(16) **Task 1**

**A** (2) Time from production of the refrigerator until failure, in years (i.e. the lifetime of refrigerators).

**B** (2) Complete information is only available for refrigerators produced after January 1, 2010. Refrigerators that failed earlier than 2010 are not included. This means that **left truncation** is represented in this study.

**C** (8)

i) Here we have an exact observation (as exact as it gets when the unit is years), which means that there is **no censoring**.

ii) **Right censoring**, since we haven't observed the event at the end of the study.

iii) Another exact observation, **not censored**.

iv) **Right censoring** again, since we haven't observed the event at the end of the study.

**D** (4) Refrigerators produced before 2010 should be considered being sampled from a truncated distribution.

i) We have an event (exact observation) at 8 years, from a non-truncated distribution.

*ii*) This is a right censored observation at 5 years, from a non-truncated distribution.

*iii*) An event at 12 years, from a truncated distribution. We have to condition on surviving at least 1 year (2010-2009=1)

*iv*) A right censored observation at 22 years, from a truncated distribution. We have to condition on surviving at least 9 years (2010-2001=9)

$$\mathbf{L} = f(8) \cdot S(5) \cdot f(12)/S(1) \cdot S(22)/S(9)$$

(26) **Task 2**

**A** (2) The Nelson-Aalen estimator has better small-sample properties, but the sample here is relatively large in both groups (76 events in the smallest group), which means that the standard Kaplan-Meier estimator can be used.

**B** (8) Standard care:
**25$^{th}$ percentile = 4 months**. This means that 25% of the patients who received the standard care have had a remission within 4 months from study start.

**Median = 11 months**. This means that 50% of the patients who received the standard care have had a remission within 11 months from study start.

**75$^{th}$ percentile= 23 months**. This means that 75% of the patients who received the standard care have had a remission within 23 months from study start.

New treatment:
**25$^{th}$ percentile = 6 months**. This means that 25% of the patients who received the new treatment have had a remission within (at or before) 4 months from study start.

**Median = 18 months**. This means that 50% of the patients who received the new treatment have had a remission within 11 months from study start.

**75$^{th}$ percentile= 31 months**. This means that 75% of the patients who received the new treatment have had a remission within 23 months from study start.

These measures all have larger values for patients who received the new treatment than for the ones receiving standard care, which means that the remission free time is larger for the new treatment.

**C** (3) 24-month remission-free probability
For patients receiving standard care: **24.1%**
For patients receiving the new treatment: **49.4 %**

The probability of not having a remission within 6 months is a lot larger for patients that received the new treatment, more than twice as large as for patients that received the standard care.

**D** (13)

$$H_0 : h_{new\_trt}(t) = h_{standard\_care}(t) \text{ for all } t \le \tau$$
$$H_a : h_{new\_trt}(t) \ne h_{standard\_care}(t) \text{ for some } t \le \tau$$

Where $\tau$ = largest time at which both groups have at least one subject at risk.

**Test**: Log-rank or Gehan's/Wilcoxon test. Motivation needed.

**Assumptions:**

- Random sample – not specified in the task. Groups are however randomized, which is positive for the comparison.
- Independent samples – OK (reasonable to assume that the times to remission in the two different groups are independent)
- Non-informative/random censoring (i.e. that the censoring times are not related to the later, unknown, remission times) – not specified in the task (should be discussed with the clinician), but the censoring plot doesn't contradict this since the censoring pattern is similar in both groups.
- Right censored data – OK (reasonable to assume from the given information)
- Survival probabilities are the same for subjects recruited early and late in the study - not specified in the task (should be discussed with the clinician), but reasonable to assume.
- Large samples (both tests are based on large-sample approximations to the distribution of the chi-square statistics) –OK, 76 events in the smallest group.

**Choice of significance level**:

Wrongly rejecting the null hypothesis here would mean that we claim that there is a difference in time to remission between the two groups, when in fact there is no difference. Depending on any side effects of the treatments, you could argue to use a lower significance level. But if you do, the power will be decreased, which also has to be taken into consideration. 5% is probably a reasonable choice in this case.

**Result:**

*P*-value = 0.0004 (if Log-rank test is chosen) or 0.0019 (if Wilcoxon is chosen).
The *P*-value is smaller than (any) $\alpha$, thus $H_0$ is rejected.

**Conclusion:**

The test suggests that there is a significant difference in time to remission between the two treatment groups in the population of patients with this type of migraine, in favor of the new treatment.

NOTE! This conclusion is only valid if the sample is random.

(38) **Task 3**

**A** (2) Time is measured in months. The ties are probably a result of imprecise measurement (the times could be measured more precise), which means that the **exact** method of handling would be the best choice.

