



Linear Models

STAT 551

Course Notes by Meredith Bartley



Copyright © 2013 John Smith

PUBLISHED BY PUBLISHER

BOOK-WEBSITE.COM

Licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License (the “License”). You may not use this file except in compliance with the License. You may obtain a copy of the License at <http://creativecommons.org/licenses/by-nc/3.0>. Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

First printing, March 2013

Contents

I	Part One	
1	Linear Regression	7
1.1	Projection in Euclidean Space	7
1.2	Cochran's Theorem	14
1.3	Gaussian Linear Regresson Model	15
1.4	Statistical Inference for β, σ^2	18
1.5	Delete One Prediciton	19
2	ANOVA (1-way)	21
2.1	Overview	21
3	Mutiway ANOVA	23
3.1	Overview	23
4	Nonorthogonal Design	25
4.1	Overview	25
5	Random Effects Model	27
5.1	Overview	27

II	Part Two	
6	Basic Concepts	31
6.1	Overview	31
7	Estimation	33
7.1	Overview	33
8	Inference	35
8.1	Overview	35
9	Residuals	37
9.1	Overview	37
10	Cetegorical Prediction	39
10.1	Overview	39
11	Some Important GLM	41
11.1	Overview	41
12	Multivariate GLM	43
12.1	Overview	43
III	Part Three	
13	Principle Component Analysis	47
13.1	Overview	47
14	Canonical Correlation Analysis	49
14.1	Overview	49
15	Independent Component Analysis	51
15.1	Overview	51
	Index	53



Part One

1	Linear Regression	7
1.1	Projection in Euclidean Space	
1.2	Cochran's Theorem	
1.3	Gaussian Linear Regression Model	
1.4	Statistical Inference for β, σ^2	
1.5	Delete One Prediction	
2	ANOVA (1-way)	21
2.1	Overview	
3	Mutlway ANOVA	23
3.1	Overview	
4	Nonorthogonal Design	25
4.1	Overview	
5	Random Effects Model	27
5.1	Overview	

1. Linear Regression

- projection
- orthongonal decomposition
- Gaussian Linear Regression
- prediction (generally of \hat{y})
- different types of errors
- influence
- lack of fit
- R^2
- Multicollinearity

1.1 Projection in Euclidean Space

Monday August 22

Definition 1.1.1 — Euclidian Space. One way to think of the Euclidean plane is as a set of points satisfying certain relationships, expressible in terms of distance and angle. **Euclidean space** is an abstraction detached from actual physical locations, specific reference frames, measurement instruments, and so on.

Let Euclidian Space be denoted by \mathbb{R}^P .

$$\mathbb{R}^X \dots \mathbb{R}^X = \{(x_1, \dots, x_p) : x_1 \in \mathbb{R}, \dots, x_p \in \mathbb{R}\}$$

Definition 1.1.2 — Inner Product. In linear algebra, an inner product space is a vector space with an additional structure called an inner product. This additional structure associates each pair of vectors in the space with a scalar quantity known as the inner product of the vectors. **Inner products** allow the rigorous introduction of intuitive geometrical notions such as the length of a vector or the angle between two vectors. They also provide the means of defining orthogonality between vectors (zero inner product).

Let $a \in \mathbb{R}^P, b \in \mathbb{R}^P$

$$a^T b = \sum_{i=1}^P a_i b_i$$

$$a^T b = \langle a, b \rangle$$

Definition 1.1.3 — Hilbert Space. The mathematical concept of a Hilbert space generalizes the notion of Euclidean space. It extends the methods of vector algebra and calculus from the two-dimensional Euclidean plane and three-dimensional space to spaces with any finite or infinite number of dimensions. A Hilbert space is an abstract vector space possessing the structure of an inner product that allows length and angle to be measured. Furthermore, Hilbert spaces are complete: there are enough limits in the space to allow the techniques of calculus to be used.

Hilbert Inner Product Space $\{\mathbb{R}^P, \langle a, b \rangle\}$

General Inner Product

Let $\Sigma \in \mathbb{R}^{P \times P}$ set of all $P \times P$ matrices. Assume Σ is a positive definite matrix.

$$x^T \Sigma x < 0$$

$$\forall x \in \mathbb{R}^P$$

$$x \neq 0$$

Then $a^T \Sigma b$ also satisfies the conditions for inner product.

$$a^T \Sigma b = \langle a, b \rangle_{\Sigma}$$

$$a^T b = a^T I b = \langle a, b \rangle_I$$

$\{\mathbb{R}^P, \langle, \rangle_{\Sigma}\}$ is a more general inner product space.

