MAT 3375
Regression Analysis

Chapter 1
Preliminaries

P. Boily (uOttawa)

Summer – 2023

Outline

1 – Preliminaries

Regression analysis is not a very complicated discipline ... assuming that its pre-requisites are mastered well. In this course, it will be useful to be familiar with a number of notions relating to:

- **random variables**;
- **multivariate calculus**;
- **linear algebra**;
- **quadratic forms**, and
- **optimization**.

1.1 – Random Variables

A **random experiment** is a **process** (together with its **sample space** $\mathcal{S}$) for which it is impossible to predict the **outcome with certainty**. The **sample space** $\mathcal{S}$ is the set of the random experiment's **possible outcomes**.

A **random variable** Y associated to this process is a **function** $Y : \mathcal{S} \to \mathbb{R}$. If the set $Y(\mathcal{S}) = \{Y(s) \mid s \in \mathcal{S}\}$ is **countable**, we say that Y is a **discrete r.v.**; if it is **uncountable**, we say that Y is a **continuous r.v.**

Each r.v. Y has a corresponding **probability function** $f(Y)$, which specifies the **probabilities of the values taken by** Y.

Y_1 and Y_2 are **independent** when their **joint probability function** $f(Y_1, Y_2)$ is the product of the **individual probability functions** $f(Y_1)f(Y_2)$.

1.1.1 – Expectation, Variance, and Covariance

The **expectation operator** $\mathrm{E}\{\cdot\}$ is defined by

$$\mathrm{E}\{Y\} = \begin{cases} \displaystyle\sum_{Y(s)} Y(s) f(Y(s)), & \text{if } Y \text{ is discrete} \\ \displaystyle\int_{\mathbb{R}} Y f(Y)\, dy, & \text{if } Y \text{ is continuous} \end{cases}$$

The expectation $\mathrm{E}\{Y\}$ is the **average value** that we would expect to observe if the experiment is repeated a large number of times.

The expectation is sometimes also called the **mean** of Y, denoted $\overline{Y}$; it is thus a measure of Y's **centrality**.

The **variance operator** $\sigma^2\{\cdot\}$ is defined by

$$\sigma^2\{Y\} = \mathrm{E}\left\{(Y - \mathrm{E}\{Y\})^2\right\} = \mathrm{E}\{Y^2\} - \left(\mathrm{E}\{Y\}\right)^2.$$

It is often denoted by Var(Y). It is a measure of Y's **dispersion** (large variances are associated with r.v. with **heavy dispersion**, and *vice-versa*).

The **covariance operator** $\sigma\{\cdot,\cdot\}$ is defined by

$$\sigma\{Y, W\} = \mathrm{E}\{(Y - \mathrm{E}\{Y\})(W - \mathrm{E}\{W\})\} = \mathrm{E}\{YW\} - \mathrm{E}\{Y\}\mathrm{E}\{W\}.$$

It is often denoted by Cov(Y, W). It is a measure of the **strength of the linear relationship** between two r.v. (large covariance magnitudes are associated with **linearity**, but "large" is a relative concept).

The **standard deviation operator** $\sigma\left\{\cdot\right\}$ is defined by

$$\sigma\left\{Y\right\} = \sqrt{\sigma^2\left\{Y\right\}}.$$

It is always non-negative.

The **correlation operator** $\rho\left\{\cdot,\cdot\right\}$ is defined by

$$\rho\left\{Y,W\right\} = \frac{\sigma\left\{Y,W\right\}}{\sigma\left\{Y\right\}\sigma\left\{W\right\}},$$

assuming that $\sigma\left\{Y\right\}\sigma\left\{W\right\} \neq 0$.

When $\rho\left\{Y,W\right\} = 0$, we say that the r.v. are **uncorrelated**.

