(*) How do we interprete this $CR_{1-\alpha}(\mu)$?
IT IS **NOT** that the probability of $\mu$ being on the $CR_{1-\alpha}$ is $1-\alpha$. THIS IS one realization of a random ellipse. There is no probability here.
There is one ellipse (we have seen the realization) and we're saying $1-\alpha$ % of the time we use this algorithm we cover $\mu$. We don't know if today (with this realization) we're covering or not.

If we toss a coin and we got head it's not that we got head with probability $1/2$. We have the probability before the realization!

We just have CONFIDENCE that the ellipse is generated by an algorithm that covers $1-\alpha$ % of the time $\mu$
(That's why we use CONFIDENCE and NOT PROBABILITY)

(*) If we say to our client that a certain realization of CR has $\mu$ inside $1-\alpha$ % of the time then the client will use it thinking that $1-\alpha$ % of the times he'll be right.
WRONG: If that realization is not covering $\mu$ then the client will be wrong 100 % of the time (!)

30/03

Lecture of 30.03.20
lunedì 30 marzo 2020 10:18

## Inference for the mean $\mu$

$\mathbb{R}^p \ni X_1, \dots, X_n$ n. vectors iid.
with $E[X] = \mu$ and $cov(X_1) = \Sigma$, $\det\Sigma > 0$.

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ estimator for $\mu$ (point) (for n large and small)

• **Large n** ($n >> p$).

$$\sqrt{n}(\bar{X}-\mu) \underset{\text{approx.}}{\sim} N_p(0, \Sigma) \quad \text{(CLT)}$$

$$(\sqrt{n}(\bar{X}-\mu))^T \Sigma^{-1} (\sqrt{n}(\bar{X}-\mu)) \sim \chi^2(p)$$

$$\boxed{n(\bar{X}-\mu)^T\Sigma^{-1}(\bar{X}-\mu) \sim \chi^2(p)}$$

If $\Sigma$ is unknown:

$S \longrightarrow \Sigma$ as $n \to \infty$ (LLN)

Hence: for large n

$$\boxed{n(\bar{X}-\mu)^T S^{-1}(\bar{X}-\mu) \sim \chi^2(p)}$$

pivotal statistics

the unknown part is only the mean, the one parameter we want to estimate (we don't know $\mu$ but we know its distribution)

$$\mathbb{P}\left[ n(\bar{X}-\mu)^T S^{-1}(\bar{X}-\mu) \le \chi^2_\alpha(p)\right] = 1-\alpha$$

$$\mathbb{P}\left[ \boxed{d^2_{S^{-1}}(\bar{X},\mu)} \le \chi^2_\alpha(p)\right] = 1-\alpha \quad \text{if } \alpha \in (0,1)$$

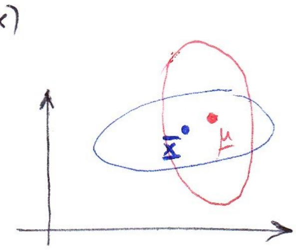
Mahalanobi's distance: it defines an ellipse

$$\mathcal{E}^\alpha_{S^{-1}}(\mu) = \{x \in \mathbb{R}^p : d^2_{S^{-1}}(x,\mu) \le \chi^2_\alpha(p)\}$$

$$\mathcal{E}^\alpha_{S^{-1}}(\bar{X}) = \{y \in \mathbb{R}^p : d^2_{S^{-1}}(y,\bar{X}) \le \chi^2_\alpha(p)\}$$

$$\bar{X} \in \mathcal{E}^\alpha_{S^{-1}}(\mu) \Longleftrightarrow \mu \in \mathcal{E}^\alpha_{S^{-1}}(\bar{X})$$

(*)




if $\bar{X}$ is less than $1-\alpha$ from $\mu$, then $\bar{X}$ belong to the ellipse centered in $\mu$ but then $\mu$ is less than $1-\alpha$ from $\bar{X}$, so it belongs to the ellipse centered on $\bar{X}$

$$\boxed{\mathbb{P}\left[ d^2_{S^{-1}}(\bar{X},\mu) \le \chi^2_\alpha(p)\right] = 1-\alpha}$$

equivalent

1. $$\boxed{\mathbb{P}\left[ \underbrace{\bar{X}}_{\text{random}} \in \underbrace{\mathcal{E}^\alpha_{S^{-1}}(\mu)}_{\text{fixed}} \right] = 1-\alpha}$$

2. $$\boxed{\mathbb{P}\left[ \underbrace{\mu}_{\text{fixed}} \in \underbrace{\mathcal{E}^\alpha_{S^{-1}}(\bar{X})}_{\text{random}} \right] = 1-\alpha}$$

**NOTE**: we need to pay attention to the difference between 1. and 2.: they're probabilities. The probability is the way we capture the uncertainty of the realization of something that is random.
$\bar{X}$ is random, $\mu$ is fixed (unknown, but fixed)

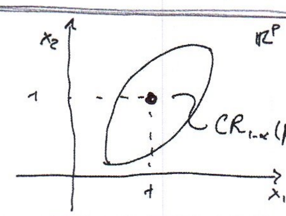
! Here the ellipse is fixed, $\bar{X}$ is random and we're saying that the probability of something random will be in a fixed ellipse is $1-\alpha$ (we don't know the fixed ellipse)

$$\boxed{CR_{1-\alpha}(\mu) = \{y \in \mathbb{R}^p : y \in \mathcal{E}^\alpha_{S^{-1}}(\bar{X})\}}$$
$$= \{y \in \mathbb{R}^p : n(y-\bar{X})^T S^{-1}(y-\bar{X}) \le \chi^2_\alpha(p)\}$$

We're saying that the random ellipse will be covering the true (fixed) value of $\mu$, $1-\alpha$ of times (with $\mathbb{P} = 1-\alpha$)

we also have the uncertainty ($\mathrm{Det}(S)$: it says how the ellipse is large)

**Ex.**
$\bar{X} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$
$S$, $\alpha$




! Here the ellipse is random. The ellipse is randomly generated once we know the data. For every set of data there will be an $\bar{X}$ and an ellipse, but before knowing the data the ellipse will be random

(*)

usually we want to prove that $H_0$ is false: we are in the business of DISPROVING $H_0$ (we want to guarantee the fact that we reject $H_0$ when $H_0$ is true only with very small probability)

**Test** $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$

$\alpha \in (0,1)$ (small).

$\boxed{T_0^2 = n(\bar{x} - \mu_0)^T S^{-1} (\bar{x} - \mu_0)}$

$\boxed{\text{If } H_0 \text{ is true} \Rightarrow T_0^2 \sim \chi^2(p)}$

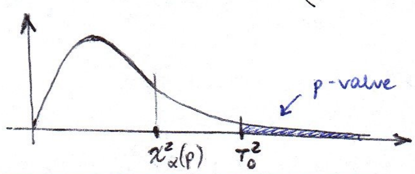
$\boxed{\text{Reject if } T_0^2 > \chi^2_\alpha(p)}$

Rejection Region$_\alpha$ = $\{ T_0^2 > \chi^2_\alpha(p) \}$

When do we have proof against $H_0$? If the value of $\bar{x}$ is very far from $\mu_0$

Means that we reject $H_0$ when $H_0$ is true only $\alpha$% of the time.

How far is "very far"? It should be so far that either a miracle occurred or the assumptions ($H_0$) are wrong.




