Chapter 6
Compare Two Groups

Section 6.5:
Power, Margin of Error and Sample Size

© 2014 by David W. Gerbing

School of Business Administration
Portland State University

- Power, Margin of Error and Sample Size
  - Power Curve
  - Needed Sample Size for Confidence Interval
    - Procedure
    - Application

6.5a
Power Curve

Power Analysis for Comparing Means

Run a power analysis after failing to reject the null

- Power analysis for the hypothesis test of a mean difference is very similar to the power calculation for a single mean
- **Key Concept**: After a failure to detect an alternative value from the null of the population mean, μ, or mean difference, $\mu_1 - \mu_2$, evaluate power against potential alternative values
- Power calculations are not from data, but are an electronic look-up of values from a "non-central t-distribution"
- The two groups can have different samples sizes, $n_1 \neq n_2$, and standard deviations, $s_1 \neq s_2$, yet the power analysis for the mean difference requires only one value of n and of s
  - For the one standard deviation, use the within-group standard deviation, s_w
  - For the one sample size, use the harmonic mean of the two sample sizes, described next

Combine Sample Sizes with the Harmonic Mean

One sample size

- The calculation of power for a t-test over both groups requires just one sample size, applied to both groups
- **Harmonic mean**:

$$\tilde{n} = \frac{2}{\frac{1}{n_1} + \frac{1}{n_2}}$$

- The desirable property of the harmonic mean, $\tilde{n}$, is that it is closer to the smaller sample size instead of evenly splitting the distance as with the usual arithmetic mean, $\bar{n}$
  - $n_1 = 50, n_2 = 50: \quad \bar{n} = 50, \ \tilde{n} = 50$
  - $n_1 = 30, n_2 = 70: \quad \bar{n} = 50, \ \tilde{n} = 42$
  - $n_1 = 10, n_2 = 90: \quad \bar{n} = 50, \ \tilde{n} = 18$
- Power analysis proceeds from the harmonic mean calculated from the actual sample sizes of the two samples

Need Range of Values for Alternative Values of $\mu_1 - \mu_2$

Obtain the power curve after failing to reject the null

- Compare Salary (1000's) for Men and Women in a company

```
Salary for Gender Males:
  n = 18,   mean = 78.1,   sd = 4.25
Salary for Gender Females:
  n = 24,   mean = 76.5,   sd = 5.87
```

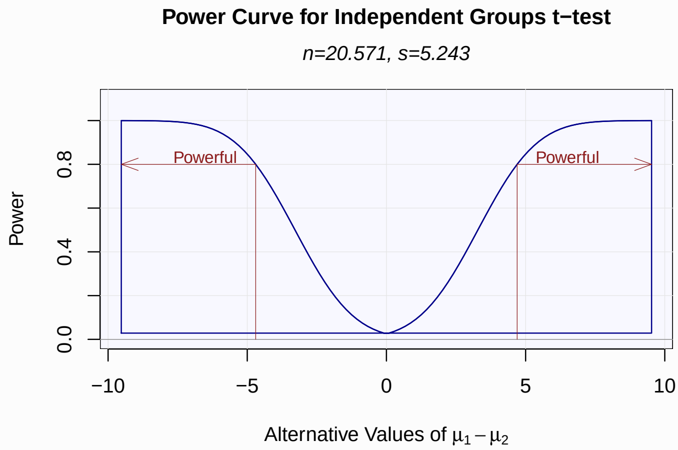
- The hypothesis test of $H_0: \mu_M - \mu_F = 0$ yielded

```
Hypothesis Test of 0 Mean Diff:
  t = 0.98,  df = 40,  p-value = 0.334
```

- No difference detected, but what if there really is a difference?
- The true value of $\mu_1 - \mu_2$ is not known, so obtain the power of a range of alternative values of the population mean difference to derive a power curve
- Use function ttestPower from the lessR package

ttestPower: Graphic Output

```
> ttestPower(n1=18, n2=24, s1=4.25, s2=5.87)
```



ttestPower: Text Output

```
---------------------------------------------------
Power Curve Analysis for Independent Groups t-test
---------------------------------------------------

Harmonic mean of the two sample sizes: n = 20.57
Within-group (pooled) Standard Deviation:  sw = 5.24
---------------------------------------------------

Mean difference to achieve power of 0.8: Diff = 4.69
---------------------------------------------------
```


Power Curve Interpretation

What values of the mean difference have low or high power?

