

FRM® PART I · BOOK 2

Quantitative Analysis

Detailed Formula Sheet

Every formula with variable definitions, intuition, worked examples and exam traps · Prepared by **RBei Classes**, Noida

What this book tests. Quantitative Analysis carries about 20% of the Part I exam and is the most *calculation-heavy* book after Valuation & Risk Models. Expect numericals on probability & Bayes, distributions, covariance/correlation, hypothesis tests, regression (t , R^2 , adjusted R^2 , F), AR/MA processes, returns & the Jarque–Bera test, and a few machine-learning calculations (Q-values, confusion matrix, rescaling).

How to read this sheet. Teal boxes = formulas Green boxes = fully worked examples (every step verified) Orange boxes = exam traps Purple boxes = must-know facts & critical values.

Notation. μ = population mean; $\bar{X}$ or $\hat{\mu}$ = sample mean; σ^2 / s^2 = population / sample variance; n or T = number of observations; k = number of slope coefficients (regressors); α = significance level (or intercept, from context).

Contents

- | | |
|---|--|
| 1. Fundamentals of Probability | 9. Regression Diagnostics |
| 2. Random Variables | 10. Stationary Time Series (AR, MA, ARMA) |
| 3. Common Univariate Distributions | 11. Non-Stationary Time Series |
| 4. Multivariate Random Variables | 12. Returns, Volatility & Correlation |
| 5. Sample Moments | 13. Simulation & Bootstrapping |
| 6. Hypothesis Testing | 14. Machine Learning Methods |
| 7. Linear Regression | 15. Machine Learning & Prediction |
| 8. Regression with Multiple Explanatory Variables | 16. Critical-value tables · 17. One-page recap |

1. Fundamentals of Probability

Events, independence, mutual exclusivity, conditional probability, total probability, Bayes' rule

1.1 Core probability rules

$$0 \leq P(A) \leq 1, \quad P(A^c) = 1 - P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (\text{addition rule, "A or B"})$$

$$P(A \cap B) = P(A | B)P(B) = P(B | A)P(A) \quad (\text{multiplication rule, joint})$$

$$P(A | B) = \frac{P(A \cap B)}{P(B)} \quad (\text{conditional probability})$$

Independent events: $P(A \cap B) = P(A)P(B)$ and $P(A | B) = P(A)$. For n independent events: $P(A_1 \cap \dots \cap A_n) = \prod P(A_i)$.

Mutually exclusive events: $P(A \cap B) = 0 \Rightarrow P(A \cup B) = P(A) + P(B)$.

Conditionally independent: $P(A \cap B | C) = P(A | C)P(B | C)$ — independence and conditional independence do *not* imply each other.

1.2 Total probability rule & Bayes' rule

If $B_1, \dots, B_n$ are mutually exclusive and exhaustive:

$$P(A) = \sum_{i=1}^n P(A | B_i)P(B_i)$$

$$P(B_i | A) = \frac{P(A | B_i)P(B_i)}{P(A)} = \frac{P(A | B_i)P(B_i)}{\sum_j P(A | B_j)P(B_j)}$$

Bayes updates a *prior* probability $P(B_i)$ into a *posterior* $P(B_i | A)$ once evidence A is observed. Numerator = joint probability; denominator = total probability of the evidence.

Worked example — Addition rule

$P(\text{stock A rises}) = 60\%$, $P(\text{stock B rises}) = 55\%$, $P(\text{both}) = 40\%$. $P(A \text{ or } B) = 60 + 55 - 40 = \mathbf{75\%}$ (not 115%).

Worked example — Bayes' rule (probability tree)

$P(\text{economy outperforms}) = 0.6$, $P(\text{underperforms}) = 0.4$; $P(\text{stock up} | \text{out}) = 0.7$, $P(\text{up} | \text{under}) = 0.2$.

Joint: $0.6 \times 0.7 = 0.42$ and $0.4 \times 0.2 = 0.08 \Rightarrow P(\text{up}) = 0.42 + 0.08 = 0.50$ (total probability).

$P(\text{out} | \text{up}) = 0.42/0.50 = \mathbf{84\%}$ (prior 60% revised upward after seeing the stock rise).

Worked example — Counting table (1,000 cars)

600 blue (150 with driver-assist, DA), 400 red (200 with DA).

$P(DA | B) = 150/600 = 25\%$; $P(DA | R) = 200/400 = 50\%$; $P(DA) = 0.25(0.6) + 0.5(0.4) = 35\%$.

$P(B | DA) = P(B \cap DA)/P(DA) = 0.15/0.35 = \mathbf{42.9\%}$. Check by counting: $150/350 = 42.9\%$ ✓

EXAM TRAP: Mutually exclusive events are **dependent** (if one happens the other cannot). “Independent” and “mutually exclusive” are different ideas. The addition rule $P(A) + P(B) - P(AB)$ holds for *all* events.

2. Random Variables

PMF, CDF, PDF, expectation, the four moments, quantiles, linear transformations

2.1 PMF, CDF, PDF & quantile function

PMF (discrete): $f(x) = P(X = x)$, $\sum_x f(x) = 1$. **CDF**: $F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$ (discrete) $= \int_{-\infty}^x f(t) dt$ (continuous).

PDF (continuous): $P(a \leq X \leq b) = \int_a^b f(x) dx$; $P(X = x) = 0$ for any single point.

Quantile function $Q(\alpha) = F^{-1}(\alpha)$: the value below which α of outcomes lie. Median $= Q(0.5)$; IQR $= Q(0.75) - Q(0.25)$.

Bernoulli PMF: $f(x) = p^x(1-p)^{1-x}$, $x \in \{0, 1\}$.

2.2 Expectation & the four population moments

$$E(X) = \mu = \sum_i x_i P(x_i) \quad (\text{discrete}); \quad E(cX) = cE(X), \quad E(X+Y) = E(X) + E(Y)$$

$$\sigma^2 = E[(X - \mu)^2] = E(X^2) - [E(X)]^2 \quad \sigma = \sqrt{\sigma^2}$$

$$\text{Skewness} = \frac{E[(X - \mu)^3]}{\sigma^3} \quad \text{Kurtosis} = \frac{E[(X - \mu)^4]}{\sigma^4} \quad \text{Excess kurtosis} = \text{Kurtosis} - 3$$

Normal distribution: skew = 0, kurtosis = 3 (excess 0). Positive (right) skew: mode < median < mean. Negative (left) skew: mean < median < mode. Kurtosis > 3 = leptokurtic / fat-tailed; < 3 = platykurtic.

2.3 Linear transformation $Y = a + bX$

Mean $E(Y) = a + bE(X)$

Variance / s.d. $\sigma_Y^2 = b^2 \sigma_X^2$, $\sigma_Y = |b| \sigma_X$ (a has no effect on dispersion)

Skewness unchanged if $b > 0$; same magnitude, **sign flips** if $b < 0$

Kurtosis unchanged

Median / IQR median_Y = $a + b$ median_X; IQR_Y = $|b|$ IQR_X

Worked example — Expected value & variance of EPS

EPS outcomes (probability): 1.80 (0.10), 1.60 (0.20), 1.20 (0.40), 1.00 (0.30).

$E(\text{EPS}) = 0.1(1.80) + 0.2(1.60) + 0.4(1.20) + 0.3(1.00) = \mathbf{1.28}$.

