

STAT2430: Calcul statistique sur ordinateur

SAS Enterprise Guide - SAS partie 2

Enterprise Guide (EG) est une interface graphique créée par SAS pour l'utilisateur final, qui n'est pas censé connaître le langage de programmation SAS. L'environnement permet d'effectuer la majorité des tâches courantes : importation de jeux de données, statistiques de base, graphiques, sorties (format HTML). Pour un statisticien, EG s'avère trop limité (principalement pour la gestion des données), aussi est-il indispensable d'apprendre le langage SAS. Néanmoins, EG allège agréablement le lourd labeur de mémorisation du code complexe pour les analyses statistiques et surtout les graphiques. Il s'avère particulièrement puissant lorsque mis en liaison avec SAS : le code généré sous EG peut être récupéré sous SAS (par copier-coller) et de très légères modifications fournissent immédiatement les analyses/graphiques souhaités sous SAS. Enfin, la liaison peut être utilisée dans l'autre sens : EG propose la possibilité d'inclure votre propre code SAS.

Nous allons nous servir du fichier texte SFOSCH.XLS, qui contient diverses informations sur les vols partant de San Francisco. Vous pouvez ouvrir ce fichier avec Excel pour avoir un aperçu de sa structure. Toutes les variables contenues dans le fichier sont reprises dans le tableau suivant:

Nom de la variable	Description	Colonne	Type de variable
FLIGHTID	Numéro d'identification du vol	1	Texte
ROUTEID	Identification du trajet	2	Texte
ORIGIN	Origine du vol	3	Texte
DESTINATION	Destination du vol	4	Texte
MODEL	Modèle d'avion	5	Texte
DEPARTDATE	Date de départ	6	Texte 01JAN2000
DEPARTDAY	Jour de départ (lundi, mardi...)	7	Numérique
FCLASSPASS	Nbre de pass. 1ere classe	8	Numérique
BCLASSPASS	Nbre de pass. classe business	9	Numérique
ECLASSPASS	Nbre de pass. classe économique	10	Numérique
TOTPASSCAP	Capacité totale de l'avion	11	Numérique
CARGOWT	Poids de l'avion	12	Numérique

Exercice 1 – Création d'un projet sous Enterprise Guide

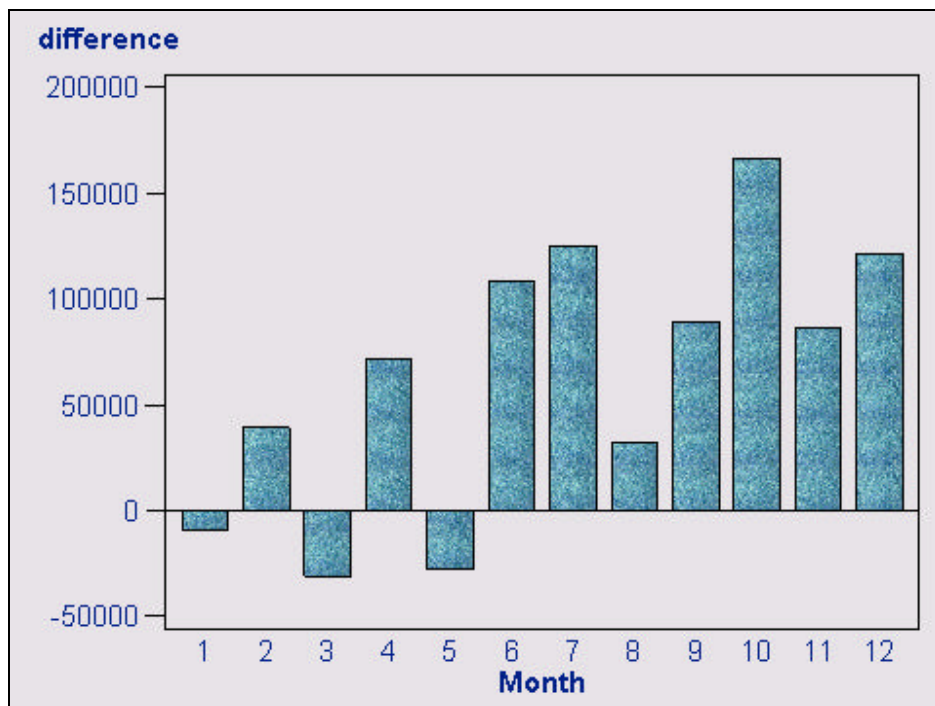
1. Lancer Enterprise Guide. Créer un nouveau projet.
2. Importer les données du fichier SFOSCH.XLS (fichier Excel), en utilisant le menu Insert... Data.
3. Etudier le résultat de l'importation. S'intéresser aux symboles de colonne (pyramide rouge ou sphère bleue). Que signifient-ils? De quel type est la variable departdate? Regarder les propriétés du fichier importé (clic droit sur le nom de fichier, Properties).