- Minimum power is usually set at 0.8, here obtained with a mean difference of around $+\$4,690$ or $-\$4,690$
- There is an 80% or larger chance that an actual population mean difference greater than about $+\$4,690$ or less than $-\$4,690$ would have been properly detected by rejection of the null hypothesis of no difference
- For a mean difference of about $3000 or less, in either direction, there is more than a 50% chance of committing a Type II error, which is a failure to detect the real difference

Minimum Mean Difference of Practical Importance

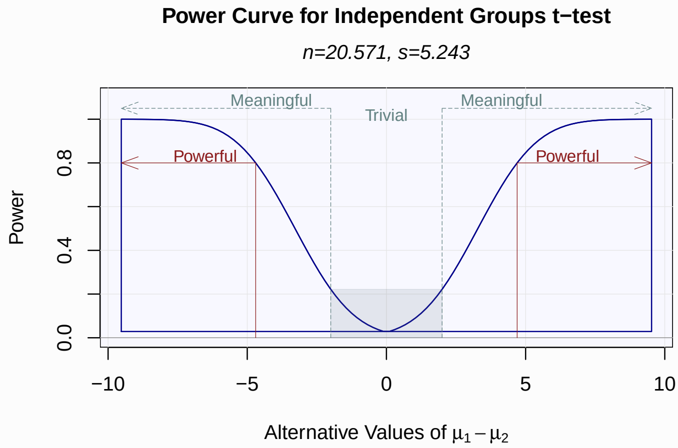
How large does the difference need be before important?

- **Minimum Mean Difference** of practical importance (mmd): Smallest difference from the null value, usually 0, considered to be of practical importance
- When no difference from the null value is detected, the question becomes, "Could a real, meaningful difference in fact exist, so that the outcome of the analysis is incorrect?"
- The power analysis provides the probabilities of detecting various meaningful differences
- Suppose that management deemed a population difference of average salary between men and women of $2000 to be the smallest difference of practical importance, that is, meaningful
- To annotate the power curve, use the minimum mean difference option, `mmd`, with the `lessR` `ttestPower` function

© 2014 by David W. Gerbing

ttestPower: Graphic Output

```
> ttestPower(n1=18, n2=24, s1=4.25, s2=5.87, mmd=2)
```



© 2014 by David W. Gerbing

Analysis of Power Curve with Meaningful Changes

What changes are detected and which are meaningful?

- Text output of this `ttestPower` analysis includes

```
Given n = 20.57 and sw = 5.24, for mmd of 2:
      Power = 0.222
Warning: Meaningful differences, from 2 to 4.696,
      have Power < 0.8
```

- If a meaningful difference of $2000 did exist from the null value of 0, then there is a low probability, 0.22, of detecting this difference
- So maybe a meaningful difference does exist, $2000 and larger, and this test failed to detect it
- If a more definitive conclusion is desired, the only alternative is to gather more data, of which `ttestPower` provides the relevant information

© 2014 by David W. Gerbing

Needed Sample Size for Desired Power

More information from ttestPower

- Text output of this ttestPower analysis, because of the specification of mmd=2 for a minimal mean difference of $2000, also includes

```
Given sw, needed n to achieve power= 0.8
       for mmd of 2: n = 109
Sample size n applies to *each* group
```

- Achieve an 80% probability of correctly detecting a true difference for the mmd of $2000 with a sample size of $n = 109$ for *each* sample
- This needed value of 109 is a considerable increase over the initial sample sizes of $n_1 = 18$ and $n_2 = 24$