$E(X^2) = 0.1(3.24) + 0.2(2.56) + 0.4(1.44) + 0.3(1.00) = 1.712 \Rightarrow \sigma^2 = 1.712 - 1.28^2 = \mathbf{0.0736}$, $\sigma = 0.2713$.

Check via deviations: $0.1(0.52)^2 + 0.2(0.32)^2 + 0.4(0.08)^2 + 0.3(0.28)^2 = 0.02704 + 0.02048 + 0.00256 + 0.02352 = 0.0736$ ✓

Worked example — Linear transformation

$E(X) = 5$, $\sigma_X^2 = 4$, skew_X = +0.8. For $Y = 2 - 3X$: $E(Y) = 2 - 15 = -13$; $\sigma_Y^2 = 9 \times 4 = \mathbf{36}$; $\sigma_Y = \mathbf{6}$; skew_Y = -0.8; kurtosis unchanged.

3. Common Univariate Random Variables

Uniform, Bernoulli, binomial, Poisson, normal, lognormal, χ^2 , t , F , exponential, beta, mixtures

3.1 Discrete & continuous distributions — summary

Distribution	PMF / PDF	Mean	Variance
Uniform [a, b]	$f(x) = \frac{1}{b-a}$; $P(x_1 \leq X \leq x_2) = \frac{x_2 - x_1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Bernoulli	$p^x(1-p)^{1-x}$	p	$p(1-p)$ (max at $p = 0.5$)
Binomial	$\binom{n}{x} p^x(1-p)^{n-x}$, $\binom{n}{x} = \frac{n!}{x!(n-x)!}$	np	$np(1-p)$
Poisson	$P(X=x) = \frac{\lambda^x e^{-\lambda}}{x!}$	λ	λ
Exponential	$f(x) = \frac{1}{\beta} e^{-x/\beta}$, $\lambda = 1/\beta$ (hazard rate)	$\beta = 1/\lambda$	$\beta^2 = 1/\lambda^2$
Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
Lognormal	$\ln X \sim N(\mu, \sigma^2)$	$e^{\mu+\sigma^2/2}$	right-skewed, $X > 0$

Poisson rate scales with time (e.g. 0.1 calls/sec $\Rightarrow \lambda = 6/\text{min}$). Number of defaults by time t with exponential waiting times is Poisson with rate t/β . Beta distribution lives on $[0, 1]$ — used for default probabilities & recovery rates.

3.2 Normal distribution & standardisation

$$z = \frac{x - \mu}{\sigma} \quad Z \sim N(0, 1), \quad N(-z) = 1 - N(z)$$

Confidence interval: $\mu \pm z\sigma$: 90% ($z = 1.65$), 95% ($z = 1.96$), 99% ($z = 2.58$)

$\approx 68\%$ within $\pm 1\sigma$, $\approx 95\%$ within $\pm 2\sigma$, $\approx 99.7\%$ within $\pm 3\sigma$. Sums of independent normals are normal. Tails never reach zero.

3.3 Sampling distributions: χ^2 , t , F

Student's t : symmetric, fatter tails than normal; one parameter $df = n - 1$; $\rightarrow N(0, 1)$ as $df \rightarrow \infty$. Use for small samples with unknown variance.

Chi-squared: $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$ ($df = n - 1$) — tests a single variance; right-skewed, ≥ 0 .

F: $F = \frac{s_1^2}{s_2^2}$ ($df_1 = n_1 - 1$, $df_2 = n_2 - 1$) — tests equality of two variances; right-skewed, ≥ 0 .

Links: $t_{n-1}^2 = F_{1, n-1}$; $k \times F_{k, m} \rightarrow \chi_k^2$ as $m \rightarrow \infty$.

3.4 Mixture distributions

Draw from distribution i with probability (weight) w_i , $\sum w_i = 1$:

$$E(X) = \sum_i w_i \mu_i, \quad \sigma^2 = \sum_i w_i (\sigma_i^2 + \mu_i^2) - \left(\sum_i w_i \mu_i \right)^2$$

Components are parametric, weights non-parametric. Mixing different **variances** raises kurtosis (fat tails); mixing different **means** creates skew or bimodality.

Worked example — Uniform, binomial, Poisson

Uniform on $[2, 12]$: $P(4 \leq X \leq 8) = (8-4)/(12-2) = 40\%$; mean = 7; variance = $100/12 = 8.33$.

Binomial: $n = 5$, $p = 0.6$, $P(X = 3) = \binom{5}{3} (0.6)^3 (0.4)^2 = 10 \times 0.216 \times 0.16 = 0.3456$; $E(X) = 3$, $\text{Var} = 5(0.6)(0.4) = 1.2$.

Poisson: 0.1 calls/second $\Rightarrow \lambda = 6$ per minute. $P(X = 5) = \frac{6^5 e^{-6}}{5!} = \frac{7776 \times 0.0024788}{120} = 0.1606$.

Worked example — Normal probabilities & confidence interval

$\text{EPS} \sim N(6, 2^2)$. $P(\text{EPS} > 9.70)$: $z = (9.70 - 6)/2 = 1.85$, $N(1.85) = 0.9678 \Rightarrow 1 - 0.9678 = \mathbf{3.22\%}$.
 Fund return mean 10.5%, $\sigma = 18\%$: 95% interval = $10.5 \pm 1.96(18) = -24.78\%$ to 45.78% .

Worked example — Mixture (regime) distribution

75% of the time returns are $N(0, 1\%^2)$, 25% of the time $N(0, 3\%^2)$. Mean = 0; $\sigma^2 = 0.75(1) + 0.25(9) = 3 \Rightarrow \sigma = \mathbf{1.73\%}$.
 Same mean & a single σ , but more probability near 0 and in the tails $\Rightarrow$ **excess kurtosis** (fat tails).

EXAM TRAP: For the binomial, define “success” carefully. For Poisson, rescale λ to the period asked. A lognormal variable cannot be negative — it suits prices (price relatives), not returns.

4. Multivariate Random Variables

Joint PMF, marginal & conditional distributions, covariance, correlation, portfolio variance, i.i.d.

4.1 Joint, marginal & conditional distributions

$$f_{X_1, X_2}(x_1, x_2) = P(X_1 = x_1, X_2 = x_2), \quad \sum \sum f = 1$$

$$\text{Marginal: } f_{X_1}(x_1) = \sum_{x_2} f_{X_1, X_2}(x_1, x_2) \quad \text{Conditional: } f_{X_1|X_2}(x_1 | x_2) = \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_2}(x_2)}$$

$$E[g(X_1, X_2)] = \sum_{x_1} \sum_{x_2} g(x_1, x_2) f_{X_1, X_2}(x_1, x_2), \quad E(X_1 | X_2 = x_2) = \sum_{x_1} x_1 f_{X_1|X_2}(x_1 | x_2)$$

In a probability matrix, the marginal of the row variable = sum across a row; of the column variable = sum down a column.

4.2 Covariance & correlation

$$\text{Cov}(X, Y) = E\{[X - E(X)][Y - E(Y)]\} = E(XY) - E(X)E(Y)$$

$$\rho_{XY} = \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}, \quad -1 \leq \rho \leq +1, \quad \text{Cov}(X, X) = \sigma_X^2$$

Linear transformations: $\text{Cov}(a + bX, c + dY) = bd \text{Cov}(X, Y)$; Corr unchanged in size (sign flips if $bd < 0$).

If $Y = a + bX$ exactly: $\rho = +1$ ($b > 0$), -1 ($b < 0$). Independence $\Rightarrow \rho = 0$, but $\rho = 0$ does **not** imply independence (e.g. $Y = X^2$).