Exercice 2 – Première tâche / exportation d'un jeu de données

1. La liste des tâches (en bas à gauche) rassemble toutes les fonctionnalités d'Enterprise Guide, rangées par ordre alphabétique ou par catégories.
2. Trouver la tâche qui permet de centrer réduire (normaliser ou encore standardiser) des variables numériques. Pour ajouter une tâche au projet en cours, double cliquer dessus. A l'aide de l'assistant, standardiser toutes les variables numériques du fichier importé.
3. Lorsque la tâche est exécutée, trois objets dépendant de la tâche sont ajoutés au projet: le code généré pour réaliser la tâche, le log de SAS où sont affichées les erreurs d'exécution et enfin un nouveau fichier de données (appelé standardized sfosch).
4. Aller voir le log. Y a-t-il des erreurs lors de l'exécution?
5. Exporter le fichier généré (au format Excel ou TXT). Pour ce faire, il faut effectuer un clic avec le bouton droit de la souris sur le nom du fichier généré, pour appeler le menu contextuel et choisir [Save as / Export].

Exercice 3 – Ajout de données / Création d'un premier graphique

1. Ajouter au projet le dataset **europa_compare**, qui fournit pour chaque mois de l'année les budgets alloués pour l'Europe et les dépenses effectuées, ainsi que la différence (pertes ou gains pour l'entreprise).
2. Réaliser un graphique en barres pour représenter la différence selon le mois de l'année. Vérifier toutes les options, pour obtenir le graphique suivant :



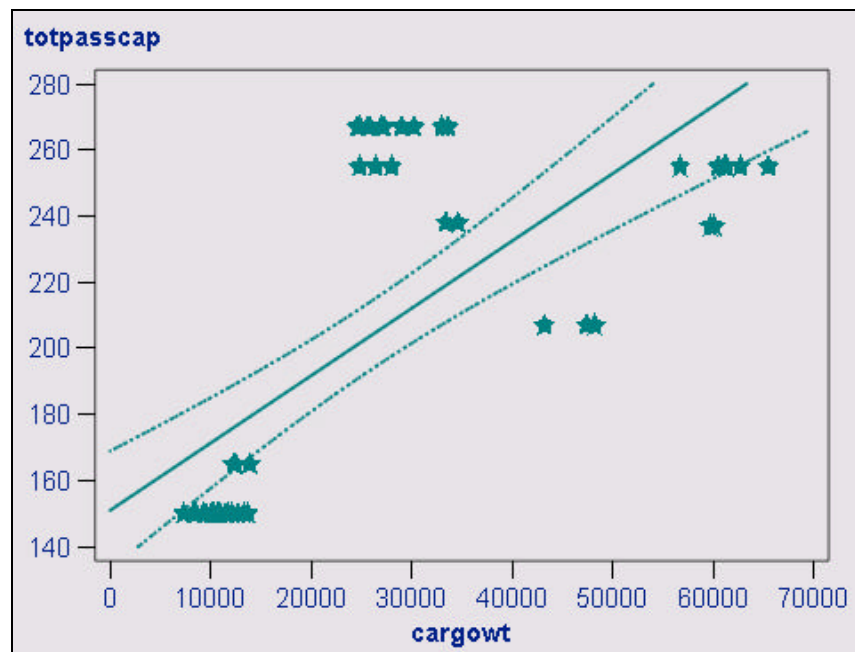
3. Copier le graphique et le coller dans un fichier Word. Attention : pour n'avoir aucun problème lors du transport du fichier sur un autre ordinateur, il est préférable d'utiliser un collage spécial (Edit... Paste Special) et de coller le graphique en tant qu'image.

