Manuel des Mnémotechniques pour Examens Data Analyst

Guide de Révision Ultra-Rapide avec Techniques de Mémorisation

---

1. TYPES DE VARIABLES - “NOIR”

Mnémotechnique: N-O-I-R

Lettre	Type	Mémoire	Exemple
N	**N**ominale	**N**oms sans ordre	Type de compte
O	**O**rdinale	**O**rdre naturel	Rating AAA>AA>A
I	**I**ntervalle	**I**ntervalles égaux	Température
R	**R**atio	**R**atios possibles	Montant en HTG

Sous-catégories:

“Binaire = Bi = 2 choix” → Oui/Non, Défaut/Non-défaut

“Polytomique = Poly = Plusieurs” → Type compte: Épargne/Courant/DAT

Règle des statistiques:

“Le Mode pour les Mots, la Médiane pour les rangs Ordonnés” - Nominale → Mode - Ordinale → Médiane - Quantitative → Moyenne (si symétrique)

---

2. STATISTIQUES DESCRIPTIVES - “MMV-VEC-QP”

Tendance Centrale: M-M-M

- **M**oyenne (sensible aux outliers)
- **M**édiane (robuste)
- **M**ode (pour catégories)

Dispersion: V-E-C-I

- **V**ariance
- **É**cart-type
- **C**oefficient de Variation
- **I**QR (InterQuartile Range)

Position: Q-P-D

- **Q**uartiles (Q1, Q2, Q3)
- **P**ercentiles
- **D**éciles

Règle 68-95-99.7:

“68 à 1, 95 à 2, 99.7 à 3” - 68% dans $\mu \pm 1\sigma$ - 95% dans $\mu \pm 2\sigma$ - 99.7% dans $\mu \pm 3\sigma$

---

3. TESTS STATISTIQUES - “2 GROUPES = t, 3+ = ANOVA”

Tableau des Tests:

Situation	Test	Mnémotechnique
2 moyennes	**t-test**	“**T**wo groups = **T**-test”
3+ moyennes	**ANOVA**	“**A**nalyse de la **V**ariance pour **A**utres groupes”
Catégories	**Chi-carré**	“**Ch**i pour **Ch**oix catégoriels”
Corrélation	**Pearson/Spearman**	“**P**earson = **P**aramétrique, **S**pearman = **S**ans normalité”

p-value:

“P petit = Preuve Présente” - $p < 0.05 \rightarrow$ Rejeter $H_0 \rightarrow$ Significatif

---

4. PROBABILITÉS - “BICU”

Formule de Bayes: “Bayes Inverse les Conditionnelles”

$$P(A|B) = P(B|A) \times P(A) / P(B)$$

Distributions - “BPNE”

- **B**ernoulli → **B**inaire (0 ou 1)
- **B**inomiale → **B**eaucoup d’essais
- **P**oisson → **P**eu fréquent (événements rares)
- **N**ormale → **N**aturelle (la plus commune)
- **E**xponentielle → **E**spacement entre événements

---

5. SQL - “CTE avant WINDOW avant GROUP”

Window Functions: “RRDN-LLF”

- **R**OW_NUMBER → Unique
- **R**ANK → Saute les rangs
- **D**ENSE_RANK → Ne saute pas
- **N**TILE → Divise en groupes
- **L**AG → Ligne précédente
- **L**EAD → Ligne suivante
- **F**IRST/LAST_VALUE

Pattern CTE:

“WITH avant WHERE”

```
WITH cte AS (SELECT ...)
SELECT * FROM cte WHERE ...
```

Optimisation - “SIEWL”

- **S**ELECT colonnes spécifiques (pas *)
- **I**NDEX sur colonnes filtrées
- **E**XISTS plutôt qu’IN
- **W**HERE avant HAVING
- **L**IMIT pour tester

---

6. MACHINE LEARNING - “SNC-CRC”

Types d’apprentissage:

- **S**upervisé → avec **S**olutions (labels)
- **N**on supervisé → **N**o labels
- **S**emi-supervisé → **S**ome labels

Modèles de Classification: “L-A-R-G-E”

- **L**ogistic Regression
- **A**rbre de décision
- **R**andom Forest
- **G**radient Boosting
- **E**nsemble methods

Métriques: “PAR-FAR”

- **P**recision = TP / (TP + **F**P) → **P**armi les prédits positifs
- **R**ecall = TP / (TP + **F**N) → **R**écupérer tous les vrais positifs
- **A**ccuracy = (TP + TN) / Total
- **F**1 = 2 × P × R / (P + R)

AUC et Gini:

“Gini = 2 AUC moins 1”

```
Gini = 2 × AUC - 1
```

---

7. RISQUE DE CRÉDIT - “PLE” (Perte Liée à l’Exposition)

Expected Loss:

“EL = PLD” (Perte Liée au Défaut)

```
Expected Loss = PD × LGD × EAD
```

Composant	Signification	Mnémotechnique
PD	Probability of Default	**P**robabilité
LGD	Loss Given Default	**L**oss (perte)
EAD	Exposure at Default	**E**xposition

Métriques de Scoring: “GAK”

- **G**ini = 2 × AUC - 1 (discrimination)
- **A**UC = Aire sous ROC
- **K**S = Kolmogorov-Smirnov (séparation)

---

8. KPIs BANCAIRES - “RRNN-NCC-CLL”

Rentabilité: “ROE-ROA-NIM-CIR”

- **ROE** = Résultat / Equity (capitaux propres)
- **ROA** = Résultat / Assets
- **NIM** = Net Interest Margin
- **CIR** = Coût / Income

Qualité: “NPC”

- **N**PL = Non-Performing Loans / Total
- **P**rovision Coverage = Provisions / NPL
- **C**ost of Risk = Dotations / Encours

