

## Synthèse Statistiques

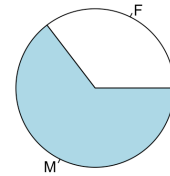
Zou Yongkang  
& Hassan Yusuf Zakaria

Répartition des medecins par leurs sexes

### ## a) Variables qualitatives

# pour **prasex** :

	F	M
distribution en effectifs	177	323
distribution en frequences	0.354	0.646

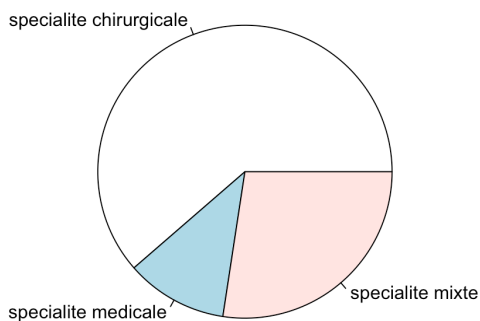


# les hommes sont majoritaires

# pour **prasep3** :

	specialite chirurgicale	specialite medicale	specialite mixte
distribution en effectifs	307	56	137
distribution en frequences	0.614	0.112	0.274

Répartition des medecins par leurs specialites



# Les spécialités chirurgicales sont les plus nombreuses : elles représentent 61.4% % de tous les médecins.

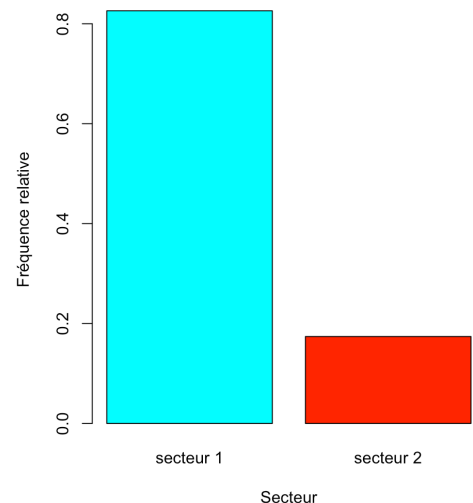
# pour **secteur** :

	secteur 1	secteur 2
distribution en effectifs	413	87
distribution en frequences	0.826	0.174