Covariance is in squared units and unbounded; correlation is unit-free.

4.3 Variance of sums & weighted sums

$$\text{Var}(X + Y) = \sigma_X^2 + \sigma_Y^2 + 2\text{Cov}(X, Y), \quad \text{Var}(X - Y) = \sigma_X^2 + \sigma_Y^2 - 2\text{Cov}(X, Y)$$

$$\text{Var}(aX + bY) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\text{Cov}(X, Y)$$

$$\text{Minimum-variance weight in } X: \quad w^* = \frac{\sigma_Y^2 - \sigma_{XY}}{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}}$$

4.4 i.i.d. random variables

If $X_1, \dots, X_n$ are independent and identically distributed with mean μ and variance σ^2 :

$$E\left(\sum X_i\right) = n\mu, \quad \text{Var}\left(\sum X_i\right) = n\sigma^2 \text{ (grows linearly)}, \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n} \text{ (falls with } n)$$

All covariances are zero, so the cross terms drop out. Sums of i.i.d. normals are normal.

Worked example — Covariance & correlation from states

States (prob.): boom 0.3, normal 0.5, slow 0.2. Returns A: 20%, 12%, 5%; B: 30%, 10%, 0%.

$E(R_A) = 0.3(0.20) + 0.5(0.12) + 0.2(0.05) = 0.13$; $E(R_B) = 0.3(0.30) + 0.5(0.10) + 0 = 0.14$.

$\text{Cov} = 0.3(0.07)(0.16) + 0.5(-0.01)(-0.04) + 0.2(-0.08)(-0.14) = 0.00336 + 0.00020 + 0.00224 = \mathbf{0.0058}$.

$\sigma_A^2 = 0.3(0.0049) + 0.5(0.0001) + 0.2(0.0064) = 0.0028$; $\sigma_B^2 = 0.3(0.0256) + 0.5(0.0016) + 0.2(0.0196) = 0.0124$.

$$\rho = \frac{0.0058}{\sqrt{0.0028}\sqrt{0.0124}} = \frac{0.0058}{0.05292 \times 0.11136} = \mathbf{0.984}$$

Worked example — Conditional expectation

Given a negative earnings announcement (marginal probability 40%), joint probabilities for returns -3% , 0% , $+3\%$ are 25% , 15% , 0% . Conditional probabilities: $25/40 = 0.625$, $15/40 = 0.375$, 0 . $E(R | \text{negative}) = -3(0.625) + 0 + 3(0) = -1.875\%$.

5. Sample Moments

Estimators, bias, BLUE, consistency, LLN & CLT, sample skewness/kurtosis, quantiles, co-moments

5.1 Sample mean & variance

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \sum (X_i - \bar{X}) = 0$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (\text{unbiased}) \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (\text{biased})$$

$$E(\hat{\sigma}^2) = \frac{n-1}{n} \sigma^2 \Rightarrow \text{Bias} = -\frac{\sigma^2}{n} \quad \text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$$

$$\text{Var}(\hat{\mu}) = \frac{\sigma^2}{n}, \quad \text{Standard error} = \frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \frac{s}{\sqrt{n}}$$

Estimator = the formula; estimate = the number it produces. The sample mean is **BLUE** (best linear unbiased estimator: minimum variance among linear unbiased estimators) when data are i.i.d. **Consistent** estimator: bias and variance $\rightarrow 0$ as $n \rightarrow \infty$.

5.2 Law of Large Numbers & Central Limit Theorem

LLN: $\bar{X} \rightarrow \mu$ as $n \rightarrow \infty$ (needs finite mean) — makes estimators consistent.

CLT: for i.i.d. data with finite variance, $\bar{X} \overset{a}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$ regardless of the population's shape (rule of thumb $n \geq 30$).

5.3 Sample skewness, kurtosis, quantiles

$$\hat{S} = \frac{\hat{\mu}_3}{\hat{\sigma}^3}, \quad \hat{\mu}_3 = \frac{1}{n} \sum (X_i - \bar{X})^3 \quad \hat{K} = \frac{\hat{\mu}_4}{\hat{\sigma}^4}, \quad \hat{\mu}_4 = \frac{1}{n} \sum (X_i - \bar{X})^4$$

Median (sorted data): n odd $\Rightarrow x_{(n+1)/2}$; n even $\Rightarrow \frac{1}{2}(x_{(n/2)} + x_{(n/2+1)})$.

α -quantile: sort, take observation at position $\alpha \times n$ (average neighbours if not an integer). Median and IQR are **robust** to outliers; the mean is not.

5.4 Sample covariance, correlation & co-moments

$$\hat{\sigma}_{XY} = \frac{1}{n-1} \sum (X_i - \bar{X})(Y_i - \bar{Y}), \quad \hat{\rho}_{XY} = \frac{\hat{\sigma}_{XY}}{\hat{\sigma}_X \hat{\sigma}_Y}$$

$$\text{Coskewness: } s(X, X, Y) = \frac{E[(X - \mu_X)^2(Y - \mu_Y)]}{\sigma_X^2 \sigma_Y}, \quad s(X, Y, Y) = \frac{E[(X - \mu_X)(Y - \mu_Y)^2]}{\sigma_X \sigma_Y^2}$$

$$\text{Cokurtosis: } k(X, X, Y, Y) = \frac{E[(X - \mu_X)^2(Y - \mu_Y)^2]}{\sigma_X^2 \sigma_Y^2} \quad (\text{also } k(X, X, X, Y), k(X, Y, Y, Y))$$

Bivariate normal: coskewness = 0; symmetric cokurtosis = $1 + 2\rho^2$ (min 1 at $\rho = 0$, max 3 at $|\rho| = 1$); asymmetric cokurtosis = 3ρ (linear from -3 to $+3$).

Worked example — Sample statistics

Returns: 24% , 34% , 18% , 54% , 10% ($n = 5$). $\bar{X} = 140/5 = 28\%$.

Deviations: -4 , 6 , -10 , 26 , -18 ; squared: 16 , 36 , 100 , 676 , 324 ; sum = $1,152$ ($\%^2$) = 0.1152 .

Biased $\hat{\sigma}^2 = 0.1152/5 = 0.02304$; unbiased $s^2 = 0.1152/4 = 0.0288 \Rightarrow s = 16.97\%$.

Median: sorted $10, 18, 24, 34, 54 \Rightarrow 24\%$. Standard error of mean = $16.97/\sqrt{5} = 7.59\%$.

EXAM TRAP: Always divide by $n - 1$ for the sample variance unless the question explicitly asks for the biased (MLE) version. LLN needs a finite mean; CLT additionally needs a finite variance — neither needs normality.

6. Hypothesis Testing

Null & alternative, one/two-tailed tests, errors, power, p -values, confidence intervals, difference in means, multiple testing

6.1 General test statistic & decision rule

$$\text{Test statistic} = \frac{\text{Sample statistic} - \text{Hypothesised value}}{\text{Standard error of the statistic}}$$

$$z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \quad (\sigma \text{ known, normal}) \quad t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}, \quad df = n - 1 \quad (\sigma \text{ unknown})$$

Two-tailed: $H_0: \mu = \mu_0$ vs. $H_A: \mu \neq \mu_0$ — reject if $|\text{stat}| > \text{critical value}$ ($\alpha/2$ in each tail).

