FALLET Laurent
PERRIN Guillaume

UV Technique des Statistiques

Etude de la répartition des salaires à la sortie de l'INSA de Rouen

Juin 2002

Tout d'abord, nous désirons adresser des remerciements chaleureux aux personnes suivantes : Peggy LEGUILLON, Machocki WOJCIECH, Rose-Marie LIEVROUW, Sylvie MULLER et Mohamed AIT OUFKIR pour leur aide à distance et leurs chaleureux encouragements. Ils nous ont en effet permis de sonder un plus grand nombre de personnes, en nous donnant de bonnes adresses électroniques ou des méthodes efficaces.

- Donc pour retrouver d'anciens camarades de promotion :

Un bon site : www.copainsdavant.com
Liste du groupe d'Insaïens (élèves et ingénieurs) sur Yahoo :
insalien@yahoogroupes.fr

- Quelques extraits de commentaires :

« Félicitations, très très bon sujet qui passionnera tous le monde :))) »

« Je suis volontaire […] sur l'île de la Réunion […] donc je ne suis pas vraiment un travailleur. »

« Je ne connaît pas mon salaire brut désolé, j ai donc arrondi à la louche »

« Ne vous lancez pas dans la chimie, il n y a pas de boulot en France!! »

« Bonne idée pour ce foutu projet a la c... !! Bon courage ! »

« Bonne journée et bien le bonjour à M barbe de ma part. » - Sylvain Lemasson -

« Bon courage les gars .... Une bise a Roger :-)) » - Anonyme -

« Que c'était beau la deuxième année!!! »

« En espérant ne pas fausser vos données... »

Sommaire

Introduction

60 % des 16-25 ans affirment avoir pour but, dans la vie, de gagner de l'argent (d'après un sondage du Ministère de la Jeunesse de des Sports). Cela révèle un intérêt certain des jeunes sur le montant futur de leur salaire, qui décidera de leur niveau de vie.

Or les passionnés de M8 n'échappent pas aux statistiques… Nous nous sommes déjà interrogés sur le fruit de notre travail assidu et acharné : combien ?

Un "combien" qui, malgré les apparences, se révèle plus ambigu que l'on ne pourrait le penser. Il nous a donc fallu étudier les différents aspects de la question et dégager une tendance générale.

Ce projet est une illustration de la définition des **Statistiques** :

Branche des mathématiques qui a pour objet la collecte, le traitement et l'analyse de données numériques relatives à un ensemble d'individus ou d'éléments. Les statistiques constituent un outil précieux pour l'expérimentation de projets, la gestion des entreprises ou encore l'aide à la décision. Une recherche statistique se décompose en quatre étapes : la définition et la collecte des données, leur présentation en tableaux, leur analyse et enfin la comparaison des résultats avec des lois statistiques connues.

********************* *Delirium* *********************

La statistique est moins une science qu'un art. Elle est la poésie des nombres. Chacun y trouve ce qu'il y met.

Brie, Albert

Il faut bien que tout le monde vive.
Et comme il faut bien que tout le monde meure, ça fait une moyenne.

Léo Campion

Nos jours sont comptés : par des statisticiens.

Stanislaw Jerzy Lec

Un sondage n'est pas un substitut à la réflexion.

Warren Buffet

****

I - L'enquête

1 – Objectif

Nous avons choisi de réaliser cette enquête sur les salaires à la sortie de l'école afin de nous rendre compte de la réalité du marché.

En effet, nous avons voulu vérifier par nous même ce que beaucoup de magazines à vocation économique ou destinés aux étudiants et lycéens proposent tout au long de l'année : une grille des salaires d'embauche.

Notre étude se base sur la récupération de "profils" (ensemble de caractères) auprès des promotions récemment sorties de l'INSA de Rouen.

Nous chercherons à déterminer un salaire moyen d'embauche, en prenant en compte des éléments non négligeables tels que la situation et le statut professionnel, le lieu de travail ou bien encore la taille des entreprises dans lesquelles les jeunes diplômés évoluent.

Nous comparerons les effectifs des différentes classes de salaires que nous avons définies avec la loi normale, ce qui nous permettra de déterminer, pour les nouveaux diplômés, la probabilité d'être payé plus de 30000 € par an.

Enfin, nous comparerons nos résultats avec une étude nationale réalisée par l'INSEE et une étude du magazine L'expansion.

Il est nécessaire de signaler que nous remercions tous les ingénieurs qui ont pris part à notre sondage. En contrepartie, ce rapport est disponible sur Internet, à l'adresse www.thelorenzo.fr.st , pour qu'ils puissent voir les résultats. A noter que cette adresse est la même que celle du sondage, et que les ingénieurs en ont été avertis (de plus ils nous ont maintes fois demandé d'avoir accès aux résultats).

2 – Moyens mis en œuvre et techniques de récupération :

En France, demander ouvertement le montant du salaire d'une personne est taxé d'indiscrétion. Il est logique que la rémunération de chacun soit tenue privée, et nous ne pouvions donc téléphoner ou démarcher une à une les personnes que nous voulions sonder. Il a donc fallu trouver un moyen pour récupérer le montant des salaires de manière totalement anonyme. Or les moyens classiques, tels que téléphone, mail, permettent sans mal de retrouver le nom d'une personne.

Nous avons de ce fait mis en place un système d'envoi de mails et de redirection vers un site que nous avons conçu (en HTML et PHP). Ce site récupère les données indiquées par les ingénieurs sondés et nous les envoie par mail. Voici le formulaire en question :

Vous êtes...		○ **Un homme** ○ **Une femme** ◉ **Ne souhaite pas répondre**
Promotion :		[▾]
Département :		[▾]
Avez vous effectué un 3ème cycle spécialisé?		○ **Oui** ◉ **Non**
A propos de votre période d'essai (ou stage ingénieur):	**Ce fut:**	○ **Une réussite (emploi à la clé)** ○ **Un échec (cause employeur)** ○ **Une expérience (vous êtes parti(e))** ○ **Autre arrangement...**
Situation professionnelle à la sortie de l'INSA :		[▾]
Expérience professionnelle :		[▾]
Secteur d'activité de l'entreprise:		[▾]
Taille de l'entreprise :		[▾]
Lieu de travail :		[▾]
Rémunération brute annuelle à la sortie de l'INSA : *(ex: 31 710 €)*	[ ]	○ **francs** ○ **euros**
Commentaires: *(si vous désirez nous laisser un message, un conseil, une dédicace...)* *N'hésitez pas à indiquer tout ce qu'il manque dans le formulaire!!!* *Ne pas indiquer votre nom svp...*		[ ]

Il va sans dire que la mise en page était plus soignée sur le site, afin d'être agréable à l'œil.

Nous avons essayé de ne pas forcer la main au niveau des réponses, c'est-à-dire de toujours laisser le choix de ne pas répondre ou la possibilité de nous expliquer en détail (cf. section commentaires).

C'est ainsi que les profils "exotiques" ont pu se loger dans les catégories « Autre » ou « Ne Se Prononce Pas ». Le salaire pouvait être indiqué en francs ou euros, cela va sans dire que nous nous sommes chargés de la conversion…

En bref, le but est de rendre le remplissage de l'enquête le plus rapide (moins de 5 minutes ici) par des méthodes simples :

- Questions courtes et simples ;
- Réponses courtes, fermées, choix limités ;
- Informations logiques, sues par cœur (base du CV) ;

Ces règles de base ne sont pas uniquement utiles pour la conception du sondage et son utilisation, mais aussi pour l'exploitation des données.

3 – Ambiguïté de l'enquête

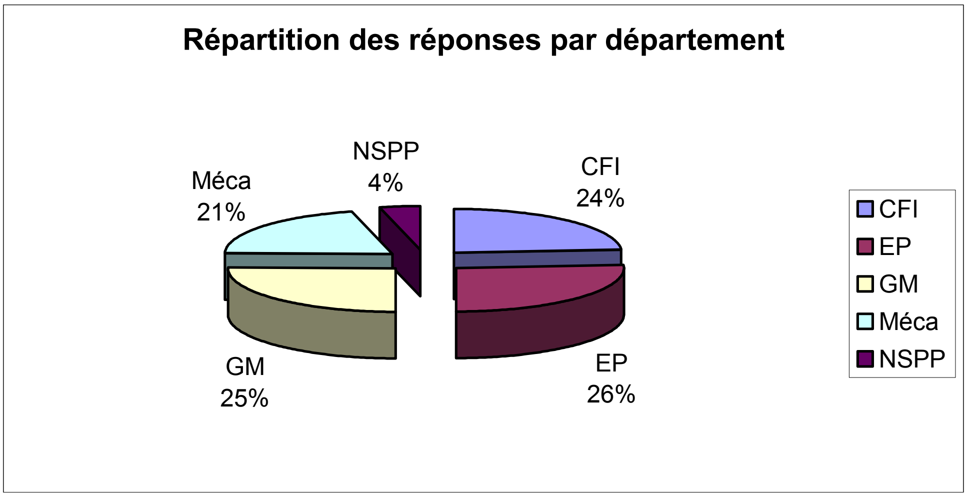
La matière première des statistiques est constitué d'ensembles de nombres. Il convient en premier lieu de s'assurer de l'exhaustivité et de la justesse des informations recueillies.

