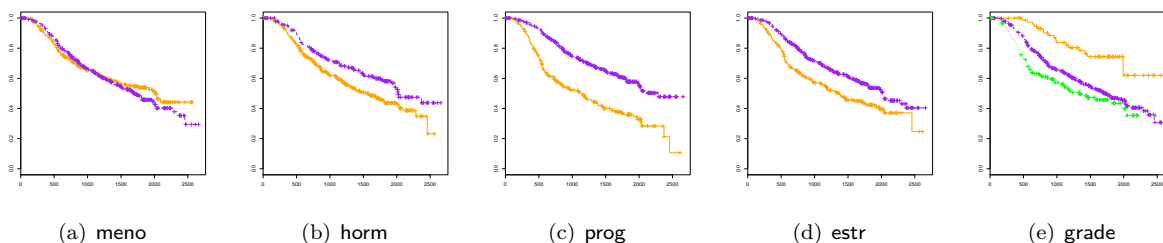


Rak piersi a czas przeżycia bez nawrotu choroby

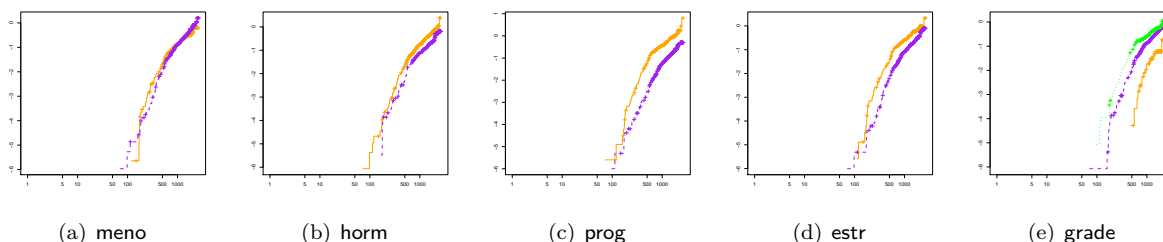
Analizie poddano 686 pacjentek cierpiących na raka piersi. Za główny cel postawiono pytanie, które zmienne mają wpływ na czas przeżycia bez nawrotu choroby. Zaproponowano model proporcjonalnych hazardów (PH).

Sprawdzenie założeń - krzywe przeżycia oraz ich transformacje.

Dla każdej ze zmiennych (poza ciągłymi) sprawdzono spełnienie założeń poprzez narysowanie krzywych Kaplana-Meiera (Rysunek 1) oraz wykresów transformacji $\log(-\log)$ krzywych przeżycia (Rysunek 2).



Rysunek 1: Krzywe przeżycia Kaplana-Meiera dla zmiennych dyskretnych.



Rysunek 2: Transformacje $\log(-\log)$ krzywych przeżycia Kaplana-Meiera.

Krzywe przeżycia dla różnych poziomów zmiennej **meno** przecinają się, co może oznaczać niespełnianie założeń modelu PH. Dla pozostałych zmiennych krzywe nie przecinają się oraz wykresy transformacji możemy uznać za nieodstające od równoległych. Z tej racji zaproponowano model proporcjonalnych hazardów warstwowy względem zmiennej **meno**, gdyż nie ma podstaw by nie twierdzić, że założenia modelu ph nie są spełnione.

W celu potwierdzenia wniosków płynących z graficznej prezentacji krzywych przeżycia, przeprowadzono formalny test logrank dla każdej zmiennej, którego podsumowanie widać w Tabeli 1. Dla zmiennej **grade** przeprowadzono test logrank dla trendu z racji na naturalne uporządkowanie poziomów. Hipoteza zerowa zakłada, że krzywe przeżycia dla zmiennych nie różnią się istotnie statystycznie. Wartości krytyczne dla zmiennych **horm**, **prog**, **estr** i **grade** (**?-nie widzę p-wartości dla tej zmiennej :(**), które są mniejsze od zakładanego

