

Iso-Linien (Contours) und kritische Werte

$$\{\mathbf{x} : (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = c^2\}$$

Die Konturen mit konstanter Wahrscheinlichkeit c einer p -dim. Normalverteilung sind Ellipsen definiert durch die Menge aller \mathbf{x} für die gilt

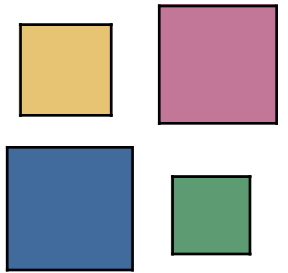
$$(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = c^2$$

Diese Ellipsen sind zentriert um $\boldsymbol{\mu}$ und haben die Achsen $c\sqrt{\lambda_i}\mathbf{e}_i$, wobei λ_i der i -te Eigenwert und \mathbf{e}_i der i -te Eigenvektor ist.

Eine (gefüllte) Ellipse aus den Werten \mathbf{x} die

$$(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \leq \chi_p^2(\alpha)$$

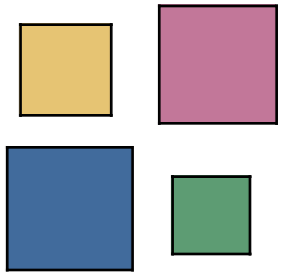
erfüllen, hat Wahrscheinlichkeit $1 - \alpha$.



Eigenschaften der multivariaten NV

- Linear Kombinationen von Komponenten von \mathbf{X} sind wiederum normalverteilt.
- Jede Teilmenge von Komponenten von \mathbf{X} ist wiederum normalverteilt.
- Eine Kovarianz von 0 impliziert, dass die jeweiligen Komponenten unabhängig sind.
- Die bedingte Verteilung von Komponenten von \mathbf{X} ist auch multivariat normalverteilt.
- Die “normierte” multiv. Normalverteilung ist χ^2 verteilt mit p Freiheitsgraden, d.h.

$$(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \sim \chi_p^2$$

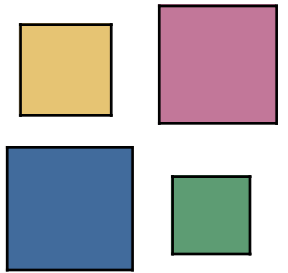


Parameterschätzung

- Der Mittelwert μ wird durch das Stichproben Mittel $\bar{\mathbf{X}}$ geschätzt, die Varianz-Covarianz Σ durch die Stichproben Varianz-Covarianz Matrix \mathbf{S} .
- Die Verteilung von $\bar{\mathbf{X}}$ ist multivariat normal $N_p(\mu, \frac{1}{n}\Sigma)$.
 $(n-1)\mathbf{S}$ besitzt eine Wishartverteilung mit $(n-1)$ Freiheitsgraden.

$$\left(W_m(\cdot|\Sigma) = \sum_{j=1}^m \mathbf{z}_j \mathbf{z}_j' \text{ mit } \mathbf{z}_j \sim N_p(0, \Sigma) \right)$$

- $\bar{\mathbf{X}}$ und \mathbf{S} sind unabhängig.



MNV Stichproben

Für $\mathbf{X} \sim N(0, I_p)$ gilt: $X_i \sim N(0, 1)$ und $\sigma_{ij} = 0$ für $1 \leq i \neq j \leq p$

Damit kann eine $N(0, I_p)$ Verteilung aus p univariaten Standardnormalverteilungen erzeugt werden.

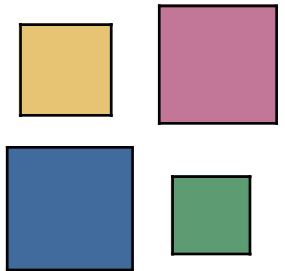
Über die Spektralzerlegung

$$\Sigma = \sum_{i=1}^p \lambda_i e_i e_i' = \mathbf{P} \Lambda \mathbf{P}'$$

der Varianz Covarianz Matrix Σ kann mittels μ und Σ eine $Y \sim N(0, I_p)$ in eine $W \sim N(\mu, \Sigma_p)$ überführt werden.

