

# Building and Checking Survival Models

David M. Roche

October 19, 2023

# hodg Lymphoma Data Set from KMsurv

This data set consists of information on 43 bone marrow transplant patients at Ohio State University (Avalos 1993). The patients had either Hodgkin's or non-Hodgkin lymphoma and were treated either with an allogenic (HLA-matched sib) or autologous bone marrow transplant. In addition to the time to death or relapse (or censored), the data set has the Karnofsky score and the waiting time to transplant in months.

Bone marrow and stem cell transplants require essentially killing the patient's white blood and stem cells by radiation or chemotherapy, so this is a fairly dangerous procedure.

# R-CHOP Therapy (since 2006)

This combination contains the chemotherapy drugs cyclophosphamide, doxorubicin hydrochloride (hydroxydaunomycin), and vincristine sulfate (Oncovin), the targeted therapy drug rituximab, and the steroid hormone prednisone. Combinations usually work better than single drugs because different drugs kill cancer cells in different ways.

---

## Drugs in the R-CHOP combination:

---

- R Rituximab
  - C Cyclophosphamide
  - H Doxorubicin Hydrochloride (Hydroxydaunomycin)
  - O Vincristine Sulfate (Oncovin)
  - P Prednisone
-

# R-CHOP: Use in Cancer

R-CHOP is used to treat B-cell non-Hodgkin lymphoma, including the following types:

- Diffuse large B-cell lymphoma (DLBCL).
- Follicular lymphoma that is advanced. It is used as the first treatment.
- Waldenström macroglobulinemia.
- This combination may also be used with other drugs or treatments or to treat other types of cancer.

R-CHOP was the first initial therapy for non-Hodgkin lymphoma that can be actually curative and not just induce temporary remission. In this case, fewer bone marrow or stem cell transplants would be needed.

# hodg Lymphoma Data Set from KMsurv

---

gtype	Graft type 1=allogenic, 2=autologous
dtype	Disease type 1=Non Hodgkin lymphoma, 2=Hodgkins disease
time	Time to death or relapse, days
delta	Death/relapse indicator 0=alive, 1=dead
score	Karnofsky score
wtime	Waiting time to transplant in months

---

# Karnofsky Score

Score 80–100: Able to carry on normal activity and to work; no special care needed.

Score 50–70: Unable to work; able to live at home and care for most personal needs; varying amount of assistance needed.

Score 10-60: Unable to care for self; requires equivalent of institutional or hospital care; disease may be progressing rapidly.

```

> hodg2 <- hodg
> hodg2$gtype <- with(hodg,factor(gtype,labels=c("Allo","Auto")))
> table (hodg2$gtype,hodg$gtype)

      1  2
Allo 16  0
Auto  0 27
> hodg2$dtype <- with(hodg,factor(dtype,labels=c("NHL","HOD")))
> table (hodg2$dtype,hodg$dtype)

      1  2
NHL 23  0
HOD  0 20

> with(hodg2,(table(gtype,dtype)))
      dtype
gtype  NHL  HOD
Allo   11   5
Auto   12  15

```

```
> hodg.surv <- with(hodg2, Surv(time, delta))
> hodg.cox1 <- coxph(hodg.surv~gtype*dtype+score+wtime, data=hodg2)
> summary(hodg.cox1)
```

n= 43, number of events= 26

	coef	exp(coef)	se(coef)	z	Pr(> z )	
gtypeAuto	0.63940	1.89534	0.59372	1.077	0.28151	
dtypeHOD	2.76033	15.80504	0.94738	2.914	0.00357	**
score	-0.04948	0.95172	0.01242	-3.984	6.77e-05	***
wtime	-0.01656	0.98357	0.01021	-1.623	0.10461	
gtypeAuto:dtypeHOD	-2.37093	0.09339	1.03548	-2.290	0.02204	*

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

	exp(coef)	exp(-coef)	lower .95	upper .95
gtypeAuto	1.89534	0.52761	0.59199	6.0682
dtypeHOD	15.80504	0.06327	2.46821	101.2066
score	0.95172	1.05073	0.92884	0.9752
wtime	0.98357	1.01670	0.96409	1.0034
gtypeAuto:dtypeHOD	0.09339	10.70738	0.01227	0.7108



```
> hodg.surv <- with(hodg2, Surv(time, delta))
> hodg.cox1 <- coxph(hodg.surv~gtype*dtype+score+wtime, data=hodg2)
> summary(hodg.cox1)
```

```
Concordance= 0.776 (se = 0.061 )
Rsquare= 0.527 (max possible= 0.983 )
Likelihood ratio test= 32.15 on 5 df, p=5.539e-06
Wald test = 27.19 on 5 df, p=5.232e-05
Score (logrank) test = 37.7 on 5 df, p=4.325e-07
```

# Proportionality

If two hazard functions are proportional, then the cumulative hazard functions are proportional. This means that the log cumulative hazard functions are parallel.

$$h_2(t) = ph_1(t)$$

$$H_2(t) = pH_1(t)$$

$$\ln[H_2(t)] = \ln[p] + \ln[H_1(t)]$$

also

$$\ln[H(t)] = \ln[-\ln(S(t))]$$

The function on the RHS of this equation,  $\ln[-\ln(\cdot)]$  is called the complementary log-log, or cloglog. The negation is required because  $S(t) \leq 1$ .