Properties of the Operators

Let Y, Y_i, W be r.v., and $a, b, c, a_i, b_i, c_i \in \mathbb{R}$, $i = 1, \ldots, n$. Then:

- $\mathrm{E}\left\{\cdot\right\}$ is **linear** on the space of r.v.: $\mathrm{E}\left\{aY + b\right\} = a\mathrm{E}\left\{Y\right\} + b$ and

$$\mathrm{E}\left\{\sum_{i=1}^{n} a_i Y_i\right\} = \sum_{i=1}^{n} a_i \mathrm{E}\left\{Y_i\right\}$$

- $\sigma^2\left\{aY + b\right\} = a^2\sigma^2\left\{Y\right\}$ and

$$\sigma^2\left\{\sum_{i=1}^{n} a_i Y_i\right\} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \sigma\left\{Y_i, Y_j\right\} = \sum_{i=1}^{n} a_i^2\sigma^2\left\{Y_i\right\} + \sum_{i\neq j} a_i a_j \sigma\left\{Y_i, Y_j\right\}$$

- $\sigma\{Y,Y\} = \sigma^2\{Y\}$ and $\sigma\{Y,W\} = \sigma\{W,Y\}$
- $\sigma\{a_1Y+b_1, a_2W+b_2\} = a_1a_2\sigma\{Y,W\}$
- $\{Y_i\}$ **uncorrelated** $\implies$

$$\sigma\left\{\sum_{i=1}^n a_iY_i, \sum_{i=1}^n c_iY_i\right\} = \sum_{i=1}^n a_ic_i\sigma^2\{Y_i\}$$

- $\sigma\{Y,W\} < 0 \iff$ observations of Y above $\overline{Y}$ tend to accompany corresponding observations of W below $\overline{W}$, and *vice-versa*.
- $\sigma\{Y,W\} > 0 \iff$ observations of Y above $\overline{Y}$ tend to accompany corresponding observations of W above $\overline{W}$, and *vice-versa*.

- $\sigma\{Y,W\}=0 \implies Y$ and W are **uncorrelated**
- Y,W **independent** $\implies \rho\{Y,W\}=0$ (uncorrelated)
- $\rho\{Y,W\}=0 \not\Longrightarrow Y,W$ **independent**, however
- $|\rho\{Y,W\}|\leq 1$ (consequence of the Cauchy-Schwartz inequality)
- $|\rho\{Y,W\}|=1 \Longleftrightarrow Y=aW+b$ for some $a,b\in\mathbb{R}$,

Random Vectors

If $Y_1, \ldots, Y_n$ are random variables, then

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$$

is a **random vector**. The **expectation** of $\mathbf{Y}$ is

$$\mathrm{E}\{\mathbf{Y}\} = \begin{pmatrix} \mathrm{E}\{Y_1\} \\ \vdots \\ \mathrm{E}\{Y_n\} \end{pmatrix}.$$

The components of $\mathbf{Y}$ need not all have identical distributions.

The **variance-covariance matrix** of $\mathbf{Y}$ is the symmetric matrix

$$\sigma^2\{\mathbf{Y}\} = (g_{i,j}), \quad \text{where } g_{i,j} = \begin{cases} \sigma^2\{Y_i\} & i=j \\ \sigma\{Y_i, Y_j\} & i \neq j \end{cases}$$

or

$$\sigma^2\{\mathbf{Y}\} = \begin{pmatrix} \sigma^2\{Y_1\} & \cdots & \sigma\{Y_1, Y_n\} \\ \vdots & \ddots & \vdots \\ \sigma\{Y_1, Y_n\} & \cdots & \sigma^2\{Y_n\} \end{pmatrix}$$

If the components of $\mathbf{Y}$ are **independent** and all have the **same variance** σ^2, then

$$\sigma^2\{\mathbf{Y}\} = \sigma^2\mathbf{I}_n.$$

In practice, we usually work with **samples** of the random variables. Let $\{(X_i, Y_i)\}_{i=1}^n$ be observed from the joint distribution of (X, Y):

- $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $\overline{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$, the **sample means**, are unbiased estimators of $\mathrm{E}\{X\}$ and $\mathrm{E}\{Y\}$, respectively;
- $s_X^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ and $s_Y^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y})^2$, the **sample variances**, are unbiased estimators of $\sigma^2\{X\}$ and $\sigma^2\{Y\}$, respectively;
- $s_{XY} = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})$, the **sample covariance**, is an unbiased estimator of $\sigma\{X, Y\}$.