If we chose $\alpha = 0.05$ and $T_0^2 > \chi^2_{0.05}(p)$ it means that or we saw something that happens only 5% of the times or we've wrong with assumptions ($H_0$).

!! If $H_0$ is true then $T_0^2$ has this distribution. But we have observed $T_0^2$ in the tail. The p-value is the area on the right of the observation. So instead of chosing an $\alpha$, we look for the p-value of the test.

$\boxed{\text{p-value} \leq \alpha \Leftrightarrow T_0^2 > \chi^2_\alpha(p)}$

---

- **Small $n$**

$X_1, \dots, X_n$ iid $\sim N_p(\mu, \Sigma)$, $\det(\Sigma) > 0$

(with $n$ large $X_i \nsim N_p$ necessarily)

$n(\bar{X} - \mu)^T S^{-1} (\bar{X} - \mu) \sim ?$

Def. $y \sim \chi^2(n)$, $w \sim \chi^2(m)$ and $y \perp\!\!\!\perp w$

$$\boxed{\frac{y/n}{w/m} \sim F(n, m)}$$

Obs.

1) $t \sim t(n)$ $\quad t = \dfrac{z}{\sqrt{\frac{w}{n}}}$ $\begin{cases} z \sim N(0,1) \\ w \sim \chi^2(n) \\ z \perp\!\!\!\perp w \end{cases}$

$t^2 = \dfrac{z^2}{w/n}$ $\begin{cases} z^2 \sim \chi^2(1) \\ w \sim \chi^2(n) \\ z^2 \perp\!\!\!\perp w \end{cases}$

$\Rightarrow \boxed{t^2 \sim F(1, n)}$

2) $\boxed{F(n, m) \longrightarrow \frac{1}{n} \chi^2(n)}$ if $m \to \infty$.

$\dfrac{y/n}{w/m}$ $\quad y \sim \chi^2(n)$, $w \sim \chi^2(m)$, $y \perp\!\!\!\perp w$

$w = \sum_{i=1}^{m} z_i^2 \quad z_1, \dots, z_m$ iid $\sim N(0,1)$

$E[z_i^2] = 1$

$\frac{1}{m} \sum_{i=1}^{m} z_i^2 \longrightarrow 1$ as $m \to \infty$ (LLN)

$\frac{w}{m} \longrightarrow 1$ as $m \to \infty$

$\dfrac{y/n}{w/m} \longrightarrow \frac{1}{n} y \sim \frac{1}{n} \chi^2(n)$ as $m \to \infty$

---

**Hotelling's Th.** (1931)

GAUSSIAN + WISHART

$X \sim N_p(\mu, \Sigma)$, $\det(\Sigma) > 0$

$W \sim Wish\left(\frac{1}{m} \Sigma, m\right)$ ($\Sigma$ $p \times p$)

$X \perp\!\!\!\perp W$

$$\Rightarrow \frac{m - p + 1}{m p} (X - \mu)^T W^{-1} (X - \mu) \sim F(p, m - p + 1)$$

**Coro** $X_1, \dots, X_n$ iid $\sim N_p(\mu, \Sigma)$

$$\boxed{n(\bar{X}-\mu)^T S^{-1}(\bar{X}-\mu) \sim \frac{(n-1)p}{n-p} F(p, n-p)}$$

Now we know what is the distribution of the square Mahalanobi's distance between the sample mean and the true mean when the random vectors generating our sample are gaussian

**proof.**

$$\left.\begin{aligned} &\sqrt{n}(\bar{X}-\mu) \sim N_p(0, \Sigma) \\ &S \sim \text{Wish}\left(\tfrac{1}{n-1}\Sigma, n-1\right) \end{aligned}\right\} \bar{X} \perp\!\!\!\perp S$$

Use Hotelling's th. $(m = n-1)$

$$\frac{n-1-p+1}{(n-1)p} n(\bar{X}-\mu)^T S^{-1}(\bar{X}-\mu) \sim F(p, n-1-p+1)$$

$$\left[\frac{n-p}{(n-1)p} n(\bar{X}-\mu)^T S^{-1}(\bar{X}-\mu) \sim F(p, n-p)\right.$$

equiv.

$$\left[n(\bar{X}-\mu)^T S^{-1}(\bar{X}-\mu) \sim \frac{(n-1)p}{n-p} F(p, n-p)\right. \quad \blacksquare$$

$$\boxed{T^2 = n(\bar{X}-\mu)^T S^{-1}(\bar{X}-\mu)} \quad \text{(def.)}$$

---

**Confidence Region**

$$\mathbb{P}\Big[\underbrace{n(\bar{X}-\mu)^T S^{-1}(\bar{X}-\mu)}_{d^2_{S^{-1}}(\bar{X}, \mu)} \leq \frac{(n-1)p}{n-p} F_\alpha(p, n-p)\Big] = 1-\alpha$$

as before $\rightarrow$ $\mathbb{P}\left[\mu \in \mathcal{E}^\alpha_{S^{-1}}(\bar{X})\right] = 1-\alpha \qquad \alpha \in (0,1)$

$$\mathcal{E}^\alpha_{S^{-1}}(\bar{X}) = \left\{ y \in \mathbb{R}^p : d^2_{S^{-1}}(\bar{X}, \mu) \leq \frac{(n-1)p}{n-p} F_\alpha(p, n-p) \right\}$$

$$\boxed{CR_{1-\alpha}(\mu) = \mathcal{E}^\alpha_{S^{-1}}(\bar{X})}$$

---

**Obs.**

$$\frac{(n-1)p}{n-p} F_\alpha(p, n-p) \rightarrow \chi^2_\alpha(p) \quad \text{as } n \rightarrow \infty$$

$\frac{(n-1)}{n-p} \rightarrow 1$ $\qquad$ $F_\alpha(p, n-p) \rightarrow \frac{1}{p}\chi^2(p)$

What if $n$ is large and the sample is gaussian? The two things are the same!

---

**Test** $\quad H_0: \mu = \mu_0 \quad$ vs $\quad H_1: \mu \neq \mu_0$

at level $\alpha \in (0,1)$

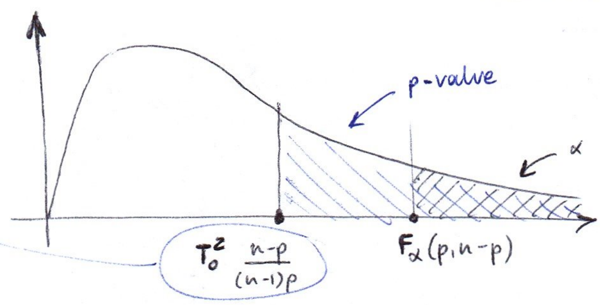
$$T_0^2 = n(\bar{X}-\mu_0)^T S^{-1}(\bar{X}-\mu_0).$$

$T^2$ statistic when $H_0$ is true

If $H_0$ is true $\Rightarrow$ $T_0^2 \sim \frac{(n-1)p}{n-p} F(p, n-p)$

$$\boxed{\text{Reject if } T_0^2 > \frac{(n-1)p}{n-p} F_\alpha(p, n-p)}$$

Be careful with the p-value!




we're comparing this quantity with the quantile, not $T_0^2$

---

**Goal:** understand how these regions change following the correlation between variables

**Ex.** $X_1, \dots, X_{10}$ iid $\sim N_2(\mu, \Sigma)$ $n=10$

and suppose $\bar{X} = 0$ and $S = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