# Proportionality

- We first graph the survival function for the four combinations of disease type and graft type.
- We graph the complimentary log-log survival for the four groups—these should be parallel with proportional hazards.
- Then we graph the observed vs. expected survival functions.
- There appear to be problems with proportionality.

```

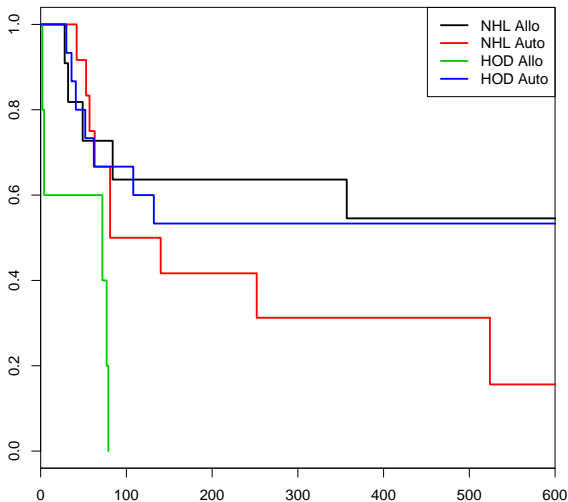
plot1 <- function(){
  plot(survfit(hodg.surv~dtype+gtype,data=hodg2),xlim=c(0,600),col=1:4,lwd=2)
  legend("topright",c("NHL Allo","NHL Auto","HOD Allo","HOD Auto"),col=1:4,lwd=2)
  title("Survival Curves for HOD/NHL and Allo/Auto Grafts")
}

plot2 <- function(){
  plot(survfit(hodg.surv~dtype+gtype,data=hodg2,type="fleming"),
       col=1:4,lwd=2,fun="cloglog")
  legend("topleft",c("NHL Allo","NHL Auto","HOD Allo","HOD Auto"),col=1:4,lwd=2)
  title("Complimentary Log-Log Survival Curves")
}

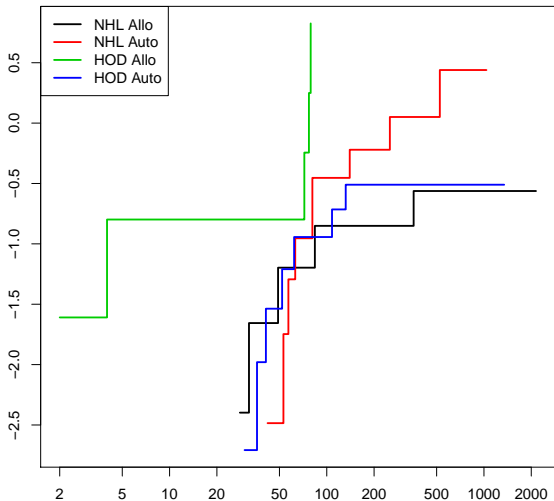
plot3 <- function(){
# score and wtime set to mean values for disease and graft types
  plot(survfit(hodg.surv~dtype+gtype,data=hodg2),xlim=c(0,600),col=1:4,lwd=2)
  lines(survfit(hodg.cox1,data.frame(gtype=c("Allo","Auto","Allo","Auto"),
    dtype=c("NHL","NHL","HOD","HOD"),score=c(75,76,56,85),
    wtime=c(17,23,59,58)),data=hodg2),col=1:4,lwd=2,lty=2)
  legend("topright",c("NHL Allo","NHL Auto","HOD Allo","HOD Auto"),col=1:4,lwd=2)
  title("Observed and Expected Survival Curves")
}

```

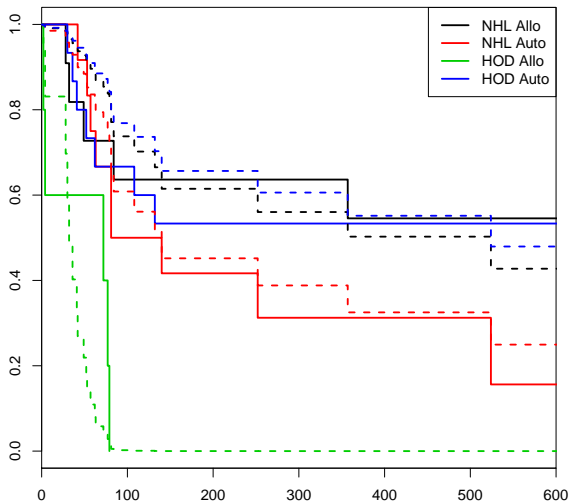
### Survival Curves for HOD/NHL and Allo/Auto Grafts



### Complimentary Log-Log Survival Curves



## Observed and Expected Survival Curves



# Predictions and Residuals

- In linear regression, we have a linear predictor for each data point  $i$

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}$$

$$\hat{y}_i = \hat{\eta}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_p x_{pi}$$

$$y_i \sim N(\eta_i, \sigma^2)$$

- This point prediction is an estimate of the conditional mean of  $y_i$  given the covariate values of point  $i$ . This together with the prediction error says that we are predicting the distribution of values of  $y$ .



# Predictions and Residuals

- The usual residual is  $r_i = y_i - \hat{y}_i$  which is the difference between the actual value of  $y$  and a prediction of its mean.
- The residuals are also the quantities the sum of whose squares is being optimized by the least squares/MLE estimation.

# Predictions and Residuals

- In survival analysis by Cox regression, the equivalent of  $y_i$  is the event time, which is not known for the censored observations.
- Because the base hazard is not defined by a continuous formula, there is no such thing as the “expected” (predicted mean) survival time for a subject, so there is no single predicted value.
- We do have a predicted distribution of survival, which is defined by the base hazard times  $\hat{\theta}_i$ , but since there is no point prediction, there is no point residual.

# Predictions and Residuals

- Even if we had a continuous distribution the nature of time-to-event data results in very wide prediction intervals.
- Suppose a cancer patient is predicted to have a mean lifetime of 5 years after diagnosis and suppose the distribution is exponential.
- If we want a 95% interval for survival, the lower end is at the 0.025 percentage point of the exponential which is 0.13 years (1.5 months), or  $1/40$  of the mean lifetime.
- The upper end is at the 0.975 point which is 18.4 years, or 3.7 times the mean lifetime.

# Predictions and Residuals

- Saying that the survival time is somewhere between 6 weeks and 18 years does not seem very useful but it may be the best we can do.
- For survival analysis, something is like a residual if it is small when the model is accurate or if the accumulation of them is in some way minimized by the estimation algorithm, but there is no exact equivalence to linear regression residuals.
- There are a number of quantities that can be thought of as like residuals in some way; they can be quite large!