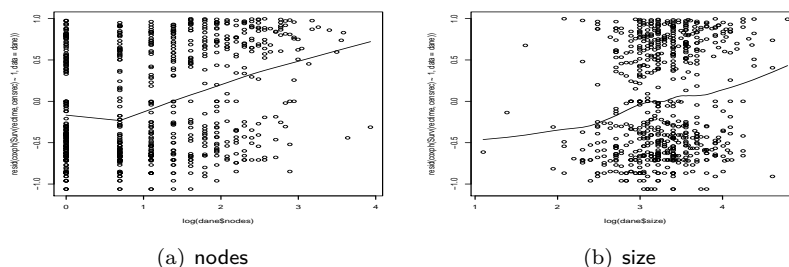
poziomu istotności $\alpha = 0.05$ dają statystycznie istotne podstawy do odrzucenia hipotez zerowych w testach logrank oraz do przyjęcia hipotez alternatywnych o tym, że krzywe przeżycia dla tych zmiennych różnią się. Dla zmiennej **meno** p-wartość 0.597 nie daje podstaw do odrzucenia hipotezy zerowej (dla przyjętego poziomu istotności $\alpha = 0.05$, mówiącej o równości krzywych przeżycia dla różnych poziomów tej zmiennej).

Tabela 1: Wyniki testów logrank.

	χ^2	St. swobody	p-wartość
meno	0.3	1	0.597
horm	8.6	1	0.003
prog	49.3	1	0.000
estr	14.3	1	0.000
grade	44.5	2	0.000

Przekształcenia zmiennych ciągłych.

W celu sprawdzenia, czy zmienne ciągłe nie powinny zostać przekształcone przed wprowadzeniem ich do modelu, wykonano wykresy reszt martyngałowych pustego modelu od każdej z tych zmiennych. Otrzymane wykresy sugerują przekształcenie zmiennych poprzez logarytm, więc dla potwierdzenia tych przypuszczeń narysowano dodatkowo wykresy reszt od logarytmów zmiennych (Rysunek 3) i zaobserwowano, że wykresy są teraz bliższe liniowym. Wprowadzono więc do modelu zmienne ciągłe przekształcone logarytmem.



Rysunek 3: Wykresy reszt martyngałowych od logarytmów zmiennych ciągłych.

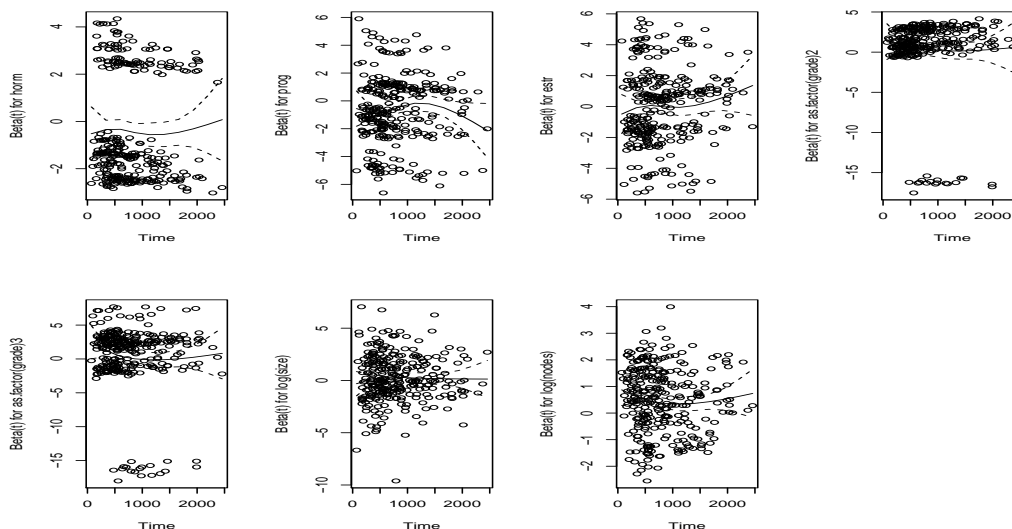
Sprawdzenie założeń - formalny test Schoenfelda.

W celu formalnego sprawdzenia czy współczynniki w modelu są stałe w czasie przeprowadzono test Schoenfelda, którego globalna p-wartość oraz pojedyncze p-wartości dla zmiennych dyskretnych oraz ciągłych przekształconych przez logarytm są większe od zakładanego poziomu istotności $\alpha = 0.05$ co nie daje podstaw do odrzucenia hipotezy zerowej w tym teście, mówiącej o stałości współczynników w czasie.

Uwagę zwrócono również na wykresy skalowanych reszt Schoenfelda dla każdej zmiennej w modelu od czasu i dopasowano do nich krzywe. Dołączone granice dla 95% obszarów ufności sugerują, że nie ma podstaw do odrzucenia hipotez, że krzywe nie różnią się od horyzontalnych.