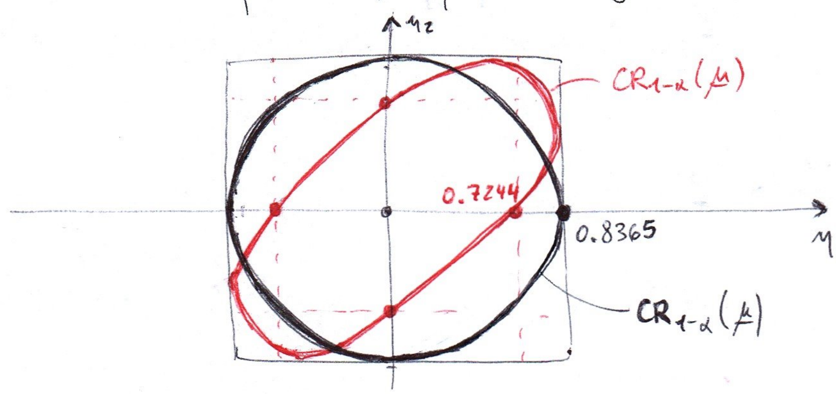
$CR_{1-\alpha}(\mu) = \left\{ \eta \in \mathbb{R}^p : 10\, \eta^T I \eta \le \frac{9 \cdot 2}{8} F_\alpha(2,8) \right\}$

$\alpha = 0.1 \quad 1-\alpha = 0.9 \qquad \eta = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}$

$F_{0.1}(2,8) = 3.11$

$\underline{CR_{0.9}(\mu)} = \left\{ \eta \in \mathbb{R}^p : 10(\eta_1^2 + \eta_2^2) \le 6.997 \right\}$

$= \left\{ \eta \in \mathbb{R}^p : \eta_1^2 + \eta_2^2 \le 0.6997 \right\}$ circle centered in 0




Assume now: $\bar{X} = 0 \quad S = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ $\longrightarrow$ now $X_1$ and $X_2$ are more correlated

$CR_{0.9}(\mu) = \left\{ \eta \in \mathbb{R}^2 : 10\, \eta^T S^{-1} \eta \le \frac{9 \cdot 2}{8} \cdot 3.11 \right\}$.

$S^{-1} = \begin{bmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{bmatrix}$

$10\, \eta^T S^{-1} \eta = 10 \cdot \frac{4}{3} (\eta_1^2 - \eta_1\eta_2 + \eta_2^2)$

$\underline{CR_{0.9}(\mu)} = \left\{ \eta \in \mathbb{R}^2 : \eta_1^2 - \eta_1\eta_2 + \eta_2^2 \le \frac{3}{4}(0.6997) \right\}$

from the drawing we can see that the more the two components are correlated, the more the ellipse is squished (schiacciata)

---

**Goal 2:** how to avoid a statistical temptation

$\bar{X} = \begin{pmatrix} \bar{X}_1 \\ \bar{X}_2 \end{pmatrix} \qquad S = \begin{bmatrix} S_{11} = 1 & S_{12} \\ S_{21} & S_{22} = 1 \end{bmatrix}$

singular CI

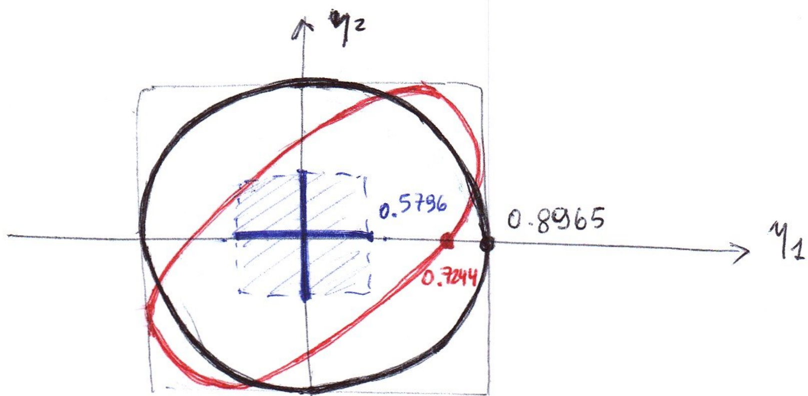
$$CI_{0.9}(\mu_1) = \left[ \bar{x}_1 \pm t_{0.95}(9)\sqrt{\frac{1}{10}} \right]$$ (from stat 101)

$$= [\pm .5796]$$

$$CI_{0.9}(\mu_2) = [\pm .5796]$$

Same CI's for $\underline{S} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\underline{S} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$.

$\Longrightarrow$




So using the tools of Stat 101 (Univariate statistics) we know with confidence 95% that $\mu_1$ will be in — and $\mu_2$ in I.

So the temptation will be to consider as a confident region the cartesian product of the two single intervals (////).

But:

$$\mathbb{P}(\mu_1 \in CI_{0.9}(\mu_1),\ \mu_2 \in CI_{0.9}(\mu_2)) = \underbrace{(0.9)^2}_{< 0.9} \quad \text{if } \bar{X}_1 \perp\!\!\!\perp \bar{X}_2$$

$\Rightarrow$ the cartesian product of the two intervals cannot generate a region that has the same confidence

Lecture of 31.03.20

martedì 31 marzo 2020 11:59

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

For large $n$: $\sqrt{n}(\bar{X}-\mu) \sim N_p(0,\Sigma)$ (CLT)

$\Rightarrow \bar{X}_n \sim N_p(\mu, \frac{1}{n}\Sigma)$

they're equivalent

Mahalanobis distance

$d^2_{(\frac{1}{n}\Sigma)^{-1}}(\bar{X},\mu) = (\bar{X}-\mu)^T(\frac{1}{n}\Sigma)^{-1}(\bar{X}-\mu) = n(\bar{X}-\mu)^T\Sigma^{-1}(\bar{X}-\mu)$

---

**Confidence Intervals** for linear combination of the mean.

(**CONFIDENCE INTERVALS** for linear combination of the mean)

We're in the small $n$ case!

$X_1, \dots, X_n$ iid $\sim N_p(\mu,\Sigma)$

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ estimator for $\mu$.

**Question**: (**Goal**): take a linear combination of the mean $\mu$ and find an estimator for that and also quantify the uncertainty about these estimates

---

**1.** One-at-the-time $CI(\mu)$ = one linear comb.

Let $a \in \mathbb{R}^p$, $\underline{CI_{1-\alpha}(a^T\mu)}$ ?

obvious estimator $\rightarrow$ Estimator for $a^T\mu$: $a^T\bar{X}$

$\mathbb{R} \ni a^T\bar{X} \sim N_1(a^T\mu, \frac{1}{n}a^T\Sigma a)$.

$\Rightarrow$ $\boxed{\frac{a^T\bar{X} - a^T\mu}{\sqrt{a^T\Sigma a}}\sqrt{n} \sim N_1(0,1)}$

can we use it to make inference on $a^T\mu$? **No**, because we don't know $\Sigma$. we need to estimate $\Sigma$.

estimator for $\Sigma$ $\Rightarrow (n-1)S \sim \text{Wishart}(\Sigma, n-1)$

$\Rightarrow$ $\boxed{(n-1)a^TSa \sim (a^T\Sigma a)\chi^2(n-1)}$

$\perp\!\!\!\perp$

Moreover: $\bar{X} \perp\!\!\!\perp S$.

