

ENSAE PARISTECH

La beauté et la couleur de peau en politique

PROJET D'ÉCONOMÉTRIE 2 PROPOSÉ PAR OPHÉLIE COUPERIER

Cursus ingénieur
2ème Année

Maxime Blin
Pierre Delanoue

Mai 2019

Table des Matières

1	Question 1	2
1.1	(i)	2
1.2	(ii)	2
2	Question 2	3
3	Question 3	3
4	Question 4	5
5	Question 5	11
5.1	(i)	11
5.2	(ii)	14
5.3	(iii)	16
5.4	(iv)	18
6	Question 6	25
6.1	(i)	25
6.2	(ii)	27
6.3	(iii)	28
6.4	(iv)	28
7	Question 7	29
7.1	(i)	29
7.2	(ii)	38
8	Question 8	40
8.1	(i)	40
8.2	(ii)	40
8.3	(iii) et (iv)	43
9	Question 9	49
9.1	(i)	49
9.2	(ii)	50
10	Question 10	57
10.1	(ii)	58
11	Question 11	65
11.1	(i)	65
11.2	(ii)	66
11.3	(iii)	68

1 Question 1

1.1 (i)

Le système de vote en Australie pour les élections fédérales est le vote préférentiel. Ce système de vote permet de simuler plusieurs tours d'élection avec un seul bulletin. Le fonctionnement du vote est le suivant:

Sur son bulletin de vote, chaque électeur classe tous les candidats par ordre de préférence: 1,2,3 etc. Si un candidat est classé premier par plus de 50 pourcents des votants, il est élu. Si ce n'est pas le cas, on retire le candidat ayant reçu le moins de vote en première position. Les électeurs ayant classé le candidat éliminé en première position voient donc leur classement modifié : le deuxième au classement passe premier, le troisième passe deuxième etc. De nouveau on regarde si un candidat est classé premier par plus de 50 pourcents des votant. Si c'est le cas il est élu, sinon on répète les étapes précédentes jusqu'à ce qu'un candidat soit élu.

Un des principaux avantages de ce système d'élection est donc que les électeurs peuvent s'exprimer sur l'ensemble des candidats.

Cette méthode présente également certains inconvénients. Tout d'abord, elle nécessite des procédures de dépouillement contraignantes. De plus, il a été montré que de nombreuses personnes classaient les candidats au hasard sur leur bulletin. Bien que l'ordre des candidats sur les bulletins de vote soit désigné au hasard avant l'élection, les candidats en haut du bulletin obtiennent donc systématiquement un résultat artificiellement gonflé. Enfin, cette méthode ne satisfait pas entièrement le critère de Condorcet qui stipule que le candidat élu doit gagner en duel de vote face à n'importe quel autre candidat.

Ces particularités du vote préférentiel ne sont pas recouvertes par nos données.

Une autre particularité du vote en Australie est qu'il est obligatoire pour les élections nationales depuis 1924 (1984 pour les autochtones). En cas de non-participation à un vote, la personne en question risque aujourd'hui une amende de 20 dollars australiens. Lors des élections de 2005, cette amende était de 110 dollars australiens.

Le taux de participation dans le Territoire du Nord reste cependant relativement faible. En 2005 il était de seulement 80 pourcents (pour une moyenne nationale de 95 pourcents). Ceci est probablement due à la faible densité de population de cette région.

1.2 (ii)

Dans l'absolu, porter son choix sur un candidat particulier lors d'une élection politique demande de mettre en perspectives de nombreuses données : propositions visant à améliorer telle ou telle situations dans tel ou tel domaines, capacité à mettre en place sa politique, compétences etc. Ceci nécessite donc le traitement d'une quantité conséquente d'informations qui se distinguent de l'attraction physique que peut exercer le candidat.

Cependant dans les faits, les votants sont généralement mal informés et la quantité trop importante

de critères à juger peuvent les inciter à extrapoler des informations limitées.

En outre, nous pensons que la beauté peut influencer positivement des électeurs peu renseignés en leur inspirant une plus grande confiance envers le candidat, et même en octroyant artificiellement un plus fort charisme au candidat.

Nous pensons aussi que les électeurs auront tendance à voter pour des candidats ayant la même couleur de peau qu'eux car ils se retrouveront superficiellement plus en ces candidats.

2 Question 2

On retranscrit dans le tableau suivant les corrélations des variables correspondants au 4 évaluateurs de la beauté.

	Homme indigène de 24 ans	Femme indigène de 39 ans	Femme non-indigène de 24 ans	Homme non-indigène de 40 ans
Homme indigène de 24 ans	1	0,58	0,60	0,46
Femme indigène de 39 ans	0,58	1	0,47	0,46
Femme non-indigène de 24 ans	0,60	0,47	1	0,46
Homme non-indigène de 40 ans	0,46	0,46	0	1

Figure 1: Tableau des corrélations au sens de Pearson entre les notes des évaluateurs

Rhodes explique dans son travail sur la psychologie évolutionniste de la beauté faciale que l'existence d'accords interculturels sur l'attractivité remettent en question une opinion de longue date selon laquelle nos préférences reflètent les normes arbitraires de beauté établies par les cultures. De la même manière, Langlois affirme dans son travail des années 2000 que "raters agree about who is and is not attractive, both within and across cultures". Ainsi selon la littérature scientifique, la notation de la beauté physique d'une personne ne devrait pas trop varier selon qui note l'individu .

Le tableau de corrélation que l'on a obtenu est en accord avec ces résultats. En effet, on remarque une corrélation positive et forte entre les notations des 4 évaluateurs. La corrélation positive traduit le fait que les évaluateurs vont dans le même "sens" au sujet de ce qui définit la beauté.

3 Question 3

Dans cette question, on s'occupe de la moyenne et la variance des notes attribuées par les quatre évaluateurs.

	Moyenne	Variance
Femme indigène de 39 ans	4,65	3,82
Homme indigène de 24 ans	5,11	1,64
Femme non-indigène de 24 ans	5,09	4,08
Homme non-indigène de 40 ans	4,70	1,90

Figure 2: Moyennes et variances des notes attribuées par évaluateur

On remarque que les moyennes sont relativement proches entre les évaluateurs. Cependant les variances semblent différer entre hommes et femmes. En effet, elles sont respectivement de 1,64 et 1,90 pour les hommes lorsqu'elles sont de 3,82 et 4,08 pour les femmes. Ainsi la répartition des notes est plus dispersée pour les femmes que pour les hommes.

On peut également noter que les moyennes des deux noteurs les plus jeunes sont sensiblement plus élevées que celles des noteurs les plus âgés. Les deux évaluateurs de 24 ans ont donc été un petit peu plus indulgents sur leur notation moyenne que les deux noteurs de 39 et 40 ans.

Pour finir on observe que les non-indigènes ont une variance sensiblement supérieure à celle des indigènes.

On se propose aussi d'observer la moyenne, l'écart type, le minimum et le maximum des notes obtenues par les hommes et les femmes en ne considérant notamment que les notes des femmes puis des hommes.

	Moyenne	Écart Type	Min	Max
Femmes total	5.59	1.12	3.5	7.75
Hommes total	4.48	1.26	2.25	7.5
Femmes notées par Femmes	5.78	1.46	3.0	8.0
Femmes notées par Hommes	5.4	1.06	4.0	7.5
Hommes notées par Femmes	4.34	1.63	1.0	8.5
Hommes notées par Hommes	4.62	1.1	3.0	7.0

Figure 3: Moyennes, variances, minimum et maximum des notes de beauté reçus selon le sexe

Les femmes sont en moyenne significativement mieux notées que les hommes avec un écart type similaire. On observe de plus que les évaluateurs sont moins exigeant que leur collègue du sexe opposé pour noter les candidats de leur sexe. Ainsi, les évaluateurs hommes notent mieux les candidats masculins que les évaluateurs femmes, et moins bien les candidates que les évaluateurs femmes.

Maintenant, on effectue pour chacune des variables une normalisation afin qu'elles soient centrées et réduites pour chaque évaluateur. Pour cela, on a soustrait à chaque variable la moyenne des notes de son évaluateur et on a divisé le tout par la racine de la variance des notes de son évaluateur. Nous nommons ces nouvelles variables les z-scores

Cette normalisation nous permet de pouvoir comparer les notes attribuées par les évaluateurs. En ef-

et chaque personne intériorise une échelle de notation propre à lui. Par exemple, attribuer un 7 peut être une excellente note pour une première personne mais une note moyenne pour une autre. En normalisant les notes, on les met sur une échelle commune qui nous permet de les comparer.

Nous proposons ainsi deux nouvelles variables synthétiques des scores de beauté :

- Une première est simplement pour chaque candidat la moyenne des z-scores lui ayant été attribués par chaque évaluateur. Cette nouvelle variable se justifie naturellement et est permise par ce que nous avons expliqué au paragraphe précédent.

- Une deuxième variable possible est la suivante. On crée d'abord un score de beauté autochtone en faisant la moyenne entre les scores de beauté attribués par les autochtones. On procède de façon similaire avec les non-autochtones afin de créer un score de beauté non-autochtone. Ensuite on crée la variable de score de beauté composite qui consiste en la moyenne entre le score de beauté autochtone et le score de beauté non-autochtone pondéré par la proportion d'autochtone dans la circonscription. Ce nouveau score se justifie par le fait que le score de beauté que nous étudions doit refléter la perception des électeurs moyens à circonscription donnée et donc prendre compte de l'identité ethnique de la circonscription.

4 Question 4

Nous allons dans un premier temps effectuer les statistiques descriptives des variables étant susceptibles d'être corrélées avec la beauté. Le graphique qui suit représente la répartition de caractéristiques faciales.

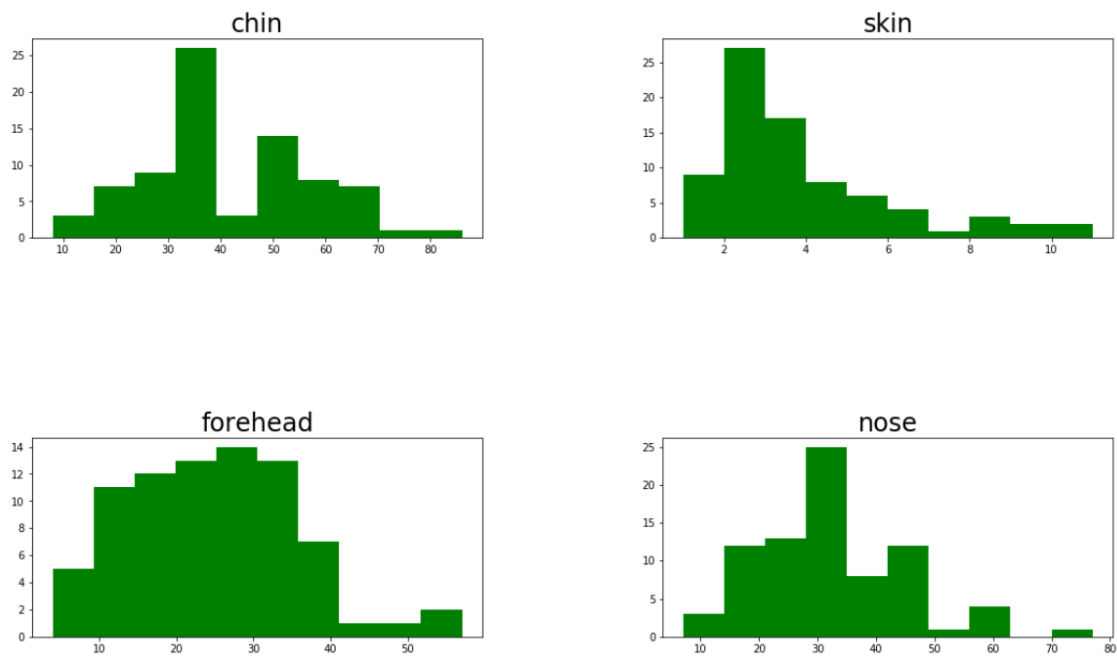


Figure 4: Histogrammes de la taille du menton, de la couleur de la peau, de la taille du front et de la taille du nez des candidats (de gauche à droite puis de haut en bas)

On remarque que les variables sur le menton, le front et le nez semblent être plutôt normalement distribuées. De plus, on sait que la répartition de la taille de ces caractéristiques physique est normalement distribuée dans la population. Ceci nous incite à penser que la taille du nez, du menton et du front ne sont pas directement discriminant pour le fait de se présenter ou non aux élections.

Par contre, au sujet de la couleur de la peau, on voit qu'il y a une prédominance de candidat de couleur blanche. Il serait intéressant de comparer cette répartition avec celle de la couleur de peau de l'ensemble des habitants de la région des territoires du nord.

On propose maintenant de regarder la moyenne, l'écart type, le minimum et le maximum de la perte de votes reçue selon différents groupes d'appartenance du candidat. On résume nos résultats dans le tableau suivant.

	Moyenne	Ecart type	Min	max
Total	0.32	0.22	0.01	0.76
Femme	0.33	0.21	0.03	0.67
Candidat en exercice	0.56	0.12	0.31	0.76
ALP	0.53	0.16	0.12	0.76
CLP	0.35	0.11	0.16	0.54
Greens	0.09	0.03	0.05	0.16
IND	0.10	0.14	0.01	0.57

Figure 5: Part des votes reçus moyens selon le groupe

Dans un premier temps il est intéressant de remarquer que la moyenne des votes reçus par les femmes est presque identique à celle de l'ensemble de la population. Les écarts types sont de même presque égaux. On en conclut qu'il ne semble pas y avoir de discrimination par le vote selon le sexe.

On observe aussi qu'un candidat en exercice reçoit significativement plus de voix que les autres candidats. Ils ont de plus un plus faible écart type et le candidat ayant reçu la plus forte part des voix était un candidat masculin en exercice du parti ALP.

