Power

Christiana Kartsonaki

December 17, 2015

Hypothesis testing

H_0: null hypothesis

H_1: alternative hypothesis

Hypothesis testing

- test statistic
- critical region
- significance level α

$\rightarrow$ reject or not reject H_0

Hypothesis testing

		test	
		not reject H_0	reject H_0
truth	H_0 correct	correct	type I error
	H_1 correct	type II error	correct

Hypothesis testing

		test	
		not reject H_0	reject H_0
truth	H_0 correct	correct true negative	type I error false positive
	H_1 correct	type II error false negative	correct true positive

Hypothesis testing

		test	
		not reject H_0	reject H_0
truth	H_0 correct	correct true negative	type I error false positive
	H_1 correct	type II error false negative	correct true positive

Power

$$\text{power} = \mathbb{P}(\text{reject } H_0 \mid H_1 \text{ true})$$

Hypothesis testing and power

		test	
		not reject H_0	reject H_0
truth	H_0 correct	correct true negative	type I error false positive α: false positive rate
	H_1 correct	type II error false negative β: false negative rate	correct true positive $1 - \beta$: power

significance level α: an upper bound for the probability of type I error

Power and sample size

- cost of observations vs loss in reaching conclusions of low precision
- calculate size of study desirable / establish whether the resources available are sufficient
- standard error of estimate of quantity of interest
- probability of detecting a preassigned departure from a null hypothesis at a specified level of statistical significance

power $\uparrow$ effect size $\uparrow$
sample size $\uparrow$
significance level $\uparrow$
variability $\downarrow$

Power and sample size

$\mathbb{P}(\text{value of test statistic in critical region} \mid H_0) = \alpha$

can solve for n (sample size)

Time-to-event outcome

Cox proportional hazards model – A single binary explanatory variable

Suppose we have a Cox proportional hazards model with a single binary covariate,

$$h(t) = h_0(t)e^{\beta_1 x}.$$

[1]Schoenfeld, D. A. (1983). Sample-size formula for the proportional-hazards regression model. *Biometrics*, **39** (2), 499–503.

Time-to-event outcome

Cox proportional hazards model – A single binary explanatory variable

Suppose we have a Cox proportional hazards model with a single binary covariate,

$$h(t) = h_0(t)e^{\beta_1 x}.$$

Suppose a two-sided test is to be performed with a significance level α and power β when the hazard ratio is HR. Assume that treatment effect is tested by an appropriate test based on the partial likelihood.

Then the total number of events required is given by

$$\frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{p(1-p)\log^2(\text{HR})},$$

where $z_{1-\alpha/2}$, $z_{1-\beta}$ are quantiles of the normal distribution and p is the proportion randomized to treatment 1 (Schoenfeld, 1983)[1].

---

[1] Schoenfeld, D. A. (1983). Sample-size formula for the proportional-hazards regression model. *Biometrics*, **39** (2), 499–503.

Cox proportional hazards model – example

Example

Suppose we want to test whether having a particular gene 'signature' has an effect on progression-free survival of colorectal cancer patients.

Suppose the proportion of the sample that is in group 1 is 0.4 (based on previous studies about 40% of patients have the signature).

Number of events required to achieve each level of power for each hazard ratio, for significance level $\alpha = 0.05$ and a two-sided test:

		power				
		0.95	0.9	0.8	0.7	0.6
	0.5	113	92	69	54	43
	0.6	208	168	126	99	79
HR	0.7	426	345	258	203	161
	0.8	1088	880	657	517	410
	0.9	4878	3944	2947	2317	1839

If the probability of an uncensored observation is p, these numbers should be divided by p to find the total sample size needed.

Cox proportional hazards model – example

Example

Suppose we want to test whether having a particular gene 'signature' has an effect on progression-free survival of colorectal cancer patients.

Suppose the proportion of the sample that is in group 1 is 0.4 (based on previous studies about 40% of patients have the signature).

Number of events required to achieve each level of power for each hazard ratio, for significance level $\alpha = 0.05$ and a two-sided test:

		power				
		0.95	0.9	0.8	0.7	0.6
	0.5	113	92	69	54	43
	0.6	208	168	126	99	79
HR	0.7	426	345	258	203	161
	0.8	1088	880	657	517	410
	0.9	4878	3944	2947	2317	1839

If the probability of an uncensored observation is p, these numbers should be divided by p to find the total sample size needed.

