Solution TMA4275, june 2017

1a) The estimate of $R(100)$ for males is

$$\hat{R}(100) = \prod_{T_{(i)} \leq 100} \left(1 - \frac{d_i}{n_i}\right)$$

$$= \left(1 - \frac{1}{7}\right)\left(1 - \frac{1}{6}\right) \cdots \left(1 - \frac{1}{2}\right)$$

$$= \frac{1}{7} = 0.1428$$

Estimates of median lifetimes are 59 days for males and 116 days for females.

For males an estimate of $ET = \int_0^\infty R_M(t)\,dt$

is $\widehat{ET} = \int_0^{161} \hat{R}_M(t)\,dt \approx 50$ days

For females, assuming that $R(t) = 0$ for $t > 365$,

$$\widehat{ET} = \int_0^{365} \hat{R}_F(t)\,dt \approx 100 \text{ days}$$

but this has negative bias.

b) We are testing

$$H_0 : R_M(t) = R_F(t)$$

vs.

$$H_1 : R_M(t) \neq R_F(t)$$

At the time of the failure of unit $i=3$, $y_i = 20$, $n_{0j} = 11$ females and $n_{1j} = 5$ males are at risk.

$d_{0j} = 1$ female and $d_{1j} = 0$ males failed.

Under H_0

$$E(d_{0j}) = e_{0j} = \frac{11}{16} \quad \text{and} \quad E(d_{1j}) = e_{1j} = \frac{5}{16}$$

Summing over all observed failures gives the numbers in the R output for the log-rank test

c) The last model assumes that the hazard function for a subject with covariate vector $\mathbf{x}_i$ is

$$z(t; \mathbf{x}_i) = z_0(t) e^{\mathbf{x}_i \beta}$$

The covariate vector

$$\mathbf{x}_i = (\text{sexmale}_i, \text{usage}_i)$$

where $\text{sexmale}_i = \begin{cases} 1 & \text{for males} \\ 0 & \text{for females.} \end{cases}$

To test for an effect of sex with usage included in the model we compare model cox1 (H_0) to model cox12 (H_1). Under H_0

$$2(l_1 - l_0) \overset{\text{approx}}{\sim} \chi^2_{p_1 - p_0}$$

Critical value: $\chi^2_{0.05, 2-1} = 3.84$.

Observed value

$$2(-33.36 + 33.47) = 0.22.$$

Conclusion: We can not reject H_0.

Without sex included usage is still significant. Thus we prefer model cox 2 as our best model.
This model also has the smallest AIC.

The hazard change by a factor

$$e^{\beta_{usage}} = e^{0.1262} = 1.135$$

per minute increase in usage, that is, 13.5 % increase.

d) The Schoenfeld residuals is the difference between the observed and expected covariate value at the i'th failure based on the fitted cox model.

If the prop. haz. assumption holds (H_0),

$$E(res_i) = 0$$

If the effect of a covariate on the hazard increase over time, this translates to a trend in the Schoenfeld residuals.

The residuals i fig 2 do not indicate any deviation from the expected behaviour under H_0.

e) A subject with covarriate vector x_i has hazard function

$$z(t; x_i) = z_0(t) e^{x_i\beta}$$

and survival function

$$R(t; x_i) = e^{-\int_0^t z(u; x_i)\, du}$$

$$= e^{-\int_0^t z_0(u) e^{x_i\beta} du}$$

$$= \left(e^{-\int_0^t z_0(u) du}\right)^{\left(e^{x_i\beta}\right)}$$

$$= R_0(t)^{\left(e^{x_i\beta}\right)}$$

Hence,

$$\widehat{P(T>100;\ x=10)} = \hat{R}(100;\ x=10)$$

$$= \hat{R}_0(100)^{e^{\hat{\beta}\cdot 10}}$$

$$= 0.38^{\left(e^{0.1262\cdot 10}\right)}$$

$$= 0.38^{3.53} = 0.033.$$

f) The model assumes that

$$\ln T_i = \beta_0 + \beta_1 x_i + \sigma U_i$$

where T_i is the lifetime of the i'th unit and $U_i \overset{iid}{\sim} N(0,1)$.