Il serait intéressant d'observer l'effet de la couleur de la peau sur le score attribué par les noteurs autochtones et non autochtones. Pour cela on a fait la moyenne des notes attribuées par des autochtones (*note_i*) et la moyenne des notes attribuées par des non-autochtones (*note_ni*) pour chaque candidat. On trace ensuite l'évolution de ces notes en fonction de la couleur de peau avec un intervalle de confiance à 95 pourcents. Du fait que la formule de la variance empirique (nécessaire pour calculer l'intervalle de confiance empirique) nécessite d'avoir strictement plus d'une observation par groupe (ce qui n'est pas le cas avec la couleur de la peau ici), on propose de créer un nouveau score de couleur de la peau qui crée 6 classes équilibrées de couleur de peau (0 étant la plus claire et 3 étant la plus foncée). On trace de même nos scores de beauté en fonction de ce nouveau score de couleur de peau (que l'on nomme *skin_g*).

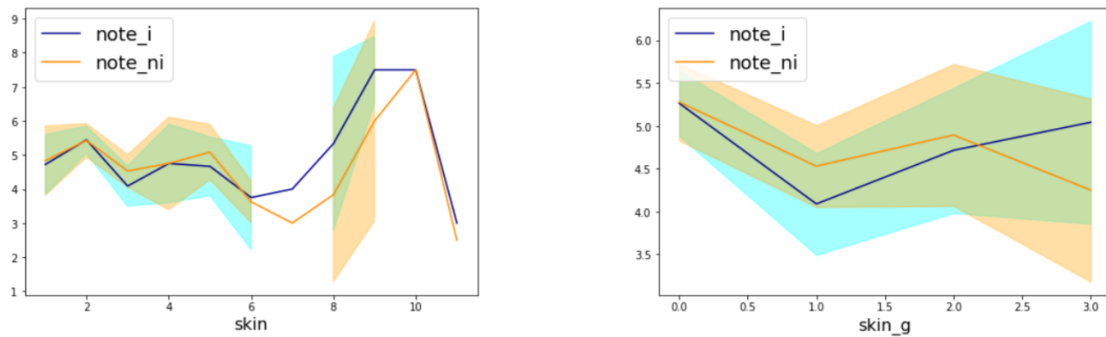


Figure 6: Évolution de la note moyenne indigène (*note_i*) et non indigène (*note_ni*) en fonction de la couleur de peau (Gauche : score peau de la base, Droite : nouveau score peau pour 6 classes équilibrées)

Les graphiques précédents semblent nous indiquer que les personnes d'origines indigènes auront tendance à attribuer un meilleur score que les non-indigènes aux candidats de peau foncée. Cette tendance s'inverse pour les candidats de peau plus claire. Ainsi les personnes non-autochtones donnent une meilleur note aux candidats de peau claire que les candidats autochtones. Cependant on voit que les intervalles de confiance sont très larges donc on ne peut pas tirer de conclusions.

Maintenant on veut regarder l'évolution de la part de votes reçus selon la couleur de la peau et le score de beauté.

Dans un premier temps on regarde l'évolution de la part de votes reçu en fonction de la couleur de la peau en procédant de la même manière que précédemment (avec les deux scores de couleurs de peau).

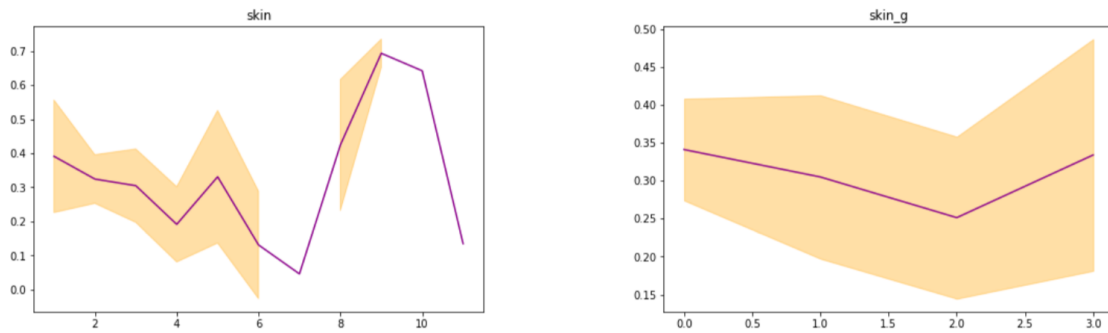


Figure 7: Évolution de la part de votes obtenus moyenne en fonction de la couleur de peau (Gauche : score peau de la base, Droite : nouveau score peau pour 6 classes équilibrées)

Le graphique de gauche pourrait être trompeur en nous indiquant qu'une peau foncée pourrait nous permettre d'obtenir beaucoup de votes. Or le deuxième graphique nous montre bien que la couleur de peau ne semble pas influencer la part de vote reçue. De plus on observe une très grande variance dans nos données. Nous ne pouvons rien conclure ici.

On propose maintenant d'afficher l'évolution de la part des votes obtenus moyenne en fonction du score de beauté obtenu par les candidats. Concernant l'intervalle de confiance de ce graphique, on a un problème similaire à celui rencontré lorsque l'on utilisait le score de la couleur de peau. Ici, pour mieux appréhender les tendances, on se propose de regarder le même graphique en prenant la partie entière du score de beauté (que l'on nomme 'score_beaute.int'). On obtient les deux graphiques suivants :

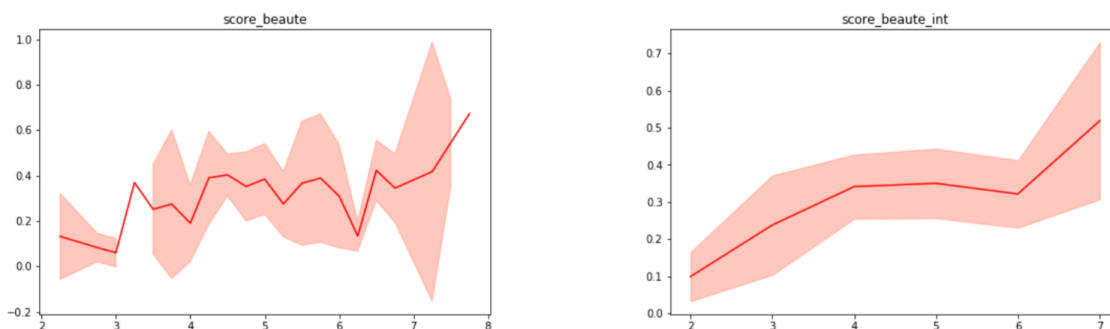


Figure 8: Évolution de la part de votes obtenus moyenne en fonction du score de beauté général (Gauche : moyenne des notes de beauté obtenues, Droite : partie entière de la moyenne des notes de beauté obtenues)

Le deuxième graphique semble nous indiquer que le nombre de part de vote est effectivement croissant en fonction du score de beauté. Tout cela avec un intervalle de confiance relativement raisonnable pour ne pas contredire nos résultats. Ceci est un premier indice en faveur de l'influence positive de la beauté sur la part de votes obtenus.

On fait les mêmes graphiques mais en utilisant cette fois-ci le score de beauté pondéré par la proportion d'aborigène dans la circonscription (comme expliqué à la question précédente).

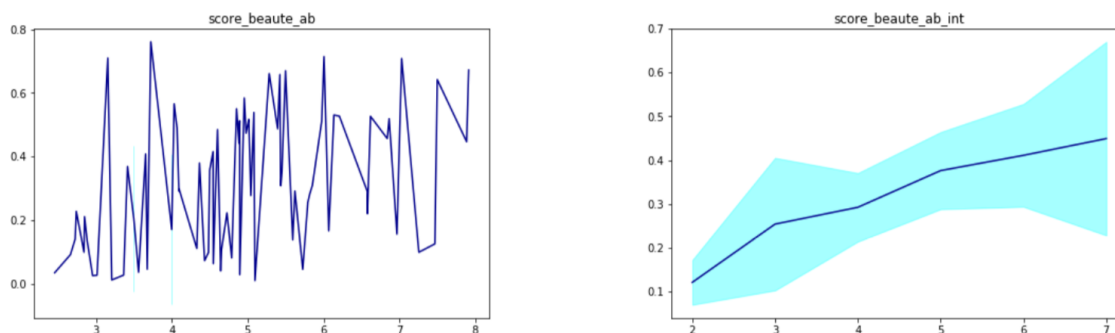


Figure 9: Évolution de la part de votes obtenus moyenne en fonction du score de beauté pondéré (Gauche : moyenne des notes de beauté obtenues, Droite : partie entière de la moyenne des notes de beauté obtenues)

On obtient des résultats similaires aux précédents mais on observe que cette fois-ci la tendance positive est encore plus franche ce qui est un résultat en faveur de notre choix de score de beauté.

On propose enfin de regarder l'ensemble des corrélations linéaires au sens de Pearson de nos variables. On résume nos résultats dans la figure suivante.

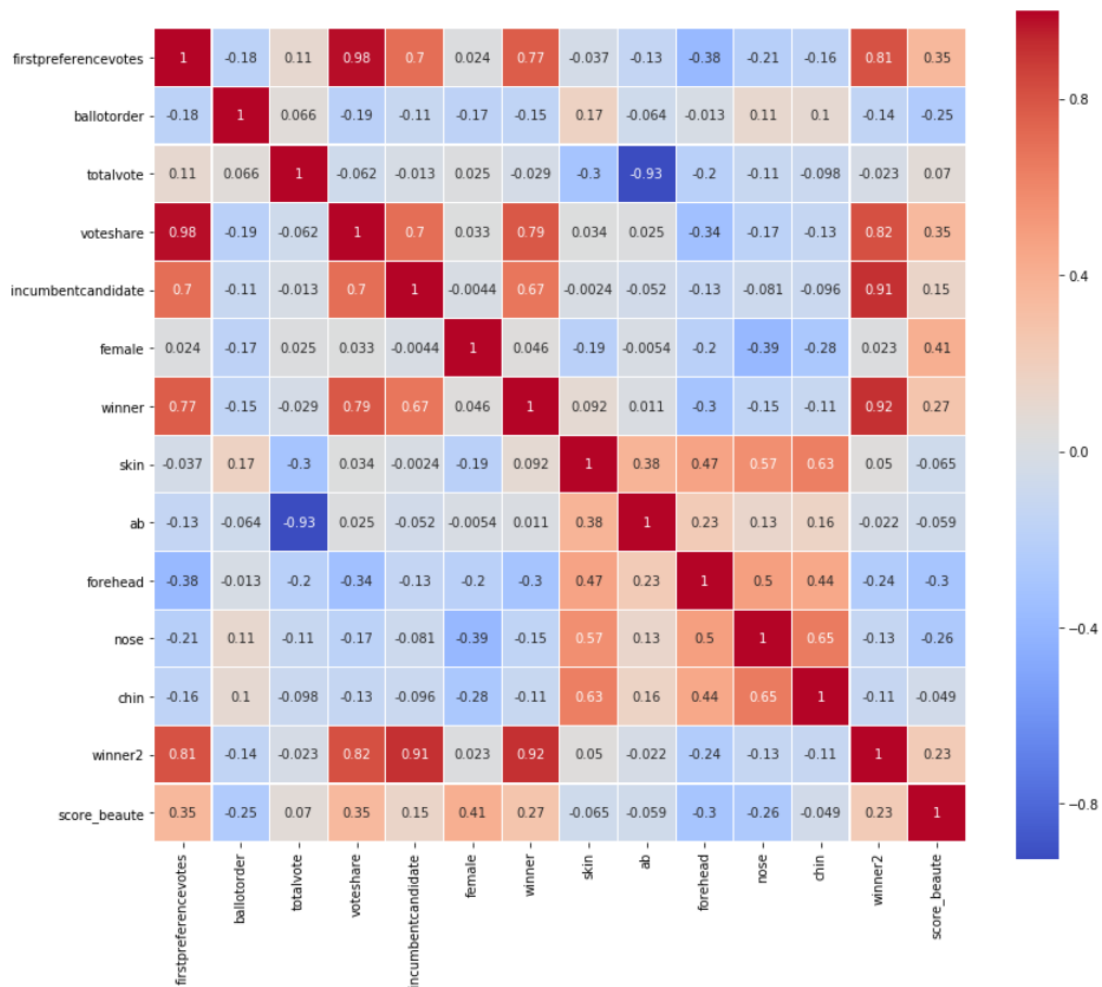


Figure 10: Corrélations linéaires au sens de Pearson

On retrouve dans ce tableau plusieurs résultats formulés précédemment. On retrouve le fait que le score de beauté et la part de votes obtenus sont positivement significativement corrélés. On retrouve aussi le fait que les femmes soient mieux notées que les hommes.

On apprend aussi plusieurs choses sur nos données. On voit notamment que la taille des caractéristiques physiques (nez, front etc.) sont très positivement corrélées. Mais on apprend surtout qu'il semble que les circonscriptions avec un grand pourcentage d'aborigènes sont très significativement des circonscriptions avec peu de votant.

5 Question 5

5.1 (i)

On cherche désormais à grapher la part de votes obtenus pour chaque candidat (variable "voteshare") en fonction du score de beauté moyen obtenu après normalisation des notes individuelles (voire question 3).

Pour cela nous allons procéder à notre première régression linéaire avec comme variable cible la part de vote obtenus et comme variable explicative le score de beauté moyen.

On considère ainsi le modèle

$$Y = X'\beta + \epsilon$$

On reprend les notation du cours. On supposera notamment toujours que les variables $(Y_i, X_i)_{i=1,\dots,79}$ seront iid de même loi que (Y, X) et que les paramètres indicés par 0 seront les vrais paramètres du modèle.

Pour estimer le modèle linéaire on fait l'hypothèse d'exogénéité de X :

$$\mathbb{E}(X\epsilon) = 0$$

On peut ainsi estimer β_0 par l'estimateur des MCO (Moindres carrés ordinaires) :

$$\hat{\beta}_{MCO} = \left(\frac{1}{n} \sum_{i=1}^n X_i X_i'\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n X_i Y_i\right)$$

On suppose aussi qu'il y a une bonne répartition des X_i (ce qui est bien le cas ici).

