$H_0: \mu = \mu_0$

$t = \frac{(\bar{x}-\mu_0)}{s/\sqrt{n}} \sim t_{n-1}$ where $s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j-\bar{x})^2$

$\left|\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right| \le t_{n-1}(\alpha/2)$. $[\bar{x} - t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}}, \bar{x} + t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}}]$.

$n\frac{(\bar{x}-\mu_0)^2}{\sigma^2} \overset{H_0}{\sim} \chi^2_1$.

$H_0: \mu = \mu_0$. $n(\bar{x}-\mu_0)^T\Sigma^{-1}(\bar{x}-\mu_0) \overset{H_0}{\sim} \chi^2_p$

$\sum z_i^2 = \sum z_i z_i^T$. $z_i \sim N(0,1)$

$(a^T\Sigma a)\sum z_i^2 = \sum[(a^T\Sigma a)^{1/2}z_i]^2$

$= \sum[(a^T\Sigma a)^{1/2}z_i][(a^T\Sigma a)^{1/2}z_i]^T$ $\quad N(0, a^T\Sigma a)$

① $\chi^2_p = Z^TZ = \sum_{i=1}^{p} z_i^2$ $\quad Z \sim N_p(0, I)$, $z_i \sim N(0,1)$

② $(x-\mu)^T\Sigma^{-1}(x-\mu) = [\underbrace{\Sigma^{-\frac{1}{2}}(x-\mu)}_{Z \sim N_p(0,I)}]^T [\underbrace{\Sigma^{-\frac{1}{2}}(x-\mu)}_{Z \sim N_p(0,I)}] = \sum_{i=1}^{p} z_i^2 = \chi^2_p$.

$(x-\mu) \sim N_p(0,\Sigma)$, $\Sigma^{-\frac{1}{2}}(x-\mu) \sim N_p(0, I)$

$R^{\frac{1}{2}}N(0,\sigma^2) \sim N(0, R\sigma^2)$, $R^{\frac{1}{2}}N(0,\Sigma) \sim N_p(0, R\Sigma)$, $\Sigma^{-\frac{1}{2}}N(0,\Sigma) = N_p(0, I_p)$

$\bar{x}-\mu_0 \sim N_p(0, \frac{\Sigma}{n})$, $\sqrt{n}(\bar{x}-\mu_0) \sim N_p(0,\Sigma)$

③ $(\underbrace{\bar{x}-\mu_0}_{N_p(0,\frac{\Sigma}{n})})^T(\frac{\Sigma}{n})^{-1}(\bar{x}-\mu_0)^T = n(\bar{x}-\mu_0)^T\Sigma^{-1}(\bar{x}-\mu_0)^T$

$= [\underbrace{\sqrt{n}(\bar{x}-\mu_0)}_{N_p(0,\Sigma)}]^T\Sigma^{-1}[\sqrt{n}(\bar{x}-\mu_0)]$. $= \chi^2_p$.

$\frac{(n-1)p}{n-p}F_{p,n-p} \to \chi^2_p$

$\frac{(n-1)p}{n-p}\frac{\chi^2_p/p}{\chi^2_{n-p}/n-p} = \frac{(n-1)\chi^2_p}{\chi^2_{n-p}}$

$= \chi^2_p \frac{1}{\frac{1}{n-1}\sum z_i^2} \to \chi^2_p$

① $T^2 = (\bar{x}-\mu_0)^T(\frac{S}{n})^{-1}(\bar{x}-\mu_0)$ $\quad \underset{(p\times 1)}{\bar{x}} = \frac{1}{n}\sum_{j=1}^{n}x_j$ $\quad \underset{(p\times p)}{S} = \frac{1}{n-1}\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})^T$

$= n(\bar{x}-\mu_0)^TS^{-1}(\bar{x}-\mu_0)$

② $T^2 \sim \frac{(n-1)p}{(n-p)}F_{p,n-p}$ $\quad T^2 \overset{H_0: \mu=\mu_0}{=} n(\bar{x}-\mu_0)^TS^{-1}(\bar{x}-\mu_0) > \frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)$ ☆

confidence region for μ: ③ $\{\mu : n(\bar{x}-\mu)^TS^{-1}(\bar{x}-\mu) \le \frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)\}$

μ: ellipsoid - centered at $\bar{x}$

distance: $\left(\frac{(n-1)pF_{p,n-p}(\alpha)}{n\cdot(n-p)}\right)^{\frac{1}{2}}$ to $\bar{x}$.

$\chi^2_m = Z^TZ = \sum_{i}^{m} z_i^2$

$= \sum^m z_iz_i^T$, $z_i \sim N(0,1)$

$= W_{1,m}(1)$

$R\chi^2_m = \sum(R^{\frac{1}{2}}z_i)(R^{\frac{1}{2}}z_i)^T$ $\quad R^{\frac{1}{2}}z_i \sim N(0,R)$

$R\,W_{1,m}(1) = W_{1,m}(R)$.

$RW_{p,m}(\Sigma) = W_{p,m}(R\Sigma)$?

$A \in R^{q\times p}$

$AW_{p,m}(\Sigma)A^T$

$= W_{q,m}(A\Sigma A^T)$

$= \sum_{j=1}^{m}(AX_j)(AX_j)^T$ $\quad AX_j \sim N_q(0, A\Sigma A^T)$

$W_{p,m}(\Sigma) = \sum_{j=1}^{m}X_jX_j^T$. $X_j \sim N_p(0,\Sigma)$

$\sqrt{n}(\bar{x}-\mu) \sim N_p(0,\Sigma)$

$(n-1)S \sim W_{p,n-1}(\Sigma)$

$W_{p,n-1}(\Sigma) = \sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})^T$

$T^2 = \underbrace{\sqrt{n}(\bar{x}-\mu_0)^T}_{N_p(0,\Sigma)^T}\underbrace{\left(\frac{\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})^T}{n-1}\right)^{-1}}_{\frac{1}{n-1}W_{p,n-1}(\Sigma)}\underbrace{\sqrt{n}(\bar{x}-\mu)}_{N_p(0,\Sigma)}$.

$T^2 \sim \frac{(n-1)p}{(n-p)}F_{p,n-p}$.

$(n_1-1)S_1 \sim W_{p,n_1-1}(\Sigma)$

$(n_2-1)S_2 \sim W_{p,n_2-1}(\Sigma)$

$(n_1+n_2-2)S_{pooled} \sim W_{p,n_1+n_2-2}(\Sigma)$

$\Sigma^{-\frac{1}{2}}W_{p,m}(\Sigma)\Sigma^{-\frac{1}{2}} = W_{p,m}(I_p)$ ✓

$N_p(0,\Sigma)^T[\frac{1}{n-1}W_{p,n-1}(\Sigma)]^{-1}N_p(0,\Sigma)$

$= N_p(0,I_p)^T[\frac{1}{n-1}W_{p,n-1}(I_p)]^{-1}N_p(0,I_p) \sim \frac{(n-1)p}{n-p}F_{p,n-p}$

$Z \sim N(0,I)$, $V\!\cdot\! \sim W_p(v,I)$: $Z^T(\frac{W}{v})^{-1}Z \sim \frac{vp}{v-p+1}F_{p,v-p+1}$

$a_1 = [0, 1, 0 \cdots 0]$ $\quad a_1^T\mu = \mu_2$.

$a_2 = [1, -1, 0 \cdots 0]$ $\quad a_2^T\mu = \mu_1 - \mu_2$.

I:

① t: $\frac{(a^T(\bar{x}-\mu))}{\sqrt{a^T\frac{S}{n}a}} \sim t_{n-1}$ $\quad a^T\frac{\Sigma}{n}a$.

$P(a^T\mu \in C(\alpha)) = 1-\alpha$, $C\alpha$?

$[a^T\bar{x} \pm t_{n-1}(\frac{\alpha}{2})\sqrt{\frac{a^TSa}{n}}]$ ☆

② T^2: $\max_{a\neq 0} \frac{n(a^T\boxed{\bar{x}-\mu})^2}{a^T\boxed{S}a} = n(\bar{x}-\mu)^TS^{-1}(x-\mu) = T^2$, at $a = c\boxed{S}^{-1}\boxed{\bar{x}-\mu}$. $\quad C^2 = \frac{p(n-1)}{n-p}F_{p,n-p}(\alpha)$

$P(\forall a^T\mu \in C) = 1-\alpha$, C?

$[a^T\bar{x} \pm \sqrt{\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)}\sqrt{\frac{a^TSa}{n}}]$ ☆

bivar confidence region

③ Bonferroni: $P(A^c \cup B^c) \le P(A^c) + P(B^c) \le \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha$.

$P[a_1^T\mu \in C_1, \ldots a_m^T\mu \in C_m] \ge 1-\alpha$, $C_1? \sim C_m?$

$\bar{x}_i \pm t_{n-1}(\frac{\alpha}{2p})\sqrt{\frac{S_{ii}}{n}}$.

$[a^T\bar{x} \pm t_{n-1}(\frac{\alpha}{2p})\sqrt{\frac{a^TSa}{n}}]$ ☆

$H_0: a_1^T\mu = 0$, $a_2^T\mu = 0$, $a_3^T\mu > 0$. at least one ↑.

$|a_1^T\bar{x} - a_1^T\mu| > t(\frac{\alpha}{2m})\sqrt{\frac{a_1^TSa_1}{n}}$, a_2^T —, a_3^T — $>$, reject H_0

H_0: >0 >0 >0. $<$ $<$ reject

X. $\bar{x} = \begin{pmatrix} x_1 \\ x_p \end{pmatrix}$ $\quad S = \begin{pmatrix} S_{11}, S_{1p} \\ S_{p1}, S_{pp} \end{pmatrix}$ $\quad \mu = \begin{pmatrix} \mu_1 \in [C_1] \\ \mu_p \in [C] \end{pmatrix}$

$n(\bar{x}-\mu)^TS^{-1}(\bar{x}-\mu) \to \chi^2_p$, as $n \to \infty$

pair $\begin{pmatrix} x_{1i1} - x_{2i1} \\ x_{1ip} - x_{2ip} \end{pmatrix}$ $\quad H_0: E d_i = 0$

$d_i = x_{1i} - x_{2i}$

$\bar{d} = \frac{1}{n}\sum_{i=1}^{n}(x_{1i}-x_{2i}) = \bar{x}_1 - \bar{x}_2$ $\quad \delta = \mu_1 - \mu_2$

$S_d = \frac{1}{n-1}\sum_{i=1}^{n}(d_i-\bar{d})(d_i-\bar{d})^T$

$d_1 \sim d_n$ ind $\sim N(\delta,\Sigma)$;

$T^2 = n(\bar{d}-\delta)^TS_d^{-1}(\bar{d}-\delta) \sim \frac{(n-1)p}{n-p}F_{p,n-p}$

$\underset{2p\times 1}{x_i} = \begin{bmatrix} x_{1i} \\ x_{2i} \end{bmatrix}$ $\quad \underset{2p\times 1}{\bar{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}$ $\quad \underset{2p\times 2p}{S} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$

$\underset{p\times 2p}{C} = \begin{bmatrix} 1 & 0 & \cdots & -1 & 0 \\ 0 & 1 & & 0 & -1 \\ & & \ddots & & \\ & & 1 & & -1 \end{bmatrix}$

$d_i = Cx_i$ $\quad \bar{d} = C\bar{x}$ $\quad S_d = CSC^T$

$T^2 = n\bar{x}^TC^T(CSC^T)^{-1}C\bar{x} \sim \frac{(n-1)p}{n-p}F_{p,n-p}$

pair

Two population

☆ $\Sigma_1 = \Sigma_2 = \Sigma$. and two are MVN

$\bar{X}_1 - \bar{X}_2 \sim N_p(\mu_1-\mu_2, (\frac{1}{n_1}+\frac{1}{n_2})\Sigma)$

$(n_1-1)S_1 \sim W_p(n_1-1, \Sigma)$

$(n_2-1)S_2 \sim W_p(n_2-1, \Sigma)$

$S_{pooled} = \frac{\sum_{j=1}^{n_1}(x_{1j}-\bar{x}_1)(x_{1j}-\bar{x}_1)^T + \sum_{j=1}^{n_2}(x_{2j}-\bar{x}_2)(x_{2j}-\bar{x}_2)^T}{n_1+n_2-2} = \frac{(n_1-1)S_1 + (n_2-1)S_2}{n_1+n_2-2}$