1.1.2 – Important Distributions

The **(cumulative) distribution function** (c.d.f.) of any continuous random variable Y is defined by

$$F_Y(y) = P(Y \leq y) = \int_{-\infty}^{y} f_Y(t)\, dt$$

viewed as a function of a real variable y. Alternatively, We can describe the **distribution** of Y *via* the following relationship between $f_Y(y)$ and $F_Y(y)$:

$$f_Y(y) = \frac{d}{dy} F_Y(y).$$

Probability Density Function

The **probability density function** (p.d.f.) of a continuous random variable Y is an **integrable** function $f_Y : Y(\mathcal{S}) \to \mathbb{R}$ such that

- $f_Y(y) > 0$ for all $y \in Y(\mathcal{S})$ and $\lim_{y\to\pm\infty} f_Y(y) = 0$;
- $\int_{\mathcal{S}} f_Y(y)\,dy = 1$;
- for any a, b,

$$P(a < Y < b) = P(a \leq Y < b) = P(a < Y \leq b) = P(a \leq Y \leq b)$$
$$= F_Y(b) - F_Y(a) = \int_a^b f(y)\,dy.$$

Normal distribution: the c.d.f. of the r.v. $Y \sim \mathcal{N}(\mu, \sigma^2)$ is

$$F_Y(y) = P(Y \leq y) = \Phi(y),$$

with

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

χ^2 **distribution:** the p.d.f. of the r.v. $Y \sim \chi^2(\nu)$ is

$$f_Y(y;\nu) = \begin{cases} \dfrac{y^{\frac{\nu}{2}-1} e^{-\frac{y}{2}}}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)}, & y > 0; \\ 0, & \text{otherwise.} \end{cases}$$

where $\Gamma(\cdot)$ is the **Gamma function**.

If $U_i \sim \chi^2(\nu_i)$, $i = 1, 2$, and U_1, U_2 are independent, then

$$U = U_1 + U_2 \sim \chi^2(\nu_1) + \chi^2(\nu_2) = \chi^2(\nu_1 + \nu_2).$$

There is an important link between the standard normal distribution and the $\chi^2(1)$ distribution: if $Z \sim \mathcal{N}(0, 1)$, then $Z^2 \sim \chi^2(1)$.

Student's distribution: if $Z \sim \mathcal{N}(0, 1)$ and $U \sim \chi^2(\nu)$, Z, U independent:

$$t = \frac{Z}{\sqrt{U/\nu}} \sim t(\nu),$$

the **Student T-distribution with ν degrees of freedom**.

Fisher's distribution: if $U_i \sim \chi^2(\nu_i)$, $i = 1, 2$ and U_1, U_2 are independent:

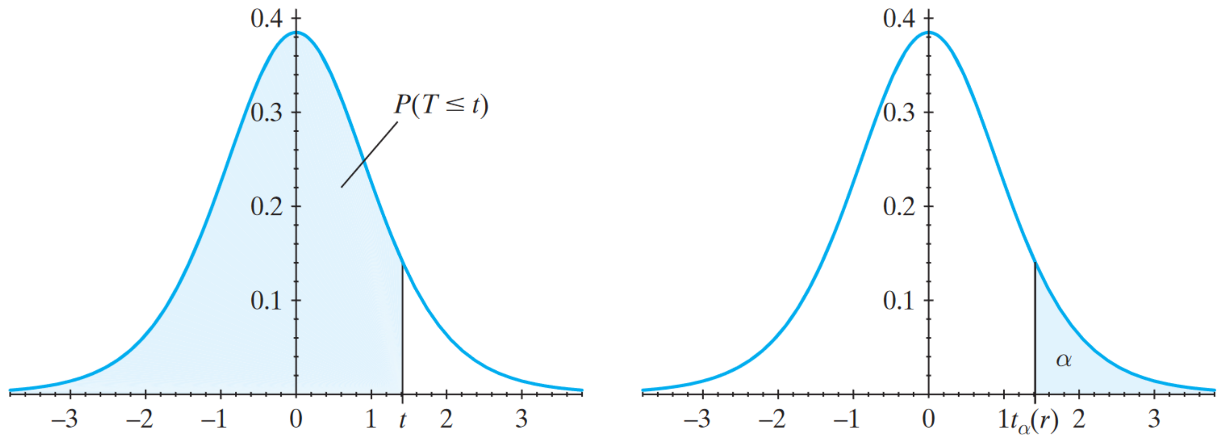
$$F = \frac{U_1/\nu_1}{U_2/\nu_2} \sim F(\nu_1, \nu_2),$$

the **Fisher's distribution with ν_1 and ν_2 degrees of freedom**.