Time-to-event outcome

Cox proportional hazards model – A single continuous explanatory variable

$$h(t) = h_0(t)e^{\beta_1 x}$$

Total number of events required:

$$\frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{\sigma^2 \log^2(\mathrm{HR})},$$

where $z_{1-\alpha/2}$, $z_{1-\beta}$ are quantiles of the normal distribution and σ^2 is the variance of x (Hsieh and Lavori, 2000)[2].

[2] Hsieh, F. Y. and Lavori, P. W. (2000). Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates. *Controlled Clinical Trials*, **21**, 552–560.

correction

$$n^* = \frac{n}{1 - \rho^2}$$

ρ: multiple correlation coefficient (proportion of variance explained by the remaining covariates)

Logistic regression

A single continuous explanatory variable

binary outcome Y
explanatory variable x

$$\log \frac{\mathbb{P}(Y=1)}{1-\mathbb{P}(Y=1)} = \beta_0 + \beta_1 x$$

OR: odds ratio comparing the odds at one standard deviation of x above the mean with the odds at the mean of x

[3] Hsieh, F. Y., Bloch, D. A. and Larsen, M. D. (1998). A simple method of sample size calculation for linear and logistic regression. *Statistics in Medicine*, **17**, 1623–1634.

Logistic regression

A single continuous explanatory variable

binary outcome Y
explanatory variable x

$$\log \frac{\mathbb{P}(Y=1)}{1-\mathbb{P}(Y=1)} = \beta_0 + \beta_1 x$$

OR: odds ratio comparing the odds at one standard deviation of x above the mean with the odds at the mean of x

Number of individuals required:

$$\frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{p(1-p)\log^2(\mathrm{OR})},$$

p: $\mathbb{P}(Y=1)$ at the mean of x (Hsieh et al., 1998)[3]

[3] Hsieh, F. Y., Bloch, D. A. and Larsen, M. D. (1998). A simple method of sample size calculation for linear and logistic regression. *Statistics in Medicine*, **17**, 1623–1634.

Logistic regression

A single continuous explanatory variable

Example

logistic regression with continuous explanatory variable, $n = 1200$

Bonferroni-corrected power calculation for 96 tests

		OR			
		1.25	1.5	1.75	2
	0.1	0.12	0.77	0.99	1.00
	0.2	0.35	0.98	1.00	1.00
p	0.3	0.53	1.00	1.00	1.00
	0.4	0.62	1.00	1.00	1.00
	0.5	0.65	1.00	1.00	1.00

p: $\mathbb{P}(Y = 1)$ at the mean of x

Logistic regression

A single binary explanatory variable

$$n = \frac{(z_{1-\alpha/2}[p(1-p)/B]^{1/2} + z_{1-\beta}[p_1(1-p_1) + p_2(1-p_2)(1-B)/B]^{1/2})^2}{(p_1-p_2)^2(1-B)}$$

$p_1 = \mathbb{P}(Y = 1 \mid x = 0)$
$p_2 = \mathbb{P}(Y = 1 \mid x = 1)$
$B = \mathbb{P}(X = 1)$
$p = (1-B)p_1 + Bp_2$

multiple logistic regression $\text{logit}(p) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$

adjustment $n^* = \frac{n}{1-\rho^2}$, where ρ^2 is the multiple correlation coefficient of the covariate of interest with the remaining covariates (Hsieh, 1989; Hsieh et al., 1998)

In R

Sample size for Cox proportional hazards model with a single binary explanatory variable

```
z_a <- qnorm(alpha/2, lower = FALSE)
z_b <- qnorm(beta, lower = FALSE)

sample_size <- function(z_a, z_b, p, HR) {
            ((z_a + z_b)^2)/(p * (1 - p) * ((log(HR))^2))
            }
```

In R

Power for Cox proportional hazards model with a binary explanatory variable

```
library(powerSurvEpi)

powerEpi.default(n, theta, p, psi, rho2, alpha = 0.05)
```

`n`: total number of subjects.
`theta`: postulated hazard ratio.
`p`: proportion of subjects taking the value one for the covariate of interest.
`psi`: proportion of subjects who had event.
`rho2`: square of the multiple correlation coefficient between the covariate of interest and other covariates.
`alpha`: type I error rate.