$= \frac{n_1-1}{n_1+n_2-2}S_1 + \frac{n_2-1}{n_1+n_2-2}S_2$

☆ $(n_1+n_2-2)S_{pooled} \sim W_p(n_1+n_2-2, \Sigma)$ $\quad E(S_{pooled}) = \Sigma$

$(n-1)\frac{S}{n} \sim W_p(n-1, \frac{\Sigma}{n})$

$(n-1)S \sim W_p(n-1, \Sigma)$

$T^2 = [\bar{x}_1 - \bar{x}_2 - (\mu_1-\mu_2)]^T[(\frac{1}{n_1}+\frac{1}{n_2})S_{pooled}]^{-1}[\bar{x}_1 - \bar{x}_2 - (\mu_1-\mu_2)]$

$\sim T^2_{p, n_1+n_2-2}$

$(n_1+n_2-2)S_{pooled} \sim W_p(n_1+n_2-2, \Sigma)$

$(\bar{x}-\mu)^T[\frac{S}{n}]^{-1}(\bar{x}-\mu)$

$(n-1)\frac{S}{n} \sim W_p(n-1, \frac{\Sigma}{n})$

$(n_1+n_2-2)(\frac{1}{n_1}+\frac{1}{n_2})S_{pooled} \sim W_p(n-1, (\frac{1}{n_1}+\frac{1}{n_2})\Sigma)$

$\frac{(n_1+n_2-2)p}{(n_1+n_2-2)-p+1}F_{p, n_1+n_2-2-p+1}$

① t: $[a^T\bar{x}_1 - a^T\bar{x}_2 \pm t_{n_1+n_2-2}(\frac{\alpha}{2})\sqrt{(\frac{1}{n_1}+\frac{1}{n_2})a^TS_{pooled}a}]$ ☆

② T^2: $[a^T\bar{x}_1 - a^T\bar{x}_2 \pm \sqrt{\frac{(n_1+n_2-2)p}{(n_1+n_2-2)-p+1}F_{p,(n_1+n_2-2)-p+1}(\alpha)}\sqrt{(\frac{1}{n_1}+\frac{1}{n_2})a^TS_{pooled}a}]$ ☆

③ Bonferroni: $[a^T\bar{x}_1 - a^T\bar{x}_2 \pm t_{n_1+n_2-2}(\frac{\alpha}{2p})\sqrt{(\frac{1}{n_1}+\frac{1}{n_2})a^TS_{pooled}a}]$ ☆

☆. $\Sigma_1 \ne \Sigma_2$ $\quad$ why MVN not enough ?

$n_1 \to \infty$, $n_2 \to \infty$. $\bar{x}_1 - \bar{x}_2 \to N_p(\mu_1-\mu_2, (\frac{\Sigma_1}{n_1}+\frac{\Sigma_2}{n_2}))$

$[\bar{x}_1 - \bar{x}_2 - (\mu_1-\mu_2)]^T(\frac{\Sigma_1}{n_1}+\frac{\Sigma_2}{n_2})^{-1}[\bar{x}_1 - \bar{x}_2 - (\mu_1-\mu_2)]^T \to \chi^2_p$

$n_1 \to \infty$, $n_2 \to \infty$, $S_1 \to \Sigma_1$, $S_2 \to \Sigma_2$

$[\bar{x}_1 - \bar{x}_2 - (\mu_1-\mu_2)]^T(\frac{S_1}{n_1}+\frac{S_2}{n_2})^{-1}[\bar{x}_1 - \bar{x}_2 - (\mu_1-\mu_2)] \to \chi^2_p$ $\quad T^2$, why?

$E(\bar{x}_1 - \bar{x}_2 - (\mu_1-\mu_2)) = 0$

$Cov(\bar{x}_1 - \bar{x}_2 - (\mu_1-\mu_2)) = Cov(\bar{x}_1 - \bar{x}_2)$

$= Cov(\bar{x}_1) - Cov(\bar{x}_1,\bar{x}_2) - Cov(\bar{x}_2,\bar{x}_1) - Cov(\bar{x}_2)$

$Cov(\frac{1}{n_1}\sum_{i}^{n_1}x_i) = \frac{1}{n_1^2}\sum_{i=1}^{n_1}Cov(x_i) + \frac{1}{n_1^2}\sum_{i\ne j}Cov(x_i,x_j) = \frac{1}{n_1}\Sigma$ ...

$Cov(\frac{1}{n_1}\sum x_{1i}, \frac{1}{n_2}\sum x_{2j}) = \frac{1}{n_1n_2}\sum\sum Cov(x_{1i}, x_{2j}) = 0$

uni: $Var(z_1+z_2) = Cov(z_1,z_1) + Cov(z_1,z_2) + Cov(z_2,z_1) + Cov(z_2,z_2)$

multi: $Cov(z_1+z_2) = E(z_1-Ez_1)(z_1-Ez_1)^T + E(z_1-Ez_1)(z_2-Ez_2)^T + E(z_2-Ez_2)(z_1-Ez_1)^T$ ...

$\bar{x}_1 = \begin{pmatrix} \bar{x}_{11} \\ \bar{x}_{1p} \end{pmatrix}$ $\quad S_1 = \begin{pmatrix} S_{11} & S_{1p} \\ S_{p1} & S_{pp} \end{pmatrix}$ $\quad \mu_1 = \begin{pmatrix} \mu_{11} \\ \mu_{1p} \end{pmatrix}$ $\quad \mu_1 - \mu_2 = \begin{pmatrix} \mu_{11}-\mu_{21} \in [C] \\ \mu_{1p}-\mu_{2p} \in [C] \end{pmatrix}$

ANOVA.

assumptions: $x_{ij} \sim N(\mu_i, \sigma^2)$ iid

$H_0: \mu_1 = \mu_2 = \cdots = \mu_g \Leftrightarrow \tau_1 = \tau_2 = \cdots = \tau_g = 0$

All population same Σ, $\Sigma_1 = \cdots = \Sigma_g$

$\mu_i = \mu + \mu_i - \mu$ $\quad \sum_{i=1}^{g} n_i\tau_i = 0$

$= \mu + \tau_i$

$x_{ij} = \mu + \tau_i + \varepsilon_{ij}$, $\varepsilon_{ij} \sim N(0,\sigma^2)$ and ind

$x_{ij} = \bar{x} + \underbrace{(\bar{x}_i - \bar{x})}_{\hat{\tau}_i} + \underbrace{(x_{ij} - \bar{x}_i)}_{\hat{\varepsilon}_{ij}}$ $\quad$ B treatment, W residual

$\bar{X} = \frac{1}{n}\sum_{i=1}^{g}\sum_{j=1}^{n_i}X_{ij} = \frac{1}{n}\sum_{i=1}^{g}n_i\bar{X}_i = \frac{1}{n}(n_1\bar{x}_1 + \cdots + n_g\bar{x}_g)$

$\bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}X_{ij}$

$\sum\sum(x_{ij}-\bar{x})(x_{ij}-\bar{x})^T = \sum\sum[(x_{ij}-\bar{x}_i) + (\bar{x}_i-\bar{x})][(x_{ij}-\bar{x}_i) + (\bar{x}_i-\bar{x})]^T$

$= \sum\sum(x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)^T + \sum\sum(\bar{x}_i-\bar{x})(\bar{x}_i-\bar{x})^T + \cdots$

$\sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})(x_{ij}-\bar{x})^T$

$= \sum\sum(\bar{x}_i-\bar{x})(\bar{x}_i-\bar{x})^T + \sum\sum(x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)^T$

$\sum_{j}^{n_i}(x_{ij}-\bar{x}_i)(\bar{x}_i-\bar{x})^T = \sum_{j}(x_{ij}-\bar{x}_i)(\bar{x}_i-\bar{x})^T = \sum(n_i\bar{x}_i - n_i\bar{x}_i)(\bar{x}_i-\bar{x})^T = 0$

One-way ANOVA

Treatment $SS_{tr} = \sum_{i=1}^{g}n_i(\bar{x}_i-\bar{x})^2$ $\quad g-1$

Residual $SS_{res} = \sum_{i}\sum_{j}(x_{ij}-\bar{x}_i)^2$ $\quad n-g$

Total $SS_{total} = \sum_{i}\sum_{j}(x_{ij}-\bar{x})^2$ $\quad n-1$

$H_0: \mu_1 = \cdots = \mu_g$ / $\tau_1 = \cdots = \tau_g$

$F = \frac{SS_{tr}/(g-1)}{SS_{res}/(n-g)} \sim F_{g-1, n-g}$

MANOVA

$B = \sum_{i=1}^{g}n_i(\bar{x}_i-\bar{x})(\bar{x}_i-\bar{x})^T = n_1(\bar{x}_1-\bar{x})(\bar{x}_1-\bar{x})^T + \cdots + n_g(\bar{x}_g-\bar{x})(\bar{x}_g-\bar{x})^T$ $\quad g-1$ (B)

$W = \sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)^T = \sum_{i=1}^{g}(n_i-1)S_i = (n_1-1)S_1 + \cdots + (n_g-1)S_g$ $\quad n-g$ (W)

$T = B + W = \sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})(x_{ij}-\bar{x})^T$ $\quad n-1$ (T)

Wilk's Lambda $\quad \Lambda^* = \frac{|W|}{|T|} = \frac{\prod_{i=1}^{p}\lambda_i(W)}{\prod_{i=1}^{p}\lambda_i(T)}$ ← W sym. $W = U\Lambda U^T$, $\Lambda = (\lambda_1 \ldots \lambda_p)$, $|W| = |U\Lambda U^T| = |U||\Lambda||U^T|$ $\quad \Lambda^* = \frac{|W|}{|W+B|}$

Two-way

$\mu_{lk} = \mu + \tau_l + \beta_k + \gamma_{lk}$ $\quad X_{lkr} = \mu + \tau_l + \beta_k + \gamma_{lk} + e_{lkr}$

$\sum_{l=1}^{g}\tau_l = \sum_{k=1}^{b}\beta_k = \sum_{l=1}^{g}\gamma_{lk} = \sum_{k=1}^{b}\gamma_{lk} = 0$

factor A 1 … g, factor B 1 2 … b, X_{lkr} … X_{gbr}

$x_{lkr} = \underset{\hat{\mu}}{\bar{x}} + \underset{\hat{\tau}_l}{(\bar{x}_{l\cdot} - \bar{x})} + \underset{\hat{\beta}_k}{(\bar{x}_{\cdot k} - \bar{x})} + \underset{\hat{\gamma}_{lk}}{(\bar{x}_{lk} - \bar{x}_{l\cdot} - \bar{x}_{\cdot k} + \bar{x})} + \underset{\hat{e}_{lkr}}{(x_{lkr} - \bar{x}_{lk})}$

MANOVA

SS_A: $\sum_{l} bn(\bar{x}_{l\cdot}-\bar{x})(\bar{x}_{l\cdot}-\bar{x})^T$ $\quad g-1$

SS_B: $\sum_{k} gn(\bar{x}_{\cdot k}-\bar{x})(\bar{x}_{\cdot k}-\bar{x})^T$ $\quad b-1$

$SS_{interaction}$: $\sum_{l}\sum_{k} n(\bar{x}_{lk} - \bar{x}_{l\cdot} - \bar{x}_{\cdot k} + \bar{x})(\bar{x}_{lk} - \bar{x}_{l\cdot} - \bar{x}_{\cdot k} + \bar{x})^T$ $\quad (g-1)(b-1)$

$SS_{residual}$: $\sum_{l}\sum_{k}\sum_{r}(x_{lkr} - \bar{x}_{lk})(x_{lkr} - \bar{x}_{lk})^T$ $\quad gb(n-1)$

SS_{total}: $\sum_{l}\sum_{k}\sum_{r}(x_{lkr} - \bar{x})(x_{lkr} - \bar{x})^T$ $\quad gbn-1$

① $\underbrace{(x-\hat{\mu}_x)^T}_{N_p(0,\Sigma_x)^T}\underbrace{\hat{\Sigma}_x^{-1}}_{\frac{W_p(df_S, \Sigma_x)}{df_S}}\underbrace{(x-\hat{\mu}_x)}_{N_p(0,\Sigma_x)} \sim \frac{df_S \cdot p}{df_S - p + 1}F_{p, df_S - p + 1}$