Familiarize yourself with the c.d.f. tables and the corresponding `R` functions

- `qt()`, `dt()`, `pt()`, `rt()`, and
- `qf()`, `df()`, `pf()`, `rf()`.

Table VI The *t* Distribution



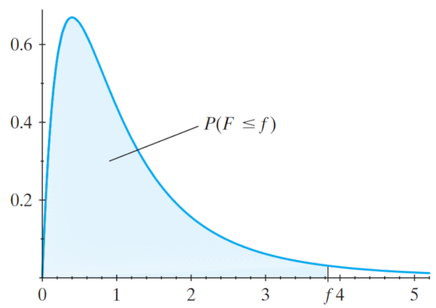
$$P(T \leq t) = \int_{-\infty}^{t} \frac{\Gamma[(r+1)/2]}{\sqrt{\pi r}\,\Gamma(r/2)(1+w^2/r)^{(r+1)/2}}\,dw$$

$$P(T \leq -t) = 1 - P(T \leq t)$$

	$P(T \leq t)$						
	0.60	0.75	0.90	0.95	0.975	0.99	0.995
r	$t_{0.40}(r)$	$t_{0.25}(r)$	$t_{0.10}(r)$	$t_{0.05}(r)$	$t_{0.025}(r)$	$t_{0.01}(r)$	$t_{0.005}(r)$
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841

Table VII The F Distribution

$$P(F \le f) = \int_0^f \frac{\Gamma[(r_1+r_2)/2](r_1/r_2)^{r_1/2} w^{r_1/2-1}}{\Gamma(r_1/2)\Gamma(r_2/2)(1+r_1 w/r_2)^{(r_1+r_2)/2}}\, dw$$



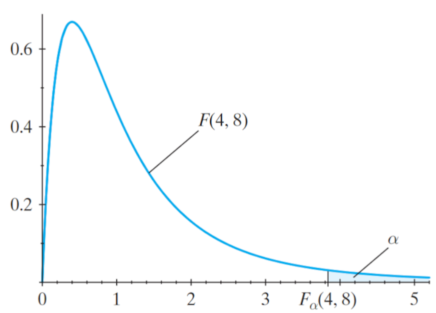


Table VII *continued*

$$P(F \le f) = \int_0^f \frac{\Gamma[(r_1+r_2)/2](r_1/r_2)^{r_1/2} w^{r_1/2-1}}{\Gamma(r_1/2)\Gamma(r_2/2)(1+r_1 w/r_2)^{(r_1+r_2)/2}}\, dw$$

		Den. d.f.	Numerator Degrees of Freedom, r_1									
α	$P(F \le f)$	r_2	1	2	3	4	5	6	7	8	9	10
0.05	0.95	1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9
0.025	0.975		647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28	968.63
0.01	0.99		4052	4999.5	5403	5625	5764	5859	5928	5981	6022	6056
0.05	0.95	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
0.025	0.975		38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40
0.01	0.99		98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40
0.05	0.95	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
0.025	0.975		17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42
0.01	0.99		34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23
0.05	0.95	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
0.025	0.975		12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84
0.01	0.99		21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55
0.05	0.95	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
0.025	0.975		10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62
0.01	0.99		16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05
0.05	0.95	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
0.025	0.975		8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46
0.01	0.99		13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87
0.05	0.95	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
0.025	0.975		8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76
0.01	0.99		12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62

Central Limit Theorem

Theorem: let $X_1, \ldots, X_n$ be independent normal random variables with mean $\mu_1, \ldots, \mu_n$ and standard deviations $\sigma_1, \ldots, \sigma_n$. Then

$$X_1 + \cdots + X_n \sim \mathcal{N}(\mu_1 + \cdots + \mu_n, \sigma_1^2 + \cdots + \sigma_n^2).$$

If $\mu_i \equiv \mu$ and $\sigma_i^2 \equiv \sigma$ for $i = 1, \ldots, n$, then $X_1 + \cdots + X_n \sim \mathcal{N}(n\mu, n\sigma^2)$.

