6

TENTAMEN (EXAMINATION)

Tentamensdatum/*Examination date:* 20-01-14
(åå-mm-dd/*yy-mm-dd)*

	Ifylles av student	Ifylles av vakt
AID-nummer ***AID number***	2 9 8 4	2 9 8 4
	Completed by student	*Completed by supervisor*

Utbildningskod/*Education code:* 732A97 **Modul/*Module:*** TENT

Kursnamn/*Course title:* Multivariate Statistical Methods

Institution/*Department:* IDA

Jag intygar att varken mobil eller något annat otillåtet hjälpmedel finns tillgängligt under tentamen.
I confirm that no mobile or other non-permitted aids are available during the examination. ☑

Inlämnat: antal lösblad 7 **tentamensformulär** ☐
Enclosed: number of sheets ***exam booklet***

Markera behandlade uppgifter med X/*Mark tasks attempted with an X*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X här/*here*	x	x	x	x											
Erhållna poäng *Points obtained*	4	5	3½	~~3~~ 3½											
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
X här/*here*															
Erhållna poäng *Points obtained*															

Anvisningar/*Instructions*

1. Skriv AID-nummer, datum, utb.kod, modul på varje blad som lämnas in/*Write AID number, date, edu.code and module on every sheet that is handed in*
2. På varje papper får högst en uppgift lösas om inget annat anges/ *Maximum one task per sheet unless otherwise instructed*
3. Skriv endast på papprets ena sida om inget annat anges/ *Use only one side of each sheet unless otherwise instructed*
4. Numrera de papper som lämnas in/*Number every sheet that is handed in*
5. Använd inte röd penna/*Do not use a red pen/pencil*

Sen inlämning ☐
Late hand in

Klockslag ____
Time

Orsak ____
Reason

Σ Poäng/*Points:* 16 Betyg/*Grade:* B

Examinator/*Examiner:* [signature]

LINKÖPINGS

1

Problem 1

Let Σ be a pxp symmetric equivcorrelation matrix, then Σ is said to be symmetric +ve semi definite if

$\boxed{x^T \Sigma x \geq 0 \text{ for all } x}$ ✓

5

$$\underset{(p \times p)}{\Sigma} = (1-\alpha)\underset{(p\times p)}{I} + \alpha \underset{(p\times 1)}{1_p}\,\underset{(1\times p)}{1_p^T}$$ ✓

Then,

$$x^T \Sigma x = x^T (I - \alpha I + \alpha 1_p 1_p^T) x$$

$$= x^T I x - \alpha x^T I x + \alpha x^T 1_p 1_p^T x$$

$$= x^T x - \alpha x^T x + \alpha (x^T 1_p).(x^T 1_p)$$ because $x^T 1_p = 1_p^T x$ (dot product of vectors)

$$= x^T x - \alpha x^T x + \alpha (x^T 1_p)^2$$

For any vector $x = [x_1, x_2 \; \cdots \; x_p]$, $x^T \Sigma x$ can be written as

$$x^T \Sigma x = (x_1^2 + x_2^2 + \cdots x_p^2) - \alpha (x_1^2 + x_2^2 + x_3^2 + \cdots x_p^2) + \alpha (x_1 + x_2 + \cdots x_p)^2$$?

Since $\alpha \geq 0$ and all the individual terms above are squared, we can conclude $x^T \Sigma x \geq 0$ for all x. ($x^T \Sigma x$ is always +ve)

you have a $-\alpha$

Specifically, when $(x_1^2 + x_2^2 + \cdots x_p^2) + \alpha (x_1 + x_2 + \cdots x_p)^2 = \alpha (x_1^2 + x_2^2 + x_3^3 + \cdots x_p^2)$, we get

$x^T \Sigma x = 0$.

Hence it is +ve semi-definite.