$N_p(0,\Sigma)^T[\frac{1}{v}W_p(v,\Sigma)]^{-1}N_p(0,\Sigma) \sim T^2_{p,v}$

$N_p(0,I)^T[\frac{1}{v}W_p(v,I)]^{-1}N_p(0,I) \sim \frac{vp}{v-p+1}F_{p,v-p+1}$

② $(\hat{\Sigma}_x^{-\frac{1}{2}}(x-\hat{\mu}_x))^T[\hat{\Sigma}_x^{-\frac{1}{2}}\hat{\Sigma}_x\hat{\Sigma}_x^{-\frac{1}{2}}]^{-1}(\hat{\Sigma}_x^{-\frac{1}{2}}(x-\hat{\mu}_x)) \sim \frac{df_S \cdot p}{df_S - p + 1}F_{p, df_S - p + 1}$

$N_p(0,I_p)^T$, $\frac{W_p(df_S, I_p)}{df_S}$, $N_p(0,I_p)$

$W_p(n_1-1,\Sigma)$ $\quad W_p(n_2-1,\Sigma)$

$= (n_1-1)S_1 + (n_2-1)S_2$

$(n_1+n_2-2)S_{pooled} \sim W_p(n_1+n_2-2, \Sigma)$

③ $(A(x-\hat{\mu}_x))^T[A\hat{\Sigma}_xA^T]^{-1}_{q\times q}(A(x-\hat{\mu}_x))_q \sim \frac{df_S \cdot q}{df_S - q + 1}F_{q, df_S - q + 1}$

$N_q(0, A\Sigma A^T)$, $\frac{W_q(df_S, A\Sigma_x A^T)}{df_S}$, $N_q(0, A\Sigma_x A^T)$

① $(\bar{x}-\mu_0)^T(\frac{S}{n})^{-1}(\bar{x}-\mu_0)$.

$N_p(0,\frac{\Sigma}{n})^T$, $\frac{W_p(n-1,\frac{\Sigma}{n})}{n-1}$, $N_p(0,\frac{\Sigma}{n})$.

$\sqrt{n}(\bar{x}-\mu_0)^T$ S^{-1} $\sqrt{n}(\bar{x}-\mu_0)^T$

$N_p(0,\Sigma)^T$, $\frac{W_p(n-1,\Sigma)}{n-1}$, $N_p(0,\Sigma)$ $\quad \frac{(n-1)p}{n-p}F_{p,n-p}$

$(\bar{x}_1-\bar{x}_2-(\mu_1-\mu_2))^T[(\frac{1}{n_1}+\frac{1}{n_2})S_{pooled}]^{-1}(\bar{x}_1-\bar{x}_2-(\mu_1-\mu_2))$

$N_p(0,(\frac{1}{n_1}+\frac{1}{n_2})\Sigma)^T$, $\frac{W_p(n_1+n_2-2, (\frac{1}{n_1}+\frac{1}{n_2})\Sigma)}{n_1+n_2-2}$, $N_p(0,(\frac{1}{n_1}+\frac{1}{n_2})\Sigma)$.

$\sim \frac{(n_1+n_2-2)p}{(n_1+n_2-2)-p+1}F_{p,(n_1+n_2-2)-p+1}$

②

(1) t. $\quad a^T\mu$: $[a^T\bar{x} \pm t_{n-1}(\alpha/2)\sqrt{\frac{a^TSa}{n}}]$

μ_i: $[\bar{x}_i \pm t_{n-1}(\alpha/2)\sqrt{\frac{S_{ii}}{n}}]$

(2) T^2: $a^T\mu$: $[a^T\bar{x} \pm \sqrt{\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)}\sqrt{\frac{a^TSa}{n}}]$

μ_i: $[\bar{x}_i \pm \sqrt{\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)}\sqrt{\frac{S_{ii}}{n}}]$

(3) Bonferroni: $a^T\mu$: $[a^T\bar{x} \pm t_{n-1}(\alpha/2m)\sqrt{\frac{a^TSa}{n}}]$

μ_i: $[\bar{x}_i \pm t_{n-1}(\alpha/2m)\sqrt{\frac{S_{ii}}{n}}]$

t: $a^T\mu_1 - a^T\mu_2$: $[a^T\bar{x}_1 - a^T\bar{x}_2 \pm t_{n_1+n_2-2}(\frac{\alpha}{2})\sqrt{(\frac{1}{n_1}+\frac{1}{n_2})a^TS_{pooled}a}]$

$\mu_{1i}-\mu_{2i}$: $[\bar{x}_{1i} - \bar{x}_{2i} \pm t_{n_1+n_2-2}(\frac{\alpha}{2})\sqrt{(\frac{1}{n_1}+\frac{1}{n_2})(S_{pooled})_{ii}}]$

T^2: $a^T\mu_1 - a^T\mu_2$: $[a^T\bar{x}_1 - a^T\bar{x}_2 \pm \sqrt{\frac{(n_1+n_2-2)p}{(n_1+n_2-2)-p+1}F_{p,(n_1+n_2-2)-p+1}(\alpha)}\sqrt{(\frac{1}{n_1}+\frac{1}{n_2})a^TS_{pooled}a}]$

$\mu_{1i}-\mu_{2i}$: $[\bar{x}_{1i} - \bar{x}_{2i} \pm \sqrt{\frac{(n_1+n_2-2)p}{(n_1+n_2-2)-p+1}F_{p,(n_1+n_2-2)-p+1}(\alpha)}\sqrt{(\frac{1}{n_1}+\frac{1}{n_2})(S_{pooled})_{ii}}]$

Bonferroni: $a^T\mu_1 - a^T\mu_2$: $[a^T\bar{x}_1 - a^T\bar{x}_2 \pm t_{n_1+n_2-2}(\frac{\alpha}{2m})\sqrt{(\frac{1}{n_1}+\frac{1}{n_2})a^TS_{pooled}a}]$

$\mu_{1i}-\mu_{2i}$: $[\bar{x}_{1i} - \bar{x}_{2i} \pm t_{n_1+n_2-2}(\frac{\alpha}{2m})\sqrt{(\frac{1}{n_1}+\frac{1}{n_2})(S_{pooled})_{ii}}]$

MANOVA: $\bar{x} = \frac{1}{n}\sum_{i=1}^{g}\sum_{j=1}^{n_i}x_{ij} = \frac{1}{n}\sum_{i=1}^{g}n_i\bar{x}_i$. $\quad \bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}x_{ij}$. $\quad x_{ij} = \bar{x} + (\bar{x}_i - \bar{x}) + (x_{ij} - \bar{x}_i)$

$B = \sum_{i=1}^{g}n_i(\bar{x}_i-\bar{x})(\bar{x}_i-\bar{x})^T = n_1(\bar{x}_1-\bar{x})(\bar{x}_1-\bar{x})^T + \cdots + n_g(\bar{x}_g-\bar{x})(\bar{x}_g-\bar{x})^T$ $\quad g-1$

$W = \sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)^T = \sum_{i=1}^{g}(n_i-1)S_i = (n_1-1)S_1 + \cdots + (n_g-1)S_g$ $\quad n-g$

$\Lambda^* = \frac{|W|}{|T|}$ $\quad T = B + W = \sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})(x_{ij}-\bar{x})^T$. $\quad n-1$.

PCA: $\frac{\hat{\lambda}_i}{1+z(\alpha/2)\sqrt{2/n}} \le \lambda_i \le \frac{\hat{\lambda}_i}{1-z(\alpha/2)\sqrt{2/n}}$

Bonferroni: $\frac{\hat{\lambda}_i}{1+z(\alpha/2m)\sqrt{2/n}} \le \lambda_i \le \frac{\hat{\lambda}_i}{1-z(\alpha/2m)\sqrt{2/n}}$.

(almost completely determined by X_k, due to large $var(X_k)$ relative to $X_?$.

Ch6 . PCA

1. reduce the dim, there is larger number of inter-related variables; retain as much as possible the variation in original set of variables.
2. variable: var↑ informative/important↑.
3. The size of the element of u_i are indicators of a variable's importance to ith PC.

S: i-th PC is essentially the X_k, because: $Var(X_k) >> Var(X^{next})$
standardize, use R not S, to remove bias caused by diff scales of Var

Can be summarized $< p$ dim:
$\square$% of the total var accounted by 1-PC ~ q-PC without much loss of information

$X_{1\times p} = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$ $\quad Y_{1\times q} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_p \end{pmatrix} = \begin{pmatrix} a_1^T x \\ \vdots \\ a_p^T x \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \dots a_{1p}x_p \\ \vdots \\ a_{p1}x_1 + \dots a_{pp}x_p \end{pmatrix}$

Standardized Variables (Z-score): $Z = D_\sigma^{-1/2}(X-\mu)$ $\quad z_i = \frac{x_i - \mu_i}{\sqrt{\sigma_{ii}}}$

$Cov(AX) = E[(AX - EAX)(AX - EAX)^T] = E[A(x-\mu)(x-\mu)^T A^T] = A\,cov(x)\,A^T$

$\begin{cases} var(Y_i) = a_i^T \Sigma_x a_i & var(Y_i) = var(a_i^T X) = E(a_i^T x - a_i^T \mu)^2 = E[a_i^T(x-\mu)\,a_i^T(x-\mu)] = E[a_i^T(x-\mu)(x-\mu)^T a_i] = a_i^T E[(x-\mu)(x-\mu)^T]a_i \\ Cov(Y_i, Y_k) = a_i^T \Sigma_x a_k & cov(Y_i,Y_k) = cov(a_i^T X, a_k^T x) = E[(a_i^T x - a_i^T\mu)(a_k^T x - a_k^T \mu)] = E[a_i^T(x-\mu)a_k^T(x-\mu)] = E[a_i^T(x-\mu)(x-\mu)^T a_k] = a_i^T \dots a_k \end{cases}$

① $\max_{a\neq 0} \frac{a^T \Sigma a}{a^T a} = \lambda_1$ $(a = u_1)$ ② $\min_{a \neq 0} \frac{a^T \Sigma a}{a^T a} = \lambda_p$ $(a = u_p)$ ③ $\max_{a \perp u_1 \dots u_k} \frac{a^T \Sigma a}{a^T a} = \lambda_{k+1}$ $(a = u_{k+1})$

①② $\frac{a^T \Sigma a}{a^T a} = \frac{a^T U \Lambda U^T a}{a^T U U^T a} \xrightarrow{d = U^T a} \frac{d^T \Lambda d}{d^T d} = \frac{\sum \lambda_i d_i^2}{\sum d_i^2}$ $\begin{cases} \leq \lambda_1 \frac{\sum d_i^2}{\sum d_i^2} = \lambda_1 & (d_2 = \dots d_p = 0) \; d = \begin{pmatrix} d_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\ \geq \lambda_p \dots = \lambda_p & (d_1 = \dots d_{p-1} = 0) \; d = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ d_p \end{pmatrix} \end{cases}$

$a = u_1 \Rightarrow d = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \Rightarrow \frac{d^T \Lambda d}{d^T d} = \lambda_1$

③ $a \perp u_1 \dots u_k$: $u_i^T a = 0 \xrightarrow{a = Ud} d_i = 0$ $(d_1 = \dots = d_k = 0)$ $\quad d = \begin{pmatrix} 0 \\ \vdots \\ d_{k+1} \\ \vdots \end{pmatrix}$

$\frac{a^T \Sigma a}{a^T a} = \frac{\sum \lambda_i d_i^2}{\sum d_i^2} = \frac{0 + \sum_{i=k+1} \lambda_i d_i^2}{0 + \sum_{i=k+1} d_i^2} \leq \lambda_{k+1}$ $\quad (d_{k+2} = \dots d_p = 0)$

Σ, R: $\begin{cases} \lambda_i \neq \tilde{\lambda}_i \\ u_i \neq \tilde{u}_i \end{cases}$

$\Sigma \to \lambda_1 \geq \dots \geq \lambda_p \geq 0$. $\quad R = D_\sigma^{-1/2} \Sigma_x D_\sigma^{-1/2} \to \tilde{\lambda}_1 \geq \dots \geq \tilde{\lambda}_p \geq 0$

