STAT 40001/STAT 50001 Statistical Computing

Lecture 11

Department of Mathematics and Statistics

Outline

- **Sample size- single mean**
- **Sample size- single proportion**
- **Power Analysis**
- **Sample size-Two sample**

Sample Size Determination

Precision in the estimation of a parameter is quantified by the size of the bound of the error of estimation $|\hat{\theta} - \theta|$, or, equivalently, by the length of the CI for θ, which is twice the size of the error bound. A shorter error bound, or shorter CI, implies more precise estimation.
Sample size determination for μ

$$n \geq \left(z_{\frac{\alpha}{2}} \frac{\sigma}{Error} \right)^2$$

Example: The estimation of a new operating system's mean response time to an editing command should have an error bound of 5 milliseconds with 95% confidence. Experience with other operating systems suggests that preliminary/pilot standard deviation of 25 is a reasonable approximation to the population standard deviation. What sample size n should be used?

```
> library(BSDA)
> nsize(b=5, sigma=25, conf.level=0.95, type="mu")
```

Sample Size Determination- Proportion

Sample size determination for p

$$n \geq \frac{z^2_{\alpha/2}\hat{p}(1-\hat{p})}{Error^2}$$

If a prior estimate of p is unavailable, the sample size required is given by

$$n = 0.25\left(\frac{z_{\frac{\alpha}{2}}}{Error}\right)^2$$

Example: A new method of pre-coating fittings used in oil, brake, and other fluid systems in heavy-duty trucks is being studied for possible adoption. In this context, the proportion of such fittings that leak must be determined to within 0.01 with 95% confidence. What sample size is needed if a preliminary sample gave $\hat{p} = 0.9$?

```
> library(BSDA)
> nsize(b=0.01,  p=0.9, conf.level=0.95, type="pi")
```

The Power of a test

Power is one of the most important things in experimental design. When a hypothesis test does not reject the null hypothesis when it's false a type II error has been made. The power of the test is the probability of rejecting the null hypothesis when it is false, in other words that the test will not make a type II error. Thus power is defined as

$$Power = 1 - P(\text{Type II Error}) = P(\text{Reject } H_0 \text{ when } H_0 \text{ is false})$$

Since power is the probability of correctly rejecting the null hypothesis when it is false it makes sense that we would like this as large as possible. To do this, as shown in the formula above, we would like the probability of a type II error to be as small as possible.
There are two different aspects of power analysis.

- Calculate the necessary sample size for a specified power
- Calculate the power when given a specific sample size

Sample Size

Suppose we want to test $H_o : \mu = \mu_o$ Vs. $H_a : \mu > \mu_o$. Given α and β we want to find the sample size n and also we want to find the point K where the rejection begins. We know that

$$\alpha = P(\text{Type I error}) = P(\text{Reject} H_0 | H_0 \text{is true})$$

$$\begin{aligned} \alpha &= P(\overline{X} > K | \mu = \mu_0) \\ &= P\left(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{K - \mu_0}{\sigma/\sqrt{n}}\right) \\ &= P(Z > z_\alpha) \end{aligned}$$

And we know that

$$\beta = P(\text{Type II error}) = P(\text{Fail to Reject} H_0 | H_0 \text{is false})$$

Therefore,

$$\begin{aligned} \beta &= P(\overline{X} \leq K | \mu = \mu_a) \\ &= P\left(\frac{\overline{X} - \mu_a}{\sigma/\sqrt{n}} \leq \frac{K - \mu_a}{\sigma/\sqrt{n}}\right) \\ &= P(Z > -z_\beta) \end{aligned}$$

Where

$$z_\alpha = \frac{K - \mu_0}{\sigma/\sqrt{n}}$$

and

$$-z_\beta = \frac{K - \mu_a}{\sigma/\sqrt{n}}$$

Sample Size

Now solving

$$z_\alpha = \frac{K - \mu_0}{\sigma/\sqrt{n}}$$

and

$$-z_\beta = \frac{K - \mu_a}{\sigma/\sqrt{n}}$$

we gat

$$\mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} = \mu_a - z_\beta \frac{\sigma}{\sqrt{n}}$$

which yields,

$$\sqrt{n} = \frac{(z_\alpha + z_\beta)\sigma}{\mu_a - \mu_0}$$

Therefore,

$$n = \frac{(z_\alpha + z_\beta)^2\sigma^2}{(\mu_a - \mu_0)^2}$$

Power Analysis- Single Sample t-test

The power of the test is directly related to the number of individuals per group (n) , the amplitude of the differences we want to detect(also known as effect size), within group variability(σ) and the type I error (α). For t-tests, use the `power.t.test` function in R

```
power.t.test(n = NULL, delta = NULL, sd = 1, sig.level = 0.05,
             power = NULL,
             type = c("two.sample", "one.sample", "paired"),
             alternative = c("two.sided", "one.sided"),
             strict = FALSE)
```