Solvabilité: “CAR-LDR-LCR”

- **CAR** = Capital / RWA (≥ 12% BRH)
- **LDR** = Loans / Deposits
- **LCR** = Liquidity Coverage Ratio

---

9. EDA - “CCANVD”

Checklist EDA:

“Comprendre, Charger, Analyser, Nettoyer, Visualiser, Documenter”

1. **C**omprendre le business
2. **C**harger les données
3. **A**nalyser la structure
4. **N**ettoyer (manquants, doublons)
5. **V**isualiser
6. **D**ocumenter

Traitement des Outliers: “IQR × 1.5”

```
Outlier si: x < Q1 - 1.5×IQR ou x > Q3 + 1.5×IQR
```

---

10. SEGMENTATION RFM - “Récent-Fréquent-Montant”

R-F-M:

- **R**ecency → Jours depuis dernière activité (inversé: 5=récent)
- **F**requency → Nombre de transactions
- **M**onetary → Montant total

Segments à retenir:

- **555** = Champions
- **5XX** = Actifs récents
- **X5X** = Fréquents
- **XX5** = Gros dépenseurs
- **111** = Perdus

---

11. DATA ENGINEERING - “EOLT”

ETL:

“Extraire-Transformer-Charger” - **E**xtract → Sources - **T**ransform → Nettoyage, agrégation - **L**oad → Data Warehouse

OLAP vs OLTP:

“OLAP = Analyse, OLTP = Transactions” - OLAP → Requêtes complexes, agrégations - OLTP → Transactions rapides, CRUD

---

12. VISUALISATION - “HBL-SPC”

Choix du graphique:

Type de données	Graphique	Mnémotechnique
Distribution continue	**H**istogramme	“**H**istogramme pour **H**istoire des valeurs”
Comparer groupes	**B**ar chart	“**B**ar pour **B**andes de catégories”
Tendance temporelle	**L**ine chart	“**L**igne pour le **L**ong terme”
Relation 2 variables	**S**catter	“**S**catter pour **S**auter entre points”
Composition	**P**ie	“**P**ie = **P**arts (< 6 catégories)”
Corrélation	**H**eatmap	“**H**eatmap pour **H**ot spots”

---

13. ENCODAGE - “One-Hot = Noms, Ordinal = Ordre”

Règle simple:

“One-Hot pour les Noms, Ordinal pour l’Ordre”

```python
# Nominale → One-Hot
pd.get_dummies(df, columns=['region'])

# Ordinale → Label Encoding ordonné
ordre = {'Faible': 1, 'Moyen': 2, 'Élevé': 3}
df['risque'] = df['risque'].map(ordre)
```

Data Leakage:

“Fit sur Train, Transform sur Test”

```python
scaler.fit_transform(X_train)  # FIT + TRANSFORM
scaler.transform(X_test)       # TRANSFORM seulement
```

---

14. FRAUDE - “Recall > Precision”

Règle d’or:

“En fraude, mieux vaut trop d’alertes que de rater une fraude”

- **Recall élevé** → Détecter toutes les fraudes
- **Precision faible** acceptable si ressources disponibles

Déséquilibre:

“SMOTE pour les PETITS” - SMOTE → Suréchantillonne la classe minoritaire

---

15. VALIDATION - “TTT-PSI”

Split temporel:

“Train sur le passé, Test sur le présent”

Monitoring:

“PSI > 0.25 = Problème” - PSI < 0.10 → OK - 0.10 ≤ PSI < 0.25 → Attention - PSI ≥ 0.25 → Action requise

---

16. FORMULES ESSENTIELLES

Statistiques:

Moyenne: $\bar{x} = \Sigma x_i / n$
Variance: $s^2 = \Sigma(x_i - \bar{x})^2 / (n-1)$
IC 95%: $\bar{x} \pm 1.96 \times (s/\sqrt{n})$
CV: $(s / \bar{x}) \times 100$

ML:

```
Gini = 2 × AUC - 1
F1 = 2 × (Precision × Recall) / (Precision + Recall)
EL = PD × LGD × EAD
```

Finance:

```
ROE = Résultat Net / Capitaux Propres
Variation % = (Nouveau - Ancien) / Ancien × 100
CAGR = (Vf/Vi)^(1/n) - 1
```

CHECKLIST FINALE - “12 Points Clés”

```
□ 1. NOIR → Types de variables
□ 2. MMM-VECI → Tendance centrale et dispersion
□ 3. t pour 2, ANOVA pour 3+ → Tests statistiques
□ 4. p < 0.05 = Significatif
□ 5. EL = PD × LGD × EAD → Risque crédit
□ 6. Gini = 2×AUC - 1 → Performance scoring
□ 7. ROE, ROA, NPL, CAR → KPIs bancaires
□ 8. One-Hot/Ordinal → Encodage
□ 9. Fit Train, Transform Test → Data leakage
□ 10. SMOTE pour minorité → Déséquilibre
□ 11. Recall pour fraude → Priorité détection
□ 12. PSI > 0.25 = Drift → Monitoring
```

RÉVISION EXPRESS (15 min avant l’examen)

Top 10 à ne pas oublier:

1. **Variable nominale** = pas d’ordre (type compte)
2. **Variable ordinale** = ordre naturel (rating)
3. **Médiane** pour outliers/asymétrie
4. **Chi-carré** pour indépendance catégories
5. **Gini = 2×AUC - 1**
6. **EL = PD × LGD × EAD**
7. **CAR ≥ 12%** (exigence BRH)
8. **NPL** = prêts > 90 jours
9. **Fit sur TRAIN uniquement**
10. **Recall** prioritaire en fraude

VOUS AVEZ LES OUTILS - CONFIANCE!