Un problème auquel nous avons été confrontés est de définir la nature et la quantité des données à recueillir.

La quantité :

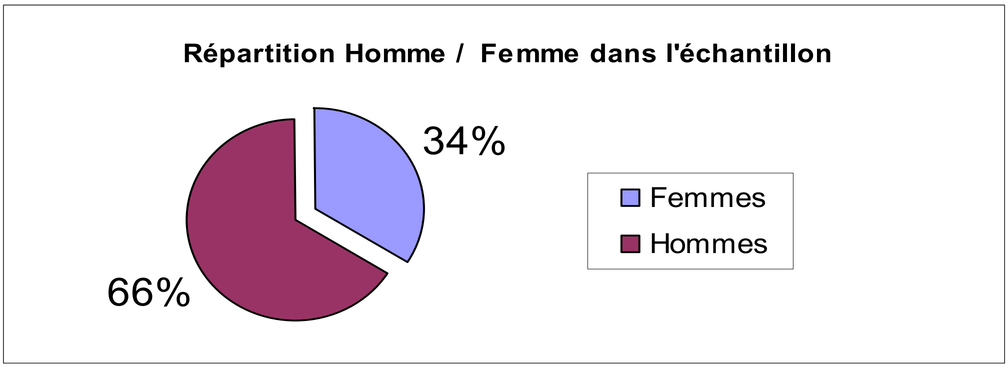
La quantité globale de réponses devait être strictement supérieure à 30, avec un minimum acceptable à 50. En effet, sans une nombre suffisant de réponses, toute étude est inutile, car non représentative ou non modélisable. Nous avons recueilli en tout 130 réponses, sur un total de 252 mails valides (l'adresse existe réellement) et envoyés, ce qui donne un pourcentage d'efficacité de 51,6%. Cela signifie que les ingénieurs sont attachés à leur école et aux travaux des élèves, et qu'ils prennent le temps de répondre à l'enquête.

Notre échantillon devait de représenter la population de l'INSA dans son ensemble, soit un nombre semblable de réponse dans les 4 départements. La méthode théorique pour obtenir un tel résultat est de considérer le pourcentage de réponse indépendant du département, et d'envoyer autant de mails dans un département que dans tous les autres. Nous avons procédé ainsi, et le nombre de réponses parvenues dans chaque département se répartit comme suit :



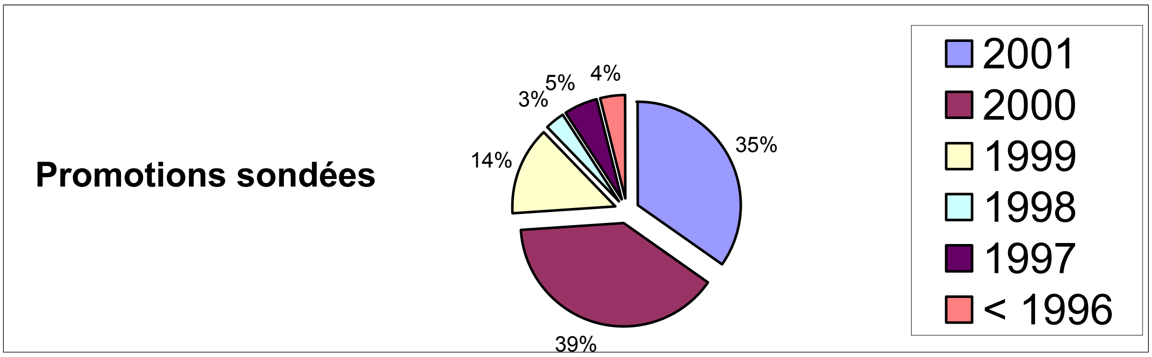
Nous considérons, au vu des résultats, que l'équité entre départements a été respectée.

D'un autre coté, la parité hommes femmes est moins respectée :



Cela peut être du au fait que les femmes sont moins touchées par la passion de l'informatique, donc elles sont un moins grand nombre à posséder une adresse mail, et de ce fait, nous en avons moins sondées. D'autre part, les promotions recrutées il y a 5 ans et plus étaient moins féminisées que maintenant.

Répartition des réponses par promotions :



La qualité

La plus grosse erreur à faire dans un sondage est de laisser libre la réponse à donner. Dans le cas de caractères qualitatifs, les modalités sont non mesurables (cf. cours). Dans la pratique, demander si l'emploi est apprécié et donner des choix tels que « beaucoup », « moyennement », « pas vraiment », « pas du tout » ne mène à rien. D'une part on ne peut tirer de conclusion ou de modélisation sur les réponses, et d'autre part les gens ne savent jamais quoi répondre.

Le traitement des données, pour être exact, exige parfois de vérifier la cohérence d'une réponse et son exactitude. Dans notre enquête, plusieurs vérifications et rectifications ont été effectuées. Par exemple, un ingénieur de la promo 2001 ne peut avoir 5 ans d'expérience ou déjà fait un 3ème cycle. Il n'a donc pas d'expérience si il est en recherche d'emploi, et si il est en 3ème cycle il est en étude, donc ni en CDD ou CDI. Ces erreurs dans les données ne peuvent être corrigées qu'en se fiant à un critère, ici la promotion.

Les données que nous avons recueillies sont du type quantitatif continu (pour les salaires). D'où la nécessité d'un découpage en classes. De plus nos données semblent suivre une loi normale, mais ceci doit être démontré.

Tout ceci sera étudié dans la section suivante…

FALLET Laurent
PERRIN Guillaume

II - Modélisation et Outils Mathématiques

1 – Vérification de l'hypothèse de la loi normale :

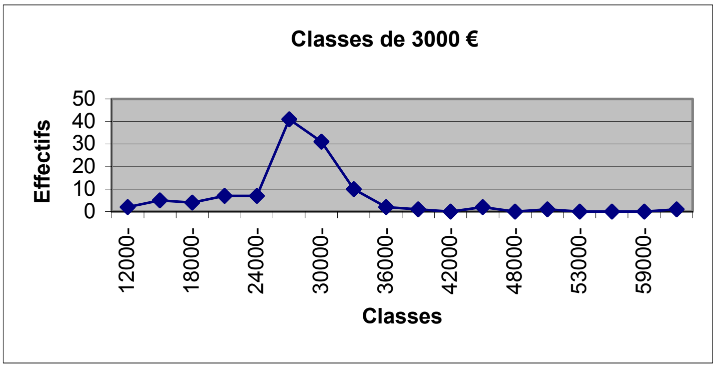
- Statistiques de base sur les salaires de notre échantillon global

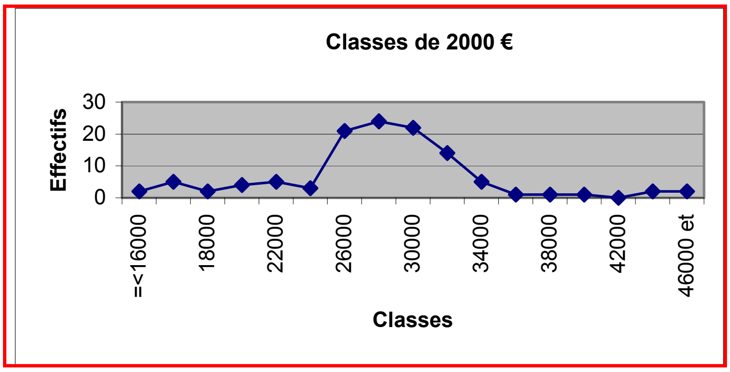
Nous n'avons, évidemment, pas indiqué nos 130 valeurs, ce qui nous semble inutile et inintéressant.

