

Référentiel d'activités, de compétences et d'évaluation

CERTIFICAT *ANALYSE DE DONNEES AVEC LE LANGAGE R*

I. Référentiel d'activités, de compétences et d'évaluation

Le titulaire de la certification sera capable, dans une démarche de data science, d'intégrer et de traiter des données, et de produire un rapport en utilisant de manière appropriée le langage de programmation R. Le titulaire de la certification sera capable de mettre en place un processus de traitement des données sous R qui soit reproductible avec différents types de jeux de données.

Compétences attestées :

Intitulé de la certification		
Analyse statistique de données avec le langage R		
Description de l'activité		
Le titulaire de la certification est capable de mettre en place un processus de traitement des données sous R qui est reproductible avec différents jeux de données		
Référentiels		
Activités	Compétences	Modalités d'évaluation
A1. Importer un jeu de donnée dans l'environnement R - Configuration de l'environnement de travail dans le logiciel R - Chargement des packages adaptés - Importation du fichier de données - Vérification de la qualité de l'import - Typage des données (ce qu'est chaque colonne : date, numérique, texte...)	Importer des données dans un environnement R adapté, en utilisant le(s) package(s) adéquat(s) et les lignes de code adaptées aux données, afin de vérifier leur qualité et réaliser leur mise en forme.	Cas pratique : Sur la base d'un jeu de données, contenant des erreurs, le candidat doit : - Importer les données ; - Corriger les erreurs ; - Décrire et compiler les données dans un rapport mis en forme selon des spécifications précises. Critère général de forme : Le candidat respecte les bonnes pratiques d'écriture et de commentaires des lignes de code.
A2. Visualiser un jeu de donnée dans l'environnement R - Production de graphiques et tableaux à partir des données	Visualiser des données importées dans l'environnement R sous forme de graphiques, tableaux et d'indicateurs statistiques adaptés à leur nature (moyennes, médianes et écarts-types...)	

<ul style="list-style-type: none"> - Description statistique des données : calcul des paramètres descriptifs pertinents (effectifs, moyennes, médianes, écarts-types...) 		
<p>A3. Nettoyer un jeu de donnée dans l'environnement R</p> <ul style="list-style-type: none"> - Interprétation des indicateurs statistiques (moyennes, médianes, écarts-types) - Identification des erreurs de saisie ou de collecte - Correction des erreurs par codage en langage R - Visualisation des données à l'aide des lignes de codes appropriés - Recherche d'erreurs non détectées lors des vérifications initiales - Corrections des nouvelles erreurs détectées - Transformation des données pertinentes pour rendre le jeu de données complètement exploitable 	<p>Interpréter les indicateurs statistiques produits dans l'environnement R pour identifier des erreurs de saisie ou de collecte et les corriger à l'aide des lignes de code appropriées, afin de rendre le jeu de données complètement exploitable</p>	
<p>A4. Produire un rapport à partir d'un jeu de donnée dans l'environnement R</p> <ul style="list-style-type: none"> - Choix des variables et indicateurs statistiques selon la commande ou l'intérêt des données - Sélection des formats de visualisation des variables et indicateurs statistiques - Production des tableaux et graphiques attendus ou pertinents pour la compréhension et l'exploitation des données - Mise en forme des tableaux et graphiques (titre, couleur, traits, proportions...) 	<p>Sélectionner des variables et d'indicateurs statistiques pertinents en fonction des données ou de la commande, et produire des tableaux et des graphiques mis en forme, permettant de comprendre et interpréter les données, en rédigeant les lignes de codes adaptées, afin de produire un rapport pour un tiers.</p>	