# Types of Residuals

- It is often hard to make a decision from graph appearances, though the process can reveal much.
- Some diagnostic tests are based on residuals as with other regression methods.
- We use Schoenfeld residuals (via `cox.zph`) to test for proportionality.
- We use Cox-Snell residuals to test for goodness of fit.
- We use martingale residuals to look for non-linearity.
- We can also look at `dfbeta` for influence.

## residuals.coxph {survival} R Documentation

### Calculate Residuals for a 'coxph' Fit

#### Description

Calculates martingale, deviance, score, or Schoenfeld residuals for a Cox proportional hazards model.

#### Usage

```
residuals(object,  
  type=c("martingale", "deviance", "score", "schoenfeld",  
  "dfbeta", "dfbetas", "scaledsch", "partial"),  
  collapse=FALSE, weighted=FALSE, ...)
```

#### Arguments

**object** an object inheriting from class `coxph`, representing a fitted Cox regression model. Typically this is the output from the `coxph` function.

```
residuals(object,  
  type=c("martingale", "deviance", "score", "schoenfeld",  
  "dfbeta", "dfbetas", "scaledsch", "partial"),  
  collapse=FALSE, weighted=FALSE, ...)
```

## Arguments

- object** an object inheriting from class `coxph`, representing a fitted Cox regression model. Typically this is the output from the `coxph` function.
- type** character string indicating the type of residual desired. Possible values are "martingale", "deviance", "score", "schoenfeld", "dfbeta", "dfbetas", and "scaledsch". Only enough of the string to determine a unique match is required.
- collapse** vector indicating which rows to collapse (sum) over. In time-dependent models more than one row data can pertain to a single individual. If there were 4 individuals represented by 3, 1, 2 and 4 rows of data respectively, then `collapse=c(1,1,1, 2, 3,3, 4,4,4,4)` could be used to obtain per subject rather than per observation residuals.

For martingale and deviance residuals, the returned object is a vector with one element for each subject (without collapse). This means that even censored observations have a martingale residual and a deviance residual. Each subject's value for the martingale residual and the deviance residual is a single value.



All observations also have a score residual, though this is a vector not a scalar. For score residuals the returned object is a matrix with one row per subject and one column per variable. The row order will match the input data for the original fit. The score residuals are each individual's contribution to the score vector. Two transformations of this are often more useful: `dfbeta` is the approximate change in the coefficient vector if that observation were dropped, and `dfbetas` is the approximate change in the coefficients, scaled by the standard error for the coefficients.

For Schoenfeld residuals, the returned object is a matrix with one row for each event and one column per variable. The rows are ordered by time within strata, and an attribute **strata** is attached that contains the number of observations in each strata. The scaled Schoenfeld residuals are used in the `cox.zph()` function. Only subjects with an observed event time have a Schoenfeld residual, which like the score residual is a vector.

# Schoenfeld Residuals

There is a Schoenfeld residual for each subject  $i$  with an event (not censored) and for each predictor  $x_k$ . At the event time  $t$  for that subject, there is a risk set  $R$ , and each subject  $j$  in the risk set has a risk coefficient  $\theta_j$  and also a value  $x_{jk}$  of the predictor. The Schoenfeld residual is the difference between  $x_{ik}$  and the risk-weighted average of all the  $x_{jk}$  over the risk set.

$$r_{ik}^S = x_{ik} - \frac{\sum_{j \in R} x_{jk} \theta_j}{\sum_{j \in R} \theta_j}$$

# Schoenfeld Residuals

This is a measure of how typical the individual subject is with respect to the covariate at the time of the event. Since subjects should fail more or less uniformly according to risk, the Schoenfeld residuals should be approximately level over time, not increasing or decreasing. We can test this by a score test for whether a linear trend of the residuals has slope 0 with respect to time on some scale, which could be the time itself, the log time, or the rank in the set of failure times. Previously, instead of the score test, a test of correlation of the Schoenfeld residuals with the time variable was used. The default is to use the KM curve as a transform, which is similar to the rank but deals better with censoring. The old correlation test and the new score test are asymptotically identical but not in small samples.

```
cox.zph(fit, transform="km", terms=TRUE, singledf=FALSE, global=TRUE)
```

### Arguments

- fit** the result of fitting a Cox regression model, using the `coxph` or `coxme` functions.
- transform** a character string specifying how the survival times should be transformed before the test is performed. Possible values are "km", "rank", "identity" or a function of one argument.
- terms** if TRUE, do a test for each term in the model rather than for each separate covariate. For a factor variable with  $k$  levels, for instance, this would lead to a  $k-1$  degree of freedom test. The plot for such variables will be a single curve evaluating the linear predictor over time.
- singledf** use a single degree of freedom test for terms that have multiple coefficients, i.e., the test that corresponds most closely to the plot. If `terms=FALSE` this argument has no effect.
- global** should a global chi-square test be done, in addition to the per-variable or per-term tests.