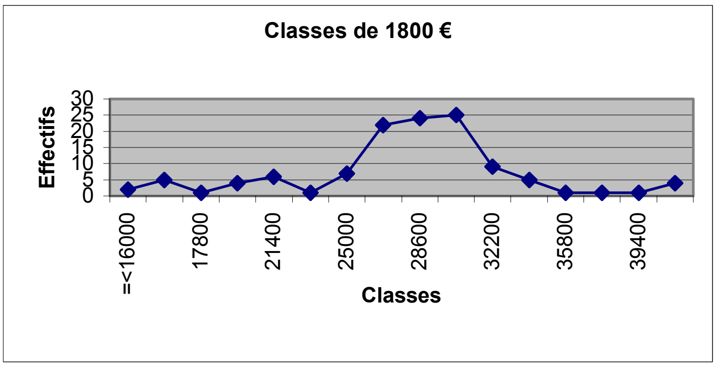
Moyenne : 29 114,19 €
Médiane : 29 050,00 €
Ecart-type : 6 624,00 € (arrondi à l'euro près car la valeur grande).
Mode : 27 400,00 €

- Les classes

Dans un premier temps, étude des différentes classes possibles :







Analyse :
On observe que les classes de 3000 et 1800 ne sont pas les plus proches de la courbe théorique (courbe de la loi normale), c'est pourquoi nous choisirons des **classes de 2000€.**

Nombre de classes : 17
Ce nombre est satisfaisant, pas excessif.

En **page suivante** : vérification de la normalité de la loi sur du papier gausso-arithmétique, à partir des effectifs cumulés.

Virer cette feuille et mettre celle du papier gausso-arithmétique.

Interprétation :

La courbe ci avant n'est pas rigoureusement droite. Cela peut être du à la taille de notre échantillon (130 éléments). En effet si nous avions eu 500 éléments la courbe eut été plus droite.

NB : le point d'effectif cumulé 100% n'est pas indiqué car la dernière classe comporte les salaires supérieurs à 46 000 €, qui sont dispersés donc non classables.

Nous avons tracé deux droites (une rouge et une verte) :

- La droite rouge prend tous les points en compte et est la moins inclinée. On observe environ 6 points au dessus de la courbe et autant en dessous. Il n'y a pas de points réellement aberrants.

- La droite verte prend en compte les classes de salaire se trouvant entre 24000 et 36000 € uniquement. La droite passe quasiment par tous les points, ce qui est très satisfaisant.

Conclusion :

Nos chiffres peuvent être modélisés par une loi normale, tout en gardant à l'esprit qu'une étude plus étendue serait plus précise.

Nous n'effectuerons pas de comparaison avec la loi log-normale car cette dernière est valable uniquement en entreprise, non pour un vivier d'ingénieurs dont certains sont encore en études (donc peu payés).

2 – Estimation d'une proportion avec un intervalle de confiance.

On choisit un intervalle de confiance à 95%.

L'écart-type du caractère **X** dans la population sera $\sigma = \sqrt{pq}$ où $q = 1 - p$. En effet, si la population est de taille n, **X** prend les valeurs x_i dans {0; 1} et on a donc :

$$\sigma^2 = \frac{\sum x_i^2}{n - p^2} = pq$$

Alors, la variable moyenne des échantillons $\widetilde{X}$ aura les caractéristiques suivantes :
Moyenne : *p*

Ecart-type : $\sqrt{\frac{pq}{n}}$

La variable centrée réduite Z de $\widetilde{X}$ est telle que l'on a :

$Z = \dfrac{\widetilde{X} - p}{\sqrt{\frac{pq}{n}}}$ suit une loi normale $N(0,1)$ si $n \geq 30$
suit une loi de Student $T(n-1)$ si $n < 30$

Soit u la valeur lue dans la table et telle que :

$$P(|Z| \leq u) = 95\% \iff P\left(\left|\widetilde{X} - p\right| \leq u\sqrt{\frac{pq}{n}}\right) = 95\%$$

On obtient alors l'intervalle de confiance sous la forme : $C = \widetilde{x} \pm u\sqrt{\frac{\widetilde{x}(1-\widetilde{x})}{n}}$ avec $\widetilde{x}$ les valeurs de la moyenne observée dans l'échantillon.

3 – Comparaison de deux moyennes

a) Pour des échantillons supérieurs à 30

On réalise un test d'égalité de deux moyennes sur des échantillons de taille différente :

Soit un échantillon A :

de moyenne m_1 et d'écart-type σ_1
de taille $n_1 \geq 30$
de variable moyenne d'échantillon $\widetilde{X}_1$

Soit un échantillon B :

de moyenne m_2 et d'écart-type σ_2
de taille $n_2 \geq 30$
de variable moyenne d'échantillon $\widetilde{X}_2$

On pose $D = \widetilde{X}_1 - \widetilde{X}_2$ suit une loi normale $N\left(m_1 - m_2 ; \sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}\right)$

On connaît σ_1 et σ_2 des échantillons, on choisi d'utiliser les valeurs estimées : $\hat{\sigma}_1^2 = \sigma_1^2 \times \frac{n_1}{n_1 - 1}$ et $\hat{\sigma}_2^2 = \sigma_2^2 \times \frac{n_2}{n_2 - 1}$

On obtient : $\sqrt{\frac{\sigma_1^2}{n_1 - 1} + \frac{\sigma_2^2}{n_2 - 1}} = \sigma_D$

Hypothèse H_0 : $m_1 = m_2 = m$

On fixe le seuil de confiance à 95%.
D suit une loi normale $N(0 ; \sigma_D)$
Si $|D| > \alpha$ on réfute H_0.

Si $|D| < \alpha \quad \Leftrightarrow \quad |Z| = \frac{|D|}{\sigma_D} > \frac{\alpha}{\sigma_D} = u$

$P(|Z| \leq u) = 95\%$ on trouve u puis α ,

Si $|d| = |\widetilde{x}_1 - \widetilde{x}_2| > \alpha$ on réfute H_0 et on sait que l'hypothèse est fausse dans 95% des cas.

b) pour des échantillons inférieurs à 30

On réalise un test d'égalité de deux moyennes sur des échantillons de taille différente :

Soit un échantillon A :

de moyenne m_1 et d'écart-type σ_1
de taille $n_1 < 30$
de variable moyenne d'échantillon $\widetilde{X}_1$

Soit un échantillon B :

de moyenne m_2 et d'écart-type σ_2
de taille $n_2 < 30$
de variable moyenne d'échantillon $\widetilde{X}_2$

On pose $D = \widetilde{X}_1 - \widetilde{X}_2$
On connaît σ_1 et σ_2 des échantillons, on choisi d'utiliser les valeurs estimées : $\hat{\sigma}_1^2 = \sigma_1^2 \times \dfrac{n_1}{n_1 - 1}$ et $\hat{\sigma}_2^2 = \sigma_2^2 \times \dfrac{n_2}{n_2 - 1}$

On obtient : $\sqrt{\dfrac{\sigma_1^2}{n_1 - 1} + \dfrac{\sigma_2^2}{n_2 - 1}} = \sigma_D$

On pose $Z = \dfrac{D}{\sqrt{\dfrac{\sigma_1^2}{n_1 - 1} + \dfrac{\sigma_2^2}{n_2 - 1}}}$ suit une loi de Student $T_{n_1 + n_{2-1}}$

Hypothèse H$_0$: $m_1 = m_2 = m$

On fixe le seuil de confiance à 95%.
Z suit une loi de Student $T_{n_1 + n_{2-1}}$
Si $|D| > \alpha$ on réfute H$_0$.
Si $|D| < \alpha \quad \Leftrightarrow \quad |Z| = \dfrac{|D|}{\sigma_D} > \dfrac{\alpha}{\sigma_D} = u$

$P(|Z| \leq u) = 95\%$ on trouve u puis α ,

Si $|d| = |\widetilde{x}_1 - \widetilde{x}_2| > \alpha$ on réfute H$_0$ et on sait que l'hypothèse est fausse dans 95% des cas.

III – Les Résultats

1 – Etude des salaires

Nous avons déjà réalisé l'étude des salaires et déduit que ceux-ci suivent une loi Normale $N\left(29114{,}2\,;\,6624\right)$.