Linear Transformation

A matrix, $A, \in \mathbb{R}^{P \times P}$ can be viewed as linear transformation

$$T_A : \mathbb{R}^P \rightarrow \mathbb{R}^P, x \mapsto Ax$$



Bing Li will denote T_A as A .

\rightarrow means maps to for a domain.

\mapsto means maps to for a value.

\Rightarrow means implies.

If $A : \mathbb{R}^P \rightarrow \mathbb{R}^P$,

$$\ker(A) = \{x \in \mathbb{R}^P, Ax = 0\}$$

$$\text{ran}(A) = \{Ax : x \in \mathbb{R}^P\}$$

Definition 1.1.4 — Kernel. In linear algebra, the kernel, or sometimes the null space, is the set of all elements v of V for which $L(v) = 0$, where 0 denotes the zero vector in W .

In coordinate plane, think of a function that crosses the x -axis. The kernel would be all points on x where $y = 0$.

Definition 1.1.5 — Range. In coordinate plane, how much of the y axis is reached with the function? Now extend this idea to more dimensions.

A linear transformation is **idempotent** if

$$\begin{aligned} A &= A^2 \\ Ax &= A(A(x)) \\ \forall x \in \mathbb{R}^P \end{aligned}$$

If A were a number it could only be 1 or 0.

Wednesday August 24

Let $T \in \mathbb{R}^{P \times P}$ then there exists a unique operator $R \in \mathbb{R}^{P \times P}$ such that $\forall x, y \in \mathbb{R}^P$,

$$\langle x, Ty \rangle = \langle Rx, y \rangle$$

(general inner product, $a^T \Sigma b$). Aside: What this states is that if you give me any operator in the first you can find one in the second.

R is called the **adjoint operator** of T . Written as T^* , that is,

$$\langle x, Ty \rangle = \langle T^*x, y \rangle$$

Derived Facts

$$\begin{aligned} \langle x, Ty \rangle &= \langle T^*, y \rangle && \text{(by the definition)} \\ &= \langle y, T^*x \rangle && \text{(inner products the order doesn't matter)} \\ &= \langle (T^*)^*y, x \rangle && \text{(Use the definition again)} \\ &= \langle x, (T^*)^*y \rangle && \text{(swap order)} \end{aligned}$$

So, $T = (T^*)^*$.

It is easy to see in our case

$$\begin{aligned} \langle x, Ty \rangle_{\Sigma} &= x^T \Sigma Ty \\ &= x^T \Sigma T \Sigma^{-1} \Sigma y \\ &= (\Sigma^{-1} T^T \Sigma x)^T \Sigma y \\ &= \langle \Sigma^{-1} T^T \Sigma x, y \rangle_{\Sigma} \end{aligned}$$

So, $T^* = \Sigma^{-1} T^T \Sigma$ when $\Sigma = I_P$ (identity) and $T^* = T^T$.

Derived Facts

An operator is **self adjoint** if its adjoint is itself. (i.e. if $T = T^*$ or $\langle x, Ty \rangle = \langle Tx, y \rangle$). In the case of \langle, \rangle_Σ ,

$$T = \Sigma^{-1} T^T \Sigma$$

if

$$\Sigma = I_P, T = T^T$$



Self adjoint implies symmetric. It's a more general case, hence the use of Σ vs I . Useful to remember in following two Theorems

Theorem 1.1.1 If $A \in \mathbb{R}^{P \times P}$ is symmetric, then there exists **eigenvalue-eigenvector pairs**. $(\lambda_1, v_1), \dots, (\lambda_P, v_P)$ such that $v_1 \perp \dots \perp v_P$. Orthogonal basis (ONB) such that

$$A = \sum_{i=1}^P \lambda_i v_i v_i^T \text{ (spectral decomposition)}$$

More generally, if A is a linear operator in \mathcal{H} (finite dimensional inner product such as $(\mathbb{R}^P, \langle, \rangle_\Sigma)$). its eigen pair (linear operator now) (λ, v) is defined by

$$\begin{cases} Av = \lambda v \\ \langle v, v \rangle = 1 \end{cases}$$

Definition 1.1.6 — Orthogonal Basis. In the following, $(\mathbb{R}^P, \langle, \rangle_\Sigma) = \mathcal{H}$ (H for Hilbert)

