多元统计分析

第4讲 多元正态分布(2)

Johnson & Wichern Ch5.1-5.5

统计学研究中心 邓婉璐
wanludeng@tsinghua.edu.cn

2018-2019春季学期

Outline

- Inference for a normal population mean
  - Hypothesis test (decision making)
  - Hotelling's T^2 vs LRT
  - Confidence regions and simultaneous comparisons of component means
- Large Sample Inferences about a population mean vector

Example – Scores for Statistical Courses

- Scores for 5 courses: n = 100.
- Interested in
  - Mean, variance
  - Whether the same mean as expected?
  - What is our confidence about the mean vector?
  - What is our confidence about the GPA?
  - What if we have multiple ways for calculating GPA?

```
> head(data)
  Prob Inference Computing MVA LinearReg
1   83        75        88  86        79
2   85        74        90  85        79
3   84        81        90  84        75
4   83        77        91  91        82
5   87        70        89  85        83
6   84        79        91  86        73
```

```
> colMeans(data)
     Prob Inference Computing       MVA LinearReg
    84.98     74.79     89.17     84.97     78.60
```

Test for One-sample Mean

Whether the same mean as expected?

Hotelling's T^2

Suppose we observe p-vectors

$$x_1, \ldots, x_n \overset{iid}{\sim} N(\mu, \Sigma)$$

with likelihood

$$L(\mu, \Sigma; x) = c|\Sigma|^{-n/2} \exp\left\{-\frac{n}{2}\left[\mathrm{tr}\Sigma^{-1}S_n + (\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu)\right]\right\}$$

Then

$$\hat{\mu} = \bar{x} \sim N(\mu, \frac{\Sigma}{n}), \qquad n\hat{\Sigma} = nS_n \sim \mathrm{Wishart}(n-1, \Sigma)$$

S为样本方差, $S_n = \frac{n-1}{n}S$.

with $\hat{\mu} \perp \hat{\Sigma}$.

Hotelling's T^2

Consider the point null hypothesis,

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0,$$

Recall 1-dim:

$$t^2 = \sqrt{n}(\bar{X} - \mu_0)(s^2)^{-1}\sqrt{n}(\bar{X} - \mu_0)$$

$$t_{n-1}^2 = \left(\frac{normal}{r.v.}\right)\left(\frac{(scaled)\chi^2 r.v.}{d.f.}\right)^{-1}\left(\frac{normal}{r.v.}\right)$$

Here s^2 is unbiased estimator for sample variance.

Hotelling's T^2

Consider the point null hypothesis,

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0,$$

For p-dim:

$$T^2 = \boxed{\sqrt{n}(\bar{X} - \mu_0)' \left(\frac{\sum_{j=1}^{n} (X_j - \bar{X})(X_j - \bar{X})'}{n-1} \right)^{-1} \sqrt{n}(\bar{X} - \mu_0)}$$

$$= \left(\begin{array}{c} \text{multivariate normal} \\ \text{random variable} \end{array} \right)' \left(\frac{\text{Wishart random matrix}}{d.f.} \right)^{-1} \left(\begin{array}{c} \text{multivariate normal} \\ \text{random variable} \end{array} \right)$$

Under H_0: $= N_p(0, \Sigma)' [\frac{1}{n-1} W_{p,n-1}(\Sigma)]^{-1} N_p(0, \Sigma)$

Hotelling's T^2

$$\frac{T^2}{n-1} = N(0,\Sigma)' W_{p,n-1}(\Sigma)^{-1} N(0,\Sigma)$$

$$= N(0,I)' W_{p,n-1}(I)^{-1} N(0,I)$$

$$\boxed{\text{Under } H_0 \text{ we have } \quad \frac{n-p}{p}\frac{T^2}{n-1} \sim F_{p,n-p},}$$

and we reject if

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

Example – Scores for Statistical Courses

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

H_{01}: μ = (85,75,89,85,79)'

```
> hotelling_T2
         [,1]
[1,] 3.824517
```

```
> qf(1-alpha,p,n-p)*p*(n-1)/(n-p)
[1] 12.03749
```

$\alpha = 0.05$

Can't reject H_{01}

Reject H_{02}

H_{02}: μ = (85,75,90,85,79)'

```
> hotelling_T2
         [,1]
[1,] 39.52191
```

```
> colMeans(data)
     Prob Inference Computing       MVA LinearReg
    84.98     74.79     89.17     84.97     78.60
```

For comparison

Properties of Hotelling's T²

Invariant to changes in the units of measurements for X:

Suppose: $\underset{(p\times 1)}{Y} = \underset{(p\times p)}{C}\underset{(p\times 1)}{X} + \underset{(p\times 1)}{d}$, C non singular

Then, $\bar{y} = C\bar{x} + d$ $\quad S_y = \frac{1}{n-1}\sum_{j=1}^{n}(y_j - \bar{y})(y_j - \bar{y})' = CSC'$

Denote $\mu_Y = C\mu + d$ $\quad \mu_{Y,0} = C\mu_0 + d$

Then,

$$
\begin{aligned}
T^2 &= n(\bar{y} - \mu_{Y,0})' S_y^{-1} (\bar{y} - \mu_{Y,0}) \\
&= n(C(\bar{x} - \mu_0))'(CSC')^{-1}(C(\bar{x} - \mu_0)) \\
&= n(\bar{x} - \mu_0)'C'(CSC')^{-1}C(\bar{x} - \mu_0) \\
&= n(\bar{x} - \mu_0)'C'(C')^{-1}S^{-1}C^{-1}C(\bar{x} - \mu_0) = \boxed{n(\bar{x} - \mu_0)'S^{-1}(\bar{x} - \mu_0)}
\end{aligned}
$$

线性变换后不变

Hotelling's T^2 vs LRT

Consider the point null hypothesis

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0,$$

and the corresponding likelihood ratio test statistic

$$\Lambda = \frac{\max_{H_0} L(\mu_0, \Sigma)}{\max_{H_0 \cup H_1} L(\mu_1, \Sigma)} = \left(1 + \frac{T^2}{n-1}\right)^{-n/2}$$

with

$$T^2 = T^2_{\mu_0} = n(\bar{x} - \mu_0)' S^{-1} (\bar{x} - \mu_0).$$

Under the hypothesis H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, the normal likelihood specializes to

$$L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} \exp\left(-\frac{1}{2}\sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_j - \boldsymbol{\mu}_0)\right)$$

The mean $\boldsymbol{\mu}_0$ is now fixed, but $\boldsymbol{\Sigma}$ can be varied to find the value that is "most likely" to have led, with $\boldsymbol{\mu}_0$ fixed, to the observed sample. This value is obtained by maximizing $L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$ with respect to $\boldsymbol{\Sigma}$.

Following the steps in (4-13), the exponent in $L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$ may be written as

$$-\frac{1}{2}\sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_j - \boldsymbol{\mu}_0) = -\frac{1}{2}\sum_{j=1}^{n}\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}(\mathbf{x}_j - \boldsymbol{\mu}_0)(\mathbf{x}_j - \boldsymbol{\mu}_0)'\right]$$
$$= -\frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu}_0)(\mathbf{x}_j - \boldsymbol{\mu}_0)'\right)\right]$$

Applying Result 4.10 with $\mathbf{B} = \sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu}_0)(\mathbf{x}_j - \boldsymbol{\mu}_0)'$ and $b = n/2$, we have

指数部分相等

$$\max_{\boldsymbol{\Sigma}} L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2}|\hat{\boldsymbol{\Sigma}}_0|^{n/2}} e^{-np/2} \tag{5-11}$$

with

$$\hat{\boldsymbol{\Sigma}}_0 = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu}_0)(\mathbf{x}_j - \boldsymbol{\mu}_0)'$$

To determine whether $\boldsymbol{\mu}_0$ is a plausible value of $\boldsymbol{\mu}$, the maximum of $L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$ is compared with the unrestricted maximum of $L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The resulting ratio is called the *likelihood ratio statistic*.