Tabela 2: Wyniki testu Schoenfelda.

	ρ	χ^2	p-wartość
-horm	-0.009	0.024	0.876
-prog	0.036	0.369	0.543
-estr	0.071	1.458	0.227
-as.factor(grade)2	-0.052	0.811	0.368
-as.factor(grade)3	-0.089	2.347	0.126
-log(size)	0.011	0.037	0.848
-log(nodes)	-0.051	0.879	0.349
-GLOBAL	NA	11.322	0.125



Rysunek 4: Wykresy skalowanych reszt Schoenfelda.

Zarówno wykresy skalowanych reszt Schoenfelda, jak i wyniki testu nie sugerują odstępstw od założenia PH.

Dopasowanie modelu proporcjonalnych hazardów.

```
model <- coxph(Surv(rectime,censrec)~horm+prog+estr+as.factor(grade)+strata(meno)+log(size)+log(nodes),
              data = dane)
summary(model)
```

Call:

```
coxph(formula = Surv(rectime, censrec) ~ horm + prog + estr +
      as.factor(grade) + strata(meno) + log(size) + log(nodes),
      data = dane)
```

n= 686, number of events= 299

	coef	exp(coef)	se(coef)	z	Pr(> z)	
horm	-0.40288	0.66839	0.12924	-3.117	0.00183	**
prog	-0.67857	0.50734	0.14434	-4.701	2.59e-06	***
estr	0.03400	1.03458	0.14314	0.238	0.81226	
as.factor(grade)2	0.50378	1.65497	0.25275	1.993	0.04624	*
as.factor(grade)3	0.52144	1.68445	0.27898	1.869	0.06161	.
log(size)	0.20252	1.22449	0.13207	1.533	0.12515	
log(nodes)	0.47991	1.61593	0.06697	7.166	7.71e-13	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
horm	0.6684	1.4961	0.5188	0.8611
prog	0.5073	1.9711	0.3823	0.6732
estr	1.0346	0.9666	0.7815	1.3696
as.factor(grade)2	1.6550	0.6042	1.0084	2.7160
as.factor(grade)3	1.6845	0.5937	0.9750	2.9102
log(size)	1.2245	0.8167	0.9452	1.5862
log(nodes)	1.6159	0.6188	1.4172	1.8426

Concordance= 0.697 (se = 0.025)

Rsquare= 0.172 (max possible= 0.99)

Likelihood ratio test= 129.7 on 7 df, p=0

Wald test = 128.6 on 7 df, p=0

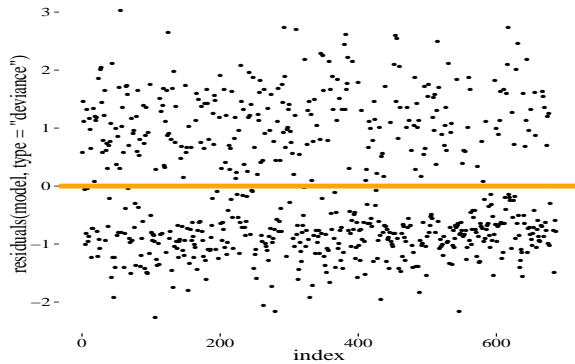
Score (logrank) test = 136.5 on 7 df, p=0

W stworzonym modelu zmiennymi istotnymi statystycznie na poziomie istotności 0.05 są: **horm**, **prog**, **as.factor(grade)2**, **log(nodes)**. Współczynniki modelu przy tych zmiennych wynoszą odpowiednio: -0.40 , -0.67 , 0.50 , 0.47 . Oznacza to, że jeśli została użyta terapia hormonalna, to hazard zgonu lub nawrotu choroby zmniejsza się $\exp(-0.40) = 0.66$ raza w stosunku do nieużycia terapii hormonalnej (gdy wszystkie inne zmienne są takie same). Jeśli wskaźnik receptorów progesteronu zmienia się z ujemnego na dodatni, to hazard zmieni się o $\exp(-0.67) = 0.5$ raza, natomiast gdy logarytm liczby węzłów chłonnych z przerzutami nowotworu wzrośnie o 1, to hazard zwiększy się $\exp(0.47) = 1.61$ raza. Zmiana z **grade** (stopień zróżnicowania komórek nowotworu (1-wysoki, 2-średni) z 1 do 2 zwiększa hazard $\exp(0.5) = 1.65$ raza, gdy wszystkie inne zmienne pozostaną na tym samym poziomie.