ONB is defined by:

1. $v_i \perp v_j, \langle v_i, v_j \rangle = 0$
2. $\|v_i\| = 1$
3. $\text{span}\{v_1, \dots, v_P\} = \mathcal{H}$

Theorem 1.1.2 Suppose $A : \mathcal{H} \rightarrow \mathcal{H}$ is a self adjoint linear operator. Then A has eigen pairs: $(\lambda_1, v_1), \dots, (\lambda_P, v_P)$ where $\{v_1, \dots, v_P\}$ is ONB of \mathbb{R} such that

$$A = \sum_{i=1}^P \lambda_i v_i v_i^T \Sigma$$

Proof. (λ, v) is eigen pair of A , which means

$$Av = \lambda v$$

$$\langle v, v \rangle = 1$$

$$v^T \Sigma v = 1$$

Let $u = \Sigma^{\frac{1}{2}} v$.



Aside: $\Sigma^\alpha = \Sigma \lambda_i^\alpha v_i v_i^T$

Let $v = \Sigma^{-\frac{1}{2}}u$.

$$A\Sigma^{-\frac{1}{2}}u = \lambda\Sigma^{-\frac{1}{2}}u$$

$$\Sigma^{-\frac{1}{2}}u = \lambda u$$

So, (λ, v) is an eigen pair of A in $(\mathbb{R}, <, >_{\Sigma}) \Leftrightarrow (\lambda, u)$ '...' of $\Sigma^{\frac{1}{2}}A\Sigma^{-\frac{1}{2}}$ in $(\mathbb{R}, <, >_I)$.

Note that, A is self adjoint in $(\mathbb{R}, <, >_{\Sigma})$. So, $A = \Sigma^{-1}A^T\Sigma$

$$\begin{aligned}\Sigma^{\frac{1}{2}}A\Sigma^{-\frac{1}{2}} &= \Sigma^{\frac{1}{2}}A^T\Sigma\Sigma^{-\frac{1}{2}} \\ &= \Sigma^{-\frac{1}{2}}A\Sigma^{\frac{1}{2}} \\ &= (\Sigma^{\frac{1}{2}}A\Sigma^{-\frac{1}{2}})^T\end{aligned}$$

Note: $\Sigma^{\frac{1}{2}}A\Sigma^{-\frac{1}{2}}$ is symmetric!! So by Theorem 1.1, $\Sigma^{\frac{1}{2}}A\Sigma^{-\frac{1}{2}} = \sum \lambda_i v_i v_i^T$ where (λ_i, v_i) eigenpairs of $\Sigma^{\frac{1}{2}}A\Sigma^{-\frac{1}{2}}$.

That means $(\lambda_i, \Sigma^{\frac{1}{2}}v_i)$ are eigen pairs of A .

$$\text{So, } \Sigma^{\frac{1}{2}}A\Sigma^{-\frac{1}{2}} = \sum_{i=1}^P \Sigma^{\frac{1}{2}}u_i u_i^T \Sigma^{\frac{1}{2}} \Rightarrow A = \sum_{i=1}^P \lambda u_i u_i^T \Sigma$$

■

Definition 1.1.7 — Projection. If P is an operator in $(\mathbb{R}^P, <, >)$ then P is called a **projection** if it is both idempotent ($P = P^2$) and self adjoint ($P = P^*$).

Proposition 1.1 If A is a linear operator then $\ker(A) = \text{ran}(A^*)^{\perp}$

Proof. Take $x \in \ker(A) (\Rightarrow Ax = 0)$.

$$\begin{aligned}\forall y \in \text{ran}(A^*), x &\perp y \\ \Rightarrow x &\perp y \forall y = A^*z, z \in \mathbb{R}^P\end{aligned}$$

Hence,

$$\begin{aligned}\langle x, y \rangle &= \langle x, A^*z \rangle \\ &= \langle Ax, z \rangle \\ &= \langle 0, z \rangle \\ &= 0\end{aligned}$$

$$\begin{aligned}\Rightarrow x &\perp y \\ \Rightarrow x &\in \text{ran}(A^*)^{\perp}\end{aligned}$$

Or vice versa.

