

Gestion des Données Industrielles Statistiques

Rapport Final

Date limite de soumission: 11/10/2019 à 23h59

Groupe 2



Maxime HERVIER
ZHANG Quan
ZHAO Yue
HE Jianjin

Table des matières

Contents

1	Introduction	3
2	Suivi et explication des travaux	4
2.1	Nettoyage de la base de données	4
2.2	Caractérisation des variables	4
2.3	Dégager la problématique	5
2.4	Aborder les statistiques descriptives	5
2.5	Statuer sur les tests à mener	7
3	Déroulement des tests	8
3.1	Test (1)	8
3.2	Test (2)	9
4	Conclusion	10
5	Perspectives	11

1 Introduction

Dans le contexte d'une étude pour le développement d'un nouveau modèle de défonceuse, il nous est demandé de mener une étude statistique sur un ensemble de clients potentiels pour optimiser l'ergonomie du futur produit.

Un sondage est donc mené sur un échantillon de 120 individus de gabarits variés portant sur les mesures de leur main, un test de force avec prise en compte de l'effort ressenti et une préférence cosmétique entre trois propositions de formes de gâchette présentées sous cette forme :

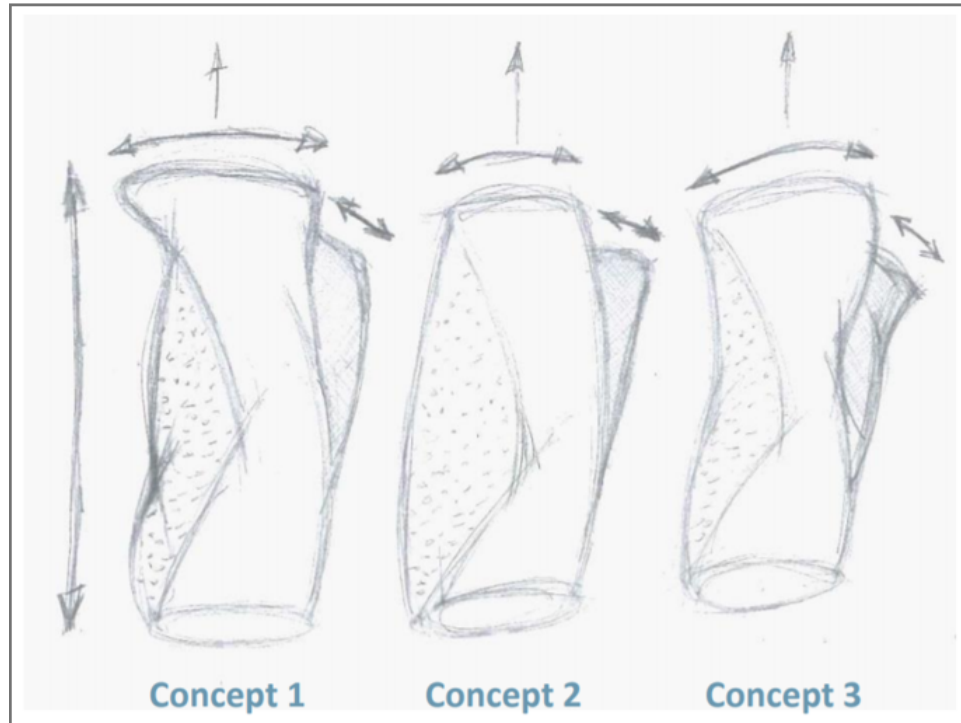


Figure 1: Les trois concepts

2 Suivi et explication des travaux

Comme chaque étude statistique, la première phase des travaux consiste en un nettoyage des données recueillies, de sorte à disposer d'une base de données que l'on puisse considérer comme fiable pour mener nos études.