Using Equations (5-10) and (5-11), we get

$$\text{Likelihood ratio} = \Lambda = \frac{\max_{\boldsymbol{\Sigma}} L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})}{\max_{\boldsymbol{\mu}, \boldsymbol{\Sigma}} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})} = \left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_0|}\right)^{n/2} \tag{5-12}$$

The equivalent statistic $\Lambda^{2/n} = |\hat{\boldsymbol{\Sigma}}|/|\hat{\boldsymbol{\Sigma}}_0|$ is called *Wilks' lambda*. If the observed value of this likelihood ratio is too small, the hypothesis H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ is unlikely to be true and is, therefore, rejected. Specifically, the likelihood ratio test of H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ against H_1: $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ rejects H_0 if

$$\Lambda = \left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_0|}\right)^{n/2} = \left(\frac{\left|\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'\right|}{\left|\sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu}_0)(\mathbf{x}_j - \boldsymbol{\mu}_0)'\right|}\right)^{n/2} < c_\alpha \tag{5-13}$$

where c_α is the lower (100α)th percentile of the distribution of Λ. (Note that the likelihood ratio test statistic is a power of the ratio of generalized variances.) Fortunately, because of the following relation between T^2 and Λ, we do not need the distribution of the latter to carry out the test.

Hotelling's T^2 vs LRT

Since by MLE, we have

$$\Lambda = \left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|} \right)^{\frac{n}{2}}$$

We can obtain T^2 easily from determinant, avoiding calculation of S^{-1}.

$$T^2 = \frac{(n-1)|\hat{\Sigma}_0|}{|\hat{\Sigma}|} - (n-1)$$

$$\boxed{= \frac{(n-1)\left|\sum_{j=1}^{n}(x_j-\mu_0)(x_j-\mu_0)'\right|}{\left|\sum_{j=1}^{n}(x_j-\overline{x})(x_j-\overline{x})'\right|} - (n-1)}$$

这里是矩阵乘积，不是内积

Hotelling's T^2 vs LRT

Proof.

$$A = \begin{bmatrix} \sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})' & \sqrt{n}(\bar{x}-\mu_0) \\ \sqrt{n}(\bar{x}-\mu_0)' & -1 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

$$|A| = |A_{22}||A_{11} - A_{12}A_{22}^{-1}A_{21}| = |A_{11}||A_{22} - A_{21}A_{11}^{-1}A_{12}|$$

$$(-1)\left|n\hat{\Sigma}_0\right| = (-1)\left|\sum_{j=1}^{n}(x_j-\mu_0)(x_j-\mu_0)'\right| = (-1)\left|\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})' + n(\bar{x}-\mu_0)(\bar{x}-\mu_0)'\right|$$

$$= \left|\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'\right|\left|-1-n(\bar{x}-\mu_0)'\left(\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'\right)^{-1}(\bar{x}-\mu_0)\right|$$

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

Confidence Interval

What is our confidence about the mean?

Example – Scores for Statistical Courses

- What is our confidence about the mean vector?
- What is our confidence about the GPA?
- What if we have multiple ways for calculating GPA?

```
> head(data)
  Prob Inference Computing MVA LinearReg
1   83        75        88  86        79
2   85        74        90  85        79
3   84        81        90  84        75
4   83        77        91  91        82
5   87        70        89  85        83
6   84        79        91  86        73
```

```
> colMeans(data)
     Prob Inference Computing       MVA LinearReg
    84.98     74.79     89.17     84.97     78.60
```

Confidence regions

Recall 1-dim:

$$z_1, \ldots, z_n \overset{iid}{\sim} N(\psi, \sigma_\psi^2)$$

$$s^2 = \frac{1}{n-1} \sum_i (z_i - \bar{z})^2 \sim \sigma_\psi^2 \chi_{n-1}^2 / (n-1)$$

$$\hat{\psi} = \bar{z} \sim N(\psi, \frac{\sigma_\psi^2}{n})$$

$$se(\hat{\psi}) = \frac{\sigma}{\sqrt{n}} \Rightarrow \hat{se}(\hat{\psi}) = \frac{s}{\sqrt{n}}$$

$$\frac{\hat{\psi} - \psi}{\hat{se}(\hat{\psi})} = \frac{\bar{z} - \psi}{s/\sqrt{n}} \sim t_{n-1}$$

So a confidence interval for the mean has the form:

$$\bar{z} \pm M_\alpha \frac{s}{\sqrt{n}} = I(z).$$

Confidence regions

The coverage property of the interval $I(z)$ is

$$
\begin{aligned}
\mathbb{P}_\psi[I(z) \text{ covers } \psi] &= \mathbb{P}\left[\bar{z} - M_\alpha \frac{s}{\sqrt{n}} \leqslant \psi \leqslant \bar{z} + M_\alpha \frac{s}{\sqrt{n}}\right] \\
&= \mathbb{P}\left[\left|\frac{\bar{z} - \psi}{s/\sqrt{n}}\right| \leqslant M_\alpha\right] \\
&= \mathbb{P}\left[|t_{n-1}| \leqslant M_\alpha\right] \\
&= 1 - \alpha
\end{aligned}
$$

if $M_\alpha = t_{n-1}(\alpha/2)$.