☆ i-th PC: $Y_i = u_i^T X = u_{i1}X_1 + \dots + u_{ip}X_p = (u_{i1} \dots u_{ip})\begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$ $\quad \tilde{Y}_i = \tilde{u}_i^T Z_i = \tilde{u}_i^T (D_\sigma^{-1/2}(x-\mu))$

$Y = U^T X = \begin{pmatrix} Y_1 = u_1^T X \\ Y_q = u_q^T X \\ Y_p = u_p X \end{pmatrix} = \begin{pmatrix} (u_{11} \dots u_{1p})\begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \\ \vdots \\ (u_{p1} \dots u_{pp})\begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \end{pmatrix} = \begin{pmatrix} u_{11}X_1 + \dots + u_{1p}X_p = Y_1 \\ u_{q1}X_1 + \dots + u_{qp}X_q = Y_q \\ u_{p1}X_1 + \dots + u_{pp}X_p = Y_p \end{pmatrix} \Rightarrow \begin{pmatrix} Y_1 \\ \vdots \\ Y_q \end{pmatrix}$

☆ var explained by ith PC
$\begin{cases} var(Y_i) = u_i^T \Sigma u_i = \lambda_i & (\Sigma u_i = \lambda_i u_i \Rightarrow u_i^T \Sigma u_i = \lambda_i) \\ Cov(Y_i, Y_k) = u_i^T \Sigma u_k = 0 & (u_i^T \Sigma u_k = u_i^T \lambda_k u_k = \lambda_k u_i^T u_k) \end{cases}$

$Cov(Y) = \Lambda$ $\left(cov(Y) \xlongequal{Y = U^T X} U^T \Sigma_x U \xlongequal{\Sigma_x = U \Lambda U^T} \Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{pmatrix} \right)$

$\sum_{i=1}^p Var(X_i) = \sum_{i=1}^p var(Y_i) = tr(\Sigma)$ $\quad (tr(\Sigma_x) = tr(U\Lambda U^T) = tr(\Lambda U^T U) = tr(\Lambda))$
$\sum \sigma_{ii}$ $\quad \sum \lambda_i$

$\sum_{i=1}^p Var(Z_i) = \sum_{i=1}^p Var(\tilde{Y}_i) = tr(R) = p$

☆ $\begin{cases} \text{Proportion of total var due to } k\text{-th PC: } \frac{\lambda_k}{\lambda_1 + \dots + \lambda_p} = \frac{\lambda_k}{\sum_{i=1}^p \lambda_i} \\ \text{Proportion of var accounted for by the first } q \text{ PCs: } \frac{\sum_{k=1}^q \lambda_k}{tr(\Sigma_x)} \end{cases}$

☆ $Cor(Y_i, X_k) = \rho_{Y_i, X_k} = \frac{u_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}$. $\quad \rho_{\tilde{Y}_i, Z_k} = \sqrt{\tilde{\lambda}_i}\,\tilde{u}_{ik}$ $\quad \left(\frac{Cov(Y_i, X_k)}{\sqrt{var(Y_i)}\sqrt{var(X_k)}} = \frac{Cov(u_i^T x, e_k^T x)}{\sqrt{\lambda_i}\sqrt{\sigma_{kk}}} = \frac{u_i^T \Sigma_x e_k}{\dots} = \frac{u_i^T \lambda_i e_k}{\dots} = \frac{\lambda_i u_i^T e_k}{\dots} = \frac{\lambda_i u_{ik}}{\dots} \right)$

$\sqrt{\lambda_i}\,u_{ik} / \sqrt{\sigma_{kk}}$

$S_{p\times p} = \frac{1}{n-1}\sum (x_i - \bar{x})(x_i - \bar{x})^T = \frac{1}{n-1}(\sum x_i x_i^T - n\bar{x}\bar{x}^T) \to \hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_p$

i-th PC: $\hat{y}_i = \hat{u}_i^T X = \hat{u}_{i1}x_1 + \dots + \hat{u}_{ip}x_p$.

$u_i = \begin{pmatrix} u_{i1} \\ \vdots \\ u_{ip} \end{pmatrix}$

$\begin{cases} var(\hat{y}_i) = \hat{\lambda}_i \\ cov(\hat{y}_i, \hat{y}_k) = 0 \end{cases}$

$tr(S) = \sum_{i=1}^p S_{ii} = \sum_{i=1}^p \hat{\lambda}_i$

proportion of total sample var ith PC: $\frac{\hat{\lambda}_i}{\sum \hat{\lambda}_k}$

correlation between $\hat{y}_i, \hat{y}_k$: $r_{\hat{y}_i, x_k} = \frac{\sqrt{\hat{\lambda}_i}}{\sqrt{s_{kk}}}\hat{u}_{ik}$ $\quad r_{\hat{\tilde{y}}_i, z_k} = \sqrt{\hat{\tilde{\lambda}}_i}\,\hat{\tilde{u}}_{ik}$

$X = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix} = \begin{pmatrix} x_{11} \dots x_{1p} \\ \vdots \\ x_{n1} \dots x_{np} \end{pmatrix}_{n\times p}$ $\quad \hat{Y} = \begin{pmatrix} \hat{u}_1^T x_1 \dots \hat{u}_q^T x_1 \\ \vdots \\ \hat{u}_1^T x_n \dots \hat{u}_q^T x_n \end{pmatrix} = \begin{pmatrix} \hat{u}_1^T \begin{pmatrix} x_{11} \\ \vdots \\ x_{1p} \end{pmatrix} \dots \hat{u}_q^T \begin{pmatrix} x_{11} \\ \vdots \\ x_{1p} \end{pmatrix} \\ \vdots \\ \hat{u}_1^T \begin{pmatrix} x_{n1} \\ \vdots \\ x_{np} \end{pmatrix} \dots \hat{u}_q^T \begin{pmatrix} x_{n1} \\ \vdots \\ x_{np} \end{pmatrix} \end{pmatrix}$

$= \begin{pmatrix} \hat{u}_{11}x_{11} + \dots \hat{u}_{1p}x_{1p} & \dots & \hat{u}_{q1}x_{11} + \dots \hat{u}_{qp}x_{1p} \\ \vdots \end{pmatrix}$

PC score

$x_1 \dots x_n \sim N_p(\mu, \Sigma)$ $\quad Y_i = u_i^T(X - \mu) \sim MVN$ $\quad \hat{Y} = \begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_p \end{pmatrix} \sim N_p(0, \Lambda)$ $\quad \Sigma$ unknown.

n large $\sqrt{n}(\hat{\lambda} - \lambda) = \sqrt{n}\left(\begin{pmatrix} \hat{\lambda}_1 \\ \vdots \\ \hat{\lambda}_p \end{pmatrix} - \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_p \end{pmatrix}\right) \approx N_p(0, 2\Lambda^2)$, $\quad \hat{\lambda} \approx N_p(\lambda, \frac{2}{n}\Lambda^2)$ $\quad \frac{\sqrt{n}(\hat{\lambda}_i - \lambda_i)}{\sqrt{2}\lambda_i} \sim N_1(0,1)$
$\sqrt{n}(\hat{\lambda}_i - \lambda_i) \approx N_1(0, 2\lambda_i^2)$. $\quad \hat{\lambda}_i \approx N_1(\lambda_i, \frac{2}{n}\lambda_i^2)$.

☆ $(1-\alpha)100\%$ CI for λ_i: $\frac{\hat{\lambda}_i}{1 + z_{\alpha/2}\sqrt{2/n}} \leq \lambda_i \leq \frac{\hat{\lambda}_i}{1 - z_{\alpha/2}\sqrt{2/n}}$. $\quad$ Bonferroni: $\frac{\hat{\lambda}_i}{1 + z_{\alpha/2m}\sqrt{2/n}} \leq \lambda_i \leq \frac{\hat{\lambda}_i}{1 - z_{\alpha/2m}\sqrt{2/n}}$

$V = \frac{1}{p}\sum_{j=1}^m \left[\sum_{i=1}^p \left(\frac{\hat{l}_{ij}^{*2}}{\hat{h}_i^2} \right)^2 - \frac{1}{p}\left(\sum_{i=1}^p \frac{\hat{l}_{ij}^{*2}}{\hat{h}_i^2} \right)^2 \right]$

$a = \sum (\frac{l_{ij}^2}{h_i^2})^2$, $\quad b = \sum \frac{l_{ij}^2}{h_i^2}$ $\quad C = \sum_{j=1}^m (a - \frac{1}{p}b^2)$ $\quad V = \frac{1}{p}C$.

Factor Analysis

PCA	approximate the cov	obtain principal components
FA		obtain observed variables

L non-random

$\begin{cases} X_{p\times 1} = \mu_{p\times 1} + L_{p\times m} F_{m\times 1} + \varepsilon_{p\times 1} \\ x_i = \mu_i + \sum_{k=1}^m l_{ik} F_k + \varepsilon_i \end{cases}$ (loading) (common) (error (specific f))

$= \mu + \begin{pmatrix} l_{11} \dots l_{1m} \\ \vdots \\ l_{p1} \dots l_{pm} \end{pmatrix}\begin{pmatrix} F_1 \\ \vdots \\ F_m \end{pmatrix} + \varepsilon$ $\quad \begin{cases} X_1 = \mu_1 + l_{11}F_1 + \dots + l_{1m}F_m + \varepsilon_1 \\ X_p = \mu_p + l_{p1}F_1 + \dots + l_{pm}F_m + \varepsilon_p \end{cases} \Rightarrow \begin{cases} X_1 = \mu_1 + \sum_{k=1}^m l_{1k} + \varepsilon_1 \\ X_p = \mu_p + \sum_{k=1}^m l_{pk} + \varepsilon_p \end{cases}$

assume: $\begin{cases} E(F) = 0, & cov(F) = I_m & (cov(F) = EFF^T) \Rightarrow cov(QF) = I_m \text{ (Q orthogonal)} = Q\,cov(F)Q^T = QQ^T \\ E(\varepsilon) = 0, & cov(\varepsilon) = \psi = \begin{pmatrix} \psi_1 & \\ & \psi_p \end{pmatrix} & (cov(\varepsilon) = E\varepsilon\varepsilon^T) \\ cov(F, \varepsilon) = 0 & (cov(F, \varepsilon) = EF\varepsilon^T - EF \cdot E\varepsilon = 0) \end{cases}$

$\begin{cases} E(X - \mu) = 0. & (E(x-\mu) = E(LF + \varepsilon) = LE(F) + E(\varepsilon) = 0) \\ Cov(x) = cov(x - \mu) = \Sigma = LL^T + \psi & (\Sigma = E[(x-\mu)(x-\mu)^T] = E(LF + \varepsilon)(LF+\varepsilon)^T = E(LF + \varepsilon)(F^T L^T + \varepsilon^T) = LE(FF^T)L^T + LE(F\varepsilon^T) + E(\varepsilon F^T)L^T + E(\varepsilon\varepsilon^T)) \\ Cov(X, F) = L. & (E(X - EX)(F - EF)^T = E(x-\mu)F^T = E(LF + \varepsilon)F^T = LE(FF^T) + E\varepsilon F^T = L) \end{cases}$

$\begin{cases} var(x_i) = \sigma_{ii} = \sum_{j=1}^m l_{ij}^2 + \psi_i & ((LL^T)_{ii} = \sum_k l_{ik}(L^T)_{ki} = \sum_k l_{ik} l_{ik}) \\ cov(x_i, x_j) = \sigma_{ij} = \sum_{k=1}^m l_{ik} l_{jk} \end{cases}$ $\quad \psi_i = \sigma_{ii} - h_i^2$

$\begin{cases} \sum_{j=1}^m : h_i^2, \psi_i \\ \sum_{i=1}^p : \text{proportion } j\text{th. } V \end{cases}$ $\quad \sum_{j=1}^m [\sum_{i=1}^p \dots]$

☆ communality: $h_i^2 = \sum_{j=1}^m l_{ij}^2$ $\quad \begin{cases} h_1^2 = \sum_{j=1}^m l_{1j}^2 = l_{11}^2 + \dots + l_{1m}^2 \\ h_p^2 = \sum_{j=1}^m l_{pj}^2 = l_{p1}^2 + \dots + l_{pm}^2 \end{cases}$

specific variance: $\hat{\psi}_i = s_{ii} - \sum_{j=1}^m \hat{l}_{ij}^2$ $\quad \begin{cases} \psi_1 = \sigma_{11} - h_1^2 \\ \psi_p = \sigma_{pp} - h_p^2 \end{cases}$ $\quad \psi = diag(\Sigma - LL^T)$