2.1 Nettoyage de la base de données

Ce processus consiste en la suppression systématique des données présentant les défauts suivants :

- Des erreurs de syntaxe, y compris de simples problèmes d'ordre de grandeur (les mesures recueillies devant être spécifiées en millimètres alors que certains participants les ont reportées en centimètres) ;

41	Femme	17	8	16
42	Homme	185	80	170

Figure 2: Cas d'un problème d'ordre de grandeur

- Des erreurs sémantiques qui, dans le cas des données recueillies, semblent être absentes mais seraient, de toute manière, inidentifiables ;
- Des erreurs "outlier" caractérisant des valeurs aberrantes, que l'on repère grâce au reste des données. Certaines de ces erreurs proviennent manifestement de la confusion entre les pouces et les centimètres au moment de la mesure (les règles distribuées possédant les repères des deux unités) ;

110	Homme	180	85	190
111	Femme	84	93	83

Figure 3: Cas d'erreur "outlier" dû aux unités

- Des valeurs manquantes dans certaines variables, ce qui n'a pas été le cas dans notre tableau ;

Cette phase, une fois achevée, entraîne la suppression de 37 lignes sur les 120 recueillies soit environ 30.8% de la base de données initiales.

2.2 Caractérisation des variables

Une fois notre base de données nettoyée, il convient, pour mener l'étude de manière efficace, d'identifier la nature des données recueillies.

On identifie alors les variables comme telles :

- Longueur de la main : quantitative continue ;
- Largeur de la paume : quantitative continue ;
- Longueur pouce-index : quantitative continue ;
- Effort couple : quantitative continue ;
- Effort ressenti : qualitative ordinale ;
- Concept préféré : qualitative nominale.

En ce qui concerne les variables qualitatives obtenues lors de la mesure, l'excès de zèle de certains participants qui ont reporté des valeurs très précises rend l'ensemble des variables quantitatives continues.

2.3 Dégager la problématique

Le but de cette étude est, rappelons-le, de collecter des données dans le but d'optimiser un nouveau modèle de défonceuse que l'on souhaite commercialiser. Dans ces conditions, il nous faut formuler explicitement la question à laquelle cette collecte de données doit nous aider à répondre. En l'occurrence:

Comment adapter les caractéristiques de la défonceuse en adéquation avec les “capacités” (performances physiques) des utilisateurs testés et leur préférence esthétique ?

2.4 Aborder les statistiques descriptives

Avant d'évaluer l'existence de tendances conjointes entre plusieurs variables, il est intéressant de mener des analyses univariées puis bivariées en fonction du sexe des participants sur les variables quantitatives de sorte à dégager des informations utiles au développement de notre produit. Par exemple, la dimension moyenne de la main des utilisateurs potentiels de la défonceuse constitue une information essentielle au dimensionnement de la gâchette des poignées, qui devront également tenir compte de la variance de cette même mesure.

Initialement, on détermine les statistiques descriptives sur l'ensemble de notre échantillon nettoyé, sans tenir compte des valeurs des variables qualitatives :