Mais il serait intéressant de connaître la probabilité d'être payé plus de 30 000 € à la sortie de l'INSA de Rouen…

Soit X la variable aléatoire du salaire :

$$p(X > 30000) = p\left(S > \frac{30000 - 29114{,}2}{6624}\right) = 1 - p(S \leq 0{,}134) = 1 - 0{,}554 = 0{,}446$$

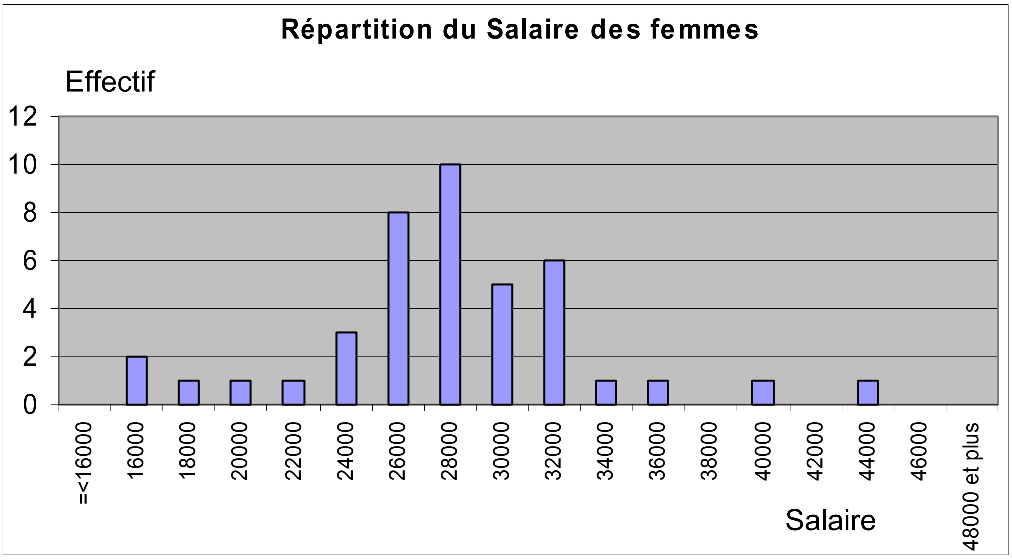
On a donc **44,6%** de "chance" d'être payé plus de 30 000 €.

Il serait maintenant bon de savoir si cette probabilité est la même dans toutes les catégories que nous avons sondées.

2 – Sexe et Salaire

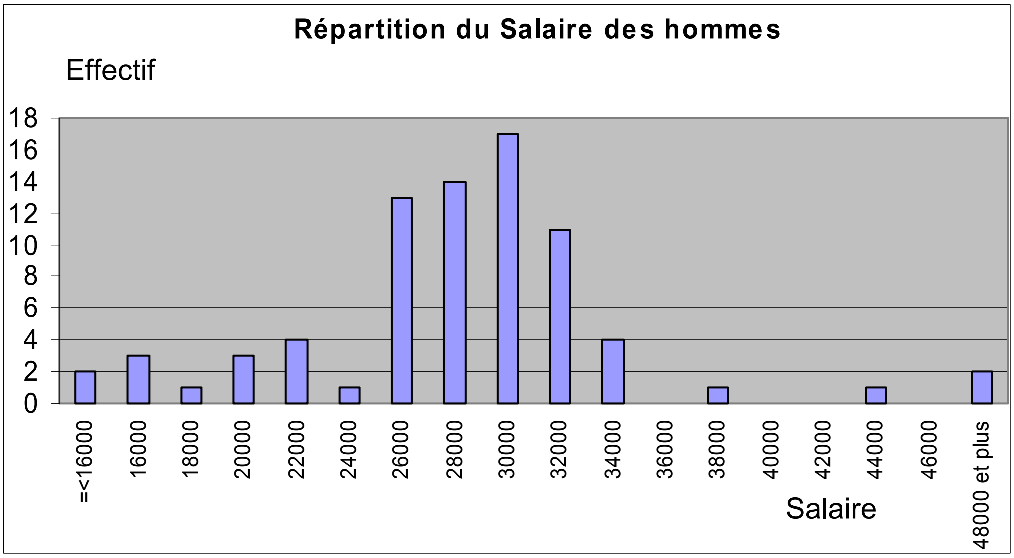
On étudie la répartition des salaires entre femmes et les hommes. Comme indiqué dans la première partie, il se pose le problème des échantillons qui ont des tailles différentes.

L'étude révèle une répartition des salaires comme suit :



Moyenne : 28899,02€
Médiane : 29 000,00€
Ecart-type : 5290,69€
Mode : 29000,00€

La classe médiane est la classe [28000 ; 30000].



Moyenne : 29 228,77€
Médiane : 29 400,00€
Ecart-type : 7264,22€
Mode : 27400,00€

La classe médiane est la classe [30000 ; 32000].

Afin de pouvoir comparer les résultats obtenus, on utilise la méthode de comparaison de deux moyennes d'échantillons pour des échantillons de taille supérieure à 30 puisque nous avons :

$n_1 = 41$ le nombre de femmes ayant répondu à cette question.
$n_2 = 77$ le nombre d'hommes ayant répondu à cette question.

On considère que les variables moyennes d'échantillons suivent des lois normales :

Alors :
$D = \left|\tilde{X}_1 - \tilde{X}_2\right| \quad suit \quad N(m_1 - m_2\ ; \sigma_D)$

$\sigma_D = 1180{,}7$ On choisit u = **1, 96** et on obtient $\alpha = 2314$ $\quad |d| = 329{,}75 \le \alpha$ donc pour un seuil de confiance à 95% la moyenne des salaires des hommes est égale à la moyenne des salaires des femmes.

On peut conclure que les entreprises tendent de nos jours à respecter la parité homme femme au niveau des salaires.

Etudions maintenant les probabilités pour une femme puis pour un homme de gagner plus de 30 000 € à la sortie de l'école :

X_1 la variable aléatoire du salaire des femmes, X_2 celle du salaire des hommes.

$$p(X_1 > 30000) = p\left(S_1 > \frac{30000 - 28899{,}02}{5290{,}69}\right) = 1 - p(S_1 \le 0{,}208) = 1 - 0{,}582 = 0{,}418$$

La probabilité d'être payé plus de 30 000 € pour une femme à la sortie de l'INSA est de **41,8%**.

$$p(X_2 > 30000) = p\left(S_2 > \frac{30000 - 29228{,}77}{7264{,}22}\right) = 1 - p(S_2 \le 0{,}106) = 1 - 0{,}542 = 0{,}458$$

La probabilité d'être payé plus de 30 000 € pour un homme à la sortie de l'INSA est de **45,8%**.

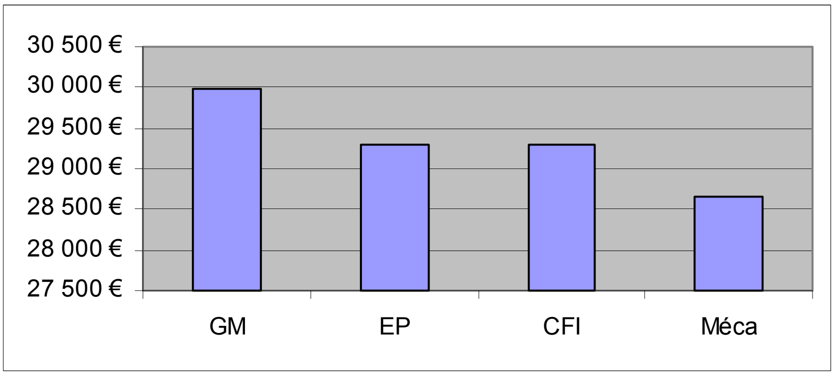
Même si la moyenne de salaire d'embauche semble la même pour les hommes et les femmes, on remarque une légère différence dans la probabilité d'être payé plus de 30 000 €.

