

Lab 3: Solutions

11/01/2026

Contents

Άσκηση 1 – Μέγεθος δείγματος για προ-καθορισμένη ακρίβεια	3
Θεωρητική λύση	3
Υπολογιστική λύση	3
Σχόλια κλινικής ερμηνείας (ενδεικτικές απαντήσεις)	4
Άσκηση 2 – One-sample t-test με pwr.t.test	5
Υποθέσεις και effect size	5
Υπολογισμός μεγέθους δείγματος + achieved power (σωστή σειρά)	5
Σχόλια κλινικής ερμηνείας (ενδεικτικές απαντήσεις)	7
Άσκηση 3 – Δύο ποσοστά (two-arm) και υπολογισμός ισχύος	8
Υποθέσεις (two-arm)	8
Effect size (Cohen's h) και μέγεθος δείγματος	8
NNT και κλινική ερμηνεία	9
Άσκηση 4 – Phase II single-arm design	11
Exact binomial design	11
Approximate design με arcsine (pwr.p.test)	12
Achieved α και power (σωστό tail)	13
Κλινικά σχόλια – κρίσιμα σημεία	14

Άσκηση 5 – Δύο ανεξάρτητες μέσες τιμές	15
Πίνακας N	15
Άσκηση 6 – Unequal allocation	17
Άσκηση 7 – Σχεδιασμός μελέτης επιβίωσης	20
Median → hazard και HR	20
Sample size με nSurv	20
Dropout και follow-up	22
Καμπύλη N(F)	24
Κλινικά σχόλια	26

Άσκηση 1 – Μέγεθος δείγματος για προ-καθορισμένη ακρίβεια

Θεωρητική λύση

Θέλουμε:

$$\Pr(|\bar{X} - \mu| < d) \geq 1 - \alpha.$$

Υποθέτουμε:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Τότε:

$$\Pr(|\bar{X} - \mu| < d) = \Pr\left(|Z| < \frac{d\sqrt{n}}{\sigma}\right).$$

Η απαίτηση

$$\Pr(|Z| < d\sqrt{n}/\sigma) \geq 1 - \alpha$$

ισοδυναμεί με:

$$\frac{d\sqrt{n}}{\sigma} \geq z_{1-\alpha/2}, \quad \Rightarrow \quad n \geq \left(\frac{\sigma z_{1-\alpha/2}}{d}\right)^2.$$

Για $d = 0.5\sigma$:

$$n \geq (2z_{1-\alpha/2})^2.$$

Υπολογιστική λύση

```
alpha_ex1 <- 0.05
z_ex1 <- qnorm(1 - alpha_ex1/2)
n_raw_ex1 <- (2 * z_ex1)^2
n_min_ex1 <- ceiling(n_raw_ex1)

z_ex1
```

```
## [1] 1.959964
```

```
n_raw_ex1
```

```
## [1] 15.36584
```

```
n_min_ex1
```

```
## [1] 16
```

Με $z_{0.975} = 1.96$:

- $n_{\text{raw}} = 15.37$
- $n_{\text{min}} = 16$

Άρα απαιτούνται **τουλάχιστον 16 άτομα**.

Σχόλια κλινικής ερμηνείας (ενδεικτικές απαντήσεις)

- Precision-based σχεδιασμός είναι πιο φυσικός σε **early-phase, exploratory, biomarker, pilot** μελέτες, όπου προέχει η ακρίβεια εκτίμησης και όχι τυπικός έλεγχος υπόθεσης.
- Μεγαλύτερο n σε early-phase σημαίνει περισσότεροι ασθενείς εκτεθειμένοι σε θεραπεία αβέβαιου οφέλους, καθυστέρηση και κόστος.
- Η υπόθεση γνωστής σ σπάνια είναι πλήρως ρεαλιστική· συνήθως βασίζεται σε μικρές προηγούμενες μελέτες ή expert opinion.
- Υποεκτιμημένη σ οδηγεί σε υπο-προγραμματισμό δείγματος και ευρύτερα διαστήματα εμπιστοσύνης, άρα μειωμένη χρησιμότητα αποτελεσμάτων.

Άσκηση 2 – One-sample t-test με pwr.t.test

Υποθέσεις και effect size

Έλεγχος:

- $H_0 : \mu \leq 3$
- $H_1 : \mu > 3$ (one-sided).