Confidence regions

For p-dim: We have corresponding confidence regions. Suppose

$$x_1, \ldots, x_n \sim N(\mu, \Sigma)$$

Then a credible region $R(x)$ is a subset of $\mathbb{R}^p$ with the property

$$\mathbb{P}_\mu[R(x) \text{ covers } \mu] = 1 - \alpha.$$

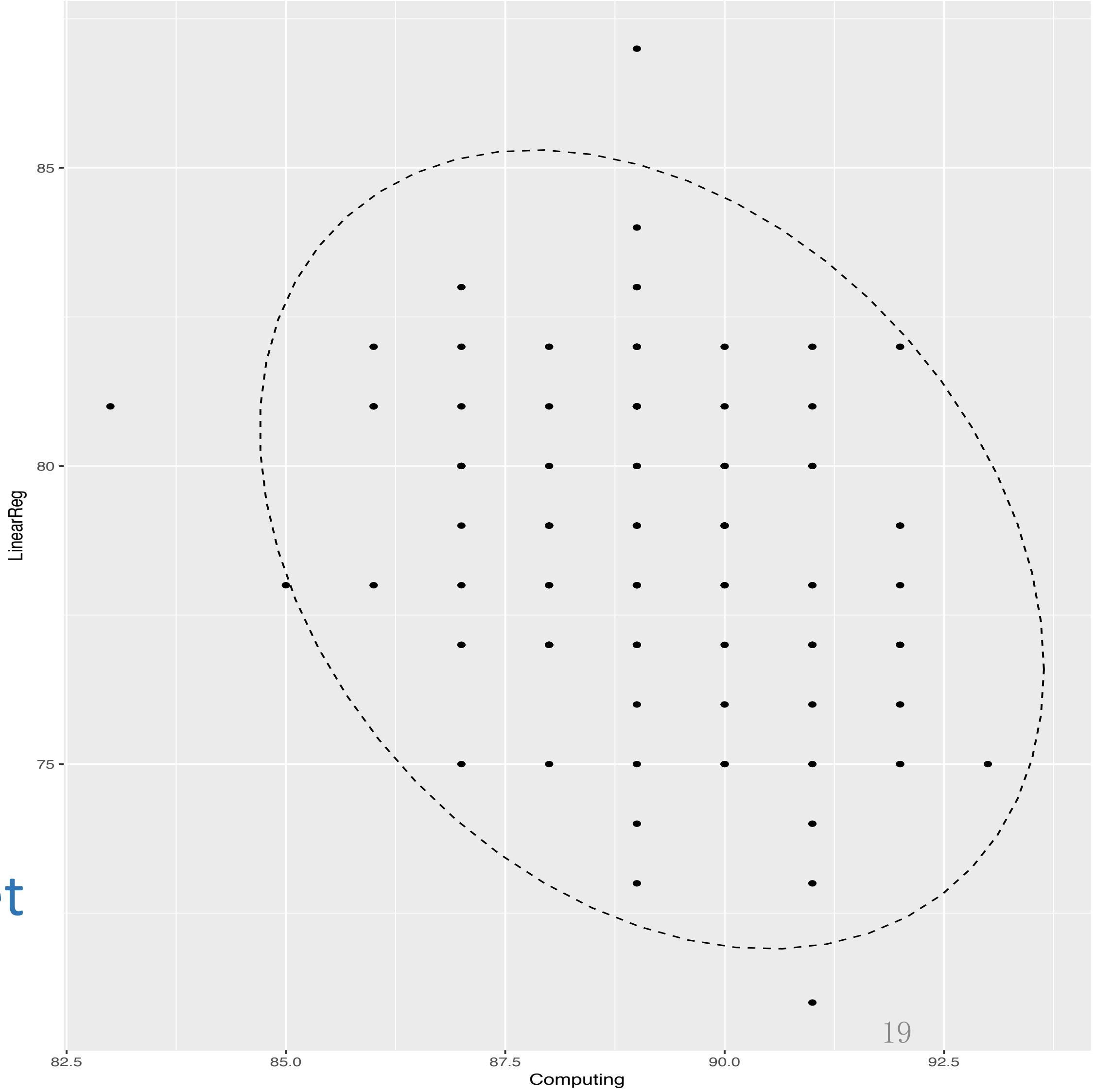
These are *ellipsoids*

$$\boxed{\begin{aligned} R(x) &= \{\mu : n(\bar{x} - \mu)' S^{-1} (\bar{x} - \mu) = c^2\} \\ &= \{\mu : T_\mu^2 \leqslant c^2\} \end{aligned}}$$

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

Example – Scores for Statistical Courses

- What is our confidence about the mean vector?
- What is our confidence about the GPA?
- What if we have multiple ways for calculating GPA?



Display for 2-variable sub-dataset
$\alpha = 0.05$, $p = 2$, $n = 100$

Example – Scores for Statistical Courses

- What is our confidence about the mean vector?
- What is our confidence about the GPA?
- What if we have multiple ways for calculating GPA?

e.g. GPA = weighted linear combination of scores by credits

```
> head(data)
  Prob Inference Computing MVA LinearReg
1   83        75        88  86        79
2   85        74        90  85        79
3   84        81        90  84        75
4   83        77        91  91        82
5   87        70        89  85        83
6   84        79        91  86        73

> colMeans(data)
     Prob Inference Computing       MVA LinearReg
    84.98     74.79     89.17     84.97     78.60
```

Individual Coverage Intervals

Consider linear combinations

$$\psi_k = a_k'\mu, \quad z_{ik} = a_k' x_i \sim N(a_k'\mu, a_k'\Sigma a_k)$$

and put $\psi_k = a_k'\mu$ and $\sigma^2_{\psi_k} = a_k'\Sigma a_k$, and

$$\hat{\psi}_k = \bar{z}_k = a_k'\bar{x}.$$

由此可以构造对于某个分量的置信区间

Then

$$\hat{se}(\hat{\psi}_k) = \frac{s_{zk}}{\sqrt{n}}, \quad s^2_{zk} = \frac{1}{n-1}\sum_i (z_{ik} - \bar{z}_k)^2 = a'Sa.$$

Then a $100(1-\beta)\%$ t-interval is

$$I_{\psi_k} : \hat{\psi}_k \pm t_{n-1}(\beta/2)\hat{se}(\hat{\psi}_k), \quad a_k'\bar{x} \pm t_{n-1}(\beta/2)\sqrt{a_k' \frac{S}{n} a_k}$$

Example – Scores for Statistical Courses

- What is our confidence about the mean vector?
- What is our confidence about the GPA?
- What if we have multiple ways for calculating GPA?

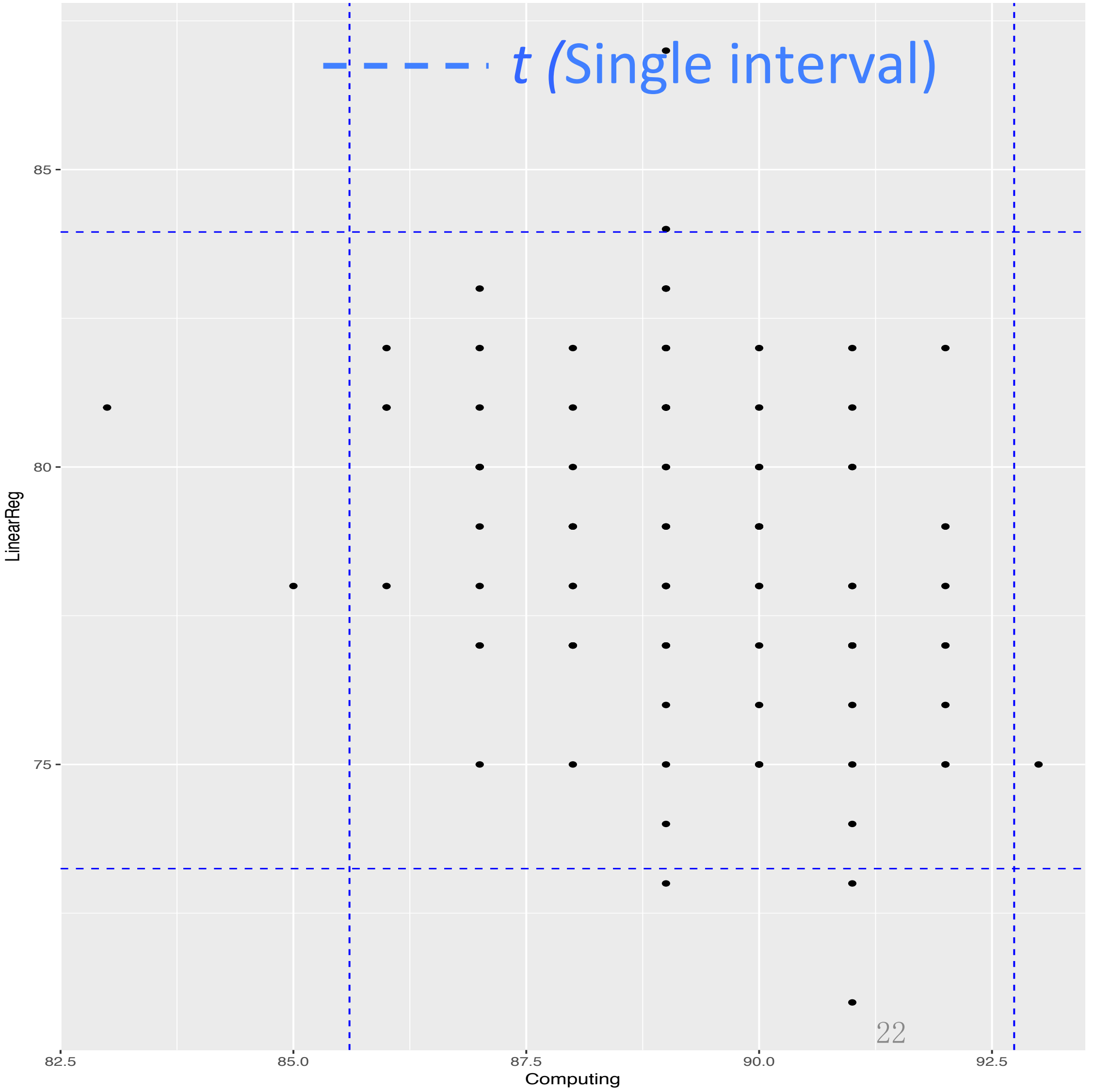
e.g. GPA = weighted linear combination of scores by credits

```
> qt(1-alpha/2,n-1)
[1] 1.984217
```