■

Friday August 26

 \perp means orthogonal complement.

$$\mathcal{S}^{\perp} = \{v \in \mathbb{R}^P, v \perp \mathcal{S}\}$$

$$v \perp w \forall w \in \mathcal{S}$$

$$\langle v, w \rangle = 0 \forall w \in \mathcal{S}$$

$$= \{v \in \mathbb{R}^P, \langle v, w \rangle = 0 \forall w \in \mathcal{S}\}$$

Recall, $\ker(A) = \text{ran}(A^*)^\perp$

So, if A is self adjoint then this is true and $\text{ran}(A)$ is also $\text{span}(A)$ which is the subspace spanned all columns of A .

Theorem 1.1.3 If P is a projection, then

1. $Pv = v, \forall v \in \text{ran}(P)$
 2. $Pv = 0, \forall v \perp \text{ran}(P)$
 3. If Q is another projections such that the $\text{ran}(Q) = \text{ran}(P)$ then $Q = P$. (The range determines the operator, because it is what decomposes the operator.)
- Asside: P acts like one on some spaces, and zero on orthogonal space.

Proof. 1. Let $v \in \text{ran}(P)$. Since $P^2 = P$ (idempotent) then

$$P^2v = Pv$$

$$\Rightarrow P^2v - Pv = 0$$

$$\Rightarrow P(Pv - v) = 0$$

$$\Rightarrow Pv - v \in \ker(P)$$

$$\Rightarrow Pv - v \perp \text{ran}(P)$$

$$\Rightarrow \langle Pv - v, Pv - v \rangle = 0$$

$$\Rightarrow \|Pv - v\| = 0$$

$$\Rightarrow Pv - v = 0$$

$$\Rightarrow Pv = v$$

2. If

$$v \perp \text{ran}(P)$$

$$\Rightarrow v \in \ker(P)$$

$$\Rightarrow Pv = 0$$

3. If Q is another operator with $\text{ran}(Q) = \text{ran}(P) = \mathcal{S}$ then $\forall v \in \mathcal{S}$

$$Qv = v = Pv \quad (\forall v \in \mathcal{S})$$

$$Qv = 0 = Pv$$

$$Qv = Pv \quad \forall v \in \mathcal{S}$$

$$Q = P$$

■

Theorem 1.1.4 Suppose \mathcal{S} is a subspace of \mathbb{R}^P , $R \ v_1, \dots, v_m$ is a basis of \mathcal{S} .

Let $v = (v_1, \dots, v_m) \in \mathbb{R}^{xM}$.

Then,

1. $A = v(v^T \Sigma v)^{-1} v^T \Sigma$ is a projection.

2. $\text{ran}(A) = \mathcal{S}$

Proof. 1. idempotent.

$$A^2 = v(v^T \Sigma v)^{-1} v^T \Sigma v(v^T \Sigma v)^{-1} v^T \Sigma$$

$$= v(v^T \Sigma v)^{-1} v^T \Sigma$$

$$= A$$

2. Self adjoint.

Let $x, y \in \mathbb{R}^P$

$$\begin{aligned}
 \langle x, Ay \rangle &= x^T \Sigma v (v^T \Sigma v)^{-1} v^T \Sigma y \\
 &= (v (v^T \Sigma v)^{-1} v^T \Sigma x)^T \Sigma y \\
 &= \langle Ax, y \rangle
 \end{aligned}$$

3. $\text{ran}(A) = \mathcal{S}$?Let $x \in \mathbb{R}^P$.

$$Ax = v (v^T \Sigma v)^{-1} v^T \Sigma x \in \text{span}(v) = \mathcal{S}$$

So let $x \in \mathcal{S}$,

$$x \in \text{ran}(v)$$

$$x = vy$$

for some $y \in \mathbb{R}^P$

$$= v (v^T \Sigma v)^{-1} v^T \Sigma vy$$

$$\in \text{ran}(A)$$

So, $\mathcal{S} \subseteq \text{ran}(A)$ and then $\mathcal{S} = \text{ran}(A)$. ■

We write A as $P_{\mathcal{S}}(\Sigma)$ (orthogonal projection on to \mathcal{S} with respect to Σ - product).

