Exercise 1:

Recall the data file `melanoma.dat` that was analysed previously. In **R**, one can use the `survdiff()` function from the package `survival` to test whether male and female patients have the same survival function.

(a) Create a graph of the Kaplan-Meier estimator stratified by gender along with 95% pointwise confidence bounds using the complementary log-log transformation (`log-log`). What do you conclude from this plot regarding the equality of the gender-specific survivor functions?

(b) Conduct both a log-rank test and a Wilcoxon test to test the null hypothesis that there is no difference in the survivor functions for men and women. Interpret the obtained output!

(c) Create a plot of the estimated Kaplan-Meier estimate of the survivor function for male patients with 95% pointwise confidence bounds using the complementary log-log transformation together with the survivor function of an exponentially distributed random variable $T \sim \mathcal{E}(0.0002)$. Use the log-rank test to test the null hypothesis that the distribution of the survival time of men is exponential with $\lambda = 0.0002$. Hint: Make use of `help(survdiff)` to find out how a one sample test can be performed.

Exercise 2:

An approximation to the log-rank test statistic $W_L = U_L^2/V_L$ that was introduced in the lecture for comparing two survivor curves, and which avoids computing the variance V_L, is as follows:

$$X^2 = \sum_{k=1}^{2} \frac{(O_k - E_k)^2}{E_k} \sim \chi_1^2 \ , \tag{1}$$

where $O_k = \sum_{j=1}^{r} d_{kj}$ and $E_k = \sum_{j=1}^{r} e_{kj}$ denotes, for group k ($k = 1, 2$), the sum of the observed and expected counts over all r distinct failure times across the two groups, respectively. Using the approximation formula (1), carry out a log-rank test for the breast cancer data example on slide 2 of the set of slides "Nonparametric methods for comparing survival distributions". Hint: You may use some of the results given on slide 9. Compare your results with the ones obtained in the lecture using the test statistic W_L.

Exercise 3:

Five hundred and ninety-five persons participate in a case control study of the association of cholesterol and coronary heart disease (CHD). Among them, 300 persons are known to have CHD and 295 are free of CHD. To find out if elevated cholesterol is significantly associated with CHD, the investigator decides to control the effects of smoking. The study subjects are then devided into two strata: smokers and nonsmokers. The following two tables provide the data.

Smokers

	With CHD	Without CHD	Total
Elevated cholesterol			
Yes	120	20	140
No	80	60	140
Total	200	80	280

Nonsmokers

	With CHD	Without CHD	Total
Elevated cholesterol			
Yes	30	60	90
No	70	155	225
Total	100	215	315

Conduct an appropriate test to judge whether elevated cholesterol is significantly associated with CHD after adjusting for the effects of smoking.

Exercise 4:

Consider two groups with sizes $n_1 = n_2 = 100$ from a population of size $n = n_1 + n_2$. Suppose that the true survival times of the first group are distributed as $T_1 \sim \mathcal{WB}(3, 0.928)$ and that the second group has true survival times T_2 with hazard rate $h_2(t) = 3t^2$. For both groups we assume that censoring times are independent and identically distributed. To begin with, execute the command `set.seed(1234)`.

(a) Generate right censored survival times for both groups separately. Assume that the censoring times in both groups are exponentially distributed with parameter $\lambda = 2/3$. Use the inverse transform sampling method (exercise 5, study sheet 1) to generate true survival times of group 2. Combine your data (`time` = observed survival times, `delta` = censoring indicator, `group` = group membership) into a data frame.

(b) Use a two-sample test of your choice to test whether the survivor functions of both groups are identical. Give an interpretation of the test result ($\alpha = 0.05$).

(c) Use a one-sample test of your choice to test whether the distribution of the survival times T of the whole population are distributed as $T \sim \mathcal{WB}(3, 0.928)$. Give an interpretation of the test result ($\alpha = 0.05$).

(d) Assume that the theoretical event times of the whole population are distributed as $T \sim \mathcal{WB}(3, 1)$. Compute the Kaplan-Meier estimator, the theoretical survivor function and the empirical survivor function for the whole population and visualise them in a single plot. How do the results obtained in (c) change if the empirical survivor function is used as an offset?

(e) For group 2, compute the mean and median lifetime and the probability to survive longer than the mean and median lifetime. Carry out calculations for both the theoretical data and the censored data.