Note that for a single sample test

$$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\mu_a - \mu_o)^2}$$

In the R code $\delta = \frac{\mu_a - \mu_o}{\sigma}$

More Reading: Cohen's method

Example

A company that manufactures light bulbs claims that a particular type of light bulb will last 850 hours on average with standard deviation of 50. A consumer protection group thinks that the manufacturer has overestimated the lifespan of their light bulbs by about 40 hours. How many light bulbs does the consumer protection group have to test in order to prove their point with reasonable confidence?

```
> power.t.test(d=40/50,power=0.9,sig.level=0.05,
type="one.sample",alternative="one.sided")

     One-sample t test power calculation

              n = 14.84346
          delta = 0.8
             sd = 1
      sig.level = 0.05
          power = 0.9
    alternative = one.sided
```

Power Analysis- Paired Sample t-test

A company markets an eight week long weight loss program and claims that at the end of the program on average a participant will have lost 5 pounds. On the other hand, you have studied the program and you believe that their program is scientifically unsound and shouldn't work at all. With some limited funding at hand, you want test the hypothesis that the weight loss program does not help people lose weight. Your plan is to get a random sample of people and put them on the program. You will measure their weight at the beginning of the program and then measure their weight again at the end of the program. Based on some previous research, you believe that the standard deviation of the weight difference over eight weeks will be 5 pounds. You now want to know how many people you should enroll in the program to test your hypothesis with power=0.8.

```
>library(pwr)
> pwr.t.test(d=(0-5)/5,power=0.8,sig.level=0.05,type="paired",
        alt="two.sided")
     Paired t test power calculation
              n = 9.93785
              d = 1
      sig.level = 0.05
          power = 0.8
    alternative = two.sided
NOTE: n is number of *pairs*
```

Power Analysis- Paired Sample t-test

A company markets an eight week long weight loss program and claims that at the end of the program on average a participant will have lost 5 pounds. On the other hand, you have studied the program and you believe that their program is scientifically unsound and shouldn't work at all. With some limited funding at hand, you want test the hypothesis that the weight loss program does not help people lose weight. Your plan is to get a random sample of people and put them on the program. You will measure their weight at the beginning of the program and then measure their weight again at the end of the program. Based on some previous research, you believe that the standard deviation of the weight difference over eight weeks will be 5 pounds. You now want to know how many people you should enroll in the program to test your hypothesis with power of the test 0.9.

```
>library(pwr)
> pwr.t.test(d=(0-5)/5,power=0.9,sig.level=0.05,type="paired",
             alt="two.sided")
     Paired t test power calculation

              n = 12.58547
              d = 1
      sig.level = 0.05
          power = 0.9
    alternative = two.sided

NOTE: n is number of *pairs*
```

Power Analysis- Paired Sample t-test

A company markets an eight week long weight loss program and claims that at the end of the program on average a participant will have lost 5 pounds. On the other hand, you have studied the program and you believe that their program is scientifically unsound and shouldn't work at all. With some limited funding at hand, you want test the hypothesis that the weight loss program does not help people lose weight. Your plan is to get a random sample of people and put them on the program. You will measure their weight at the beginning of the program and then measure their weight again at the end of the program. Based on some previous research, you believe that the standard deviation of the weight difference over eight weeks will be 5 pounds. You now want to know how many people you should enroll in the program to test your hypothesis with power of the test 0.9 at 0.01 level of significance.

```
>library(pwr)
> pwr.t.test(d=(0-5)/5,power=0.9,sig.level=0.01,type="paired",
        alt="two.sided")

     Paired t test power calculation
              n = 18.30346
              d = 1
      sig.level = 0.01
          power = 0.9
    alternative = two.sided

NOTE: n is number of *pairs*
```

Power Analysis- Paired Sample t-test

A human factors researcher wants to study the difference between dominant hand and the non-dominant hand in terms of manual dexterity. She designs an experiment where each subject would place 10 small beads on the table in a bowl, once with the dominant hand and once with the non-dominant hand. She measured the number seconds needed in each round to complete the task. She has also decided that the order in which the two hands are measured should be counter balanced. She expects that the average difference in time would be 5 seconds with the dominant hand being more efficient with standard deviation of 10. She collects her data on a sample of 35 subjects. The question is, what is the statistical power of her design with an N of 35 to detect the difference in the magnitude of 5 seconds.

```
>library(pwr)
> pwr.t.test(d=(0-5)/10,n=35,sig.level=0.01,type="paired",alt="two.sided")

     Paired t test power calculation
              n = 35
              d = 0.5
      sig.level = 0.01
          power = 0.5939348
    alternative = two.sided

NOTE: n is number of *pairs*
```

Power Analysis- Paired Sample t-test, one tailed

A human factors researcher wants to study the difference between dominant hand and the non-dominant hand in terms of manual dexterity. She designs an experiment where each subject would place 10 small beads on the table in a bowl, once with the dominant hand and once with the non-dominant hand. She measured the number seconds needed in each round to complete the task. She has also decided that the order in which the two hands are measured should be counter balanced. She expects that the average difference in time would be 5 seconds with the dominant hand being more efficient with standard deviation of 10. She collects her data on a sample of 35 subjects. The question is, what is the statistical power of her design with an N of 35 to detect the difference in the magnitude of 5 seconds.

```
>library(pwr)
> pwr.t.test(d=(0-5)/10,n=35,sig.level=0.01,type="paired",alt="less")

    Paired t test power calculation

              n = 35
              d = -0.5
      sig.level = 0.01
          power = 0.6961194  # note that the power has been increased
    alternative = less

NOTE: n is number of *pairs*
```