One-tailed: $H_0: \mu \leq \mu_0$ vs. $H_A: \mu > \mu_0$ (reject if $\text{stat} > +\text{c.v.}$) or $H_0: \mu \geq \mu_0$ vs. $H_A: \mu < \mu_0$ (reject if $\text{stat} < -\text{c.v.}$). The null always contains the “=” sign; the alternative is what you hope to prove. Never “accept” H_0 — only fail to reject.

6.2 Confidence intervals, p -values, errors & power

$$\text{Two-sided CI: } \bar{X} \pm c_{\alpha/2} \frac{s}{\sqrt{n}} \quad \text{One-sided: } (\bar{X} - c_{\alpha} \frac{s}{\sqrt{n}}, \infty) \text{ or } (-\infty, \bar{X} + c_{\alpha} \frac{s}{\sqrt{n}})$$

Reject H_0 iff the hypothesised value lies *outside* the $(1 - \alpha)$ CI.

p -value: smallest significance level at which H_0 can be rejected. Reject if $p < \alpha$. Two-tailed $p = 2 \times [1 - N(|z|)]$.

	H_0 true	H_0 false
Reject H_0	Type I error (prob. = α , size)	Correct (prob. = power = $1 - \beta$)
Fail to reject	Correct ($1 - \alpha$)	Type II error (prob. = β)

Lower $\alpha \Rightarrow$ more Type II errors / less power (for fixed n). Only a larger sample raises power without raising α .

6.3 Which test? & difference in means

Population	Small sample ($n < 30$)	Large sample ($n \geq 30$)
Normal, σ known	z	z
Normal, σ unknown	t	t (or z)
Non-normal, σ unknown	no valid test	t (or z)

Equality of two means ($H_0: \mu_X = \mu_Y$), paired i.i.d. series of length n :

$$T = \frac{\hat{\mu}_X - \hat{\mu}_Y}{\sqrt{\frac{\hat{\sigma}_X^2 + \hat{\sigma}_Y^2 - 2\hat{\sigma}_{XY}}{n}}} \sim N(0, 1)$$

Multiple testing: running many tests on one data set inflates the true Type I error rate (data snooping) — the nominal α is valid only for a single test.

Worked example — z/t -test and confidence interval

250 daily option-portfolio returns: $\bar{X} = 0.1\%$, $s = 0.25\%$. Test $H_0: \mu = 0$ vs. $H_A: \mu \neq 0$ at 5%.

$SE = 0.25/\sqrt{250} = 0.25/15.811 = 0.01581\%$; test stat = $0.1/0.01581 = 6.32 > 1.96 \Rightarrow$ **reject** H_0 .

95% CI: $0.1 \pm 1.96(0.01581) = 0.069\% \text{ to } 0.131\%$ — excludes 0, consistent with the test $\checkmark$. One-tailed ($H_A: \mu > 0$): $6.32 > 1.645 \Rightarrow$ reject.

Worked example — Two-tailed z -test (quality control)

Gizmo length: $\mu_0 = 2.5$, $\sigma = 0.021$ (known), $n = 49$, $\bar{X} = 2.49$. $z = \frac{2.49 - 2.50}{0.021/7} = \frac{-0.01}{0.003} = -3.33 < -1.96 \Rightarrow$ reject; machine needs adjustment.

Worked example — p -value & one-tailed test

(a) Two-tailed test, $z = 2.3$: $N(2.3) = 0.9893 \Rightarrow p = 2(1 - 0.9893) = 2.14\%$. Reject at 5%, 3%; not at 2% or 1%.

(b) House prices: $H_A: \mu > 145,000$; $n = 36$, $\bar{X} = 149,750$, $\sigma = 24,000$. $z = \frac{4,750}{24,000/6} = 1.19 < 2.33$ (1%) $\Rightarrow$ fail to reject.

EXAM TRAP: Type I error probability = significance level α . Power = $1 - P(\text{Type II})$, not $1 - \alpha$. The test statistic is the same for one- and two-tailed tests; only the critical value changes.

7. Linear Regression (One Explanatory Variable)

OLS estimates, interpretation, assumptions, hypothesis tests & confidence intervals, R^2

7.1 The model & OLS estimators

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad \text{OLS chooses } \hat{\alpha}, \hat{\beta} \text{ to minimise } \sum \hat{\varepsilon}_i^2 = \sum (Y_i - \hat{\alpha} - \hat{\beta} X_i)^2$$

$$\hat{\beta} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\hat{\sigma}_{XY}}{\hat{\sigma}_X^2} = \hat{\rho}_{XY} \frac{\hat{\sigma}_Y}{\hat{\sigma}_X} \quad \hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}$$

$\hat{\beta}$ = change in Y for a one-unit change in X ; $\hat{\alpha}$ = value of Y when $X = 0$. The fitted line always passes through $(\bar{X}, \bar{Y})$. “Linear” means linear in the *parameters* — X may be transformed (e.g. $\ln X$, X^2); dummy variables take values 0/1.

7.2 Inference on a coefficient

$$\text{Var}(\hat{\beta}) \approx \frac{\sigma_\varepsilon^2}{n \hat{\sigma}_X^2} \quad (\text{larger error variance } \uparrow, \text{ more spread in } X \downarrow)$$

$$t = \frac{\hat{\beta} - \beta_0}{SE(\hat{\beta})}, \quad df = n - 2 \quad \text{CI} = \hat{\beta} \pm t_c \times SE(\hat{\beta})$$

Steps: (1) state $H_0 : \beta = \beta_0$ vs. $H_A : \beta \neq \beta_0$; (2) compute t ; (3) compare with critical t (or p -value with α , or check whether β_0 lies in the CI).

$$R^2 = \rho_{XY}^2 \quad \Rightarrow \quad |\rho_{XY}| = \sqrt{R^2} \quad (\text{sign of } \rho = \text{sign of } \hat{\beta})$$

OLS assumptions (single regression)

(1) $E(\varepsilon_i | X_i) = 0$ — violated by survivorship/sample-selection bias, simultaneity bias, omitted variables, attenuation (measurement-error) bias; (2) (X_i, Y_i) i.i.d.; (3) $\text{Var}(X) > 0$; (4) constant error variance (homoskedasticity); (5) no large outliers. Under these, $\hat{\alpha}, \hat{\beta}$ are unbiased, consistent and asymptotically normal.

Worked example — Estimating a regression

Stock vs. market returns: $\bar{Y} = 11\%$, $\bar{X} = 8.4\%$, $\text{Cov}(X, Y) = 0.0073$, $\text{Var}(X) = 0.0100$.

$\hat{\beta} = 0.0073/0.0100 = \mathbf{0.73}$; $\hat{\alpha} = 0.11 - 0.73(0.084) = 0.11 - 0.0613 = \mathbf{0.0487 (4.87\%)}$.

Interpretation: a 1% rise in the market lifts the stock by 0.73%; with a 0% market return the stock earns $\approx 4.9\%$.

Worked example — Testing the slope

$n = 46$, $\hat{\beta} = 0.76$, $SE = 0.33$, $t_c(44 df, 5\%) = 2.02$. $t = 0.76/0.33 = \mathbf{2.30} > 2.02 \Rightarrow \text{reject } H_0 : \beta = 0$.

95% CI = $0.76 \pm 2.02(0.33) = \mathbf{0.093 \text{ to } 1.427}$; 0 is outside $\checkmark$; 0.20 is inside $\Rightarrow \text{cannot reject } H_0 : \beta = 0.20$.