In the following, let $I : \mathbb{R}^P \rightarrow \mathbb{R}^P$ be the identity mapping. ($x \mapsto x$)

Let \mathcal{S} be a subspace in \mathbb{R}^P .

$$\text{Let } Q_{\mathcal{S}}(\Sigma) = I - P_{\mathcal{S}}(\Sigma)$$

Proposition 1.2 $Q_{\mathcal{S}}(\Sigma) = P_{\mathcal{S}^\perp}(\Sigma)$

Proof. Show $Q_{\mathcal{S}}(\Sigma)$ is projection.

1. Idempotent

$$\begin{aligned}
 Q_{\mathcal{S}}^2(\Sigma) &= Q_{\mathcal{S}}(\Sigma) Q_{\mathcal{S}}(\Sigma) \\
 &= (I - P_{\mathcal{S}}(\Sigma))(I - P_{\mathcal{S}}(\Sigma)) \\
 &= I - P_{\mathcal{S}}(\Sigma) - P_{\mathcal{S}}(\Sigma) + P_{\mathcal{S}} P_{\mathcal{S}} \\
 &= Q_{\mathcal{S}}(\Sigma)
 \end{aligned}$$

2. Self-adjoint

$$x, y \in \mathbb{R}^P$$

$$\begin{aligned}
 \langle x, Q_{\mathcal{S}}(\Sigma)y \rangle &= \langle x, (I - P_{\mathcal{S}}(\Sigma))y \rangle \\
 &= \langle x, y \rangle - \langle x, P_{\mathcal{S}}(\Sigma)y \rangle \\
 &= \langle x, y \rangle - \langle P_{\mathcal{S}}(\Sigma)x, y \rangle \\
 &= \langle (I - P_{\mathcal{S}}(\Sigma))x, y \rangle \\
 &= \langle Q_{\mathcal{S}}(\Sigma)x, y \rangle
 \end{aligned}$$

3. Range

$$\text{ran}(Q_{\mathcal{S}}(\Sigma)) = \mathcal{S}^\perp. \text{ Take } x \perp \mathcal{S} = \text{ran}(P_{\mathcal{S}}(\Sigma))^\perp = \text{ker}(P_{\mathcal{S}}(\Sigma)).$$

$$\Rightarrow P_{\mathcal{S}}(\Sigma) = 0$$

$$\Rightarrow Q_{\mathcal{S}}(\Sigma)x = x - P_{\mathcal{S}}(\Sigma)x = x$$

$$X \in \text{ran}(Q_{\mathcal{S}}(\Sigma))$$

$$\Rightarrow \mathcal{S}^{\perp} \subseteq \text{ran}(Q_{\mathcal{S}}(\Sigma))$$

Take $x \in \text{ran}(Q_{\mathcal{S}}(\Sigma))$, $\forall y \in \mathcal{S} = \text{ran}(P_{\mathcal{S}}(\Sigma))$

$$y = P_{\mathcal{S}}(\Sigma)z \text{ for some } z \in \mathbb{R}^P$$

$$\langle x, y \rangle = \langle x, P_{\mathcal{S}}(\Sigma)z \rangle = \langle P_{\mathcal{S}}(\Sigma)x, z \rangle = 0$$

$$\Rightarrow x \in \mathcal{S}^{\perp}$$

$$\Rightarrow \text{ran}(Q_{\mathcal{S}}(\Sigma)) = \mathcal{S}^{\perp}$$

■

1.2 Cochran's Theorem

This section will be about the distribution of the squared norm of a projection of a Gaussian random vector.

Proposition 1.3 If A is idempotent, then its eigenvalues are either 0 or 1.

Proof. λ is eigenvalue of A .

$$\Rightarrow Av = \lambda v (||v|| = 1)$$

$$\lambda = Av = A^2v = \lambda Av = \lambda^2$$

So, λ is 0 or 1.

■

Monday August 29

Lemma 1.1 Suppose $V \sim N(0, \sigma^2 I_P)$.

