Ideas for Introducing Power in the Service Statistics Course

Alan T. Arnholt and Jose Almer T. Sanqui, Appalachian State University, Boone, NC, USA

Introduction

Power is one of the more important and least covered topics in an Introductory Statistics course. This poster shows how power, even with non-central distributions, can be covered for students in a basic statistics course.

Prerequisite Concepts

Once a student understands just a few concepts, the *power of a test* can be introduced.

- $\alpha = \mathbb{P}(\text{type I error}) = \text{level of significance} = \mathbb{P}(\text{reject } H_0 | H_0 \text{ is true}) = \mathbb{P}(\text{accept } H_1 | H_0 \text{ is true})$.
- $\beta = \mathbb{P}(\text{type II error}) = \mathbb{P}(\text{fail to reject } H_0 | H_0 \text{ is false}) = \mathbb{P}(\text{accept } H_0 | H_1 \text{ is true})$.
- Given a composite alternative hypothesis $H_1 : \theta \in \Theta_1$, $\textit{Power}(\theta) = \mathbb{P}(\text{reject } H_0 | H_0 \text{ is false}) = \mathbb{P}(\text{accept } H_1 | H_1) = 1 - \beta(\theta)$, where $\beta(\theta)$ is the probability of a type II error at a given θ.

Normal Distribution and Power

Computing the power for a particular alternative or finding the power function when working with normal distributions is covered in most texts and is easily done with R.

Problem: Given a normal distribution with unknown mean μ and known standard deviation $\sigma = 3$, for a test of the null hypothesis $H_0 : \mu = 40$ versus the alternative hypothesis $H_1 : \mu = 48$ using an α level of 0.05

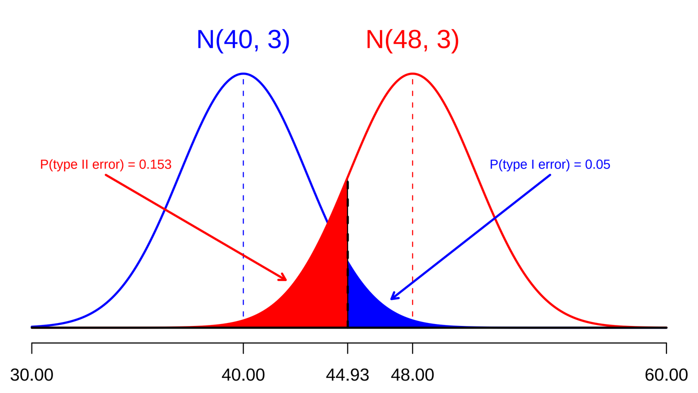
A. With a sample of size one, compute the probability of a type II error.

B. Graph the $\textit{Power}(\mu)$ for values of μ from 25 to 55 for testing a two tailed alternative hypothesis using samples of size one and nine.

Answers:

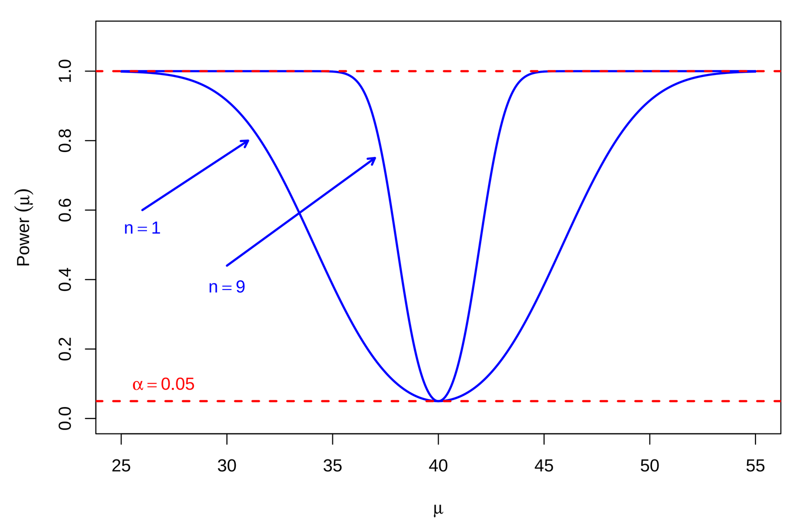
A. To find the the probability of a type II error, use the R commands

```
> cv <- qnorm(0.95, 40, 3)
> typeIIerror <- pnorm(cv, 48, 3)
> typeIIerror
[1] 0.1534347
```



