ENV 710 – Applied Data Analysis
Fall 2014
Lab 5: T-tests and Power Analysis

The goal of this lab is to practice some of the concepts that we have been studying in lecture over the last two weeks, including testing for normality, transforming data, conducting two-sample tests, and understanding p-values. The emphasis of the lab is on understanding Type I and Type II error, statistical power, and the elements of study design that affect power.

After this lab, you should be able to:

- Conduct t-tests two-sample and paired t-tests
- Appropriately report statistics and p-values
- Conduct power analysis for t-tests.

For this lab, there are four questions to answer. Please type your answers to each problem, including any requested graphs, in a Word document. *Submit your answers and your **R**-code to the class Sakai site under the folder Assignments before 5 pm on either Mon., Sep. 29 (Section 07) or Wed., Oct., 1 (Sections 01 and 02).*

More functions in R

In this lab, we introduce a few new **R** commands.

`array()`- An array is a vector that can have one or two dimensions. A two dimensional array is the same as a matrix. Within array, `dim` gives the maximum indices in each dimension.
`ceiling ()`- rounds up to the nearest integer.
`floor ()` - rounds down to the nearest integer.
`dim ()`- determines the dimensions of a data frame or array.
`lines()`- adds lines to an existing plot.
`points()`- adds points to an existing plot.
`abline()`- adds one or more straight lines through the current plot. The line can be specified by providing the intercept and slope.
`legend()`- adds a legend to a plot, by giving the x, y coordinates for the legend position and supplying the legend content (see graphing details `?legend`).

Here are some examples using the above functions.

```
array(1:50, dim = c(5,10), dimnames = c("a", "b"))
array(1:24, dim=c(3,4,2))

a <- matrix(8, 2, 3)
b <- matrix(9, 2, 3)
my.array <- array(c(a,b), c(2,3,2))
```

```
dim(my.array)
a[1]
b[2,]
```

What do these functions do to this vector of data?

```
ceiling(c(4.2, 5.5, 6.2))
floor(c(4.2, 5.5, 6.2))
```

We will make a quick (meaningless) plot using the power law (not to be confused with statistical power) to employ `lines()`, `abline()` and `points()`.

```
x <- c(1:20)
a <- 3
b <- 2
plot(x, y=a*x^b, xlab = "Counts", ylab = "Response",
las=1, pch=21, bg="darkblue")

lines(x, y=a*x^b, lwd=2, col="darkblue")
abline(h=300, lty=2)
abline(v=10, lty=2)
points(10, 300, pch=21, cex=2, col="red", bg ="red")
legend(4, 1000, c("target point"), pt.bg="red", pch=21,
pt.cex=2)
```

Understanding Power Analysis

Statistical tests are typically used to decide between two hypotheses: the null hypothesis and the alternative hypothesis. We generally want statistical tests to provide us with evidence that we can reject the null, and conclude that the impact or factor that we are studying had an effect. Of course, we are really interested in the strength of the effect, not just whether there is a difference between the null and the alternative hypothesis. With any statistical test, however, there is always the possibility that we will find a difference between groups when one does not actually exist. This is called a *Type I* error, and it is denoted by α. Likewise, it is possible that when a difference does exist, the test does not detect it. This type of mistake is called a *Type II* error, and it is denoted by β.

Statistical *power* is the ability to detect a statistically significant difference when the null hypothesis is false. It is your ability to find a difference when the difference exists. The power of an experiment or study depends on the sample size, the effect size, and the alpha level. Generally speaking, as your sample size increases, so does the power of your test. This should intuitively make sense as a larger sample means that you have collected more information -- which makes it easier to correctly reject the null hypothesis when you should.

Power analyses are often used prior to a study or experiment to determine an appropriate sample size. In this case, if we know the power of our test (or our desired power), we can find any of the other three quantities (given that we know the other two).

Knowing the desired power and alpha is pretty easy, but the *effect size* might be more challenging. An effect size is a quantitative measure of the strength of a phenomenon.

A common measure of effect size is Cohen's *d*, which can be used when comparing two means, such as when you do a t-test. It is calculated by taking the difference between the two groups (e.g., the mean of treatment group *minus* the mean of the control group) and dividing it by the pooled standard deviation: $d = \frac{|\mu_1 - \mu_2|}{\sigma_p}$, where $\sigma_p = \sqrt{(\sigma_1^2 + \sigma_2^2)/2}$. Thus, Cohen's *d* simply scales the difference in means by the number of standard deviations. With a *d* of 1, we know that the two groups' means differ by one standard deviation; a *d* of 0.5 tells us that the two groups' means differ by half a standard deviation; and so on. Cohen (1977, 1988) justifies three levels of effect sizes (small, medium, and large) for different types of statistical tests (see table below). Following this logic, if two groups' means differ by 0.2 standard deviations or less, the difference is trivial, even if it is statistically significant. Note that *d* can be larger than 1 if there is a really big difference between two means.

	Effect size Index	Small	Medium	Large
t-test on means	d	0.20	0.50	0.80
t-test on correlations	r	0.10	0.30	0.50
F-test ANOVA	f	0.10	0.25	0.40
F-test regression	f^2	0.02	0.15	0.35
Chi-square test	w	0.10	0.30	0.50

Because effect size can only be calculated after you collect data, an estimate is often used for power analysis. It is typical to use an estimate from a prior study or to use a value of 0.5 as it indicates a moderate to large difference.

The **R** library, *pwr*, performs power analysis for proportions, t-tests, one-way ANOVA, correlation, and linear models. Download the `pwr` package and install it in your **R** library for use in this lab.

T-tests and Power Analysis

Today we will focus on power analysis of t-tests, employing the functions: `pwr.t.test()` or `pwr.t2n.test()`. The first function is used if both samples have the same sample size, the second is used if they have different sample sizes. Type `?pwr` to see the other available functions for power calculations.