Value  
an object of class "cox.zph", with components:

table a matrix with one row for each variable, and optionally a last row for the global test. Columns of the matrix contain a score test of for addition of the time-dependent term, the degrees of freedom, and the two-sided p-value.

x the transformed time axis.

time the untransformed time values; there is one entry for each event time in the data

strata for a stratified coxph model, the stratum of each of the events

y the matrix of scaled Schoenfeld residuals. There will be one column per term or per variable (depending on the terms option above), and one row per event. The row labels are a rounded form of the original times.

var a variance matrix for the covariates, used to create an approximate standard

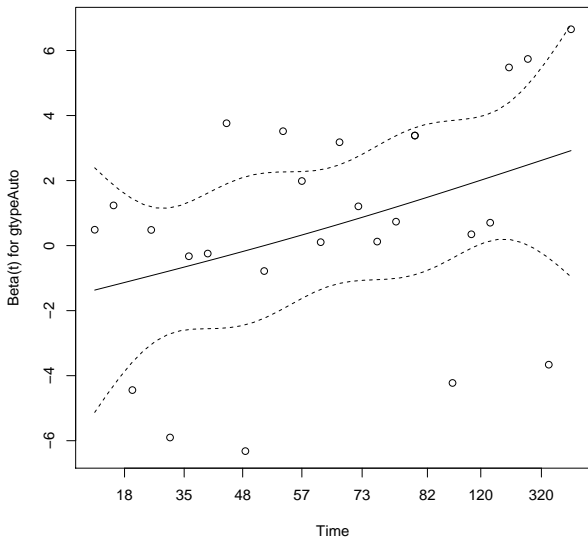
transform the transform of time that was used

call the calling sequence for the routine.

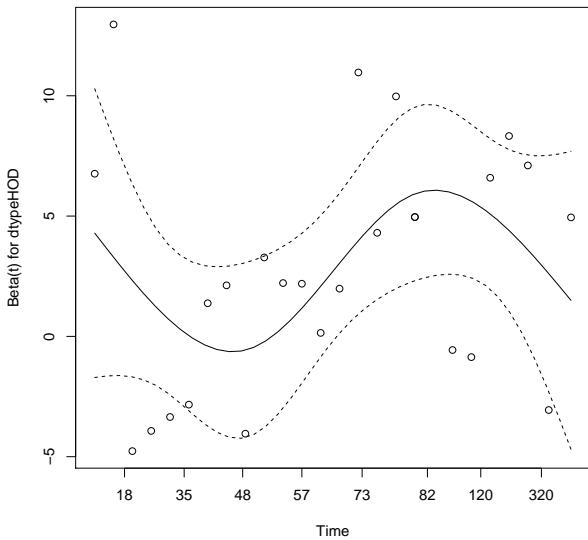
```
> hodg.zph <- cox.zph(hodg.cox1)
> print(hodg.zph)
```

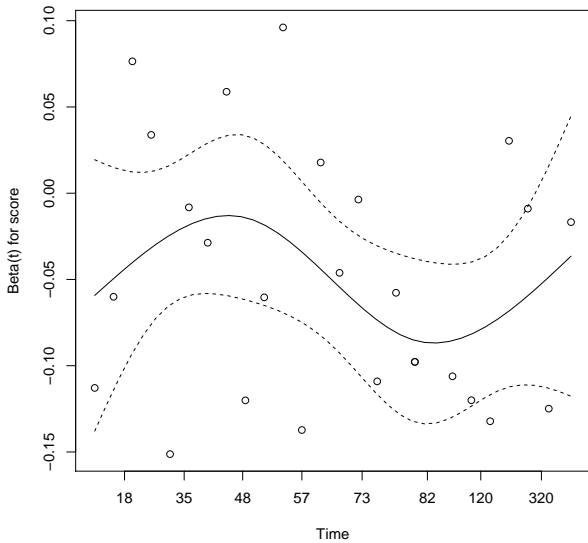
	chisq	df	p
gtype	0.5400	1	0.462
dtype	1.8012	1	0.180
score	3.8805	1	0.049
wtime	0.0173	1	0.895
gtype:dtype	4.0474	1	0.044
GLOBAL	13.7573	5	0.017

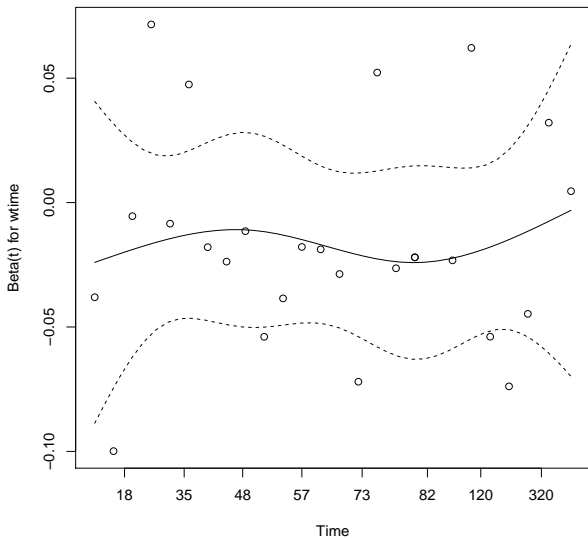
```
pdf("hodgzph1.pdf")
plot(hodg.zph[1])
dev.off()
pdf("hodgzph2.pdf")
plot(hodg.zph[2])
dev.off()
pdf("hodgzph3.pdf")
plot(hodg.zph[3])
dev.off()
pdf("hodgzph4.pdf")
plot(hodg.zph[4])
dev.off()
pdf("hodgzph5.pdf")
plot(hodg.zph[5])
dev.off()
```

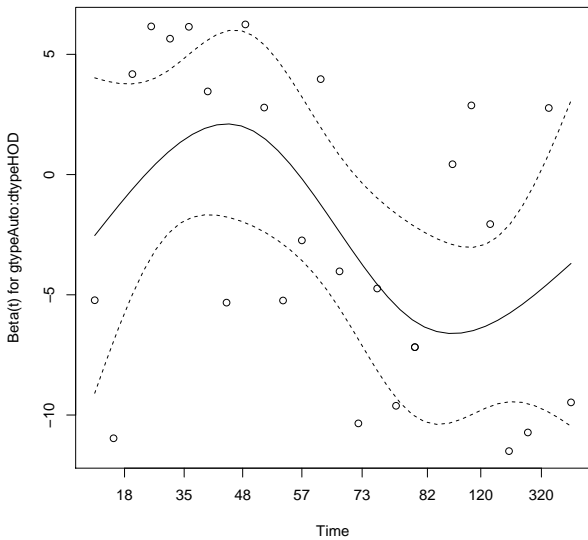












- From the score test, the Karnofsky score and the gtype:dtype interaction have modest but statistically significant non-proportionality, the global test is also statistically significant.
- The sample size here is relatively small (26 events in 43 subjects). If the sample size is large, very small amounts of non-proportionality can induce a significant result.
- As time goes on, autologous grafts are over-represented at their own event times, but those from HOD patients become less represented.
- Both the statistical tests and the plots are useful.