Power Analysis- Two sample t-test

A clinical dietician wants to compare two different diets, A and B, for diabetic patients. She hypothesizes that diet A (Group 1) will be better than diet B (Group 2), in terms of lower blood glucose. She plans to get a random sample of diabetic patients and randomly assign them to one of the two diets. At the end of the experiment, which lasts 6 weeks, a fasting blood glucose test will be conducted on each patient. She also expects that the average difference in blood glucose measure between the two group will be about 10 mg/dl. Furthermore, she also assumes the standard deviation of blood glucose distribution for diet A to be 15 and the standard deviation for diet B to be 17. The dietician wants to know the number of subjects needed in each group assuming equal sized groups.

We need the information below to perform the power analysis:

a) The expected difference in the average blood glucose; in this case it is set to 10.

b) The standard deviations of blood glucose for Group 1 and Group 2; in this case, they are set to 15 and 17 respectively.

c) The alpha level, or the Type I error rate. A common practice is to set it at the .05 level.

d) The pre-specified level of statistical power for calculating the sample size; this will be set to 0.8.

Power Analysis- Two sample t-test

Remark: Since what really matters is the difference, instead of means for each group, we can enter a mean of zero for Group 1 and 10 for the mean of Group 2, so that the difference in means will be 10. Next, we need to specify the pooled standard deviation, which is the square root of the average of the two standard deviations. In this case, it is $\sqrt{(15^2 + 17^2)/2} = 16.03$.

```
> pwr.t.test(d=(0-10)/16.03,power=.8,sig.level=.05,type="two.sample", a

     Two-sample t test power calculation

              n = 41.31968
              d = 0.6238303
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

Note that we could get the same results with different means with difference 10. For example, d=(10-20)/16.03

Power Analysis- Two sample t-test

Remark: Since what really matters is the difference, instead of means for each group, we can enter a mean of zero for Group 1 and 10 for the mean of Group 2, so that the difference in means will be 10. Next, we need to specify the pooled standard deviation, which is the square root of the average of the two standard deviations. In this case, it is $\sqrt{(15^2 + 17^2)/2} = 16.03$. Suppose we have total of 50 individuals we want to know corresponding power

```
> pwr.t.test(d=(0-10)/16.03,n=25,sig.level=.05,type="two.sample", alt="

     Two-sample t test power calculation

              n = 25
              d = 0.6238303
      sig.level = 0.05
          power = 0.5798042
    alternative = two.sided

NOTE: n is number in *each* group
```

Power Analysis- Tests of Proportions

We can calculate the power of test using pwr library and code below

```
pwr.2p.test(h = , n = , sig.level =, power = )
```

where h is the effect size and n is the common sample size in each group.

$$h = 2\arcsin(\sqrt{p_1}) - 2\arcsin(\sqrt{p_2})$$

Cohen suggests that h values of 0.2, 0.5, and 0.8 represent small, medium, and large effect sizes respectively.

For unequal n's use

```
pwr.2p2n.test(h = , n1 = , n2 = , sig.level = , power = )
```

To test a single proportion use

```
pwr.p.test(h = , n = , sig.level = power = )
```

Example

```
> library(pwr)
> pwr.2p.test(h =0.2,  sig.level =0.05, power =0.8 )
 Difference of proportion power calculation
              h = 0.2
              n = 392.443
      sig.level = 0.05
          power = 0.8
```

Example: Power Analysis- Tests of Proportions

Suppose two different sample reveal $p_1 = 0.299$ and $p_2 = 0.249$. What sample sizes should we need if we want a power of 90% to test the equality of proportion.

```
> power.prop.test(p1=.299, p2=.249, sig.level=.05, power=.9,
+                 alternative="two.sided")

     Two-sample comparison of proportions power calculation

              n = 1670.065
             p1 = 0.299
             p2 = 0.249
      sig.level = 0.05
          power = 0.9
    alternative = two.sided

NOTE: n is number in *each* group
```

Effect Size Vs. Sample size

Create a table showing the effect size(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0) versus sample size

```
> library(pwr)
> samplesize=cbind(NULL,NULL)
> for (i in c(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0))
+ {
+ power=pwr.t.test(d=i,power=0.8,sig.level=0.05,
+ type="two.sample",alt="two.sided")
+ samplesize=rbind(samplesize,cbind(power$d,power$n))
+ }
> samplesize
      [,1]       [,2]
 [1,]  0.1 1570.73305
 [2,]  0.2  393.40570
 [3,]  0.3  175.38467
 [4,]  0.4   99.08032
 [5,]  0.5   63.76561
 [6,]  0.6   44.58577
 [7,]  0.7   33.02457
 [8,]  0.8   25.52458
 [9,]  0.9   20.38631
[10,]  1.0   16.71472
>
```