B. To graph the $\textit{Power}(\mu)$ for values of μ from 25 to 55 for testing a two tailed alternative hypothesis using samples of size one and nine, use the R commands:

```
> mu <- seq(25, 55, 0.01)
> powerONE <- pnorm(qnorm(.025,40,3), mu, 3) +
+ pnorm(qnorm(0.975,40,3),mu,3,lower=FALSE)
> powerNINE <- pnorm(qnorm(.025,40,1), mu, 1) +
+ pnorm(qnorm(0.975,40,1),mu,1,lower=FALSE)
> plot(mu, powerONE, type="l", lwd=2, col="blue",
+ ylim=c(0,1.1))
> lines(mu, powerNINE, col="blue", lwd=2)
```



Binomial Distribution and Power

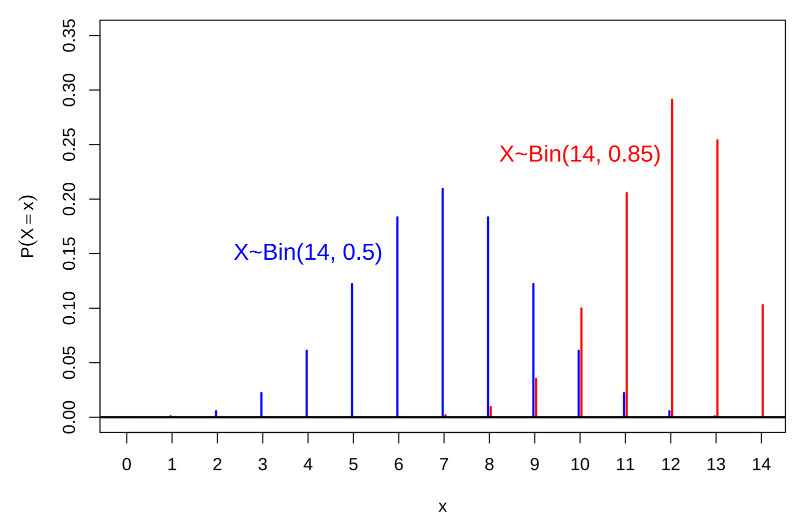
So that students do not think that all power can be computed by shifting a **central** sampling distribution either to the right or left, introduce power using the binomial distribution.

Problem: Suppose $X \sim \textit{Bin}(n = 14, \theta = 0.5)$. Determine the $\textit{Power}(\theta = 0.85)$ when testing $H_0 : \theta = 0.5$ versus $H_1 : \theta > 0.5$ with $\alpha = 0.0897$ (a decision to reject the null hypothesis when $X > 9$).

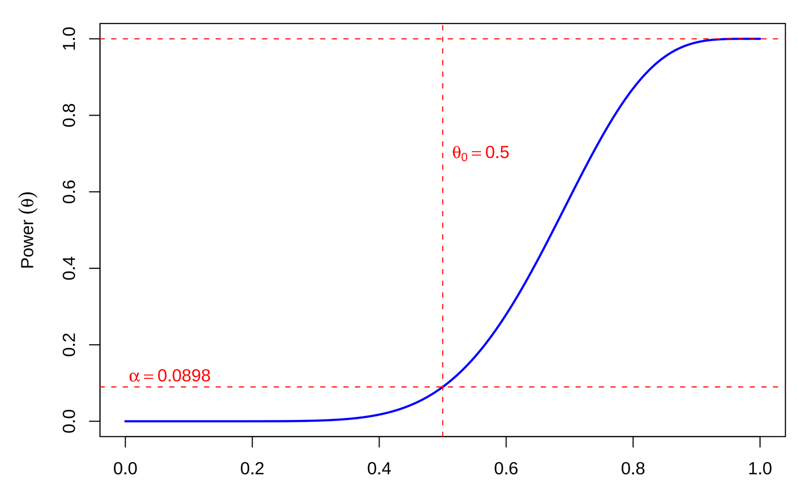
Answer: By graphing a $X \sim \textit{Bin}(n = 14, \theta = 0.5)$ and a $X \sim \textit{Bin}(n = 14, \theta = 0.85)$, as shown below, it is very easy for the students to visualize the $\textit{Power}(\theta = 0.85)$ with an asymmetric distribution.