8. Regression with Multiple Explanatory Variables

Partial slopes, TSS = ESS + RSS, R^2 , adjusted R^2 , SER, t - and F -tests

8.1 Multiple regression & interpretation

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i$$

β_j = **partial slope**: change in Y for a one-unit change in X_j , *holding all other X s constant*. Extra (6th) assumption: no **perfect collinearity** among the X s. OLS can be seen as regressing Y on the part of X_1 not explained by X_2 (Frisch–Waugh).

8.2 Decomposition of variation & goodness of fit

$$\underbrace{\sum (Y_i - \bar{Y})^2}_{TSS} = \underbrace{\sum (\hat{Y}_i - \bar{Y})^2}_{ESS \text{ (explained)}} + \underbrace{\sum (Y_i - \hat{Y}_i)^2}_{RSS \text{ (residual)}}$$

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n - 1}{n - k - 1}$$

$$SER = \sqrt{\frac{RSS}{n - k - 1}}$$

R^2 never falls when a variable is added; $\bar{R}^2$ penalises extra regressors, can fall, and can be negative; $\bar{R}^2 \leq R^2$. R^2 is not comparable across models with different Y s and a high R^2 does not prove causation or a complete model. In multiple regression $R^2 = [\text{Corr}(Y, \hat{Y})]^2$.

8.3 Hypothesis tests in multiple regression

Single coefficient: $t = \frac{\hat{\beta}_j - \beta_{j0}}{SE(\hat{\beta}_j)}$, $df = n - k - 1$; $CI = \hat{\beta}_j \pm t_c SE(\hat{\beta}_j)$.

All slopes jointly zero ($H_0: \beta_1 = \dots = \beta_k = 0$; H_A : at least one $\neq 0$):

$$F = \frac{ESS/k}{RSS/(n-k-1)} \quad df = (k, n-k-1)$$

Subset of q coefficients — unrestricted/full (U) vs. restricted/partial (R) model:

$$F = \frac{(RSS_R - RSS_U)/q}{RSS_U/(n-k_U-1)} \quad df = (q, n-k_U-1)$$

The F -test is always one-tailed (reject if $F > F_c$). Note $F = \frac{R^2/k}{(1-R^2)/(n-k-1)}$ for the all-slopes test.

Worked example — Prediction & partial effect

$\hat{R}_P = 1.70 + 1.03R_m - 0.23R_z + 0.32R_v$. With $R_m = 8$, $R_z = 2$, $R_v = 3$: $1.70 + 8.24 - 0.46 + 0.96 = \mathbf{10.44\%}$. If R_z falls by 1: $\Delta R_P = -0.23(-1) = +0.23\%$. All factors zero: $\mathbf{1.70\%}$.

Worked example — R^2 and adjusted R^2

$n = 60$, $k = 5$, $TSS = 460$, $RSS = 170 \Rightarrow ESS = 290$.

$R^2 = 290/460 = \mathbf{0.630}$; $\bar{R}^2 = 1 - (0.370)\frac{59}{54} = 1 - 0.4043 = \mathbf{0.596}$.

Add 4 variables ($k = 9$) and R^2 rises to 0.65: $\bar{R}^2 = 1 - 0.35 \times \frac{59}{50} = 1 - 0.413 = \mathbf{0.587} < 0.596 \Rightarrow$ prefer the 5-variable model.

Worked example — F -tests

(a) All slopes: $n = 46$, $k = 5$, $ESS = 290$, $RSS = 170$. $F = \frac{290/5}{170/40} = \frac{58}{4.25} = \mathbf{13.65} > F_c(5, 40) = 2.45 \Rightarrow$ at least one slope $\neq 0$.

(b) Full (3 vars) vs. CAPM (1 var): $n = 64$, $RSS_U = 6,650$, $RSS_R = 7,140$, $q = 2$.

$F = \frac{(7,140 - 6,650)/2}{6,650/(64 - 3 - 1)} = \frac{245}{110.83} = \mathbf{2.21} < F_c(2, 60) = 3.15 \Rightarrow$ the two extra variables do not add jointly.

EXAM TRAP: t -tests in multiple regression use $df = n - k - 1$ (not $n - 2$). A dummy's coefficient is the difference in intercept versus the omitted category, holding other variables constant.

9. Regression Diagnostics

Heteroskedasticity, multicollinearity, omitted/irrelevant variables, bias–variance, residual plots, outliers, BLUE

9.1 Heteroskedasticity

Unconditional (not related to X): usually harmless. **Conditional** (variance related to X): coefficients still **unbiased & consistent**, but standard errors unreliable $\Rightarrow t/F$ tests invalid.

White test: regress $\hat{\epsilon}_i^2$ on the X s, their squares and cross-products; then

$$\chi^2 = nR_{\text{aux}}^2, \quad df = \frac{k(k+3)}{2} \quad (\text{reject } H_0 \text{ of homoskedasticity if } \chi^2 > \chi_c^2)$$

Fix: use White (heteroskedasticity-consistent, “robust”) standard errors; alternatively WLS or transform the data.

9.2 Multicollinearity

$$VIF_j = \frac{1}{1 - R_j^2} \quad R_j^2 = R^2 \text{ from regressing } X_j \text{ on the other } X\text{s}; \quad VIF_j > 10 \text{ (} R_j^2 > 0.90 \text{)} \Rightarrow \text{problem}$$

Perfect collinearity (exact linear combination) violates an assumption — OLS cannot be computed. **Multicollinearity** (high but imperfect correlation) is not a violation, but inflates SEs $\Rightarrow$ Type II errors. Classic symptom: high R^2 & significant F , yet insignificant individual t s.

Fix: drop a variable / stepwise selection.

9.3 Model specification, bias–variance, outliers

Omitted variable bias requires both: omitted Z is correlated with an included X and Z determines $Y \Rightarrow$ biased, inconsistent coefficients (does not vanish with more data).

Irrelevant regressor: coefficients unbiased, but less precise; $\bar{R}^2$ falls.

Bias–variance trade-off: large models = low bias, high variance (overfit); small models = high bias, low variance. Select via **general-to-specific** (drop smallest $|t|$ repeatedly) or **m -fold cross-validation** (fit on $m - 1$ folds, validate on the held-out fold; pick lowest out-of-sample error).

Cook's distance (influence of observation j):

$$D_j = \frac{\sum_{i=1}^n (\hat{Y}_i^{(-j)} - \hat{Y}_i)^2}{k s^2} \quad D_j > 1 \Rightarrow \text{outlier}$$

$\hat{Y}_i^{(-j)}$ = fitted value with observation j dropped; s^2 = residual variance of the full model. Standardised residuals beyond ± 4 flag outliers in residual plots.

OLS is BLUE (Gauss–Markov) when the model is linear, errors have mean zero, are homoskedastic and uncorrelated (normality *not* required for BLUE).

Worked example — White test & VIF

Two regressors ($k = 2$), $n = 200$, auxiliary $R^2 = 0.06$: $\chi^2 = 200 \times 0.06 = 12$; $df = 2(5)/2 = 5$; $\chi_c^2(5, 5\%) = 11.07 \Rightarrow$ **reject** homoskedasticity $\Rightarrow$ use White SEs.

Regressing X_2 on the other regressors gives $R^2 = 0.92$: $VIF = 1/0.08 = 12.5 > 10 \Rightarrow$ multicollinearity in X_2 .