Theorem: let $X_1, \ldots, X_n$ be independent normal random variables with mean μ and standard deviation σ. Let $\overline{X}$ be the sample mean. Then

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

Theorem: let $X_1, \ldots, X_n$ be independent random variables with mean μ and standard deviation σ. Let $\overline{X}$ be the sample mean. Then

$$Z_n = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \to Z \sim \mathcal{N}(0,1), \quad \text{as } n \to \infty.$$

Theorem: let $X_1, \ldots, X_n$ be independent normal random variables with mean μ and common variance. Let $\overline{X}$ and s^2 be the sample mean and the sample variance, respectively. Then the random variable

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1),$$

follows a **Student t-distribution with $\nu = n-1$ degrees of freedom**.

1.2 – Multivariate Calculus

Let $f:\mathbb{R}^n\to\mathbb{R}$ be a **differentiable** function. If $\mathbf{Y}=(Y_1,\ldots,Y_n)$, the **derivative** of f with respect to $\mathbf{Y}$ is

$$\nabla_{\mathbf{Y}}f(\mathbf{Y})=\begin{pmatrix}\dfrac{\partial f(\mathbf{Y})}{\partial Y_1}\\ \vdots\\ \dfrac{\partial f(\mathbf{Y})}{\partial Y_n}\end{pmatrix}.$$

The gradient is a **linear operator**:

$$\nabla_{\mathbf{Y}}(af+bg)(\mathbf{Y})=a\nabla_{\mathbf{Y}}f(\mathbf{Y})+b\nabla_{\mathbf{Y}}g(\mathbf{Y}).$$

If $f(\mathbf{Y})\equiv a$, then $\nabla_{\mathbf{Y}}f(\mathbf{Y})=\mathbf{0}$. If $f(\mathbf{Y})=\mathbf{Y}^\top\mathbf{v}$, then $\nabla_{\mathbf{Y}}f(\mathbf{Y})=\mathbf{v}$.

1.3 – Matrix Algebra

Let $A \in M_{m,n}(\mathbb{R})$ and $\mathbf{Y}$ be a random vector. Consider $\mathbf{W} = A\mathbf{Y}$. Then

$$\mathrm{E}\left\{\mathbf{W}\right\} = A\mathrm{E}\left\{\mathbf{Y}\right\} \quad \text{and} \quad \sigma^2\{\mathbf{W}\} = A\sigma^2\{\mathbf{Y}\}A^\top.$$

Furthermore, if $\mathbf{Y} \sim \mathcal{N}\left(\mathrm{E}\left\{\mathbf{Y}\right\}, \sigma^2\{\mathbf{Y}\}\right)$, then

$$\mathbf{W} \sim \mathcal{N}\left(\mathrm{E}\left\{\mathbf{W}\right\}, \sigma^2\{\mathbf{W}\}\right) = \mathcal{N}\left(A\mathrm{E}\left\{\mathbf{Y}\right\}, A\sigma^2\{\mathbf{Y}\}A^\top\right).$$

If $A \in M_{n,n}(\mathbb{R})$, the **trace** of A is $\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii} = a_{11} + a_{22} + \cdots + a_{nn}$.

The trace is a **linear operator**: $\mathrm{tr}(kA + B) = k\,\mathrm{tr}\,A + \mathrm{tr}\,B$; we also have $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ (when the matrices are **compatible**).

The **transpose** of a matrix A, denoted by $A^\top$, is obtained by interchanging its **rows** and its **columns**, or simply by **reflecting** the matrix along its **primary diagonal**.

Properties: if $A \in M_{m,n}(\mathbb{R})$ and $k \in \mathbb{R}$, then

- $(A^\top)^\top = A$
- $k^\top = k$
- $(kA + B)^\top = kA^\top + B^\top$
- $(AB)^\top = B^\top A^\top$

1.4 – Quadratic Forms

A **symmetric quadratic form** in $Y_1, \ldots, Y_n$ is an expression of the form

$$Q_A(\mathbf{Y}) = \mathbf{Y}^\top A \mathbf{Y} = \sum_{i,j=1}^{n} a_{i,j} Y_i Y_j,$$

where A is an $n \times n$ **symmetric matrix** ($A^\top = A$).