To find the $\textit{Power}(\theta = 0.85)$, use the R commands

```
> POWER <- sum(dbinom(10:14, 14, 0.85))
> POWER
[1] 0.9532597
```



With a little creativity, students can write a few lines of code to create a graph like the one below showing the $\textit{Power}(\theta)$.



Central and Non-central t Distribution

Simulation is an effective way to reinforce the concept of a sampling distribution.

Central t Distribution

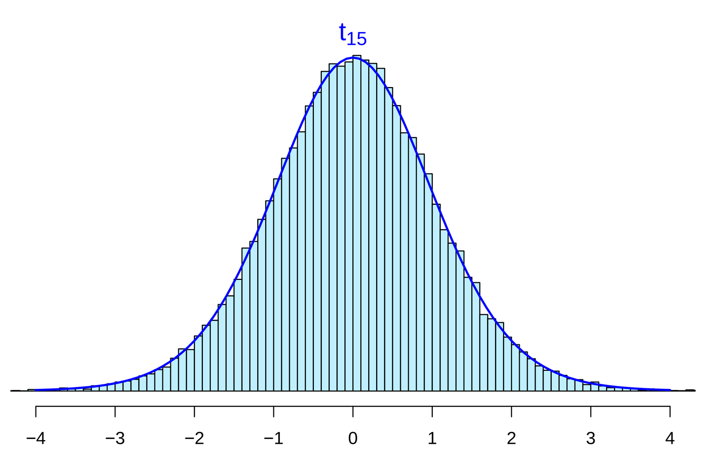
Example: Have students simulate the quantity

$$\text{tc} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

when sampling from a normal distribution. Compare the quantiles of the simulated sampling distribution versus the theoretical quantiles of a t_{n-1}.

The R simulation of 50,000 samples of size 16 from a normal distribution with mean of 100 and standard deviation of 20 is in the online script.

A density histogram of the quantity tc with a superimposed density of a t_{15} along with the theoretical and simulated quantiles suggest the simulation is a quite accurate representation of a t_{15} distribution:



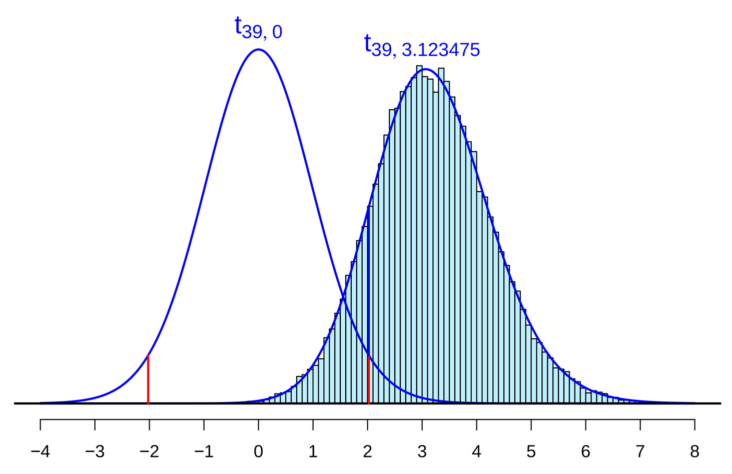
Non-central t Distribution

To introduce the non-central t distribution with non-centrality parameter γ $(t^{\star}_{\nu;\gamma})$, have the students simulate the quantity

$$\text{tnc} = \frac{(\bar{Y}_{1\bullet} - \bar{Y}_{2\bullet})}{S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}. \tag{1}$$

The online script takes $m = 50,000$ samples from $N(\mu_1 = 120, \sigma_1 = 20)$ of size $n_1 = 16$ and from $N(\mu_2 = 100, \sigma_2 = 20)$ of size $n_2 = 25$.