Display for 2-variable sub-datase

$\alpha = 0.05$, $p = 2$, $n = 100$

$a_1=[1,0]$, $a_2=[0,1]$



Example – Scores for Statistical Courses

- What is our confidence about the mean vector?
- What is our confidence about the GPA?
- What if we have multiple ways for calculating GPA?

```
> head(data)
  Prob Inference Computing MVA LinearReg
1   83        75        88  86        79
2   85        74        90  85        79
3   84        81        90  84        75
4   83        77        91  91        82
5   87        70        89  85        83
6   84        79        91  86        73

> colMeans(data)
     Prob Inference Computing       MVA LinearReg
    84.98     74.79     89.17     84.97     78.60
```

Individual vs Simultaneous Coverage Intervals

An individual confidence interval has coverage property defined by probability statements of the form

$$\mathbb{P}[I_{\psi_k} \text{ covers } \psi_k],$$

while *simultaneous* coverage concerns the probability

$$\mathbb{P}[I_{\psi_k} \text{ covers } \psi_k \text{ for all } k = 1, \ldots, m].$$

Simultaneous Coverage Intervals

Let C_k denote the event $\{I_{\psi_k}$ covers $\psi_k\}$. By Bonferroni's inequality

$$\mathbb{P}[\text{some } C_k \text{ false}] \leqslant \sum_{k=1}^{m} \mathbb{P}[C_k \text{ false}] = \sum_{k=1}^{m} \beta = m\beta,$$

assuming each I_{ψ_k} is a $100(1-\beta)\%$ interval. So if we put $\beta = \alpha/m$ then

$$\mathbb{P}[\text{all } C_k \text{ true}] \geqslant 1 - m\beta = 1 - \alpha,$$

so the simultaneous interval is *conservative*. These are called *Bonferroni* intervals

$$\hat{\psi}_k \pm t_{n-1}(\alpha/(2m))\hat{se}(\hat{\psi}_k).$$

Example – Scores for Statistical Courses

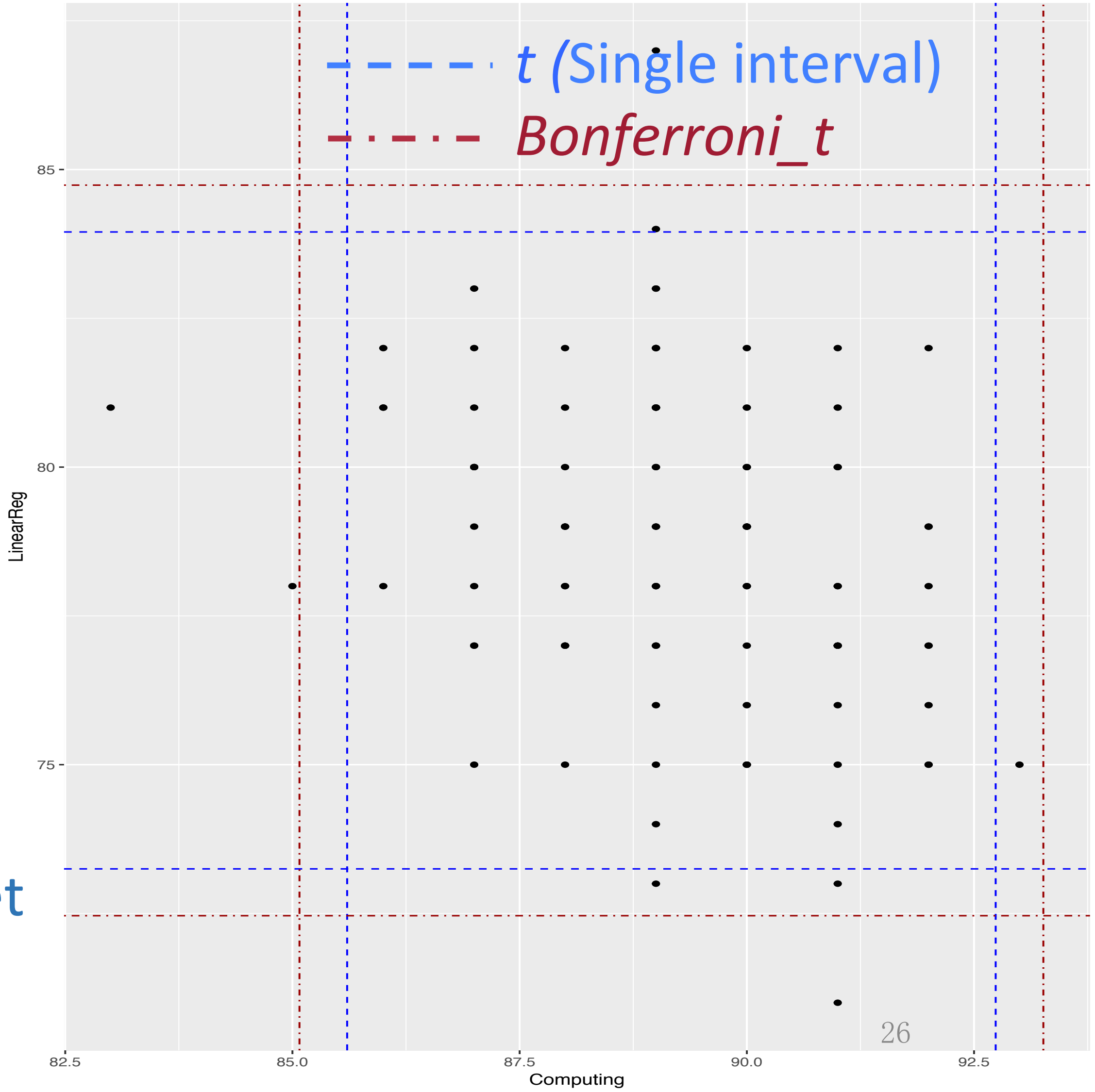
- What is our confidence about the mean vector?
- What is our confidence about the GPA?
- What if we have multiple ways for calculating GPA?

```
> qt(1-alpha/2/m,n-1)
[1] 2.276003
```