Hence

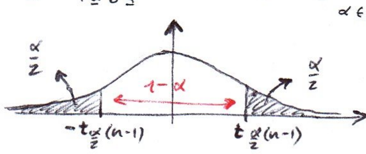
$$\frac{\frac{a^T\bar{X} - a^T\mu}{\sqrt{a^T\Sigma a}}\sqrt{n} \sim N(0,1)}{\sqrt{\frac{\cancel{(n-1)}a^TSa}{\cancel{(n-1)}(a^T\Sigma a)}} \sim \sqrt{\frac{\chi^2(n-1)}{n-1}}} \sim t(n-1)$$

we can use this to compute $CI(\mu)$ since now we have a random quantity which depends only on the param. we want to estimate (and the distr. does not depend on that parameter

$\rightarrow$ PIVOTAL QUANTITY)

$$\boxed{\frac{a^T\bar{X} - a^T\mu}{\sqrt{a^TSa}}\sqrt{n} \sim t(n-1)}$$

$\mathbb{P}\left[\frac{|a^T\bar{X} - a^T\mu|}{\sqrt{a^TSa}}\sqrt{n} < t_{\frac{\alpha}{2}}(n-1)\right] = 1-\alpha$ $\quad \alpha \in (0,1)$




$\Rightarrow \mathbb{P}\left[a^T\mu \in \left[a^T\bar{X} \pm t_{\frac{\alpha}{2}}(n-1)\sqrt{\frac{a^TSa}{n}}\right]\right] = 1-\alpha$

$\Rightarrow$ $\boxed{CI_{1-\alpha}(a^T\mu) = \left[a^T\bar{X} \pm t_{\frac{\alpha}{2}}(n-1)\sqrt{\frac{a^TSa}{n}}\right]}$

Ex. • $a = (0 - 0\ 1\ 0 - 0)^T$ ($1$ in position $i$), $i = 1 \dots p$

$a^T\mu = \mu_i$

$CI_{1-\alpha}(\mu_i) = \left[\bar{X}_i \pm t_{\frac{\alpha}{2}}(n-1)\sqrt{\frac{S_{ii}}{n}}\right]$

• $a = (0 - 0\ 1\ 0 - 0\ \text{-}1\ 0 - 0)^T$ ($1$ in position $i$, $-1$ in position $j$)

$a^T\mu = \mu_i - \mu_j$

$CI_{1-\alpha}(\mu_i - \mu_j) = \left[\bar{X}_i - \bar{X}_j \pm t_{\frac{\alpha}{2}}(n-1)\sqrt{\frac{S_{ii} - 2S_{ij} + S_{jj}}{n}}\right]$

If our problem is not IC but testing:

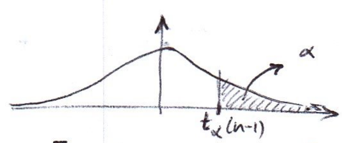
Testing:

$H_0: a^T\mu = \delta_0$ vs $H_1: a^T\mu > \delta_0$

If $H_0$ is true:

$\frac{a^T\bar{X} - \delta_0}{\sqrt{a^TSa}}\sqrt{n} \sim t(n-1)$

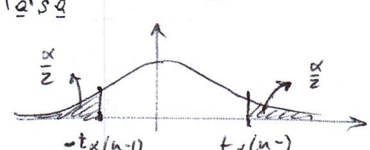
Reject if at level $\alpha \in (0,1)$ $\frac{a^T\bar{X} - \delta_0}{\sqrt{a^TSa}}\sqrt{n} > t_\alpha(n-1)$




$H_0: a^T\mu = \delta_0$ vs $H_1: a^T\mu \neq \delta_0$

Reject at level $\alpha \in (0,1)$:

$\frac{|a^T\bar{X} - \delta_0|}{\sqrt{a^TSa}}\sqrt{n} > t_{\frac{\alpha}{2}}(n-1)$




IC and testing are naturally related

We proved :

$$\boxed{\mathbb{P}\left[\underline{a}^T\underline{\mu} \in \underbrace{\left[\underline{a}^T\bar{\underline{X}} \pm t_{\frac{\alpha}{2}}(n-1)\sqrt{\frac{\underline{a}^T S\,\underline{a}}{n}}\right]}_{CI_{1-\alpha}(\underline{a}^T\underline{\mu})\ \text{one-at-the-time}}\right] = 1-\alpha, \quad \forall \underline{a} \in \mathbb{R}^p}$$

there is a huge difference to put $\forall \underline{a} \in \mathbb{R}^p$ inside or outside the prob.

Did we prove :

$$\cancel{\mathbb{P}\left[\underline{a}^T\underline{\mu} \in \left[\underline{a}^T\bar{\underline{X}} + t_{\frac{\alpha}{2}}(n-1)\sqrt{\frac{\underline{a}^T S\,\underline{a}}{n}}\right], \ \forall \underline{a} \in \mathbb{R}^p\right] = 1-\alpha} \quad ? \quad \text{NO.}$$

Did we prove that simultaneously taken, all these confidence intervals together cover all possible linear combinations of $\mu$ with prob. $1-\alpha$? **NO.**

**TOSSING A COIN**

$$X_i = \begin{cases} 0 & 1/2 \\ 1 & 1/2 \end{cases}$$

$X_1, \dots, X_n \overset{iid}{\sim} Be\left(\frac{1}{2}\right)$

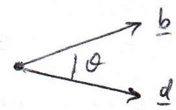
1. $\mathbb{P}(X_i = 1) = \frac{1}{2} \quad \forall i = 1, \dots, 100$
2. $\mathbb{P}(X_i = 1 \ \forall i = 1, \dots, 100) = \left(\frac{1}{2}\right)^{100}$

## 2. Simultaneous CI for linear comb. of $\mu$

$$\mathbb{P}\left[\underline{a}^T\underline{\mu} \in \left[\underline{a}^T\bar{\underline{X}} \pm (?)\sqrt{\frac{\underline{a}^T S\,\underline{a}}{n}}\right], \ \forall \underline{a} \in \mathbb{R}^p\right] = 1-\alpha$$

**A little excursus in Algebra**

$\underline{b}, \underline{d} \in \mathbb{R}^p$




$$\frac{\langle \underline{b}, \underline{d} \rangle}{\|\underline{b}\| \, \|\underline{d}\|} = \cos(\theta)$$

$$\frac{(\langle \underline{b}, \underline{d} \rangle)^2}{\|\underline{b}\|^2 \|\underline{d}\|^2} \le 1$$

equality holds if $\underline{b} \in \text{span}(\underline{d})$ i.e. $\underline{b} \propto \underline{d}$

Hence :

$$\boxed{(\underline{b}^T\underline{d})^2 \le \left(\sum b_i^2\right)\left(\sum d_i^2\right)} \quad \text{C-S inequality}$$

and eq. holds if $\underline{b} \in \text{span}(\underline{d})$

$$\int fg \le \left(\int f^2\right)\left(\int g^2\right)$$

Take $B$ $p \times p$ matrix pos. def.

$$\Rightarrow \boxed{(\underline{b}^T\underline{d})^2 \le (\underline{b}^T B\, \underline{b})(\underline{d}^T B^{-1} \underline{d})} \quad \text{Extended C-S}$$

proof.