En rajoutant comme hypothèses que les ϵ_i sont homoscédastiques et qu'ils sont gaussiens, on peut tester les alternatives $H_0 : \beta_0 = 0$ et $H_1 : \beta_0 \neq 0$ en faisant un t-test avec la statistique $t_{\hat{\beta}} = \frac{\hat{\beta}}{se(\hat{\beta})}$ ($se(\hat{\beta})$ est l'écart type empirique). On utilise ainsi le p-valeur pour rejeter ou ne pas rejeter notre hypothèse de significativité du test.

Dans notre cas, on a les résultats suivants:

Table 1: Regression de la part de vote sur le score de beauté moyen

	<i>Dependent variable:</i>
	voteshare
beaute_mean	10.428*** (3.074)
Constant	40.000*** (2.424)
Observations	79
R ²	0.130
Adjusted R ²	0.119
Residual Std. Error	21.544 (df = 77)
F Statistic	11.508*** (df = 1; 77)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

On voit qu'on rejette H_0 au niveau 1 pourcent et que la beauté a un effet positif sur la part de voix récoltés.

On obtient ainsi le graphique suivant:

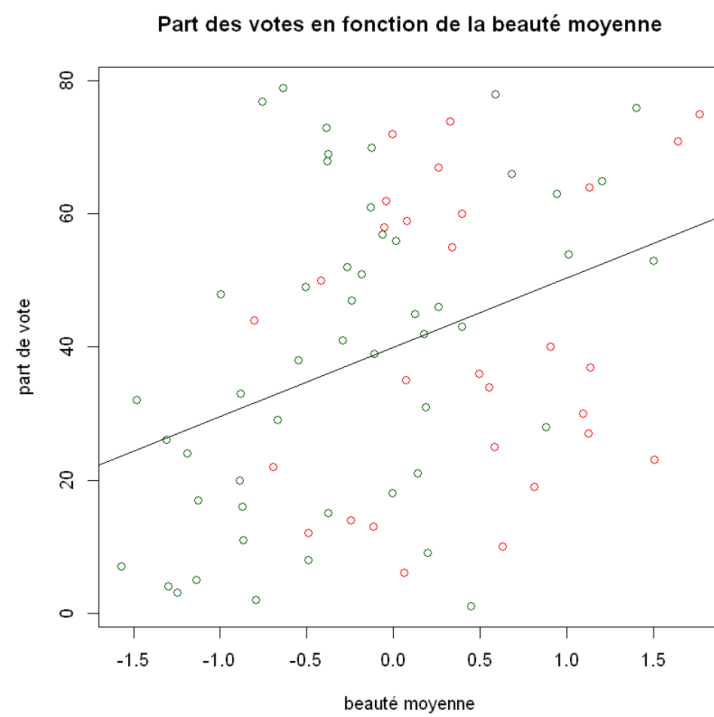


Figure 11: Part de votes obtenus en fonction du score de beauté et sa droite de régression (Rouge : Femmes, Vert: Hommes)

5.2 (ii)

On effectue le même travail que précédemment mais cette fois-ci en séparant les candidats sortants des opposants :

Table 2: Résultats Régression part de votes sur beauté moyenne pour les candidats en exercice

	<i>Dependent variable:</i>
	voteshare
beaute_mean	1.928 (3.263)
Constant	64.865*** (2.320)
Observations	22
R ²	0.017
Adjusted R ²	-0.032
Residual Std. Error	10.496 (df = 20)
F Statistic	0.349 (df = 1; 20)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 3: Résultats Régression part de votes sur beauté moyenne pour les candidats qui ne sont pas en exercice

	<i>Dependent variable:</i>
	voteshare
beaute_mean	9.226*** (2.797)
Constant	30.932*** (2.284)
Observations	57
R ²	0.165
Adjusted R ²	0.150
Residual Std. Error	17.175 (df = 55)
F Statistic	10.878*** (df = 1; 55)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Concernant l'influence de la beauté sur la part de vote on observe plusieurs choses. Dans un premier

temps, la variable de score de beauté moyen ne semble pas significative pour les candidats en exercice (on ne rejette pas H_0) mais l'est au niveau 1 pourcent pour les candidats qui ne sont pas en exercice. Dans un deuxième temps on observe que l'effet marginale d'une augmentation de la variable de beauté est significativement plus forte pour les candidats en exercice que pour les autres candidats, tout en étant positive dans les deux cas.

On interprète ces résultats en argumentant que la beauté peut certainement influencer davantage le vote pour les nouveaux candidats que pour les anciens. En effet, les anciens candidats ont déjà leur électorat et sont mieux identifiés, ce qui peut diminuer la nécessité d'extrapoler des facteurs tels que la beauté physique. Par contre la beauté des nouveaux candidats semble être un atout pour attirer une attention électorale.

On a ainsi les graphiques suivants :

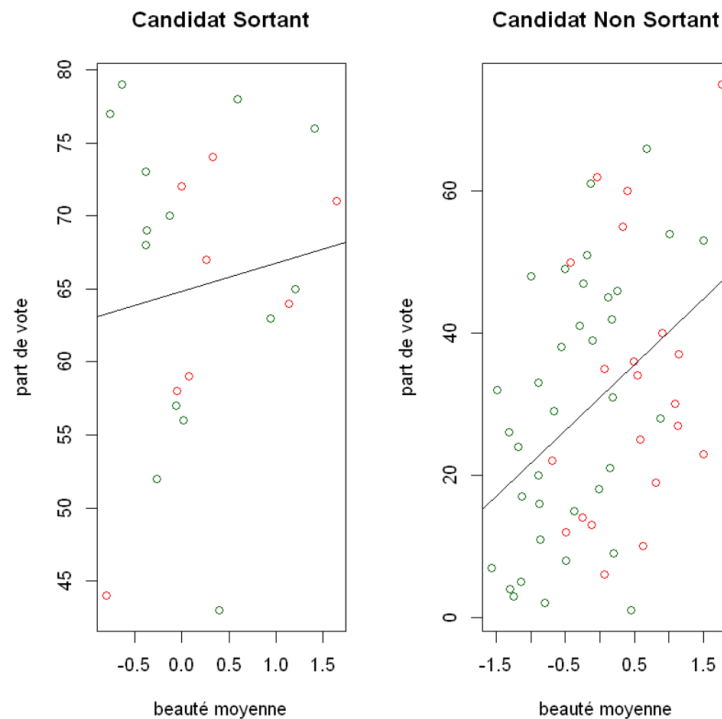


Figure 12: Part de votes obtenus en fonction du score de beauté et sa droite de régression pour les candidats en exercice (gauche) et ceux qui ne le sont pas (droite) (Rouge : Femmes, Vert: Hommes)

5.3 (iii)

Dans cette question on s'intéresse à l'influence de la couleur de la peau sur la part de vote obtenue pour respectivement tous les candidats, les candidats en exercice et les candidats qui ne sont pas en exercice. En utilisant le même modèle et les mêmes hypothèses qu'avant on obtient les résultats suivants:

Table 4: Regression de la part de vote sur la couleur de la peau

	<i>Dependent variable:</i>
	voteshare
skin	0.270 (1.164)
Constant	39.079*** (4.752)
Observations	79
R ²	0.001
Adjusted R ²	-0.012
Residual Std. Error	23.090 (df = 77)
F Statistic	0.054 (df = 1; 77)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 5: Résultats Régression part de votes sur la couleur de la peau pour les candidats en exercice

	<i>Dependent variable:</i>
	voteshare
skin	0.798 (0.860)
Constant	62.507*** (3.671)
Observations	22
R ²	0.041
Adjusted R ²	-0.007
Residual Std. Error	10.367 (df = 20)
F Statistic	0.861 (df = 1; 20)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 6: Résultats Régression part de votes sur beauté moyenne pour les candidats qui ne sont pas en exercice

	<i>Dependent variable:</i>
	voteshare
skin	−0.013 (1.194)
Constant	30.309*** (4.783)
Observations	57
R ²	0.00000
Adjusted R ²	−0.018
Residual Std. Error	18.797 (df = 55)
F Statistic	0.0001 (df = 1; 55)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Dans chacune des trois situations, on ne rejette pas le fait que la variable qui indique la couleur de peau ne soit pas significative. L'estimation du coefficient associé est, de plus, à chaque fois proche de zéro, avec un écart type ne nous permettant pas d'être convaincu de la positivité voire de la négativité du coefficient.

On a les graphiques suivants :

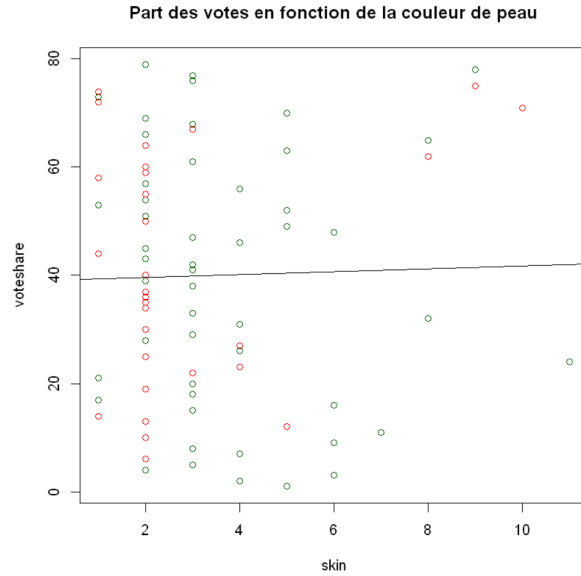


Figure 13: Part de votes obtenus en fonction de la couleur de peau et sa droite de régression pour tous les candidats (Rouge : Femmes, Vert: Hommes)

Ces résultats ne sont pas similaires à ceux de la question précédente. Ils semblent nous inciter à croire que la couleur de peau n'a pas d'influence directe sur la part des votes de façon générale. Il serait intéressant de tester en quelle mesure la couleur de peau peut avoir une influence sur la part des votes dans le cas où le candidat provient d'une circonscription où il y a beaucoup d'indigènes. C'est l'objet de la question suivante.

5.4 (iv)

Dans cette question on sépare les candidats en deux groupes de même taille: les candidats provenant d'une circonscription avec un faible pourcentage d'indigène et ceux venant d'une circonscription avec un fort pourcentage d'indigène. Pour cela on retrouve la médiane des valeurs de la part d'individus autochtones dans l'électorat (environ 13,4 pourcent d'autochtones) et on fait deux groupes grâce à celle-ci. On réalise ensuite les mêmes tests que précédemment.

Pour le groupe de candidats venant d'une circonscription avec un faible pourcentage d'indigène, on obtient les résultats suivants:

On observe que dans cette situation la couleur de peau est négativement significative au taux 1 pourcent pour l'ensemble des candidats. Cela semble en fait être particulièrement le cas pour les candidats qui ne sont pas en exercice (28 candidats sur les 40) qui présentent les mêmes résultats. D'un autre côté il semblerait que la couleur de peau ne soit pas significative pour les candidats qui sont en exercice ici. Les raisons à cela doivent être similaires à celles que l'on a évoqué pour expliquer le fait que la beauté semble ne pas être significative pour les candidats en place.

Table 7: Résultats Régression part de votes sur beauté moyenne pour les candidats dans une circonscription avec peu d'indigènes

	<i>Dependent variable:</i>
	voteshare
skin	-7.410*** (2.159)
Constant	62.623*** (7.053)
Observations	40
R ²	0.237
Adjusted R ²	0.217
Residual Std. Error	19.876 (df = 38)
F Statistic	11.782*** (df = 1; 38)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 8: Résultats Régression part de votes sur beauté moyenne pour les candidats qui sont en exercice dans une circonscription avec peu d'indigènes

	<i>Dependent variable:</i>
	voteshare
skin	-1.880 (1.642)
Constant	70.033*** (4.741)
Observations	12
R ²	0.116
Adjusted R ²	0.027
Residual Std. Error	8.212 (df = 10)
F Statistic	1.310 (df = 1; 10)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

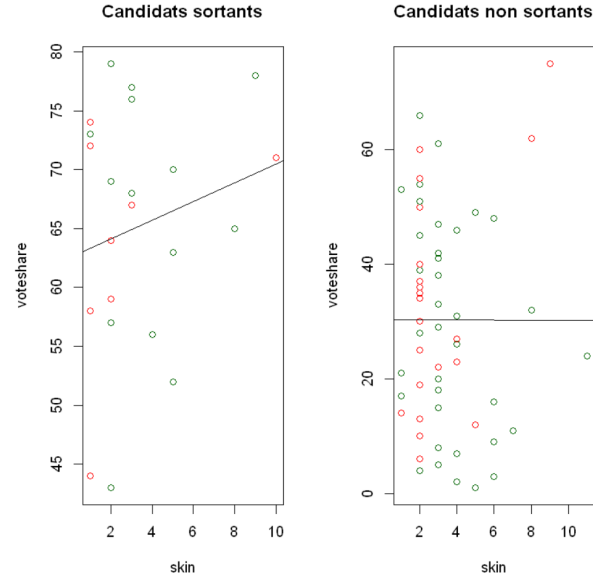


Figure 14: Part de votes obtenus en fonction de la couleur de peau et sa droite de régression pour les candidats en exercice (gauche) et ceux qui ne le sont pas (droite) (Rouge : Femmes, Vert: Hommes)

Table 9: Résultats Régression part de votes sur beauté moyenne pour les candidats qui ne sont pas en exercice dans une circonscription avec peu d'indigènes

<i>Dependent variable:</i>	
	voteshare
skin	-7.119*** (1.990)
Constant	52.620*** (6.802)
Observations	28
R ²	0.330
Adjusted R ²	0.304
Residual Std. Error	14.985 (df = 26)
F Statistic	12.792*** (df = 1; 26)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

On peut ainsi avoir les graphiques suivants :

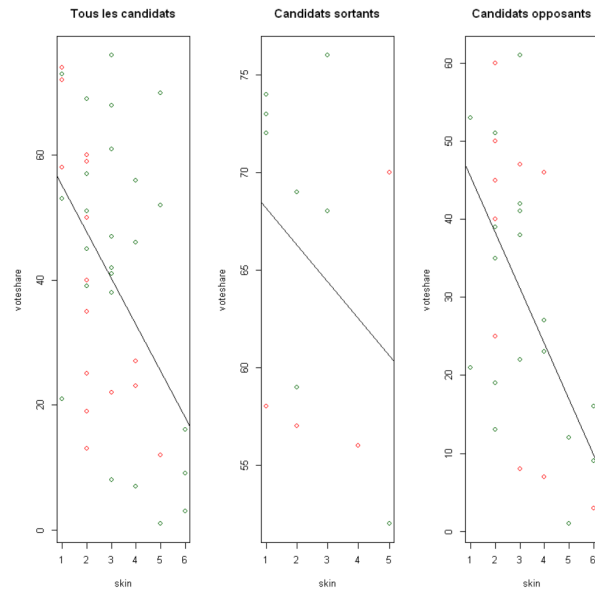


Figure 15: Part de votes obtenus en fonction de la couleur de peau dans les circonscription avec peu d'indigènes et sa droite de régression pour respectivement de gauche à droite tous les candidats, les candidats en exercice et ceux qui ne le sont pas (Rouge : Femmes, Vert: Hommes)

On procède maintenant à la même étude mais avec les circonscriptions avec une part importante d'indigènes.