Display for 2-variable sub-dataset

$\alpha = 0.05$, $p = 2$, $n = 100$, $m = 2$

$a_1=[1,0]$, $a_2=[0,1]$



Example – Scores for Statistical Courses

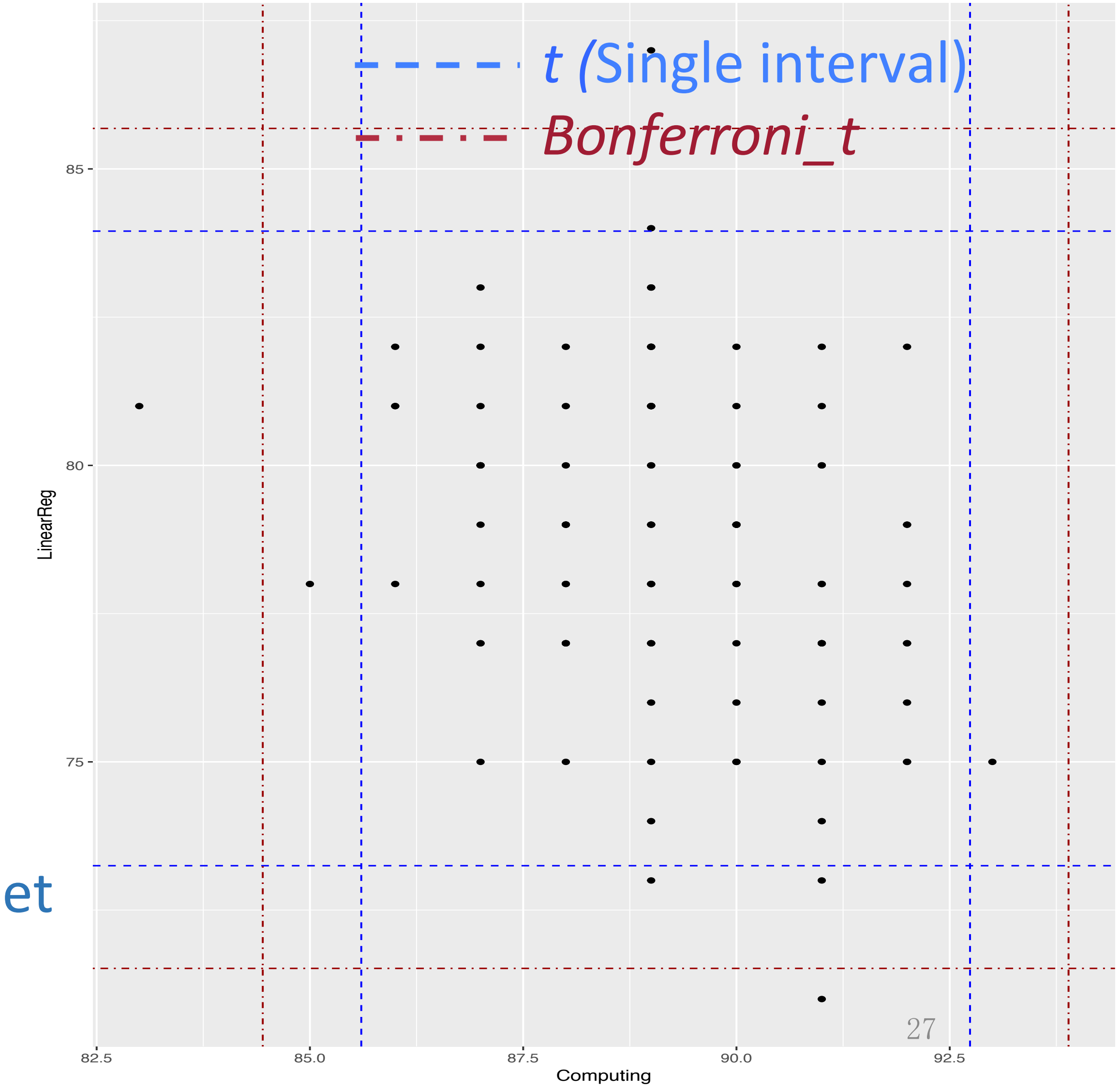
- What is our confidence about the mean vector?
- What is our confidence about the GPA?
- What if we have multiple ways for calculating GPA?

```
> qt(1-alpha/2/m,n-1)
[1] 2.626405
```

Display for 2-variable sub-dataset

$\alpha = 0.05$, $p = 2$, $n = 100$, $m = 5$

$a_1=[1,0]$, $a_2=[0,1]$



Simultaneous Coverage Intervals

On the other hand, can we find a lower bound for c, such that

$$I_a(c) = a'\bar{x} \pm c\sqrt{a'\frac{S}{n}a} \quad \text{covers } a'\mu \text{ for all } a \neq 0 \quad ?$$

$$\Longleftrightarrow \quad \max_a t^2 = \max_a \frac{n(a'(\bar{x}-\mu))^2}{a'Sa}$$

By maximization lemma (2-50) in Chapter 2

$$= n\max_a \frac{(a'(\bar{x}-\mu))^2}{a'Sa} = n(\bar{x}-\mu)'S^{-1}(\bar{x}-\mu) = T^2$$

Simultaneous Coverage Intervals

Another type of interval with a simultaneous coverage property under the normal likelihood is the T^2 interval

$$I_a(c) = a'\overline{x} \pm c\sqrt{a'\frac{S}{n}a}$$

with the property that if

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

then

$$\mathbb{P}[I_a(c) \text{ covers } a'\mu \text{ for all } a \neq 0] = 1 - \alpha.$$

$1-\alpha = P(T^2 \le c^2) = P(t_a^2 \le c^2, \forall a)$

Individual Coverage Intervals

Consider linear combinations

$$\psi_k = a_k'\mu, \quad z_{ik} = a_k' x_i \sim N(a_k'\mu, a_k'\Sigma a_k)$$

and put $\psi_k = a_k'\mu$ and $\sigma^2_{\psi_k} = a_k'\Sigma a_k$, and

$$\hat{\psi}_k = \bar{z}_k = a_k'\bar{x}.$$

由此可以构造对于某个分量的置信区间

Then

$$\hat{se}(\hat{\psi}_k) = \frac{s_{zk}}{\sqrt{n}}, \quad s^2_{zk} = \frac{1}{n-1}\sum_i (z_{ik} - \bar{z}_k)^2 = a'Sa.$$