$$\underline{b}^T\underline{d} = \underline{b}^T B^{1/2} B^{-1/2}\underline{d} \qquad (\text{we know } \exists B^{-1/2} \text{ since it's positive def.})$$

$$(\underline{b}^T\underline{d})^2 = \left(\underline{b}^T B^{1/2} \, B^{-1/2}\underline{d}\right)^2$$

$$\overset{C\text{-}S}{\le} (\underline{b}^T B^{1/2} B^{1/2} \underline{b})(\underline{d}^T B^{-1/2} B^{-1/2} \underline{d})$$

$$= (\underline{b}^T B\, \underline{b})(\underline{d}^T B^{-1} \underline{d})$$

and equality holds if $\underline{b}^T B^{1/2} \in \text{span}(B^{-1/2}\underline{d})$

$\underline{b} \in \text{span}(B^{-1}\underline{d})$ ∎

Max lemma: Let $B$ $p \times p$ pos. def., $\underline{d} \in \mathbb{R}^p$

$$\boxed{\max_{\underline{x} \in \mathbb{R}^p, \, \underline{x} \neq \underline{0}} \frac{(\underline{x}^T\underline{d})^2}{\underline{x}^T B\, \underline{x}} = \underline{d}^T B^{-1} \underline{d}}$$

proof.

$$(\underline{x}^T\underline{d})^2 \le (\underline{x}^T B\, \underline{x})(\underline{d}^T B^{-1} \underline{d})$$

if $\underline{x} \neq \underline{0} \Rightarrow \underline{x}^T B\, \underline{x} > 0$ (pos. def.)

$$\frac{(\underline{x}^T\underline{d})^2}{\underline{x}^T B\, \underline{x}} \le \underline{d}^T B^{-1} \underline{d} \quad \forall \underline{x} \neq \underline{0}$$

equality holds if $\underline{x} \in \text{span}(B^{-1}\underline{d})$ ∎

**Back to Stat** :

$\underline{X}_1 \dots \underline{X}_n$ iid $\sim N_p(\underline{\mu}, \Sigma)$

$$\bar{\underline{X}} = \frac{1}{n}\sum_{i=1}^{n} \underline{X}_i$$

let $\underline{a} \in \mathbb{R}^p$

$$\frac{\left(\underline{a}^T(\bar{\underline{X}} - \underline{\mu})\right)^2}{\underline{a}^T S\, \underline{a}} \, n$$

By Max Lemma:

$$\max_{\underline{a} \in \mathbb{R}^p} \frac{(\underline{a}^T(\bar{\underline{x}} - \underline{\mu}))^2}{\underline{a}^T S \underline{a}} n = n(\bar{\underline{x}} - \underline{\mu})^T S^{-1} (\bar{\underline{x}} - \underline{\mu})$$

$$\underbrace{n(\bar{\underline{x}} - \underline{\mu})^T S^{-1} (\bar{\underline{x}} - \underline{\mu})}_{:= T^2}$$

$$T^2 \sim \frac{(n-1)p}{n-p} F(p, n-p)$$ (we know it by Hotelling)

back to the starting problem

$$\Longrightarrow \mathbb{P}\left[ \frac{|\underline{a}^T(\bar{\underline{x}} - \underline{\mu})|}{\sqrt{\underline{a}^T S \underline{a}}} \sqrt{n} \le c, \ \forall \underline{a} \in \mathbb{R}^p, \underline{a} \ne \underline{0} \right] = 1-\alpha$$

(We want to find $c$. We already know that the $t$ quantile is wrong)

$$\Longleftrightarrow \mathbb{P}\left[ \max_{\substack{\underline{a} \in \mathbb{R}^p \\ \underline{a} \ne \underline{0}}} \frac{|\underline{a}^T(\bar{\underline{x}} - \underline{\mu})|}{\sqrt{\underline{a}^T S \underline{a}}} \sqrt{n} \le c \right] = 1-\alpha$$

asking everybody to be less than $c$ means ask the longest to be lest than $c$

$$\Longleftrightarrow \mathbb{P}\left[ \boxed{\max_{\substack{\underline{a} \in \mathbb{R}^p \\ \underline{a} \ne \underline{0}}} \frac{(\underline{a}^T(\bar{\underline{x}} - \underline{\mu}))^2}{\underline{a}^T S \underline{a}} n} \le c^2 \right] = 1-\alpha$$

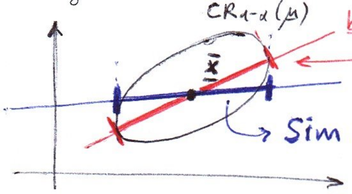
$$\sim \frac{(n-1)p}{n-p} F(p, n-p)$$

$$\Longrightarrow c^2 = \frac{(n-1)p}{n-p} F_\alpha(p, n-p)$$

$$\Longrightarrow \boxed{\mathbb{P}\left[ \frac{|\underline{a}^T(\bar{\underline{x}} - \underline{\mu})|}{\sqrt{\underline{a}^T S \underline{a}}} \sqrt{n} \le \sqrt{\frac{(n-1)p}{n-p} F_\alpha(p, n-p)}, \ \forall \underline{a} \in \mathbb{R}^p \right] = 1-\alpha}$$

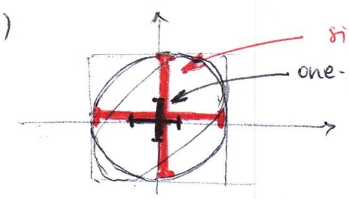
$$\boxed{\underline{\text{Sim}}\, CI_{1-\alpha}(\underline{a}^T\underline{\mu}) = \left[ \underline{a}^T \bar{\underline{x}} \pm \sqrt{\frac{(n-1)p}{n-p} F_\alpha(p, n-p)} \sqrt{\frac{\underline{a}^T S \underline{a}}{n}} \right]}$$

Geometry:




: projection of $CR_{1-\alpha}(\underline{\mu})$ on the given direction $\underline{a}$

<u>Ex.</u> (yesterday's)




= projection of the ellipse on the direction for which we want the confidence interval

$\Longrightarrow$ Sim $CI_{0.9}(\mu_1) = [\pm 0.83\ldots] =$ Sim$_{0.9}$ $CI(\mu_2)$

---

**Bonferroni's** method for simult. CI for a <u>finite</u> number of lin. comb. of $\underline{\mu}$.

We're not testing on one single mean, we're testing on more means

**Problem:** we fix k directions

Given $\underline{a}_1, \ldots, \underline{a}_k \in \mathbb{R}^p$
find $CI(\underline{a}_1^T\underline{\mu}), \ldots, CI(\underline{a}_k^T\underline{\mu})$
with a simultaneous confidence of $1-\alpha$
$\alpha \in (0,1)$

Let $CI_{1-\alpha}(\underline{a}^T\underline{\mu})$ is a one-at-the-time conf. interv. for $\underline{a}^T\underline{\mu}$ of level $1-\alpha$

E.g. $CI_{1-\alpha}(\underline{a}^T\underline{\mu}) = \left[ \underline{a}^T \bar{\underline{x}} \pm t_{\frac{\alpha}{2}}(n-1) \sqrt{\frac{\underline{a}^T S \underline{a}}{n}} \right]$

$$\mathbb{P}\left[ \bigcap_{i=1}^{k} \{ \underline{a}_i^T \underline{\mu} \in CI_{1-\alpha}(\underline{a}_i^T\underline{\mu}) \} \right] =$$

$$= 1 - \mathbb{P}\left[ \bigcup_{i=1}^{k} \{ \underline{a}_i^T \underline{\mu} \notin CI_{1-\alpha}(\underline{a}_i^T\underline{\mu}) \} \right] \qquad (*)$$