Με $\sigma = 2$:

$$d = \frac{4 - 3}{2} = 0.5.$$

Άρα έχουμε Cohen's $d = 0.5$ (μέτριο effect).

Υπολογισμός μεγέθους δείγματος + achieved power (σωστή σειρά)

```
library(pwr)

mu0_ex2 <- 3
mu1_ex2 <- 4
sigma_ex2 <- 2
d_ex2 <- (mu1_ex2 - mu0_ex2) / sigma_ex2

res_one_ex2 <- pwr.t.test(
  d = d_ex2,
  sig.level = 0.05,
  power = 0.90,
  alternative = "greater",
  type = "one.sample"
)

n_required_ex2 <- res_one_ex2$n
n_int_ex2 <- ceiling(n_required_ex2)
```

```
res_power_ex2 <- pwr.t.test(
  d = d_ex2,
  sig.level = 0.05,
  n = n_int_ex2,
  alternative = "greater",
  type = "one.sample"
)
```

```
res_one_ex2
```

```
##
##      One-sample t test power calculation
##
##              n = 35.65269
##              d = 0.5
##      sig.level = 0.05
##      power = 0.9
##      alternative = greater
```

```
n_required_ex2
```

```
## [1] 35.65269
```

```
n_int_ex2
```

```
## [1] 36
```

```
res_power_ex2
```

```
##
##      One-sample t test power calculation
##
##              n = 36
##              d = 0.5
##      sig.level = 0.05
##      power = 0.9025746
##      alternative = greater
```

Η pwr.t.test δίνει $n_{\text{required}} = 35.653$ και μετά τη στρογγυλοποίηση έχουμε $n = 36$.

Achieved power με αυτό το n : 0.903.

Σχόλια κλινικής ερμηνείας (ενδεικτικές απαντήσεις)

- Η διαφορά $\mu_1 - \mu_0 = 1$ πρέπει να αξιολογείται ως προς τη μονάδα μέτρησης και τη φυσική μεταβλητότητα.
- Η υπόθεση σταθερής/γνωστής σ είναι απλοποίηση· καλό practice: sensitivity analysis σε εναλλακτικές τιμές σ .
- Αν επιλεγεί $\alpha = 0.10$, το απαιτούμενο n μειώνεται, αλλά αυξάνει ο κίνδυνος Type I.
- One-sided έλεγχος: μόνο όταν έχει δηλωθεί εκ των προτέρων και μόνο η “σωστή” κατεύθυνση έχει κλινικό νόημα.

Άσκηση 3 – Δύο ποσοστά (two-arm) και υπολογισμός ισχύος

Υποθέσεις (two-arm)

Ορισμοί: - p_C = ποσοστό ανταπόκρισης στην ομάδα ελέγχου, - p_T = ποσοστό ανταπόκρισης στην πειραματική ομάδα.

Υποθέσεις (one-sided):

$$H_0 : p_T \leq p_C \quad \text{έναντι} \quad H_1 : p_T > p_C,$$

με $p_C = 0.20$, $p_T = 0.30$, $\alpha = 0.10$, $\text{power} = 0.80$.

Effect size (Cohen's h) και μέγεθος δείγματος

Χρησιμοποιούμε το Cohen's h (arcsine transformation) ως measure του effect size:

$$h = 2 \arcsin(\sqrt{p_T}) - 2 \arcsin(\sqrt{p_C}).$$

R

```
library(pwr)

pC_ex3 <- 0.20
pT_ex3 <- 0.30
alpha_ex3 <- 0.10
power_target_ex3 <- 0.80

# Cohen's h (arcsine difference)
h_ex3 <- 2 * asin(sqrt(pT_ex3)) - 2 * asin(sqrt(pC_ex3))
h_ex3

## [1] 0.2319843

# Two-arm test for two independent proportions (equal allocation)
res_ex3 <- pwr.2p.test(
  h = h_ex3,
```



```

sig.level = alpha_ex3,
power = power_target_ex3,
alternative = "greater"
)

```

```

n_per_group_ex3 <- ceiling(res_ex3$n)
n_total_ex3 <- 2 * n_per_group_ex3

```

```
res_ex3
```

```

##
##      Difference of proportion power calculation for binomial distribution (arcsine tr
##
##              h = 0.2319843
##              n = 167.5265
##      sig.level = 0.1
##              power = 0.8
##      alternative = greater
##
## NOTE: same sample sizes

```

```
n_per_group_ex3
```

```
## [1] 168
```

```
n_total_ex3
```

```
## [1] 336
```

Απαιτείται $n = 168$ ανά ομάδα, άρα συνολικά $N = 336$.