P is projection with I_P - inner product. Then $V^T P V \sim \sigma^2 \chi_S^2$ where $\text{df} = \text{rank}(P)$.

Proof. P is symmetric, and it has spectral decomposition,

$$A R A^T$$

where the A 's are orthogonal and R is diagonal with diagonal entries 0 or 1.

Then,

$$A^T V \sim N_P(0, A^T (\sigma^2 I_P) A) = N_P(0, \sigma^2 I_P)$$

Let,

$$Z = R A^T V$$

then,

$$Z \sim N_P(0, \sigma^2 R^2) = N_P(0, \sigma^2 R)$$

That means among the components of Z , some are distributed as $N(0, 1)$ and the rest are zero and they are independent. So,

$$Z^T Z \sim \chi_S^2 = V^T P V$$

■

Corollary 1.1 Suppose $X \sim N(0, \Sigma)$. Consider the Hilbert space $(\mathbb{R}^P, \langle, \rangle_{\Sigma^{-1}})$.

$$\langle a, b \rangle_{\Sigma^{-1}} = a^T \Sigma^{-1} b$$

Let \mathcal{S} be a subspace of \mathbb{R}^P and $P_{\mathcal{S}}(\Sigma^{-1})$ be the projection onto \mathcal{S} with respect to $\langle, \rangle_{\Sigma^{-1}}$ (special case of Fisher information inner product)

Then,

$$\|P_{\mathcal{S}}(\Sigma^{-1})x\|_{\Sigma^{-1}}^2 \sim \chi_r^2$$

where $r = \dim(\mathcal{S})$.

Proof. Let V be a basis matrix of \mathcal{S} (i.e. the col of V form basis in \mathcal{S}).

$$\begin{aligned} \|P_{\mathcal{S}}(\Sigma^{-1})x\|_{\Sigma^{-1}}^2 &= \langle P_{\mathcal{S}}(\Sigma^{-1})x, P_{\mathcal{S}}(\Sigma^{-1})x \rangle \\ &= x^T P_{\mathcal{S}}(\Sigma^{-1}) \Sigma^{-1} P_{\mathcal{S}}(\Sigma^{-1}) x \\ &= x^T (V(V^T \Sigma^{-1} V)^{-1} V^T \Sigma^{-1})^T \Sigma^{-1} (V(V^T \Sigma^{-1} V)^{-1} V^T \Sigma^{-1})^T x \\ &= x^T \Sigma^{-1} V(V^T \Sigma^{-1} V)^{-1} V^T \Sigma^{-1} V(V^T \Sigma^{-1} V)^{-1} V^T \Sigma^{-1} x \\ &= (\Sigma^{-\frac{1}{2}} x)^T \square \end{aligned}$$

But,

$$\Sigma^{-\frac{1}{2}} x \sim N(0, I_P)$$

$$\Sigma^{-\frac{1}{2}} V(V^T \Sigma^{-1} V)^{-1} V^T \Sigma^{-\frac{1}{2}}$$

is a projection with respect to I_P -inner product (idempotent, self adjoint, YES).

By Lemme 1.1, $\sim \chi_r^2$. ■

It is then easy to derive Cochran's Theorem. (see proof in Homework 1)

Theorem 1.2.1 Let $X \sim N(0, \Sigma)$ and $\mathcal{H} = \{\mathbb{R}^P, \langle, \rangle_{\Sigma^{-1}}\}$. Let $\mathcal{S}_1, \dots, \mathcal{S}_k$ be linear subspaces of \mathbb{R}^P such that $\mathcal{S}_i \perp \mathcal{S}_j$ in $\langle, \rangle_{\Sigma^{-1}}$

Let $r_i = \dim(\mathcal{S}_i)$.

Let $w_i = \|P_{\mathcal{S}_i}(\Sigma^{-1})X\|_{\Sigma^{-1}}^2$

Then,

1. $W_i \sim \chi_{r_i}^2$
2. $W_1 \perp, \dots, \perp W_k$ where \perp indicates independence.

1.3 Gaussian Linear Regression Model

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$X = \begin{pmatrix} x_{11} & \dots & x_{1P} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nP} \end{pmatrix} \in \mathbb{R}^{n \times P}$$

Consider the linear model,

$$y = X\beta + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

where X has full column rank ($n \geq p$).