Für $W = \mathbf{P} \Lambda^{1/2} Y + \mu$ gilt:

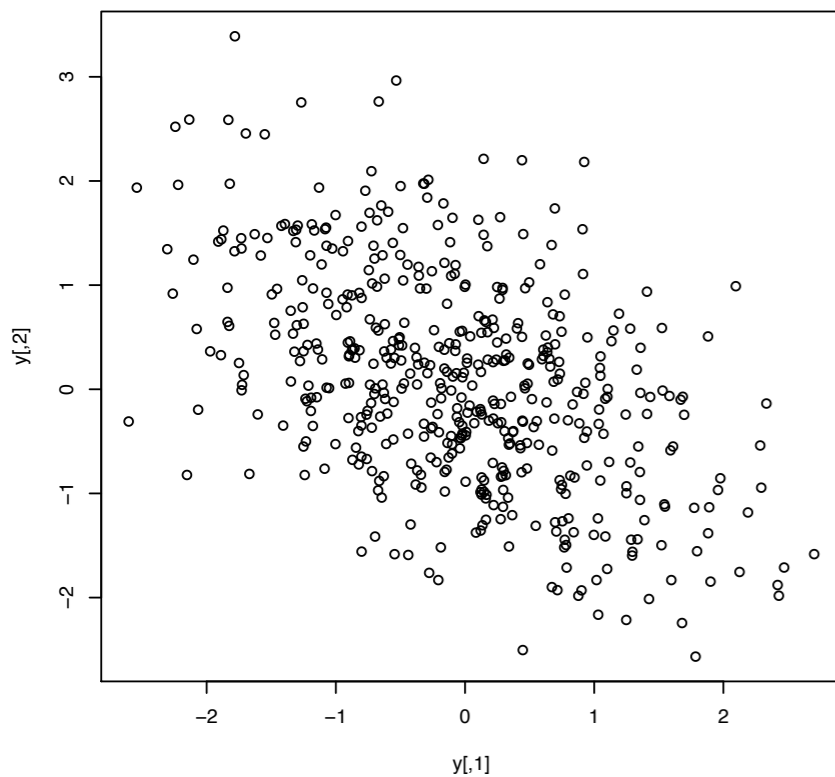
$$W \sim N(\mathbf{P} \Lambda^{1/2} 0 + \mu, \mathbf{P} \Lambda^{1/2} I_p \Lambda^{1/2} \mathbf{P}') = N(\mu, \Sigma_p)$$



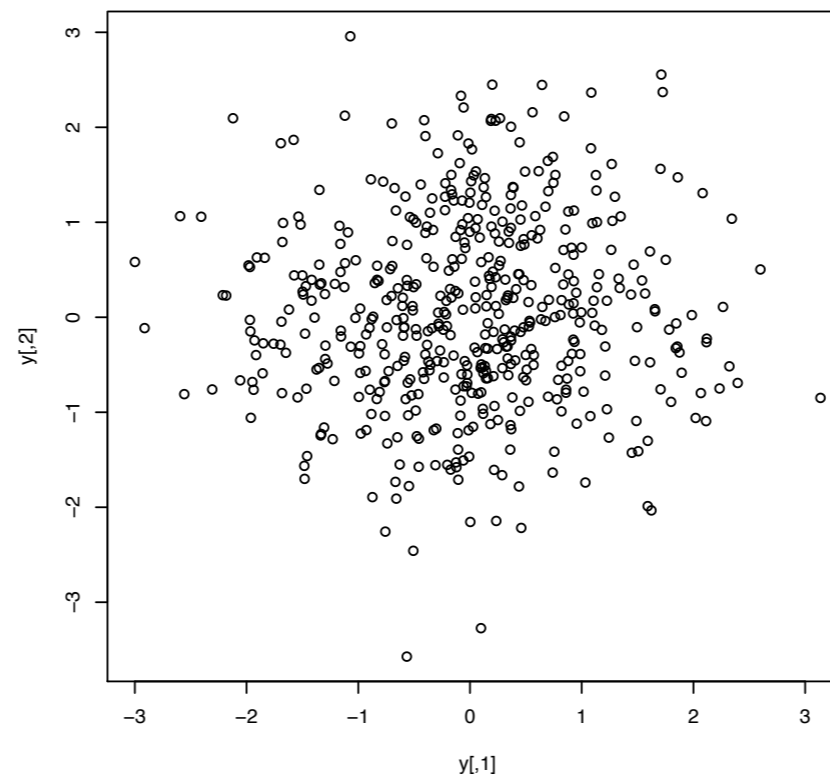
Beispiele in R

```
x<-f.gen.std.norm(n=500,p=2)
vc<-matrix(c(1,-0.5,-0.5,1),ncol=2,byrow=T)
evc<-eigen(vc)
vc2<-(evc$vector)%*%diag(sqrt(evc$value))%*%t(evc$vector)
y<-x%*%vc2
plot(y)
```