$$p = P(T>t) = P(\ln T > \ln t)$$

$$= P(\beta_0 + \beta_1 x + \sigma U > \ln t)$$

$$= 1 - P\left(U \le \frac{\ln t - \beta_0 - \beta_1 x}{\sigma}\right)$$

$$= 1 - \Phi\left(\frac{\ln t - \beta_0 - \beta_1 x}{\sigma}\right)$$

$$\hat{p} = 1 - \Phi\left(\frac{\ln 100 - 5.4 + 0.08 \cdot 10}{1.02}\right)$$

$$= 0.49$$

Using the delta-method,

$$\mathrm{Var}(\hat{p}) \approx \left(\frac{\partial p}{\partial \beta_0}\right)^2 \mathrm{Var}(\hat{\beta}_0) + \left(\frac{\partial p}{\partial \beta_1}\right)^2 \mathrm{Var}(\hat{\beta}_1) + \left(\frac{\partial p}{\partial \ln \sigma}\right)^2 \mathrm{Var}(\ln \hat{\sigma})$$

$$+ 2\left(\frac{\partial p}{\partial \beta_0}\right)\left(\frac{\partial p}{\partial \beta_1}\right) \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1)$$

$$+ 2\left(\frac{\partial p}{\partial \beta_1}\right)\left(\frac{\partial p}{\partial \ln \sigma}\right) \mathrm{Cov}(\hat{\beta}_0, \ln \hat{\sigma})$$

$$+ 2\left(\frac{\partial}{\partial \beta_1}\right)\left(\frac{\partial p}{\partial \ln \sigma}\right) \mathrm{Cov}(\hat{\beta}_1, \ln \hat{\sigma})$$

where

$$\frac{\partial p}{\partial \beta_0} = \phi\left(\frac{\ln t - \beta_1 - \beta_1 x}{\sigma}\right) \frac{1}{\sigma},$$

$$\frac{\partial p}{\partial \beta_1} = \phi\left(\frac{\ln t - \beta_0 - \beta_1 x}{\sigma}\right)\frac{x}{\sigma}$$

$$\frac{\partial p}{\partial \ln\sigma} = \frac{\partial}{\partial \ln\sigma}\left[1 - \Phi\left((\ln t - \beta_0 - \beta_1 x)e^{-\ln\sigma}\right)\right]$$

$$= -\phi\left(\frac{\ln t - \beta_0 - \beta_1 x}{\sigma}\right)e^{-\ln\sigma}(-1)$$

$$= \phi\left(\frac{\ln t - \beta_0 - \beta_1 x}{\sigma}\right)\frac{1}{\sigma}$$

9) $ET = E e^{\beta_0 + \beta_1 x + \sigma U}$

$$= e^{\beta_0 + \beta_1 x} E(e^{\sigma U})$$

$$= e^{\beta_0 + \beta_1 x} M_U(\sigma)$$

$$= e^{\beta_0 + \beta_1 x} e^{\sigma^2/2}$$

$$= e^{\beta_0 + \beta_1 x + \sigma^2/2}$$

$$\widehat{ET} = e^{5.4 - 0.08 \cdot 10 + 1.02^2/2} = 167 \text{ days.}$$

If phone usage $X \sim \exp(\frac{1}{\theta})$, $M_X(t) = \frac{1}{1-\theta t}$,

$$ET = E e^{\beta_0 + \beta_1 X + \sigma U}$$

$$= e^{\beta_0} E(e^{\beta_1 X}) E(e^{\sigma U})$$

$$= e^{\beta_0} M_X(\beta_1) M_U(\sigma)$$

$$= e^{\beta_0} \frac{1}{1 - \theta\beta_1} e^{\sigma^2/2} = \frac{e^{5.4 + 1.02^2/2}}{1 - (-0.08) \cdot 10} = 207 \text{ days}$$

2a)

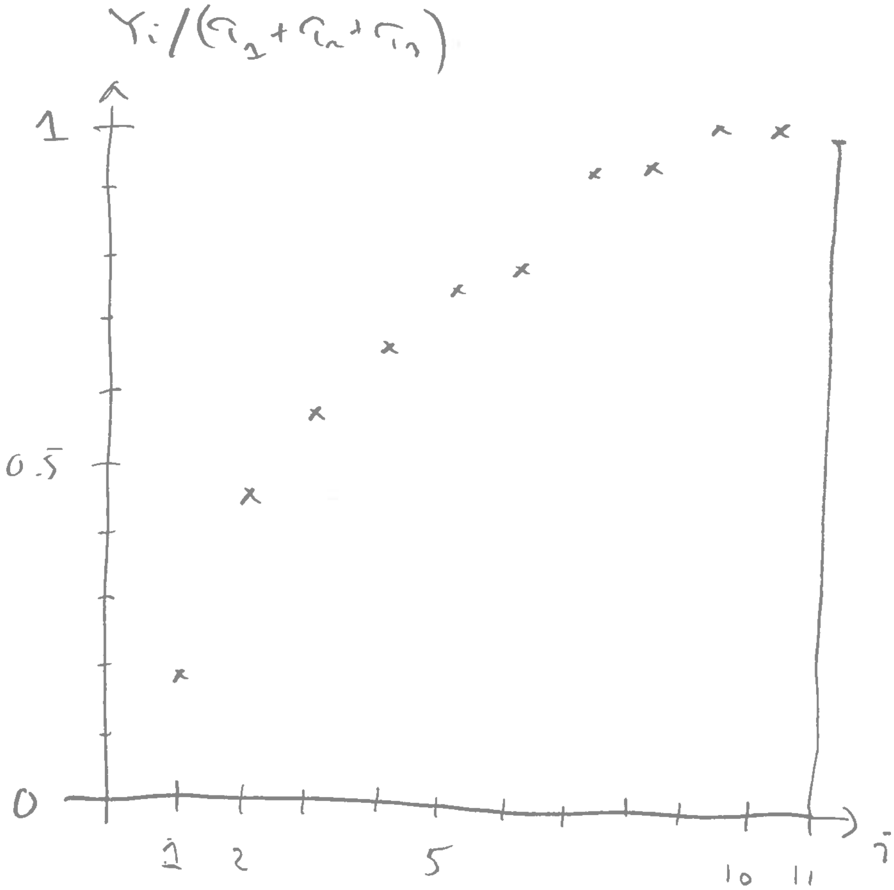
$t_{(i)}$	$Y_i = \tau(t_{(i)})$		$\frac{Y_i}{\tau_1+\tau_2+\tau_3} = \frac{Y_i}{35\ldots}$
21	$3 \cdot 21$	$= 63$	0.18
55	$150 + 2 \cdot 5$	$= 160$	0.45
75	$150 + 2 \cdot 25$	$= 200$	0.57
92	$150 + 2 \cdot 42$	$= 234$	0.66
122	$250 + 22$	$= 272$	0.77
125		$= 275$	0.78
173		323	0.92
178		328	0.93
190		340	0.97
195		345	0.98

