S.2.3 2001

Beyond the Formula V

Tom Short
Department of Mathematical Sciences
Villanova University
800 Lancaster Ave.
Villanova, PA 19085-1699
thomas.short@villanova.edu
www.villanova.edu/~tshort/

August 2001

Statistical Power without and with a Graphing Calculator

1. An Activity about Power in a Test for One Proportion

2. Statistical Power and the Graphing Calculator

An Activity about Power in a Test for One Proportion

Tom Short
thomas.short@villanova.edu

August 2001

Ingredients (per pair of participants)

1. One six-sided die
2. A small screen (a folder or a book) to hide the roll of the die
3. Two "data recording forms" (see page 6 of this handout)

Instructions

As a class:

Choose a number, *n*, to represent the number of die rolls in each set (a good number to begin with is $n = 15$).

Within each pair of participants:

- Decide which partner will be the "Roller" and which will be the "Guesser."

- The "Roller" should roll the die once, out of sight of the "Guesser." If the die comes up with an even number proceed to *Case A*. If the die comes up with an odd number skip to *Case B*.

Case A - Even Number

1. The "Roller" should record out of sight of the "Guesser" that *Case A* has been selected.

2. The "Roller" rolls the die once.

3. If the number on the die is 1, 2, or 3, the "Roller" says enthusiastically "success!" If the number on the die is 4, 5, or 6, the "Roller" says slowly and sadly "failure."

4. The "Guesser" keeps track of how many successes and failures result.

5. The "Roller" keeps rolling until *n* rolls have been made.

6. Proceed to the *Guessing Stage* on page 4 of this handout.

Case B - Odd Number

1. The "Roller" should record out of sight of the "Guesser" that *Case B* has been selected.

2. The "Roller" rolls the die once.

3. If the number on the die is 1 or 2, the "Roller" says enthusiastically "success!" If the number on the die is 3, 4, 5, or 6, the "Roller" says slowly and sadly, "failure."

4. The "Guesser" keeps track of how many successes and failures result.

5. The "Roller" keeps rolling until *n* rolls have been made.

6. Proceed to the *Guessing Stage* on page 4 of this handout.

Guessing Stage

The "Guesser" can analyze the data in any manner, but eventually, the "Guesser" must guess which of the two ***Cases*** (*A* or *B*) was being used for this set of *n* trials.

After the "Guesser" guesses either ***Case A*** or ***Case B***, the "Roller" should record whether the guess was correct or not. At this point the "Roller" may reveal whether the guess was correct or incorrect.

The partners can repeat the whole activity again, including the "odd or even" roll at the beginning. The instructor will eventually request that the partners submit how many guesses were correct out of how many times the experiment was conducted.

Analysis

The instructor (or a participant) collects the results of the guesses and tallies them in a table as follows:

		Status of Guess	
		Guess was ***Case A***	Guess was ***Case B***
Selected Case	***Case A*** actually selected	Number of times *Case A* was selected and the "Guesser" guessed *Case A* *a*	Number of times *Case A* was selected and the "Guesser" guessed *Case B* *b*
	Case B actually selected	Number of times *Case B* was selected and the "Guesser" guessed *Case A* *c*	Number of times *Case B* was selected and the "Guesser" guessed *Case A* *d*

Assume that the null hypothesis for a test is:

H_0 : ***Case A*** was selected,

and that the alternative hypothesis is

H_a : ***Case B*** was selected.

The **parameter** in this study is the true probability of "success," which we will denote by p.

The hypotheses can be restated in terms of the parameter:

$$H_0 : p = \frac{1}{2}$$

and

$$H_a : p = \frac{1}{3}.$$

Let α be the empirical probability of Type I error, which is the observed proportion of the trials on which the null hypothesis was rejected when the null hypothesis was true. The formula for this empirical version of the significance probability is: $\alpha = \frac{b}{a+b}$.

The empirical power of the test is the observed proportion of rejections of the null hypothesis when the null hypothesis was false, which is Power = $\frac{d}{c+d}$.

A Summary of the Instructions for this Activity

"Roller" rolls the die once at the beginning.

- If even, proceed to *Case A*.
- If odd, skip to *Case B*.

Case A – Even Number

1, 2, or 3 = "success"
4, 5, or 6 = "failure"

Case B – Odd Number

1 or 2 = "success"
3, 4, 5, or 6 = "failure"

Data Recording Form

Guesser			Roller		
Initial die roll (1 to 6)	True case selected (A or B)	Number of rolls in each turn (n)	Number of successes ($1 \le x \le n$)	Number of failures ($n - x$)	Guessed case (A or B)

Statistical Power and the Graphing Calculator

Tom Short
thomas.short@villanova.edu
August 2001

Let X count the number of successes in a Binomial experiment, let p be the true probability of success on each trial, and let n be the number of trials.

