

Statistique inférentielle

TP bilan

Un compte rendu est à rendre à la fin du TP. Il s'agira d'un rapport au format PDF, incluant les programmes informatiques R. Le TP est à réaliser en binôme. La qualité de la rédaction (programmes et rapport) sera prise en compte. Le compte rendu sera envoyé à l'adresse **gwandji@polytech-lille.fr**.

- Nom du fichier : **TP_SI_Binôme1_Binôme2.pdf**
- Intitulé du mail : **BILAN SI 2018**

Exercice 1 : maximum de vraisemblance

Les données suivantes représentent la durée de vie expérimentale d'un composant électronique observé sur un échantillon de mesures :

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
durée de vie	48.80	7.30	27.30	53.30	25.70	59.40	42.90	63.70	68.40	19.80

On suppose que cette durée de vie notée $D \sim \mathcal{W}(\beta, k)$ de densité à support sur \mathbb{R}_+ .

1. On suppose k connu, donner un estimateur de β
2. Estimer β, k en utilisant la méthode du maximum de vraisemblance.

Exercice 2 : Contrôle de fabrication par mesures

Les données ci-dessous représentent les poids en grammes de 99 pots de chocolats sortis d'une machine à conditionner :

	CL 1	CL 2	CL 3	CL 4	CL 5	CL 6	CL 7	CL 8	CL 9	CL 10	CL 11
effectifs	1	2	6	13	20	21	18	11	5	1	1
poids	500.50	501.00	501.50	502.00	502.50	503.00	503.50	504.00	504.50	505.00	505.50

1. Peut-on assimiler les données à une loi normale ?
2. Proposer un intervalle de confiance de niveau 95% pour :
 - Le poids moyen des boîtes
 - La variance du poids des boîtes

Préciser à chaque fois la statistique utilisée pour construire l'intervalle de confiance, ainsi que sa loi.

Exercice 3 : Gravité des blessures d'accidents de la route

Une étude sur des accidents de voiture a noté la gravité des blessures en fonction du port de la ceinture (Oui vs. Non).

	aucune	minimale	légère	grave
Oui	128213	647	359	42
Non	65963	4000	2642	303

1. Calculer le pourcentage des différents types d'accidents en fonction du port de la ceinture de sécurité, et représenter graphiquement les données ?
2. Le port de la ceinture de sécurité influence-t-il la gravité des blessures ? Préciser le test statistique mis en oeuvre ici.

Exercice 4 : Suivi de sinistres des clients d'une banque

Le fichier `incident.csv` contient des informations concernant des clients fragiles d'une institution financière. Un chargé d'études statistiques de l'organisme de crédit cherche à estimer la durée moyenne en mois entre le moment où un client a été jugé fragile et la survenance du premier incident de paiement. On suppose que ces durées sont distribuées normalement.

1. Importer le fichier de données et effectuer les analyses statistiques (statistiques descriptives et représentations graphiques) des différentes variables.
2. Peut-on affirmer au risque $\alpha = 5\%$ que la durée moyenne avant la survenance du premier incident de paiement est :
 - supérieure à 2 mois ?
 - inférieure à 4 mois ?

En déduire un encadrement de la durée moyenne avant le premier incident de paiement. Donner un intervalle de confiance de niveau 95% de la durée moyenne avant la survenance du premier incident de paiement. Discuter.

La population est constituée de trois catégories de clients, les clients dit "surendettés", les clients dit "restructurés" et les clients "standards".

3. Donner des intervalles de confiance de niveau 99% de la moyenne et de l'écart-type de la durée avant la survenance du premier incident de paiement pour chaque type de client.
4. Peut-on dire qu'il y a une différence entre la variance des durées avant la survenance du premier incident de paiement pour :
 - Les clients surendettés et les clients restructurés ?
 - Les clients surendettés et les clients standards ?
 - Les clients restructurés et les clients standards ?
5. Peut-on affirmer qu'il y a une différence entre la durée moyenne avant la survenance du premier incident de paiement pour :
 - Les clients surendettés et les clients restructurés ?
 - Les clients surendettés et les clients standards ?
 - Les clients restructurés et les clients standards ?

Exercice 5 : Données accouchements prématurés

Le fichier `prematures.csv` contient des observations des facteurs prénataux (médicaux et de personne) liés à un accouchement prématuré chez des femmes déjà en travail prématuré.

Les variables du fichier sont les suivantes :

GEST : l'âge gestationnel en semaines à l'entrée dans l'étude

DILATE : la dilatation du col en cm

EFFACE : l'effacement de col (en pourcentages)

CONSIS : la consistance du col (1=mou, 2=moyen, 3=ferme)

CONTR : la présence (=1) ou non (=2) de contraction

MEMBRAN : les membranes ruptures (=1) ou non (=2) ou incertain (=3)

AGE : âge de la patiente

GRAVID : la gestite (nombre de grossesses antérieures y compris celle en cours).

DIAB : la présence (=1) ou non (=2) d'un problème de diabète, ou valeur manquante (=3).

TRANSF : transfert (1) ou non (2) dans un hôpital de soins spécialisés

GEMEL : grossesse simple (=1) ou multiple (2)

PREMATURE : accouchement prématuré (positif ou négatif)

1. Effectuer une analyse statistique (statistiques descriptives et représentations graphiques) des variables du jeu de données.
2. Estimer à partir des données brutes la proportion de femmes ayant une période de gestation de moins de 31 semaines.
3. Peut-on considérer que la moitié des femmes ayant déjà connues un accouchement prématuré ont une période de gestation inférieure à 31 semaines ? Donner la région critique du test, sa p-valeur.
4. La durée moyenne de gestation est-elle différente entre les femmes ayant déjà connues un accouchement prématuré et celle qui n'en ont jamais eu ?
5. Existe-il une relation entre :

- l'accouchement prématurée et la type de grossesses (gémellaires oui/non) ?
- Les membranes rupturées et la consistance du col ?