		3	
3	6	3	

I constructed stem-plots because they display the exact data values and they also provide a clearly defined picture since the values can be grouped. I chose to sort the data into intervals of two because it showed a good amount of detail. The stem-plot for Treatment 1 does not reveal a prominent peak; instead it shows that the majority of the data are between 24 and 33 points. The stem-plot for Treatment 2 reveals a single peak and is relatively symmetric about that peak. These pictures agree with the summary statistics below.

Feature	*Double Shot*	*Milwaukee Bucks*
Measures of Center		
Mean	28	24
Median	27.5	24
Measures of Spread		
Range	15	14
IQR	6.5 ($= 31.5 - 25$)	4.5 ($= 26 - 21.5$)
Standard deviation	3.91	3.61

After collecting the data, I found that the mean on Treatment 1 is 28 points and the mean on Treatment 2 is 24 points. The difference in means is 4 points. These data support my conjecture and are good proof because it is an average of 2 more baskets made which is relatively a lot considering the allotted time. If you also look at the quartiles of each treatment, in addition to the means, you will see even greater support for my conjecture. Quartile 1 signifies that 25% of the data is below that point; Quartile 3 signifies that 75% of the data is below that point. In Treatment 1, Q1= 25 and $Q_3 = 31.5$. In Treatment 2, $Q_1 = 21.5$ and $Q_3 = 26$. Comparing the quartiles to the means of each treatment shows that at least 75% of the scores in Treatment 1 are larger than the average score of Treatment 2. This support can also be shown in terms of the IQR, which represents 50% of the data. The majority of the scores represented in the IQR for Treatment 1 are higher than the scores represented in the IQR for Treatment 2. The Range in each treatment is approximately equal which reveals that my accuracy was consistent in each treatment. There is a significant difference in the maximum and minimum points I scored in each treatment though (Treatment 1 having higher endpoints). This difference is shown in the stem and leaf plots and also proves my alternative hypothesis. Finally, the standard deviation of this data is 3.91 for Treatment 1 and 3.61 for Treatment 2. The standard deviation of each treatment supports the empirical rule that 68% of the data lies within one standard deviation of the center (the sample mean of each treatment). This works well because the distributions of the basketball scores are bell-shaped, as shown by the stem-plots.

Model Project Number 4:
Free Throws

Eric P. Seiler and Gesina M. Seiler

We performed 100 dichotomous trials on free throw shooting. Ena was our shooter since she was a high school (⋆ star ⋆) shooting guard. She hasn't played in 13 years and we wondered if she could still shoot 82% like she used to. This project requires the use (or attempted use) of Bernoulli trials. We considered shooting the free throws to be Bernoulli trials and so proceeded to the basketball court. The following is a record of the shots taken.

First 100 Trials:

SSSFS SFSFF FFSSF SSSSS FFFSF FFSFF FSFSF FSSFS SSSFS SSSFS

SFFFS SSFSF SFFFS FSFSF FSSSF SSSSF FSSFS SFFFS SFFFF SSSSF

⋆**Note**⋆ During the performance of our study, we thought we detected a pattern where it is obvious that there is no such thing.

The following three assumptions must be true in order for a study to be Bernoulli trials. The possibility of only two outcomes, success or failure, in this case either a sunk shot or a missed shot.

Our data are summarized in the following table:

TREATMENT	SUCCESS	FAILURE	TOTAL
FREETHROWS	54	46	100

$\hat{p} = 0.54$

The second assumption is that p is constant, if p increases the shooter is improving with practice, if p decreases the shooter is either tired or bored. Our values of $\hat{p}$ are within 0.04, which we consider close enough to be Bernoulli trials; there is little evidence of improvement or decline.

	SUCCESS	FAILURE	TOTAL
FIRST HALF	28	22	50
SECOND HALF	26	24	50
	54	46	100

$\hat{p}_1 = 0.56$ and $\hat{p}_2 = 0.52$.

The third assumption is that the trials are independent of each other. The previous outcome doesn't influence the current outcome, i.e. there is no memory.

PREVIOUS	CURRENT TRIAL		
TRIAL	SUCCESS	FAILURE	TOTAL
SUCCESS	29	25	54
FAILURE	24	21	45
	53	46	99

$\hat{p}_1 = 0.54$ and $\hat{p}_2 = 0.53$.