Let $\hat{p}$ be the observed proportion of successes on n trials, so $\hat{p} = \frac{x}{n}$.

Suppose that the hypotheses being tested are:

$H_0 : p = \frac{1}{2}$ and $H_a : p < \frac{1}{2}$, and that the alternative value of particular interest is $H_a : p = \frac{1}{3}$.

Let α = Pr(Type I Error) = Pr(Reject H_0 when H_0 is true).

Let Power = 1 - Pr(Type II Error) = Pr(Reject H_0 when H_0 is false).

Program PPOWERA

1. Program PPOWERA requests values for the null hypothesis, a specific alternative value, the significance level α, and the number of trials in each sample, n.

2. The program generates simulated values for the number of successes in a Binomial experiment using the null hypothesis probability of success, in this case $p = \frac{1}{2}$. It displays the simulated numbers of successes in a histogram.

3. The program then generates and displays simulated numbers of successes from a Binomial experiment using the alternative value for p, in this case $\frac{1}{3}$.

4. Next, the two histograms corresponding to the null and alternative values of p are overlaid. Normal distributions that approximate the two histograms are superimposed on the graph as well.

5. The program then converts the two Normal distributions from the number of successes scale to the proportion of successes scale, but the distributions look exactly the same.

6. The program shades α, the probability of rejecting H_0 when H_0 is true, under the Normal distribution corresponding to the null hypothesis.

7. The program also shades the value of power, the probability of rejecting H_0 when H_0 is false, under the Normal distribution corresponding to the alternative hypothesis.

8. Finally, the program reports the value of power at the specific alternative value of interest.

Program PPOWERB

1. Given the value of the true probability of success that is specified in the null hypothesis H_0, the significance level α, and the sample size, n, program PPOWERB graphs a **power curve**.

2. Alternative values of the probability of success are plotted along the x-axis, and corresponding values of power are plotted along the y-axis.

3. This program allows the user to find the value of the power at various alternative values while keeping the sample size fixed.

Program PPOWERC

1. Program PPOWERC requests the null and alternative values of the probability of success p, the significance level α, and the sample size, n.

2. The program first computes the power at the specific value of the alternative hypothesis for the given sample size.

3. Then the program constructs a graph of the sample size, n, on the x-axis and the power at the specific value of the alternative hypothesis on the y-axis.

4. This program allows the user to determine the sample size necessary to achieve a particular level of power at the fixed alternative value of interest.

All of the explicit probability calculations in the programs rely on the large sample distribution of the observed proportion, $\hat{p}$.

When data are sampled in sets of n trials from a Binomial experiment with true probability of success p, then the statistic:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

follows an approximately Normal distribution with mean 0 and standard deviation 1.

The trick to calculating power is to understand whether the correct value of p in any particular calculation should be the null hypothesis value or the alternative hypothesis value.

S.4.2 2001

Beyond the Formula V

Tom Short
Department of Mathematical Sciences
Villanova University
800 Lancaster Ave.
Villanova, PA 19085-1699
thomas.short@villanova.edu
www.villanova.edu/~tshort/

August 2001

Activities for the Design of Experiments

1. Writing a Protocol for a Clinical Study

2. The Value of Randomization

Writing a Protocol for a Clinical Study

Tom Short
thomas.short@villanova.edu

August 2001

Choose one of the following common sayings. Write a **protocol** or plan for the design of a study to verify that the saying is true.

Do your best to design an **experiment** to explore the validity of the saying. If it is impossible to design an experiment, then design an **observational study**. As a last resort, write a plan for a **survey** to investigate the saying.

The evaluation of your design will be based on your ability to consider the following concepts and to incorporate them into the protocol:

1. Identify a specific **population** that you plan to examine.
2. Describe how a representative **sample** will be selected.
3. Establish a **baseline**, if the experiment will continue over time.
4. Include a **control group**, if possible, or explain why you cannot.
5. Include a **placebo treatment**, if possible, or explain why you cannot.
6. Assure the **blindness** of subjects, or if possible, the **double-blindness** of subjects and investigators, or explain why you cannot.
7. Use **randomization** and **matching** where appropriate.
8. List exactly which variables will you measure and compare.
9. Describe specifically which displays and statistics will you use to make comparisons and to reach a conclusion in your study.