Table 10: Résultats Régression part de votes sur beauté moyenne pour les candidats dans une circonscription avec beaucoup d'indigènes

	<i>Dependent variable:</i>
	voteshare
skin	2.674* (1.343)
Constant	28.537*** (6.413)
Observations	39
R ²	0.097
Adjusted R ²	0.072
Residual Std. Error	22.825 (df = 37)
F Statistic	3.962* (df = 1; 37)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 11: Résultats Régression part de votes sur beauté moyenne pour les candidats qui sont en exercice dans une circonscription avec beaucoup d'indigènes

	<i>Dependent variable:</i>
	voteshare
skin	1.680 (1.234)
Constant	57.539*** (6.773)
Observations	10
R ²	0.188
Adjusted R ²	0.087
Residual Std. Error	12.252 (df = 8)
F Statistic	1.853 (df = 1; 8)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 12: Résultats Régression part de votes sur beauté moyenne pour les candidats qui ne sont pas en exercice dans une circonscription avec beaucoup d'indigènes

	<i>Dependent variable:</i>
	voteshare
skin	2.176 (1.401)
Constant	21.931*** (6.306)
Observations	29
R ²	0.082
Adjusted R ²	0.048
Residual Std. Error	19.090 (df = 27)
F Statistic	2.414 (df = 1; 27)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Cette fois ci l'influence de la peau ne semble être significative que pour l'ensemble des candidats et seulement au taux 10 pourcents. De plus les coefficient sont cette fois ci positifs mais en valeur absolue environ 4 fois plus petits que les coefficients des circonscriptions avec peu d'indigènes.

On a entre autre on a les graphiques suivants :

En conclusions: Ainsi il semblerait que la couleur de peau a une plus grande influence sur les votes lorsque la circonscription a peu d'aborigènes et que le candidat n'est pas en exercice. Dans ce cas précis l'influence est négative, plus notre peau est foncée moins on est susceptible de recevoir de votes. Dans les autres cas, la couleur de la peau semble peu significative mais négative dans les circonscriptions avec peu d'aborigènes et positive dans les circonscriptions avec un fort taux d'aborigènes.

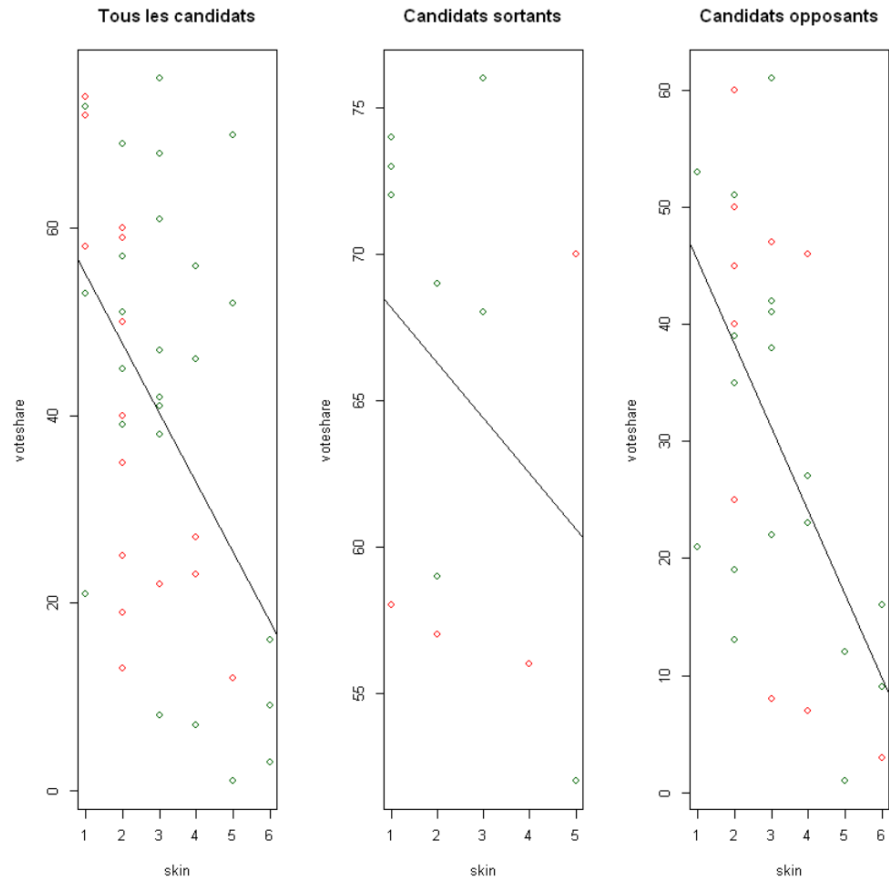


Figure 16: Part de votes obtenus en fonction de la couleur de peau dans les circonscription avec peu d'indigènes et sa droite de régression pour respectivement de gauche à droite tous les candidats, les candidats en exercice et ceux qui ne le sont pas (Rouge : Femmes, Vert: Hommes)

6 Question 6

6.1 (i)

Pour la suite de la question, on admet qu'une variable aléatoire X qui suit une loi de Gompertz de paramètres η et b , $X \sim G(\eta, b)$, a pour densité $f_{G(\eta, b)}(x) = b\eta e^{-bx} e^{-\eta e^{-bx}}$ sur \mathbb{R} et pour fonction de répartition $\mathbb{F}_{G(\eta, b)}(t) = e^{-\eta e^{-bt}}$ sur \mathbb{R} . On suppose que les $(\epsilon_{ij})_{j \in \llbracket 0, J \rrbracket}$ sont iid de loi $G(\eta, b)$ avec $\eta > 0$ et $b > 0$.

Soit i l'indice d'un électeur. Soit $k \in \llbracket 0, J \rrbracket$.

On cherche à calculer $\mathbb{P}(Y_i = k | X_1, \dots, X_J, \zeta_1, \dots, \zeta_J)$ que l'on notera parfois $\mathbb{P}(Y_i = k | X, \zeta)$.

Dans un premier temps on a clairement que

$$\mathbb{P}(Y_i = k | X_1, \dots, X_J, \zeta_1, \dots, \zeta_J) = \mathbb{P}(\max_{j \in \llbracket 0, J \rrbracket / \{k\}} U_{ij} < U_{ik} | X_1, \dots, X_J, \zeta_1, \dots, \zeta_J)$$

On a pour tous $j \in \llbracket 0, J \rrbracket$, $U_{ij} = X'_j \beta + \zeta_j + \epsilon_{ij}$.

Si $Z \sim G(\eta, b)$ et $c \in \mathbb{R}$. Quelle loi suit $Z + c$?

Observons: Soit $t \in \mathbb{R}$

$$\begin{aligned} \mathbb{P}(Z + c < t) &= \mathbb{P}(Z < t - c) \\ &= \mathbb{F}(t - c) \\ &= e^{-\eta e^{-b(t-c)}} \\ &= e^{-\eta e^{bc} e^{-bt}} \end{aligned}$$

Donc $Z + c \sim G(\eta e^{bc}, b)$.

Ainsi on a $\forall j \in \llbracket 0, J \rrbracket$, $U_{ij} | X_1, \dots, X_J, \zeta_1, \dots, \zeta_J \sim G(\eta e^{b(X'_j \beta + \zeta_j)}, b)$.

Quelle est la loi de $\max_{j \in \llbracket 0, J \rrbracket / \{k\}} U_{ij}$?

Soient $Z_1 \sim G(\eta_1, b)$ et $Z_2 \sim G(\eta_2, b)$ indépendantes.

Observons: Soit $t \in \mathbb{R}$

$$\begin{aligned} \mathbb{P}(\max(Z_1, Z_2) < t) &= \mathbb{P}(Z_1 < t, Z_2 < t) \\ &= \mathbb{P}(Z_1 < t) \mathbb{P}(Z_2 < t) \quad (\text{indépendance}) \\ &= e^{-\eta_1 e^{-bt}} e^{-\eta_2 e^{-bt}} \\ &= e^{-(\eta_1 + \eta_2) e^{-bt}} \end{aligned}$$

Donc $\max(Z_1, Z_2) \sim G(\eta_1 + \eta_2, b)$

De plus $\forall(a, b, c) \in \mathbb{R}^3$, $\max(a, b, c) = \max(a, \max(b, c))$.

Par indépendance des $(U_{ij})_{j \in \llbracket 0, J \rrbracket / \{k\}}$ on a donc :

$$\max_{j \in \llbracket 0, J \rrbracket / \{k\}} U_{ij} | X_1, \dots, X_J, \zeta_1, \dots, \zeta_J \sim G\left(\sum_{j=0, j \neq k}^J \eta e^{b(X'_j \beta + \zeta_j)}, b\right)$$

On a U_{ik} et $\max_{j \in \llbracket 0, J \rrbracket / \{k\}} U_{ij}$ indépendants.

Soient $Z \sim G(\eta_Z, b)$ et $Y \sim G(\eta_Y, b)$ indépendantes. Que vaut $\mathbb{P}(Z < Y)$?

Observons:

$$\begin{aligned} \mathbb{P}(Z < Y) &= \int_{-\infty}^{+\infty} \int_{-\infty}^y f_{G(\eta_Y, b)}(y) f_{G(\eta_Z, b)}(z) dz dy \quad (\text{indépendance}) \\ &= \int_{-\infty}^{+\infty} f_{G(\eta_Y, b)}(y) \mathbb{F}_{G(\eta_Z, b)}(y) dy \\ &= \int_{-\infty}^{+\infty} b \eta_Y e^{-by} e^{-\eta_Y e^{-by}} e^{-\eta_Z e^{-by}} dy \\ &= \int_{-\infty}^{+\infty} b \eta_Y e^{-by} e^{-(\eta_Z + \eta_Y) e^{-by}} dy \\ &= - \int_0^{+\infty} b \eta_Y z e^{-(\eta_Z + \eta_Y) z} \left(-\frac{1}{zb}\right) dz \quad (*) \\ &= \int_0^{+\infty} \eta_Y e^{-(\eta_Z + \eta_Y) z} dz \\ &= \eta_Y \left[-\frac{1}{\eta_Z + \eta_Y} e^{-(\eta_Z + \eta_Y) z} \right]_0^{+\infty} \\ &= \frac{\eta_Y}{\eta_Z + \eta_Y} \end{aligned}$$

(*) où on a fait le changement de variable $z = e^{-by}$ où l'on a $dy = -\frac{1}{bz} dz$ et les bornes $-\infty$ et $+\infty$ deviennent $+\infty$ et 0 .

Comme $U_{ik} | X, \zeta \sim G(\eta e^{b(X'_k \beta + \zeta_k)}, b)$ et $\max_{j \in \llbracket 0, J \rrbracket / \{k\}} U_{ij} | X, \zeta \sim G(\sum_{j=0, j \neq k}^J \eta e^{b(X'_j \beta + \zeta_j)}, b)$ et sont indépendants on a :

$$\begin{aligned}
\mathbb{P}(Y_i = k|X, \zeta) &= \mathbb{P}(\max_{j \in \llbracket 0, J \rrbracket / \{k\}} U_{ij} < U_{ik} | X, \zeta) \\
&= \frac{\eta e^{b(X'_k \beta + \zeta_k)}}{\sum_{j=0}^J \eta e^{b(X'_j \beta + \zeta_j)}} \\
&= \frac{e^{b(X'_k \beta + \zeta_k)}}{\sum_{j=0}^J e^{b(X'_j \beta + \zeta_j)}}
\end{aligned}$$

On a $X'_0 \beta + \zeta_0 = 0$ donc $\mathbb{P}(Y_i = k|X, \zeta) = \frac{1}{\sum_{j=0}^J e^{b(X'_j \beta + \zeta_j)}}$

Donc

$$\frac{\mathbb{P}(Y_i = k|X, \zeta)}{\mathbb{P}(Y_i = 0|X, \zeta)} = e^{b(X'_k \beta + \zeta_k)}$$

D'où

$$\ln(\mathbb{P}(Y_i = k|X, \zeta)) - \ln(\mathbb{P}(Y_i = 0|X, \zeta)) = \ln\left(\frac{\mathbb{P}(Y_i = k|X, \zeta)}{\mathbb{P}(Y_i = 0|X, \zeta)}\right) = b(X'_k \beta + \zeta_k)$$

6.2 (ii)

Soit I le nombre total d'électeurs.

On pose $\forall k \in \llbracket 0, J \rrbracket$, V_k le nombre d'électeurs ayant voté pour le candidat k .

On a donc $\forall k \in \llbracket 0, J \rrbracket$, $V_k = \sum_{i=1}^I \mathbb{1}_{Y_i=k}$

Ainsi on a $\forall k \in \llbracket 0, J \rrbracket$:

$$\begin{aligned}
\mathbb{E}[V_k|X, \zeta] &= \mathbb{E}\left[\sum_{i=1}^I \mathbb{1}_{Y_i=k}\right] \\
&= \sum_{i=1}^I \mathbb{P}(Y_i = k|X, \zeta) \\
&= \sum_{i=1}^I \frac{e^{b(X'_k \beta + \zeta_k)}}{\sum_{j=0}^J e^{b(X'_j \beta + \zeta_j)}} \\
&= \frac{I e^{b(X'_k \beta + \zeta_k)}}{\sum_{j=0}^J e^{b(X'_j \beta + \zeta_j)}}
\end{aligned}$$

Où l'on a fait l'hypothèse que pour tous les $i \in \llbracket 1, I \rrbracket$ les $(\epsilon_{ij})_{j=0, \dots, J}$ partagent le même paramètre

b.