	A	B	C	D	E	F	G	H	I	J	K	L
1	Longueur de la main - L_Main (mm)			Largeur de la paume - L_Paume (mm)			Longueur Pouce-Index - L_PouceIndex (mm)			Couple - E_Couple (mm)		
2												
3	Moyenne	192,01013		Moyenne	88,818987		Moyenne	193,42785		Moyenne	29,149367	
4	Erreur-type	1,3300409		Erreur-type	0,8779366		Erreur-type	1,6412989		Erreur-type	1,1725159	
5	Médiane	190		Médiane	90		Médiane	195		Médiane	30	
6	Mode	190		Mode	90		Mode	190		Mode	30	
7	Écart-type	11,821662		Écart-type	7,8032714		Écart-type	14,588184		Écart-type	10,421549	
8	Variance de l'échantillon	139,75169		Variance de l'échantillon	60,891045		Variance de l'échantillon	212,81511		Variance de l'échantillon	108,60869	
9	Kurtosis (Coefficient d'aplatissemen	0,0667345		Kurtosis (Coefficient d'aplatissemen	0,3388371		Kurtosis (Coefficient d'aplatissemen	-0,3925246		Kurtosis (Coefficient d'aplatissemen	-0,3280258	
10	Coefficient d'asymétrie	-0,0312373		Coefficient d'asymétrie	-0,1745517		Coefficient d'asymétrie	-0,0956385		Coefficient d'asymétrie	0,0443678	
11	Plage	60		Plage	40		Plage	70		Plage	45	
12	Minimum	160		Minimum	70		Minimum	160		Minimum	10	
13	Maximum	220		Maximum	110		Maximum	230		Maximum	55	
14	Somme	15168,8		Somme	7016,7		Somme	15280,8		Somme	2302,8	
15	Nombre d'échantillons	79		Nombre d'échantillons	79		Nombre d'échantillons	79		Nombre d'échantillons	79	
16	Niveau de confiance(95,0%)	2,647908		Niveau de confiance(95,0%)	1,7478376		Niveau de confiance(95,0%)	3,2675752		Niveau de confiance(95,0%)	2,3342998	
17												
18												

Figure 4: Statistiques descriptives générales par variables

On obtient ainsi des tendances très générales qui méritent du raffinement par analyses bivariées. On conduit alors les mêmes calculs pour chaque variable en dissociant les données relatives aux participantes et aux participants:

	A	B	C	D	E
1	ueur de la main - L_Main (mm)		Longueur de la main - L_Main (mm)		
2					
3	Moyenne	178.75		Moyenne	194.385075
4	Erreur-type	3.14576435		Erreur-type	1.26770635
5	Médiane	180		Médiane	193
6	Mode	170		Mode	190
7	Écart-type	10.8972474		Écart-type	10.3766237
8	Variance de l'	118.75		Variance de l'	107.674319
9	Kurstosis (Coe	-0.78065978		Kurstosis (Coe	-0.30580428
10	Coefficient d'	0.00395166		Coefficient d'	0.34164024
11	Plage	35		Plage	45
12	Minimum	160		Minimum	175
13	Maximum	195		Maximum	220
14	Somme	2145		Somme	13023.8
15	Nombre d'éch	12		Nombre d'éch	67
16	Niveau de cor	6.92378065		Niveau de cor	2.53105739
17					
18	Femme			Homme	

Figure 5: Variable : Longueur de la main

	A	B	C	D	E
1	eur de la paume - L_Paume (mm)		Largeur de la paume - L_Paume (mm)		
2					
3	Moyenne	79.25		Moyenne	90.5328358
4	Erreur-type	1.96223821		Erreur-type	0.81617354
5	Médiane	80		Médiane	90
6	Mode	80		Mode	90
7	Écart-type	6.79739255		Écart-type	6.68066836
8	Variance de l'	46.2045455		Variance de l'	44.6313297
9	Kurstosis (Coe	0.08846439		Kurstosis (Coe	0.74000102
10	Coefficient d'	0.39167532		Coefficient d'	0.11465754
11	Plage	23		Plage	35
12	Minimum	70		Minimum	75
13	Maximum	93		Maximum	110
14	Somme	951		Somme	6065.7
15	Nombre d'éch	12		Nombre d'éch	67
16	Niveau de cor	4.31885718		Niveau de cor	1.62954305
17					
18	Femme			Homme	

Figure 6: Variable : Largeur de la paume

	A	B	C	D	E	F
1	ur Pouce-Index - L_PouceIndex (mm)		Longueur Pouce-Index - L_PouceIndex (mm)			
2						
3	Moyenne	182.583333		Moyenne	195.370149	
4	Erreur-type	3.82467732		Erreur-type	1.71490528	
5	Médiane	183		Médiane	195	
6	Mode	190		Mode	190	
7	Écart-type	13.2490709		Écart-type	14.0371047	
8	Variance de l'	175.537879		Variance de l'	197.040308	
9	Kurstosis (Coe	-0.97448014		Kurstosis (Coe	-0.42895264	
10	Coefficient d'	-0.28521384		Coefficient d'	-0.08264799	
11	Plage	40		Plage	65	
12	Minimum	160		Minimum	165	
13	Maximum	200		Maximum	230	
14	Somme	2191		Somme	13089.8	
15	Nombre d'éch	12		Nombre d'éch	67	
16	Niveau de cor	8.41805801		Niveau de cor	3.42391886	
17						
18	Femme			Homme		