# Goodness of Fit using the Cox-Snell Residuals

Suppose that the  $i^{\text{th}}$  individual has a survival time  $T_i$  which has survival function  $S_i(t)$ , meaning that  $\Pr(T_i > t) = S_i(t)$ . Then  $S_i(T_i)$  has a uniform distribution on  $(0, 1)$ .

$$\begin{aligned}\Pr(S_i(T_i) \leq u) &= \Pr(T_i > S_i^{-1}(u)) \\ &= S_i(S_i^{-1}(u)) \\ &= u\end{aligned}$$

# Goodness of Fit using the Cox-Snell Residuals

Also, if  $U$  has a uniform distribution on  $(0, 1)$ , then what is the distribution of  $-\ln(U)$ ?

$$\begin{aligned}\Pr(-\ln(U) < x) &= \Pr(U > \exp(-x)) \\ &= 1 - e^{-x}\end{aligned}$$

which is the CDF of an exponential distribution with parameter  $\lambda = 1$ .

# Goodness of Fit using the Cox-Snell Residuals

So,  $r_i^{CS} = \hat{\Lambda}_i(t_i) = -\ln[\hat{S}_i(t_i)] = -\ln[\hat{S}(t_i|\text{covariates})]$  should have an exponential distribution with constant hazard  $\lambda = 1$  if the estimate  $\hat{S}_i$  is accurate, which means that these values should look like a censored sample from this exponential distribution. These values are called *generalized residuals* or *Cox-Snell residuals*.



# Martingale Residuals

The *martingale residuals* are a slight modification of the Cox-Snell residuals. If the censoring indicator is  $\delta_i$ , then

$$r_i^M = \delta_i - r_i^{CS}$$

These residuals can be interpreted as an estimate of the excess number of events seen in the data but not predicted by the model. We will use these to examine the functional form of covariates.

# Martingale

Originally, a martingale referred to a betting strategy where you bet \$1 on the first play, then you double the bet if you lose and continue until you win. This seems like a sure thing, because at the end of each series when you finally win, you are up \$1. For example,  $-1 - 2 - 4 - 8 + 16 = 1$ . But this assumes that you have infinite resources. Really, you have a large probability of winning \$1, and a small probability of losing everything you have, kind of the opposite of a lottery.

# Martingale

In probability, a *martingale* is a sequence of random variables such that the expected value of the next event at any time is the present observed value, and that no better predictor can be derived even with all past values of the series available. At least to a close approximation, the stock market is a martingale. Under the assumptions of the proportional hazards model, the martingale residuals ordered in time form a martingale.

# Using Martingale Residuals

Martingale residuals can be used to examine the functional form of a numeric variable. We fit the model without that variable and compute the martingale residuals. We then plot these martingale residuals against the values of the variable. We can see curvature, or a possible suggestion that the variable can be discretized. We will use this to examine the `score` and `wtime` variables in the `hodg` data set.

```

hodg.mart <- residuals(hodg.cox1,type="martingale")
hodg.cs <- hodg$delta-hodg.mart

plot1r <- function(){
  surv.csr = survfit(Surv(hodg.cs,hodg2$delta)~1,type="fleming-harrington")
  plot(surv.csr,fun="cumhaz")
  abline(0,1)
  title("Cumulative Hazard of Cox-Snell Residuals")
}

plot2r <- function(){
  mres <- residuals(coxph(hodg.surv~gtype*dtype+wtime,data=hodg2),type="martingale")
  plot(hodg2$score,mres,xlab="Karnofsky Score",ylab="Martingale Residuals")
  lines(lowess(hodg2$score,mres))
  title("Martingale Residuals vs. Karnofsky Score")
}

```

```

hodg.mart <- residuals(hodg.cox1,type="martingale")
hodg.cs <- hodg$delta-hodg.mart

plot3r <- function(){
  mres <- residuals(coxph(hodg.surv~gtype*dtype+score,data=hodg2),type="martingale")
  plot(hodg2$wtime,mres,xlab="Waiting Time",ylab="Martingale Residuals")
  lines(lowess(hodg2$wtime,mres))
  title("Martingale Residuals vs. Waiting Time")
  print(head(cbind(hodg2$wtime,mres)[order(hodg2$wtime,decreasing=T),]))
}

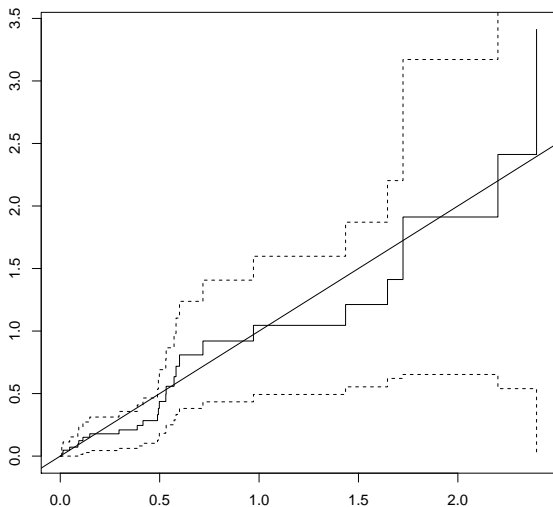
```

```

          mres
41 171 -0.6099433
15 102 -1.2045188
43  98 -1.0541449
28  84 -0.5916094
40  84 -0.5065709
29  73  0.9774249

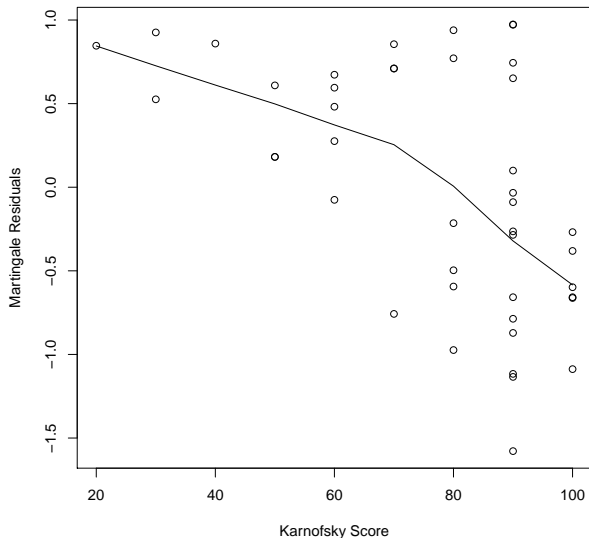
```