The simulated values are displayed in a density histogram and a non-central t with non-centrality parameter $\gamma = 3.123475$ is superimposed over the simulated values.



The simulated values are counted to compute $\textit{Power}(\mu_1 - \mu_2 = 20)$ for $H_0 : \mu_1 - \mu_2 = 0$ versus $H_1 : \mu_1 - \mu_2 \neq 0$ at the $\alpha = 0.05$ level (critical values shown in red above).

$$\widehat{\textit{Power}}(\mu_1 - \mu_2 = 20) = \frac{\#\left[(\text{tnc} > t_{0.975,39}) \cup (\text{tnc} < t_{0.025,39})\right]}{m} = 0.862$$

This agrees well with the theoretical power of the test (0.861).

Power could also be approximated using the simulation approach when the variances for the two populations are unknown and unequal (Behrens-Fisher problem).

The Non-centrality Parameter

For the non-centrality parameter, γ, and the t statistic, t,

$$\gamma = \frac{(\mu_1 - \mu_2)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2}}{\sigma} \quad \text{and} \quad t = \frac{(\bar{Y}_{1\bullet} - \bar{Y}_{2\bullet})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2}}{S_p}, \tag{2}$$

t measures the statistical differences between the **sample** means and γ is used to measure the statistical differences between the **population** means.

Squaring both quantities in (2), gives

$$F = t^2 = \frac{(\bar{Y}_{1\bullet} - \bar{Y}_{2\bullet})^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1}}{S_p^2} = \frac{\textit{MS}_{\text{Treatment}}}{\textit{MS}_{\text{Error}}}$$

and

$$\lambda = \gamma^2 = \frac{(\mu_1 - \mu_2)^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1}}{\sigma^2} = \frac{\textit{SS}_{\text{Hypothesis}}(\text{population})}{\sigma^2}$$

where $\textit{SS}_{\text{Hypothesis}}(\text{population})$ is the sum of squares for treatments obtained by replacing $\bar{Y}_{1\bullet}$ with μ_1, $\bar{Y}_{2\bullet}$ with μ_2, and $\bar{Y}_{\bullet\bullet}$ with $\frac{n_1\mu_1 + n_2\mu_2}{n_1 + n_2}$.

When λ is the ratio of $\textit{SS}_{\text{Hypothesis}}(\text{population})$ to σ^2, the calculation of λ is straightforward: The $\textit{SS}_{\text{Hypothesis}}(\text{population})$ will always be the sum of squares formula for the H_0 being tested.

This method of computing λ extends to any hypothesis the user would like to test. It is not limited merely to the equality of treatment means nor to equal sample sizes.

To compute the power of the test when $\mu_1 - \mu_2 = 20$, $\sigma_1 = \sigma_2 = 20$, $n_1 = 16$, and $n_2 = 25$ using a two-sided alternative with $\alpha = 0.05$, compute the non-centrality parameter to be

$$\gamma = \frac{(\mu_1 - \mu_2)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2}}{\sigma} = \frac{(120 - 100)\left(\frac{1}{16} + \frac{1}{25}\right)^{-1/2}}{20} = 3.123475.$$