Common Sayings

1. An apple a day keeps the doctor away
2. Babies tend to be born during a full moon
3. Blonds have more fun
4. Carrots improve vision
5. Chicken soup cures a cold
6. Crest prevents cavities
7. Early to bed, early to rise, keeps a person healthy, wealthy, and wise
8. Feed a cold and starve a fever
9. Left-handed people are more creative than right-handed people
10. Oat bran reduces cholesterol
11. Things go better with Coke
12. Toads cause warts
13. Think of your own common saying (but get the instructor's approval)

References

Short, T. H., and Pigeon, J. G. (1998) "Protocols and Pilot Studies: Taking Data Collection Projects Seriously" *Journal of Statistics Education* [Online], 6(1). www.amstat.org/publications/jse/v6n1/short.html

Talbot, M. (2000) "The Placebo Prescription" *The New York Times* Sunday, January 9, 2000, Section 6, p. 34.

The Value of Randomization

Tom Short
thomas.short@villanova.edu

August 2001

Consider an experiment intended to compare two hybrid varieties of corn, labeled **A** and **B**. An available field has been partitioned into 36 plots arranged in six rows and six columns. Our task is to consider the design of the experiment given the layout of the field. The **response variable** is the yield of corn measured in bushels per acre.

Convenience Assignment

With modern equipment it is fast and easy to plant adjacent plots with the same type of seed. So one design that is practical from the planting point of view is to plant the 18 plots on the left side of the field with variety **A** and the 18 plots on the right side of the field with variety **B**.

Here are the data that result from such a planting scheme:

A 130	A 149	A 139	B 155	B 137	B 145
A 149	A 133	A 152	B 131	B 147	B 136
A 141	A 156	A 137	B 146	B 132	B 148
A 150	A 142	A 155	B 136	B 152	B 133
A 139	A 155	A 139	B 147	B 137	B 153
A 155	A 138	A 150	B 137	B 145	B 136

Does one variety appear to have a larger mean yield than the other?

Systematic Assignment

In case there are differences in soil quality, drainage, or other characteristics in the field, it may be wise to mix up the varieties among the plots more than they were mixed in the convenience assignment.

One design that many people perceive to be "random" is to alternate planting variety **A** and then **B** across the rows of the field. This is called a systematic assignment, and beginning with variety **A**, it results in the following data:

A 130	B 137	A 139	B 155	A 149	B 145
B 137	A 133	B 140	A 143	B 147	A 148
A 141	B 144	A 137	B 146	A 144	B 148
B 138	A 142	B 143	A 148	B 152	A 145
A 139	B 143	A 139	B 147	A 149	B 153
B 143	A 138	B 138	A 149	B 145	A 148

Do the data resulting from the systematic assignment of varieties to plots provide evidence of a difference in the mean yields?

Random Assignment

We'll have to talk a little bit about how to make the assignments of varieties to plots at random, but you can use the table below to record whatever yields result from our random assignment.

Some constraints we may wish to consider for our randomization include assigning equal numbers (18) of plots to each of the two varieties and balancing the number of plots for each variety across each row and down each column.

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36

The Truth

For your reference, here is the true underlying distribution of corn yields for each of the two varieties across all of the plots in the field.

A 130 B 118	A 149 B 137	A 139 B 127	A 167 B 155	A 149 B 137	A 157 B 145
A 149 B 137	A 133 B 121	A 152 B 140	A 143 B131	A 159 B 147	A 148 B 136
A 141 B 129	A 156 B 144	A 137 B 125	A 158 B 146	A 144 B 132	A 160 B 148
A 150 B 138	A 142 B130	A 155 B 143	A 148 B 136	A 164 B 152	A 145 B 133
A 139 B 127	A 155 B 143	A 139 B 127	A 159 B 147	A 149 B 137	A 165 B 153
A 155 B 143	A 138 B 126	A 150 B 138	A 149 B 137	A 157 B 145	A 148 B 136

Reference

Stephenson, W. R., and Stern, H. (2000) "AP Statistics" *STATS*, 28, p. 23-27.

Other Interesting Activities for the Design of Surveys and Experiments

Anderson-Cook, C. M. and Dorai-Raj, S. (2001) "An Active Learning In-Class Demonstration of Good Experimental Designs" *Journal of Statistics Education* [Online], 9(1). www.amstat.org/publications/jse/v9n1/anderson-cook.html

Schwarz, C. J. (1997) "StatVillage: An On-Line, WWW-Accessible, Hypothetical City Based on Real Data for Use in an Introductory Class in Survey Sampling" *Journal of Statistics Education* [Online], 5(2). www.amstat.org/publications/jse/v5n2/schwarz.html