

## Power and Sample Size Determination

Bret Hanlon and Bret Larget

Department of Statistics  
University of Wisconsin—Madison

October 28, 2010

- To this point in the semester, we have largely focused on methods to analyze *the data that we have* without any regard to *from where the data comes*.
- *Design of Experiments* is the area of statistics that examines plans on how to gather data to achieve good (or optimal) inference.
- Here, we will focus on the question of sample size:
  - ▶ how large does a sample need to be so that a confidence interval will be no wider than a given size?
  - ▶ how large does a sample need to be so that a hypothesis test will have a low p-value if a certain alternative hypothesis is true?
- Sample size determination is just one aspect of good design of experiments: *we will encounter additional aspects next lecture*.

## Proportions

- Recall methods for inference about proportions: confidence intervals

### Confidence Interval for $p$

A  $P\%$  confidence interval for  $p$  is

$$p' - z^* \sqrt{\frac{p'(1-p')}{n'}} < p < p' + z^* \sqrt{\frac{p'(1-p')}{n'}}$$

where  $n' = n + 4$  and  $p' = \frac{X+2}{n+4} = \frac{X+2}{n'}$  and  $z^*$  is the critical number from a standard normal distribution where the area between  $-z^*$  and  $z^*$  is  $P/100$ . (For 95%,  $z^* = 1.96$ .)

## Proportions

- ... and hypothesis tests.

### The Binomial Test

If  $X \sim \text{Binomial}(n, p)$  with null hypothesis  $p = p_0$  and we observe  $X = x$ , the p-value is the probability that a new random variable  $Y \sim \text{Binomial}(n, p_0)$  would be at least as extreme (either  $P(Y \leq x)$  or  $P(Y \geq x)$  or  $P(|Y - np_0| \geq |x - np_0|)$  depending on the alternative hypothesis chosen.)

## Sample size for proportions

- On November 2, Wisconsin will hold an election with races for the U.S. Senate and the Governor.
- A news organization plans to take a final poll over the weekend.
- One population proportion of interest is the proportion of voters who will cast a ballot for incumbent senator Russ Feingold.
- Assuming that they can take a random sample of likely voters:

*How large of a sample is needed for a 95% confidence interval to have a margin of error of no more than 4%?*

## Calculation

- Notice that the margin of error depends on both  $n$  and  $p'$ , but we do not know  $p'$ .

$$1.96\sqrt{\frac{p'(1-p')}{n+4}}$$

- However, the expression  $p'(1-p')$  is maximized at 0.5; if the value of  $p'$  from the sample turns out to be different, the margin of error will just be a bit smaller, which is even better.
- So, it is conservative (in a statistical, not political sense) to set  $p' = 0.5$  and then solve this inequality for  $n$ .

$$1.96\sqrt{\frac{(0.5)(0.5)}{n+4}} < 0.04$$

- Show on the board why  $n > \left(\frac{(1.96)(0.5)}{0.04}\right)^2 - 4 \doteq 621$ .

## General Formula

### Sample size for proportions

$$n > \left(\frac{(1.96)(0.5)}{M}\right)^2 - 4$$

where  $M$  is the desired margin of error.

## Significance Level and Power

### Definition

The *significance level of a test* is the probability of rejecting the null hypothesis when it is true and is denoted  $\alpha$ .

### Definition

The *power* of a hypothesis test for a specified alternative hypothesis is  $1 - \beta$ , which is the probability of rejecting a specific true alternative hypothesis.

- Note that as there are many possible alternative hypotheses, for a single  $\alpha$  there are many values of  $\beta$ .
- It is helpful to plot the probability of rejecting the null hypothesis against the parameter.

## Example

- Consider a population with proportion  $p$ .
- Let  $X$  be the number of successes in a random sample of size 100
- with model  $X \sim \text{Binomial}(100, p)$ .
- Consider the hypotheses  $H_0: p = 0.3$  versus  $H_A: p < 0.3$  and the decision to reject if  $X \leq 22$ .
  - 1 Find  $\alpha$
  - 2 Find  $\beta$  if  $p = 0.2$ .
  - 3 Plot the probability of rejecting the null hypothesis versus  $p$ .