Under H_0,
$Y_1, Y_2, \ldots$ are distributed as the arrival times in a HPP.

Conditional on 10 failures occurring $Y_1, Y_2, \ldots, Y_n$ are distributed as the order statistics of 10 iid $\mathrm{Unif}(0, \tau_1+\tau_2+\tau_3)$

Hence,

$$E\left(\frac{Y_i}{\tau_1+\tau_2+\tau_3}\right) = \frac{i}{n+1}$$



The concave shape of the plot indicates an increasing intensity $w(t)$

Under H_0, $\sum Y_i$ is approximately normal with

$$E\left(\sum Y_i\right) = E\left(\sum U_{(i)}\right) = E\left(\sum U_i\right) = n\frac{\tau}{2}$$

and

$$\mathrm{Var}\left(\sum Y_i\right) = \frac{n\tau^2}{12}$$

whee $\tau = \tau_1 + \tau_2 + \tau_3 = 350$

Laplace test:

$$Z = \frac{\sum \tau_i - n\tau/2}{\tau\sqrt{n/12}} = \frac{2540 - 1750}{319}$$

$= 2.47.$

Reject H_0 if $|z| > z_{\alpha/2} = 1.96$

Conclusion: Reject H_0.

b) Likelihood function

$$L(\theta) = \prod_{j=1}^{3} e^{-W(\tau_j;\theta)} \prod_{i=1}^{n_j} w(s_{ij};\theta)$$

If $w(t) = e^{\beta_0+\beta_1 t}$

we have $W(t) = \int_0^t w(u)\,du = \frac{1}{\beta_1} e^{\beta_0+\beta_1 t}$

and

$$\ell(\beta_0,\beta_1) = -\frac{1}{\beta_1}\prod_{j=1}^{3} e^{\beta_0+\beta_1\tau_j} + \sum_{j=1}^{3}\sum_{i=1}^{n_j}(\beta_0+\beta_1 s_{ij})$$

$$= -\frac{1}{\beta_1} e^{\beta_0+\beta_1\sum_{j=1}^{3}\tau_j} + \sum_{j=1}^{3} n_j\beta_0 + \beta_1\sum_{j=1}^{3}\sum_{i=1}^{n_j} s_{ij}$$

c) Wald test:

$$Z = \frac{0.018}{0.00435} = 4.13$$

Since $|Z| > z_{\alpha/2} = 1.96$ we reject $H_0: \beta_2 = 0$ (constant rate)

Likelihood ratio test:

Under H_0

$w(t) = w$,

the MLE of $\hat{w} = \frac{\sum n_j}{\sum \tau_j} = \frac{n}{\tau}$, and

the maximum log likelihood is

$$l_0 = \ln\left(e^{-\hat{w}\tau}\hat{w}^n\right)$$

$$= -\hat{w}\tau + n\ln w$$

$$= -\frac{n}{\tau}\cdot\tau + n\ln\frac{n}{\tau}$$

$$= -n + n(\ln n - \ln\tau)$$

$$= -45.55$$

This gives a likelihood ratio test statistic

$$2(\ell_1 - \ell_0) = 2(-43.17 + 45.55)$$

$$= 4.76 > \chi^2_{0.05,\, 2-1} = 3.84$$

Conclusion: Again, we reject H_0.