# les secteurs 1 sont les plus nombreux

# et les secteurs 2 sont très peu nombreux : environ 17 % des médecins.

Répartition des medecins par leurs secteurs

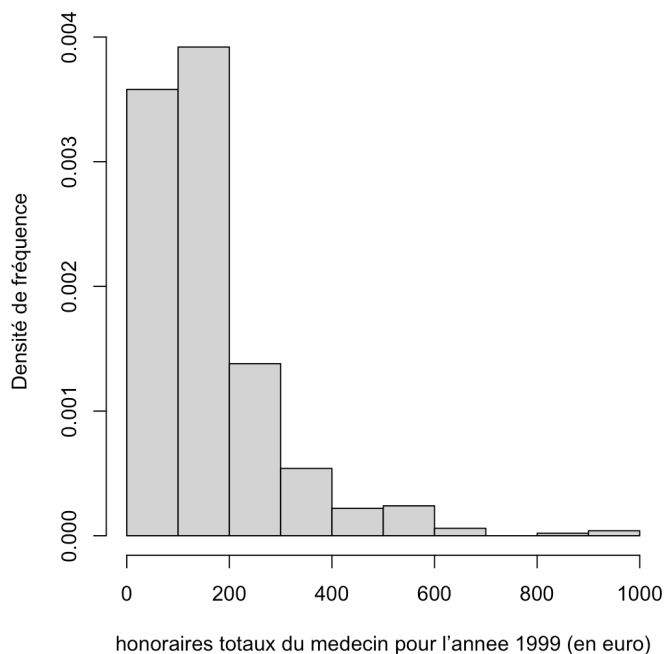


## ## b) Variables quantitatives

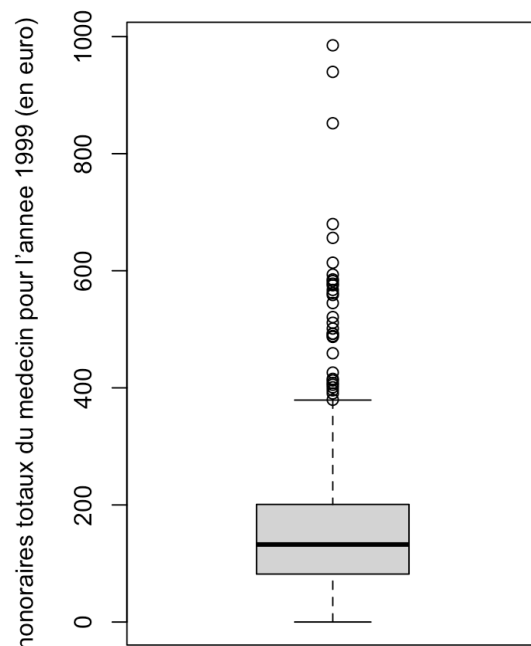
# Pour **honormkf** :

	min	Q1	mediane	moyenne	Q3	maximum	variance	ecart-type	coefficient de variance
honormkf	0.056	81.922	132.417	165.635	200.444	985.082	17051.05	130.580	0.788

**Histogramme des honoraires**



**Boîte à moustaches du honormkf**



# L'histogramme des honoraires nous montre une distribution très asymétrique avec une surreprésentation

# des bas honoraires et des valeurs extrêmes correspondant à des hauts honoraires, # ce que nous confirme la boîte à moustaches. Plusieurs médecins pratiquent des honoraires très élevés.

# l'honoraire moyen que pratique un médecin de notre échantillon est de 165,65 euros

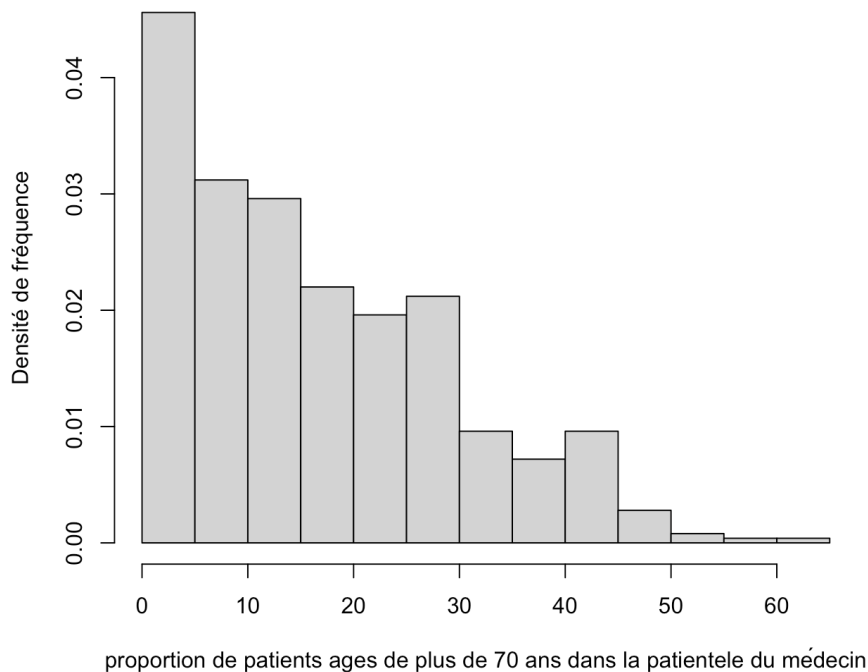
# et le médian est de 132,41 euros.

# Le coefficient de variation, égal à 0.78, révèle une forte dispersion des honormkf.

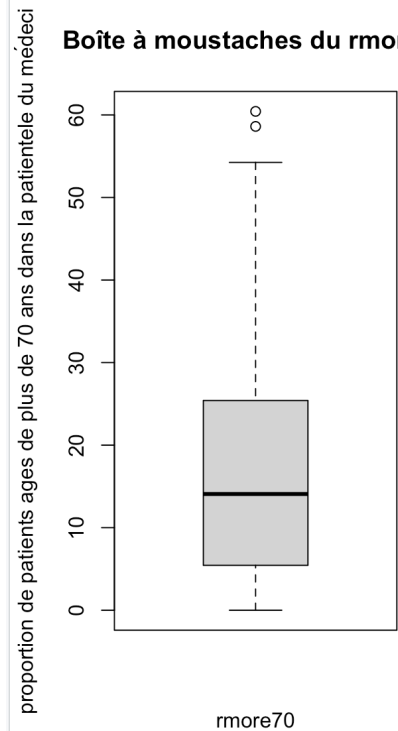
# Pour **rmore70** :

	min	Q1	mediane	moyenne	Q3	maximum	variance	ecart-type	coefficient de variance
rmore70	0.00	5.46	14.08	16.66	25.41	60.43	162.66	12.75	0.766