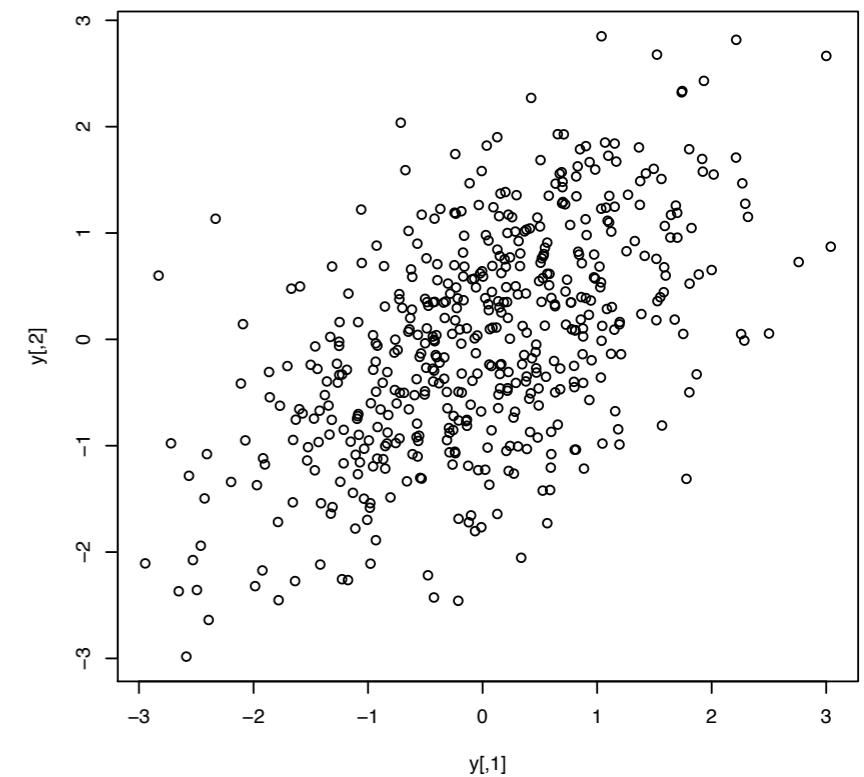
$c(1, -0.5, -0.5, 1)$

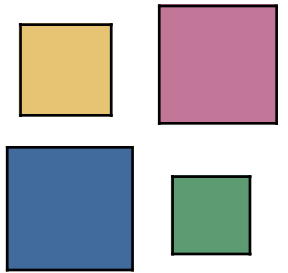


$c(1, 0, 0, 1)$



$c(1, 0.5, 0.5, 1)$





MNV: Mittelwert Tests

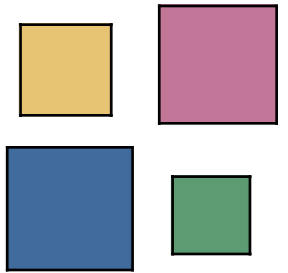
Fragestellung:

Enstammt ein vorgegebener Wert für einen Mittelwert dem Mittelwert der vorliegenden Stichprobe.