Figure 7: Variable : Longueur pouce-index

	A	B	C	D	E
1	Effort ressenti - E_Ressenti		Effort ressenti - E_Ressenti		
2					
3	Moyenne	7		Moyenne	5.91044776
4	Erreur-type	0.65133895		Erreur-type	0.21971091
5	Médiane	7.5		Médiane	6
6	Mode	9		Mode	7
7	Écart-type	2.2563043		Écart-type	1.79841128
8	Variance de l'	5.09090909		Variance de l'	3.23428313
9	Kurstosis (Coe	-0.99336735		Kurstosis (Coe	-0.20492758
10	Coefficient d'	-0.51284876		Coefficient d'	0.02522943
11	Plage	7		Plage	8
12	Minimum	3		Minimum	2
13	Maximum	10		Maximum	10
14	Somme	84		Somme	396
15	Nombre d'éch	12		Nombre d'éch	67
16	Niveau de cor	1.43358736		Niveau de cor	0.43866698
17					
18	Femme			Homme	

Figure 8: Variable : Effort Ressenti

on obtient ainsi des données plus spécifiques au sexe des participants qui permettent, dans un premier temps, avant de mener des tests plus précis, de percevoir grossièrement l'existence de tendances spécifiques à chaque sexe de sorte à déterminer s'il faut prévoir des tests de corrélation spécifiques.

Aussi, dans une optique purement informative, on peut établir une répartition de la préférence du concept au sein de l'échantillon:

4	Nombre de Concept préféré - P_Concept	Étiquettes de colonnes			
5	Étiquettes de lignes	Concept 1	Concept 2	Concept 3	Total général
6	Femme	5	1	6	12
7	Homme	28	3	36	67
8	Total général	33	4	42	79
9					

Figure 9: La répartition du concept préféré parmi les sujets de l'échantillon en fonction de leur sexe

On observe notamment que le Concept 3 emporte 53% des préférences générales (50% des préférences

féminines et 54% des préférences masculines), ce qui en fait le concept à privilégier lors de la future phase de conception du produit.

2.5 Statuer sur les tests à mener

Avant de proposer des hypothèses à tester il convient de mentionner les limites de l'échantillon dont nous disposons.

En premier lieu, comme l'indique l'illustration précédente, dans notre échantillon seulement 4 participants ont choisi le concept 2, ce qui, dans l'éventualité d'un test de corrélation entre variables qualitatives, ne permet pas de satisfaire au critère du test Chi^2 exige un effectif minimum de 5 par modalité de la variable de réponse. On extrapole alors en concluant que notre échantillon ne nous permet pas de tester des liens entre variables qualitatives avec la préférence du concept.

Dans la mesure où l'on mène cette étude pour dimensionner un futur produit (notre défonceuse initiale), on cherche en priorité à déterminer si l'adaptabilité aux différents gabarits des utilisateurs doit également s'accommoder d'une adaptabilité mécanique en relation directe avec cette dernière ; en des termes plus simples : est-il nécessaire d'adapter l'effort de couple nécessaire à l'actionnement de l'appareil en relation avec le gabarit de l'utilisateur ?

Dans le contexte de notre étude, on peut ramener les tests de corrélation gabarit/effort déployé à une unique variable dans la mesure où la longueur de la main, la largeur de la main et la longueur pouce-index varient globalement proportionnellement.