☆ residual matrix. $S - (\hat{L}\hat{L}^T + \hat{\psi})$ $\quad \begin{bmatrix} 0 & \\ & 0 \end{bmatrix}$

☆ proportion of total sample var due to j-th factor

j-th prop: $\frac{\sum_{i=1}^p \hat{l}_{ij}^2}{trS}$ $\begin{cases} 1\text{-th}: \frac{\sum_{i=1}^p \hat{l}_{i1}^2}{trS} = \frac{\hat{l}_{11}^2 + \dots + \hat{l}_{p1}^2}{trS} \\ m\text{-th}: \frac{\sum_{i=1}^p \hat{l}_{im}^2}{trS} = \frac{\hat{l}_{1m}^2 + \dots + \hat{l}_{pm}^2}{trS} \end{cases}$

PCA: $\Sigma = U\Lambda U^T = (u_1 \dots u_p)\begin{pmatrix} \lambda_1 & \\ & \lambda_p \end{pmatrix}\begin{pmatrix} u_1 \\ \vdots \\ u_p \end{pmatrix} = \sum_{i=1}^p \lambda_i u_i u_i^T$
$= (\sqrt{\lambda_1}u_1 \dots \sqrt{\lambda_p}u_p)\begin{pmatrix} \sqrt{\lambda_1}u_1 \\ \vdots \\ \sqrt{\lambda_p}u_p \end{pmatrix}$

$L = (\sqrt{\lambda_1}u_1 \dots \sqrt{\lambda_m}u_m)$ $\quad \psi = diag(\Sigma - LL^T)$

$S = \sum_{i=1}^p \hat{\lambda}_i \hat{u}_i \hat{u}_i^T$
$\hat{L} = (\sqrt{\hat{\lambda}_1}\hat{u}_1 \dots \sqrt{\hat{\lambda}_m}\hat{u}_m)$ $\quad \hat{\psi}$: $diag(S - \hat{L}\hat{L}^T)$

$\hat{L} = \begin{pmatrix} \sqrt{\hat{\lambda}_1}\hat{u}_{11} \dots \sqrt{\hat{\lambda}_m}\hat{u}_{m1} \\ \vdots \\ \sqrt{\hat{\lambda}_1}\hat{u}_{1p} \dots \sqrt{\hat{\lambda}_m}\hat{u}_{mp} \end{pmatrix}$

$\hat{L}\hat{L}^T = \sum_{i=1}^m \hat{\lambda}_i u_i u_i^T = \hat{U}_1\hat{\Lambda}_1\hat{U}_1^T$
$S = \sum_{i=1}^p \hat{\lambda}_i u_i u_i^T = \hat{U}\hat{\Lambda}\hat{U}^T$
$S - \hat{L}\hat{L}^T = \hat{U}_2\hat{\Lambda}_2\hat{U}_2^T$
$\|S - \hat{L}\hat{L}^T\|_{HS}^2 = tr(\hat{U}_2\hat{\Lambda}_2\hat{U}_2^T)(\hat{U}_2\hat{\Lambda}_2\hat{U}_2^T)^T = tr(\hat{\Lambda}_2\hat{\Lambda}_2^T) = \hat{\lambda}_{m+1}^2 + \dots + \hat{\lambda}_p^2$

$\|S - (\hat{L}\hat{L}^T + \hat{\psi})\|_{HS} \leq \sum_{i=m+1}^p \hat{\lambda}_i^2$

proportion to total sample var due to jth factor: $\frac{\hat{\lambda}_j}{trS}$

	add factor	S or R	test
PCA	=	≠	
MLE	≠	=	✓

MLE: $x_1 \dots x_n$ from $N_p(\mu, \Sigma)$. $\quad S_n = \frac{n-1}{n}S$.
$L(\mu, \Sigma) = -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\Sigma| - \frac{n}{2}[tr(\Sigma^{-1}S_n) + (\bar{x} - \mu)^T\Sigma^{-1}(\bar{x} - \mu)]$
$\hat{\mu} = \bar{x}$. $\Sigma = LL^T + \psi$ $\quad \min_{\Sigma, \psi} \log|LL^T + \psi| + tr(S_n(LL^T + \psi)^{-1})$, $\quad \hat{L}^T\hat{\psi}^{-1}\hat{L}$ diag

Test whether m is adequate? fit a factor model with m factors?

$H_0: \Sigma = L_{p\times m}L^T + \psi_{p\times p}$
$-2\log\Lambda = -2\log\frac{\max L(\mu, \Sigma)}{\max L(\mu, \Sigma)} = -2\log\frac{L(\hat{\mu}, \hat{L}\hat{L}^T + \hat{\psi})}{L(\bar{x}, S_n)} = nF$
$F = \log\left(\frac{|\hat{L}\hat{L}^T + \hat{\psi}|}{|S_n|}\right) + tr(S_n(\hat{L}\hat{L}^T + \hat{\psi})^{-1}) - p$.
$n \to \infty$ $\quad (n - 1 - \frac{1}{6}(2p+5) - \frac{2}{3}m)F \to \chi^2_v$ $\quad v = \frac{1}{2}[(p-m)^2 - (p+m)]$
if $\dots > q_\alpha(\chi^2_v)$ reject. $m = ?$

Factor rotation
$\begin{cases} L \to LQ^T \\ F \to QF \end{cases}$ Q orthogonal $\quad X = \mu + LF + \varepsilon = \mu + LQ^T QF + \varepsilon$

☆ varimax: $V = \frac{1}{p}\sum_{j=1}^m \left[\sum_{i=1}^p \left(\frac{\hat{l}_{ij}^{*2}}{\hat{h}_i^2}\right)^2 - \frac{1}{p}\left(\sum_{i=1}^p \frac{\hat{l}_{ij}^{*2}}{\hat{h}_i^2}\right)^2\right]$ $\quad \begin{cases} \hat{l}_{ij} = \hat{L}_{ij} & \text{(unrotated)} \\ \hat{l}_{ij}^* = (\hat{L}Q^T)_{ij} & \text{(rotated)} \end{cases}$

Factor scores. (estimated values of common factors)

☆ $\begin{cases} \hat{f}_j = \hat{L}^T\Sigma^{-1}(x_j - \bar{x}) = \hat{L}^T S^{-1}(x_j - \bar{x}) \\ \hat{f}_j = (I_m + \hat{L}^T\hat{\psi}^{-1}\hat{L})^{-1}\hat{L}^T\hat{\psi}^{-1}(x_j - \bar{x}) \end{cases}$ $\quad \hat{f}_j = \hat{L}^T S^{-1} z_j$

$x - \mu = LF + \varepsilon$ $\quad \varepsilon \sim (0, \psi)$
$\begin{pmatrix} x - \mu \\ F \end{pmatrix} \sim N_{p+m}\left(0, \begin{pmatrix} LL^T + \psi & L \\ L^T & I_m \end{pmatrix}\right)$
$F | X = x \sim N_m(L^T\Sigma^{-1}(x - \mu), I_m - L^T\Sigma^{-1}L)$

PCA: $\begin{pmatrix} \sqrt{\lambda_1} & \\ & \sqrt{\lambda_m} \end{pmatrix} u_m^T S^{-1}(x_j - \bar{x})$.

$\hat{L}^T\Sigma^{-1}(x_j - \bar{x}) = \hat{L}^T(\hat{L}\hat{L}^T + \hat{\psi})^{-1}_{p\times p}(x_j - \bar{x}) = (I_m + \hat{L}^T\hat{\psi}^{-1}\hat{L})^{-1}_{m\times m}\hat{L}^T\hat{\psi}^{-1}(x_j - \bar{x})$

Classification $\quad P(X = x | Y = k)$ density or $P(X = x | Y = k)$?

Classification

$P(X = x | Y = k)$ density or $P(X = x | Y = k)$?

☆ $P(Y = k | X = x) = \frac{P(X = x | Y = k)P(Y = k)}{P(X = x)} = \frac{f_k(x)\pi_k}{P(X = x)}$. $\quad \begin{cases} P(Y = k) = \pi_k \\ f_k(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x - \mu_k)^T\Sigma^{-1}(x - \mu_k)} \end{cases}$ $\quad \sum_{k=1}^K \pi_k = 1$

☆ $f_k(x)\pi_k$ $\quad f_1(x)\pi_1 > f_2(x)\pi_2$

$\frac{\pi_k}{c \cdot |\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x - \mu_k)^T\Sigma_k^{-1}(x - \mu_k)}$

$\delta_1(x) > \delta_2(x)$

$\begin{cases} LDA: \log\pi_k - \frac{1}{2}(x - \mu_k)^T\Sigma^{-1}(x - \mu_k) \xRightarrow{x^T\Sigma^{-1}\mu_k = \mu_k^T\Sigma^{-1}x} \delta_k(x) = \log\pi_k - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k + x^T\Sigma^{-1}\mu_k \\ QDA: \log\pi_k - \frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(x - \mu_k)^T\Sigma_k^{-1}(x - \mu_k) \Rightarrow \delta_k(x) = \log\pi_k - \frac{1}{2}\mu_k^T\Sigma_k^{-1}\mu_k + x^T\Sigma_k^{-1}\mu_k - \frac{1}{2}\log|\Sigma_k| - \frac{1}{2}x^T\Sigma_k^{-1}x \end{cases}$

LDA: $\delta_k(x) = \log\pi_k - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k + x^T\Sigma^{-1}\mu_k$. $\quad$ (the inputs of every class have same Σ)

decision boundary: $\delta_k(x) = \delta_l(x)$.

clf x as 1: $\delta_1(x) - \delta_2(x) > 0$ ① $\delta_{12} = (\mu_1 - \mu_2)^T\Sigma^{-1}x - \frac{1}{2}(\mu_1 - \mu_2)^T\Sigma^{-1}(\mu_1 + \mu_2) + \log(\frac{\pi_1}{\pi_2})$. ☆

$X = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$ $\quad$ (π_k given π_k or $\hat{\pi}_k$) ② $(\mu_1 - \mu_2)^T\Sigma^{-1}x > C$; $\quad C: \frac{1}{2}(\mu_1 - \mu_2)^T\Sigma^{-1}(\mu_1 + \mu_2) + \log(\frac{\pi_2}{\pi_1})$ ☆

☆ $\hat{\pi}_k = \frac{n_k}{n}$ $\quad \hat{\mu}_k = \bar{x}_k$ $\quad \hat{\Sigma} = \frac{(n_1 - 1)S_1 + \dots + (n_K - 1)S_K}{n - K} = S_{pooled}$ $\quad S_{pooled} = \frac{(n_1 - 1)S_1 + \dots + (n_K - 1)S_K}{n_1 + \dots + n_K - K}$

☆ $\hat{\delta}_k(x) = \log\hat{\pi}_k - \frac{1}{2}\hat{\mu}_k^T\hat{\Sigma}^{-1}\hat{\mu}_k + \hat{\mu}_k^T\hat{\Sigma}^{-1}x = \log\hat{\pi}_k - \frac{1}{2}\bar{x}_k^T S_{pooled}^{-1}\bar{x}_k + \bar{x}_k^T S_{pooled}^{-1}x$

$\hat{\delta}_1(x) - \hat{\delta}_2(x) > 0$. $\quad (\bar{x}_1 - \bar{x}_2)^T S_{pooled}^{-1} x > C$; $\quad C = \frac{1}{2}(\bar{x}_1 - \bar{x}_2)^T S_{pooled}^{-1}(\bar{x}_1 + \bar{x}_2) + \log\frac{\hat{\pi}_2}{\hat{\pi}_1}$. ☆

☆ $TPM = \pi_1 P(2|1) + \pi_2 P(1|2)$ $\quad \Leftarrow P(1)P(\text{clf as } 2 | \text{true } 1) + P(2)P(\text{clf as } 1 | \text{true } 2)$.

