University of California, Los Angeles
Department of Statistics

Statistics 100C **Instructor: Nicolas Christou**

Power analysis and non-central distributions in simple regression

Consider the simple regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, ..., n$. The Gauss-Markov conditions hold and also $\epsilon_i \sim N(0, \sigma)$. We want to test the hypothesis $H_0 : \beta_1 = 0$ against the alternative $H_a : \beta_1 \neq 0$. The test can be performed either using the t statistic or the F statistic. Under H_0 these two statistics follow the central t_{n-2} and the central $F_{1,n-2}$.

$$t = \frac{\hat{\beta}_1}{\frac{S_e}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}} \quad \text{or} \quad F = \frac{\hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}{S_e^2}.$$

If H_0 is not true then these two statistics follow the non-central t distribution with non-centrality parameter $\frac{\beta_1 \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}{\sigma}$ and the non-central F distribution with non-centrality parameter $\frac{\beta_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2}$. Why?

Example
We will the following data:

```
#Read the data:
a <- read.table("http://www.stat.ucla.edu/~nchristo/statistics100C/soil.txt", header=TRUE)
#Use only n=7:
a1 <- a[3:9,]
```

We run the regression `lead` on `zinc`

```
#Run simple regression of lead on zinc:
q <- lm(a1$lead ~ a1$zinc)

#Summary of the regression:
summary(q)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 65.12581    8.31592   7.831 0.000545
a1$zinc      0.20743    0.02166   9.577 0.000210

#Using alpha=0.05 and degrees of freedom 5 we obtain the critical t value= 2.570582:
qt(0.975, 5)
```

We see that $t = 9.577$. The critical value using $\alpha = 0.05$ is $t_{0.975,5} = 2.570582$, therefore H_0 is rejected. We can also use the p-value=0.000210. Since the p-value $< \alpha = 0.05$ we reject H_0.

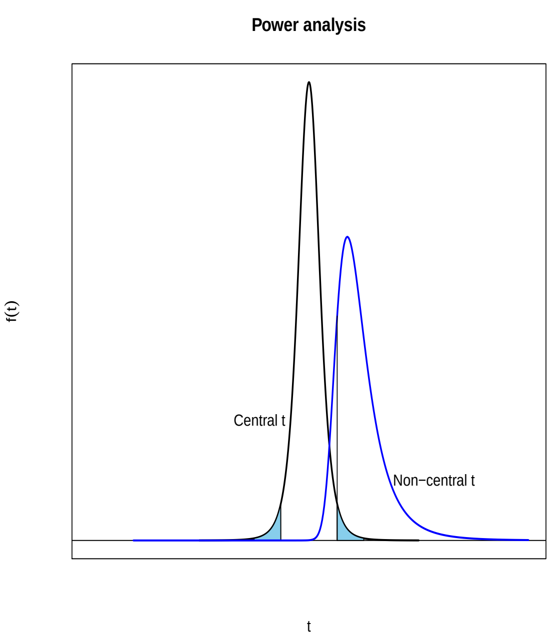
Now suppose we want to compute the power of the test if $\beta_1 = 0.3$ and $\sigma^2 = 625$. We will use either the non-central t or the non-central F distributions.

A. Using the non-central t distribution. The `R` calculations follow:

```
#We need to compute the non-centrality parameter:
beta1 <- 0.3
sigma <- 25
ncp1 <- beta1*sqrt(6*var(a1$zinc)) / sigma

#Compute the power:
pt(2.570582, 5, ncp=ncp1, lower.tail=FALSE) + pt(-2.570582, 5, ncp=ncpar1)

#It is equal to 0.8729043.
```



B. Using the non-central F distribution. The `R` calculations follow:

```
#Critical value using the F distribution is equal to 6.607891 obtained as follows:
qf(0.95, 1,5)

#We need the non-centrality parameter for the chi-squared distribution:
ncp2 <- beta1^2*6*var(a1$zinc)/sigma^2

#Compute the power:
1-pf(6.607891, 1,5, ncp=ncp2)

#It is equal to 0.8729043
```