FALLET Laurent
PERRIN Guillaume

3 – Département et Salaire

Tout d'abord voici un tableau récapitulatif des statistiques générales sur les 4 départements de l'INSA :

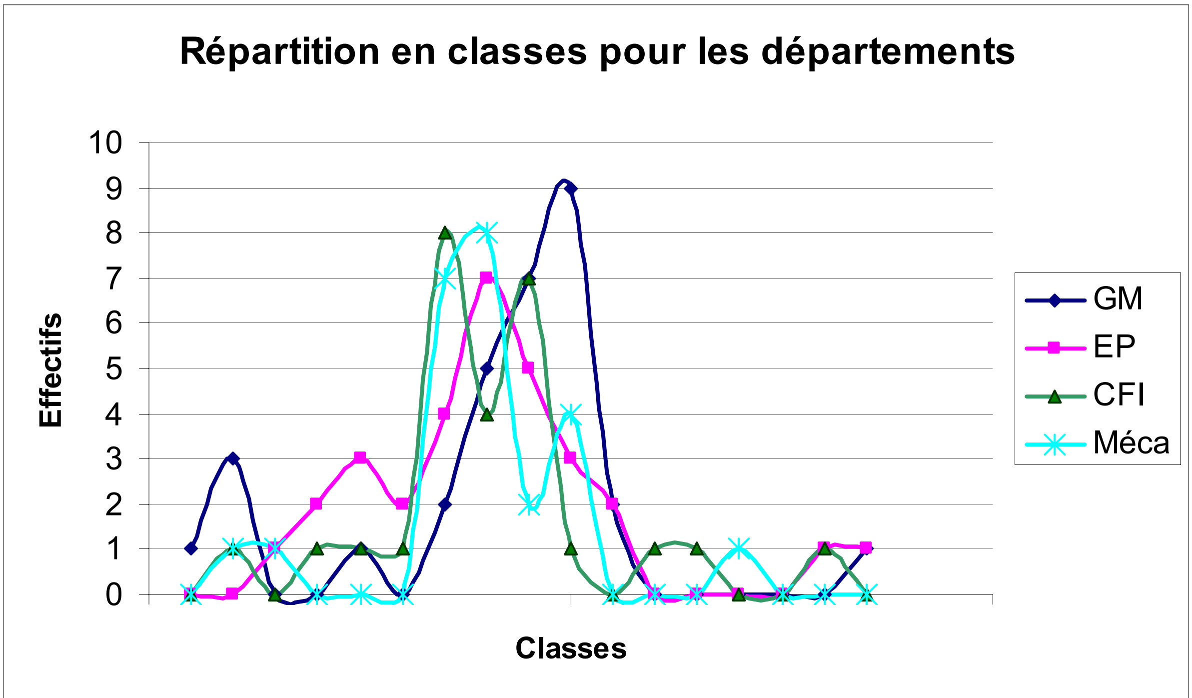
	GM	EP	CFI	Méca	Total	Global (salaires sans filtre)
Effectif	**31**	**31**	**27**	**24**	**113**	**118**
Moyenne	29 980,84 €	29 286,74 €	29 284,44 €	28 650,00 €		29 114,19 €
Médiane	30 500,00 €	29 000,00 €	29 500,00 €	29 000,00 €		29 050,00 €
Ecart-type	8 600,00 €	6 585,00 €	5 325,00 €	4 355,00 €		6 624,00 €



Sur ce graphe on observe que les étudiants de GM sont plus payés que ceux des autres départements, et que les Méca sont en dernière position.

La médiane confirme cette tendance.

Nous avons ensuite reclassé les séries de salaires de chaque département avec le même principe que celui page 9. On obtient au final un graphe qui permet de comparer les répartitions des 4 départements à la fois.



Les étudiants GM sont là encore les mieux payés même si une pointe dans les classes inférieures se dessine. Cela est du au fait que les élèves de ce département poursuivent assez souvent leurs études en 3ème cycle, qui est rémunéré à temps partiel.

D'après le graphique et les effectifs supérieurs à 30 pour les GM et EP nous considérerons que leurs salaires suivent une loi normale.

Etudions maintenant les probabilités pour un étudiant de chacun de ces départements de gagner plus de 30 000 € à la sortie de l'INSA :

$$p(X_{GM} > 30000) = p\left(S > \frac{30000 - 29980{,}8}{8600}\right) = 1 - p(S \leq 0{,}0022) = 1 - 0{,}501 = 0{,}499$$

Donc la probabilité pour un élève de GM de gagner plus de 30 000 € quand il débute est de **49,9 %.**

Notons que la médiane est de 30 500 €; cela signifie que 50% des ingénieurs sondés gagnent cette somme ou plus. Pourquoi trouvons-nous donc une probabilité aussi basse? La raison est que la médiane considère les élèves ingénieurs encore en études (3ème cycle, rémunérés 20 000 € ou moins) au même régime que d'autres gagnant entre 24 000 et 28 000 €. La moyenne, elle, prend en compte cette différence significative, qui se répercute sur notre approximation de loi normale.

Pour les EP:

$$p(X_{EP} > 30000) = p\left(S > \frac{30000 - 29286{,}7}{6585}\right) = 1 - p(S \leq 0{,}108) = 1 - 0{,}544 = 0{,}456$$

La probabilité pour un ancien élève d'EP de gagner plus de 30 000 € quand il débute est de **45,6 %.**

Pour les chimistes (CFI), le nombre de données récoltées ne permet d'approximer que par une loi de Student. Nombre de degrés de liberté : 27-1=26.

$$p(X_{CFI} > 30000) = p\left(S > \frac{30000 - 29284{,}7}{5325}\right) = 1 - p(S \leq 0{,}134)$$

La valeur 0,134 n'étant pas dans notre appendice de loi de Student, nous allons interpoler cette valeur (nous vous épargnons les calculs).

$$p(X_{CFI} > 30000) = 1 - p(S \leq 0{,}134) = 1 - 0{,}553 = 0{,}447$$

Cette fois la probabilité prend la valeur **44,7 %**.

En procédant de même pour nos camarades mécaniciens (avec une loi de Student également), on obtient dans **38 %** des cas un salaire supérieur à 30 000 € par an.

Le hit-parade des départements est : GM, EP, CFI puis Méca. Cet ordre est le même que celui observé avec les moyennes.

Une remarque intéressante: la moyenne faite à partir des moyennes des départements vaut 29 300,51 €. Cette valeur est supérieure à celle calculée directement à partir des réponses récoltées (qui vaut 29 114,19 €).

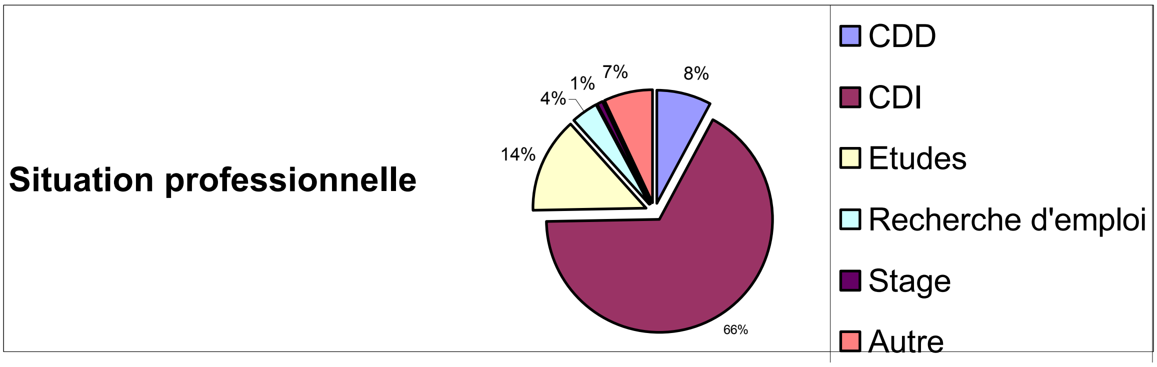
4 – Situation Professionnelle et Salaire

a) Présentation des résultats

Nous avons proposé différentes rubriques pour limiter le choix dans la rubrique "situation professionnelle". Cependant, vu la diversité des statuts des jeunes diplômés, nous avons donc un pourcentage assez important de réponses "autre".

Il faut préciser qu'en général, les personnes nous ont ajouté en commentaire qu'elles travaillaient en volontariat ou en coopération à l'étranger.

La répartition des différentes situations professionnelles existantes est la suivante :



Au vu de ce graphique, il est clair que les personnes en recherche d'emploi sont en minorité ce qui est assez rassurant pour les promotion à venir ! Cependant, il faudrait réaliser l'étude sur un échantillon plus important pour tirer des conclusions vraiment valables. (Notre échantillon est de 130 personnes).