OER when $\pi_1 = \pi_2 = \frac{1}{2}$

$\delta_1(x) > \delta_2(x)$: $(\mu_1 - \mu_2)^T\Sigma^{-1}x > \frac{1}{2}(\mu_1 - \mu_2)^T\Sigma^{-1}(\mu_1 + \mu_2) + \log\frac{\pi_2}{\pi_1}$.
$\delta_1(x) < \delta_2(x)$: $<$
$Y = (\mu_1 - \mu_2)^T\Sigma^{-1}x$. $\quad m_1 = E(Y | \text{true } 1) = (\mu_1 - \mu_2)^T\Sigma^{-1}\mu_1$. $\quad \Delta^2 = Cov(Y) = (\mu_1 - \mu_2)^T\Sigma^{-1}(\mu_1 + \mu_2)$
$m_2 = E(Y | \text{true } 2) = (\mu_1 - \mu_2)^T\Sigma^{-1}\mu_2$. $\quad m_1 - m_2 = \Delta^2$.
$P(\text{clf as } 1 | \text{true } 2) = P(Y > \frac{1}{2}(\mu_1 - \mu_2)^T\Sigma^{-1}(\mu_1 + \mu_2)|2)$
$= P\left(\frac{Y - EY_2}{\sqrt{Cov(Y)}} > \frac{\dots - m_2}{\sqrt{Cov(Y)}}\right) = P\left(\frac{Y - EY}{\Delta} > \frac{\Delta}{2}\right) = \Phi(-\frac{\Delta}{2})$
$OER = \Phi(-\frac{\Delta}{2})$.

		Predict N−	Predict P+	
True	N−	TN	FP	N
True	P+	FN	TP	P
		N^*	P^*	n

		True N−	True P+	
Predict	N−	TN	FN	N^*
Predict	P+	FP	TP	P^*
		N	P	n

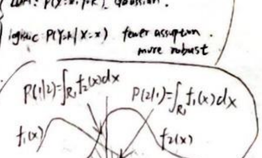
$TPM = \pi(1)\pi(0|1) + \pi(0)\pi(1|0)$
$\approx \frac{P}{n}\frac{FN}{P} + \frac{N}{n}\frac{FP}{N} = \frac{FN + FP}{n}$

$\Omega = \mathbb{R}^p$. all possible x
R_1: $\{x | \text{clf as } C_1\}$ $\quad R_2 = \{x | \text{clf as } C_2\}$
$P(x \in R_2 | \text{true } C_1) = \int_{R_2 = \Omega - R_1} f_1(x)dx$.
$P(x \in R_1 | \text{true } C_2) = \int_{R_1 = \Omega - R_2} f_2(x)dx$.
$TPM = \pi_1 P(2|1) + \pi_2 P(1|2)$ $\quad$ density
$= \pi_1\int_{R_2} f_1(x)dx + \pi_2\int_{R_1} f_2(x)dx$
$= \pi_1(1 - \int_{R_1} f_1(x)dx) + \pi_2\int_{R_1} f_2(x)dx$
$= \int_{R_1}[\pi_2 f_2(x) - \pi_1 f_1(x)]dx + \pi_1$
< 0, $\forall x \in R_1$

Fisher's: C_1 with n_1 points $\quad C_2$ with n_2 points.
$\hat{\mu}_1 = \frac{1}{n_1}\sum_{i \in C_1} x_i$. $\quad \hat{\mu}_2 = \frac{1}{n_2}\sum_{i \in C_2} x_i$ $\quad$ (π_1, π_2 separated as much as possible)
$m_2 - m_1 = a^T(\hat{\mu}_2 - \hat{\mu}_1)$
$S^2 = \frac{1}{n_1 + n_2 - 2}\sum_{k=1}^2\sum_{i \in C_k}(a^T x_i - a^T\hat{\mu}_k)^2$
$\frac{\uparrow \text{Between-class separation}}{\downarrow \text{Within-class variability}} = \frac{(m_2 - m_1)^2}{S^2} = \frac{[a^T(\hat{\mu}_2 - \hat{\mu}_1)]^2}{a^T S_{pooled}\,a}$ $\quad S_{pooled} = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$
max: $a = (\hat{\mu}_1 - \hat{\mu}_2)^T S_{pooled}^{-1}$ $\quad$ max: $(\hat{\mu}_1 - \hat{\mu}_2)^T S_{pooled}^{-1}(\hat{\mu}_1 - \hat{\mu}_2)$
$\begin{cases} x \to 1: (\hat{\mu}_1 - \hat{\mu}_2)S_{pooled}^{-1}x \geq \frac{1}{2}(\hat{\mu}_1 - \hat{\mu}_2)S_{pooled}^{-1}(\hat{\mu}_1 + \hat{\mu}_2) \\ x \to 2: < \end{cases}$

Logistic regression: $P(Y = 1 | X) = \frac{\exp\{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p\}}{1 + \exp\{\beta_0 + \dots + \beta_p x_p\}}$
$P(Y = 0 | X) = \frac{1}{1 + \exp\{\beta_0 + \dots + \beta_p x_p\}}$
$\log\frac{p}{1 - p} = \log\frac{P(Y = 1 | X)}{P(Y = 0 | X)} = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$
$x^* \to 1$: $\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_p x_p > 0$. $\quad P(Y = 1 | x) > P(Y = 0 | x)$.

$\begin{cases} LDA: P(X = x, Y = k) \text{ Gaussian.} \\ \text{logistic } P(Y = k | X = x) \text{ fewer assumption. more robust} \end{cases}$



Matrix:

$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ $\quad A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

$A \in \mathbb{R}^{k\times k}$ $\forall v \in \mathbb{R}^k$. $v^T A v \geq 0 \Rightarrow A$. psd

$A = \lambda_1 u_1 u_1^T + \dots \lambda_k u_k u_k^T$ $\quad u_1^T A u_1 = \lambda_1$ $\quad v^T X^T X v = (Xv)^T(Xv) \geq 0$

orthogonal: $Q = \begin{pmatrix} q_1^T \\ \vdots \\ q_k^T \end{pmatrix}$
$QQ^T = Q^T Q = I$.
$\begin{cases} q_i^T q_i = 1 \\ q_i^T q_j = 0 \end{cases}$

$Ax = \lambda x$. $\quad A \in \mathbb{R}^{k\times k}$
$|A - \lambda I| = 0 \Rightarrow \lambda_1 \sim \lambda_k$ $\quad \begin{vmatrix} a_{11} - \lambda & & a_{1k} \\ & \ddots & \\ a_{k1} & & a_{kk} - \lambda \end{vmatrix} = 0$
$(A - \lambda_i I)x = 0 \Rightarrow x_i / Lx_i \Rightarrow u_i$ $\quad Lx = \sqrt{x_1^2 + \dots + x_n^2}$ $\quad \begin{pmatrix} a_{11} - \lambda_i & & a_{1k} \\ & \ddots & \\ a_{k1} & & a_{kk} - \lambda_i \end{pmatrix} \Rightarrow \begin{pmatrix} 0 \\ 0\;0\;0\;0 \end{pmatrix}$

sym: $A = U\Lambda U^T$

sym: $\begin{cases} det(A) = |U\Lambda U^T| = |U||U^T||\Lambda| = |\Lambda| = \prod\lambda_i \\ tr(A) = tr(U\Lambda U^T) = tr(UU^T\Lambda) = tr(\Lambda) = \sum\lambda_i \end{cases}$ $\quad \begin{cases} u_j^T A u_i = u_i^T A u_j \\ \lambda_i u_j^T u_i = \lambda_j u_i^T u_j. \; \lambda_i \neq \lambda_j. \; u_i^T u_j = 0. \end{cases}$

$x^T A x = a_{11}x_1^2 + a_{22}x_2^2 + 2a_{12}x_1 x_2 = c^2$
$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ $\quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

$\begin{bmatrix} a & & 0 \\ & b & \\ 0 & & c \end{bmatrix}^{-1} = \begin{bmatrix} 1/a & & \\ & 1/b & \\ & & 1/c \end{bmatrix}$.

$A = U\Lambda U^T$.
$x^T A x = \lambda_1(\underbrace{x^T u_1}_{y_1})^2 + \lambda_2(\underbrace{x^T u_2}_{y_2}) = c^2$.
$x_1 = \frac{c}{\sqrt{\lambda_1}}u_1$ $\quad x_2 = \frac{c}{\sqrt{\lambda_2}}u_2$. $\quad \frac{c}{\sqrt{\lambda_2}}u_2$ $\quad \frac{c}{\sqrt{\lambda_1}}u_1$

$R = \rho = \begin{pmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \dots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \ddots & & \vdots \\ \vdots & & & \\ \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} & \dots & & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{pp}}} \end{pmatrix}$.
$= D_\sigma^{-1/2}\Sigma D_\sigma^{-1/2}$.
$= \begin{pmatrix} \frac{1}{\sqrt{\sigma_{11}}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sqrt{\sigma_{pp}}} \end{pmatrix}\begin{pmatrix} \sigma_{11} & \dots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \dots & \sigma_{pp} \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{\sigma_{11}}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sqrt{\sigma_{pp}}} \end{pmatrix}$

$S = \frac{1}{n-1}\sum_{i=1}^n(x_i - \bar{x})(x_i - \bar{x})^T$
$= \frac{1}{n-1}(\sum_{i=1}^n x_i x_i^T - n\bar{x}\bar{x}^T)$. $\quad X = (x_{i1} \dots x_{ip})\; x_i^T$
$= \frac{1}{n-1}(X^T X - n\bar{x}\bar{x}^T)$

A.B.MVN. A.B ind: $= 0$
$Cov\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} varA & cov(A, B) \\ cov(B, A) & varB \end{pmatrix}$
$= 0$

$cov\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \left(\begin{array}{cc|c} \checkmark & \checkmark & \checkmark \\ \checkmark & \checkmark & \checkmark \\ \hline \checkmark & \checkmark & \checkmark \end{array}\right)$ $\quad \Sigma_{12,12}$, $\quad \Sigma_{12,3} = cov(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, x_3)$, $\quad \to cov(x_2, x_3)$, $\quad \Sigma_{3,12}$, $\quad \Sigma_{3,3}$

$Cov\begin{pmatrix} x_2 - 3 \\ 2x_1 - x_2 + x_3 + 1 \end{pmatrix} = Cov\begin{pmatrix} x_2 \\ 2x_1 - x_2 + x_3 \end{pmatrix}$.

$f_X(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\{-\frac{1}{2}(x - \mu)^T\Sigma^{-1}(x - \mu)\}$, $\quad$ if $\Sigma = \begin{pmatrix} * & & 0 \\ & \ddots & \\ 0 & & * \end{pmatrix}$, $\quad f_X\begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} = f_{X_1}(x_1) \cdot \dots \cdot f_{X_p}(x_p)$. ind.

$E(X+Y)$

$E(AXB) = AE(X)B$. A, B constant.

$E(AX) = \begin{pmatrix} E\sum_i A_{1i}X_{i1} & E\sum_i A_{1i}X_{i2} \\ E\sum_i A_{2i}X_{i1} & E\sum_i A_{2i}X_{i2} \end{pmatrix} = \begin{pmatrix} \sum A_{1i}EX_{i1} & \sum A_{1i}EX_{i2} \\ \sum A_{2i}EX_{i1} & \sum A_{2i}EX_{i2} \end{pmatrix} = AE(X)$.

$E(AXB) = \left(E\sum_i\sum_j A_{1i}X_{ij}B_{j1} \;\; \sim \right) = AE(X)B$

$\mu_i = \begin{cases} \int_{-\infty}^{\infty} x_i f_i(x_i)\,dx_i \\ \sum_{\text{all } x_i} x_i\, p_i(x_i) \end{cases}$

$\sigma_i^2 = \begin{cases} \int_{-\infty}^{\infty} (x_i-\mu_i)^2 f_i(x_i)\,dx_i \\ \sum_{\text{all } x_i} (x_i-\mu_i)^2\, p_i(x_i) \end{cases}$

$\sigma_{ik} = E(X_i-\mu_i)(X_k-\mu_k) = \begin{cases} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x_i-\mu_i)(x_k-\mu_k) f_{ik}(x_i,x_k)\,dx_i\,dx_k \\ \sum_{\text{all } x_i}\sum_{\text{all } x_k} (x_i-\mu_i)(x_k-\mu_k)\, p_{ik}(x_i,x_k) \end{cases}$

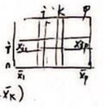
Independence. $P[X_i \le x_i \text{ and } X_k \le x_k] = P[X_i \le x_i]$ and $P[X_k \le x_k]$

$f_{ik}(x_i, x_k) = f_i(x_i)\cdot f_k(x_k)$

X_i, X_k independent $\underset{\not\Leftarrow}{\Longrightarrow}$ $\mathrm{Cov}(X_i, X_k) = 0$

$\mu = E(X)$.