$\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$
$$\leq \mathbb{P}(A) + \mathbb{P}(B)$$
$$(*) \geq 1 - \sum_{i=1}^{k} \underbrace{\mathbb{P}\left[ \underline{a}_i^T \mu \notin CI_{1-\alpha}(\underline{a}_i^T \mu) \right]}$$
$$= 1 - k\alpha$$

↳ what is the prob. that $CI_{1-\alpha}$ does not cover the parameter that is estimating? $\alpha$

Hence
$$\mathbb{P}\left[ \bigcap_{i=1}^{k} \{ \underline{a}_i^T \mu \in CI_{1-\frac{\alpha}{k}}(\underline{a}^T \mu) \} \right] \geq 1-\alpha$$

Bonferroni Sim. Conf Int for $\underline{a}_1^T \mu, \dots, \underline{a}_k^T \mu$

$$\boxed{\text{Bonf } CI_{1-\alpha}(\underline{a}_i^T \mu) = \left[ \underline{a}_i^T \bar{\underline{x}} \pm t_{\frac{\alpha}{2k}}(n-1) \sqrt{\frac{\underline{a}^T S \underline{a}}{n}} \right]}$$

(Yesterday's) Ex. $\text{Bonf } CI_{0.9}(\mu_i) = [\pm 0.7153]$ $i = 1,2$

**NOTE**: if $k \to \infty$ (or $k$ too big) then the quantile goes to $+\infty$ so the interval becomes $-\infty, +\infty$ and so we cannot use this to extend to all possible linear combinations (only a small number of them)

---

Testing. (simult. Bonferroni way).

$H_0: \underline{a}_1^T \mu = \delta_1, \; \vdots \; \underline{a}_k^T \mu = \delta_k$ vs $H_1$: At least for one $i$ $\underline{a}_i^T \mu \neq \delta_i$

Reject at level $\alpha \in (0,1)$ if for at least one $i$:
$$\frac{|\underline{a}_i^T \bar{\underline{x}} - \delta_i|}{\sqrt{\underline{a}_i^T S \underline{a}_i}} \sqrt{n} > t_{\frac{\alpha}{2k}}(n-1)$$

In fact: if $H_0$ is true
$$\mathbb{P}\left[ \text{reject at least one hyp } \underline{a}_i^T \mu = \delta_i \,\middle|\, \text{all of them are true} \right]$$
$$= \mathbb{P}\left[ \bigcup_{i=1}^{k} \left\{ \frac{|\underline{a}_i^T \bar{\underline{x}} - \delta_i|}{\sqrt{\underline{a}_i^T S \underline{a}_i}} > t_{\frac{\alpha}{2k}}(n-1) \right\} \,\middle|\, H_0 \right]$$
$$\leq \sum_{i=1}^{k} \mathbb{P}\left[ \frac{|\underline{a}^T \bar{\underline{x}} - \delta_i|}{\sqrt{\underline{a}_i' S' \underline{a}_i}} \sqrt{n} > t_{\frac{\alpha}{2k}}(n-1) \,\middle|\, H_0 \right]$$
$$= \sum_{i=1}^{k} \frac{\alpha}{k} = \alpha$$

= probability of rejecting at least one of the assumptions when they're all true is $\alpha$

**NOTE**: Bonferroni's way of simultaneous testing is very conservative (very strict), with BIG DATA we're not going to use this.
(Because of the NOTE above)

Lecture of 02.04.20

giovedì 2 aprile 2020 11:25

## Goal

we want to test $k$ linear assumptions ($k$ linear hypothesis) on the mean $\mu$ (unknown) of this distribution

$X_1, \dots, X_n$ iid $\sim N_p(\mu, \Sigma)$ $\det(\Sigma) > 0$

Given $a_1, \dots, a_k$ in $\mathbb{R}^p$

$H_{0i}: a_i^T \mu = \mu_{0i}$ $i = 1 \dots k$ $\mu_{0i} \in \mathbb{R}$

vs

$H_{1i}: a_i^T \mu \neq \mu_{0i}$ $i = 1 \dots k$

**Bonferroni's** strategy:

Reject at level $\alpha \in (0,1)$ $H_{0i}$ if:

$$\boxed{\frac{|a_i^T \bar{X} - \mu_{0i}|}{\sqrt{a_i^T S a_i}} \sqrt{n} > t_{\frac{\alpha}{2k}}(n-1)}$$

If $\bigcap_{i=1}^{k} \{H_{0i} \text{ is true}\}$ then

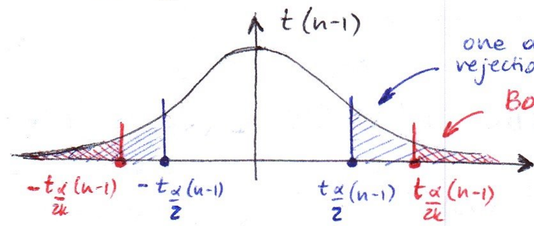
$$\mathbb{P}[\text{rejecting at least one } H_{0i} \mid \cap H_{0i} \text{ true}]$$

$$= \mathbb{P}\Big[\bigcup_{i=1} \{\text{reject } H_{0i}\} \mid \cap H_{0i} \text{ true}\Big]$$

$$\leq \sum_{i=1}^{k} \boxed{\mathbb{P}[\text{reject } H_{0i} \mid H_{0i} \text{ is true}]}$$ since we're taking $t_{\frac{\alpha}{2k}} \Rightarrow \frac{\alpha}{k}$

$$\leq \sum_{i=1}^{k} \frac{\alpha}{k} = \alpha.$$

**Problem:** the more simultaneous tests we make, the more conservative the procedure becomes (because we're taking quantiles that are closer and closer to $\infty$)




The problem is not the procedure of the testing, the point is: because we're testing $k$ hypothesis all together (simultaneously) and we want to have an overall $\alpha$ level for the entire procedure, then we need for each singular test to take a level that is way smaller.

(if we're making 1000 hypothesis simultaneously and we want a level $\alpha$ of the overall procedure, we have to take each trashold for each test equal to $t_{\frac{\alpha}{1000}}$, means that we're saying that the trasholds are $-\infty$ and $+\infty$ so we never reject $H_{0i}$ (so all the $H_{0i}$ hp. are true simultaneously), but when we're doing a test we try to reject $H_0$)

## FALSE DISCOVERY RATE (FDR) : Benjamin-Hochberg (1995)

2014: 21'000 citations

2019: 53'000 "

2020: 63'000 "

Suppose we have $k$ test to be performed:

$k$ test simultaneous:

$H_{0i}$ vs $H_{1i}$ $i = 1 \dots k$

Let $\mathcal{D}$ be a strategy for testing:

Errors of strat. $\mathcal{D}$

| | not-rejecting $H_0$ | reject $H_0$ | |
|---|---|---|---|
| Truth: $H_0$ | $U$ | $V$ false positive | $k_0$ # true hypothesis |
| Truth: $H_1$ | $T$ false negative | $S$ | $k - k_0$ # false hypothesis |
| | $k - R$ # not rejected | $R$ # rejected | $k$ # testing |

only observable things ($U, V, T, S, k_0$ are not)

Suppose $\mathcal{D}$ is Bonferroni's strategy:

$I_0$: the set of true $H_{0i}$'s, $|I_0| = k_0$

what Bonferroni does is:

Given $\alpha \in (0,1)$ $\Rightarrow$ reject $H_{0i}$ at level $\frac{\alpha}{k}$

$$\mathbb{P}[V \geq 1] = \mathbb{P}\Big[\bigcup_{j \in I_0} \{H_{0j} \text{ is rejected}\}\Big]$$

$$\leq \sum_{j \in I_0} \mathbb{P}[H_{0j} \text{ is rejected}]$$

$$\leq \sum_{j \in I_0} \frac{\alpha}{k} = k_0 \frac{\alpha}{k} \leq k \frac{\alpha}{k} = \alpha$$

So if we use BONFERRONI we are sure that the probability that we will reject one or more of the true hypothesis is controlled by $\alpha$.