In Sprache der Hypothesentests:

$$H_o : \mu = \mu_0 \quad vs \quad H_A : \mu \neq \mu_0$$

wo $\mu = (\mu_1, \mu_2, \dots, \mu_p)$ der wahre Stichprobenmittelwert und $\mu_0 = (\mu_{01}, \mu_{02}, \dots, \mu_{0p})$ der zu testende Mittelwert ist.



Univariater Fall

Für

$$H_o : \mu = \mu_0 \quad vs \quad H_A : \mu \neq \mu_0$$

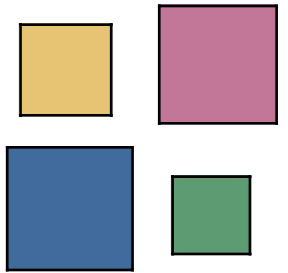
berechnen wir die t -Statistik, wie folgt:

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}.$$

Wir vergleichen $|t|$ mit den Werten der t -Verteilung mit $(n - 1)$ Freiheitsgraden, und verwerfen H_o genau dann wenn $|t|$ größer als der kritischer Wert ist.

Analog können wir das $100(1 - \alpha)\%$ Konfidenz Intervall für das Stichprobenmittel μ berechnen mit:

$$\bar{X} \pm t_{n-1}(\alpha/2) \frac{s}{\sqrt{n}}$$



Hotellings T^2

Das multivariate Analogon der t -Statistik ist:

$$T^2 = n(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)' \mathbf{S}_{n-1}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}_0)$$

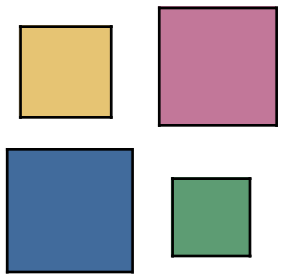
Diese Statistik wird *Hotellings T^2* genannt.

T^2 genügt einer $\frac{(n-1)p}{n-p} \mathcal{F}_{p, n-p}$ Verteilung.

(Zur Erinnerung:

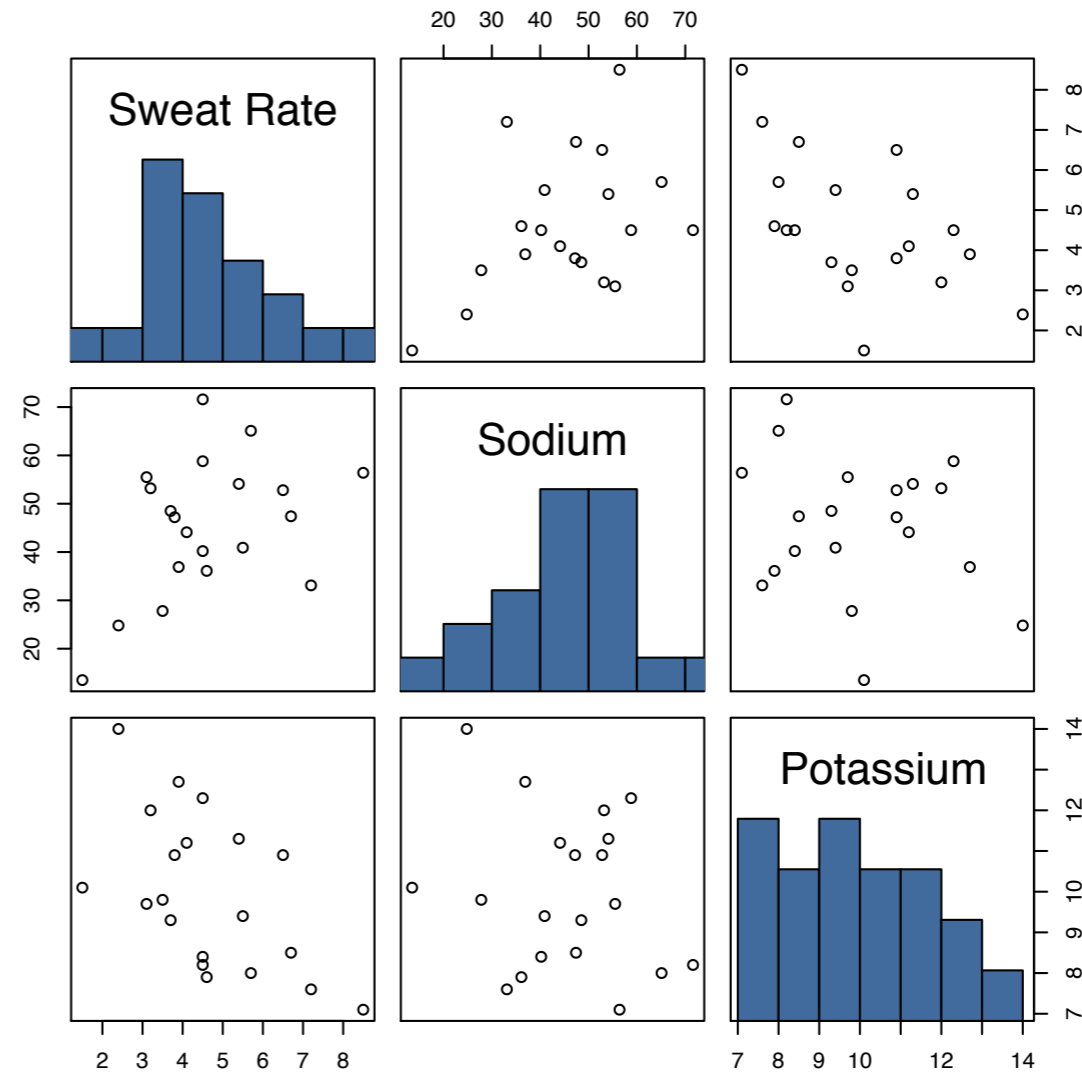
Im univariaten Fall gilt: wenn $t \sim t_{n-1}$, dann ist $t^2 \sim \mathcal{F}_{1, n-1}$.

Im multivariaten Fall fließt die Dimensionalität in die Freiheitsgrade der Verteilung ein.)



Beispiel: Transpiration

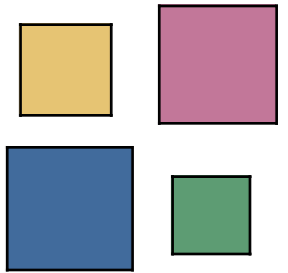
- 20 Beobachtungen:
 - Transpirations Rate
 - Natrium (Sodium, Na)
 - Kalium (Potassium, K)



Zu testen:

$$H_o : \mu = (4 \ 50 \ 10)' \quad vs \quad H_A : \mu \neq (4 \ 50 \ 10)'$$

$$\bar{\mathbf{X}} = \begin{bmatrix} 4.61 \\ 45.6 \\ 9.97 \end{bmatrix} \quad \mathbf{S}_{n-1} = \begin{bmatrix} 2.75 & 9.37 & -1.72 \\ 9.37 & 190.8 & -5.35 \\ -1.72 & -5.35 & 3.45 \end{bmatrix}$$



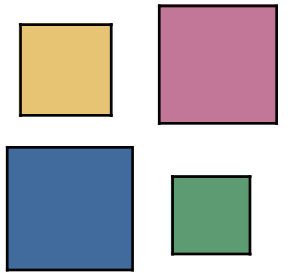
Transpiration (cont.)

