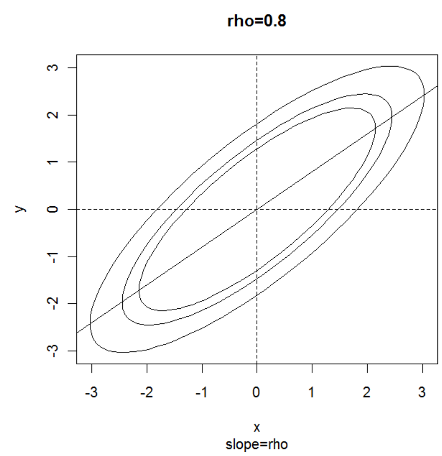
WS11/12
Multivariate Analysis
K. Utikal

1) isoklinen einer bivariaten normalVerteilung

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\left(1-\rho^2\right)}\left(\frac{\left(x-\mu_1\right)^2}{\sigma_1^2} - 2\rho\frac{\left(x-\mu_1\right)}{\sigma_1}\frac{\left(y-\mu_2\right)}{\sigma_2} + \frac{\left(y-\mu_2\right)^2}{\sigma_2^2}\right)\right\}$$

set $\mu_1 = \mu_2 = 0,\ \sigma_1^2 = \sigma_2^2 = 1 \Longrightarrow$isoclines $x^2 - 2\rho xy + y^2 = c$,



plot by `ellipse(ellipse)`(r-package)

2) für die **raucherDaten**

2a) erklären sie wie die marginale verteilung von überlebenStatus und raucherStatus erhalten wurde. was vermuten sie nachdem sie diese daten gesehen haben?

solution: X : smoker, Y : survival, Z : age,

2b) teste auf **marginale unabhängigkeit** von überlebenStatus und raucherStatus mit dem LR-test likelihood function (multinomial for fixed n): $L(\mathbf{p}|\ \mathbf{n}) = \prod_{ij} p_{ij}^{n_{ij}}$

distribution:

p_{11}	p_{12}	$p_{1\cdot}$
p_{21}	p_{22}	$p_{2\cdot}$
$p_{\cdot 2}$	$p_{\cdot 2}$	1

$\Longrightarrow$ sample:

n_{11}	n_{12}	$n_{1\cdot}$
n_{21}	n_{22}	$n_{2\cdot}$
$n_{\cdot 2}$	$n_{\cdot 2}$	$n_{\cdot\cdot}$

$\Longrightarrow$likelihood: $L(\mathbf{p}|\ \mathbf{n}) = \prod_{ij} p_{ij}^{n_{ij}} \Longrightarrow \widehat{p}_{ij} = \frac{n_{ij}}{n_{\cdot\cdot}}$

Null: $H_0: \ p_{ij} = p_{i\cdot}p_{\cdot j} \Longrightarrow L(\mathbf{p}|\ \mathbf{n}) = \prod_{ij} (p_{i\cdot}p_{\cdot j})^{n_{ij}} \Longrightarrow \widehat{p}_{i\cdot} = \frac{n_{i\cdot}}{n_{\cdot\cdot}},\ \widehat{p}_{\cdot j} = \frac{n_{\cdot j}}{n_{\cdot\cdot}}$

LR-test: $-2\sum n_{ij}(\log \widehat{p}_{i\cdot}\widehat{p}_{\cdot j} - \log \widehat{p}_{ij})$

likelihood function (binomial for fixed row marginals $n_{1\cdot},\ n_{2\cdot}$):

p_1	$1-p_1$	1
p_2	$1-p_2$	1

$\Longrightarrow$

n_{11}	$n_{1\cdot} - n_{11}$	$n_{1\cdot}$
n_{21}	$n_{2\cdot} - n_{21}$	$n_{2\cdot}$

$\Longrightarrow L(\mathbf{p}|\ \mathbf{n}) = \prod_i p_i^{n_{i1}}(1-p_i)^{n_{i\cdot}-n_{i1}} \Longrightarrow \widehat{p}_{ij} = \frac{n_{ij}}{n_{i\cdot}}$

Null: $H_0: \ p_i = p$ for some $p \Longrightarrow L(\mathbf{p}|\ \mathbf{n}) = p^{n_{\cdot 1}}(1-p)^{n_{\cdot\cdot}-n_{\cdot 1}} \Longrightarrow \widehat{p} = \frac{n_{1\cdot}}{n_{\cdot\cdot}}$

case: fixed row marginals + fixed column marginals$\Longrightarrow$Fishers **exact** (hypergeometric)-test

2c) was schliessen sie aus der verteilung von altersGruppe und raucherStatus?