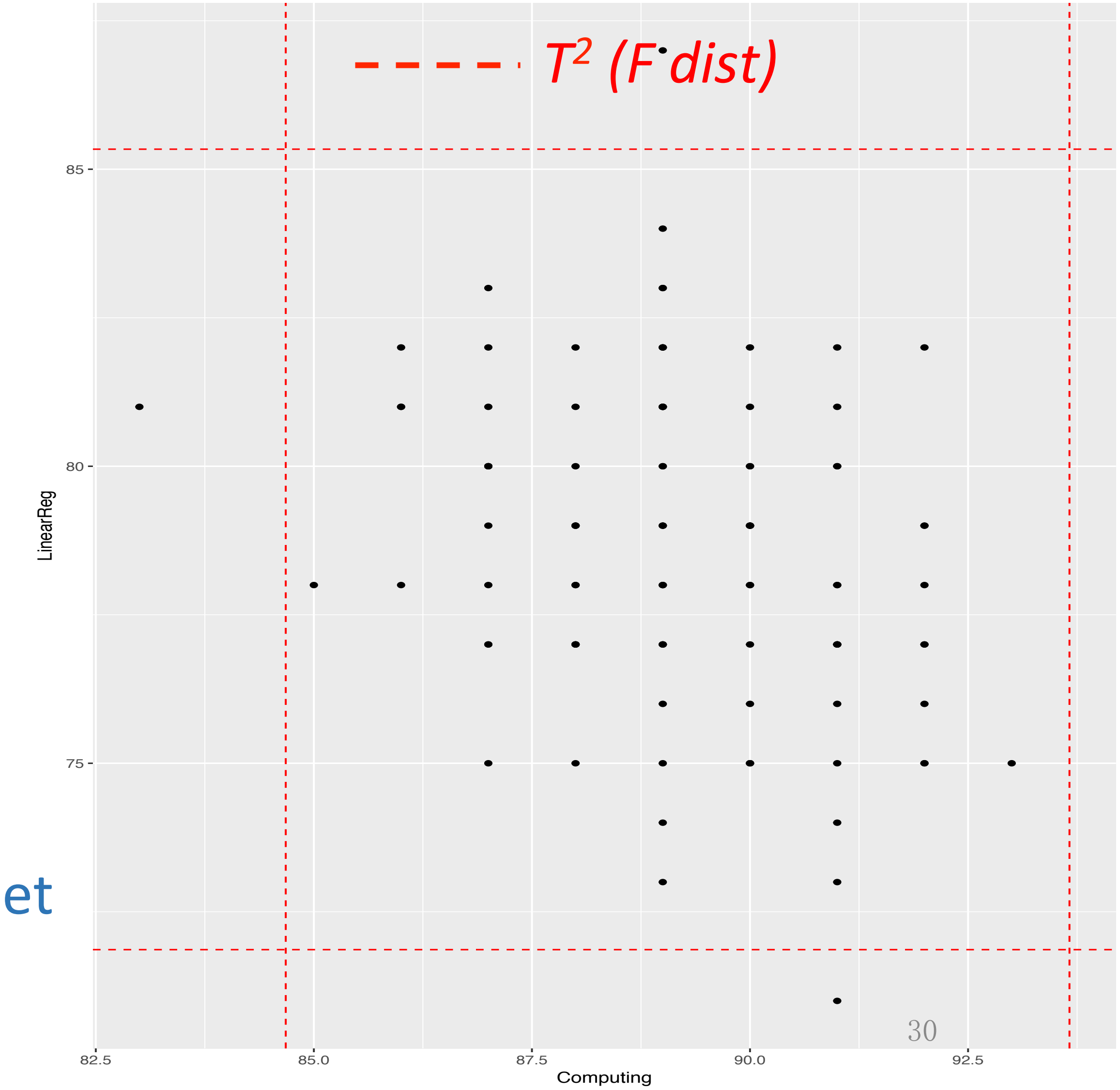
Then a $100(1-\beta)\%$ t-interval is

$$I_{\psi_k} : \hat{\psi}_k \pm t_{n-1}(\beta/2)\hat{se}(\hat{\psi}_k), \quad a_k'\bar{x} \pm t_{n-1}(\beta/2)\sqrt{a_k' \frac{S}{n} a_k}$$

Example – Scores for Statistical Courses

- What is our confidence about the mean vector?
- What is our confidence about the GPA?
- What if we have multiple ways for calculating GPA?

```
> sqrt(qf(1-alpha,p,n-p)*p*(n-1)/(n-p))
[1] 2.49829
```



Display for 2-variable sub-dataset

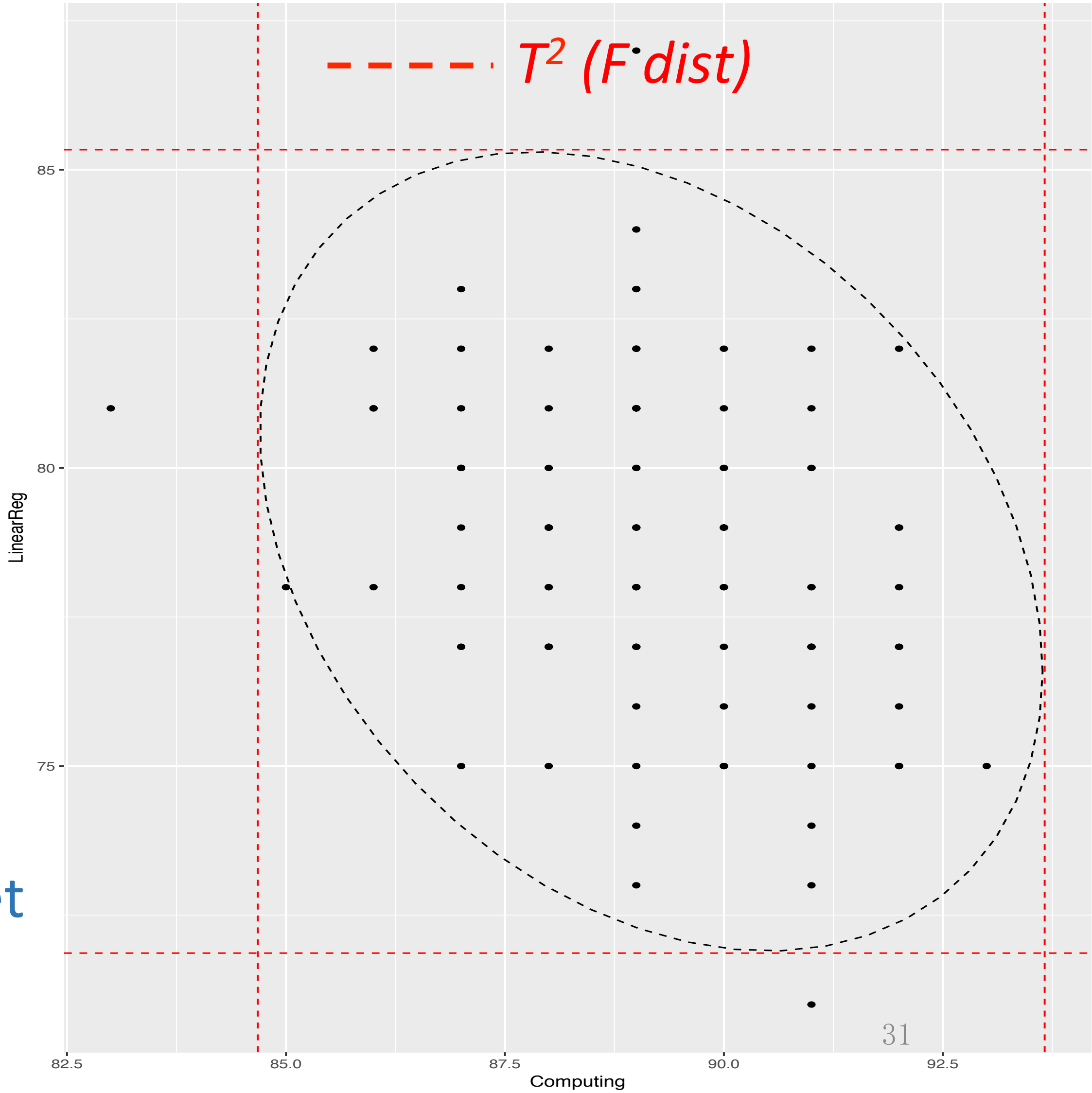
$\alpha = 0.05$, $p = 2$, $n = 100$

$a_1=[1,0]$, $a_2=[0,1]$

Example – Scores for Statistical Courses

It happens to be the projection of T^2 Interval on corresponding direction of a (the linear combination vector).

