Exercise 1

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.

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

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 χ^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.

Is the χ^2-test valid in this case?

Yes. All expected values are above 5.

What is the odds ration of survival for genders?

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

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

First the standard error of the natural logarithm transformed odds ratio is calculated:

$$\text{SE}\left(\ln\left(or\right)\right) = \sqrt{\tfrac{1}{a}+\tfrac{1}{b}+\tfrac{1}{c}+\tfrac{1}{d}} = \sqrt{\tfrac{1}{30}+\tfrac{1}{12}+\tfrac{1}{15}+\tfrac{1}{14}} = 0.5$$

Then the odds ratio is log transformed

$\ln(or) = 0.847$

The 95% CI of the log transformed odds ratio is then

$95\%\text{CI} = 0.847 - 1.96 * 0.50 \text{ to } 0.847 + 1.96 * 0.50 = -0.142 \text{ to } 1.837$

The 95% CI of the odds ratio is then:

$95\%\text{CI} = e^{-0.142} \text{ to } e^{1.837} = 0.87 \text{ to } 6.28$

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.

Exercise 2

Kan forskerne bruge Chi-kvadrat testen?

Det kan de ikke fordi 2 af 4 forventede værdier er under 5:

	Treatment A	Treatment B	
2 arms	7	7	14
3 arms	3	3	6
	10	10	20

Hvilken test bør de bruge?

De skulle have brugt Fisher's exact test

Er der en signifikant forskel på antallet af patienter med bivirkninger i de to behandlingsgrupper?

For et beregne Fisher's exact test skal man beregne sandsynligheden for at finde netop de observerede værdier fra Tabel 2:

	Treatment A	Treatment B	Total
2 arms	5	9	14
3 arms	5	1	6
Total	10	10	20

$$p = \frac{14!\,6!\,10!\,10!}{20!\,5!\,5!\,9!\,1!} = 0{,}065$$

Ligeledes skal sandsynligheden for at finde mere ekstreme observationer beregnes. I dette tilfælde er det bare en mere ekstrem situation, som give de samme totaler:

	Treatment A	Treatment B	Total
2 arms	4	10	14
3 arms	6	0	6
Total	10	10	20

$$p = \frac{14!\,6!\,10!\,10!}{20!\,4!\,10!\,6!\,0!} = 0{,}0054$$

Sandsynligheden for at forskerne har opnået deres observationer ved et tilfælde er altså p=0,07.

Solutions for contingency exercises

Hvad er odds ratio mellem de to behandlingsgrupper for at gro en tredje arm?

$$or = \frac{5 \cdot 9}{5 \cdot 1} = 9$$

Hvad er konfidensintervallet for odds ratioen?

$$\ln(or) = 2{,}20$$

$$SE(\ln(or)) = \sqrt{\frac{1}{5} + \frac{1}{9} + \frac{1}{5} + \frac{1}{1}} = 1{,}23$$

$$2{,}20 - 1{,}96 \cdot 1{,}23 \; to \; 2{,}20 + 1{,}96 \cdot 1{,}23 = -0{,}211 \; to \; 4{,}61$$

The confidenceinterval is thus

$$e^{-0{,}211} \; to \; e^{4{,}61} \Rightarrow 0{,}81 \; to \; 100$$

Kan odds ratioen vise en statistisk signifikant forskel mellem de to behandlinger?

Da konfidensintervallet overlapper 1 kan men heller ikke vha odds ratio vise en signifikant forskel på bivirkninger mellem de to behandlingsgrupper.

Opgave 3

Først skal man lave apgarscore om til en variabel som indikerer om scoren er over 7. Derefter laves en krydstabel:

smoking * apgarOver7 Crosstabulation

			apgarOver7		
			,00	1,00	Total
smoking	no	Count	10	10	20
		Expected Count	12,0	8,0	20,0
	yes	Count	8	2	10
		Expected Count	6,0	4,0	10,0
Total		Count	18	12	30
		Expected Count	18,0	12,0	30,0

Man kan se at en af celllerne er under 4, så man bør lave en Fishers exact test i stedet for en Chi-kvadrat test. Her er resultatet for dem begge:

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	2,500[a]	1	,114	,235	,117	
Continuity Correction[b]	1,406	1	,236			
Likelihood Ratio	2,647	1	,104	,141	,117	
Fisher's Exact Test				,235	,117	
Linear-by-Linear Association	2,417[c]	1	,120	,235	,117	,096
N of Valid Cases	30					

a. 1 cells (25,0%) have expected count less than 5. The minimum expected count is 4,00.

b. Computed only for a 2x2 table

c. The standardized statistic is -1,555.

Vi har altså ikke kunnet vise at der er en sammenhæng mellem rygning og Apgar scoren, men det er IKKE det samme som at pastå at en sådan sammenhæng ikke er til stede.