Nous avons la répartition suivante pour les salaires selon la situation professionnelle :

b) Comparaison CDI - Poursuite d'études

On utilise la méthode de comparaison de deux moyennes d'échantillons pour des échantillons de taille inférieure à 30 :

$n_1 = 13$ le nombre de personnes en Etudes.
$n_2 = 87$ le nombre de personnes en CDI.
$m_1 = 23078{,}31 \quad \sigma_1 = 7309{,}44$
$m_2 = 30388{,}01 \quad \sigma_2 = 6024{,}05$

On considère que les variables moyennes d'échantillons suivent des lois de Student :

Alors :

$$Z = \frac{\left|\tilde{X}_1 - \tilde{X}_2\right|}{\sqrt{\frac{\sigma_1^2}{n_1 - 1} + \frac{\sigma_2^2}{n_2 - 1}}} \quad suit \quad T_{n_1+n_2-1}$$ et $n_1 + n_2 - 1 = 98$ > 30 donc $Z\ suit\ N(0\,;1)$.

$\sigma_D = 2207{,}78$ On choisit *u* = **1, 96** et on obtient $\alpha = 4327{,}25 \quad |d| = 7309{,}7 \geq \alpha$ donc on réfute l'hypothèse de l'égalité des moyennes de salaires pour les personnes en poursuite d'étude et celles en CDI. Ce faisant, on est sûr à 95% que les moyennes sont différentes.

On peut conclure que les étudiants en 3ème cycle ne sont pas payés comme des ingénieurs en CDI ce qui paraît normal ! On peut cependant préciser que les personnes faisant un 3ème cycle à l'étranger et plus particulièrement en Allemagne sont assimilés ingénieurs.

Etudions maintenant les probabilités, pour un Etudiant en 3ème cycle, puis pour une personne en CDI de gagner plus de 30 000 € à la sortie de l'école :

X_1 la variable aléatoire du salaire des Etudiants en 3ème cycle, X_2 celle du salaire des personnes en CDI.

$$p(X_1 > 30000) = p\left(S_1 > \frac{30000 - 23078{,}31}{7309{,}44}\right) = 1 - p(S_1 \leq 0{,}947) = 1 - 0{,}79 = 0{,}21$$

La probabilité d'être payé plus de 30 000 € pour un Etudiant en 3ème cycle à la sortie de l'INSA est de **21%**. Ce résultat n'est pas très représentatif du fait du faible nombre de réponses.

$$p(X_2 > 30000) = p\left(S_2 > \frac{|30000 - 30388{,}01|}{6024{,}05}\right) = 1 - p(S_2 \leq 0{,}065) = 1 - 0{,}5259 = 0{,}4741$$

La probabilité d'être payé plus de 30 000 € pour une personne en CDI à la sortie de l'INSA est de **47,5%.**

c) Comparaison CDI - CDD

On utilise la méthode de comparaison de deux moyennes d'échantillons pour des échantillons de taille inférieure à 30 :

$n_1 = 10$ le nombre de personnes en CDD.
$n_2 = 87$ le nombre de personnes en CDI.
$m_1 = 27800 \quad \sigma_1 = 6581{,}79$
$m_2 = 30388{,}01 \quad \sigma_2 = 6024{,}05$

On considère que les variables moyennes d'échantillons suivent des lois de Student :

Alors :

$$Z = \frac{\left|\widetilde{X}_1 - \widetilde{X}_2\right|}{\sqrt{\frac{\sigma_1^2}{n_1 - 1} + \frac{\sigma_2^2}{n_2 - 1}}} \quad suit \quad T_{n_1+n_{2-1}}$$ et $n_1 + n_2 - 1 = 96$ > 30 donc $Z\ suit\ N(0\,;1)$.

$\sigma_D = 2288{,}08$ On choisit *u* = **1, 96** et on obtient $\alpha = 4484{,}63 \quad |d| = 2588{,}01 \le \alpha$ donc on accepte l'hypothèse de l'égalité des moyennes de salaires pour les personnes en CDD et celles en CDI. On ne peut pas préciser à 95% que ces deux moyennes sont différentes.

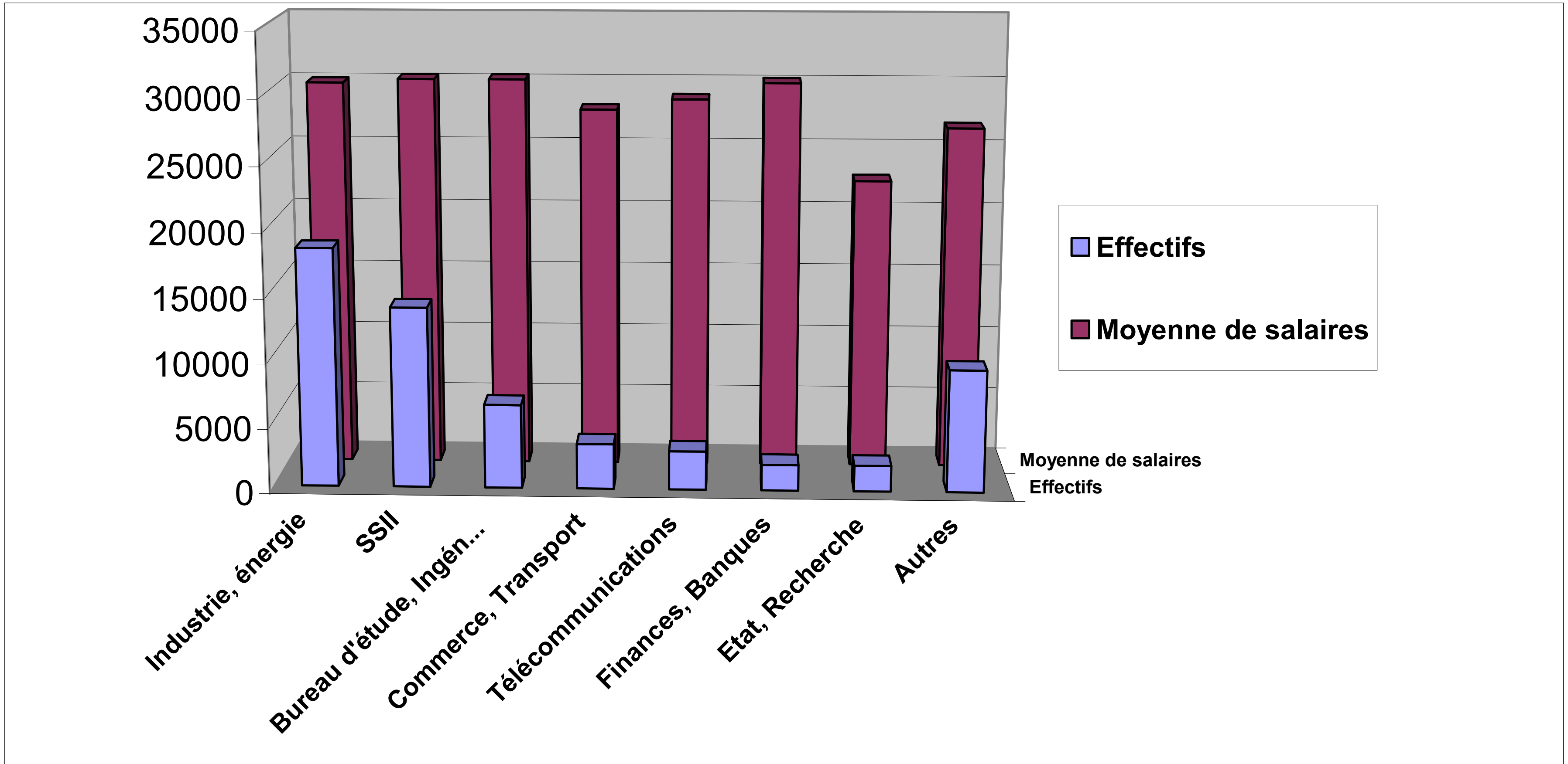
Etudions maintenant la probabilité pour une personne en CDD de gagner plus de 30 000 € à la sortie de l'école :

X_1 la variable aléatoire du salaire des personnes en CDD. On utilise la loi de Student pour la valeur de $p(S_1 \le 0{,}335)$

$$p(X_1 > 30000) = p\left(S_1 > \frac{30000 - 27800}{6581{,}79}\right) = 1 - p(S_1 \le 0{,}335) = 1 - 0{,}626 = 0{,}374$$

La probabilité d'être payé plus de 30 000 € pour une personne en CDD à la sortie de l'INSA est de **37,4%**.