La part espérée d'électeurs ayant voté pour le candidat k conditionnellement à X et ζ est donc

$$R_k = \frac{\mathbb{E}[V_k|X, \zeta]}{I} = \frac{e^{b(X'_k\beta + \zeta_k)}}{\sum_{j=0}^J e^{b(X'_j\beta + \zeta_j)}}$$

Le ratio entre la part de vote pour les deux candidats k et k' est donc $\frac{R_k}{R_{k'}} = \frac{e^{b(X'_k\beta + \zeta_k)}}{e^{b(X'_{k'}\beta + \zeta_{k'})}}$ qui ne dépend pas de J .

Ainsi dans un système à J candidats ou $J + 1$ candidats ce ratio reste identique. Cette propriété s'appelle l'IANP, l'indépendance aux alternatives non pertinentes.

6.3 (iii)

Dans un contexte de scrutin uninominal, cette hypothèse ne semble pas très raisonnable. Cette hypothèse découle notamment du fait que les $(\epsilon_{ij})_{j=0,\dots,J}$ soient iid. Or la sensibilité politique propre à chaque électeur est probablement un facteur important lors du choix d'un candidat.

Prenons l'exemple suivant où $J = 2$. Pour faire simple, on imagine une situation où un candidat est situé à droite de l'échiquier politique et l'autre candidat à gauche. Chacun recueille la moitié des votes des électeurs (on néglige le vote blanc).

Le ratio parts de vote à gauche sur parts à droite est donc $\frac{R_G}{R_D} = 1$.

Passons maintenant à système où $J = 3$ en ajoutant un nouveau candidat de gauche. On a maintenant $R_D = 0.5, R_G^1 = 0.25$ et $R_G^2 = 0.25$ car les électeurs de gauche se partagent entre les deux candidats de gauche.

Les ratios sont complètement changés : $\frac{R_G}{R_D} \neq \frac{R_G^1}{R_D}$

Cette hypothèse n'est donc pas très raisonnable.

6.4 (iv)

On note I_d le nombre total de votant dans le district d et \tilde{I}_d le nombre total de votant n'ayant pas voté blanc dans le district d (en particulier $\tilde{I}_d = \sum_{j=1}^{J_d} V_j$).

Dans notre base de données nous n'avons accès qu'à $\tilde{R}_{jd} = \frac{V_{jd}}{\tilde{I}_d}$ qui est la proportion des votes recueillis par le candidat j du district d parmi les votes non blancs. Dans la base de données nous n'avons donc pas accès aux $(R_j)_{j=0,\dots,J_d}$ des district d qui sont nécessaires dans notre modèle.

Selon Shephard, $R_0 = 0.0375$ en moyenne dans tout le territoire du nord au élections générales de

2005. On fait donc ici l'hypothèse que $R_0 = 0.0375$ dans chaque circonscription. Ainsi on a $\frac{\tilde{I}_d}{I_d} = \frac{I - V_0}{I} = 1 - R_0 = 0,9625$ pour tous d .

En observant avec $R_{jd} = \frac{V_{jd}}{I_d} = \frac{V_{jd}}{I_d} \frac{\tilde{I}_d}{I_d} = \tilde{R}_{jd} \frac{\tilde{I}_d}{I_d}$ pour tous d .

On peut donc retrouver les R_{jd} en utilisant les \tilde{R}_{jd} de notre base et en les multipliant par 0,9625

On a par la question précédente que $\frac{R_{jd}}{R_0} = e^{b(X'_{jd}\beta + \zeta_{jd})}$ d'où $\ln(R_{jd}) - \ln(R_0) = b(X'_{jd}\beta + \zeta_{jd})$.

Quitte à normaliser le modèle on peut prendre $b = 1$ sans perte de généralité.

Ainsi on se retrouve avec un modèle de la forme $Y = X'\beta + \epsilon$ avec ici $Y_j = \ln(R_j) - \ln(R_0)$ et $\epsilon_j = \zeta_j$.

En supposant donc que les composantes de tous les X_j soient linéairement indépendantes, que les $(\ln(R_j) - \ln(R_0), X_j)_{j=1, \dots, J}$ sont iid et que l'hypothèse d'exogénéité ($\mathbb{E}(X_j \zeta_j) = 0$) est satisfaite on peut utiliser l'estimateur des moindres carrés ordinaires et obtenir ainsi l'estimateur convergent de β :

$$\hat{\beta} = \left(\frac{1}{J} \sum_{j=1}^J X_j X'_j \right)^{-1} \left(\frac{1}{J} \sum_{j=1}^J X_j (\ln(R_j) - \ln(R_0)) \right)$$

.

7 Question 7

7.1 (i)

Dans cette partie on essaie d'expliquer l'impact de la couleur de peau et de la beauté sur la part de vote. Pour cela, nous allons utiliser la formule trouvée à la question précédente en utilisant différent choix de variables pour les caractéristiques des candidats.

Tout d'abord on commence par créer la variable expliquée en effectuant les transformations nécessaires, comme expliqué précédemment.

Dans la suite, on notera $Y_j = \ln(R_j) - \ln(R_0)$

Même si nous allons seulement étudier l'impact de la beauté et de la couleur de peau sur la part de vote, nous allons tout de même inclure d'autres variables par soucis d'endogénéité. Par exemple, dans tous les modèles nous incluons la variable indicatrice si le candidat est un candidat en exercice ("incumbentcandidate") ainsi que la variable du sexe de l'individu ("female"). En effet, ces deux variables sont certainement corrélées avec la beauté du candidat, on aurait donc un problème d'endogénéité en cas d'exclusion de ces variables. Nous avons décidé de ne pas inclure la variable du pourcentage d'indigène dans la circonscription dans les modèles 1,2 et 3 car celle-ci est déjà prise en compte dans le calcul de la moyenne pondérée de la beauté.

De même pour les modèles traitant de la couleur de peau (modèles 4,5 et 6) nous incluons la variable traitant du pourcentage d'indigène dans la circonscription car nous aimerions traiter de l'influence de la

couleur de peau sur les élections à circonscription donné (pour prendre en compte les différents impacts de la couleur de peau dans les circonscriptions où le pourcentage d'indigène diffère).

Nous incluons aussi la variable qui indique l'ordre du candidat sur le bulletin de vote car nous voulons voir si être positionné le plus haut fait vraiment gagner significativement des voix à l'élection.

On résume dans le tableau suivant les 6 modèles que nous nous proposons de tester.

Modèle Numéro :	1	2	3	4	5	6
Beauté Pondérée	•	•	•			•
Couleur de Peau				•	•	•
Part d'indigène				•	•	•
Candidat en place	•	•	•	•	•	•
Femme	•	•	•	•	•	•
Ordre sur bulletin	•	•	•	•	•	•
Beauté Pondérée \times Candidat en place		•				
Beauté Pondérée \times Femme			•			
Couleur peau \times Part d'indigène					•	•
Intercept	•	•	•	•	•	•

Figure 17: Les 6 modèles pour modéliser la variable Y (• indique la présence de cette variable dans le modèle).

Dans toute la suite nous effectuerons des tests de Student pour tester la significativité des variables des modèles. Nous utiliserons les p-valeurs afin de conclure de la significativité.

Modèle 1:

$$Y = \beta_0 + \beta_1 \text{beaute_pond} + \beta_2 \text{incumbentcandidate} + \beta_3 \text{female} + \beta_4 \text{ballotorder} + \epsilon$$

Dans ce premier modèle on s'intéresse à l'impact de la beauté sur la part des votes. Pour cela, nous allons utiliser comme seule variable explicative la note pondérée de beauté que nous avons introduite à la question 3 et que les statistiques descriptives de la question 4 nous incite à utiliser. Le résultat de la régression linéaire obtenue sur cette variable est le suivant:

Table 13: Résultats Modèle 1

	<i>Dependent variable:</i>
	Y
beaute_pond	0.430** (0.168)
incumbentcandidate	1.329*** (0.278)
female	-0.037 (0.273)
ballotorder	-0.148 (0.107)
Constant	9.587*** (0.317)
Observations	79
R ²	0.364
Adjusted R ²	0.330
Residual Std. Error	1.082 (df = 74)
F Statistic	10.590*** (df = 4; 74)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

On a un résultat positif et significatif (au niveau 5%) pour le coefficient devant la beauté. L'appréciation de la beauté semble donc jouer un rôle important pour la part des votes de manière général.

Modèle 2:

$$Y = \beta_0 + \beta_1 \text{beaute_pond} + \beta_2 \text{incumbentcandidate} + \beta_3 \text{female} + \beta_4 \text{ballotorder} + \beta_5 \text{beaute_pond} \times \text{incumbentcandidate} + \epsilon$$

Dans ce modèle, nous allons nous intéresser à l'impact de la beauté en différenciant le cas d'un candidat opposant et sortant. Pour cela on croise la variable de la moyenne pondérée des notes avec la variable incumbentcandidate.

Table 14: Résultats Modèle 2

	<i>Dependent variable:</i>
	Y
beaute_pond	0.550*** (0.189)
incumbentcandidate	1.409*** (0.283)
female	-0.095 (0.275)
ballotorder	-0.142 (0.106)
beaute_pond:incumbentcandidate	-0.522 (0.387)
Constant	9.606*** (0.316)
Observations	79
R ²	0.379
Adjusted R ²	0.337
Residual Std. Error	1.077 (df = 73)
F Statistic	8.929*** (df = 5; 73)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Le coefficient devant la variable du croisement n'est pas significatif (on ne peut pas rejeter l'hypothèse nulle de nullité du coefficient). Ce coefficient représente la différence de l'impact de la beauté entre un candidat sortant et un candidat opposant. On peut donc en déduire que l'impact de la beauté est la même pour les candidats sortants et pour les candidats opposants.

Modèle 3:

$$Y = \beta_0 + \beta_1 \text{beaute_pond} + \beta_2 \text{incumbentcandidate} + \beta_3 \text{female} + \beta_4 \text{ballotorder} + \beta_5 \text{beaute_pond} \times \text{female} + \epsilon$$

Dans ce modèle on aimerait distinguer l'influence de la beauté sur la part de vote d'un candidat homme et d'un candidat femme.

Table 15: Résultats Modèle 3

	<i>Dependent variable:</i>
	Y
beaute_pond	0.532** (0.211)
incumbentcandidate	1.284*** (0.284)
female	0.010 (0.280)
ballotorder	-0.140 (0.108)
beaute_pond:female	-0.281 (0.353)
Constant	9.604*** (0.319)
Observations	79
R ²	0.370
Adjusted R ²	0.326
Residual Std. Error	1.085 (df = 73)
F Statistic	8.556*** (df = 5; 73)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

De même, on observe que le coefficient devant la variable croisée n'est pas significatif (on ne peut pas rejeter l'hypothèse nulle). Ainsi l'influence de la beauté sur la part de vote obtenue est la même pour les hommes que pour les femmes.

Modèle 4:

$$Y = \beta_0 + \beta_1 skin + \beta_2 ab + \beta_3 incumbentcandidate + \beta_4 female + \beta_5 ballotorder + \epsilon$$

Dans ce modèle on s'intéresse à l'impact de la couleur de peau sur les parts de vote. Pour cela on utilise le modèle le plus basique correspondant au modèle où la variable de la couleur de peau n'est pas croisé. On obtient le résultat suivant:

Table 16: Résultats Modèle 4

	<i>Dependent variable:</i>
	Y
skin	-0.011 (0.064)
ab	0.224 (0.539)
incumbentcandidate	1.462*** (0.288)
female	0.196 (0.273)
ballotorder	-0.192* (0.113)
Constant	9.542*** (0.389)
Observations	79
R ²	0.309
Adjusted R ²	0.262
Residual Std. Error	1.136 (df = 73)
F Statistic	6.535*** (df = 5; 73)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Le coefficient devant la variable de la couleur de peau n'est pas significatif (on ne peut pas rejeter l'hypothèse nulle au niveau 5%). La couleur de peau ne semble pas avoir d'impact de façon moyenne dans la population sur la part de vote obtenue.

Modèle 5:

$$Y = \beta_0 + \beta_1 skin + \beta_2 ab + \beta_3 incumbentcandidate + \beta_4 female + \beta_5 ballotorder + \beta_6 skin \times ab + \epsilon$$

Dans ce modèle on étudie l'influence de la couleur de peau sur la part de vote en fonction de la part d'indigène dans la population.

Table 17: Résultats Modèle 5

	<i>Dependent variable:</i>
	Y
skin	-0.254** (0.116)
ab	-1.724* (0.942)
incumbentcandidate	1.436*** (0.278)
female	0.018 (0.274)
ballotorder	-0.163 (0.109)
skin:ab	0.498** (0.201)
Constant	10.331*** (0.492)
Observations	79
R ²	0.364
Adjusted R ²	0.311
Residual Std. Error	1.098 (df = 72)
F Statistic	6.858*** (df = 6; 72)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Le coefficient devant la variable de croisement est positif et significatif au niveau 1%. Ainsi il y a un impact différent de la couleur de peau suivant la circonscription. On remarque également que désormais le coefficient devant la variable de la couleur de peau est significatif au niveau 5% et négatif.

Le fait que le coefficient de la variable de croisement soit positif semble tout à fait logique car dans les circonscriptions où l'on a un haut pourcentage d'indigène, on pourrait penser que la part des votes pour les candidats de couleur de peau plus foncée sera plus importante. Cette intuition est vérifiée par nos résultats. En effet, on comprend que dans les circonscriptions où l'on a un haut pourcentage d'indigène, l'impact de l'indice de la couleur de peau sur la part de vote est plus important que dans les circonscriptions à faible pourcentage d'indigène.