Here X is treated as fixed

- Maximum Likelihood Estimator

$$E(y) = X\beta \in \mathbb{R}^n$$

$$\text{Var}(y) = \sigma^2 I_n$$

$$y \sim N_p(X\beta, \sigma^2 I_n)$$

Multivariate normal density

$$y \sim N(\mu, \Sigma)$$

$$f_Y(y) = \frac{1}{(2\pi)^{\frac{n}{2}} [\det(\Sigma)]^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(y - \mu)^T \Sigma^{-1}(y - \mu)\right)$$

In our case,

$$\Sigma = \sigma^2 I_n$$

$$\det(\Sigma) = \det(\sigma^2 I_n) = \sigma^{2n} \det(I_n) = \sigma^{2n}$$

So,

$$f_Y(y) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\sigma^{2n}}} \exp\left(-\frac{1}{2\sigma^2} \|y - \mu\|^2\right)$$

$$\log(f_Y(y)) = \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \|y - \mu\|^2 = \ell(\beta, \sigma^2, y)$$

$$\frac{\partial}{\partial \beta} = \dots = -\frac{1}{2\sigma^2} 2X^T(y - X\beta) = 0$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y \in \mathbb{R}^p$$

$$\frac{\partial}{\partial \sigma^2} \ell(\beta, \sigma^2, y) = \dots = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \|y - X\beta\|^2 = 0$$

$$\hat{\sigma}^2 = \frac{1}{n} \|y - X\hat{\beta}\|^2$$

In summary, the MLE for (β, σ^2) in Gaussian Linear Model are

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

$$\hat{\sigma}^2 = \frac{1}{n} \|y - X\hat{\beta}\|^2$$

Note that

$$X\hat{\beta} = X(X^T X)^{-1} X^T y = \hat{y}$$

So, $\hat{y} = P_{\text{span}(X)}(I_P) = P_X$.

Now,

$$\hat{\sigma}^2 = \frac{1}{n} \|(I_n - P_X)y\|^2 = \frac{1}{n} \|Q_X y\|^2$$

where $(I_n - P_X)$ is projection on to $\text{span}(X)^\perp$.

It turns out that $(X^T y, y^T y)$ is complete, sufficient statistic for this Gaussian linear model.

Wednesday August 31

Recall,

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

$$\hat{\sigma}^2 = \frac{1}{n} \|y - X\hat{\beta}\|^2$$

$$Q_X = I_n - P_X$$

$$P_X = X(X^T X)^{-1} X^T$$

Several properties,

$$E(\hat{\beta}) = \beta \text{ (unbiased)}$$

$$\text{Var}(\hat{\beta}) = (X^T X)^{-1} X^T \sigma^2 I_n X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1}$$

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$$

Because P_X has rank p and Q_X has rank $(n - p)$, then

$$\|Q_X y\|^2 \sim \chi_{(n-p)}^2$$

Let's find an unbiased estimator for σ^2

$$\begin{aligned} E(\hat{\sigma}^2) &= E\left(\frac{1}{n} \|Q_X y\|^2\right) \\ &= \frac{n-p}{n} \sigma^2 \end{aligned}$$

$$E\left(\frac{n}{n-p} \hat{\sigma}^2\right) = \sigma^2$$

Moreover, $\hat{\beta}$ has one-to-one transformation with

$$(X^T X)^{-1} X^T y \leftrightarrow X(X^T X)^{-1} X^T y = P_X y$$

$$\begin{aligned} \text{Cov}(P_X y, Q_X y) &= P_X \sigma^2 I_n Q_X \\ &= \sigma^2 P_X Q_X \\ &= 0 \end{aligned}$$

$$P_X y \perp\!\!\!\perp Q_X y \text{ (due to normality)}$$

$$\hat{\beta} \leftrightarrow P_X y$$

$$\hat{\sigma}^2 \text{ is a function of } Q_X y, \text{ so } \hat{\beta} \perp\!\!\!\perp \hat{\sigma}^2$$

In your homework, $\hat{\beta}, \hat{\sigma}^2 \leftrightarrow$ complete sufficient.

$\hat{\beta}, \hat{\sigma}^2$ is UMVUE (Lehmann-Sheffe).