On propose donc de tester les hypothèses suivantes :

- La longueur de la main n'a pas d'influence sur l'effort ressenti.(1)
- La longueur de la main a une influence sur l'effort ressenti.(2)

Le problème (1) est traité en vertu des hypothèses suivantes :

H0 : La longueur de la main n'a pas d'influence sur l'effort ressenti.

H1 : La longueur de la main a une influence sur l'effort ressenti.

Conformément aux variables du problème, le test à effectuer est une corrélation de Pearson.

Le problème (2) est traité en vertu des hypothèses suivantes :

H0 : La longueur de la main n'a pas d'influence sur l'effort ressenti.

H1 : La longueur de la main a une influence sur l'effort ressenti.

Conformément aux variables du problème, le test à effectuer est une corrélation de Spearman.

4 Conclusion

Au regard de l'étude menée, les constatations suivantes émergent :

- La tendance générale indique une certaine préférence pour le troisième concept de gâchette proposé, qu'il convient de prendre en compte lors de la phase de conception du produit ;
- La vaste disparité des gabarits des mains des participants de notre échantillon (présentant une plage de 3.5 cm) nous laisse entrevoir la nécessité de prévoir un dispositif susceptible d'être utilisé confortablement par des mains de dimensions assez variables ;
- Les tests statistiques menés n'indiquent pas de corrélation manifeste entre le gabarit des mains des sujets testés et les performances de force de ces derniers. En conséquence, il ne paraît pas nécessaire de prévoir des modes d'utilisations ajustant la résistance des gâchettes en fonction de la taille de l'utilisateur.

Au regard du processus de l'étude statistique elle-même :

- Il apparaît qu'une grande partie du temps consacré à l'étude consiste en le nettoyage de la base de données qui, de surcroît, exige des critères de sélection préalables. La conséquence de l'aboutissement de ses critères est qu'un fort niveau d'exigence sur la fiabilité des données entraîne une réduction significative (dans notre cas pratiquement un tiers) de la base de données initiales ;
- À ce titre il paraît donc préférable, et surtout plus efficace, de disposer d'une importante base de données de sorte à rendre l'étude la plus représentative possible. Notons à ce titre qu'une base de données trop restreinte peut même contraindre à renoncer à certains tests en cause d'une trop faible représentativité (se référer au cas du test χ^2 en partie 2.5).

5 Perspectives

À partir des conclusions tirées dans la partie précédente, certaines pistes de raffinement des données de l'étude apparaissent :

- Si la préférence générale pour le troisième concept de gâchette proposé apparaît suffisamment significative pour le privilégier lors de la conception du produit, il serait intéressant de mener un sondage plus détaillé sur les raisons de cette apparente préférence et déterminer en fonction de quels critères les utilisateurs potentiels sont attirés par l'apparat cosmétique de l'objet. En particulier, discuter de ce genre de détails auprès d'ergonomes paraît essentiel pour délivrer un produit optimal ;
- En termes de représentativité, on aurait tendance à vouloir mener une étude similaire (aux hypothèses identiques pour le moins) auprès d'un échantillon plus conséquent pour préciser davantage les résultats. Une base de données exploitables de 83 éléments est en effet assez restreinte au regard de la portée des utilisateurs potentiels.

En ce qui concerne le processus d'établissement des statistiques en lui-même, on comprend, à la sortie de cette étude, l'importance de définir préalablement un problème à traiter de sorte à bien définir les tests à mener sur les variables étudiées : mener une étude statistique rigoureusement permet de préciser davantage les résultats obtenus.

Aussi, une fois ces résultats recueillis, il est très important de les communiquer aussi clairement et précisément que possible, de sorte à laisser le moins de possibilités de spéculation possibles sur le sens des résultats auprès des lecteurs de l'étude qui n'ont pas accès aux démarches de celles-ci. Lors de la communication des résultats, se faire comprendre adéquatement est aussi important que de disposer de données fiables.