Nous pouvons également remarquer que le coefficient devant la variable de la couleur de peau est désormais significatif. Ce coefficient représente l'impact de la couleur de peau sur la part de votes dans les circonscriptions hypothétiques où il n'y aurait pas d'indigène. En réalité ce coefficient n'a pas du tout la même interprétation que celui dans le modèle précédent.

Modèle 6:

Le problème des modèles précédents est que la variable de la couleur de peau pourrait être corrélée avec la variable de la beauté. Ainsi il se pourrait que les estimateurs des moindres carrés trouvés précédemment soient biaisés. Pour supprimer ce biais il serait utile d'inclure toutes ces deux variables dans la régression. Dans ce dernier modèle, nous incluons la variable croisée de skin et de ab ainsi que celle de la beauté. En réalité ce modèle est un mélange du modèle 1 et du modèle 5 afin de prendre en compte les corrélations des variables. On obtient le résultat de la régression suivant:

$$Y = \beta_0 + \beta_1 \text{beaute_pond} + \beta_2 \text{skin} + \beta_3 \text{ab} + \beta_4 \text{incumbentcandidate} + \beta_5 \text{female} + \beta_6 \text{ballotorder} + \beta_7 \text{skin} \times \text{ab} + \epsilon$$

Table 18: Résultats Modèle 6

	<i>Dependent variable:</i>
	Y
beaute_pond	0.445*** (0.165)
skin	-0.276** (0.111)
ab	-1.610* (0.904)
incumbentcandidate	1.312*** (0.271)
female	-0.245 (0.280)
ballotorder	-0.103 (0.107)
skin:ab	0.498** (0.193)
Constant	10.373*** (0.473)
Observations	79
R ²	0.423
Adjusted R ²	0.366
Residual Std. Error	1.053 (df = 71)
F Statistic	7.431*** (df = 7; 71)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Tout d'abord, on voit que les effets trouvés dans les modèles précédents n'ont pas été modifiés (même significativité et même signe des coefficients). Mais on remarque que certains coefficients ont augmenté et que d'autres ont diminué.

Le coefficient devant la moyenne pondérée de la beauté est plus élevé que celui dans le modèle 1. On aurait donc une corrélation négative générale entre la beauté et la couleur de la peau. Les candidats de couleur plus foncée semblent être notés un petit peu moins bien que les autres. Notons tout de même que le biais est seulement de 0.015 (0.445-0.430).

Pour finir il faut tout de même réaliser que ces résultats tiennent seulement si les variables que nous avons utilisé dans les régressions ne sont pas corrélées avec les autres variables (hypothèse d'exogénéité). Si cette hypothèse n'est pas respectée, alors les coefficients sont biaisés.

7.2 (ii)

La question de l'impact significatif de la beauté et de la couleur de peau a été traitée à la question précédente. Notons que pour étudier la significativité des coefficients nous avons utilisé les p valeurs de la sortie de la régression. En réalité cela correspond à effectuer un test de Student au niveau de significativité de 5 pourcent.

Nous avons trouvé que la moyenne pondérée de la beauté ainsi que la variable croisée de la couleur de peau et de la part d'indigène dans la circonscription ont un impact significatif sur la part des votes. Par contre, la couleur de peau simple n'a pas d'impact significatif sur la part de vote.

Notons également que le R^2 ajusté le plus élevé est pour le modèle 6 (où la beauté et la couleur de peau sont dans la régression). On a, en effet, un R^2 ajusté de 0.366. Le R^2 ajusté est très utile pour savoir si on devrait ou pas ajouter une variable.

En effectuant la régression de la variable expliquée sur toutes les variables que nous avons utilisé dans le modèle 6 à l'exception de la beauté et de la couleur de peau croisée avec la part d'indigène dans la circonscription, on obtient un R^2 ajusté de 0.3123. Ainsi cela conforte notre idée que précédente.

En effectuant la régression de la variable expliquée sur la beauté, la couleur de peau ainsi que la couleur de peau croisée avec ab, on obtient le résultat suivant:

Table 19: Régression de voteshare sur la beauté et la couleur de la peau

<i>Dependent variable:</i>	
	Y
beaute_pond	0.559*** (0.170)
skin	-0.273** (0.119)
ab	-1.714* (1.025)
skin:ab	0.501** (0.213)
Constant	10.428*** (0.412)
Observations	79
R ²	0.204
Adjusted R ²	0.161
Residual Std. Error	1.211 (df = 74)
F Statistic	4.753*** (df = 4; 74)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

On obtient un R^2 de 0.204. On peut donc conclure que dans ce modèle la beauté et la couleur de peau explique environ 20 pourcent de la part des votes obtenus par le candidat.

8 Question 8

8.1 (i)

Comme nous l'avons vu dans les questions précédentes, le score de beauté est un score déterminé par quatre individus représentatifs de la population. De ce fait, l'hypothèse d'exogénéité est difficilement vérifiée. La moyenne pondérée des notes de beauté, qui est censé représenter au mieux la beauté du candidat dépend en réalité de la circonscription du candidat (suivant la part d'indigène de la circonscription).

Ainsi, afin d'éviter d'avoir un estimateur biaisé il est nécessaire d'utiliser des variables instrumentales. Une variable instrumentale doit être corrélée avec la variable exogène et ne doit pas être corrélée avec les résidus.

8.2 (ii)

Les variables correspondants aux mesures faciales du candidat (taille du front, taille du nez, taille du menton) semblent être une bonne idée de variables instrumentales. Elles sont, en effet, très certainement corrélées à la beauté du candidat, et donc aux notes qu'il a obtenu, et elles ne sont sûrement pas corrélées aux résidus étant donné que ce sont des composantes physiologiques du candidat.

Vérifions une à une les hypothèses nécessaires pour considérer ces variables comme des variables instrumentales.

Relevance (corrélation avec la moyenne pondérée des notes):

On normalise les variables de mesure en soustrayant la moyenne et en divisant pour l'écart-type.

Pour tester la corrélation de ces variables avec la beauté, on effectue la régression de la note moyenne pondérée de la beauté sur ces trois variables ainsi que les variables exogènes et on effectue des tests de significativité.

Modèle:

$$beaute_pond = \alpha_0 + \alpha_1 forehead + \alpha_2 chin + \alpha_3 nose + \alpha_4 incumbentcandidate + \alpha_5 ballotorder + \alpha_6 female + \epsilon$$

Table 20: Choix de l'instrument pour la beauté

	<i>Dependent variable:</i>
	beaute_pond
forehead	−0.018** (0.009)
chin	0.015** (0.007)
nose	−0.008 (0.009)
incumbentcandidate	0.260 (0.185)
ballotorder	−0.134* (0.071)
female	0.524*** (0.185)
Constant	0.102 (0.359)
Observations	79
R ²	0.269
Adjusted R ²	0.208
Residual Std. Error	0.723 (df = 72)
F Statistic	4.420*** (df = 6; 72)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

On remarque que le coefficient de la mesure du nez ne semble pas significatif dans la régression (test de Student). Ainsi, la condition de rang pour cet instrument n'est pas vérifié.

Exclusion restriction (non-corrélation avec les résidus):

Il est possible d'effectuer un test de Sargan dans notre cas car nous avons deux instruments pour la beauté (le menton et le front). Ce test permettra de conclure sur la non-corrélation des instruments avec

les résidus. Notons que nous effectuerons ce test sur le premier modèle. Le test de Sargan se déroule en trois phases:

1) On récupère le résidu \hat{u} issu du modèle 2SLS.

2) On régresse \hat{u} sur une constante et sur toutes les variables exogènes et on obtient $R_{\hat{u}}^2$.

3) Pour tester l'hypothèse nulle $H_0: E(z_1u) = 0$ et $E(z_2u) = 0$ où z_1 et z_2 sont les deux instruments de la variable de beauté, on utilise la statistique de test $nR_{\hat{u}}^2$. Sous H_0 $nR_{\hat{u}}^2 \sim \chi^2$ où χ^2 est une loi chi 2 à 1 degré de liberté.

On effectue ces trois étapes sur R et on obtient un résultat pour l'étape 3 qui est le suivant:

$$\hat{u}_i = \alpha_0 + \alpha_1 \text{incumbentcandidate}_i + \alpha_2 \text{ballotorder}_i + \alpha_3 \text{female}_i + \epsilon_i$$

Table 21: Test de Sargan

	<i>Dependent variable:</i>
	res
incumbentcandidate	−0.000 (0.280)
ballotorder	0.000 (0.108)
female	0.000 (0.263)
Constant	−0.000 (0.325)
Observations	79
R ²	0.000
Adjusted R ²	−0.040
Residual Std. Error	1.109 (df = 75)
F Statistic	0.000 (df = 3; 75)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

On obtient $R_{\hat{u}}^2 = 0.000$, c'est à dire que $nR_{\hat{u}}^2 = 0.000$. Notons que ce résultat est arrondi à la troisième décimale est donc pas vraiment égal à zéro. Or on voit sur la table du χ^2 que le quantile pour un test à

5 pourcent à 1 degré de liberté se trouve à 0.004 qui est très largement supérieur à notre valeur. Ainsi on accepte fortement l'hypothèse H_0 .

Le test de Sargan est concluant sur l'exogénéité de nos deux instruments.

Conclusion: On garde donc comme instrument de la variable de la beauté: la variable de la mesure du menton et la variable de la mesure du front.

8.3 (iii) et (iv)

Dans cette section nous allons refaire les modèles 1,2,3 et 6 de la question 7 en utilisant la mesure du front et la mesure du menton comme variables instrumentales de la moyenne du score pondéré de beauté.

Pour cela nous devons d'abord effectuer une première régression de première étape afin d'obtenir un estimateur de la beauté en fonction des variables instrumentales et des variables exogènes.

$$\widehat{beaute_pond} = \hat{\alpha}_0 + \hat{\alpha}_1 forehead + \hat{\alpha}_2 chin + \hat{\alpha}_3 incumbentcandidate + \hat{\alpha}_4 ballotorder + \hat{\alpha}_5 female$$

Une fois cette variable récupéré, nous effectuons les régressions des modèles 1,2,3 et 6 mais en remplaçant la beauté par cette variable. Nous obtenons les résultats suivants.

Modèle 1 avec variables instrumentales:

$$Y_j = \gamma_0 + \gamma_1 \widehat{beaute_pond}_j + \gamma_2 incumbentcandidate_j + \gamma_3 female_j + \gamma_4 ballotorder_j + \epsilon_j$$

Table 22: Modèle 1 instrumental

	<i>Dependent variable:</i>
	Y
instruments	0.772 (0.588)
incumbentcandidate	1.229*** (0.330)
female	-0.227 (0.421)
ballotorder	-0.107 (0.129)
Constant	9.594*** (0.328)
Observations	79
R ²	0.323
Adjusted R ²	0.287
Residual Std. Error	1.117 (df = 74)
F Statistic	8.841*** (df = 4; 74)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Tout d'abord, on remarque que le coefficient devant la variable instrumentale de la beauté n'est plus significatif (on ne peut pas rejeter l'hypothèse nulle). De plus on remarque que ce coefficient a augmenté. L'hypothèse d'endogénéité de la variable de beauté du modèle de la question 7 est ainsi vérifié par ce modèle.

Modèle 2 avec variables instrumentales:

$$Y_j = \gamma_0 + \gamma_1 \widehat{beaute_pond}_j + \gamma_2 incumbentcandidate_j + \gamma_3 female_j + \gamma_4 ballotorder_j + \gamma_5 \widehat{beaute_pond}_j \times incumbentcandidate_j + \epsilon_j$$

Table 23: Modèle 2 instrumental

	Dependent variable:
	Y
instruments	0.836 (0.593)
incumbentcandidate	1.369*** (0.368)
female	-0.156 (0.429)
ballotorder	-0.115 (0.129)
instruments:incumbentcandidate	-0.716 (0.833)
Constant	9.591*** (0.328)
Observations	79
R ²	0.330
Adjusted R ²	0.284
Residual Std. Error	1.119 (df = 73)
F Statistic	7.195*** (df = 5; 73)
Note:	*p<0.1; **p<0.05; ***p<0.01

Dans ce modèle on obtient que, comme à la question 7 la variable croisée n'est pas significative (on ne peut pas rejeter l'hypothèse nulle). Par contre on voit que la variable de la beauté instrumenté ne semble toujours pas significative (comme au modèle 1).

Modèle 3 avec variables instrumentales:

$$Y_j = \gamma_0 + \gamma_1 \widehat{beaute_pond}_j + \gamma_2 incumbentcandidate_j + \gamma_3 female_j + \gamma_4 ballotorder_j + \gamma_5 \widehat{beaute_pond}_j \times female_j + \epsilon_j$$

Table 24: Modèle 3 instrumental

	Dependent variable:
	Y
instruments	1.514** (0.737)
incumbentcandidate	1.151*** (0.330)
female	-0.121 (0.421)
ballotorder	-0.072 (0.129)
instruments:female	-1.437 (0.877)
Constant	9.698*** (0.330)
Observations	79
R ²	0.347
Adjusted R ²	0.303
Residual Std. Error	1.104 (df = 73)
F Statistic	7.770*** (df = 5; 73)
Note:	*p<0.1; **p<0.05; ***p<0.01

Dans ce modèle on obtient des résultats quelque peu différents par rapport à la question précédente. Tout d'abord le coefficient devant la variable instrumentale est beaucoup plus élevé que celui de la question précédente (0.532 pour 1.514 maintenant). Par contre, comme pour le modèle de la question 7 la variable de croisement n'est toujours pas significative (on ne peut pas rejeter l'hypothèse nulle). D'après ce modèle, la beauté aurait impact significatif (au niveau 5%) et équivalent comme pour les hommes que pour les femmes.