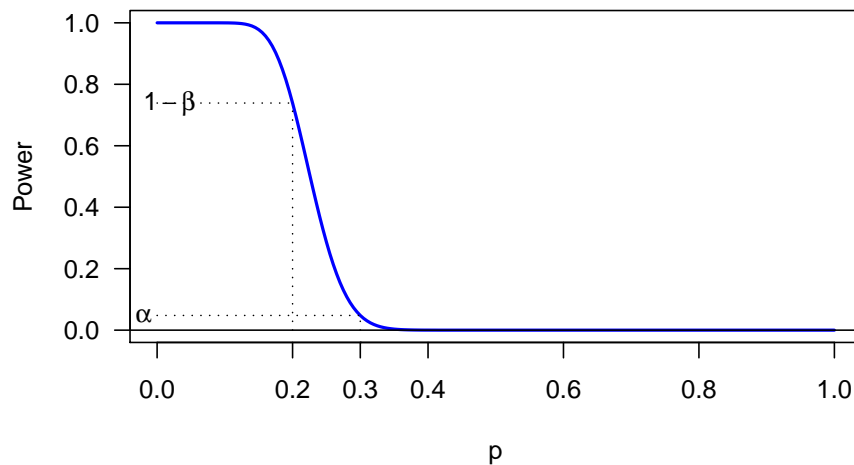
## Solution

$$\begin{aligned}\alpha &= P(X \leq 22 | p = 0.3) \\ &= \sum_{k=0}^{22} \binom{100}{k} (0.3)^k (0.7)^{100-k} \\ &\doteq 0.0479\end{aligned}$$

$$\begin{aligned}1 - \beta(p) &= P(X \leq 22 | p) \\ &= \sum_{k=0}^{22} \binom{100}{k} p^k (1-p)^{100-k}\end{aligned}$$

$$1 - \beta(0.2) = 0.7389$$

## Graph of Power



## Sample Size

- Suppose that we wanted a sample size large enough so that we could pick a rejection rule where  $\alpha$  was less than 0.05 and the power when  $p = 0.2$  was greater than 0.9.
- How large would  $n$  need to be?
- Simplify by letting  $n$  be a multiple of 25.
- We see in the previous example that 100 is not large enough: if the critical rejection value were more than 22, then  $\alpha > 0.05$  and the power is 0.7389 which is less than 0.9.
- There is no simple formula, but we can use R to find the answer by trial and error.

## Calculation

- Here we consider  $n = 150$  and find the best rejection  $k$  such that  $P(X \leq k | p = 0.3) < 0.05$  when  $n = 150$ . The answer is  $k = 35$ .

```
> cbind(30:40, pbinom(30:40, 150, 0.3))
```

```
      [,1]      [,2]
[1,] 30 0.003819432
[2,] 31 0.006657905
[3,] 32 0.011181720
[4,] 33 0.018114321
[5,] 34 0.028338450
[6,] 35 0.042860887
[7,] 36 0.062742795
[8,] 37 0.088996125
[9,] 38 0.122454316
[10,] 39 0.163633628
[11,] 40 0.212607595
```

## Calculation

- Here is a trickier way to do it.

```
> k = max(which(pbinom(1:150, 150, 0.3) < 0.05))
```

```
> k
```

```
[1] 35
```

- Find  $\alpha$  and  $1 - \beta$  for this value of  $k$ .

```
> alpha = pbinom(k, 150, 0.3)
```

```
> power = pbinom(k, 150, 0.2)
```

```
> c(alpha, power)
```

```
[1] 0.04286089 0.86831829
```

- The power is not quite high enough. Try  $n = 175$ .

## Calculation

- This code finds the largest value  $k$  such that  $P(X \leq k | p = 0.3) < 0.05$  when  $n = 175$ .

```
> k = max(which(pbinom(1:175, 175, 0.3) < 0.05))
```

```
> k
```

```
[1] 42
```

- Find  $\alpha$  and  $1 - \beta$  for this value of  $k$ .

```
> alpha = pbinom(k, 175, 0.3)
```

```
> power = pbinom(k, 175, 0.2)
```

```
> c(alpha, power)
```

```
[1] 0.04733449 0.91935706
```

- We see here that  $n = 175$  is large enough for a test with significance level less than 0.05 when the null  $p = 0.3$  is true will have power at least 0.9 to reject  $H_0$  when  $p = 0.2$ .

## Normal Populations

- The previous problems were for the binomial distribution and proportions, which is tricky because of the discreteness and necessary sums of binomial probability calculations.
- Answering similar problems for normal populations is easier.
- However, we need to provide a guess for  $\sigma$ .