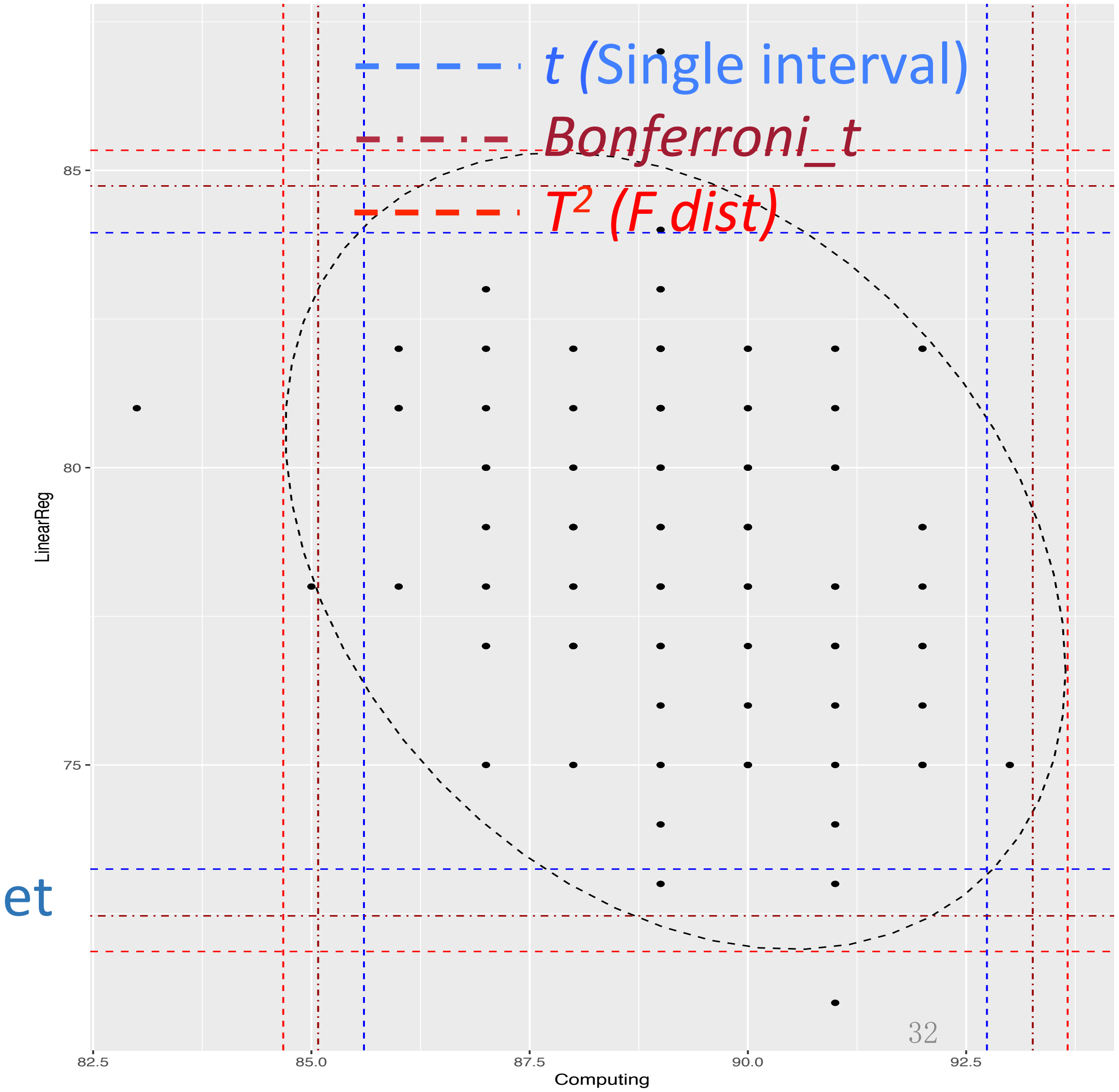
Not a coincidence!



Display for 2-variable sub-dataset
$\alpha = 0.05$, $p = 2$, $n = 100$
$a_1=[1,0]$, $a_2=[0,1]$

Example – Scores for Statistical Courses

- What can you find from the comparisons?