Modèle 6 avec variables instrumentales:

$$Y_j = \gamma_0 + \gamma_1 \widehat{note_{pond}_j} + \gamma_2 skin_j + \gamma_3 ab_j + \gamma_4 female_j + \gamma_5 incumbentcandidate_j + \gamma_6 ballotorder_j + \gamma_7 skin_j \times ab_j + \epsilon_j$$

Table 25: Modèle 6 instrumental

	<i>Dependent variable:</i>
	Y
instruments	0.544 (0.595)
skin	-0.241** (0.117)
ab	-1.520 (0.969)
female	-0.274 (0.421)
incumbentcandidate	1.282*** (0.325)
ballotorder	-0.099 (0.130)
skin:ab	0.461** (0.205)
Constant	10.280*** (0.496)
Observations	79
R ²	0.371
Adjusted R ²	0.309
Residual Std. Error	1.099 (df = 71)
F Statistic	5.984*** (df = 7; 71)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Tout d'abord, nous pouvons remarquer que la beauté (ou son instrument) n'est plus significatif dans cette régression (on ne peut pas rejeter l'hypothèse nulle). Par contre les variables contenant la couleur de peau sont toujours significatives au niveau 5% (comme dans le modèle 6 de la question précédente).

Le fait que la variable instrumentale n'est plus significative est certainement dû au fait que l'estimateur du coefficient, dans le cas de la régression avec l'instrument, a une variance très élevée dû à une faible corrélation entre la beauté et les variables instrumentales. Du point de vu théorique, nous avons les résultats suivants:

$$\widehat{Avar}(\hat{\gamma}_6^{IV}) = \frac{\hat{\sigma}^2}{n\hat{\sigma}_x^2\hat{\rho}_{x,z}^2}$$

Où $x = \text{beaute_pond}$ et $z = \widehat{\text{beaute_pond}}$

Cette variance empirique augmente lorsque la beauté est faiblement corrélée avec son estimateur obtenu à l'aide des variables instrumentales. Or dans la régression de la beauté sur les deux instruments ainsi que les variables exogènes on avait obtenu $R^2 = 0.269$. Ce résultat étant très faible, on en déduit que ces variables n'explique pas très bien la beauté. La corrélation entre la beauté et les instruments n'est donc certainement pas très élevée. Cela expliquerait donc la forte variance de l'estimateur et donc sa non-significativité dans les modèles.

Conclusion:

Les résultats que nous avons obtenu en utilisant les variables instrumentales changent nos résultats. Nous obtenons, cette fois-ci des résultats non-significatif de la beauté sur la part de vote.

9 Question 9

9.1 (i)

On souhaite étudier la variable *winner* qui est une variable binaire prenant 0 ou 1 comme valeur.

On propose de modéliser cette variable en termes de variable latente.

En notant Y notre variable cible, on suppose qu'il existe une variable continue Y^* suivant un modèle linéaire $Y^* = X'\beta_0 + \epsilon$ où $-\epsilon$ est indépendante de $X = (X_1, X_2, \dots, X_k)'$ (avec $X_1 = 1$) et ayant pour fonction de répartition F telle que $Y = \mathbb{1}_{Y^* > 0}$.

On observe que l'on a donc :

$$\mathbb{E}(Y|X) = \mathbb{P}(Y = 1|X) = \mathbb{P}(Y^* > 0|X) = \mathbb{P}(X'\beta_0 + \epsilon > 0|X) = \mathbb{P}(X'\beta_0 > -\epsilon|X) = F(X'\beta_0)$$

Dans cette catégorie de modèles binaires on a donc l'expression suivante de l'effet marginal pour x appartenant à \mathbb{R}^k avec $f = F'$:

$$\frac{\partial \mathbb{E}(Y|X = x)}{\partial x_j} = f(x'\beta_0)\beta_{0j} \quad \forall j \in \llbracket 1, k \rrbracket$$

Ainsi on ne peut plus interpréter quantitativement l'effet marginal de X_j pour tous j mais on peut comparer l'influence des effets marginaux en observant que pour tous $(i, j) \in \llbracket 1, k \rrbracket^2$ on a

$$\frac{\beta_{0i}}{\beta_{0j}} = \frac{\partial \mathbb{E}(Y|X = x)/\partial x_i}{\partial \mathbb{E}(Y|X = x)/\partial x_j}$$

Pour nos six spécifications différentes de la variables *winner* nous devons choisir une fonction F et les variables explicatives X .

Pour l'ensemble de nos modèles, nous choisirons $F(x) = \frac{1}{1 + \exp(-x)}$ qui correspond à un modèle logit.

Ce choix, bien qu'arbitraire, repose sur trois arguments. Premièrement le modèle logit possède une solution unique au calcul de maximum de vraisemblance conditionnelle. On a donc

$$\hat{\beta} = \operatorname{argmax}_{\beta \in \mathbb{R}^k} \mathbb{L}_n(Y|X; \beta)$$

où

$$\mathbb{L}_n(Y|X; \beta) = \prod_{i=1}^n F(X'_i \beta)^{Y_i} (1 - F(X'_i \beta))^{1-Y_i}$$

Deuxièmement le modèle de logit possède une propriété intéressante des odds-ratios pour les variables explicatives binaire. En effet l'odd-ratio correspondant à une variable binaire X_j valant 0 ou 1 est tel que

$$e^{\beta_{0j}} = \frac{\mathbb{P}(Y = 1|X_{-j} = x_{-j}, X_j = 1)/\mathbb{P}(Y = 0|X_{-j} = x_{-j}, X_j = 1)}{\mathbb{P}(Y = 1|X_{-j} = x_{-j}, X_j = 0)/\mathbb{P}(Y = 0|X_{-j} = x_{-j}, X_j = 0)}$$

ce qui nous offre une certaine interprétation qualitative possible.

Troisièmement des méthodes sont implémentées en R dans le package *glm* pour nous permettre de tester facilement le modèle logit.

On résume dans le tableau suivant les 6 modèles que nous nous proposerons de tester dans la question suivante.

Modèle Numéro :	1	2	3	4	5	6
Beauté Pondérée	•	•	•			•
Couleur de Peau				•	•	•
Part d'indigène				•	•	•
Candidat en place	•	•	•	•	•	•
Femme	•	•	•	•	•	•
Ordre sur bulletin	•	•	•	•	•	•
Beauté Pondérée \times Candidat en place		•				
Beauté Pondérée \times Femme			•			
Couleur peau \times Part d'indigène					•	•
Intercept	•	•	•	•	•	•

Figure 18: 6 modèles logit pour modéliser la variable *winner* (• indique la présence de cette variable dans le modèle).

Le lecteur observera que ces 6 modèles ont un choix de variables explicatives similaire à celui de la question 7. Il pourra donc se référer à la question 7 pour voir une justification de ce choix de variables.

9.2 (ii)

On teste ici les 6 modèles de logit formulés à la questions précédente pour voir si la beauté et la couleur de peau semble significatives dans le fait de gagner une élection.

Pour chaque modèle on affiche dans un premier temps les résultats puis on les commente.

Modèle Logit 1:

Table 26: Modèle Logit 1

	<i>Dependent variable:</i>
	winner
beaute_pond	1.021** (0.469)
incumbentcandidate	3.540*** (0.741)
female	-0.152 (0.717)
ballotorder	-0.165 (0.314)
Constant	-1.711* (0.877)
Observations	79
Log Likelihood	-28.566
Akaike Inf. Crit.	67.132
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

On observe que la beauté (beaute_pond) est significative au taux 5 pourcents et a un effet positif sur le fait de gagner une élection. Le fait d'être un candidat en exercice est significatif au taux 1 pourcent. De plus cette dernière variable est binaire donc on peut interpréter son odd-ratio. Ici $e^{\hat{\beta}_{incumbentcandidate}}$ est largement supérieur à 1 donc l'événement "gagner l'élection" est plus fréquent dans le cas où on est un candidat en exercice que si on ne l'est pas.

On observe de plus que les effets marginaux de la variable qui indique si le candidat est en exercice est plus forte que ceux de la variable du score de beauté pondéré. En effet le rapport des coefficient du premier sur le second est supérieur à 1.

On ne rejette pas les hypothèses que les variables être une femme et ordre sur le bulletin de vote ne soient pas significatifs.

Modèle Logit 2:

Table 27: Modèle Logit 2

	<i>Dependent variable:</i>
	winner
beaute_pond	1.297** (0.568)
incumbentcandidate	3.686*** (0.766)
female	-0.247 (0.724)
ballotorder	-0.147 (0.319)
beaute_pond:incumbentcandidate	-1.014 (1.019)
Constant	-1.828** (0.916)
Observations	79
Log Likelihood	-28.090
Akaike Inf. Crit.	68.179
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

On a ici qualitativement des résultats similaires à ceux du modèle précédent.

On apprend de plus que l'on ne rejette pas la nullité du coefficient devant la variable du croisement entre le score de beauté pondéré et le fait d'être un candidat en exercice.

Modèle Logit 3:

Table 28: Modèle Logit 3

	<i>Dependent variable:</i>
	winner
beaute_pond	1.150* (0.628)
incumbentcandidate	3.505*** (0.749)
female	-0.066 (0.763)
ballotorder	-0.155 (0.316)
beaute_pond:female	-0.291 (0.924)
Constant	-1.725* (0.884)
Observations	79
Log Likelihood	-28.516
Akaike Inf. Crit.	69.032
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Otre le fait qu'on rejette la nullité du coefficient de la variable beauté au niveau 10 pourcents et celui du fait d'être en exercice au niveau 1 pourcent, on a concernant ces variables des considérations semblables à celles des modèles 1 et 2.

On ne rejette pas la nullité du coefficient devant la variable du croisement entre le score de beauté pondéré et le fait d'être une femme.

Modèle Logit 4:

Table 29: Modèle Logit 4

	<i>Dependent variable:</i>
	winner
skin	0.208 (0.166)
ab	-0.077 (1.364)
incumbentcandidate	3.612*** (0.734)
female	0.505 (0.704)
ballotorder	-0.305 (0.300)
Constant	-2.224** (1.026)
Observations	79
Log Likelihood	-30.179
Akaike Inf. Crit.	72.359
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

On a souhaité ici tester la qualité de l'impact de la couleur de peau dans le fait de gagner une élection ou pas.

Il se trouve qu'on ne rejette pas la nullité des coefficient des variable couleur de peau (skin) , part d'aborigènes (ab), être une femme (female) et l'ordre sur le bulletin de vote (ballotorder).

On ne rejette que la nullité du coefficient devant la variable exprimant le fait d'être en exercice ou pas. Nos commentaires qualitatifs concernant cette variable sont similaires à ceux que nous avons formulé pour le premier modèle.

Modèle Logit 5:

Table 30: Modèle Logit 5

	<i>Dependent variable:</i>
	winner
skin	−0.153 (0.300)
ab	−3.093 (2.623)
incumbentcandidate	3.671*** (0.744)
female	0.253 (0.748)
ballotorder	−0.320 (0.320)
skin:ab	0.722 (0.514)
Constant	−0.944 (1.307)
Observations	79
Log Likelihood	−29.160
Akaike Inf. Crit.	72.319
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Hormis le fait que l'on ne rejette pas la nullité du coefficient de la variable de la couleur de peau croisée avec la part d'aborigène dans la circonscription , on a ici les mêmes résultats que pour le modèle 5.

Modèle Logit 6:

Table 31: Modèle Logit 6

	<i>Dependent variable:</i>
	winner
beaute_pond	1.069** (0.510)
skin	-0.255 (0.338)
ab	-3.177 (2.629)
incumbentcandidate	3.768*** (0.791)
female	-0.287 (0.785)
ballotorder	-0.238 (0.337)
skin:ab	0.893 (0.574)
Constant	-0.902 (1.332)
Observations	79
Log Likelihood	-26.634
Akaike Inf. Crit.	69.269
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Dans ce modèle 6 on retrouve les résultats qualitatifs des modèles précédents. On rejette respectivement les coefficients de beauté au niveau 5 pourcents et le fait d'être en exercice au niveau 1 pourcent. On ne rejette pas la nullité des coefficients de la couleur de la peau, de la proportion d'aborigène, d'être une femme, de l'ordre sur le bulletin de vote et la variable croisée entre la couleur de la peau et la proportion d'aborigènes.

Conclusion des modèles:

Ici on a testé dans différents modèles l'impact qualitatif et relatif de plusieurs variables sur le fait de gagner ou non une élection.

Nos résultats semblent nous indiquer le fait que la beauté a un impacte positif sur le fait de gagner une élection. Il en va de même pour le fait d'être un candidat en exercice.

Nos résultats semblent aussi nous indiquer ici que la couleur de peau n'influencerait pas le fait de gagner ou non l'élection, même mise en perspective avec la part d'aborigène dans la circonscription.

10 Question 10

Rappelons la méthode formulée par Rivers et Vuong (1989).

On considère le modèle

$$Y = \mathbb{1}_{X_1\beta_1 + X_2\beta_2 + \epsilon > 0}$$

où $X_1 \in \mathbb{R}^k$ est exogène ($X_1 \perp \epsilon$) mais $X_2 \in \mathbb{R}$, supposé ici continue, ne l'est pas.

On suppose de plus que $\epsilon \sim \mathcal{N}(0, 1)$ donc on se retrouve avec un modèle probit.

On suppose de plus avoir à notre disposition un instrument Z qui joue sur X_2 mais pas sur Y . En d'autres termes on a :

$$X_2 = X_1\gamma_1 + Z\gamma_2 + \eta$$

Où $(X_1, Z) \perp (\epsilon, \eta)$, $\epsilon = \rho\eta + \zeta$ et

$$\zeta|\eta \sim \mathcal{N}(0, \sigma^2)$$

La méthode procède en deux étapes. Dans un premier temps on estime $\hat{\eta}$ dans l'équation $X_2 = X_1\gamma_1 + Z\gamma_2 + \eta$ en utilisant $\hat{\eta} = X_2 - X_1\hat{\gamma}_1 - Z\hat{\gamma}_2$. Dans un deuxième temps, on utilise donc $\hat{\eta}$ pour estimer le probit

$$Y = \mathbb{1}_{X_1\beta_1 + X_2\beta_2 + \rho\hat{\eta} + \zeta > 0}$$

On a ainsi un test d'exogénéité en testant la significativité de coefficient devant $\hat{\eta}$ dans le modèle probit par un test de Student.

Dans notre sujet nous aurons X_1 qui représente les variables exogènes. Nous aurons également X_2 qui sera le score de beauté pondéré. Nous pouvons choisir deux instruments pour la variable X_2 , la mesure du menton ou bien la mesure du front. Nous choisirons dans cette question la mesure du menton comme

variable instrumentale Z de X_2 . Finalement, nous aurons Y qui sera la variable qui dit si le candidat a gagné les élections ou non ("winner")

10.1 (ii)

Dans cette question nous appliquons la méthode de River et Vuong aux 6 modèles que nous avons sélectionné dans les questions précédentes.