The power of the test is then

$$\begin{aligned}
\textit{Power}(\mu_1 - \mu_2 = 20) &= \mathbb{P}(\text{Reject } H_0 | H_1) \\
&= \mathbb{P}\left((T < t_{\alpha/2;n_1+n_2-2}) \middle| T \sim t^{\star}_{n_1+n_2-2;\gamma}\right) + \\
&\quad \mathbb{P}\left((T > t_{1-\alpha/2;n_1+n_2-2}) \middle| T \sim t^{\star}_{n_1+n_2-2;\gamma}\right) \\
&= \mathbb{P}\left((t^{\star}_{39;3.123475} < t_{0.025;39})\right) + \mathbb{P}\left((t^{\star}_{39;3.123475} > t_{0.975;39})\right) \\
&= \mathbb{P}\left((t^{\star}_{39;3.123475} < -2.022691)\right) + \mathbb{P}\left((t^{\star}_{39;3.123475} > 2.022691)\right) = 0.8612027
\end{aligned}$$

R commands

To find $\textit{Power}(\mu_1 - \mu_2 = 20)$ with R, one can use the standard commands `pt()` and `qt()` as follows:

```
> cvl <- qt(0.025, 39)
> cvu <- qt(0.975, 39)
> Power <- pt(cvl, 39, 3.123475) +
+ pt(cvu, 39, 3.123475, lower.tail=FALSE)
> Power
[1] 0.8612027
```

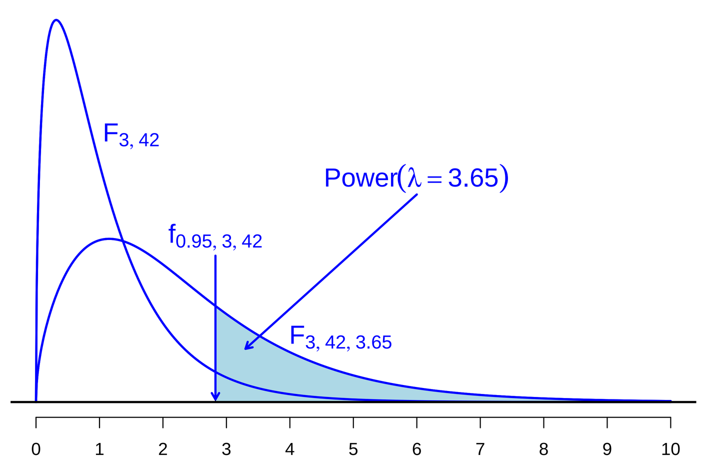
The function `power.t.test()` will return power for one- and two-sample t tests (when each sample is the same size) and `power.anova.test()` will return the power for one-way analysis of variance when the sample sizes are equal. It would not be hard to modify the current code to either function to accommodate unequal sample sizes as special cases.

Non-central F Distribution

Suppose the true mean grade for University A's students using teaching methods 1, 2, 3, and 4 have means of 71, 73, 75, and 80 with a common standard deviation of $\sigma = 12$. If $n_1 = 11$, $n_2 = 13$, $n_3 = 10$, and $n_4 = 12$, determine the probability a difference among the means will be detected using $\alpha = 0.05$.

$$\begin{aligned}
\lambda &= \frac{\textit{SS}_{\text{Hypothesis}}}{\sigma^2} = \frac{\sum_{i=1}^{4} n_i(\mu_{i\bullet} - \bar{\mu}_{\bullet\bullet})^2}{\sigma^2} \\
&= \frac{11(71 - 74.78)^2 + 13(73 - 74.78)^2 + 10(75 - 74.78)^2 + 12(80 - 74.78)^2}{12^2} \\
&= 3.65157
\end{aligned}$$

```
> Power <- 1 - pf(qf(0.95, 3, 42), 4, 42, 3.65157)
> Power
[1] 0.2204405
```



Conclusion

With appropriate examples, power can be covered with R in an introductory statistics course. When students use simulation, they gain an intuitive understanding of power, even for non-central distributions.

References

Ugarte, M. D., Militino, A. F., and Arnholt A. T. 2008. *Probability and Statistics with R.* Chapman & Hall/CRC. Boca Raton, FL.

†An R script for all graphs in the poster is available at http://www1.appstate.edu/ arnholta/UseR2009