10. Stationary Time Series

Covariance stationarity, ACF/PACF, white noise, AR, MA, ARMA, lag operator, Q -statistics, forecasting

10.1 Covariance stationarity & autocorrelation

Stationary if: (1) constant mean, (2) constant finite variance, (3) autocovariance depends only on the lag τ , not on t .

$$\gamma(\tau) = \text{Cov}(Y_t, Y_{t-\tau}), \quad \gamma(0) = \text{Var}(Y_t), \quad \rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}, \quad \rho(0) = 1, \quad \rho(\tau) = \rho(-\tau)$$

White noise $\varepsilon_t \sim WN(0, \sigma^2)$: mean 0, constant variance, zero autocorrelation. Independent WN; Gaussian (normal) WN.

Lag operator: $LY_t = Y_{t-1}$, $L^k Y_t = Y_{t-k}$.

10.2 AR(p), MA(q), ARMA(p, q)

Model	Equation	Key properties
AR(1)	$Y_t = \delta + \phi Y_{t-1} + \varepsilon_t$	Stationary if $ \phi < 1$. $\mu = \frac{\delta}{1-\phi}$, $\gamma(0) = \frac{\sigma^2}{1-\phi^2}$, $\rho(\tau) = \phi^{ \tau }$ (geometric decay). PACF cuts off after lag 1.
AR(p)	$Y_t = \delta + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t$	Mean-reverting level $\mu = \frac{\delta}{1-\phi_1-\dots-\phi_p}$. ACF decays; PACF cuts off after lag p .
MA(1)	$Y_t = \mu + \theta \varepsilon_{t-1} + \varepsilon_t$	Always stationary. $E(Y_t) = \mu$, $\gamma(0) = (1+\theta^2)\sigma^2$, $\rho(1) = \frac{\theta}{1+\theta^2}$, $\rho(\tau) = 0$ for $\tau \geq 2$.
MA(q)	$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$	ACF cuts off after lag q ; PACF decays.
ARMA	$Y_t = \delta + \sum \phi_i Y_{t-i} + \varepsilon_t + \sum \theta_j \varepsilon_{t-j}$	Both ACF & PACF decay gradually. Seasonal: $\text{ARMA}(p, q) \times (p_s, q_s)$, e.g. lag 4 (quarterly) or 12 (monthly).

10.3 Testing residual autocorrelation & forecasting

$$Q_{BP} = T \sum_{\tau=1}^h \hat{\rho}^2(\tau)$$

$$Q_{LB} = T(T+2) \sum_{\tau=1}^h \frac{\hat{\rho}^2(\tau)}{T-\tau} \sim \chi_h^2$$

H_0 : all h autocorrelations are zero. Ljung–Box is better in small samples ($T \leq 100$). 95% band for a single sample autocorrelation $\approx \pm 1.96/\sqrt{T}$.

AR(1) forecasts: $E_T(Y_{T+1}) = \delta + \phi Y_T$; $E_T(Y_{T+h}) = \mu + \phi^h(Y_T - \mu) \rightarrow \mu$ as $h \rightarrow \infty$ (mean reversion). MA(q) forecasts equal μ beyond q steps.

Worked example — AR(1) & MA(1)

AR(1): $\delta = 0.5$, $\phi = 0.8$, $\sigma^2 = 1$. $\mu = 0.5/0.2 = \mathbf{2.5}$; $\gamma(0) = 1/(1 - 0.64) = \mathbf{2.78}$; $\rho(2) = 0.8^2 = \mathbf{0.64}$.

If $Y_T = 4$: $E(Y_{T+1}) = 0.5 + 0.8(4) = \mathbf{3.70}$; $E(Y_{T+2}) = 2.5 + 0.8^2(4 - 2.5) = \mathbf{3.46}$ (check: $0.5 + 0.8 \times 3.70 = 3.46 \checkmark$).

MA(1) with $\theta = 0.5$: $\rho(1) = 0.5/1.25 = \mathbf{0.40}$; $\rho(2) = 0$.

11. Non-Stationary Time Series

Trends, seasonality, random walks & unit roots, forecasts

11.1 Trends, seasonality & unit roots

Linear trend: $Y_t = \delta_0 + \delta_1 t + \varepsilon_t$ Log-linear (constant growth): $\ln Y_t = \delta_0 + \delta_1 t + \varepsilon_t$

Quadratic: $Y_t = \delta_0 + \delta_1 t + \delta_2 t^2 + \varepsilon_t$ Seasonal dummies: $Y_t = \delta_0 + \sum_{j=1}^{s-1} \gamma_j D_{jt} + \varepsilon_t$

Random walk: $Y_t = Y_{t-1} + \varepsilon_t$ (with drift: $Y_t = \delta + Y_{t-1} + \varepsilon_t$) $\text{Var}(Y_t) = t\sigma^2 \rightarrow \infty$

Unit-root (Augmented Dickey–Fuller) test: $\Delta Y_t = \gamma Y_{t-1} + \dots + \varepsilon_t$; $H_0: \gamma = 0$ (unit root) vs. $H_A: \gamma < 0$ (stationary). Fix a unit root by **differencing**. Use $s-1$ dummies with an intercept (avoid the dummy-variable trap). Unit roots cause spurious regressions and non-standard (t -stat) distributions.

Forecasts: $\hat{Y}_{T+h} = \hat{\delta}_0 + \hat{\delta}_1(T+h)$ (+ the seasonal dummy for period $T+h$); 95% interval = $\hat{Y}_{T+h} \pm 1.96 \hat{\sigma}$. Log model: $\hat{Y} = \exp(\hat{\delta}_0 + \hat{\delta}_1 t + \hat{\sigma}^2/2)$.

Worked example — Trend forecast with interval

$\hat{Y}_t = 10 + 0.5t$, $\hat{\sigma} = 2$, last observation $t = 20$. One-step forecast $t = 21$: $10 + 0.5(21) = \mathbf{20.5}$; 95% interval = $20.5 \pm 1.96(2) = \mathbf{16.58 \text{ to } 24.42}$.

12. Measuring Returns, Volatility & Correlation

Simple vs. log returns, volatility scaling, Jarque–Bera, power law, correlation measures

12.1 Returns & volatility

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

$$r_t = \ln \frac{P_t}{P_{t-1}} = \ln(1 + R_t)$$

$$R_t = e^{r_t} - 1$$

Multi-period: $1 + R_T = \prod_{t=1}^T (1 + R_t)$ $r_T = \sum_{t=1}^T r_t$ (log returns add)

$\sigma_{\text{horizon}} = \sigma_{\text{period}} \sqrt{T}$ e.g. annual = daily $\times \sqrt{252}$, monthly $\times \sqrt{12}$

Continuously compounded return $\leq$ simple return. Volatility = s.d. of returns; variance rate = volatility². **Implied volatility** is backed out of option prices (e.g. BSM, VIX) and is forward-looking.

12.2 Testing normality — Jarque–Bera

$$JB = (T-1) \left[\frac{\hat{S}^2}{6} + \frac{(\hat{K}-3)^2}{24} \right] \sim \chi_2^2 \quad \text{critical: 5.99 (5\%), 9.21 (1\%)}$$

H_0 : $S = 0$ and $K = 3$ (normal). Large JB $\Rightarrow$ reject normality. Financial returns usually show $S \neq 0$ and $K > 3$.

Power law tails: $P(X > x) = kx^{-\alpha}$ — tails decline slowly (fat tails, e.g. Student's t); smaller α = fatter tail.