NNT και κλινική ερμηνεία

Η απόλυτη διαφορά:

$$\Delta = p_T - p_C = 0.10,$$

άρα

$$\text{NNT} = 1/\Delta = 10.$$

Ερμηνεία: - Πρέπει να θεραπεύσουμε ~10 ασθενείς με τη νέα θεραπεία για να έχουμε 1 επιπλέον ανταπόκριση σε σχέση με την standard. - Το αν είναι αποδεκτό εξαρτάται από νόσο, ασφάλεια, κόστος, εναλλακτικές.

Άσκηση 4 – Phase II single-arm design

Exact binomial design

Υποθέσεις: - $H_0 : p \leq 0.15$

- $H_A : p > 0.15$.

Κανόνας απόφασης (με το r όπως επιστρέφεται από `ph2single`):

Απορρίπτουμε το H_0 αν $X > r$ (ισοδύναμα, αν $X \geq r + 1$).

```
library(clinfun)

p0_ex4 <- 0.15
p1_ex4 <- 0.40
alpha_ex4 <- 0.10
beta_ex4 <- 0.20

out_ex4 <- ph2single(pu = p0_ex4, pa = p1_ex4, ep1 = alpha_ex4, ep2 = beta_ex4, nsoln =
out_ex4
```

```
##      n r Type I error Type II error
## 1  16 4   0.07905130   0.16656738
## 2  17 4   0.09871000   0.12599913
## 3  19 5   0.05369611   0.16292248
## 4  20 5   0.06730797   0.12559897
## 5  21 5   0.08273475   0.09574016
## 6  22 5   0.09994525   0.07222635
## 7  22 6   0.03684009   0.15844366
## 8  23 6   0.04630586   0.12395674
## 9  24 6   0.05719150   0.09596147
## 10 24 7   0.01986931   0.19194524
```

Επιλογή ελάχιστου n :

```
out_smallest_ex4 <- out_ex4[order(out_ex4$n), ][1, ]
out_smallest_ex4
```

```
##      n r Type I error Type II error
## 1 16 4      0.0790513      0.1665674
```

```
n_ex4 <- out_smallest_ex4$n
r_ex4 <- out_smallest_ex4$r
n_ex4
```

```
## [1] 16
```

```
r_ex4
```

```
## [1] 4
```

Approximate design $\mu\epsilon$ arcsine (pwr.p.test)

```
library(pwr)

h_arc_ex4 <- ES.h(p1_ex4, p0_ex4)

out_arc_ex4 <- pwr.p.test(
  h = h_arc_ex4,
  sig.level = alpha_ex4,
  power = 1 - beta_ex4,
  alternative = "greater"
)

n_arc_ex4 <- ceiling(out_arc_ex4$n)

h_arc_ex4

## [1] 0.5740396
```

```
out_arc_ex4
```

```
##
##      proportion power calculation for binomial distribution (arcsine transformation)
##
##              h = 0.5740396
##              n = 13.68003
##      sig.level = 0.1
##      power = 0.8
##      alternative = greater
```

```
n_arc_ex4
```

```
## [1] 14
```

Achieved α και power (σωστό tail)

```
alpha_ach_ex4 <- 1 - pbinom(r_ex4, size = n_ex4, prob = p0_ex4) #  $P(X \geq r+1)$ 
power_ach_ex4 <- 1 - pbinom(r_ex4, size = n_ex4, prob = p1_ex4) #  $P(X \geq r+1)$ 
```

```
alpha_ach_ex4
```

```
## [1] 0.0790513
```

```
power_ach_ex4
```

```
## [1] 0.8334326
```

Γενικά: - $\alpha_{\text{ach}} \leq 0.1$ (συχνά < λόγω διακριτότητας), - $\text{Power}_{\text{ach}} \geq 0.8$.

