This second part of the examination consists of 3 exercises with a number of sub-questions. Please mark which exercise and sub-question you are answering. Motivate your answers! (By considerations, calculations, sketches, etc.)

A group of researchers wanted to predict the death from ventilator associated pneumonia in patients admitted to an intensive care unit. The authors measures the blood concentration of proANP on day 0 (D0) and day 4 (D4), and recorded the age and gender of the patients. They used the data to predict survival, in the group of 71 patients in which 26 died.

1) The percent of males and females distributed between survivors and non-survivors are shown below:

Gender (%)	survivors (n = 45)	non-survivors (n = 26)	Total (n = 71)
Male	66.7	46.2	59.2
Female	33.3	53.8	40.8
Total	100	100	100

    a. Use the χ^2-test to test for association between gender and survival. Is there a significant association?

First we need to convert the percentages to original numbers.

Gender (%)	survivors (n = 45)	non-survivors (n = 26)	Total (n = 71)
Male	30	12	42
Female	15	14	29
Total	45	26	71

Then the expected numbers are calculated:

Gender (%)	survivors (n = 45)	non-survivors (n = 26)	Total (n = 71)
Male	26.6	15.4	42
Female	18.4	10.6	29
Total	45	26	71

Then the normalized differences between the observed and expected values are calculated and summed:

Gender (%)	survivors (n = 45)	non-survivors (n = 26)	Total (n = 71)
Male	0.429242	0.742919	1.172161
Female	0.621661	1.075952	1.697613
Total	1.050903	1.818871	2.869774

Therefore $\chi^2 = 2.87$. From the table in the appendix the p value is between 0.10 and 0.05. This is usually not considered statistically significant.

    b. Is the χ^2-test valid in this case?

Yes. All expected values are above 5.

c. What is the odds ration of survival for genders?

$$or = \frac{ad}{bc} = \frac{30*14}{12*15} = 2.33$$

d. What is the 95% confidence interval of the odds ratio?

First the standard error of the natural logarithm transformed odds ratio is calculated:
$$SE\left(\ln(or)\right) = \sqrt{\tfrac{1}{a} + \tfrac{1}{b} + \tfrac{1}{c} + \tfrac{1}{d}} = \sqrt{\tfrac{1}{30} + \tfrac{1}{12} + \tfrac{1}{15} + \tfrac{1}{14}} = 0.5$$
Then the odds ratio is log transformed
$\ln(or) = 0.847$
The 95% CI of the log transformed odds ratio is then
$95\%CI = 0.847 - 1.96*0.50 \text{ to } 0.847 + 1.96*0.50 = -0.142 \text{ to } 1.837$
The 95% CI of the odds ratio is then:
$95\%CI = e^{-0.142} \text{ to } e^{1.837} = 0.87 \text{ to } 6.28$

e. Is the odds ratio in agreement with the χ^2-test and does it provide the same information.

The odds ratio is in agreement with the χ^2-test because 1 is included in the 95% confidence interval we did not find a significant difference. If we believe that there is a difference between genders the odds ratio give an indication of the strength of the association between gender and survival. That is that the odds for males to survive is 2.33 times as high as the odds for males to survive.

2) The authors measures the blood concentration of proANP on day 0 (D0) and day 4 (D4). They constructed a ROC curve for each day for the proANP measurement with respect to mortality.
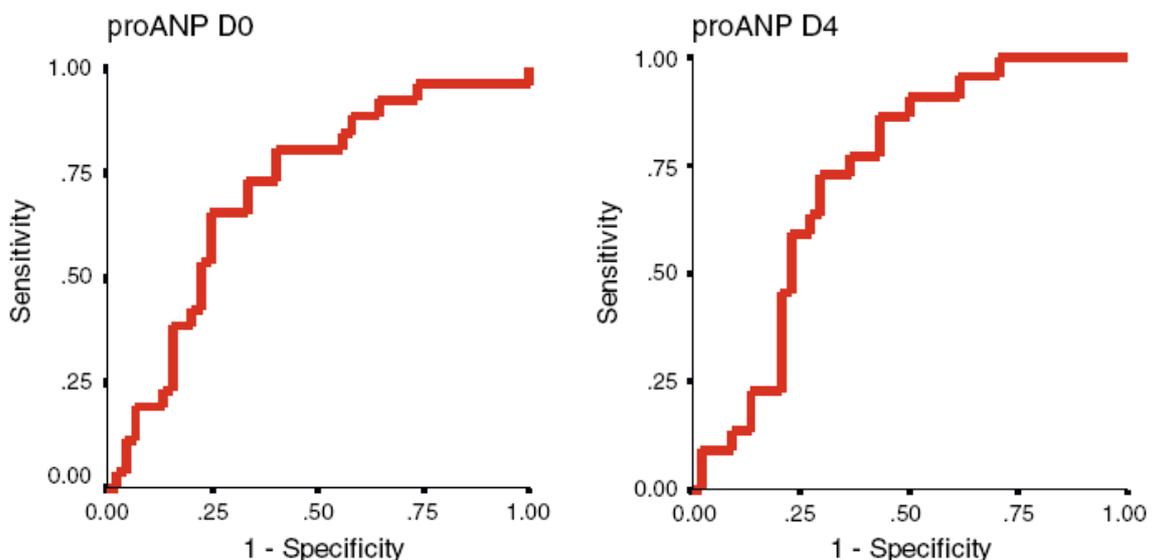
a. What does sensitivity and specificity mean?