A number of important quantities in regression analysis can be expressed as symmetric quadratic forms.

The **degrees of freedom** for a symmetric quadratic form $Q_A(\mathbf{Y})$ can be obtained by computing the **rank** of the associated matrix A.

For example, the symmetric matrix associated with the symmetric quadratic form $Q_A(\mathbf{Y}) = 4Y_1^2 + 7Y_1Y_2 + 2Y_2^2$ is

$$A = \begin{pmatrix} 4 & 7/2 \\ 7/2 & 2 \end{pmatrix}; \quad Q_A \text{ has } 2 \text{ degrees of freedom.}$$

Theorem: let $Q_1, \ldots Q_K$ be symmetric quadratic forms of $\mathbf{Y}$ with respective symmetric matrices $A_1, \ldots, A_K$. If $a_i \in \mathbb{R}$ for $i = 1, \ldots, K$, then

$$Q = a_1Q_1 + \cdots + a_KQ_K$$

is a symmetric quadratic form of $\mathbf{Y}$ with symmetric matrix

$$A = a_1A_1 + \cdots + a_KA_K.$$

For a general $n \times n$ matrix B, we have

$$\nabla_{\mathbf{Y}} \left(\mathbf{Y}^{\top} B \mathbf{Y} \right) = (B^{\top} + B)\mathbf{Y}.$$

Thus the gradient of a symmetric quadratic form $Q_A(\mathbf{Y})$ is

$$\nabla_{\mathbf{Y}} Q_A(\mathbf{Y}) = 2A\mathbf{Y}.$$

It can be shown that **every** expression of the form $\mathbf{Y}^{\top} B \mathbf{Y}$ can be associated to a symmetric matrix A. So we assume every such form is symmetric.

The role played by quadratic forms in multi-variable calculus is analogous to the role played by $f(x) = ax^2$ in calculus.

The **eigenvalues** of an $n \times n$ matrix A are the roots of the **characteristic polynomial** $p_A(\lambda)$ of A: $p_A(\lambda) = \det(A - \lambda \mathbf{I}_n) = 0$.

There are n such (complex) roots, not necessarily distinct.

If λ is an eigenvalue of A, then there exists $\mathbf{v} \neq \mathbf{0}$ such that $A\mathbf{v} = \lambda\mathbf{v}$. If A is symmetric, all of its eigenvalues are **real**.

Consider a quadratic form $Q_A(\mathbf{Y})$, with eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$.

- If $\lambda_i > 0$ for all i, we say that $Q_A(\mathbf{Y})$ and A are **positive definite**;
- If $\lambda_i < 0$ for all i, we say that $Q_A(\mathbf{Y})$ and A are **negative definite**;
- If $\lambda_i\lambda_j < 0$ for some i, j, we say that $Q_A(\mathbf{Y})$ and A are **indefinite**.

1.4.1 – Cochran's Theorem

Let $\mathbf{Y} = (Y_1, \ldots, Y_n) \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$.

Suppose that

$$\mathbf{Y}^\top \mathbf{Y} = Q_1(\mathbf{Y}) + \cdots + Q_K(\mathbf{Y}),$$

where the Q_k are positive (semi-)definite quadratic forms with $r_k (= \text{rank}(A_k))$ degrees of freedom, $k = 1, \ldots, K$.

If $r_1 + \cdots + r_K = n$, then $Q_1(\mathbf{Y}), \ldots, Q_K(\mathbf{Y})$ are **independent** random variables and

$$\frac{Q_k(\mathbf{Y})}{\sigma^2} \sim \chi^2(r_k), \quad k = 1, \ldots, K.$$

In particular, if $K = 2$ and $r_1 = r$, then $Q_2(\mathbf{Y})/\sigma^2 \sim \chi^2(n - r)$.