### Cumulative Hazard of Cox-Snell Residuals



The line with slope 1 and intercept 0 fits the curve relatively well, so we don't see lack of fit using this procedure.

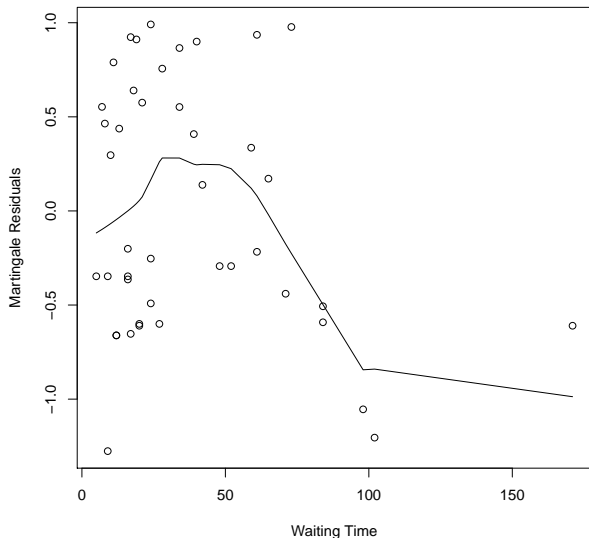
### Martingale Residuals vs. Karnofsky Score



The line is almost straight. It could be some modest transformation of the Karnofsky score would help, but it might not make much difference. Maybe a 5pt difference above 70 is more important than below 70.



## Martingale Residuals vs. Waiting Time



The line could suggest a step function. To see where the drop is, we can look at the largest waiting times and the associated martingale residual.

```

hodg.mart <- residuals(hodg.cox1,type="martingale")
hodg.cs <- hodg$delta-hodg.mart

plot3r <- function(){
  mres <- residuals(coxph(hodg.surv~gtype*dtype+score,data=hodg),type="martingale")
  plot(hodg$wtime,mres,xlab="Waiting Time",ylab="Martingale Residuals")
  lines(lowess(hodg$wtime,mres))
  title("Martingale Residuals vs. Waiting Time")
  print(head(cbind(hodg$wtime,mres)[order(hodg$wtime,decreasing=T),]))
}

```

```

          mres
41 171 -0.6099433
15 102 -1.2045188
43  98 -1.0541449
28  84 -0.5916094
40  84 -0.5065709
29  73  0.9774249

```

The martingale residuals are all negative for wtime >83 and positive for the next smallest value.

A reasonable cut-point is 80 days.

We reformulate the model with dichotomized wtime.

```
wt2 <- cut(hodg2$wtime,c(0,80,200),labels=c("short","long"))
hodg.cox2 <- coxph(hodg.surv~gtype*dtype+score+wt2,data=hodg2)
print(drop1(hodg.cox1,test="Chisq"))
```

Model:

```
hodg.surv ~ gtype * dtype + score + wtime
      Df    AIC    LRT Pr(>Chi)
<none>      152.36
score      1 167.60 17.2365 3.3e-05 ***
wtime      1 153.64 3.2792 0.07016 .
gtype:dtype 1 155.80 5.4357 0.01973 *
```

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

```
print(drop1(hodg.cox2,test="Chisq"))      #New model has better AIC
                                           #and smaller p-values.
```

Model:

```
hodg.surv ~ gtype * dtype + score + wt2
      Df    AIC    LRT Pr(>Chi)
<none>      149.03
score      1 168.64 21.6042 3.351e-06 ***
wt2        1 153.64 6.6081 0.01015 *
gtype:dtype 1 152.00 4.9697 0.02580 *
```

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

# Checking for Outliers and Influential Observations

We will check for outliers using the deviance residuals. The martingale residuals show excess events or the opposite, but highly skewed, with the maximum possible value being 1, but the smallest value can be very large negative. Martingale residuals can detect unexpectedly long-lived patients, but patients who die unexpectedly early show up only in the deviance residual. Influence will be examined using  $dfbeta$  in a similar way to linear regression, logistic regression, or Poisson regression.