**Histogramme des patients de plus de 70 ans**



**Boîte à moustaches du rmore70**



- # L'histogramme des patients de plus de 70 ans nous montre une distribution très asymétrique avec une surreprésentation.
- # On a très peu de médecins qui ont une grande proportion de patients de plus de 70 ans.
- # En moyenne, dans notre échantillon, les médecins ont 16.66 % de patients de plus de 70 ans
- # et le médian est de 14.08 %
- # Le coefficient de variation, égal à 0.76, révèle une forte dispersion des rmore70.

### **## Liaison prasex/praspe3**

i) Tableau de contingence du couple (prasex, praspe3) (\*addmarge\*)

prasex	specialite chirurgicale	specialite medicale	specialite mixte	Sum
F	106	3	68	177
M	201	53	69	323
Sum	307	56	137	500

ii) Tableau profils-lignes

prasex	specialite chirurgicale	specialite medicale	specialite mixte
F	0.599	0.017	0.384
M	0.622	0.164	0.214

- # 59,9% des femmes sont specialites chirurgicales
- # 62,2% des hommes sont specialites chirurgicales

## Commentaires :

# On constate que quelque soit le sexe, la specialite medicale est la moins representée

# et la spécialité chirurgicale est la plus représentée.

## Test du Khi-deux d'indépendance

prasex	specialite chirurgicale	specialite medicale	specialite mixte
F	108.68	19.82	48.5
M	198.32	36.18	88.5

Pearson's Chi-squared test

X-squared	df	p-value
34.344	2	3.486e-08

Contributions au Khi2

prasex	specialite chirurgicale	specialite medicale	specialite mixte
F	0.07	14.28	7.84
M	0.04	7.82	4.30

## Commentaire :

# 3 forte contributions : (F,specialite medicale), (F, specialite mixte) et (M, specialite medicale)

# On constate que la majorité des femmes ont pour spécialité la chirurgie médicale comme leurs confrères

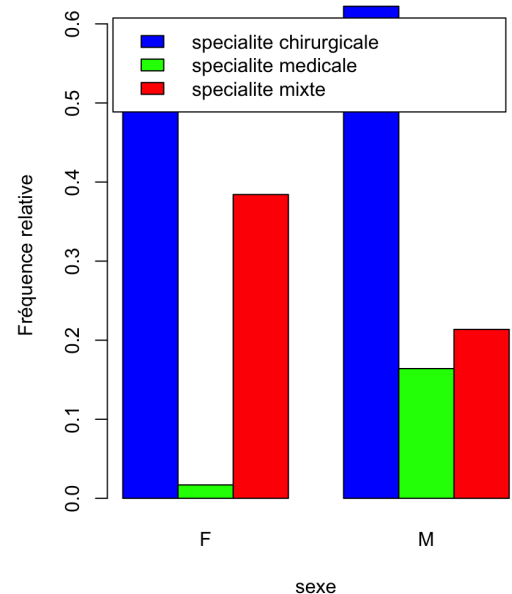
# cependant on voit que l'écart d'effectif qu'il y a entre les femmes qui ont pour spécialité <médicale> et mixte