$$\mathbf{S}_{n-1}^{-1} = \begin{bmatrix} 0.606 & -0.0222 & 0.268 \\ -0.0222 & 0.00630 & -0.00132 \\ 0.268 & -0.00132 & 0.422 \end{bmatrix}$$

$$\begin{aligned} T^2 &= 20 \begin{bmatrix} 4.61 - 4 \\ 45.6 - 50 \\ 9.97 - 10 \end{bmatrix}' \begin{bmatrix} 0.606 & -0.0222 & 0.268 \\ -0.0222 & 0.00630 & -0.00132 \\ 0.268 & -0.00132 & 0.422 \end{bmatrix} \begin{bmatrix} 4.61 - 4 \\ 45.6 - 50 \\ 9.97 - 10 \end{bmatrix} \\ &= 9.74 \end{aligned}$$

Für $\alpha = 0.10$, $\frac{(n-1)p}{n-p} \mathcal{F}_{p, n-p}(0.10) = \frac{19 \times 3}{17} \times 2.44 = 8.18$ würden wir H_0 auf einem 10% Signifikanz Niveau verwerfen.

Für $\alpha = 0.05$, $\frac{(n-1)p}{n-p} \mathcal{F}_{p, n-p}(0.05) = \frac{19 \times 3}{17} \times 3.2 = 10.7$. würden wir H_0 auf einem 5% Signifikanz Niveau NICHT verwerfen.



Konfidenz Regionen

Eine $100(1 - \alpha)\%$ Konfidenz Region für den Mittelwert einer p -dim. Normalverteilung ist die Ellipse definiert durch alle μ die

$$n(\bar{\mathbf{X}} - \boldsymbol{\mu})' \mathbf{S}_{n-1}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}) \leq \frac{p(n-1)}{n-p} \mathcal{F}_{p, n-p}(\alpha)$$

erfüllen, wobei $\mathbf{X}_1, \dots, \mathbf{X}_n$ Beobachtungen einer $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ verteilten Zufallsgröße sind.

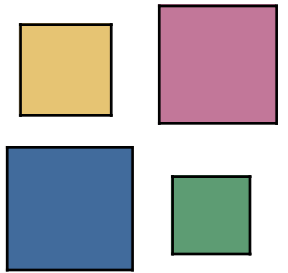
Die Form der Konfidenz Ellipse ist definiert durch \mathbf{S} .

Die Ausrichtung bestimmen die Eigenvektoren von \mathbf{S} ,

die Länge der Hauptachsen der Ellipse wird bestimmt durch

$$\sqrt{\lambda_i} \times \sqrt{\frac{p(n-1)}{n(n-p)} \mathcal{F}_{p, n-p}(\alpha)}$$

wobei λ_i der i -te Eigenwert von \mathbf{S} ist.