Sensitivity is the ability to detect patients <u>with</u> the condition; this is the true positive ratio. Specificity is the ability to detect patients <u>without</u> the condition; this is one minus the false positive rate.

      b. The authors write that the area under the ROC curve for day 0 is 0.71 (SD 0.06; P = 0.004) and 0.73 (SD 0.06; P = 0.004). What does this tell us?

The area under the curve is a measure of accuracy. This tells us how good measuring ProANP may be in detecting the survival of these patients. The low p values indicate that the area indet the ROC curve is different from 0.5 which means that using ProANP as a predictor is better than guessing. Adding and subtracting 1.96 * SD will indicate if the 95% CI's of the area under the curves is overlapping, but we are lacking decimals on AUC measure to conclude anything.

3) The authors constructed a logistic regression model to predict non-survivors in the patient group.
      a. Which parameter is the best single predictor for survival?

The level of proANP at day 4 has the highest odds ratio and is therefore the best predicting factor.

      b. Why is to odds ratio calculated from the logistic regression different from you calculated by the χ^2-test?

The two tests differ because the logistic regression takes more factors into consideration. However, in this case the conclusion is the same for the univariate χ^2-test and the multivariate logistic regression.

      c. Will it be enough to measure the level of proANP at day 0 in order to predict death by this model? (why? / why not?)

No. There is a significant impact from measurement of proANP at day 0 and day4. Both measurements add to the model significantly. The interaction terms were not reported. They may have provided more information.

      d. Calculate the regression coefficients for the logistic regression model.

The regression coefficients can be calculated by taking the logarithm to the odds ratios:

Parameter	β
Age	0
Gender, female	0.59
ln(proANP) D0	0.85
ln(proANP) D4	1.32

      e. What is the probability that a 39 year old male patent in the study will die, if his proANP concentration is 600 pmol/L and 550 pmol/L at day 0 and 4, respectively. Assume that the regression constant, β_0, is -13.0.

z = 39*0+0*0.59+ln(600)*0.85+ln(550)*1.32-13.8 = 0.766

$p = 1/(1+\exp(-z)) = 0.68$

Parameter	odds ratio (95 % CI)	P
Age	1.00 (0.96 - 1.04)	0.89
Gender, female	1.80 (0.62 - 5.26)	0.28
ln(proANP) D0	2.35 (1.05 - 5.26)	0.04
ln(proANP) D4	3.76 (1.39 - 10.18)	0.01

Tabel 1.

Appendix

Table A–5. Percentage points or critical values for the χ^2 distribution corresponding to commonly used areas under the curve.

	Area in Upper Tail			
Degrees of Freedom	**0.10**	**0.05**	**0.01**	**0.001**
1	2.706	3.841	6.635	10.828
2	4.605	5.991	9.210	13.816
3	6.251	7.815	11.345	16.266
4	7.779	9.488	13.277	18.467
5	9.236	11.071	15.086	20.515
6	10.645	12.592	16.812	22.458
7	12.017	14.067	18.475	24.322
8	13.362	15.507	20.090	26.125
9	14.684	16.919	21.666	27.877
10	15.987	18.307	23.209	29.588
11	17.275	19.675	24.725	31.264
12	18.549	21.026	26.217	32.909
13	19.812	22.362	27.688	34.528
14	21.064	23.685	29.141	36.123
15	22.307	24.996	30.578	37.697
16	23.542	26.296	32.000	39.252
17	24.769	27.587	33.409	40.790
18	25.989	28.869	34.805	42.312
19	27.204	30.144	36.191	43.820
20	28.412	31.410	37.566	45.315
21	29.615	32.671	38.932	46.797
22	30.813	33.924	40.289	48.268
23	32.007	35.173	41.638	49.728
24	33.196	36.415	42.980	51.179
25	34.382	37.653	44.314	52.620
26	35.563	38.885	45.642	54.052
27	36.741	40.113	46.963	55.476
28	37.916	41.337	48.278	56.892
29	39.088	42.557	49.588	58.302
30	40.256	43.773	50.892	59.703
40	51.805	55.759	63.691	73.402
50	63.167	67.505	76.154	86.661
60	74.397	79.082	88.379	99.607
70	85.527	90.531	100.425	112.317
80	96.578	101.879	112.329	124.839
90	107.565	113.145	124.116	137.208
100	118.498	124.342	135.807	149.449