# est très élevé comparé à ses homologues hommes.

## Conclusion :

# Le choix de la specialite par le medecin est lie a son sexe.

Distributions conditionnelles du specialites sachant le sexe

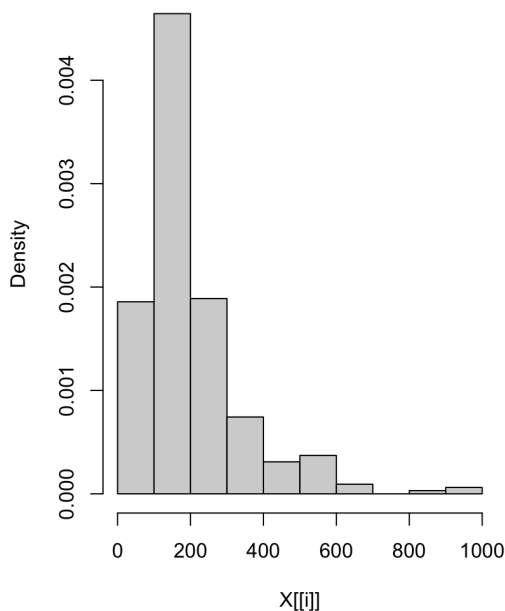


## ## Liason honormkf/prasex

	F	M
Moyennes par groupe	97.37	203.04
Variances par group	4344.39	20084.23

# Commentaires pour les deux parties (i et ii)  
 # On constate que la moyenne des honoraires des hommes est beaucoup plus grande que celle des femmes.  
 # c'est vrai aussi pour les médianes et les dispersions (visible sur les variances mais aussi sur  
 # les intervalles interquartiles des boîtes à moustaches). On voit en particulier que la médiane  
 # des honoraires des femmes est inférieure au 1er quartile des honoraires des hommes, ce qui signifie que

Histogram of X[[i]]



# la moitié des femmes ont des honoraires qui n'atteignent pas les honoraires que touchent

# 75 % des hommes. Les deux distributions présentent des valeurs extrêmes (de très hauts honoraires)

### ## test statistiques

# F test to compare two variances (data: honormkf by prasex)

F = 0.21631, num df = 176, denom df = 322,

p-value < 2.2e-16

alternative hypothesis:

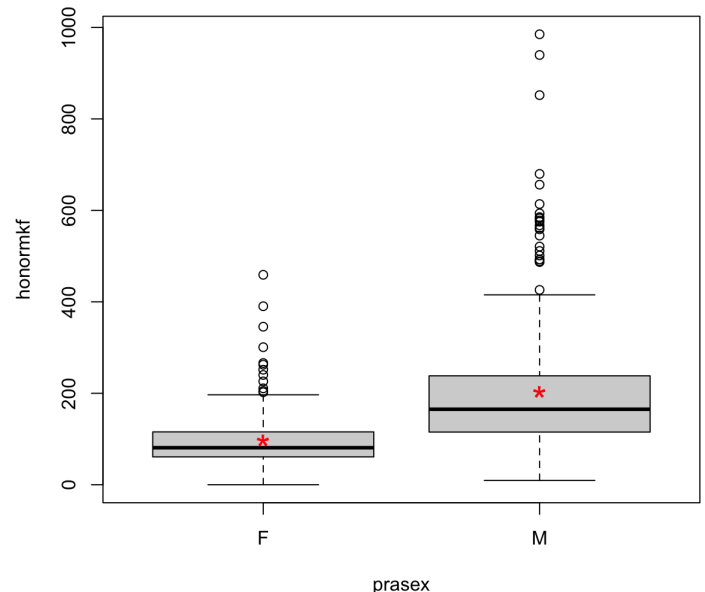
true ratio of variances is not equal to 1

95 percent confidence interval : [0.1675393, 0.2821020]

sample estimates:

ratio of variances : 0.2163085

Boîtes à moustaches juxtaposées honormkf en fonction du prasex



# Welch Two Sample t-test (data: honormkf by prasex)

t = -11.347, df = 487.43, p-value < 2.2e-16

alternative hypothesis:

true difference in means between group F and group M is not equal to 0

95 percent confidence interval: [-123.96704, -87.37138]

sample estimates:

mean in group F : 97.37281  
mean in group M : 203.04203

# on conclut que les honoraires des hommes sont en moyenne  
# significativement plus élevés que ceux des femmes  
# (203.04 euros pour les médecins hommes versus 97.37 euros pour leurs homologues femmes)

## Conclusion : Les médecins spécialistes masculins gagnent mieux leur vie que leurs consœurs.

### ## Liaison secteur/praspe3

i) Tableau de contingence du couple (secteur, praspe3)

secteur	specialite chirurgicale	specialite medicale	specialite mixte
secteur 1	289	18	106
secteur 2	18	38	31

ii) Tableau des profils-colonnes

secteur	specialite chirurgicale	specialite medicale	specialite mixte
secteur 1	0.941	0.321	0.774
secteur 2	0.059	0.679	0.226