5 – Secteur et Salaire



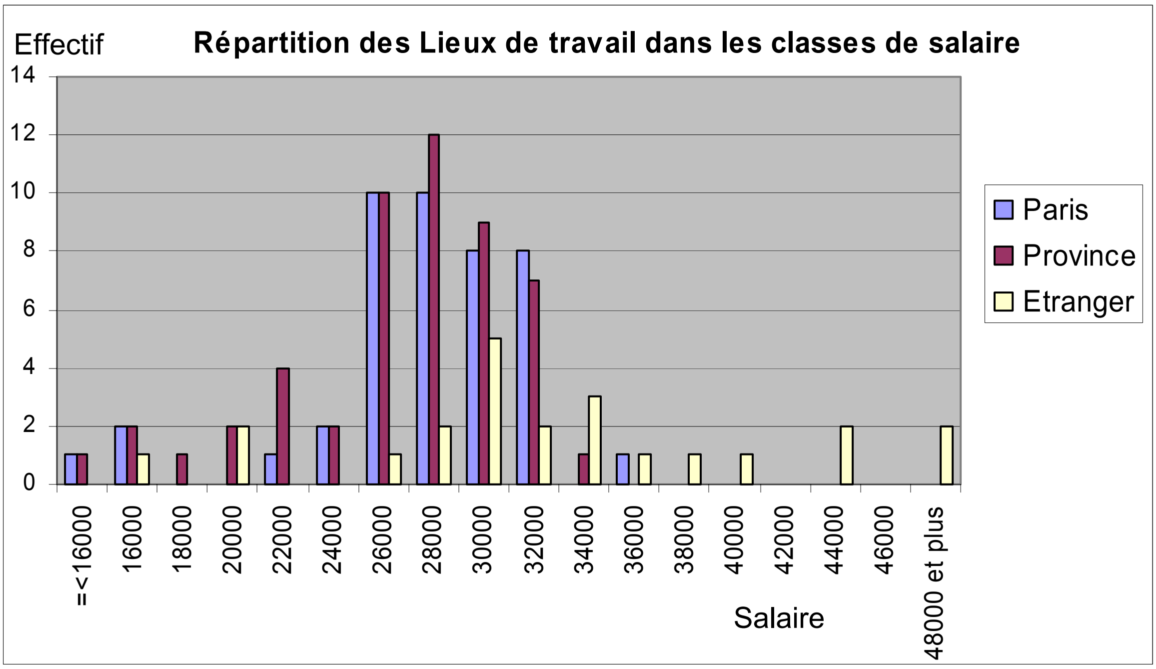
Nous observons que les secteurs les moins rémunérés sont dans l'état, la recherche et le commerce. A l'inverse, les secteurs les mieux payés sont les SSII et les bureaux d'études. Les barres bleues permettent de juger la confiance à accorder aux résultats dans chaque secteur. Les probabilités d'avoir un salaire supérieur à 30 000 € dans les secteurs: Industrie, SSII, Bureau d'étude, Finances, sont respectivement les suivants: 50%, 48,8%, 48,8%, 48,8% (ce qui est très similaire).

Nous n'effectuerons pas de tests ici car les échantillons ne sont pas représentatifs.

6 – Lieu de travail, Taille de l'entreprise et Salaire

a) Lieu de travail et salaire

Présentation des données :



Nous remarquons que les diplômés travaillant à l'étranger paraissent un peu mieux payés.

Cette tendance semble se confirmer lorsque nous regardons les données de base pour chaque catégorie :

	Paris	Province	Etranger	Total
Effectif	**43**	**51**	**23**	**117**
Moyenne	28 386,23 €	27 617,22 €	34 234,30 €	
Médiane	29 000,00 €	29 000,00 €	32 000,00 €	
Ecart-type	4 576,36 €	4 581,83 €	10 360,31 €	

Cependant, nous avons un effectif de personnes à l'étranger un peu faible.

Utilisons maintenant les comparaisons de moyennes pour déterminer si la différence entre les moyennes est vraiment représentative.

On pose :

$\widetilde{X}_1$ *et* $\widetilde{X}_2$ les variables moyenne des échantillons "Paris et "Province"

Nous considérons que les variables moyennes d'échantillons suivent des lois normales :

Alors :

$$D = \left|\widetilde{X}_1 - \widetilde{X}_2\right| \quad suit \quad N(m_1 - m_2 \; ; \sigma_D)$$

$\sigma_D = 958{,}39$ on choisit *u* = **1, 96** et on obtient $\alpha = 1878 \quad |d| = 769{,}01 \leq \alpha$ donc pour un seuil de confiance à 95% la moyenne des salaires des personnes travaillant à Paris n'est pas différente de la moyenne des salaires des personnes en Province.

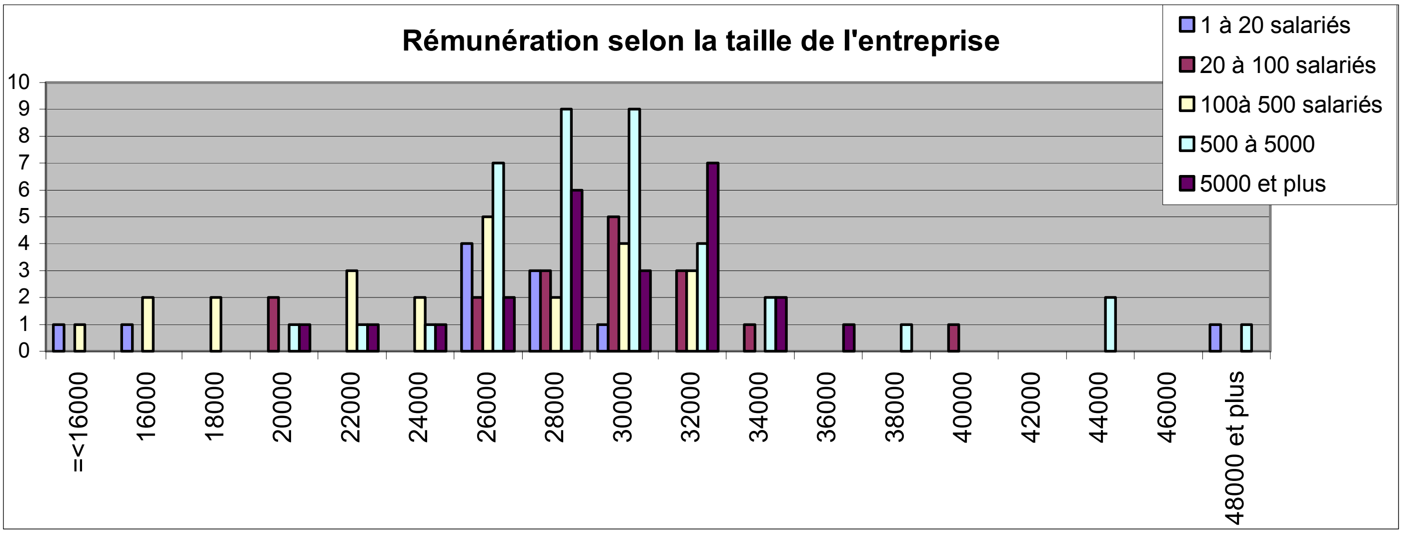
Même si à priori, on semble gagner moins en province qu'à Paris, cette tendance est à la baisse et la différence sur les moyenne n'est plus flagrante.

Pour la comparaison entre la province et l'étranger, on a pour un seuil à 95% :

$$\alpha = 4511{,}7 \quad |d| = 3848{,}07 \leq \alpha$$

Cependant, ces valeurs étant proches, on réalise une comparaison avec un seuil à 88%, alors $\alpha = 3581{,}77 \quad |d| = 3848{,}07 \geq \alpha$ donc on est sûr à 88% que la moyenne des salaires des personnes en province est différente de la moyenne des salaires des personnes à l'étranger.

b) Taille de l'entreprise et salaire



La tendance générale révèle que les entreprises de plus de 5000 salariés proposent une rémunération plus importante que les entreprises de taille moindre.

Il semble que plus l'entreprise est grande, plus le salaire à l'embauche est important. Cependant, il faudrait réaliser une étude après 3 ou 4 ans d'expérience pour savoir si oui ou non un écart se creuse au niveau des salaires entre les différentes tailles d'entreprises.