**B** (6) The two variables *gender* and *employment status* have to be recoded to 0/1 variables, or denoted as "class" variables in proc phreg. (In the output, *female* is used instead of *gender*, where 1=female and 0=male. Also, *employment status* has been recoded to 1=employed, 0=unemployed)

The Martingale residual plot suggests that the *age* variable should not be used in its original form. Also, the test of the PH assumption as well as the Score and Arjas plots show that the assumption of proportional hazard doesn't hold for *age*. Thus, some kind of transformation of age should be used. Further investigation is needed to conclude which transformation should be used.

**C** (12) Any chosen model would have to fulfill the **assumptions** for the Cox model:
• Random sample (for inference to be correct). No information in task.
• Non-informative/random censoring. Reasonable to assume (see Task 2 for the treatment covariate).
• Right-censored or left truncated data. Right censoring reasonable to assume, see Task 2.
• Large sample (common rule of thumb: $\geq 10$ events per covariate). Fulfilled, we have a total of 486 events and 4 covariates.
• Proportional hazards (to be checked when building the model)

Check of <u>proportionality (PH) assumption</u>:

<u>*1) include time-dependent covariate in model*</u>
To test the assumption of proportional hazards you can include the time-dependent covariate ln(t)*covariate in the model (if significant, the PH assumption is rejected)

$H_0$: The hazards for different values of covariate *i* are proportional
(all *i*=1 to *p* covariates are to be examined)
$H_a$: The hazards are not proportional

Significance level $\alpha$=5% fine to use (no serious consequences if we claim that the hazards are not proportional when they in fact are)

The test above is rejected for the *age* covariate, but fine (non-significant) for the rest of the covariates.

<u>*2) Score plots/Arjas plots*</u>
You should always make use of a graphical method in addition to the test above. Score plots and Arjas plots have been provided, at least one of them should be used.

The score plots show that the PH assumption looks okay for *trt*, *female*, and *employment status*, but suggests non-proportional hazards for *age*.

The Arjas plots indicate that the PH assumption is violated for all variables (except *age* which is not included).

This concludes either that the PH assumption holds for all variables but *age* (based on the Score plots), or that it is violated for all variables (based on the Arjas plots)**.**

Possible ways of dealing with the non-proportionality:

In this case, since the Martingale plot suggests that *age* should be transformed, that should be done first. That could very well improve the proportionality. If the hazards are still deemed to be non-proportional, any of the standard methods below could be used.

1) Stratify the model on the covariates that do not have proportional hazards. This provides a model where the effect of the covariate in question is accounted for in the baseline hazard, but you don't get any hazard ratio estimates for the covariate.

2) Include time-dependent covariates adjusting for the fact that the hazard ratio is not constant over time.

If stratification doesn't work, and the covariate is not very important, it can also be excluded from the model.

Or, if it is okay to interpret average effects over time, the model can be estimated without any changes (since it has been shown that the estimates represent a geometric average over time).

**D** (10) All covariates but *employment status* are significant at the 5% level. The marginal effects of the covariates are presented below, i.e. the effect of each covariate holding the other covariates constant.

Patients receiving the new treatment have a 27.4% lower remission risk of on average, compared to patients receiving standard care (hazard ratio 0.726, 95% confidence interval 0.563 to 0.922).

Female patients have a 25.9% lower remission risk of on average, compared to patients receiving standard care (hazard ratio 0.741, 95% confidence interval 0.619 to 0.887). The confidence interval is relatively narrow which means that the precision of this estimate is rather high.

The risk of remission decreases with 10.2% on average for every year older the individual is (hazard ratio 0.898, 95% confidence interval 0.884 to 0.910). The confidence interval is quite narrow which means that the precision of this estimate is high.

If a patient is employed the risk of remission is negatively affected. The remission risk is 12% larger on average for employed patients, compared to unemployed

patients (hazard ratio 1.120, 95% confidence interval 0.934 to 1.346). This effect is however not significant, which also can be seen by the inclusion of the value 1 in the confidence interval.

**E** (4) According to the Cox-Snell residuals the model doesn't fit the data extremely well.

If we take into account that the PH assumption doesn't hold for all variables, that makes sense.

Generalized $R^2 = 1 - \exp(-LRT/n)$

where $LRT = -2\log L(0) - [-2\log L(p)] = 3538.501 - 3189.377 = 349.124$

$R^2 = 1 - \exp(-349.124/754) = 1 - \exp(-0.463029) = 0.371$

According to the generalized $R^2$ the covariates are fairly associated with time to remission.

(4) **F** As always, it is a good thing to explain to your employer that you are following the declaration on professional ethics for statisticians.

Among other things it is stated there that you as a statistician should pursue objectivity, and strive to produce results that reflect the observed phenomena in an impartial manner. Not using the best functional form of a covariate despite the implication of the linear form being incorrect would violate this principle.