1.4.2 – Important Quadratic Forms

For any positive integer n, we define two **special matrices**:

$$\mathbf{J}_n = \mathbf{J} = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{1}_{n\times 1} = \mathbf{1}_n = \mathbf{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Note that $\mathbf{1}_n^\top \mathbf{1}_n = n$ and $\mathbf{1}_n \mathbf{1}_n^\top = \mathbf{J}_n$. Let $\mathbf{Y} = (Y_1, \ldots, Y_n) \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ be a random vector. What are the symmetric matrices associated with:

$$Q_A(\mathbf{Y}) = \sum_{i=1}^n Y_i^2, \quad Q_B(\mathbf{Y}) = n\overline{Y}^2, \quad \text{and} \quad Q_C(\mathbf{Y}) = \sum_{i=1}^n (Y_i - \overline{Y})^2?$$

We re-write the quadratic forms in $\mathbf{Y}$ to obtain (next page):

$$Q_A(\mathbf{Y}) = \mathbf{Y}^\top\mathbf{Y} = \mathbf{Y}^\top\mathbf{I}_n\mathbf{Y} \implies A = \mathbf{I}_n;$$

$$Q_B(\mathbf{Y}) = n\left(\frac{1}{n}\sum_{i=1}^n Y_i\right)^2 = \frac{1}{n}\sum_{i,j=1}^n Y_iY_j = \frac{1}{n}\mathbf{Y}^\top\mathbf{1}_n\mathbf{1}_n^\top\mathbf{Y} \implies B = \frac{1}{n}\mathbf{J}_n;$$

$$Q_C(\mathbf{Y}) = \sum_{i=1}^n Y_i^2 - n\overline{Y}^2 = \mathbf{Y}^\top\mathbf{I}_n\mathbf{Y} - \frac{1}{n}\mathbf{Y}^\top\mathbf{J}_n\mathbf{Y} \implies C = \mathbf{I}_n - \frac{1}{n}\mathbf{J}_n.$$

Since $\mathrm{rank}(A) = n$, $\mathrm{rank}(B) = 1$, and $\mathrm{rank}(C) = n-1$, Cochran's Theorem implies that $Q_B(\mathbf{Y})$, and $Q_C(\mathbf{Y})$ are **independent** r.v., and that

$$\frac{Q_A(\mathbf{Y})}{\sigma^2} = \frac{\mathbf{Y}^\top\mathbf{Y}}{\sigma^2} \sim \chi^2(n),\ \frac{Q_B(\mathbf{Y})}{\sigma^2} = \frac{n\overline{Y}^2}{\sigma^2} \sim \chi^2(1), \frac{Q_C(\mathbf{Y})}{\sigma^2} = \frac{\mathrm{SST}}{\sigma^2} \sim \chi^2(n-1).$$

1.5 – Optimization

Let A be a symmetric $n \times n$ matrix, $\mathbf{v} \in \mathbb{R}^n$, $c \in \mathbb{R}$. Consider the function

$$f(\mathbf{Y}) = \frac{1}{2}\mathbf{Y}^\top A\mathbf{Y} - \mathbf{Y}^\top \mathbf{v} + c.$$

Note that f is **differentiable**. The **critical points** of f satisfy

$$\nabla_{\mathbf{Y}} f(\mathbf{Y}) = A\mathbf{Y} - \mathbf{v} = \mathbf{0} \implies A\mathbf{Y} = \mathbf{v}.$$

If A is **invertible** ($\det(A) \neq 0$), there is a **unique** critical point $\mathbf{Y}^* = A^{-1}\mathbf{v}$.

If A is **singular** ($\det(A) = 0$), there is **no** critical point (if $\mathbf{v} \notin \mathbf{range}(A)$) or **infinitely many** critical points (if $\mathbf{v} \in \mathbf{range}(A)$).

When A is **invertible**:

- if A is **positive definite**, then f reaches its **global minimum** at $\mathbf{Y}^* = A^{-1}\mathbf{v}$;
- if A is **negative definite**, then f reaches its **global maximum** at $\mathbf{Y}^* = A^{-1}\mathbf{v}$;
- if A is **indefinite** (if A has positive **and** negative eigenvalues), then $\mathbf{Y}^* = A^{-1}\mathbf{v}$ is a **saddle point** for f.

If the eigenvalues can be **zero**, we replace "definite" by "semi-definite" throughout.