# 94,1% des spécialités chirurgicales sont dans le secteur 1

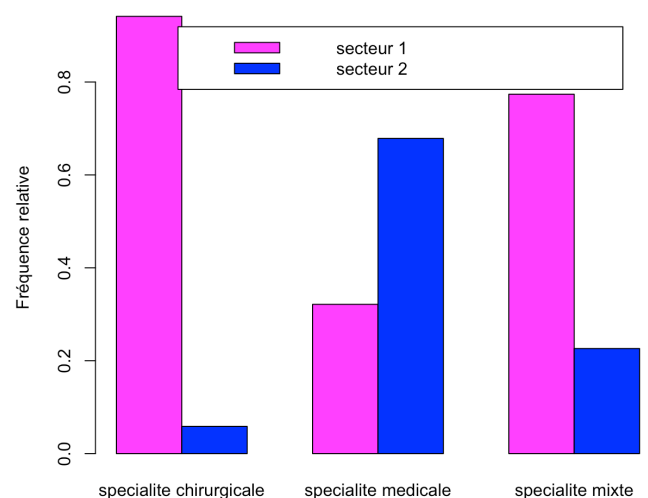
# 67,9% des spécialités médicales sont dans le secteur 2

# 77,4% des spécialités mixtes dans le secteur 1

## Commentaire:

# On constate que pour la spécialité chirurgicale et la spécialité mixte, le secteur 1 est majoritaire,  
# et pour la spécialité médicale, le secteur 2 est majoritaire.

Distributions conditionnelles du secteur sachant le specialite



iii) Test du Khi-2 d'indépendance  
conditions de validite

	specialite chirurgicale	specialite medicale	specialite mixte
secteur 1	253.58	46.26	113.16
secteur 2	53.42	9.74	23.84

Pearson's Chi-squared test (data: tableauSP)  
X-squared = 130.23, df = 2, p-value < 2.2e-16

Contributions au Khi

	specialite chirurgicale	specialite medicale	specialite mixte
secteur 1	4.95	17.26	0.45
secteur 2	23.48	81.94	2.15

3 fortes contributions : (secteur 2, spécialité médicale), (Secteur 2, spécialité chirurgicale) et (secteur 1, spécialité médicale)

## Commentaire :

# On constate que les médecins ayant pour spécialité la spécialité médicale et chirurgicale et faisant partie du secteur 2 contribuent beaucoup au khi 2 et dans une moindre mesure,  
# Les médecins du secteur 1 ayant partie pour spécialité médicale contribue aussi au khi2.

## Conclusion :

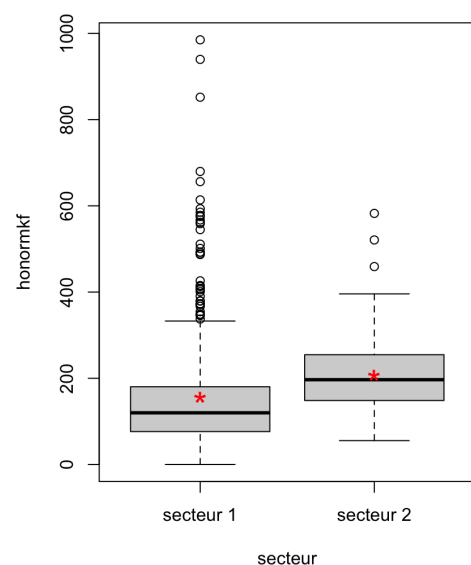
# La répartition par secteur n'est pas la même dans les trois types de spécialités.

### ## Liaison secteur/honoraire

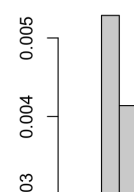
	secteur 1	secteur 2
moyennes par groupe	156.8466	207.3553
variances par groupe	18345.079	8918.298

# Commentaires pour les deux questions (i et ii)  
# On constate que la moyenne des honoraires dans le secteur 2 est plus grande que celle dans le secteur 1.  
# c'est vrai aussi pour les médianes et les dispersions.  
# On voit en particulier que la médiane des honoraires dans le secteur 1 est inférieure au 1er quartile des honoraires dans le secteur 2,  
# ce qui signifie que la moitié du secteur 1 ont des honoraires qui n'atteignent pas les honoraires que touchent

à moustaches juxtaposées honormkf en fonction



Histogram of X[[i]]



# 75 % du secteur 2. La distribution pour le secteur 1 présente des valeurs extrêmes (de très hauts honoraires)

### ## Test statistique

F test to compare two variances (data: honormkf by secteur)