This is called FAMILY WISE ERROR RATE (FWER)

$$\boxed{FWER = \mathbb{P}(V \geq 1)}$$

With Bonf. one is controlling FWER.

A different strategy: control

| | | NOT reject $H_0$ | reject $H_0$ |
|---|---|---|---|
| Truth: | $H_0$ | U | V |
| | $H_1$ | T | S |

V → FALSE DISCOVERIES

S → TRUE DISCOVERIES

$\boxed{\mathbb{E}\left[\frac{V}{R}\right]}$ = FALSE DISCOVERY RATE (FDR)

(convention: if $R=0 \Rightarrow \frac{V}{R} := 0$)

$$Q := \begin{cases} 0 & \text{if } R=0 \\ \frac{V}{R} & \text{if } R>0 \end{cases}$$

$$FDR = \mathbb{E}[Q]. = \mathbb{E}\left[\frac{\text{false discoveries}}{\text{total discoveries}}\right]$$

Obs. Assume $\boxed{k_0 = k}$ (no discoveries to be made). : all $H_{0_i}$ are true

$\Rightarrow S=0$ and $V=R$

$$Q := \begin{cases} 0 & \text{if } R=0 \\ 1 & \text{if } R>0 \end{cases}$$

$$\underbrace{\mathbb{E}[Q]}_{FDR} = \mathbb{P}[R>0] = \mathbb{P}[V>0] = \mathbb{P}[V \geq 1] = \text{FWER}$$

Obs. Assume $\boxed{k_0 < k}$

$V=0 \Rightarrow Q=0$

$V>0 \Rightarrow R>0 \Rightarrow V/R \leq 1$

$$FDR = \mathbb{E}[Q] \leq \mathbb{E}[\mathbb{1}_{\{V>0\}}] = \mathbb{P}(V>0) = \mathbb{P}(V \geq 1) = \text{FWER}$$

$$\mathbb{1}_{\{V>0\}} := \begin{cases} 0 & \text{if } V=0 \Rightarrow Q=0 \\ 1 & \text{if } V>0 \Rightarrow Q \leq 1 \end{cases}$$

We therefore proved that:

$$\boxed{FDR \leq FWER}$$

Controlling FWER with Bonferroni is too conservative, maybe we can control FDR without being so conservative as Bonferroni (NOTE: Since Bonferroni control FWER then it also control FDR)

Let's consider a procedure s.t. $FDR \leq \alpha$. (No longer FWER < $\alpha$, we relax a bit the conditions.)

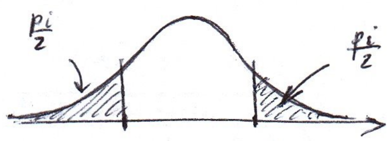
---

## B & H STRATEGY FOR CONTROLLING FDR

Benjamini & Hochberg

Let $p_i$ be the p-value of $H_{0i}$ vs $H_{1i}$

Ex.




$$t_i = \frac{|a_i^T \bar{x} - \mu_{0i}|}{\sqrt{a_i^T S a_i}}$$

So for each of the k test compute the p-value and then Order the p-values:

$$p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(k)}.$$

Let $\alpha \in (0,1)$ and compute

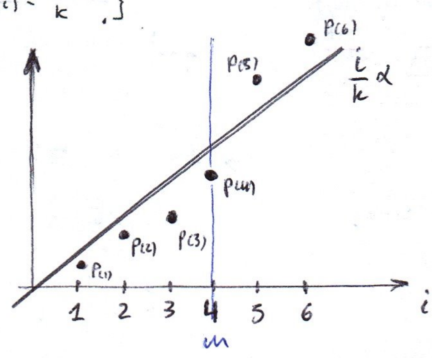
$$m = \max\left\{ i \in \{1 \dots k\} : p_{(i)} \leq \frac{i}{k} \cdot \alpha \right\}$$

We compute this trashold $m$ and then we reject all the hypothesis with index smaller than $m$

Ex.

$m = 4$ ;

Reject: $H_{0(1)}, H_{0(2)}, \dots, H_{0(m)}$




In the example we reject: $H_{0(1)}, H_{0(2)}, H_{0(3)}, H_{0(4)}$

($\neq H_{01}, H_{02}, H_{03}, H_{04}$) !

**Teo** (B&H) (1995)

If $p_1, \dots, p_k$ are independent
$\Rightarrow$ **FDR** $\le \alpha$.

By following the precedent procedure, if we have independent test, we're controlling **FDR**.

**In conclusion:**
we're not controlling FWER but if we're willing to be more relaxed about FWER then we'll have other procedures that keep things under control and allow to reject $H_0$ more often.

**Teo** (Benjamini & Yekutieli, 2001)

If $p_1, \dots, p_k$ are $\boxed{\text{pos. correl.}}$
$\Rightarrow$ reject $H_{0(i)}$ if $i \le m = \max\{j : p_{(j)} \le \frac{j\alpha}{k}\}$

If $p_1, \dots, p_k$ are $\boxed{\text{neg. correl.}}$
$\Rightarrow$ reject $H_{0(i)}$ if
$$i \le m^* = \max\{j : p_{(j)} \le \frac{j\alpha}{k\,C(k)}\}$$
$$C(k) = \sum_{j=1}^{k} \frac{1}{j}$$

Now there is correlation among the tests: first positive correlation, then negative

$$m = \max\{j : p_{(j)} \le \frac{j\alpha}{k}\}$$
$$m^* = \max\{j : p_{(j)} \le \frac{j\alpha}{k\,C(k)}\}$$

---

## COMPARING MEANS OF GAUSSIAN DISTRIBUTIONS

- **Paired data**

$n$ stat. units
each unit is observed 2 times:

$$\underline{X}_{1i} = \begin{bmatrix} x_{1i1} \\ x_{1i2} \\ \vdots \\ x_{1ip} \end{bmatrix} \qquad \underline{X}_{2i} = \begin{bmatrix} x_{2i1} \\ x_{2i2} \\ \vdots \\ x_{2ip} \end{bmatrix} \qquad i = 1, \dots, n$$

They're called paired data because every vector is paired with an other vector since both are observed on the same statistical unit.
**NOTE**: the pairing is established before of the measurements. It's not that after we got the data of two population we can pair the two tallest and so on