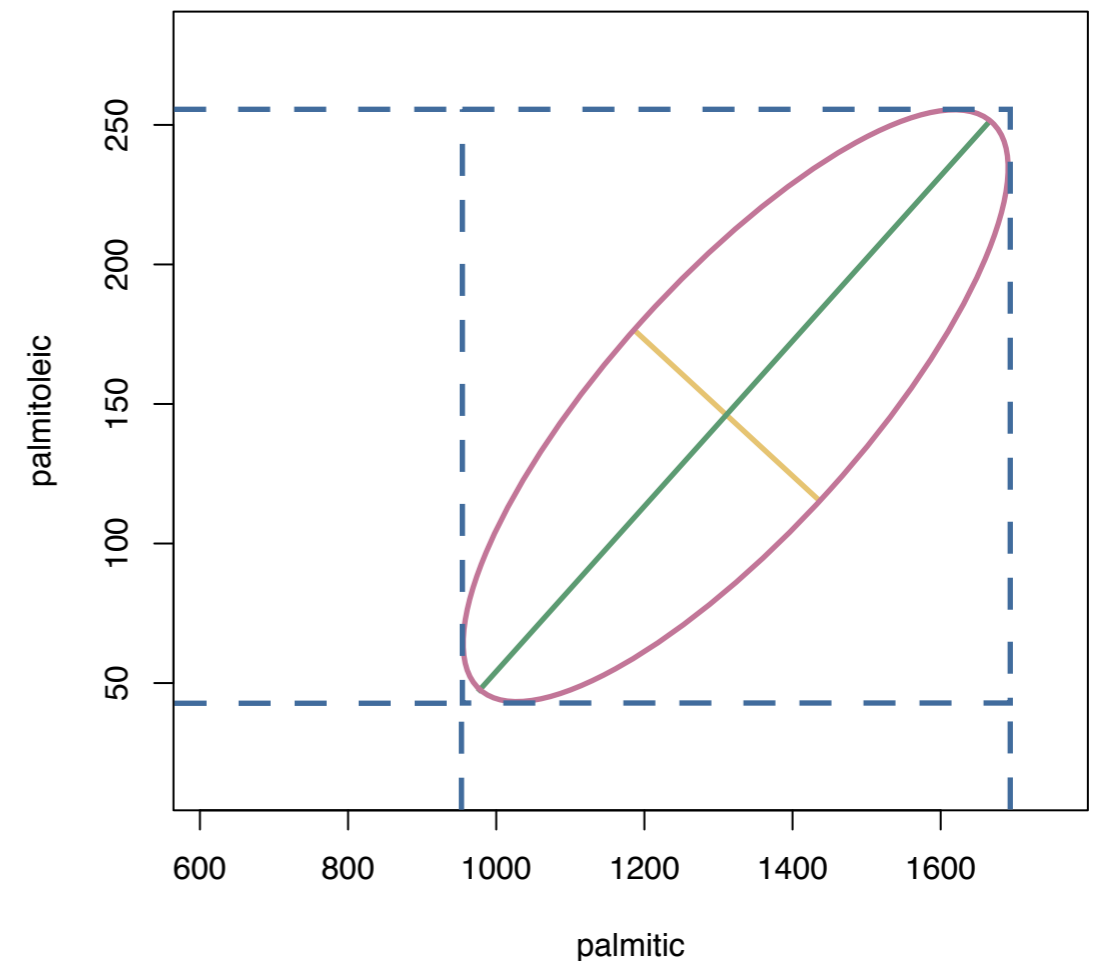
Simultane Konfidenzintervalle

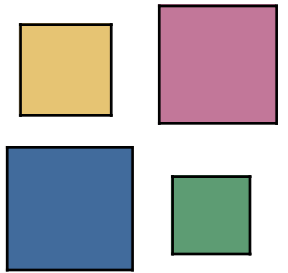
Als simultane Konfidenzintervalle erhalten wir:

$$\bar{x}_i - \sqrt{\frac{p(n-1)}{(n-p)} \mathcal{F}_{p,n-p}(\alpha)} \sqrt{\frac{s_{ii}}{n}} \leq \mu_i \leq \bar{x}_i + \sqrt{\frac{p(n-1)}{(n-p)} \mathcal{F}_{p,n-p}(\alpha)} \sqrt{\frac{s_{ii}}{n}}$$

Geometrische Interpretation

Die Intervalle ergeben sich aus den Projektionen der Ellipse auf die Achsen:





Bonferroni Korrektur

Sei C_i eine Konfidenzaussage über die Linearkombination $\mathbf{a}'_i \boldsymbol{\mu}$ mit $i = 1, \dots, m$ und $P(C_i \text{ wahr}) = 1 - \alpha$, so gilt:

$$\begin{aligned} P(\text{alle } C_i \text{ wahr}) &= 1 - P(\text{mindestens ein } C_i \text{ falsch}) \\ &\geq 1 - \sum_{i=1}^m P(C_i \text{ falsch}) = 1 - \sum_{i=1}^m (1 - P(C_i \text{ wahr})) \\ &= 1 - m\alpha \end{aligned}$$

Dann ist das simultane Konfidenzintervall für den i -ten Mittelwert mit $100(1 - \alpha)\%$ Konfidenz

$$\bar{\mathbf{X}}_i \pm t_{n-1}(\alpha/2p) \sqrt{s_{ii}/n}$$

Bonferroni Regionen definieren ein Rechteck dessen Seitenlängen kleiner sind als die entsprechenden Projektionen der T^2 Ellipse.