$\Sigma = E(X-\mu)(X-\mu)^T = E\left(\begin{bmatrix} X_1-\mu_1 \\ \vdots \\ X_p-\mu_p \end{bmatrix} [X_1-\mu_1 \sim X_p-\mu_p] \right)$



$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{bmatrix}$

$S = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})^T = \frac{1}{n-1}\sum_{i=1}^{n}\left(\begin{bmatrix} x_{i1}-\bar{x}_1 \\ \vdots \\ x_{ip}-\bar{x}_p \end{bmatrix} [x_{i1}-\bar{x}_1 \sim x_{ip}-\bar{x}_p] \right)$

$\begin{cases} S_{kk} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)^2 \\ S_{jk} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)(x_{ik}-\bar{x}_k) \end{cases}$

$\rho = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \\ & \sim & \end{bmatrix} = D_\sigma^{-\frac{1}{2}}\Sigma D_\sigma^{-\frac{1}{2}}$. $D_\sigma = \begin{pmatrix} \sigma_{11} & & 0 \\ & \ddots & \\ 0 & & \sigma_{pp} \end{pmatrix}$ $D_\sigma^{-\frac{1}{2}} = \begin{pmatrix} \frac{1}{\sqrt{\sigma_{11}}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sqrt{\sigma_{pp}}} \end{pmatrix}$

sample corr $=$ $D_s^{-\frac{1}{2}} S D_s^{-\frac{1}{2}}$

① $c \in \mathbb{R}^p$. $E(c^TX) = c^T\mu \in \mathbb{R}$

$\mathrm{Var}(c^TX) = c^T\Sigma c \in \mathbb{R} \leftarrow$ $= E(c^TX - c^T\mu)^2$ $= E[c^T(x-\mu)c^T(x-\mu)]$ $= E[c^T(x-\mu)(x-\mu)^Tc]$ $= c^TE[(x-\mu)(x-\mu)^T]\cdot c$

$\mathrm{cov}(b^TX, c^TX) = b^T\Sigma c \in \mathbb{R}$

$\overline{c^TX} = c^T\bar{x}$

$S_{c^TX} = c^TSc \leftarrow$ $= \frac{1}{n-1}\sum(c^Tx_i - c^T\bar{x})^2$ $= c^T\left[\frac{1}{n-1}\sum(x_i-\bar{x})(x_i-\bar{x})^T\right]c$

$\mathrm{cov}(b_1X_1 + \dots + b_pX_p,\, c_1X_1 + \dots + c_pX_p)$
$= E[b_1(x_1-\mu_1) + \dots + b_p(x_p-\mu_p)]\cdot[c_1(x_1-\mu_1) + \dots + c_p(x_p-\mu_p)]$
$= E[\sum_i^p b_i(x_i-\mu_i)][\sum_j^p c_j(x_j-\mu_j)]$
$= \sum_i^p\sum_j^p b_ic_jE(x_i-\mu_i)(x_j-\mu_j)$
$= \sum_i^p\sum_j^p \sigma_{ij} = [b_1 \sim b_p]\,\Sigma\,[c_1 \sim c_p]^T$

$= E(b^TX - Eb^TX)(c^TX - Ec^TX)$.
$= E[b^T(x-EX)c^T(x-EX)]$
$= E[b^T(x-EX)(x-EX)^Tc]$
$= b^TE(X-EX)(X-EX)^Tc$

② $C \in \mathbb{R}^{q\times p}$. $E(CX) = C\mu$.

$\mathrm{Cov}(CX) = C\Sigma C^T \leftarrow$ $= E[(CX - E(CX))(CX - E(CX))^T]$ $= E[C(X-EX)(X-EX)^TC^T]$ $= CE[(X-EX)(X-EX)^T]C^T$

$\overline{CX} = C\bar{X}$

$S_{CX} = CSC^T$

random X, Y.

$\mathrm{cov}(X,Y) = E[(X-EX)(Y-EY)^T] = E[XY^T] - E(X)E(Y)^T$.

$\mathrm{cov}(AX, BY) = A\,\mathrm{cov}(X,Y)B^T$

$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{(2\pi)^{\frac{1}{2}}(\sigma^2)^{\frac{1}{2}}}e^{-\frac{1}{2}(x-\mu)\frac{1}{\sigma^2}(x-\mu)}$

$f(x) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$. $N_p(\mu, \Sigma)$

if Σ diag. $\Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_{11}} & & \\ & \ddots & \\ & & \frac{1}{\sigma_{pp}} \end{bmatrix}$. $|\Sigma| = (\sigma_{11}) \sim (\sigma_{pp})$.

$f(x) = f_1(x_1)\cdot f_2(x_2) \sim \cdot f_p(x_p)$

$p=2$.
if $\rho_{12}=0$. $f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = f_1(x_1)f_2(x_2)$

dist of $Y_1 + 2Y_2 - Y_3 + 3Y_4$:
$a = (1\ 2\ -1\ 3)^T$
$N(a^T\mu, a^T\Sigma a)$

X MVN. if $\exists\, \mu \in \mathbb{R}^k, A \in \mathbb{R}^{k\times l}$ st. $X = AZ + \mu$. where $Z_i \sim N(0,1)$.

X MVN $\Leftrightarrow$ $\forall a \in \mathbb{R}^k, a \neq 0$. a^TX is normal

P1: linear comb of normal vector is normal.

$X \sim N_p(\mu, \Sigma)$

$AX + d \sim N_p(A\mu + d, A\Sigma A^T)$

B. Subsets of components of normal are normal.

$\underset{(p\times1)}{X} = \begin{bmatrix} X_1 \\ (q\times1) \\ X_2 \\ ((p-q)\times1) \end{bmatrix}$ $\underset{(p\times1)}{\mu} = \begin{bmatrix} \mu_1 \\ (q\times1) \\ \mu_2 \\ ((p-q)\times1) \end{bmatrix}$. $\underset{(p\times p)}{\Sigma} = \begin{bmatrix} \underset{q\times q}{\Sigma_{11}} & \underset{q\times(p-q)}{\Sigma_{12}} \\ \underset{(p-q)\times q}{\Sigma_{21}} & \underset{(p-q)\times(p-q)}{\Sigma_{22}} \end{bmatrix}$

$X_1 \sim N_q(\mu_1, \Sigma_{11})$.

P3. If X_1 and X_2 independent. $\mathrm{cov}(X_1, X_2) = 0$.

$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N_{q_1+q_2}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12}=0 \\ \Sigma_{21}=0 & \Sigma_{22} \end{bmatrix} \right)$.

$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$. $\mathrm{Var}(X_1) = \mathrm{Var}(X_2)$.

X_1+X_2 and X_1-X_2 independent : $\mathrm{cov}\begin{pmatrix} X_1+X_2 \\ X_1-X_2 \end{pmatrix}$ $\begin{pmatrix} * & 0 \\ 0 & * \end{pmatrix}$

independent or not ?

① X_1+X_2 and X_3 : $\mathrm{Cov}(X_1+X_2, X_3) = \mathrm{Cov}(X_1, X_3) + \mathrm{Cov}(X_2, X_3)$

② $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ and X_3 : $\begin{pmatrix} ✓ & ✓ & ✓ \\ ✓ & ✓ & ✓ \\ ✓ & ✓ & ✓ \end{pmatrix} \to \Sigma_{12,3}$

$\Sigma_{3,12}$

P4. Conditional dist of components are normal

Let $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N_p(\mu, \Sigma)$ $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$. $|\Sigma_{22}| > 0$.

$X_1 | X_2 = x_2$ $\begin{cases} \text{mean}: \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2) \\ \mathrm{cov}: \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \end{cases}$.

$A = \begin{bmatrix} I_q & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0_{(p-q)\times p} & I_{(p-q)} \end{bmatrix}$

$AX = \begin{pmatrix} X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2 \\ X_2 \end{pmatrix}$ $A\mu = \begin{pmatrix} \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2 \\ \mu_2 \end{pmatrix}$ $A\Sigma A^T = \begin{pmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0_{q\times(p-q)} \\ 0_{(p-q)\times q} & \Sigma_{22} \end{pmatrix}$

$X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2$ and X_2 are independent.

$\mathrm{dist}(X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2 | X_2 = x_2) = \mathrm{dist}(X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2) = N_q(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2,\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$

$\mathrm{dist}(X_1 | X_2 = x_2) = N_q(\mu_1 + \Sigma_{12}\Sigma^{-1}(x_2-\mu_2),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$.

$X = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix}$ $\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \Big| \begin{pmatrix} Y_3 \\ Y_4 \end{pmatrix} = \begin{pmatrix} y_3 \\ y_4 \end{pmatrix}$: $\mu = \begin{pmatrix} ✓ \\ ✓ \\ ✓ \\ ✓ \end{pmatrix}$ $\begin{matrix} \leftarrow \mu_1 \\ \leftarrow \mu_2 \end{matrix}$ $\Sigma = \begin{pmatrix} ✓ & ✓ & ✓ & ✓ \\ ✓ & ✓ & ✓ & ✓ \\ ✓ & ✓ & ✓ & ✓ \\ ✓ & ✓ & ✓ & ✓ \end{pmatrix}$ Σ_{11}, Σ_{12}, Σ_{21}, Σ_{22}

$\begin{cases} \text{any}: & \sqrt{n}(\bar{x}-\mu) \to N(0, \Sigma) \\ X_i \overset{\text{iid}}{\sim} N(\mu, \Sigma): & \sqrt{n}(\bar{x}-\mu) \sim N(0, \Sigma) \end{cases}$

$x_1 \sim x_n$ iid from $N_p(\mu, \Sigma)$.

$f(x_1 \sim x_n) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n}\frac{1}{(2\pi)^{\frac{p}{2}}}|\Sigma|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\} = \frac{1}{(2\pi)^{\frac{np}{2}}|\Sigma|^{\frac{n}{2}}}\exp\{-\frac{1}{2}\sum_i^n(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\} = L(\mu, \Sigma)$.

$\sum_{i=1}^{n}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu) = \sum_{i=1}^{n}\mathrm{tr}\{(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\} = \sum_{i=1}^{n}\mathrm{tr}\{\Sigma^{-1}(x_i-\mu)^T(x_i-\mu)\} = \mathrm{tr}\{\Sigma^{-1}\sum_{i=1}^{n}(x_i-\mu)^T(x_i-\mu)\}$

$= \mathrm{tr}\{\Sigma^{-1}[\sum_{i=1}^{n}[(x_i-\bar{x}) + (\bar{x}-\mu)][(x_i-\bar{x}) + (\bar{x}-\mu)]^T\} \overset{\text{cross}=0}{=\!=} \mathrm{tr}\{\Sigma^{-1}[\sum(x_i-\bar{x})(x_i-\bar{x})^T + n(\bar{x}-\mu)(\bar{x}-\mu)^T]\}$

$L(\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{np}{2}}|\Sigma|^{\frac{n}{2}}}\exp\{-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}[\sum(x_i-\bar{x})(x_i-\bar{x})^T + n(\bar{x}-\mu)(\bar{x}-\mu)^T])\}$

to min $\mathrm{tr}\{n\Sigma^{-1}(\bar{x}-\mu)(\bar{x}-\mu)^T\} = n(\bar{x}-\mu)^T\Sigma^{-1}(\bar{x}-\mu) > 0$. unless $\mu = \bar{x}$. $\Rightarrow \hat{\mu} = \bar{x}$.

$L(\hat{\mu}, \Sigma) = \frac{1}{(2\pi)^{\frac{np}{2}}|\Sigma|^{\frac{n}{2}}}\exp\{-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}[\sum_j(x_j-\bar{x})(x_j-\bar{x})^T])\}$.

$\Lambda = \frac{\max L(\theta_0)}{\max L(\theta)}$ $\Lambda \in [0,1]$. Λ small reject H_0

$\max_{\mu, \Sigma} L(\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{np}{2}}|\hat{\Sigma}|^{\frac{n}{2}}}e^{-\frac{np}{2}}$. with $\hat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})^T$, $\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{j=1}^{n}x_j$

Under $H_0: \mu = \mu_0$. $\max_{\Sigma} L(\mu_0, \Sigma) = \frac{1}{(2\pi)^{\frac{np}{2}}|\hat{\Sigma}_0|^{\frac{n}{2}}}e^{-\frac{np}{2}}$. with $\hat{\Sigma}_0 = \frac{1}{n}\sum_{j=1}^{n}(x_j-\mu_0)(x_j-\mu_0)^T$.