As an example, let's find the power of a one-tailed t-test, with a significance level of 0.05, 25 subjects in each group and an effect size *d* equal to 0.75. Is this a small or large effect size?

```
pwr.t.test(n = 25, d = 0.75, sig.level=.05, alternative =
            "greater")
```

What is the power of this test? Note that you could also specify, `alternative = "two.sided"`, `"less"`, or `"greater"` to indicate a two-tailed, or one-tailed test. A two-tailed test is the default. Change the code to `"two.sided"` and see how this affects the power.

Visualizing Power, Effect Size and Sample Size

What is the relationship between power, sample size, and effect size? This should be fairly intuitive by now. In case it isn't, let's graph sample size and effect size over a range of statistical power. Run the below code line-by-line so that you understand exactly what it is doing.

Set the range of effect sizes, *d*

```
d <- seq(from = 0.1, to = 0.9, by = 0.01)
nd <- length(d)
```

Set the range of power values, `p`

```
p <- seq(from = 0.4, to = 0.9, by = 0.1)
np <- length(p)
```

This loops over 6 power values and 81 effect sizes to calculate the sample sizes. What does the `samsize` array contain?

```
samsize <- array(numeric(nd*np), dim=c(nd,np))
for (i in 1:np){
 for (j in 1:nd){
     result <- pwr.t.test(n=NULL, d = d[j], sig.level =
               .05, power = p[i], alternative =
               "two.sided")

     samsize[j,i] <- ceiling(result$n)
 }
}
```

Now we set up the graph.

```
xrange <- range(d)
yrange <- round(range(samsize))
colors <- rainbow(length(p))
plot(xrange, yrange, type="n", xlab="Effect size (d)",
     ylab="Sample Size (n)", las = 1)
```

Add the power curves and make the graph pretty.

```
for (i in 1:np){
 lines(d, samsize[,i], type="l", lwd=2, col=colors[i])
}

abline(v=0, h=seq(0,yrange[2],100), lty=2, col="grey89")
abline(h=0, v=seq(xrange[1],xrange[2],.02), lty=2,
```

```
        col="grey89")
title("Sample Size Estimation for T-test\n Sig=0.05 (Two-
tailed)")
legend("topright", title="Power", as.character(p),
        fill=colors)
```

Problem #1: Analyze the following data sets. Write a brief paragraph that includes the following information: (a) your null and alternate hypotheses, (b) the type of test required (one-sample, two-sample, paired), (c) your t_{crit}, (d) your correctly reported results (e.g. (t = 2.10, df = 32, p = 0.22), and (e) a clear statement of your conclusion from the analysis. If you made an error in statistical hypothesis testing, what type of error would you have made and why?

A research study was conducted to examine the differences between older and younger adults on perceived life satisfaction. A pilot study was conducted to examine this hypothesis. Ten older adults (over the age of 50) and ten younger adults (between 20 and 30) were given a life satisfaction test (known to have high reliability and validity). Scores on the measure range from 0 to 60 with high scores indicative of high life satisfaction; low scores indicative of low life satisfaction. The data are presented below. Compute the appropriate t-test.

```
young <- c(34, 22, 15, 27, 37, 41, 24, 19, 26, 36)

old <- c(41, 24, 46, 39, 21, 33, 43, 40, 50, 41)
```

Problem #2: Using the life satisfaction data, now calculate the following: (a) calculate the effect size, *d*, for the study, (b) conduct a power analysis and report the statistical power. Write a sentence explaining what the power tells you.

Problem #3: Analyze the following data sets. Write a brief paragraph that includes the following information: (a) your null and alternate hypotheses, (b) the type of test required (one-sample, two-sample, paired), (c) your t_{crit}, (d) your correctly reported results (e.g. (t = 2.10, df = 32, p = 0.22), and (e) a clear statement of your conclusion from the analysis. If you made an error in statistical hypothesis testing, what type of error would you have made and why?

A researcher hypothesizes that electrical stimulation of the lateral habenula will result in decreased food intake in rats. Rats undergo stereotaxic surgery and an electrode is implanted in the right lateral habenula. Following a ten-day recovery period, rats (kept at 80 percent body weight) are tested for the amount of food consumed during a 10-minute period of time both with and without electrical stimulation. The testing conditions are counter balanced. Compute the appropriate t-test for the data provided below.

```
nostimulation <- c(18.4, 16.1,  9.2, 32.2, 13.8, 16.1,
27.6, 11.5, 11.5, 18.4, 22.3, 21.1, 16.4, 29.5, 27.9)

stimulation <- c(27.6, 16.1,  6.9, 25.3, 18.4, 11.5,
```

```
32.2, 16.1, 20.7, 23.0, 24.7, 18.1, 26.5, 23.4, 8.3)
```

Problem #4: For your Master's project you have chosen to study whether exposure of Cardinal eggs to direct sunlight increases the mass of the hatchlings. You plan to locate Cardinal nests and randomly open the vegetation around half of them to increase exposure to sunlight. A previous study on the nesting of the Eastern Bluebird found the effect size to be 0.4.

What sample size should you aim for in order to have sufficient power (0.80) to detect a difference? Alter the above power graph to highlight the needs of your project, assuming that the effect size could vary between 0.3 and 0.5 and that your desired statistical power is between 0.60 and 0.80. Remember that your hypothesis is a one-sided test, and so you need to take this into account in your graph. Add a horizontal black line on your power graph showing that crosses the power curves at the required sample size.

Problem #5: Change the power curve graph so that it plots the range of statistical power with sample sizes from 10 to 100. Make sure that sample size is on the x-axis and effect size is on the y-axis of your graph. What conclusion can you draw about your ability to conduct experiments with 80% power?