Κλινικά σχόλια – κρίσιμα σημεία

- Single-arm Phase II είναι αποδεκτός όταν το ιστορικό benchmark είναι αξιόπιστο και η standard-of-care δεν είναι επαρκής.
- Λάθος επιλογή p_0 δημιουργεί bias: υπερβολικά χαμηλό $p_0 \rightarrow$ θεραπεία «φαίνεται» καλή εύκολα, υπερβολικά υψηλό $p_0 \rightarrow$ καταλήγουμε να απορρίψουμε πραγματικά ωφέλιμες θεραπείες.
- Type I error = 0.10 σημαίνει ότι 10% θεραπειών με πραγματικό $p = p_0$ μπορεί να “περάσουν” σε επόμενο στάδιο. Αυτό είναι συνειδητός συμβιβασμός σε Phase II, αλλά φορτώνει Phase III με ίσως αναποτελεσματικές θεραπείες.

Άσκηση 5 – Δύο ανεξάρτητες μέσες τιμές

Πίνακας N

```
library(pwr)
library(knitr)

ES_ex5 <- seq(0.25, 1.50, by = 0.25)
power_target_ex5 <- 0.90

tab_ex5 <- data.frame(
  d = ES_ex5,
  `alpha=0.05 (Total N)` = NA_integer_,
  `alpha=0.10 (Total N)` = NA_integer_
)

for (i in seq_along(ES_ex5)) {
  d_val <- ES_ex5[i]

  n05 <- pwr.t.test(d = d_val, sig.level = 0.05, power = power_target_ex5,
                    type = "two.sample", alternative = "two.sided")$n
  n10 <- pwr.t.test(d = d_val, sig.level = 0.10, power = power_target_ex5,
                    type = "two.sample", alternative = "two.sided")$n

  tab_ex5[i, 2] <- 2 * ceiling(n05)
  tab_ex5[i, 3] <- 2 * ceiling(n10)
}

kable(tab_ex5, align = "c")
```

d	alpha.0.05..Total.N.	alpha.0.10..Total.N.
0.25	676	550
0.50	172	140
0.75	78	64
1.00	46	36
1.25	30	24
1.50	22	18

Άσκηση 6 – Unequal allocation

```
library(gsDesign)

d_ex6 <- 0.5
alpha_ex6 <- 0.05
beta_ex6 <- 0.10

k_vals_ex6 <- 1:10

ar_ex6 <- data.frame(
  k = k_vals_ex6,
  ratio = paste0("1:", k_vals_ex6),
  N = NA_integer_,
  n1 = NA_integer_,
  n2 = NA_integer_
)

for (i in seq_along(k_vals_ex6)) {

  k <- k_vals_ex6[i]

  out <- nNormal(delta1 = d_ex6, delta0 = 0,
    sd = 1, sd2 = 1,
    alpha = alpha_ex6, beta = beta_ex6,
    sided = 2,
    ratio = 1/k)

  N_total <- as.numeric(out)
  N_int <- ceiling(N_total)

  n1 <- ceiling(N_int / (1 + k))
  n2 <- N_int - n1

  ar_ex6$N[i] <- N_int
```

```

  ar_ex6$n1[i] <- n1
  ar_ex6$n2[i] <- n2
}