# Deviance Residuals

The deviance residuals are defined by

$$r_i^D = \text{sign}(r_i^M) \sqrt{-2 [r_i^M + \delta_i \ln(\delta_i - r_i^M)]}$$

$$r_i^D = \text{sign}(r_i^M) \sqrt{-2 [r_i^M + \delta_i \ln(r_i^{CS})]}$$

Roughly centered on 0 with approximate standard deviation 1.

```

hodg.mart <- residuals(hodg.cox2,type="martingale")
hodg.dev <- residuals(hodg.cox2,type="deviance")
hodg.dfb <- residuals(hodg.cox2,type="dfbeta")
hodg.preds <- predict(hodg.cox2)                #linear predictor

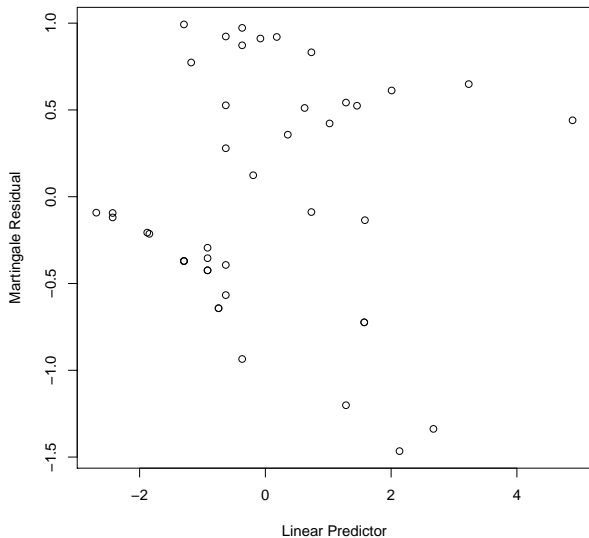
plotr21 <- function(){
  plot(hodg.preds,hodg.mart,xlab="Linear Predictor",ylab="Martingale Residual")
  title("Martingale Residuals vs. Linear Predictor")
}

plotr22 <- function(){
  plot(hodg.preds,hodg.dev,xlab="Linear Predictor",ylab="Deviance Residual")
  title("Deviance Residuals vs. Linear Predictor")
}

plotr23 <- function(){
  plot(hodg.dfb[,1],xlab="Observation Order",ylab="dfbeta for Graft Type")
  title("dfbeta Values by Observation Order for Graft Type")
}
.....

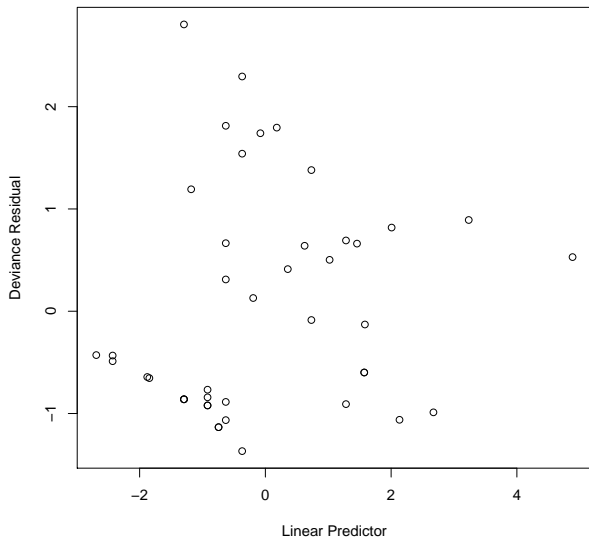
```

## Martingale Residuals vs. Linear Predictor



The smallest three martingale residuals in order are observations 1, 29, and 18.

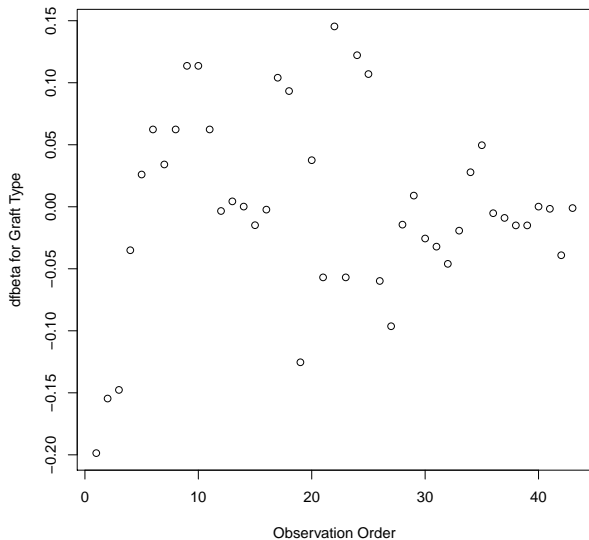
## Deviance Residuals vs. Linear Predictor



The two largest deviance residuals are observations 1 and 29. Worth examining.

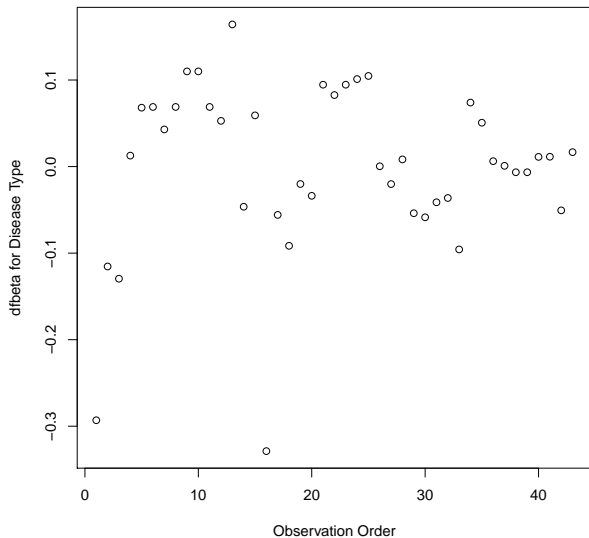


dfbeta Values by Observation Order for Graft Type



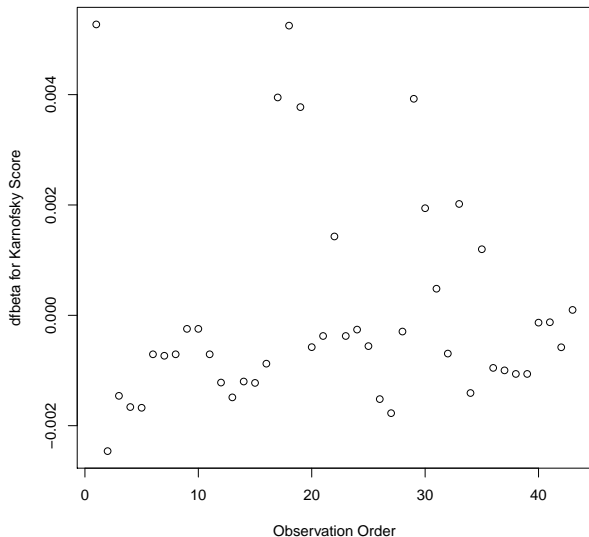
The smallest dfbeta for graft type is observation 1.