$H_0: \mu = \mu_0$

$\Lambda = \left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|}\right)^{\frac{n}{2}} = \left(\frac{|\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})^T|}{|\sum_{j=1}^{n}(x_j-\mu_0)(x_j-\mu_0)^T|}\right)^{\frac{n}{2}} < c_\alpha$ reject H_0 .

$\Lambda^{\frac{2}{n}} = \left(1 + \frac{T^2}{n-1}\right)^{-1}$

$x_1 \sim x_{20}$. $n=20$ $p=6$. $N_6(\mu, \Sigma)$.

1. $(x_1-\mu)^T\Sigma^{-1}(x_1-\mu) \sim \chi_6^2$.
2. $\bar{x} = \frac{1}{n}\sum x_i \sim N_6(\mu, \frac{1}{20}\Sigma)$.
3. $\sqrt{n}(\bar{x}-\mu) \sim N_6(0, \Sigma)$.
4. $(n-1)S \sim W_6(19, \Sigma)$.
5. $B(19S)B^T$ ($B = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$) $\sim W_2(19, B\Sigma B^T)$
$= \begin{pmatrix} \sigma_{11} & \sigma_{13} \\ \sigma_{31} & \sigma_{33} \end{pmatrix}$.

$X = (X_1\ X_2\ X_3)^T$ $N_3(\mu, \Sigma)$ $\Sigma = \begin{pmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix}$

1. X_1 and X_2 : $\mathrm{Cov}(X_1, X_2) = \Sigma_{12} = -2 \neq 0$ not ind.
2. X_2 and X_3 : $\mathrm{cov}(X_2, X_3) = \Sigma_{23} = 0$ ind .
3. $(X_1, X_2)^T$ and X_3 : $\mathrm{cov}\left(\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, X_3\right)$ $\begin{pmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. ind
4. X_2 and $-2X_1 + X_2 - X_3$: $\mathrm{cov}(X_2, -2X_1 + X_2 - X_3)$
$= \mathrm{Cov}(X_2, -2X_1) + \mathrm{cov}(X_2, X_2) + \mathrm{cov}(X_2, -X_3)$

$Y = (Y_1, Y_2, Y_3, Y_4)^T$. $N_4(\mu, \Sigma)$ $\mu = \begin{pmatrix} 1 \\ 2 \\ 3 \\ -2 \end{pmatrix}$. $\Sigma = \begin{pmatrix} 4 & 2 & -1 & 2 \\ 2 & 6 & 3 & -2 \\ -1 & 3 & 5 & -4 \\ 2 & -2 & -4 & 4 \end{pmatrix}$.

1. dis $(Y_1, Y_3)^T$: $\begin{pmatrix} 1 \\ ✓ \\ 3 \\ ✓ \end{pmatrix}$. $\begin{pmatrix} 4 & ✓ & -1 & ✓ \\ ✓ & ✓ & ✓ & ✓ \\ -1 & ✓ & 5 & ✓ \\ ✓ & ✓ & ✓ & ✓ \end{pmatrix}$ $N_2\left(\begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 4 & -1 \\ -1 & 5 \end{pmatrix}\right)$

2. $Y_2 = N(2, 6)$.

3. $Y_1 + 2Y_2 - Y_3 + 3Y_4$: $a = (1\ 2\ -1\ 3)^T$
$N(a^T\mu, a^T\Sigma a)$.

4. $(Y_1, Y_2)^T | (Y_3 = Y_4)^T = (y_3, y_4)^T$: $\begin{matrix} \mu_1 \\ \mu_2 \end{matrix}\begin{pmatrix} ✓ \\ ✓ \\ ✓ \\ ✓ \end{pmatrix}$. $\begin{array}{cc|cc} ✓ & ✓ & ✓ & ✓ \\ ✓ & ✓ & ✓ & ✓ \\ \hline ✓ & ✓ & ✓ & ✓ \\ ✓ & ✓ & ✓ & ✓ \end{array}$ Σ_{11}, Σ_{12}, Σ_{21}, Σ_{22}

length of x. $L_x = \sqrt{x_1^2 + \dots + x_n^2}$

inner product: $x^T y = x_1y_1 + x_2y_2 + \dots + x_ny_n$

angle: $\cos(\theta) = \frac{x^Ty}{L_xL_y} = \frac{x^Ty}{\sqrt{x^Tx}\sqrt{y^Ty}}$

$\vec{x}_1 \dots \vec{x}_k$ { dependent: $\exists c_1 \dots c_k$ not all 0, st. $c_1x_1 + \dots + c_kx_k = 0$
independent: $c_1 = \dots = c_k = 0$.

Proj of x on $y = \frac{x^Ty}{y^Ty}y = \frac{x^Ty}{L_y}\frac{y}{L_y}$. Length $= \frac{|x^Ty|}{L_y} = L_x\left|\frac{x^Ty}{L_xL_y}\right| = L_x|\cos\theta|$.

AB: $c_{ij} = \sum_{l=1}^{n} a_{il}b_{lj}$.

transpose: $(A^T)_{ji} = (A)_{ij}$

determinant: { $|A| = a_{11}$ if $k=1$
$|A| = \sum_{j=1}^{k} a_{1j}|A_{1j}|(-1)^{1+j}$. if $k>1$.

$\exists A^{-1} \iff$ col/row of A independent $\iff$ A has full col rank $\iff |A| \neq 0$

2x2: $A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

symmetric: $A = A^T$ $a_{ij} = a_{ji}$

Q orthogonal: $QQ^T = Q^TQ = I$. $Q^T = Q^{-1}$ { $\det(QQ^T) = \det(Q)\det(Q^T) = \det(Q)^2 = 1$, $\det(Q) = \pm 1$

$Q = \begin{pmatrix} q_1^T \\ \vdots \\ q_k^T \end{pmatrix}$ $QQ^T = \begin{pmatrix} q_1^T \\ \vdots \\ q_k^T \end{pmatrix}(q_1 \dots q_k) = I$ { $q_i^Tq_i = 1$, $q_i^Tq_j = 0$

eigen: $Ax = \lambda x$. $x \neq \vec{0}$.

$|A - \lambda I| = 0 \to \lambda_1 \dots \lambda_k$.

$(A - \lambda_i I)x = 0 \to x_i / L_{x_i} \to u_i$

symmetric:
$A = U\Lambda U^T = (u_1 \dots u_k)\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_k \end{pmatrix}\begin{pmatrix} u_1^T \\ \vdots \\ u_k^T \end{pmatrix}$
$= \sum_{i=1}^{k} \lambda_i u_iu_i^T = \lambda_1u_1u_1^T + \dots + \lambda_ku_ku_k^T$

Symm A: $A = U\Lambda U^T$

① $\det(A) = \det(U\Lambda U^T) = \det(U)\det(\Lambda)\det(U^T)$
$= \det(U)\det(U^T)\det(\Lambda)$
$= \det(\Lambda) = \lambda_1 \dots \lambda_p$

② $tr(A) = tr(U\Lambda U^T) = tr(UU^T\Lambda)$
$= tr(\Lambda) = \lambda_1 + \dots + \lambda_p$

③ $Au_i = \lambda_iu_i \to u_j^TAu_i = \lambda_iu_j^Tu_i$
$Au_j = \lambda_ju_j \to u_i^TAu_j = \lambda_ju_i^Tu_j$
$u_j^TAu_i = u_i^TAu_j \to \lambda_iu_j^Tu_i = \lambda_ju_i^Tu_j$
$\lambda_i \neq \lambda_j \to u_i^Tu_j = 0$

Positive: symmetric: A.
{ positive semi-def: $x^TAx \geq 0$. for $\forall x \in \mathbb{R}^k$
positive def: $x^TAx > 0$ for $\forall x \in \mathbb{R}^k / \{0\}$

X^TX semi: $X \in \mathbb{R}^{n\times p}$ $\forall v \in \mathbb{R}^p$
$v^TX^TXv = (Xv)^T(Xv) \geq 0$

lok symm A is positive def $\iff \forall \lambda_i$ of A is positive!

$A = \lambda_1u_1u_1^T + \dots + \lambda_ku_ku_k^T$.
let $x = u_1$. $u_1^TAu_1 = \lambda_1 \underbrace{u_1^Tu_1 \cdot u_1^Tu_1}_{=\lambda_1} + \lambda_2 \underbrace{u_1^Tu_2 u_2^Tu_1}_{=0} + \dots$
if $\lambda_1 \leq 0$. $x^TAx = u_1^TAu_1 \leq 0$

distance$^2 = a_{11}x_1^2 + \dots + a_{pp}x_p^2 + 2(a_{12}x_1x_2 + \dots + a_{p-1,p}x_{p-1}x_p)$.

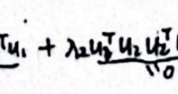
$0 <$ $= [x_1 \dots x_p]\begin{bmatrix} a_{11} & & a_{1p} \\ & \ddots & \\ a_{p1} & & a_{pp} \end{bmatrix}\begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} = x^TAx$. for $x \neq 0$ $a_{ij} = a_{ji}$

$x^TAx = a_{11}x_1^2 + a_{22}x_2^2 + 2a_{12}x_1x_2 = c^2$

$A = U\Lambda U^T = \lambda_1u_1u_1^T + \lambda_2u_2u_2^T$.

$xAx^T = xAx^T = \lambda_1(x^Tu_1)^2 + \lambda_2(x^Tu_2)^2 = c^2$

$x_1 = \frac{c}{\sqrt{\lambda_1}}u_1$ $x_2 = \frac{c}{\sqrt{\lambda_2}}u_2$ $u_1^Tu_2 = 0$



$k(AB) = (kA)B$.
$A(BC) = (AB)C$
$A(B+C) = AB + AC$ } $A\Sigma()A^T = \Sigma A()A$
$(B+C)A = BA + CA$
$(AB)^T = B^TA^T$

$(A^{-1})^T = (A^T)^{-1}$
$(AB)^{-1} = B^{-1}A^{-1}$
A.B square kxk:
$|A| = |A^T|$.
$|A| = \frac{1}{|A^{-1}|}$. $|A||A^{-1}| = 1$
$|AB| = |A||B|$.
$|rA| = r^k|A|$. $|rI| = r^k|I|$
$\begin{bmatrix} a & & 0 \\ & b & \\ 0 & & c \end{bmatrix} = \begin{bmatrix} 1/a & & 0 \\ & 1/b & \\ 0 & & 1/c \end{bmatrix}$

A: symmetric positive semi-def:
$A^{\frac{1}{2}} = \sum_{i=1}^{k} \sqrt{\lambda_i}u_iu_i^T = U\Lambda^{\frac{1}{2}}U^T$
$(A^{\frac{1}{2}})^T = A^{\frac{1}{2}}$
$A^{\frac{1}{2}}A^{\frac{1}{2}} = A$. $A^{\frac{1}{2}}A^{-\frac{1}{2}} = I$.
$(A^{\frac{1}{2}})^{-1} = \sum_{i=1}^{k} \frac{1}{\sqrt{\lambda_i}}u_iu_i^T = U\Lambda^{-\frac{1}{2}}U^T$

A square: $tr = \sum a_{ii}$
$tr(rA) = r\,tr(A)$.
$tr(A \pm B) = tr(A) \pm tr(B)$
$tr(AB) = tr(BA)$
$tr(B^{-1}AB) = tr(B^{-1}BA) = tr(A)$
$tr(AA^T) = \sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij}^2$ ← $A \in \mathbb{R}^{m\times n}$, $AA^T = \sum_{i=1}^{m}(AA^T)_{ii} = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}(A^T)_{ji} = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2$

A symmetric: $tr(A) = \sum_{i=1}^{k}\lambda_i$
$tr(A) = tr(U\Lambda U^T) = tr(UU^T\Lambda) = tr(\Lambda)$

A symm: $A^n = U\Lambda^nU^T$
$A^{\frac{1}{n}} = U\Lambda^{\frac{1}{n}}U^T$