$\underline{X}_{1i}$ iid with mean $\mu_1$
$\underline{X}_{2i}$ iid with mean $\mu_2$
**Goal**: inference on $\mu_1 - \mu_2$

this is used for instance to check if a treatment have an effect

Consider

$$\underline{D}_i = \underline{X}_{1i} - \underline{X}_{2i} \qquad i = 1 \dots n$$
$$\Rightarrow \mathbb{E}[\underline{D}_i] = \mu_1 - \mu_2$$

**Goal**: inference on $\mathbb{E}[\underline{D}_i]$ $\longrightarrow$ we reduce the problem in one set of vectors

We won't assume anything on the distribution of $\underline{X}$'s (we don't need it). We've assuming (and working) only on $\underline{D}_i$

Assume $\underline{D}_1, \dots, \underline{D}_n$ iid $\sim N_p(\underline{\delta}, \Sigma_D)$
$\underline{\delta} = \mu_1 - \mu_2$

Let
$$\bar{\underline{D}} = \frac{1}{n}\sum_{i=1}^{n} \underline{D}_i$$
$$S_D = \frac{1}{n-1}\sum_{i=1}^{n} (\underline{D}_i - \bar{\underline{D}})(\underline{D}_i - \bar{\underline{D}})^T$$

$$\Longrightarrow \boxed{n(\bar{\underline{D}} - \underline{\delta})^T S_D^{-1} (\bar{\underline{D}} - \underline{\delta}) \sim \frac{(n-1)p}{n-p} F(p, n-p)}$$

**PIVOTAL QUANTITY** $\longrightarrow$ once we have this we can create everything (IC, testing, ...)

**Conf. Region** of level $1-\alpha$, $\alpha \in (0,1)$:

$$CR_{1-\alpha}(\mu_1 - \mu_2) = \{\underline{\delta} \in \mathbb{R}^p : n(\bar{\underline{D}} - \underline{\delta})^T S_D^{-1}(\bar{\underline{D}} - \underline{\delta}) \le \frac{(n-1)p}{n-p} F_\alpha(p, n-p)\}$$

**Test** $H_0: \mu_1 - \mu_2 = \underline{\delta}_0$ vs $H_1: \mu_1 - \mu_2 \ne \underline{\delta}_0$
Reject at level $\alpha \in (0,1)$ if

$$n(\bar{\underline{D}} - \underline{\delta}_0)^T S_D^{-1}(\bar{\underline{D}} - \underline{\delta}_0) > \frac{(n-1)p}{n-p} F_\alpha(p, n-p)$$

**Sim CI$_{1-\alpha}$** $(\mu_{1i} - \mu_{2i}) \qquad i = 1 \dots p$

if the treatment made a difference, we want to know in which factor it made the difference

$$= \left[ \bar{D}_i \pm \sqrt{\frac{(n-1)p}{n-p} F_\alpha(p, n-p)} \sqrt{\frac{S_{D\,ii}}{n}} \right]$$

Bonferroni $CI_{1-\alpha}$ $i = 1 \dots p$

$$\left[ \bar{D}_i \pm t_{\frac{\alpha}{2p}}(n-1) \sqrt{\frac{S_{D\,ii}}{n}} \right]$$

- **Repeated univariate measures** : each statistical unit is observed $q$ times (in the previous case $q=2$)

$\underline{X}_i = (X_{i1}, \dots, X_{iq})^T$ $i = 1 \dots n$

$X_{ij}$ the measure of the same quantity in instance $j$ for unit $i$ $i=1\dots n$, $j = 1 \dots q$

For example $\underline{X}_1$ = blood pressure
$\underline{X}_2$ = example 2

$\underline{X}_1 = (X_{11}, \dots, X_{1q})$ : $X_{11}$ = b.p. at moment $t=1$, $X_{1q}$ = b.p. at moment $t=q$

Let $\mu = E[\underline{X}_i]$ $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_q \end{pmatrix}$

Goal: $H_0: \mu_1 = \mu_2 = \dots = \mu_q$ vs $H_1: \exists i,j$ s.t. $\mu_i \neq \mu_j$

**NOTE**: $q$ can also be space, for example: measure the purity of iron in a lama

$\bar{X}$ estimator for $\mu$

Contrast matrix : $C$ $(q-1)\times q$ s.t.

1) the rows are lin. indep.
2) $C\underline{1} = \underline{0}$

$$C = \begin{bmatrix} \underline{c}_1^T \\ \underline{c}_2^T \\ \vdots \\ \underline{c}_{q-1}^T \end{bmatrix} \quad \underline{c}_i \in \mathbb{R}^q$$

$C$ is contrast if $\underline{c}_1, \dots, \underline{c}_{q-1}$ are independent and $\underline{c}_i^T \underline{1} = 0$ for $i = 1 \dots q-1$

$(\equiv \underline{c}_i^T \perp \underline{1}$ for $i = 1 \dots q-1)$

$(\equiv \text{span}(\underline{c}_1, \dots, \underline{c}_{q-1}) = \mathcal{L}^{\perp}(\underline{1}))$

Ex.
$$C = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & \cdots & 1 & -1 \end{bmatrix} \qquad C = \begin{bmatrix} 1 & 0 & \cdots & 0 & -1 \\ 0 & 1 & 0 \cdots & 0 & -1 \\ \vdots & & & & \\ 0 & \cdots & & 1 & -1 \end{bmatrix}$$

If we apply the first we're taking $\mu_1 - \mu_2, \mu_2 - \mu_3, \dots$. In the second case we have a reference and we want to check $\mu_i - \bar{\mu}$ $i = \dots$ (where $\bar{\mu}$ = reference)

Test: $H_0: \mu_1 = \dots = \mu_q$ vs $H_1: \exists i,j$ s.t. $\mu_i \neq \mu_j$

Rephrase
$H_0: \mu \in \mathcal{L}(\underline{1})$ vs $H_1: \mu \notin \mathcal{L}(\underline{1})$

Let contrast matrix
$H_0: C\mu = \underline{0}$ vs $H_1: C\mu \neq \underline{0}$

If $\mu \in \mathcal{L}(\underline{1}) \Rightarrow C\mu = \underline{0}$ because we're projecting $\mu$ on $\mathcal{L}(\underline{1})^{\perp}$

Consider $C\bar{X}$ estimator of $C\mu$

$$C\bar{X} \sim N_p\left(C\mu, \frac{1}{n} C\Sigma C^T\right)$$ ← we need to estimate $C\Sigma C^T$

$$(n-1)CSC^T \sim Wish(C\Sigma C^T, n-1)$$

$\bar{X} \perp\!\!\!\perp S$

$\Longrightarrow$ Hotelling's Th.

$$\boxed{n(C\bar{X} - C\mu)^T (CSC^T)^{-1} (C\bar{X} - C\mu) \sim \frac{(n-1)(q-1)}{n-q+1} F(q-1, n-q+1)}$$

Reject $H_0$ at level $\alpha \in (0,1)$ if

$$n(C\bar{X} - C\mu)^T (CSC^T)^{-1} (C\bar{X} - C\mu) \geq \frac{(n-1)(q-1)}{n-q+1} F_{\alpha}(q-1, n-q+1)$$

Q: What if we take $C_1$ contrast matrix different from $C$? Does it make a difference?
NO (prove it)

If we take $C$ and $C_1$ the $T^2$ and $T_1^2$ tests are the same.
(Check the answer in the book.)

**NOTE**: The contrast matrix is not given by the problem. It's a way for us to define a basis on a linear space orthogonal to $\underline{1}$ $\Rightarrow$ we can chose it how we want