dfbeta Values by Observation Order for Disease Type



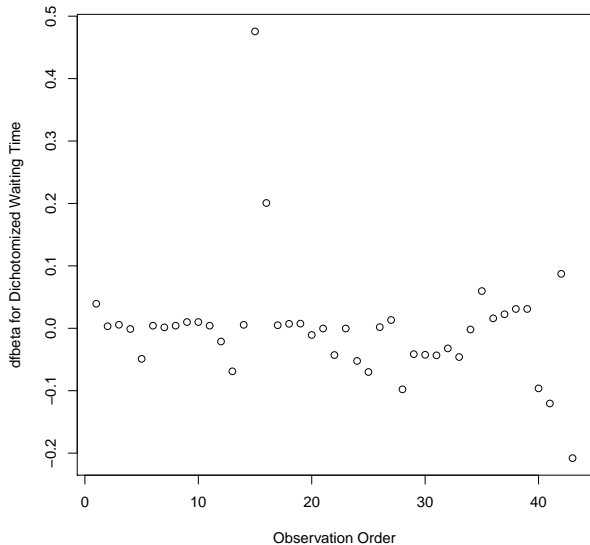
The smallest two dfbeta values for disease type are observations 1 and 16.

### dfbeta Values by Observation Order for Karnofsky Score



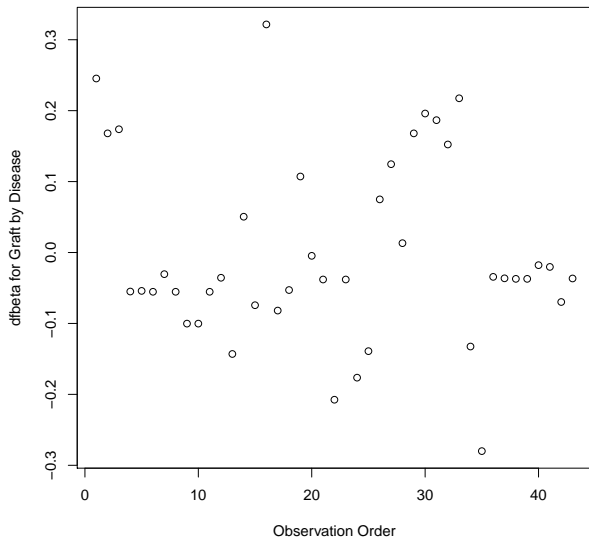
The two highest dfbeta values for score are observations 1 and 18. The next three are observations 17, 29, and 19. The smallest value is observation 2.

### dfbeta Values by Observation Order for Dichotomized Waiting Time



The two large values of dfbeta for dichotomized waiting time are observations 15 and 16. This may have to do with the discretization of waiting time.

dfbeta Values by Observation Order for Graft by Disease



The two largest values are observations 1 and 16. The smallest value is observation 35.

**Table:** Observations to Examine by Residuals and Influence

---

Martingale Residuals	1, 29, 18
Deviance Residuals	1, 29
Graft Type Influence	1
Disease Type Influence	1, 16
Karnofsky Score Influence	1, 18 (17, 29, 19)
Waiting Time Influence	15, 16
Graft by Disease Influence	1, 16, 35

---

The most important observations to examine seem to be 1, 15, 16, 18, and 29.

```

> with(hodg,summary(time[delta==1]))
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.00  41.25   62.50   97.62  83.25  524.00
> with(hodg,summary(wtime))
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  5.0   16.0   24.0   37.7   55.5   171.0
> with(hodg,summary(score))
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 20.00  60.00   80.00   76.28  90.00  100.00

```

```
> hodg.cox2
```

	coef	exp(coef)	se(coef)	z	p
gtypeAuto	0.6651	1.9447	0.5943	1.12	0.2631
dtypeHOD	2.3273	10.2505	0.7332	3.17	0.0015
score	-0.0550	0.9464	0.0123	-4.46	8.2e-06
wt2long	-2.0598	0.1275	1.0507	-1.96	0.0499
gtypeAuto:dtypeHOD	-2.0668	0.1266	0.9258	-2.23	0.0256

```
> hodg[c(1,15,16,18,29),]
```

	gtype	dtype	time	delta	score	wtime	
1	1	1	28	1	90	24	#early death, good score, low risk grp
15	1	2	77	1	60	102	#high risk grp, long wait, poor score
16	1	2	79	1	70	71	#high risk grp, short wait, poor score
18	2	1	53	1	90	17	#early death, good score, med risk grp
29	2	2	30	1	90	73	#early death, good score, med risk grp

# Action Items

- Unusual points may need checking, particularly if the data are not completely cleaned. In this case, observations 15 and 16 may show some trouble with the dichotomization of waiting time, but it still may be useful.
- The two largest residuals seem to be due to unexpectedly early deaths, but unfortunately this can occur.



- If hazards don't look proportional, then we may need to use strata, between which the base hazards are permitted to be different. For this problem, the natural strata are the two diseases, because they could need to be managed differently anyway.
- A main point that we want to be sure of is the relative risk difference by disease type and graft type.

Table: Linear Risk Predictors for Lymphoma

Disease	Graft Type	Linear Predictor
Non-Hodgkin	Allogenic	0
Non-Hodgkin	Autologous	0.6651
Hodgkin's	Allogenic	2.3273
Hodgkin's	Autologous	0.9256

For Non-Hodgkin, the allogenic graft is better. For Hodgkin's, the autologous graft is much better.