2d) schreiben sie ein multinomialModell für alle die variablen raucherStatus, überlebenStatus, altersGruppe

multinomial

$$P(n_{111}, ..., n_{IJK}|\ n_{...}) = \binom{n_{...}}{n_{111}, ..., n_{IJK}} p_{111}^{n_{111}} \cdots p_{IJK}^{n_{IJK}}$$

with $p_{ijk} = P(X = i,\ Y = j,\ Z = k)$

2e) schreiben sie ein multinomialModell für die variablen von raucherStatus und überlebenStatus

2f) konstruieren sie einen LR test auf *bedingte unabhängigkeit* von raucherStatus und überlebenStatus bedingt auf altersGruppe

likelihood

$$L(\mathbf{p}|\ \mathbf{n}) = p_{111}^{n_{111}} \cdots p_{IJK}^{n_{IJK}}$$

with $p_{ijk} = P(X = i,\ Y = j,\ Z = k)$

$H_0:\ p_{ij/k} = p_{i\cdot/k}p_{\cdot j/k}$

$$L(\mathbf{p}|\ \mathbf{n}) = (p_{11/1}p_{\cdot\cdot 1})^{n_{111}} \cdots (p_{IJ/K}p_{\cdot\cdot K})^{n_{IJK}}$$

note that $p_{\cdot\cdot 1}, ..., p_{\cdot\cdot K}$ have no information on H_0 and can be ignored.
use

$$L(\mathbf{p}|\ \mathbf{n}) = \left(p_{11/1}{}^{n_{111}} ... p_{IJ/1}{}^{n_{IJ1}}\right) \cdots \left(p_{11/K}{}^{n_{11K}} .... p_{IJ/K}{}^{n_{IJK}}\right)$$

There are K decoupled $I \times J$ tables now. Use test of marginal independence (LR-test cond on $n_{\cdot\cdot k}$)

LR-test$= -2\sum_{ijk} n_{ijk}(\log \widehat{p}_{i\cdot/k}\widehat{p}_{\cdot j/k} - \log \widehat{p}_{ij/k})$ where $\widehat{p}_{i\cdot/k} = \frac{n_{i\cdot k}}{n_{\cdot\cdot k}},\ \widehat{p}_{\cdot j/k} = \frac{n_{\cdot jk}}{n_{\cdot\cdot k}},\ \widehat{p}_{ij/k} = \frac{n_{ijk}}{n_{\cdot\cdot k}}$

alternatively, cond on row totals or on row and column totals / each table

```
femsmoke(faraway) R Documentation
Mortality due to smoking according age group in women
Description
In 1972-74, a survey of one in six residents of Whickham, near Newcastle,
England was made.  Twenty years later, this data recorded in a follow-up
study.  Only women who are current smokers or who have never smoked
are included.

> xtabs(y~smoker+dead,data=femsmoke)
 dead
smoker yes no
   yes 139 443           139/(139+443)=0.24
    no 230 502           230/(230+502)=0.31

> xtabs(y~smoker+dead+age,data=femsmoke)
, , age = 18-24
 dead
smoker yes no
   yes 2 53               2/(2+53)=0.05
    no 1 61               1/(1+61)=0.02
```

```
, , age = 25-34
 dead
smoker yes no
   yes 3 121           3/(3+121)=0.02
    no 5 152           5/(5+152)=0.03
, , age = 35-44
 dead
smoker yes no
   yes 14 95           14/(14+95)=0.13
    no 7 114           7/(7+114)=0.06
, , age = 45-54
 dead
smoker yes no
   yes 27 103           27/(27+103)=0.21
    no 12  66           12/(12+66) =0.15
, , age = 55-64
 dead
smoker yes no
   yes 51 64           51/(51+64)=0.44
    no 40 81           40/(40+81)=0.33
, , age = 65-74
 dead
smoker yes no
 yes  29 7            29/(29 + 7)=0.80
  no 101 28          101/(101+28)=0.76
, , age = 75+
 dead
smoker yes no
 yes 13 0
  no 64 0

> xtabs(y~smoker+age,data=femsmoke)
 age
smoker 18-24 25-34 35-44 45-54 55-64 65-74 75+
   yes 55    124   109  130  115   36   13
    no 62    157   121   78    121   129   64
```