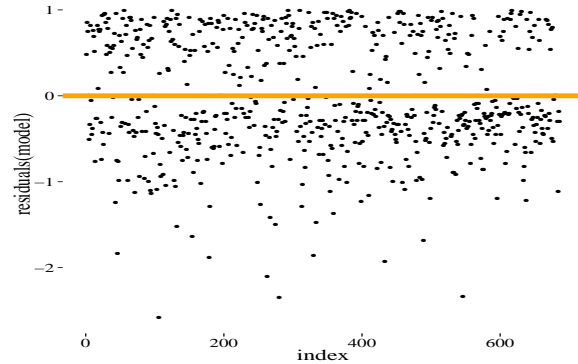
Testy Likelihood, Wald i Score dają p-wartość mniejszą od 0.05 (przyjętego poziomu istotności), co wskazuje, że odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej, zatem dopasowany model jest istotnie lepszy od modelu pustego.

Sprawdzenie dopasowania modelu proporcjonalnych hazardów.

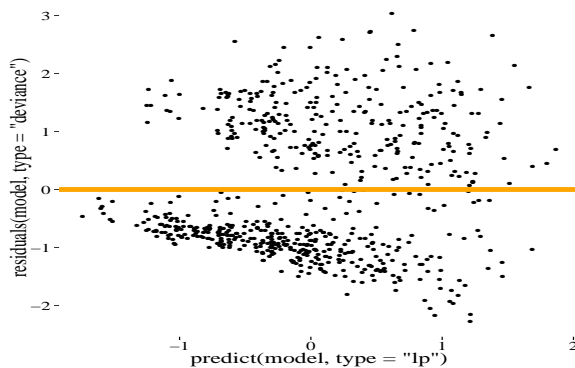
W celu sprawdzenia dopasowania modelu wygenerowano wykresy reszt dewiancji/martyngałowych od liniowej kombinacji zmiennych/indeksów. Na niektórych wykresach można wskazać wyraźnie odstające reszty. Wykres reszt dewiancji od indeksów można uznać za symetryczny, pozostałe niekoniecznie, co sugeruje, że dopasowany model nie jest perfekcyjny.



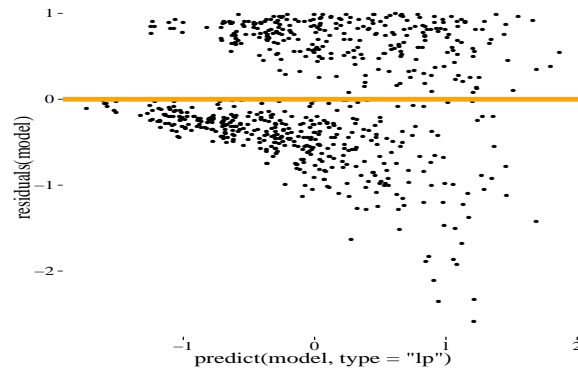
(a) Reszty dewiancji a indeks.



(b) Reszty martyngałowe a indeks.



(c) Reszty dewiancji a liniowe kombinacje.



(d) Reszty martyngałowe a liniowe kombinacje.

Rysunek 5: Wykresy reszt martyngałowych i dewiancji.

Kody.

```
library(foreign)
dane <- read.dta("gbcs_short.dta")
library(survival)
plot(survfit(Surv(rectime,censrec)~meno, data = dane), col=c("orange","purple"), lty=c(1:2), lwd=3)
#...
plot(log(dane$size), resid(coxph(Surv(rectime,censrec)~1, data = dane)))
lines(lowess(log(dane$size), resid(coxph(Surv(rectime,censrec)~1, data = dane))), iter = 0, f = 0.6)
#...

par(mfrow=c(2,3))
plot(cox.zph(model, transform = "identity"), df=4,nsmo=10, se=TRUE)
model <- coxph(Surv(rectime,censrec)~horm+prog+estr+grade+strata(meno)+log(size)+log(nodes),
              data = dane)
summary(model)

#...
qplot(1:686,residuals(model, type="deviance"))+xlab("index")+theme_tufte(base_size=20)+geom_hline()+
  geom_hline(yintercept=0, col ="orange", size = 3)
```