Our $\hat{p}$ values are 0.54 and 0.53. These are very close and show there is very little dependence on the previous action. The free throw shooter shot almost exactly the same after a success as a failure.

Interval Calculations

90% Confidence Interval

$x = 54, n = 100, \hat{p} = 0.54$ and $\hat{q} = 0.46$.

$$0.54 \pm 1.645\sqrt{\frac{0.54(0.46)}{100}} = 0.54 \pm 0.082.$$

$$0.54 + 0.082 = 0.622 \text{ and } 0.54 - 0.082 = 0.458.$$

The confidence interval is $[0.458, 0.622]$.

80% Prediction Interval

$m = 100, n = 100, x = 54, \hat{p} = 0.54$, and $\hat{q} = 0.46$.

The point prediction is $m\hat{p} = 100(0.54) = 54$.

The prediction interval is

$$54 \pm 1.282\sqrt{54(0.46)}\sqrt{1 + (100/100)} = 54 \pm 9.04 = [45, 63].$$

Later in the day, we performed 100 more trials and Ena obtained 54 successes. Both the point and interval predictions were correct.

Model Project Number 5:
Darts

Katie Ledvina and Carrie Skrip

As a dart-throwing junkie, I convinced Carrie that studying my accuracy in throwing darts would be an interesting subject to study for our statistics project. Carrie agreed, since there had been many nights where we have plugged quarters into the dart machine trying so desperately to inflate our egos over a game of cricket. Also, since darts is a very common activity among bar dwellers, we thought that it may be an interesting topic overall.

In order to set up a fixed environment, we used the electronic dartboard that I have in my apartment. I was designated dart thrower for both rounds, and Carrie kept track of all the data we collected. I stood a set distance from the board, marked by black tape, and used the same darts (accidently taken from a bar in Fond du Lac) throughout the entire experiment. The darts were thrown in a series of three, until I had made 100 throws. The red and green areas on the board were designated successes, (areas that count for doubles, triples, and bulls eye), and the black and white areas were designated failures, (areas that only count for single points). The results of the first 100 throw were as follows:

FFFFF SFFSF FSFFF FFFSS SFFFF FFFSF FSSFS FSSFF FFSFF FFSFF

FFFSF FFFFF FFSFS FFSFF FFSFF SSSFS FFFFS FFFFS SSSSF FFFFF

	Success	Failure	Total
First Half	14	36	50
Second Half	15	35	50
Total	29	71	100

$\hat{p}_1 = 14/50 = 0.28$ and $\hat{p}_2 = 15/50 = 0.30$. Therefore, there is evidence that Katie's ability improved during the course of the study, but not by very much.

Next,

	Current		
Previous	Success	Failure	Total
Success	10	19	29
Failure	19	51	70
Total	29	70	99

$\hat{p}_1 = 10/29 = 0.34$ and $\hat{p}_2 = 19/70 = 0.27$. Therefore, $\hat{p}_1 > \hat{p}_2$. This means that there is evidence that Katie's performance was better after a success rather than after a failure.

Do we have Bernoulli trials?

1. Each trial must result in 1 of 2 possibilities? YES
2. The probability of success p remains constant? The difference between $\hat{p}_1$ and $\hat{p}_2$ is only 0.02; thus, this assumption seems reasonable.
3. Are the trials independent? The difference between $\hat{p}_1$ and $\hat{p}_2$ is only 0.07; thus, this assumption seems reasonable.

On the assumption that we do have Bernoulli trials the following analysis was applied.
90% Confidence interval:

$$\hat{p} \pm z\sqrt{\hat{p}\hat{q}/n} = 0.29 \pm 1.645\sqrt{0.29(0.71)/100} = [0.22, 0.36].$$

Point prediction: $m(x/n) = 100(29/100) = 29$.
80% prediction interval:

$$m\hat{p} \pm z\sqrt{m\hat{p}\hat{q}}\sqrt{1 + m/n} = 29 \pm 1.282\sqrt{29(0.71)}\sqrt{1 + 100/100} = [20, 37].$$

After this analysis we then performed another 100 trials and compared that data to the prediction interval from our prior trials. Katie threw 21 successes that did not match the point prediction. However, the 21 successes did fall into the 80% prediction interval.