12.3 Correlation & dependence

Pearson ρ captures only *linear* dependence. **Rank correlations** capture monotonic dependence:

$$\text{Spearman: } \rho_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad \text{Kendall: } \tau = \frac{n_c - n_d}{n(n-1)/2}$$

d_i = difference in ranks; n_c/n_d = number of concordant/discordant pairs.

One-factor model: $U_i = a_i F + \sqrt{1 - a_i^2} Z_i$ (F, Z_i independent $N(0, 1)$) $\Rightarrow \text{Var}(U_i) = 1$, $\text{Corr}(U_i, U_j) = a_i a_j$.

Worked example — Returns & scaling

Price 100 $\rightarrow$ 110: $R = 10\%$; $r = \ln 1.10 = 9.53\%$. A 5% log return $= e^{0.05} - 1 = 5.13\%$ simple.
Daily volatility 1% $\Rightarrow$ annual $= 1\% \times \sqrt{252} = 15.87\%$.

Worked example — Jarque–Bera & Spearman

60 monthly returns, $\hat{S} = 0.30$, $\hat{K} = 3.50$: $JB = 59 \left[\frac{0.09}{6} + \frac{0.25}{24} \right] = 59(0.015 + 0.01042) = 1.50 < 5.99 \Rightarrow$ cannot reject normality.
Spearman: $n = 5$, $\sum d_i^2 = 4$: $\rho_s = 1 - \frac{6(4)}{5(24)} = 1 - 0.2 = 0.80$.

13. Simulation & Bootstrapping

Monte Carlo steps, sampling error & its reduction, bootstrapping, pseudo-random numbers

13.1 Monte Carlo & sampling error

Steps: (1) specify the data-generating process & parameters; (2) generate random draws; (3) compute the quantity of interest for each replication; (4) repeat b times; (5) average (or take quantiles) across replications.

$$SE(\hat{g}) = \frac{\hat{\sigma}_g}{\sqrt{b}} \quad \text{halving the error needs } 4 \times \text{ the replications}$$

Variance reduction: antithetic variates — pair each draw u with $-u$ (works because the pair is negatively correlated); **control variates** — use a correlated variable with a known mean to correct the estimate.

Bootstrapping & random numbers

Bootstrap: resample the *actual* data *with replacement* (i.i.d. or block bootstrap for dependent data). Advantage: no distributional assumption; uses real data. Fails when: the sample is small, data have **structural breaks** / regime changes, or the past is unrepresentative.

Pseudo-random numbers: deterministic algorithms that look random; the *seed* makes results reproducible.

Disadvantages of simulation: results depend on the assumed model (model risk), can be computationally costly, and give only approximate answers.

Worked example — Monte Carlo standard error

Simulated payoff s.d. = 20, $b = 10,000$ replications: $SE = 20/\sqrt{10,000} = 0.20$. To reach 0.10 you need **40,000** replications.

14. Machine Learning Methods

ML vs. econometrics, rescaling, data splits, under/overfitting, PCA, K-means, NLP, learning types, Q-values

14.1 Data preparation & sample splits

$$\text{Standardisation: } \tilde{x} = \frac{x - \mu}{\sigma} \quad (\text{mean } 0, \text{ s.d. } 1) \quad \text{Normalisation: } \tilde{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (\text{range } [0, 1])$$

Training set — estimate parameters; **validation** set — tune hyperparameters / choose among models; **test** set — final, untouched out-of-sample evaluation.

Underfitting = high bias (too simple); **overfitting** = high variance (fits noise; great in-sample, poor out-of-sample). Remedies: more/less complexity, regularisation, more data, cross-validation.

14.2 Unsupervised, supervised & reinforcement learning

PCA — reduces dimensionality: a few uncorrelated components (linear combinations of the features) explaining most variance.
K-means — choose K , assign each point to the nearest centroid, recompute centroids, repeat; minimise **inertia** (within-cluster sum of squared distances). Choose K at the “elbow”.

NLP — tokenise (lowercase, strip punctuation) → remove stop-words → stemming/lemmatisation → N -grams → “bag of words” (e.g. sentiment scoring).

Reinforcement learning — states S , actions A , rewards R ; $Q(S, A)$ = expected value of action A in state S . Choose the best-known action (*exploitation*, prob. p) or a new one (*exploration*, $1 - p$); p rises as learning progresses.

$$\text{Monte Carlo update: } Q_{\text{new}}(S, A) = Q_{\text{old}}(S, A) + \alpha [R - Q_{\text{old}}(S, A)]$$

$$\text{Temporal-difference update: } Q_{\text{new}}(S, A) = Q_{\text{old}}(S, A) + \alpha [R + Q(S_{\text{next}}, A_{\text{next}}) - Q_{\text{old}}(S, A)]$$

α = learning rate (e.g. 0.01 or 0.05). TD looks only one decision ahead, assuming the best strategy so far is followed.

Worked example — Rescaling & Q-values

Feature value 60 with min 20, max 120, mean 50, s.d. 20: normalised = $(60 - 20)/100 = \mathbf{0.40}$; standardised = $(60 - 50)/20 = \mathbf{0.50}$.

$Q(4, 3) = 0.5$; take action 3 in state 4, reward 1.2, $\alpha = 0.01$. Monte Carlo: $0.5 + 0.01(1.2 - 0.5) = \mathbf{0.507}$.

TD: reward to next decision 0.3, value of next state–action 0.8: $0.5 + 0.01(0.3 + 0.8 - 0.5) = \mathbf{0.506}$.

15. Machine Learning & Prediction

Logistic regression, encoding, regularisation, trees, ensembles, KNN, SVM, neural networks, confusion matrix

15.1 Logistic regression & encoding

$$P(Y = 1 | X) = \frac{1}{1 + e^{-(\alpha + \beta_1 X_1 + \dots + \beta_k X_k)}} = \frac{e^z}{1 + e^z} \quad \ln \frac{P}{1 - P} = \alpha + \sum \beta_j X_j \text{ (log-odds)}$$

Estimated by maximum likelihood; output lies in $(0, 1)$ and is classified with a threshold (e.g. 0.5). Linear regression can predict probabilities outside $[0, 1]$.

Categorical variables: dummy (one-hot) encoding with $c - 1$ dummies when an intercept is present (avoid the dummy trap); label encoding for ordinal categories.

15.2 Regularisation

$$\text{Ridge (L2): } \min \sum (Y_i - \hat{Y}_i)^2 + \lambda \sum \beta_j^2 \quad \text{LASSO (L1): } \min \sum (Y_i - \hat{Y}_i)^2 + \lambda \sum |\beta_j|$$

Both shrink coefficients to cut overfitting (standardise features first). Ridge shrinks but keeps all variables; LASSO can set coefficients exactly to zero (feature selection). Elastic net combines both. λ chosen by cross-validation.

15.3 Decision trees, ensembles, KNN, SVM, neural networks

$$\text{Gini} = 1 - \sum_i p_i^2 \quad \text{Entropy} = - \sum_i p_i \log_2 p_i \quad \text{Information gain} = \text{Entropy}_{\text{parent}} - \sum \frac{n_{\text{child}}}{n} \text{Entropy}_{\text{child}}$$

Ensembles: bagging (bootstrap samples, average/vote), **random forest** (bagging + random feature subsets), **boosting** (sequential learners fixing earlier errors). Wisdom of crowds lowers variance.

KNN: classify by majority of the K nearest points (distance-based; scale features). **SVM:** find the widest-margin hyperplane separating classes (support vectors on the margin).