AID-nummer: AID-number: 2984	Datum: Date: 20-01-14
Utbildningskod: Education code: 732A97	Modul: Module: TENT

Blad nummer: Sheet number: 2

2 Problem 2

If $X \sim N_p(\mu, \Sigma)$, then any set of linear combinations of components of X, $AX \sim N_p(A\mu, A\Sigma A')$

Hence, if $X \sim N_p(\mu, \Sigma)$, then

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

when $\mu = 0$ and $\Sigma = \sigma^2 I$, then

$G\mu = 0$

$$
\begin{aligned}
G\Sigma G' &= G(\sigma^2 I)G' \\
&= \sigma^2(GIG') \\
&= \sigma^2(GG') \quad (\text{Because } G \text{ is orthogonal, } GG' = I) \\
&= \sigma^2 I \\
&= \Sigma
\end{aligned}
$$

Therefore, GX has distribution $N_p(0, \sigma^2 I)$ which is the same as X.

5

AID-nummer: / AID-number: 2984	Datum: / Date: 20-01-14		Blad nummer: / Sheet number: 3
Utbildningskod: / Education code: 732A97	Modul: / Module: TENT		

LINKÖPINGS UNIVERSITET

$\boxed{3}$ Problem 3

3½ a) Assuming that the sample data comes from a normal distribution, ~~i.e. X~~ $N(\mu, \Sigma_1)$, the statistical distance (or "closeness") of the sample mean $\bar{x}$ to the population mean μ has an F-distribution

$$(\bar{x}-\mu)^T \left(\frac{1}{n}S\right)^{-1}(\bar{x}-\mu) \sim \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha) \quad (T^2 \text{ test statistic})$$

No! You know Σ!

NULL Hypothesis:

We assume that the observed sample with a sample mean $\bar{x} = [185.72, 183.84]^T$ and sample variance S, comes from a normal population with a population mean $[182, 182]^T$.

$$H_0: \mu = \begin{bmatrix} 182 \\ 182 \end{bmatrix}$$

The Null hypothesis can be tested by calculating the test statistic T^2 as mentioned above.

$$\bar{x} = \begin{bmatrix} 185.72 \\ 183.84 \end{bmatrix} \quad \mu = \begin{bmatrix} 182 \\ 182 \end{bmatrix} \quad \bar{x}-\mu = \begin{bmatrix} 3.72 \\ 1.84 \end{bmatrix} \quad S^{-1} = \begin{bmatrix} 0.022 & -0.015 \\ -0.015 & 0.021 \end{bmatrix}$$

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

$$= 25 \begin{bmatrix} 3.72 & 1.84 \end{bmatrix} \begin{bmatrix} 0.022 & -0.015 \\ -0.015 & 0.021 \end{bmatrix} \begin{bmatrix} 3.72 \\ 1.84 \end{bmatrix}$$

$$= 25 \begin{bmatrix} 0.082 - 0.028, & -0.056 + 0.039 \end{bmatrix} \begin{bmatrix} 3.72 \\ 1.84 \end{bmatrix}$$

$$= 25 \begin{bmatrix} 0.054 & -0.017 \end{bmatrix} \begin{bmatrix} 3.72 \\ 1.84 \end{bmatrix}$$

$$= 25 \begin{bmatrix} 0.2 + 0.03 \end{bmatrix}$$

$$= 25 \times \cancel{0.006}\ 0.17$$

$$= 4.25$$

$$\frac{(n-1)p}{n-p} F_{p,n-p}(\alpha) = \frac{24 \times 2}{23} \times F_{2,23}(0.05) \qquad \left(\alpha = 5\% = 0.05\right)$$

$$= 2.087 \times 3.42$$

$$= 7.14$$

1

Since $T^2 < 7.14$, we cannot reject the NULL hypothesis.

b) The 95% confidence ellipsoid for the mean $\mu = \begin{bmatrix} 182 \\ 182 \end{bmatrix}$ is given by

$$n(\bar{x}-\mu)' S^{-1} (\bar{x}-\mu) \le \frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)$$

The half-lengths of the axis are given by:

$\pm\sqrt{\lambda_i} \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)}$ along the direction of eigen vector e_i where $Se_i = \lambda_i e_i$

Axis 1: $\lambda_1 = 161.055$

$$\text{half length} = \sqrt{161.055} \times \sqrt{\frac{24 \times 2}{25 \times 23} \times 3.42}$$

$$= 12.69 \sqrt{0.2855}$$

$$= 12.69 \times 0.53$$

$$\sim 6.78$$

$$\sim 7$$

in the direction of $e_1 = \begin{bmatrix} 0.693 \\ 0.721 \end{bmatrix}$ and $\theta_1 = \tan^{-1}\left(\frac{0.721}{0.693}\right) \approx 46^\circ$

Axis 2: $\lambda_2 = 27.201$

$$\text{half length} = \sqrt{27.201} \sqrt{0.2855}$$

$$= 5.22 \times 0.53$$

$$\approx 2.78$$

$$\sim 3$$

in the direction of $e_2 = \begin{bmatrix} -0.721 \\ 0.693 \end{bmatrix}$ and $\theta_2 = \tan^{-1}\left(\frac{-0.693}{0.721}\right) \approx -44^\circ$

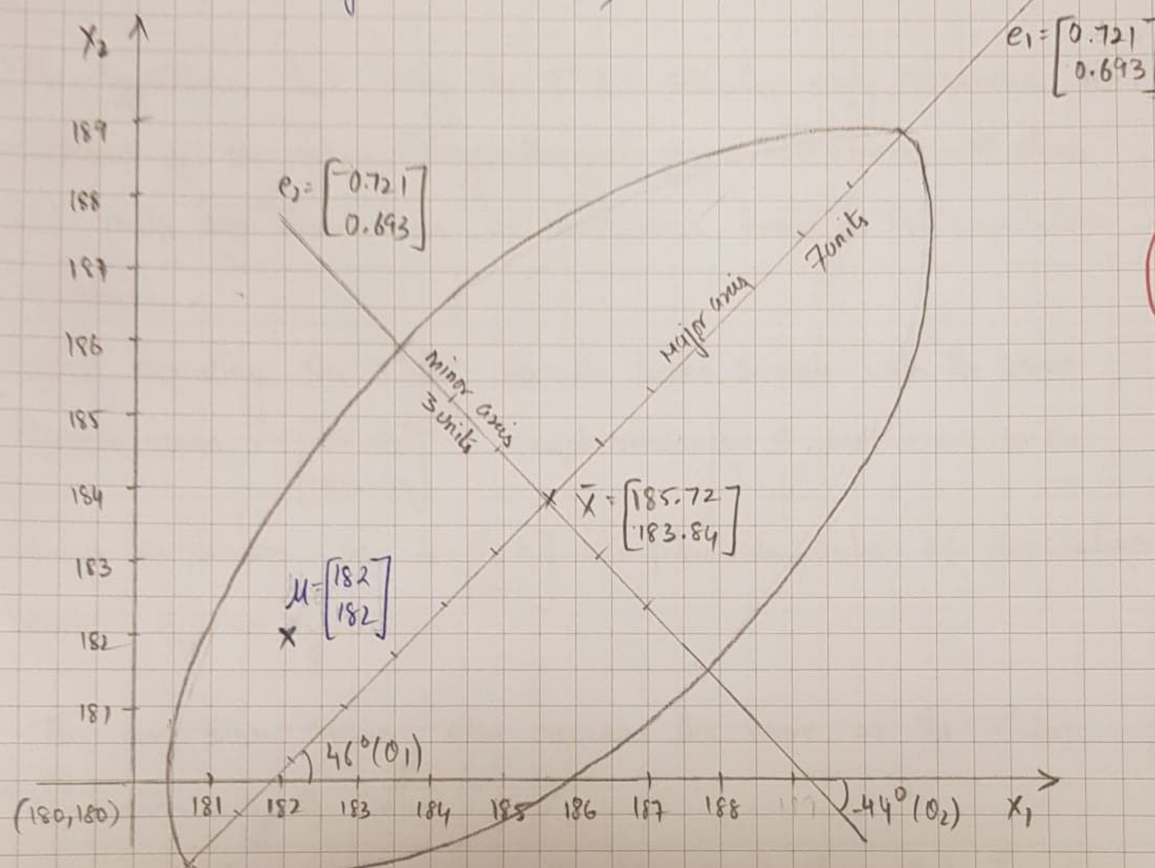
The ellipsoid is centered at $\bar{x} = \begin{bmatrix} 185.72 \\ 183.84 \end{bmatrix}$

