Math 150 - Methods in Biostatistics - project Part#2

Barah Makhdum

Due: Friday, April 19, 2019, in class

```
library(readr)
AIDSdata <- read_csv("~/Math-150/AIDSdata.csv")
```

```
## Parsed with column specification:
## cols(
##   id = col_integer(),
##   time = col_integer(),
##   censor = col_integer(),
##   time_d = col_integer(),
##   censor_d = col_integer(),
##   tx = col_integer(),
##   txgrp = col_integer(),
##   strat2 = col_integer(),
##   sex = col_integer(),
##   raceth = col_integer(),
##   ivdrug = col_integer(),
##   hemophil = col_integer(),
##   karnof = col_integer(),
##   cd4 = col_double(),
##   priorzdv = col_double(),
##   age = col_integer()
## )
```

```
library(survival)
library(FDRsampsize)
library(powerSurvEpi)
```

Power Analysis

(1) What is Power Analysis?

While conducting tests of hypotheses, the power analysis is how often do commit reject the null hypothesis when shouldn't.

	Truth: T	Truth: F
Null hypothesis	α	$1-\alpha$
Alternative hypothesis	$1-\beta$	β

The power analysis is: $1-\beta$

75% power means you have an 75% chance of getting a significant result when the effect is real.

(2)How to Calculate the power?

To calculare the power there is a common formula we can used it.

$$\beta = \Phi((\frac{|\mu_t - \mu_c|\sqrt{N}}{2\sigma}) - \Phi^{-1}(1 - \frac{\alpha}{2})$$

β is our measure of power. $0 < \beta < 1$

Φ is the CDF of the normal distribution, Φ^{-1} is the inverse.

$\mu_t$is the avarage outcome in the treatment group. Suppose it's 65

μ_c is the avarge outcome in the control group. Suppose it's 60

σ is the standard deviation of outcomes.

α is our significance level

The assumption of power analysis is, involves random sampling. This means that the sample on which power analysis is being conducted is drawn by the process of random sampling.

(3)What are ingredients of Statistical Power? There are three ingredients of that.

First,strength of the treatment. There is positive relationship between the power and strength of the treatment. That mean, when the strength of your treatment increases, the power of your experiment increases.

Second, background noise. There is opposite relationship between the power and background noise. That mean, when the background noise of your outcome variables increases, the power of your experiment decreases.

Third, experimental Design. Traditional power analysis focuses on one element of experimental design: the number of subjects in each experimental group.

(4)Power analysis for survival analysis In survival analysis, the power is directly related to the number of events observed in the study. The required sample size is therefore determined by the observed number of events. Survival data are commonly analyzed using the log-rank test or the Cox proportional hazards model.

Null Hypothesis to be Tested

$H_0 : HR = 1$

where $HR = \frac{h_0(t)}{h_1(t)}$ for all t assuming proportional hazards

Test Statistic

HR estimated from Cox model

Effect Size

HR = 1 implies no difference between treatments

HR > 1 implies "survival" is longer on treatment 2

HR < 1 implies "survival" is longer on treatment 1

Significance Level

$\alpha = 0.05$, $z_{\frac{\alpha}{2}} = 1.96$

Power

Typically desire power of at least 80%,90% or 95%. Recall that for means and proportions, power is a function of sample size.However, for survival data, power is entirely driven by number of events

Power	β	z_β
80%	0.20	0.842
90%	0.10	1.282
95%	0.05	1.645

Required Number of Events

$$events = \frac{(z_{\frac{\alpha}{2}} + z_\beta)^2}{\pi_1 \pi_2 (logHR)^2}$$

where $z_{\frac{\alpha}{2}}$ and z_β are dtandard normal percentiles, $\pi_1 and \pi_2$ are the proportion to be allocated to grpups 1 and 2

(for equal allocation π_1 and π_2=1/2)

Probability of an Event

$$pevent = 1 - (\pi_1 s_1(T) + \pi_2 s_2(T)$$

simulation

First: we have to select subset from our data. The subset has to be significant to get the meaning of power analysis (reject H_0 when shouldn't)

we need to choose 80% from 851. That is equal 681

```
library(data.table)
AIDSdata_1<-data.table(AIDSdata)
y=AIDSdata_1[sample(.N,681)]
```

we keep to that untill we get what we looking for. Then we do the Cox Test.

```
coxph(Surv(tx,censor==1)~cd4,data=y)
```

```
## Call:
## coxph(formula = Surv(tx, censor == 1) ~ cd4, data = y)
##
##         coef exp(coef) se(coef)    z       p
## cd4 -0.01938   0.98081  0.00353 -5.5 3.9e-08
##
## Likelihood ratio test=50.6  on 1 df, p=1.13e-12
## n= 681, number of events= 57
```

Since p-value < 0.05 is significant. Reject H_0 and accept $H_1 : HR \neq 1$

That mean the both groups (treatment and control) are not experiencing an equal number of events at any point in time.

Now, we will do power analysis to do not reject $H_0 : HR = 1$

```
power.cox(681,0.05,0,1)
```

```
## [1] 0.05
```

References:

http://egap.org/methods-guides/10-things-you-need-know-about-statistical-power Yulia Marchenko, 2007. "Power analysis and sample-size determination in survival models with the new stpower command