Neural network: input → hidden layers (weighted sum + bias → activation, e.g. sigmoid/ReLU) → output; weights learned by **gradient descent** with **back-propagation**, learning rate controls step size.

15.4 Confusion matrix

	Predicted positive	Predicted negative
Actual positive	TP	FN
Actual negative	FP	TN

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad \text{Precision} = \frac{TP}{TP + FP} \quad \text{Recall (sensitivity, TPR)} = \frac{TP}{TP + FN} \quad \text{Error rate} = 1 - \text{Accuracy}$$

$$\text{FPR} = \frac{FP}{FP + TN}, \quad F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

ROC curve plots TPR vs. FPR as the threshold moves; AUC = 0.5 is random, 1 is perfect. Use the *test* set for comparing logistic regression vs. a neural network.

Worked example — Confusion matrix & tree impurity

Default model on 200 loans: TP = 40, FN = 10, FP = 20, TN = 130.

Accuracy = $170/200 = 85\%$; precision = $40/60 = 66.7\%$; recall = $40/50 = 80\%$; error rate **15%**; $F_1 = \frac{2(0.667)(0.8)}{1.467} = 0.727$.

Node with 6 defaults & 4 non-defaults ($p = 0.6, 0.4$): Gini = $1 - (0.36 + 0.16) = 0.48$; entropy = $-(0.6 \log_2 0.6 + 0.4 \log_2 0.4) = 0.442 + 0.529 = 0.971$.

EXAM TRAP: Precision is about predicted positives (“how many flagged were right”); recall is about actual positives (“how many true cases we caught”). Accuracy can mislead with imbalanced classes (e.g. rare defaults).

16. Critical Values to Memorise

Standard normal, t and χ^2 values used most often in exam questions

Standard normal (z)

Significance / confidence	10% / 90%	5% / 95%	1% / 99%
One-tailed critical z	1.28	1.645	2.33
Two-tailed critical z	1.65	1.96	2.58

Useful CDF values: $N(1.00) = 0.8413$, $N(1.28) = 0.8997$, $N(1.645) = 0.9500$, $N(1.85) = 0.9678$, $N(1.96) = 0.9750$, $N(2.00) = 0.9772$, $N(2.33) = 0.9901$, $N(2.58) = 0.9951$.

Student's t (two-tailed 5%, i.e. 2.5% per tail) & χ^2

df	5	10	20	30	40	60	∞
$t_{0.025}$	2.571	2.228	2.086	2.042	2.021	2.000	1.960

χ^2 upper 5% critical values: $df = 1$: 3.84; $df = 2$: 5.99; $df = 3$: 7.81; $df = 4$: 9.49; $df = 5$: 11.07. ($df = 2$ at 1%: 9.21.)

17. One-Page Formula Recap

Tear-out summary for last-minute revision

Addition / joint

$$P(A \cup B) = P(A) + P(B) - P(AB); \quad P(AB) = P(A | B)P(B)$$

Bayes

$$P(B | A) = \frac{P(A | B)P(B)}{P(A)}; \quad P(A) = \sum P(A | B_i)P(B_i)$$

Variance

$$\sigma^2 = E[(X - \mu)^2] = E(X^2) - \mu^2; \quad Y = a + bX \Rightarrow \sigma_Y^2 = b^2 \sigma_X^2$$

Skew / kurtosis

$$E[(X - \mu)^3] / \sigma^3; \quad E[(X - \mu)^4] / \sigma^4 \text{ (normal} = 0, 3)$$

Binomial / Poisson

$$\binom{n}{x} p^x (1-p)^{n-x} \quad (np, np(1-p)); \quad \lambda^x e^{-\lambda} / x! \quad (\lambda, \lambda)$$

Uniform / exponential

$$\frac{x_2 - x_1}{b - a}, \text{ mean } \frac{a+b}{2}, \text{ var } \frac{(b-a)^2}{12}; \quad \frac{1}{\beta} e^{-x/\beta}, \text{ mean } \beta$$

Covariance / correlation

$$E(XY) - E(X)E(Y); \quad \rho = \frac{\text{Cov}}{\sigma_X \sigma_Y}; \quad \text{Cov}(a + bX, c + dY) = bd \text{Cov}$$

Weighted sum

$$\text{Var}(aX + bY) = a^2 \sigma_X^2 + b^2 \sigma_Y^2 + 2ab \text{Cov}$$

Sample variance / SE

$$s^2 = \frac{\sum (X - \bar{X})^2}{n-1}; \quad SE = s / \sqrt{n}; \quad E(\hat{\sigma}_{biased}^2) = \frac{n-1}{n} \sigma^2$$

Test statistic

$$t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}} \quad (n-1); \quad \text{CI } \bar{X} \pm t_c s / \sqrt{n}; \quad \chi^2 = \frac{(n-1)s^2}{\sigma_0^2}; \quad F = \frac{s_1^2}{s_2^2}$$

OLS

$$\hat{\beta} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}; \quad \hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}; \quad t = \frac{\hat{\beta} - \beta_0}{SE}$$

Fit

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}; \quad \bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k-1}; \quad SER = \sqrt{\frac{RSS}{n-k-1}}$$

F-tests

$$\frac{ESS/k}{RSS/(n-k-1)}; \quad \frac{(RSS_R - RSS_U)/q}{RSS_U/(n-k_U-1)}$$

Diagnostics

$$\text{White } nR^2 \text{ (df} = \frac{k(k+3)}{2}); \quad VIF = \frac{1}{1-R_j^2}; \quad \text{Cook's } D_j = \frac{\sum (\hat{Y}^{(-j)} - \hat{Y})^2}{ks^2}$$

AR(1) / MA(1)

$$\mu = \frac{\delta}{1-\phi}, \gamma(0) = \frac{\sigma^2}{1-\phi^2}, \rho(\tau) = \phi^\tau; \quad \rho(1) = \frac{\theta}{1+\theta^2}$$

Q-statistics

$$Q_{BP} = T \sum \hat{\rho}^2(\tau); \quad Q_{LB} = T(T+2) \sum \frac{\hat{\rho}^2(\tau)}{T-\tau}$$

Returns

$$r = \ln(1 + R); \quad \sigma_T = \sigma \sqrt{T}; \quad JB = (T-1) \left[\frac{s^2}{6} + \frac{(K-3)^2}{24} \right]$$

Rank correlation

$$\rho_s = 1 - \frac{6 \sum d^2}{n(n^2-1)}; \quad \tau = \frac{n_c - n_d}{n(n-1)/2}$$

Simulation

$$SE = \hat{\sigma} / \sqrt{b}$$

ML

$$\frac{x - \mu}{\sigma}; \quad \frac{x - x_{min}}{x_{max} - x_{min}}; \quad Q_{new} = Q_{old} + \alpha(R - Q_{old}); \quad \text{Gini } 1 - \sum p^2$$

Confusion matrix

$$\text{Accuracy } \frac{TP+TN}{N}; \quad \text{Precision } \frac{TP}{TP+FP}; \quad \text{Recall } \frac{TP}{TP+FN}$$

Disclaimer. Compiled by RBei Classes from the GARP FRM Part I *Quantitative Analysis* curriculum for revision purposes. Some worked examples use illustrative numbers. GARP® and FRM® are trademarks of the Global Association of Risk Professionals, which does not endorse this material. Always confirm notation against the current official GARP readings; exam dates, fees and topic weights should be checked on garp.org.