

## **PROJET D'ANALYSE STATISTIQUE DE DONNEES REELLES**

Dr. Patricia CONDE-CESPEDES

### **1. PREAMBLE**

#### **1.1 OBJECTIVE**

L'objectif de ce projet est d'appliquer vos connaissances théoriques acquises en statistique inférentielle avec un jeu de données réelles.

Le projet comprend 3 parties :

1. Statistiques descriptives
2. Estimation ponctuelle et par intervalles de confiance
3. Test d'hypothèses

#### **1.2 CRITERES D'EVALUATION**

Votre apport personnel, motivation et contribution seront prises en compte pour la notation. Elles peuvent même donner lieu à des points supplémentaires.

#### **1.3 LE RAPPORT**

Une attention particulière sera accordée à la présentation du rapport final. La langue du rapport peut être l'anglais ou le français. Dans tous les cas, la qualité de rédaction sera appréciée. Les phrases doivent être claires, explicites et bien compréhensibles.

S'il y a une preuve de copie, la note finale sera divisée parmi tous les auteurs.

Le rapport final au minimum doit contenir les parties suivantes :

- Une page couverture,
- Une table des matières, une introduction,
- Le corps du rapport expliqué dans les sections suivantes (résultats, figures, tableaux, etc.),
- La conclusion et
- Les annexes pour le code.

Dans le corps du rapport, toutes les figures, tables, sorties logiciel doivent être **commentées, interprétées et bien expliquées**. Aucun résultat non-interprété sera pris en compte.

Le nombre de pages ne doit pas dépasser 25 pages et 10 pages pour l'annexe.

La page de garde doit contenir le prénom, le nom et le numéro ISEP de tous les auteurs.

#### 1.4 REMISE DU RAPPORT

Le rapport final doit être envoyé à l'adresse : [patricia.conde-cespedes@isep.fr](mailto:patricia.conde-cespedes@isep.fr) au plus tard le **26 janvier 2018**. Le rapport final doit être envoyé en format pdf. Le nom du fichier doit être le suivant :

*NomEtudiant1\_NomEtudiant2\_NomEtudiant3.pdf*

#### 1.5 SOUTENANCE DU PROJET

Une soutenance orale aura du projet aura lieu le **1<sup>er</sup> février 2018** matin. Les détails vous seront communiqués ultérieurement.

### 2. DESCRIPTION DES DONNEES

Les données contenues dans le fichier «*Data.txt*» ont été extraites de la base de données d'une entreprise fabricante de lunettes de correction. Les individus sont les verres sortis de l'atelier de surfacage. Le but est de modéliser le résultat du contrôle de qualité en fonction des paramètres de fabrication.

Le surfacage consiste à donner la forme géométrique adéquate au verre pour qu'il puisse avoir la correction demandée. La matière première est un verre semi-fini sur lesquels il faut surfer la face arrière de telle sorte que le changement de direction des rayons lumineux traversant le verre satisfasse la correction commandée.

Après le surfacage tous les verres sont lavés et passent après par un robot de contrôle (il y en a deux). Le robot de contrôle juge la qualité du verre en fonction de ses caractéristiques optiques et cosmétiques.

Les données mises à disposition concernent une journée de fabrication. Chaque ligne correspond à un verre. Le jeu des données contient les résultats du contrôle d'environ 1000 verres.

## 2.2 PREPARATION DES DONNEES

Pour juger la qualité des verres les robots doit mesurer les variables suivantes (entre autres) : sphère, cylindre, axe, Addition et épaisseur. Pour chacune de ces variables le jeu de données contient la valeur commandée par le client et puis la valeur mesurée par le robot.

Par exemple, dans le jeu de données on trouvera la variable *Sphere* et *MesureSphere*, toutes les deux mesurées en dioptries.

Calculez pour chaque caractéristique optique une nouvelle variable qui correspond à l'écart entre la valeur mesurée et la valeur demandée, par exemple :

- $\Delta Sphere = MesureSphere - Sphere$

Ignorez les valeurs manquantes.

## 2.1 STATISTIQUES DESCRIPTIVES

Réalisez une analyse de statistiques descriptives complète. Calculez pour cela des statistiques descriptives de position et de dispersion pour les variables *DeltaSphere* et *DeltaCylindre* pour chaque type de verre indiqué par la variable *Type*. Commentez et interprétez les résultats obtenus.

## 2.2 ESTIMATION PONCTUELLE

**Questions de cours :** Quelles méthodes d'estimation ponctuelle connaissez-vous ? Faire une brève description de chaque méthode.

Que peut-on dire sur les propriétés des estimateurs ? Analysez en fonction du biais et de la convergence des estimateurs.

Faites une estimation ponctuelle de l'espérance et des variables *DeltaSphere* et *DeltaCylindre* pour les deux types de verre 'ORG' et 'OXL'. Nous supposons que les variables sont issues d'une population Gaussienne.

## 2.3 ESTIMATION PAR INTERVALLES DE CONFIANCE

Construisez un intervalle de confiance pour l'espérance et la variance des variables *DeltaSphere* et *DeltaCylindre* par type de verre (donc 4 intervalles de confiance). Vous utiliserez un niveau de signification  $\alpha=5\%$  et  $\alpha=1\%$ . Interprétez les résultats et commentez les différences obtenues avec les différents niveaux de signification. Vous devez clairement expliquer toutes les hypothèses que vous utiliserez pour la construction de ces intervalles.

## 2.4 TEST D'HYPOTHESES

La variable *DeltaSphere*, représentant l'écart entre la valeur de demandée et la valeur mesurée, représente une mesure de la qualité du verre. Dans le cas idéal, *DeltaSphere* est nul. Si la valeur mesurée s'écarte significativement de la valeur demandée, le responsable de l'atelier de surfacage doit mettre en place une vérification de caractéristiques techniques de fabrication fournies à la machine.

Pour les deux types de verre (ORG et OXL), construisez un test d'hypothèse pour rejeter/pas rejeter l'hypothèse nulle suivante :

- $H_0$  : l'écart entre la valeur demandée et la valeur fabriquée est nul.

Concluez si l'écart est statistiquement différent de zéro et en fonction de vos résultats donnez un conseil au responsable surfacage.

Vous devez clairement expliquer toutes les hypothèses sur la loi de l'estimateur que vous utiliserez pour la construction de ces tests d'hypothèse.

## 2.5 TEST DE COMPARAISON DE MOYENNES

Dans cette partie nous voulons savoir si la distribution de la variable *DeltaCylindre* dépend du type de verre. Pour cela vous allez effectuer un test de Student de comparaison de moyennes pour les verres de type ORG et les verres de type OXL.

Dans un premier temps vous allez donner l'expression de la statistique à calculer et sa loi. Vous allez décrire toutes les hypothèses que les deux populations doivent vérifier pour pouvoir appliquer ce test.

## 2.6 CONCLUSIONS

Dans cette partie vous allez faire un résumé des résultats trouvés et des conclusions.

Nous avons supposé que les observations étaient dépendantes et qu'elles étaient issues d'une population Gaussienne. Comment changerait l'analyse si les variables provenaient d'une distribution non-gaussienne ?

## 3. BIBLIOGRAPHIE

N'oubliez pas d'inclure la bibliographie dans votre rapport. Toute copie totale ou partielle d'une source externe sera considérée du plagiat.