6.5b
Needed Sample Size for Confidence Interval

Desired Margin of Error

Distinguish between what you got and what you want

- Question: How large a sample for *each* group is needed to obtain a smaller, desired margin of error?
- **Key Concept**: Calculate the needed sample size for each group, n_{needed}, that will, with a .90 probability, obtain the desired margin of error, $E_{desired}$, for the new 95% confidence interval

---

- E is the margin of error *obtained* with the group sample size of n
- $E_{desired}$ is the margin of error *desired*, which requires an increase of the current n to the larger value n_{needed}

Procedure

Overview of Sample Size Procedure

Move from initial sample to final, larger sample

- Specify the desired margin of error (precision), $E_{desired}$
- Obtain initial data samples
- Calculate margin of error, E, then the confidence interval
- If $E_{desired} < E$, calculate n_{needed}
- Gather new data for larger samples
- Re-calculate the margin of error, E, and the confidence interval

Two-step Procedure to Calculate the Needed Sample Size

Two-step procedure for obtaining needed sample size

- 1. From the initial sample, get the standard deviation, s_w, and then calculate preliminary estimate of sample size, n_s :

$$n_s = 2\left[\frac{(1.96)(s_w)}{E_{desired}}\right]^2$$

- 2. Revise the initial estimate upward to obtain the estimate actual of sample size: s may underestimate σ, so revise n_s upward for a given probability[1] of obtaining $E_{desired}$ with a 95% level of confidence
  .70 probability: $n_{needed} = 1.039n_s + 2.291$
  .90 probability: $n_{needed} = 1.099n_s + 4.863$ ▷ Most often used
  .99 probability: $n_{needed} = 1.175n_s + 7.526$

[1] These coefficients come from analysis of a paper by Kupper and Hafner in the *American Statistician*, 43(2):101-105, 1989

Application

Reconsidering Ship Times

Reduce the margin of error

- Previously constructed a confidence interval about the mean difference of delivery time for two different suppliers
- Supplier 2 was shown faster, on average, than Supplier 1
- Problem: Obtained margin of error was large, $E = 1.062$ days
- One manager prefers Suppler 1 for other reasons, so if Supplier 2 is not that much faster, the manager would prefer to retain Supplier 1
- How much faster is Suppler 2, on average?
- Answer to within a margin of error of half a day
- Managerial Question: How many shipments for each group need be sampled to reach a .90 probability of obtaining a margin of error of half a day or less at the 95% confidence level of the mean difference?

Analysis of Sample Size

Step 1: Initial sample size of each group

- From previous analysis of the confidence interval of the mean difference, $s_w = 1.51$, with $E_{desired}$ specified as 0.5

$$n_s = 2\left[\frac{(z_{.025})(s_w)}{E_{desired}}\right]^2 = 2\left[\frac{(1.96)(1.51)}{0.5}\right]^2 = 70.06$$

Step 2: Upward adjusted actual sample size of each group

- $n_{needed} = 1.099 n_s + 4.863$
  $= 1.099(70.06) + 4.863$
  $= 81.86$
- Now round up, 81.86 to: $n_{needed} = 82$

Results

82 shipments per group

- Now gather measurements of Ship Time for 82 shipments for each supplier, a dramatic increase from the 15 and 19 shipments currently in each group
- There is a .90 probability that when the revised 95% confidence interval is calculated over all 82 shipments in each group, the resulting margin of error will be 0.5 or less

Conclusion

Potential problems

- 82 shipments per supplier are not available
- The older shipments are from so long ago the underlying process has changed, rendering many of them invalid for estimating the *current* population mean shipping time, μ

Extent of knowledge

- The previous inferential analysis demonstrated that, on average, Supplier 2 is faster than Supplier 1
- And, moreover, anywhere from about $\frac{1}{2}$ to $2\frac{1}{2}$ days faster
- Obtaining more information may simply not be practical, but fortunately, much is already known

- The End