F = 2.057, num df = 412, denom df = 86, p-value = 8.6e-05

alternative hypothesis:

true ratio of variances is not equal to 1

95 percent confidence interval: [1.451908, 2.811056]

sample estimates:

ratio of variances : 2.057016

Welch Two Sample t-test (data: honormkf by secteur)

t = -4.1669, df = 170.01, p-value = 4.905e-05

alternative hypothesis: true difference in means between group secteur 1 and group secteur 2 is not equal to 0

95 percent confidence interval: [-74.43650, -26.58085]

sample estimates:

mean in group secteur 1 : 156.8466

mean in group secteur 2 : 207.3553

# on conclut que les honoraires dans le secteur 2 sont en moyenne

# significativement plus élevés que ceux dans le secteur 1

# (207,35 euros pour les medecins dans le secteur 2 versus 156,84 euros pour leurs homologues dans le secteur 1)

### ## Conclusion :

# Le secteur tarifaire auquel le médecin appartient influe sur les honoraires du médecin :

# Les médecins dans le secteur 2 sont plus susceptibles de gagner mieux que leurs homologues du secteur 1.

### ## liaison rmore70/honoraires

# Nuage de points :

# le nuage semble réparti autour d'une droite très faiblement croissante, cependant le nuage de point semble dispersé.

# coefficient de corrélation linéaire

empirique : 0.3 (plutôt faible)

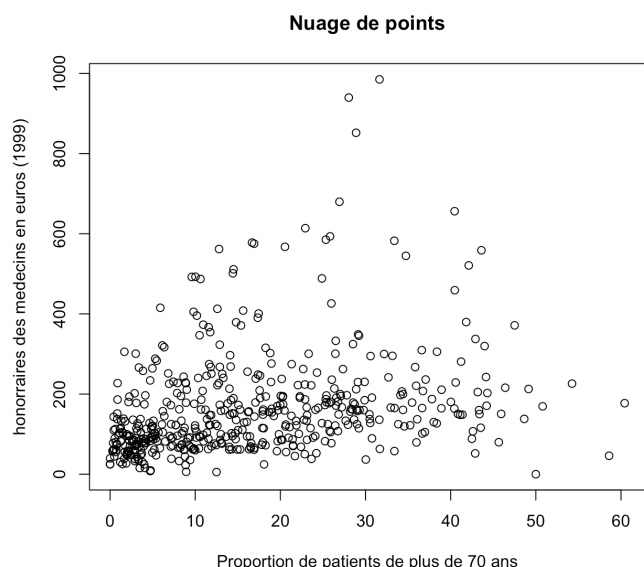
# Test statistique

Pearson's product-moment correlation

data: rmore70 and honormkf

t = 7.1297, df = 498, p-value = 3.557e-12

alternative hypothesis: true correlation is not equal to 0





95 percent confidence interval: [0.2225850, 0.3818358]

sample estimates:

cor : 0.3043354

#Conclusion

# La p-valeur ( $3.557e-12$ ) du test est inférieure au niveau de 5 % donc on rejette  $H_0$ .

# On conclut donc qu'il y a une liaison linéaire positive (car le coefficient

# de corrélation linéaire empirique de Pearson est  $0.304 > 0$ )

# entre la proportion de patients de plus de 70 ans et les honoraires des médecins.

# Donc, sans surprise, plus on a les patients de plus de 70 ans, meilleurs sont les honoraires des médecins.

### ## liaison spécialité/honoraires

	specialite chirurgicale	specialite medicale	specialite mixte
moyennes par groupe	167.5875	209.1194	143.4854
variances par groupe	22232.936	6220.467	8741.159

# Commentaires :

# On voit trois boîtes à moustaches très différentes. La médiane des spécialités médicale est bien

# au-dessus de celles des deux autres spécialités et cela est aussi vrai pour les moyennes.

# Les médecins de la spécialité médicale et les médecins de la spécialité mixte ont des honoraires médians et moyens assez similaires (6220 et 8741 euros annuels respectivement).