Theorem 1.3.1 — Gaussian Regression Model. Under this model:

1. $\hat{\beta}, \tilde{\sigma}^2$ UMVUE for β, σ^2
2. $\hat{\beta} \sim N(\beta, \sigma^2(X^T X)^{-1})$
3. $(n-p)\tilde{\sigma}^2 \sim \sigma^2 \chi_{(n-p)}^2$
4. $\hat{\beta} \perp \tilde{\sigma}^2$

1.4 Statistical Inference for β, σ^2

Suppose we want to test

$$H_0 : \beta_1 = \beta_{i0}$$

$$\text{Let } M = (X^T X)^{-1}.$$

Then,

$$\hat{\beta} \sim N(\beta, \sigma^2 M)$$

where, $M_{ii} \leftarrow (i, i)^{th}$ entry of M

$$\text{Also, } \frac{(n-p)\tilde{\sigma}^2}{\sigma^2} \sim \chi_{(n-p)}^2$$

$$\hat{\beta} \perp \tilde{\sigma}^2$$

$$\frac{\frac{\hat{\beta}_i - \beta_{i0}}{\sqrt{\sigma^2 M_{ii}}}}{\sqrt{\frac{(n-p)\tilde{\sigma}^2/\sigma^2}{n-p}}} \sim t_{(n-p)}$$

$$T = \frac{\hat{\beta}_i - \beta_{i0}}{\sigma \sqrt{M_{ii}}} \sim t_{(n-p)} = (*)$$

Reject H_0 if

$$\left| \frac{\hat{\beta}_i - \beta_{i0}}{\sigma \sqrt{M_{ii}}} \right| > t_{\frac{\alpha}{2}, (n-p)}$$

Recall,

$$X \sim N(\mu, 1) \quad y \sim \chi_r^2 \quad X \perp y$$

$$\frac{X}{\sqrt{\frac{y}{r}}} \sim t_n(\mu)$$

Power at β_{i1}


$$\hat{\beta}_i \sim N(\beta_{i1}, \sigma^2 M_{i1})$$

So,

$$\frac{\hat{\beta}_i - \beta_{i0}}{\tilde{\sigma} \sqrt{M_{ii}}} \sim t_{(n-p)} \left(\frac{\beta_{i1} - \beta_{i0}}{\sigma \sqrt{M_{ii}}} \right)$$

(alternative distribution of T)

By this (*),



2. ANOVA (1-way)

2.1 Overview

- General linear models
- Scheffe's simultaneous confidence
- Singular decomposition
- Non Gaussian error

3. Mutiway ANOVA

3.1 Overview

- Orthogonal design
- Additive 2 way ANOVA
- simultaneous intervals
- nonadditive
- decomposition of sum of squares
- Latin square
- nested design

4. Nonorthogonal Design

4.1 Overview

- $\bar{X}_j - \bar{X}_i$



5. Random Effects Model

5.1 Overview



Part Two

6	Basic Concepts	31
6.1	Overview	
7	Estimation	33
7.1	Overview	
8	Inference	35
8.1	Overview	
9	Residuals	37
9.1	Overview	
10	Cetegorical Prediction	39
10.1	Overview	
11	Some Important GLM	41
11.1	Overview	
12	Multivariate GLM	43
12.1	Overview	



6. Basic Concepts

6.1 Overview



7. Estimation

7.1 Overview

8. Inference

8.1 Overview

- deviance \leftrightarrow sum of squares



9. Residuals

9.1 Overview



10. Categorical Prediction

10.1 Overview



11. Some Important GLM

11.1 Overview



12. Multivariate GLM

12.1 Overview



Part Three

13	Principle Component Analysis	47
13.1	Overview	
14	Canonical Correlation Analysis	49
14.1	Overview	
15	Independent Component Analysis . . .	51
15.1	Overview	
	Index	53



13. Principle Component Analysis

13.1 Overview



14. Canonical Correlation Analysis

14.1 Overview



15. Independent Component Analysis

15.1 Overview

Index

Cochran's Theorem, 14

Delete One Prediction, 19

Gaussian Linear Regression Model, 15

Overview, 21, 23, 25, 27, 31, 33, 35, 37, 39, 41,
43, 47, 49, 51

Projection, 7

Statistical Inference for β , σ^2 , 18