Données de base pour les différentes tailles d'entreprise :

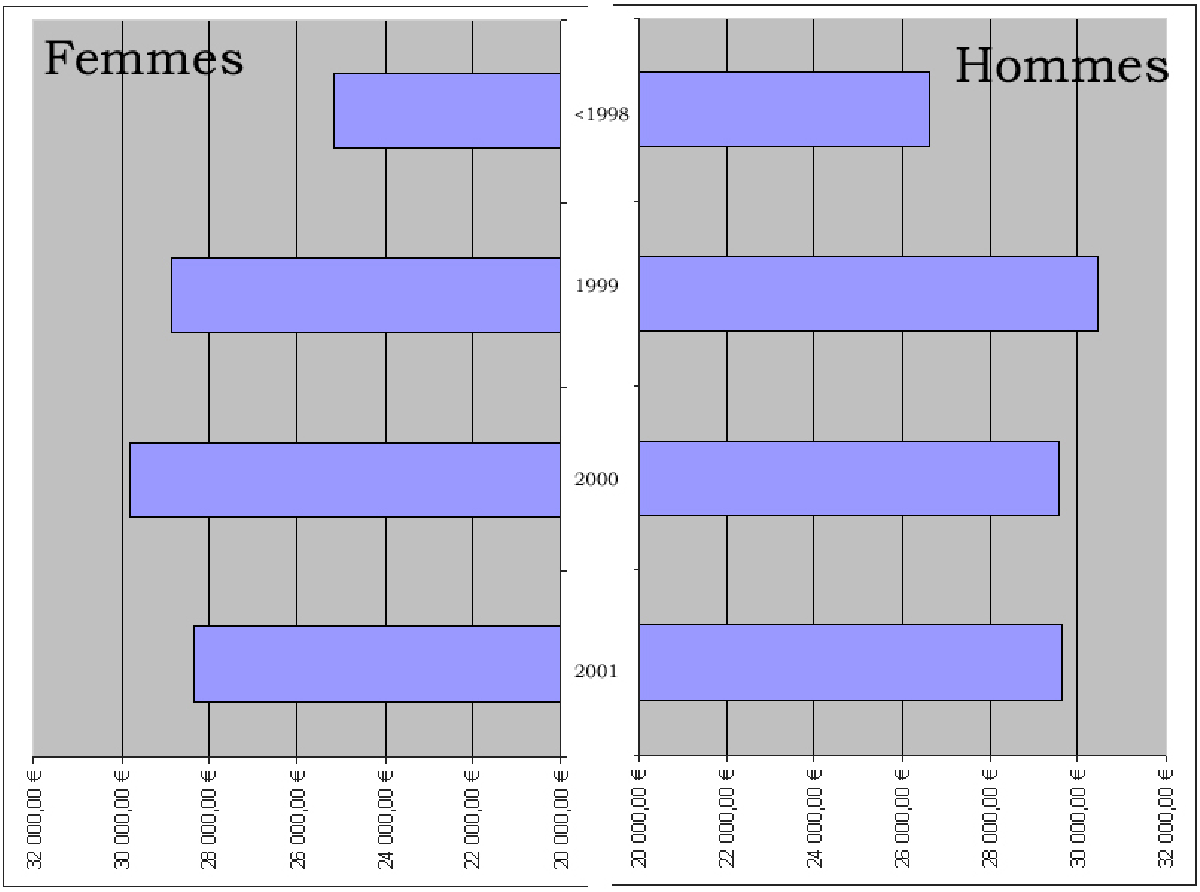
	1 à 20 salariés	20 à 100 salariés	100 à 500 salariés	500 à 5000 salariés	Plus de 5000 salariés	total
Effectif	**11**	**17**	**24**	**38**	**24**	**114**
Moyenne	29 000 €	29 989 €	25 700 €	30 945 €	30 162 €	
Médiane	27 400 €	30 500 €	27 400 €	29 940 €	30 000 €	
Ecart-type	13 240 €	4 418 €	5 507 €	5 836 €	3 850 €	

En réalisant une comparaison de moyennes pour les différentes catégories de tailles on remarque qu'à 95%, la moyenne des salaires dans les entreprises de 100 à 500 salariés est moins élevée que dans les entreprises de 500 à 5000 salariés, de plus de 5000, et de 20 à 100 salariés.

IV - Comparaisons diverses

1 – Différence de salaire Homme – Femme

D'après le département de l'Emploi et des Revenus d'activité de l'Insee (n°801 août 2001 d'Insee Première) l'écart entre les salaires de hommes et des femmes s'élève à 9,4% en 1995 sur toutes les catégories de salaires (incluant les cadres).



Voici un abaque des moyennes des salaires H / F des années 2001 à 1998.

Notre étude révèle une différence sur la moyenne des salaires homme / femme de 1,1% au bénéfice des hommes.

Deux interprétations sont possibles :

- Les salaires des ingénieurs sont moins touchés par la ségrégation homme / femme.
- L'évolution durant ces dernières années a rétabli l'équité entre les deux sexes au niveau des salaires.

2 - Valeurs extraites du supplément "Jeunes diplômés" de l'Expansion de mai 2002

L'expansion exprime la rémunération annuelle à l'embauche par la médiane des données collectées. Cette médiane est comprise dans une fourchette. Nous considérerons que le centre de cette fourchette représente la moyenne (approximation nécessaire pour les calculs). Aucune information sur la variance des informations. Elles sont, je cite, « représentatives à $\pm 5\%$ *du marché actuel ».*

Pour l'INSA de Rouen, la médiane est comprise entre 32000 et 32800 €. Nous en déduisons que :

$\hat{m}_{Rouen} = 32400$ €

Pour l'INSA de Lyon, la médiane est comprise entre 34500 et 35800 €. Nous en déduisons que :

$\hat{m}_{Lyon} = 35150$ €

Donc, l'écart observé entre les deux écoles s'élève à :

$$\frac{|\hat{m}_{Rouen} - \hat{m}_{Lyon}|}{\hat{m}_{Lyon}} = \frac{|32400 - 35150|}{35150} = 7{,}8\%$$

Ces 7,8% sont, évidemment, au bénéfice de l'INSA de Lyon.

3 - Valeurs extraites de l'INSEE Première n°812 de Novembre 2001:

Grâce à ce document qui donne une approximation du salaire d'un ingénieur, en fonction de son école, de ses caractéristiques individuelles (expérience, profession du père, profil sociodémographique), des caractéristiques de l'entreprise (taille de l'entreprise, lieu d'emploi, secteur d'activité), nous avons estimé le salaire d'un Insaïen lyonnais.

Nous avons trouvé 32 840 € environ (l'erreur peut être estimée à 2 % ou moins).

Dans le même document, un calcul du même type pour le profil d'un Insaïen rouennais donne 30 869 €. Il y a en effet 6 % d'écart, au bénéfice de l'INSA de Lyon.

Notre valeur moyenne est 29 114 €, ce qui représente encore 6 % de décalage, la valeur supérieure étant celle de l'INSEE.

La question est la suivante: pourquoi les données extérieures sont-elles supérieures? Une des raisons pourrait être que leurs valeurs n'englobent que les ingénieurs en CDI. Dans ce cas, on la compare à notre valeur (filtre: uniquement les CDI) qui est: 30 388 €.

L'écart est réduit, mais il persiste!

Conclusion

Cette étude sur les salaires se voulait essentiellement descriptive et empirique. En effet, nous constatons que les études professionnelles (par exemple menées par des instituts de sondages) portent sur des échantillons d'au moins 1 000 personnes pour un caractère donné et jusqu'à 20 000 personnes pour un phénomène à causes multiples.

L'analyse des salaires se trouve dans cette dernière catégorie. Dégager l'influence de chaque caractère indépendamment des autres est difficile sur un petit échantillon tel que le notre.

Nous avons, au delà des formules mathématiques, compris que la modélisation efficace d'un phénomène passe d'abord par un sondage exploitable sur des échantillons représentatifs de la population à étudier.

De plus, il est préférable de garder une certaine réserve quant aux valeurs des études statistiques sur les salaires, souvent revues à la hausse.

********************* *Delirium* *********************

En moyenne, chaque personne possède un testicule.

Un enfant sur sept étant Chinois, nous nous sommes arrêté à six.

Le loto, c'est un impôt sur les gens qui ne comprennent pas les statistiques.

Le statisticien est un homme qui fait un calcul juste en partant de prémices douteuses pour aboutir à un résultat faux.

Jean Delacour

A la question : faites vous encore confiance aux instituts de sondage ? 64% des Français répondent non. Et 59% répondent oui.

Philippe Geluck

Dans toute statistique, l'inexactitude du nombre est compensée par la précision des décimales.

Alfred Sauvy

Les chiffres sont aux analystes ce que les lampadaires sont aux ivrognes : ils fournissent bien plus un appui qu'un éclairage.

Jean Dion

****