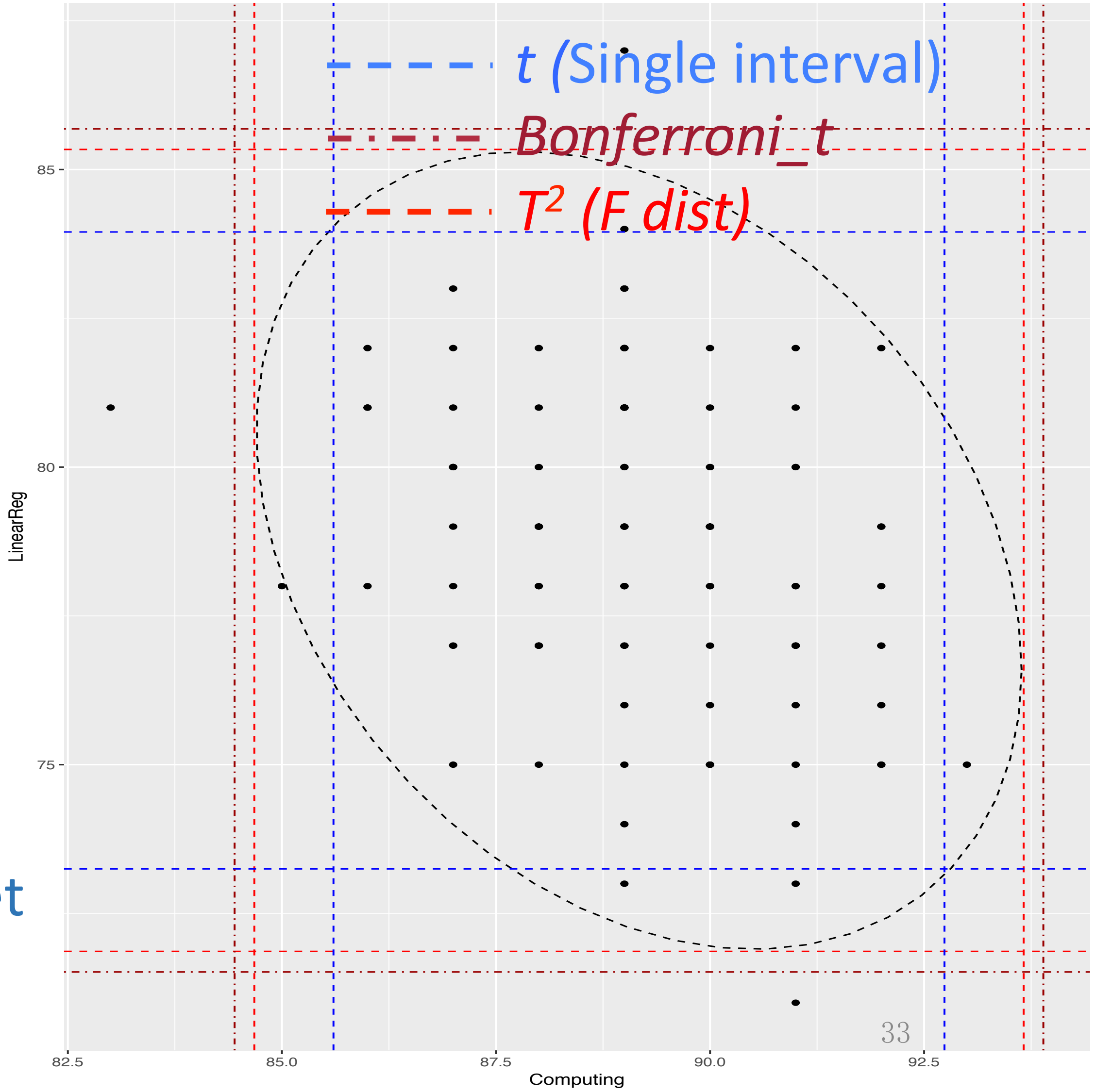
Display for 2-variable sub-dataset

$\alpha = 0.05$, $p = 2$, $n = 100$, $m = 2$

$a_1=[1,0]$, $a_2=[0,1]$

Example – Scores for Statistical Courses

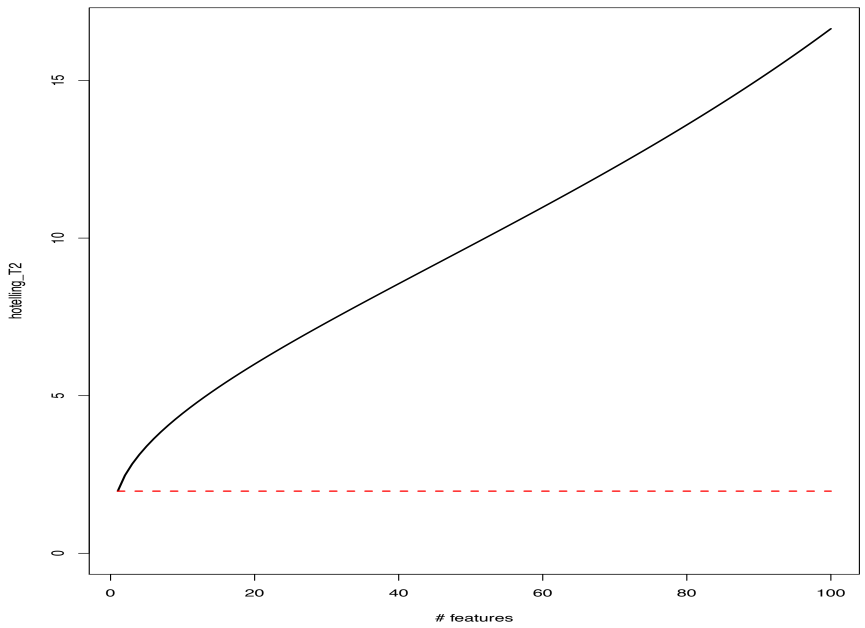
Only Bonferroni Interval changes with m (the number of linear combinations)



Display for 2-variable sub-dataset
$\alpha = 0.05$, $p = 2$, $n = 100$, $m = 5$
$a_1=[1,0]$, $a_2=[0,1]$

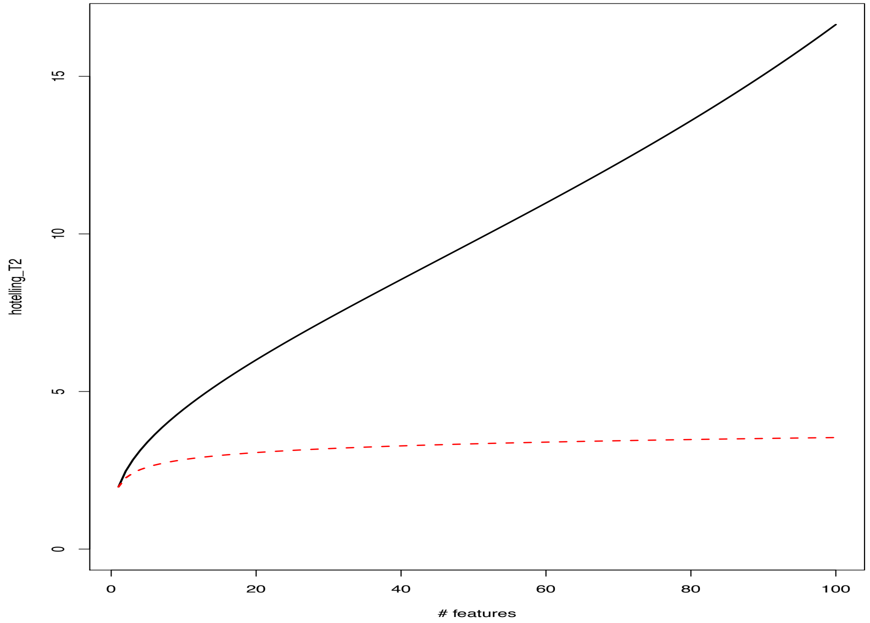
Comparison between T2 interval and Marginal

```
n = 200; p = seq(1, 100, by=1)
alpha = 0.05
one_at_a_time = rep(qt(1 - 0.5 * alpha, n-1), length(p))
hotelling_T2 = sqrt(p * (n - 1) / (n - p) * qf(1 - alpha, p, n - p))
plot(p, hotelling_T2, col='black', type='l', lwd=2, ylim=c(0, max(hotelling_T2)), xlab='# features')
lines(p, one_at_a_time, col='red', lty=2, lwd=2)
```



Comparison between Bonferroni and T2 interval

```
n = 200; p = seq(1, 100, by=1)
alpha = 0.05
bonferroni = qt(1 - 0.5 * alpha / p, n-1)
hotelling_T2 = sqrt(p * (n - 1) / (n - p) * qf(1 - alpha, p, n - p))
plot(p, hotelling_T2, col='black', type='l', lwd=2, ylim=c(0, max(hotelling_T2)), xlab='# features')
lines(p, bonferroni, col='red', lty=2, lwd=2)
```



Large Sample Inference about a Population Mean Vector

Large Sample Inference about Mean Vector

Res 5.4
Hypothesis Testing

Let $X_1, \ldots, X_n$ be a random sample from a population with mean μ and positive definite covariance matrix Σ. When n-p is large, the hypothesis $H_0: \mu = \mu_0$ is rejected in favor of $H_1: \mu \neq \mu_0$, at a level of significance approximately α, if the observed

$$\boxed{n(\bar{x} - \mu_0)' S^{-1} (\bar{x} - \mu_0) > \chi_p^2(\alpha)}$$

利用中心极限定理

Here $\chi_p^2(\alpha)$ is the upper (100α)th percentile of a chi-square distribution with p d.f.

样本量大时 T^2近似服从χ_p^2分布

- Appropriate when n is large relative to p.
- When appropriate, the results are the same as for normal case, since $(n-1)p\, F_{p,n-p}(\alpha)/(n-p)$ and $\chi^2_p(\alpha)$ are approximately equal then.

Large Sample Inference about Mean Vector

Res 5.5
Confidence Regions

Let $X_1,\ldots,X_n$ be a random sample from a population with mean μ and positive definite covariance matrix Σ. When n-p is large,

$$a'\overline{x} \pm \sqrt{\chi_p^2(\alpha)}\sqrt{\frac{a'Sa}{n}}$$

will contain $a'\mu$, for every a, with probability approximately 1-α.

Summary

Summary

- Inference for a normal population mean, analogue to 1-dim case:
  - Hypothesis test (decision making): Hotelling's T^2
  - Properties of T^2:
    - Invariant to scales
    - Connections between Hotelling's T^2 and LRT
  - Confidence regions
  - Simultaneous comparisons of component means:
    - linear combinations
    - two choices: Bonferroni, T^2
- Large Sample Inferences about a population mean vector
  - based on large sample theory of sample statistics (Ch. 4)
  - comparison with normal cases