In R

Sample size for Cox proportional hazards model with a binary explanatory variable

```
library(powerSurvEpi)

ssizeEpi.default(power, theta, p, psi, rho2, alpha = 0.05)
```

`power`: postulated power.
`theta`: postulated hazard ratio.
`p`: proportion of subjects taking the value one for the covariate of interest.
`psi`: proportion of subjects who had event.
`rho2`: square of the multiple correlation coefficient between the covariate of interest and other covariates.
`alpha`: type I error rate.

In R

Power for Cox proportional hazards model with a continuous explanatory variable

```
library(powerSurvEpi)

powerEpiCont.default(n, theta, sigma2, psi, rho2, alpha = 0.05)
```

`n`: total number of subjects.
`theta`: postulated hazard ratio.
`sigma2`: variance of the covariate of interest.
`psi`: proportion of subjects who had event.
`rho2`: square of the multiple correlation coefficient between the covariate of interest and other covariates.
`alpha`: type I error rate.

In R

Sample size for Cox proportional hazards model with a continuous explanatory variable

```
library(powerSurvEpi)

ssizeEpiCont.default(power, theta, sigma2, psi, rho2, alpha = 0.05)
```

`power`: postulated power.
`theta`: postulated hazard ratio.
`sigma2`: variance of the covariate of interest.
`psi`: proportion of subjects who had event.
`rho2`: square of the multiple correlation coefficient between the covariate of interest and other covariates.
`alpha`: type I error rate.

In R

Power for logistic regression with a single binary explanatory variable

```
library(powerMediation)

powerLogisticBin(n, p1, p2, B, alpha = 0.05)
```

`n`: total number of sample size.

`p1`: $\mathbb{P}(Y = 1|x = 0)$, i.e. the event rate at x = 0 in logistic regression $\text{logit}(p) = \beta_0 + \beta_1 x$, where x is the binary explanatory variable.

`p2`: $\mathbb{P}(Y = 1|x = 1)$, i.e. the event rate at x = 1 in logistic regression $\text{logit}(p) = \beta_0 + \beta_1 x$, where x is the binary explanatory variable.

`B`: pr(X = 1), i.e. proportion of the sample with X = 1

`alpha`: Type I error rate.

In R

Power for logistic regression with a single continuous explanatory variable

```
library(powerMediation)

powerLogisticCon(n, p1, OR, alpha = 0.05)
```

`n`: total sample size.

`p1`: the event rate at the mean of the continuous explanatory variable x in logistic regression $\text{logit}(p) = \beta_0 + \beta_1 x$.

`OR`: assumed odds ratio.

`alpha`: Type I error rate.

Other

```
?power.t.test
```

Simulation.

References

Hsieh, F. Y. (1989). Sample size tables for logistic regression. *Statistics in Medicine*, **8**, 795–802.

Hsieh, F. Y., Bloch, D. A. and Larsen, M. D. (1998). A simple method of sample size calculation for linear and logistic regression. *Statistics in Medicine*, **17**, 1623–1634.

Hsieh, F. Y. and Lavori, P. W. (2000). Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates. *Controlled Clinical Trials*, **21**, 552–560.

Qiu, W., Chavarro, J., Lazarus, R., Rosner, B. and Ma, J. (2012). powerSurvEpi: Power and sample size calculation for survival analysis of epidemiological studies. R package version 0.0.6. `http://CRAN.R-project.org/package=powerSurvEpi`

Qiu, W. (2013). powerMediation: Power/Sample size calculation for mediation analysis, simple linear regression, logistic regression, or longitudinal study. R package version 0.1.7. `http://CRAN.R-project.org/package=powerMediation`

Schmoor, C., Sauerbrei, W. and Schumacher, M. (2000). Sample size considerations for the evaluation of prognostic factors in survival analysis. *Statistics in Medicine*, **19**, 441–452.

Schoenfeld, D. A. (1983). Sample-size formula for the proportional-hazards regression model. *Biometrics*, **39** (2), 499–503.