Exercice 4 : récupération d'un code généré par EG pour réaliser un graphique

On s'intéresse maintenant à nouveau au fichier **sfosh.xls**, qui reprend les caractéristiques de vols partant de l'aéroport de San Francisco pour le mois de décembre 1999.

1. Dans Enterprise Guide, effectuer un graphique de régression de la capacité de l'avion (totpasscap) (appeler le menu Graph... Line..., dans la liste des graphiques, choisir le graphique de régression). Demander l'affichage des bandes de confiance autour de la droite de régression.

Votre graphique sous EG doit ressembler à ceci :



2. Lancer SAS. Ecrire le code SAS pour importer le fichier excel dans l'environnement SAS. Il est possible d'utiliser l'assistant d'importation SAS (File... Import Data) et de récupérer le code généré.
3. Copier le code généré par Enterprise Guide pour effectuer le dernier graphe dans votre programme SAS. Le modifier pour qu'il fonctionne avec votre dataset (changer le dataset utilisé dans la procédure GPLOT). Exécuter le code. Le graphique ressemble-t-il au graphique généré par Enterprise Guide ?
4. Supprimer les options de symbole et relancer. Que se passe-t-il ? Isoler l'option qui permet d'ajouter à un graphique la droite de régression et trouver dans l'aide d'autres fonctions similaires. Comment faire un boxplot en se servant d'une telle option ?

Exercice 5 – Outil statistique : le test t de Student

1. Ajouter au projet le dataset **employee**. Visualiser son contenu.
2. Avant d'entamer une grève, certains membres du personnel aimeraient trouver une bonne raison (!). L'argument avancé est une disparité de salaires entre hommes et femmes. Effectuer sous SAS EG un test t de Student avec graphiques pour comparer sur base des données les salaires des hommes et des femmes. (Utiliser t-test dans la liste des tâches) Quelle est la conclusion ?

3. Préparer pour le directeur des ressources humaines un programme SAS exécutable sur son ordinateur qui lui fournit un rapport en HTML lorsque exécuté.

Exercice 6 – Ajouter du code SAS à un projet EG

Retour au fichier **sfosh.xls**, déjà importé dans votre projet.

1. Insérer dans votre projet le code SAS pour ajouter aux données une nouvelle variable TOTPASS, qui compte le nombre total de passagers (ECLASSPASS+BCLASSPASS+FCLASSPASS). Calculer aussi la variable PERTES qui renseigne le nombre de places inoccupées par vol (TOTPASSCAP - TOTPASS).
2. Lancer une procédure descriptive sur la variable PERTES (statistiques + graphiques). S'assurer qu'il y a des avions avec des pertes négatives. Comment interpréter ces valeurs?
3. Ajouter un nouveau code SAS pour éliminer ces observations aberrantes.
4. Enfin, préparer un programme SAS qui prépare un rapport avec des statistiques descriptives sur les pertes (moyenne, médiane, ... + graphiques).

Exercice récapitulatif

L'exercice est à réaliser dans un projet EG. Incorporer le dataset **retiree_close**, recensant les informations des employés de la compagnie proches de la retraite. Le directeur des ressources humaines souhaite disposer d'un rapport comptabilisant pour les années à venir le nombre d'employés qui va partir à la retraite et le coût total pour la compagnie (qui doit verser 3 mois de salaire en prime de départ à chaque employés).

Directives: chaque employé travaille jusqu'à 65 ans. En se servant de la date de naissance, créer une nouvelle variable qui donne l'année de départ de l'employé. Puis effectuer les calculs suivants pour chaque année (instruction BY). Pour le rapport du DRH, ajouter des graphiques permettant de visualiser l'évolution des coûts engendrés par ces départs à la retraite.