```

```

ar_ex6

```

```

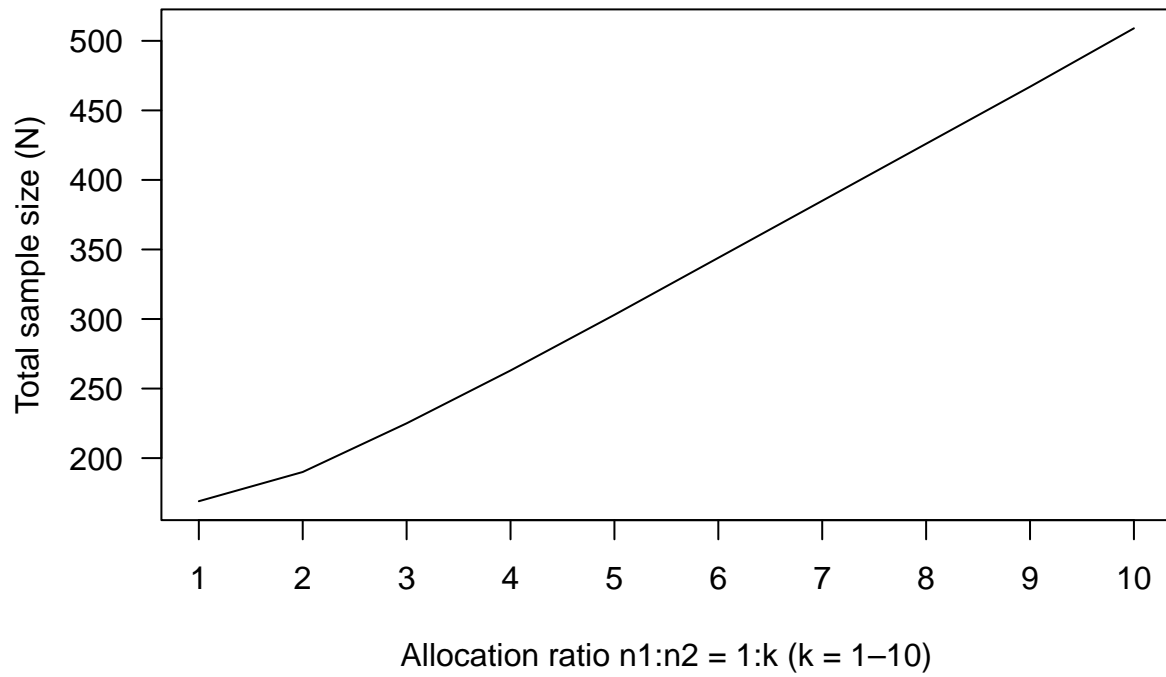
##      k ratio   N n1  n2
## 1    1   1:1 169 85  84
## 2    2   1:2 190 64 126
## 3    3   1:3 225 57 168
## 4    4   1:4 263 53 210
## 5    5   1:5 303 51 252
## 6    6   1:6 344 50 294
## 7    7   1:7 385 49 336
## 8    8   1:8 426 48 378
## 9    9   1:9 467 47 420
## 10  10  1:10 509 47 462

```

```

plot(ar_ex6$k, ar_ex6$N,
     xlab = "Allocation ratio n1:n2 = 1:k (k = 1-10)",
     ylab = "Total sample size (N)",
     type = "l", xaxt = "n",
     las = 1)
axis(1, at = 1:10)

```



Θεωρητικά:

$$\text{Var}(\bar{X}_1 - \bar{X}_2) \propto \frac{1}{n_1} + \frac{1}{n_2},$$

η οποία ελαχιστοποιείται όταν $n_1 = n_2$.

Άσκηση 7 – Σχεδιασμός μελέτης επιβίωσης

Median → hazard και HR

Υπό εκθετική κατανομή:

$$S(t) = e^{-\lambda t}, \quad S(m) = 0.5 \Rightarrow e^{-\lambda m} = 0.5,$$

οπότε:

$$\lambda = \frac{\log 2}{m}.$$

```
m0_ex7 <- 15
```

```
m1_ex7 <- 25
```

```
lambda0_ex7 <- log(2) / m0_ex7
```

```
lambda1_ex7 <- log(2) / m1_ex7
```

```
HR_ex7 <- lambda1_ex7 / lambda0_ex7
```

```
lambda0_ex7
```

```
## [1] 0.04620981
```

```
lambda1_ex7
```

```
## [1] 0.02772589
```

```
HR_ex7
```

```
## [1] 0.6
```

```
m0_ex7 / m1_ex7
```

```
## [1] 0.6
```

Sample size με nSurv

```

library(gsDesign)

R_ex7 <- 55
F_ex7 <- 10
T_ex7 <- R_ex7 + F_ex7

out1_ex7 <- nSurv(lambdaC = lambda0_ex7,
                  hr = HR_ex7,
                  hr0 = 1,
                  eta = 0,
                  alpha = 0.05,
                  beta = 0.20,
                  sided = 2,
                  ratio = 1,
                  R = R_ex7,
                  minfup = F_ex7,
                  T = T_ex7)

out1_ex7

```

```

## Fixed design, two-arm trial with time-to-event
## outcome (Lachin and Foulkes, 1986).
## Solving for: Accrual rate
## Hazard ratio                H1/H0=0.6/1
## Study duration:             T=65
## Accrual duration:           55
## Min. end-of-study follow-up: minfup=10
## Expected events (total, H1): 119.2209
## Expected sample size (total): 172.4299
## Accrual rates:
##      Stratum 1
## 0-55      3.1351
## Control event rates (H1):
##      Stratum 1
## 0-Inf     0.0462
## Censoring rates:

```

```
##          Stratum 1
## 0-Inf          0
## Power:                100*(1-beta)=80%
## Type I error (1-sided): 100*alpha=5%
## Equal randomization:      ratio=1
```

```
N1_ex7 <- out1_ex7$n
accrual_rate_ex7 <- N1_ex7 / R_ex7
```

```
N1_ex7
```

```
## [1] 172.4299
```

```
accrual_rate_ex7
```

```
## [1] 3.13509
```

Dropout and follow-up

```
out2_ex7 <- nSurv(lambdaC = lambda0_ex7,
  hr = HR_ex7,
  hr0 = 1,
  eta = 0.009,
  alpha = 0.05,
  beta = 0.20,
  sided = 2,
  ratio = 1,
  R = R_ex7,
  minfup = F_ex7,
  T = T_ex7)
```

```
out2_ex7
```

```
## Fixed design, two-arm trial with time-to-event
```

```

## outcome (Lachin and Foulkes, 1986).
## Solving for: Accrual rate
## Hazard ratio                H1/H0=0.6/1
## Study duration:              T=65
## Accrual duration:            55
## Min. end-of-study follow-up: minfup=10
## Expected events (total, H1): 119.3854
## Expected sample size (total): 196.3055
## Accrual rates:
##      Stratum 1
## 0-55      3.5692
## Control event rates (H1):
##      Stratum 1
## 0-Inf      0.0462
## Censoring rates:
##      Stratum 1
## 0-Inf      0.009
## Power:                    100*(1-beta)=80%
## Type I error (1-sided):    100*alpha=5%
## Equal randomization:       ratio=1

```

```

F2_ex7 <- 100
T2_ex7 <- R_ex7 + F2_ex7

out3_ex7 <- nSurv(lambdaC = lambda0_ex7,
                  hr = HR_ex7,
                  hr0 = 1,
                  eta = 0,
                  alpha = 0.05,
                  beta = 0.20,
                  sided = 2,
                  ratio = 1,
                  R = R_ex7,
                  minfup = F2_ex7,
                  T = T2_ex7)

```

```
out3_ex7
```

```
## Fixed design, two-arm trial with time-to-event
## outcome (Lachin and Foulkes, 1986).
## Solving for: Accrual rate
## Hazard ratio                      H1/H0=0.6/1
## Study duration:                    T=155
## Accrual duration:                  55
## Min. end-of-study follow-up: minfup=100
## Expected events (total, H1):       119.7092
## Expected sample size (total):      121.8807
## Accrual rates:
##      Stratum 1
## 0-55      2.216
## Control event rates (H1):
##      Stratum 1
## 0-Inf     0.0462
## Censoring rates:
##      Stratum 1
## 0-Inf      0
## Power:                      100*(1-beta)=80%
## Type I error (1-sided):      100*alpha=5%
## Equal randomization:         ratio=1
```

Καμπύλη N(F)

```
fp_ex7 <- seq(5, 100, by = 0.5)
N_vec_ex7 <- rep(NA_real_, length(fp_ex7))

for (i in seq_along(fp_ex7)) {
  Fi <- fp_ex7[i]
  Ti <- R_ex7 + Fi

  tmp <- nSurv(lambdaC = lambda0_ex7,
               hr = HR_ex7,
```