## Example

- We want to know the mean percentage of butterfat in milk produced by area farms.
- We can sample multiple loads of milk.
- Previous records indicate that the standard deviation among loads is 0.15 percent.
- How many loads should we sample if we desire the margin of error of a 99% confidence interval for  $\mu$ , the mean butterfat percentage, to be no more than 0.03?

## Confidence Intervals

- We will use the  $t$  distribution, not the standard normal for the actual confidence interval, but it is easiest to use the normal distribution for planning the design.
- If the necessary sample size is even moderately large, the differences between  $z^*$  and  $t^*$  is tiny.
- Recall the confidence interval for  $\mu$ .

### Confidence Interval for $\mu$

A  $P\%$  confidence interval for  $\mu$  has the form

$$\bar{Y} - t^* \frac{s}{\sqrt{n}} < \mu < \bar{Y} + t^* \frac{s}{\sqrt{n}}$$

where  $t^*$  is the critical value such that the area between  $-t^*$  and  $t^*$  under a  $t$ -density with  $n - 1$  degrees of freedom is  $P/100$ , where  $n$  is the sample size.

## Calculation

- For a 99% confidence interval, we find  $z^* = 2.58$ .
- We need  $n$  so that

$$2.58 \times \frac{0.15}{\sqrt{n}} < 0.03$$

- Work on the board to show  $n > \left( \frac{(2.58)(0.15)}{0.03} \right)^2 \doteq 167$ .

## General Formula

### Sample size for a single mean

$$n > \left( \frac{(z^*)(\sigma)}{M} \right)^2$$

where  $M$  is the desired margin of error.

## Example

- Graph the power for a sample size of  $n = 25$  for  $\alpha = 0.05$  for  $H_0: \mu = 3.35$  versus  $H_A: \mu \neq 3.35$ .
- Note that the p-value is less than 0.05 approximately when

$$\left| \frac{\bar{X} - 3.35}{0.15/\sqrt{25}} \right| > 1.96$$

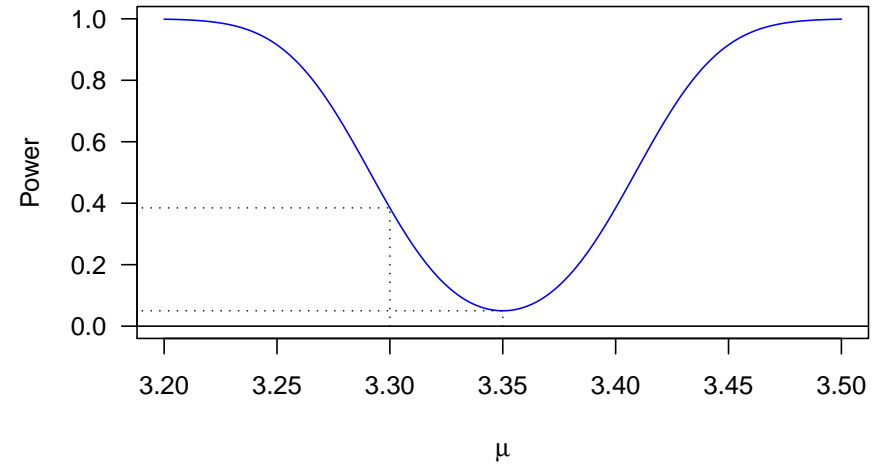
- The rejection region is

$$\bar{X} < 3.35 - 1.96 \frac{0.15}{\sqrt{25}} \doteq 3.291$$

or

$$\bar{X} > 3.35 + 1.96 \frac{0.15}{\sqrt{25}} \doteq 3.409$$

Power:  $n = 25, H_0: \mu = 3.35$



## Problem

- Long-run average percent butterfat in milk at a farm is 3.35 and the standard deviation is 0.15, with measurements taken by the load.
- How large should a sample size be to have an 80% chance of detecting a change to  $\mu = 3.30$  at a significance level of  $\alpha = 0.05$  with a two-sided test?
- We can explore this informally by graphing power curves for different  $n$  until we find a solution.
- Homework will guide you to an algebraic solution.

## What you should know

You should know:

- what the definitions of power and significance level are;
- how to find sample sizes for desired sizes of confidence intervals;
- how to find sample sizes with desired power for specific alternative hypotheses;
- how to examine and interpret a power curve;
- how power curves change as  $n$  changes.