LINKÖPINGS UNIVERSITET

AID-nummer: / AID-number: 2984	Datum: / Date: 20-01-14
Utbildningskod: / Education code: 732A97	Modul: / Module: TENT

Blad nummer: / Sheet number: 5

95% Confidence Region (ellipsoid)



1½

We can see Why? from the graph that the population Mean $\mu = \begin{bmatrix} 182 \\ 182 \end{bmatrix}$ lies inside the confidence region. Hence, it is a plausible estimate of μ based on the Sample mean $\bar{X}$. at 95% confidence level.

c) The assumed population covariance matrix $\Sigma_1 = \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix}$ claims that the two variables X_1 and X_2 are uncorrelated. (i.e $\sigma_{12} = 0$) and hence linearly independent from each other.

However, we can see from the Sample variance S (where $\sigma_{12} \neq 0$) and from the confidence region graph (~~or the constant density contour~~) (ellipse is not aligned to X_1, X_2 axis) that the variables X_1, X_2 are correlated.

Hence, it does NOT seem reasonable to assume that the population variance is Σ_1.

1

AID-nummer: *AID-number:* 2984	Datum: *Date:* 20-01-14
Utbildningskod: *Education code:* 732A97	Modul: *Module:* TENT

Blad nummer: *Sheet number:* 6

3½

Problem 4

a) The hypothesis test using the T^2 test statistic only measures the closeness of the assumed/hypothesized population mean. It does not make any assumptions or conclusions ~~based~~ on the population variance. 1

You know Σ!

Hence, repeating the T^2 test for the same sample with the same sample mean $\bar{x} = \begin{bmatrix} 185.72 \\ 183.84 \end{bmatrix}$ & sample variance S, will result in the same conclusion. i.e. $\mu = \begin{bmatrix} 182 \\ 182 \end{bmatrix}$ is a plausible value of population mean. H_0 is not rejected.

b) The confidence region also remains the same as the T^2 test was essentially the same. 1½

how do you know if (184, 181) is there!

c) The assumed new covariance $\Sigma_2 = \begin{bmatrix} 100 & 50 \\ 50 & 100 \end{bmatrix}$ claims that the two variables X_1 and X_2 are not linearly independent because they have $cov(X_1, X_2) \neq 0$. 1

This seems like a more reasonable assumption for the population variance given the data sample and sample variance.

The sample variance of a random sample from the population with variance Σ_2 can be estimated as (from the MLE estimate of population variance $\hat{\Sigma}$)

$$\hat{\Sigma} = \left(\frac{n-1}{n}\right) S \quad \text{(MLE estimate)}$$

$$\Rightarrow S = \left(\frac{n}{n-1}\right) \Sigma_2$$

$$= \frac{25}{24} \begin{bmatrix} 100 & 50 \\ 50 & 100 \end{bmatrix}$$

$$= \begin{bmatrix} 96 & 48 \\ 48 & 96 \end{bmatrix}$$ which is approximately the similar to the observed sample variance

AID-nummer: AID-number: 2984	Datum: Date: 20-01-14
Utbildningskod: Education code: 732A97	Modul: Module: TENT

Blad nummer: Sheet number: 7

Hence, it is more probable that the population variance is Σ_2 when compared to Σ_1