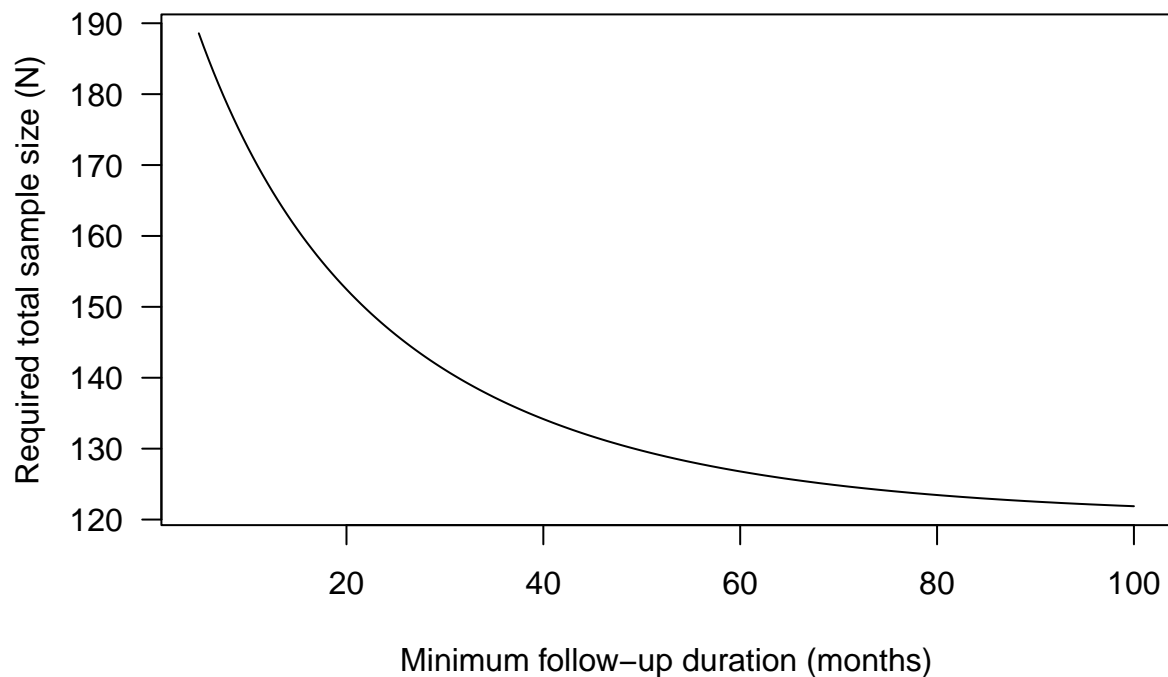
```

        hr0 = 1,
        eta = 0,
        alpha = 0.05,
        beta = 0.20,
        sided = 2,
        ratio = 1,
        R = R_ex7,
        minfup = Fi,
        T = Ti)

N_vec_ex7[i] <- tmp$n
}

plot(fp_ex7, N_vec_ex7,
     xlab = "Minimum follow-up duration (months)",
     ylab = "Required total sample size (N)",
     type = "l",
     las = 1)

```



Κλινικά σχόλια

- Η υπόθεση proportional hazards μπορεί να είναι χονδρική: delayed effects, non-proportional hazards κ.λπ. μειώνουν την καταλληλότητα του log-rank και το νόημα του HR.
- Η ισχύς σε μελέτες επιβίωσης καθορίζεται πρωτίστως από τον αριθμό γεγονότων: για δεδομένο HR και α , υπάρχει ένας απαιτούμενος αριθμός events D . Το N προκύπτει από την ισορροπία μεταξύ accrual, follow-up και dropout.
- Δεν αρκεί να «γεμίσουμε» τη μελέτη με πολλούς ασθενείς αν δεν υπάρξει επαρκής follow-up ώστε να παρατηρηθούν αρκετά γεγονότα.

Διδακτική Σημείωση: Η λογική πίσω από τα νούμερα

Είναι κρίσιμο να κατανοήσουμε ότι σε μελέτες επιβίωσης, **το νόμισμα είναι τα events, όχι οι ασθενείς**.

1. **Events First:** Ο αλγόριθμος (nSurv) υπολογίζει πρώτα τον αριθμό των απαιτούμενων συμβάντων (D) που χρειάζονται για να επιτευχθεί η ισχύς (Power) δεδομένου του HR. Ο

αριθμός αυτός είναι **σταθερός** (fixed) και δεν αλλάζει όσα άτομα και να βάλουμε ή όσο και να περιμένουμε.

2. **Το “Παζάρι” (Trade-off):** Το συνολικό N προκύπτει ως συνάρτηση του D και του χρόνου παρακολούθησης.

- Μπορούμε να βάλουμε **πολλούς ασθενείς** και να τους παρακολουθήσουμε για **λίγο χρόνο**.
- Μπορούμε να βάλουμε **λιγότερους ασθενείς** και να τους παρακολουθήσουμε για **περισσότερο χρόνο**.

3. **Το Όριο:** Καθώς αυξάνουμε το follow-up (όπως φαίνεται στο γράφημα), το απαιτούμενο N μειώνεται. Θεωρητικά, αν είχαμε άπειρο follow-up, το N θα συνέκλινε στον αριθμό των events, καθώς όλοι οι ασθενείς τελικά θα εμφάνιζαν το γεγονός.