# La dispersion des honoraires des spécialités chirurgicales est aussi beaucoup plus grande.

# La dispersion des honoraires des spécialités médicales et des spécialités mixtes est très faible.

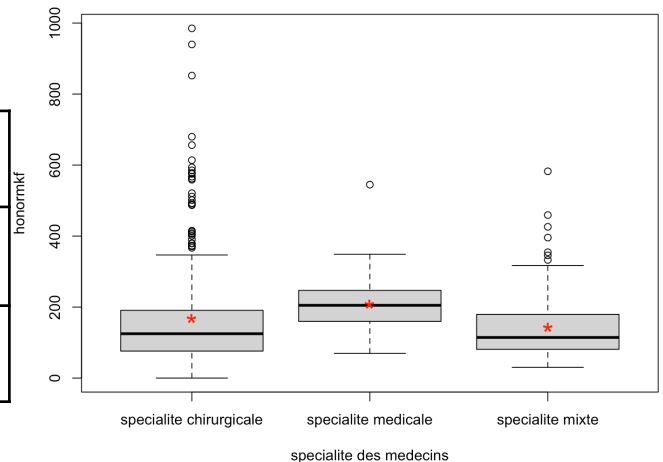
# Test statistique

Conditions de validite

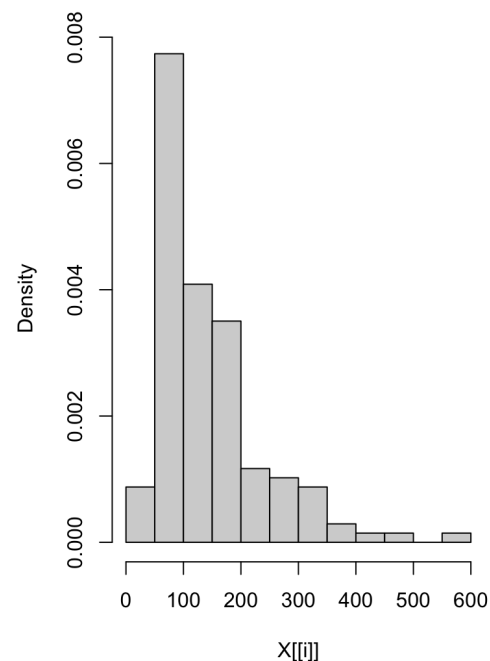
specialite chirurgicale	specialite medicale	specialite mixte
307	56	137

# on n'a donc pas la normalité pour les trois groupes d'après les histogrammes

Boîtes à moustaches juxtaposées du honoraires en fonction du specialite



Histogram of X[[i]]



# mais les effectifs sont élevés  
 # on peut donc considérer qu'on peut approcher les moyennes empiriques par groupe par  
 # des lois gaussiennes

# 2 tests possibles pour comparer les moyennes : le test classique de l'analyse  
 # de variance ou sa version modifiée en cas de variances inégales  
 # pour choisir quel test faire :  
 # on teste l'homoscédasticité = homogénéité des variances par groupe  
 # test de Brown-Forsythe :  $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2$

Levene's Test for Homogeneity of Variance (center = median)			
	Df	F value	Pr(>F)
group	2	4.5369	0.01115

# on rejette l'égalité des variances (p-valeur est inférieure à 0.05)

# il faut donc utiliser la variante de l'analyse de la variance pour variances inégales  
 One-way analysis of means (not assuming equal variances)  
 (data: honormkf and praspe3)  
 F = 12.272, num df = 2.00, denom df = 182.68, p-value = 9.986e-06  
 # si on conclut à des moyennes différentes (= rejeter  $H_0$ ), alors on continue pour savoir  
 quelles  
 # moyennes diffèrent deux à deux (sinon, on a fini)  
 # donc là, on continue avec les tests de comparaison multiples  
 # Les tests vus en cours estiment un  $\sigma^2$  commun (car hypothèse d'homoscédasticité).  
 # c'est l'argument pool.sd=TRUE  
 # si variances inégales (c'est le cas ici), il vaut mieux prendre une version  
 # "non pooled" avec pool.sd=FALSE

# test LSD (pas d'ajustement du risque alpha)  
 Pairwise comparisons using t tests with non-pooled SD (data: honormkf and praspe3)

	specialite chirurgicale	specialite medicale
specialite medicale	0.0026	-
specialite mixte	0.0396	2.3e-06

P value adjustment method: none

# on conclut ici que les trois populations sont différentes deux à deux car toutes les  
 # p-valeurs sont inférieures à 0.05  
 # et en s'aidant des valeurs des moyennes empiriques, on peut conclure à l'ordre suivant :  
 #  $\mu_E < \mu_A < \mu_M$   
 # c'est-à-dire que les salaires des managers sont significativement plus élevés que ceux des  
 # agents de sécurité qui sont significativement plus élevés que ceux des employés de  
 bureau.