Modèle 1 probit:

X_1 prend en compte "candidate en place" (incumbentcandidate), "femme" (female), "ordre sur le bulletin" (ballotorder).

Dans un premier temps nous estimons $\hat{\eta}$ en régressant X_2 sur X_1 et Z en utilisant R.

Ensuite, nous appliquons un modèle probit sur R et nous obtenons le résultat suivant:

$$P(\text{winner} = 1/X) = \Phi(\beta_{10} + \beta_{11}\text{ballotorder} + \beta_{12}\text{female} + \beta_{13}\text{incumbentcandidate} + \beta_2\text{beaute_pond} + \rho\hat{\eta})$$

Où Φ est la fonction de répartition de la loi normale centré réduite.

Table 32: Modèle 1 probit

	<i>Dependent variable:</i>
	winner
ballotorder	−0.349 (0.325)
female	1.325 (1.427)
incumbentcandidate	2.824*** (0.874)
beaute_pond	−1.817 (2.398)
eta	2.474 (2.473)
Constant	−1.210** (0.519)
Observations	79
Log Likelihood	−27.953
Akaike Inf. Crit.	67.906
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

On remarque que la beauté n'est pas significative dans ce modèle. Seule la variable "candidat en place" (incumbentcandidate) est significative dans le premier modèle.

Modèle 2 probit:

X_1 prend en compte "candidate en place" (incumbentcandidate), "femme" (female), "ordre sur le bulletin" (ballotorder). On rajoute la variable de croisement entre "candidat en place" et la beauté. Nous appliquons un modèle probit sur R et nous obtenons le résultat suivant:

$$P(\text{winner} = 1/X) = \Phi(\beta_{10} + \beta_{11}\text{ballotorder} + \beta_{12}\text{female} + \beta_{13}\text{incumbentcandidate} + \beta_2\text{beaute_pond} + \rho\hat{\eta} + \beta_3\text{beaute_pond} \times \text{incumbentcandidate})$$

Table 33: Modèle 2 probit

	<i>Dependent variable:</i>
	winner
ballotorder	−0.312 (0.329)
female	1.137 (1.455)
incumbentcandidate	2.832*** (0.877)
beaute_pond	−1.447 (2.454)
eta	2.232 (2.504)
incumbentcandidate:beaute_pond	−0.480 (0.567)
Constant	−1.239** (0.531)
Observations	79
Log Likelihood	−27.631
Akaike Inf. Crit.	69.263
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Dans ce modèle, on obtient également la non-significativité de la variable de la beauté mais également de la variable de croisement entre la beauté et "candidat en place".

Modèle 3 probit:
On considère cette fois ci :

$$P(\text{winner} = 1/X) = \Phi(\beta_{10} + \beta_{11}\text{ballotorder} + \beta_{12}\text{female} + \beta_{13}\text{incumbentcandidate} + \beta_2\text{beaute_pond} + \rho\hat{\eta} + \beta_3\text{beaute_pond} \times \text{female})$$

Table 34: Modèle 3 probit

	<i>Dependent variable:</i>
	winner
ballotorder	−0.341 (0.328)
female	1.332 (1.436)
incumbentcandidate	2.795*** (0.882)
beaute_pond	−1.717 (2.430)
eta	2.427 (2.485)
female:beaute_pond	−0.120 (0.529)
Constant	−1.211** (0.521)
Observations	79
Log Likelihood	−27.928
Akaike Inf. Crit.	69.856
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

De même on trouve que la beauté n'est pas significative.

Pour le modèle 4 et 5 la variable X_2 correspond non plus à la beauté mais plutôt à la couleur de la peau. On aimerait donc instrumentaliser la couleur de la peau. Pour cela, nous avons vu dans le tableau des corrélations que la couleur de peau semble corrélé avec les variables physiologiques (la mesure du front, du menton et du nez). Ceci est un indice pour les utiliser comme instrument. Afin de trouver l'instrument

de la couleur de la peau nous effectuons la régression linéaire de la couleur de la peau sur ces trois mesures ainsi que sur les variables exogènes des modèles 4 et 5.

$$skin = \gamma_0 + \gamma_1 forehead + \gamma_2 nose + \gamma_3 chin + \gamma_4 ballotorder + \gamma_5 female + \gamma_6 incumbentcandidate + \gamma_7 ab + \epsilon$$

Table 35: Choix de l'instrument de la couleur de la peau

	<i>Dependent variable:</i>
	skin
forehead	0.031 (0.019)
nose	0.039* (0.020)
chin	0.057*** (0.015)
ballotorder	0.272* (0.152)
female	0.294 (0.399)
incumbentcandidate	0.520 (0.397)
ab	2.322*** (0.695)
Constant	-2.411*** (0.776)
Observations	79
R ²	0.564
Adjusted R ²	0.521
Residual Std. Error	1.555 (df = 71)
F Statistic	13.096*** (df = 7; 71)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

On remarque que la mesure du nez et du menton sont significatif dans la régression. De plus on voit que le coefficient le plus élevé mais également le plus significatif et celui devant la mesure du menton. On choisit donc $Z = chin$ comme instrument.

Modèle 4 probit:

Tout d'abord, on détermine $\hat{\eta}_2$ comme le résidu de la régression suivante:

$$skin = \gamma_0 + \gamma_1 incumbentcandidate + \gamma_2 female + \gamma_3 ballotorder + \gamma_4 ab + \gamma_5 chin + \eta_2$$

Ensuite nous pouvons appliquer le modèle probit suivant:

$$P(winner = 1/X) = \Phi(\beta_{10} + \beta_{11} ballotorder + \beta_{12} female + \beta_{13} incumbentcandidate + \beta_{14} ab + \beta_2 skin + \rho \hat{\eta}_2)$$

Table 36: Modèle 4 probit

	<i>Dependent variable:</i>
	winner
ballotorder	−0.166 (0.169)
female	0.021 (0.412)
incumbentcandidate	2.262*** (0.419)
ab	0.638 (0.863)
skin	−0.095 (0.155)
eta2	0.338 (0.210)
Constant	−0.760 (0.594)
Observations	79
Log Likelihood	−28.638
Akaike Inf. Crit.	71.277
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Dans ce modèle, la couleur de peau n'est pas significative.

Modèle 5 probit:

$$P(\text{winner} = 1/X) = \Phi(\beta_{10} + \beta_{11}\text{ballotorder} + \beta_{12}\text{female} + \beta_{13}\text{incumbentcandidate} + \beta_{14}\text{ab} + \beta_2\text{skin} + \rho\hat{\eta}_2 + \beta_3\text{ab} \times \text{skin})$$

Table 37: Modèle 5 probit

	<i>Dependent variable:</i>
	winner
ballotorder	−0.169 (0.182)
female	−0.238 (0.459)
incumbentcandidate	2.376*** (0.449)
ab	−1.421 (1.520)
skin	−0.418* (0.252)
eta2	0.421* (0.235)
ab:skin	0.555* (0.318)
Constant	0.246 (0.831)
Observations	79
Log Likelihood	−27.080
Akaike Inf. Crit.	70.160
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

On obtient un coefficient significatif et positif de la variable de croisement entre la part d'indigène dans la circonscription et la couleur de la peau. Pour comprendre l'impact de la couleur de la peau sur la probabilité de gagner l'élection il faut bien comprendre que le coefficient devant la couleur de la peau

dépend de la part d'indigène dans la circonscription. On comprend que dans la circonscription avec beaucoup d'indigènes, la couleur de la peau influence positivement la probabilité de gagner l'élection. Mais à partir de "ab" tel que: $ab \times 0.555 - 0.418 = 0$ c'est à dire $ab = 0.753$ l'influence de la couleur de peau est négative. Ainsi, on comprend que lorsque la proportion d'indigène est supérieur à 0.753, la couleur de la peau a un effet positif sur la probabilité de gagner et lorsque la proportion d'indigène est inférieur à 0.753 la couleur de la peau à un effet négatif.

Nous ne pouvons pas effectuer le modèle 6 avec cette méthode probit puisque $X_2 \in \mathbb{R}$. On ne peut donc pas instrumenter deux variable en même temps.

Conclusion de l'approche probit:

Avec cette méthode nous pouvons seulement conclure que la couleur de la peau influence positivement la probabilité du candidat de remporter l'élection dans les circonscriptions où la proportion d'indigène est forte. Par contre, nous ne pouvons pas conclure sur l'impact de la beauté sur cette probabilité.

En réalité nous remarquons que l'utilisation d'une variable instrumentale engendre la non-significativité de la beauté dans nos modèle (dans les modèles de cette question et de la question 8). Les trois variables de mesure physiologique ne sont surement pas de bons instruments pour la beauté à cause de leur faible corrélation avec la variable de la beauté pondérée. On remarque en effet une variance empirique élevée pour le coefficient de la beauté lorsque l'on utilise ces instruments. Par contre utiliser l'instrument du menton pour la couleur de la peau semble être une bonne idée puisque nous avons obtenu $R^2 = 0.564$ dans la régression de la couleur de la peau sur les trois mesures physiologiques et les variables exogènes (table 35). Ce R^2 est beaucoup plus élevé que celui que l'on avait obtenu pour le choix des instruments pour la beauté. Il semblerait donc que la corrélation de la mesure du menton soit plus important avec la couleur de la peau qu'avec la beauté.

11 Question 11

11.1 (i)

La variable objectif Y , ici *winner2*, est une variable discrète qui prend les valeurs 0,1 ou 2. C'est une variable discrète ordonnée dans le sens où elle représente le nombre de victoire aux deux dernière élections.

On suppose que l'on a une variable latente Y^* telle que

$$Y^* = X'\beta_0 + \sigma_0\epsilon$$

où $\epsilon \perp X$ et ϵ est de loi connue F .

On suppose qu'on a

$$Y = \sum_{k=0}^2 k \times \mathbb{1}_{\alpha_{0,k} < Y^* < \alpha_{0,k+1}}$$

où l'on pose par convention $\alpha_{0,0} = -\infty$ et $\alpha_{0,3} = +\infty$. On a donc $\alpha_{0,1}$, $\alpha_{0,2}$ et β inconnus.

En effet, nos seuils ne sont pas connus car nous n'avons, entre autre, pas de données sur les élections

précédentes.

Dans un souci d'identification de nos modèles, on choisit de fixer le seuil $\alpha_{0,1}$ à 0 et σ_0 à 1.

On procède ici aussi par maximum de vraisemblance. On maximise donc

$$l_n(Y|X, \alpha, \beta) = \sum_{i=1}^n \ln[F(\alpha_{Y_i+1} - X'_i\beta) - F(\alpha_{Y_i} - X'_i\beta)]$$

qu'on maximise en (α, β)

11.2 (ii)

On propose d'utiliser comme variables explicatives le score de beauté pondérée (ici *beaute_pond*), la couleur de la peau (*skin*), la proportion d'aborigènes dans la circonscription (*ab*), le fait d'être une femme ou non (*female*) puis la couleur de peau croisée avec la proportion d'aborigènes et la beauté pondérée croisée avec le fait d'être une femme ou pas.

Le lecteur observera que ce choix de variable est semblable à celui que nous avons fait pour le modèle logit 6 pour la variable *winner* en dehors du fait que nous avons exclus les variables d'ordre d'entrée sur le bulletin de vote et la variable d'indicatrice égale à 1 si le candidat est sortant. Notre variable cible ne concernant que les deux dernières élections, nous avons exclus la variable d'ordre d'entrée sur le bulletin de vote car celui-ci survient après ces événements cibles et est censé de plus ne pas être corrélé à cette donnée (ordre aléatoire). On choisit de plus de ne pas inclure la variable d'indicatrice égale à 1 si le candidat est sortant car on ne veut pas distinguer les candidats par rapport à cette donnée. On ne veut pas entre autre comparer les effets en sachant si on a gagné la dernière élection ou pas.

Pour le formuler autrement, on se place dans une situation semblable à si on devait prédire et expliquer le nombre de victoires aux deux prochaines élections où on a fait l'hypothèse que tous les candidats actuels étaient susceptibles d'être candidat aux deux dernières élections. Nous faisons entre autre l'hypothèse que la beauté, la couleur de peau et la proportion d'aborigène dans les circonscriptions ne varient pas d'une élection à l'autre.

Les résultats de notre modèle sont les suivants :

Table 38: Modèle Logit Multinomiale Ordonné pour *winner2*

	<i>Dependent variable:</i>
	winner2
beaute_pond	1.381*** (0.452)
skin	-0.095 (0.225)
ab	-1.862 (1.939)
female	-0.055 (0.565)
skin:ab	0.453 (0.392)
beaute_pond:female	-1.555** (0.749)
Observations	79
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Il se trouve que la beauté pondérée et la variable croisée entre la beauté pondérée et le fait d'être une femme sont significatifs respectivement aux niveaux 1 pourcent et 5 pourcents.

De part la valeur du coefficient croisé entre la beauté pondérée et le fait d'être une femme, on voit que le score de beauté a un impact positif sur les candidats masculins et un impact négatif sur les femmes ($1.381 - 1.555 = -0.2369$). Mais on voit que l'effet de la beauté est relativement plus fort pour les hommes (1.381) que pour les femmes (-0.2369).

Conclusion du logit multinomial ordonné:

Ici on a testé dans un modèle de logit multinomial ordonné l'impact qualitatif et relatif de plusieurs variables sur le nombre d'élections qu'on gagnerait parmi deux élections.

Nos résultats semblent nous indiquer le fait que la beauté a un impacte positif sur le fait de gagner plusieurs élections pour un homme et négatif pour une femme.

Nos résultats semblent aussi nous indiquer ici que la couleur de peau n'influencerait pas le fait de gagner plusieurs élections ou non, même mise en perspective avec la part d'aborigène dans la circonscription.

11.3 (iii)

Selon notre modèle des questions précédentes on a :

$$\mathbb{P}(\text{winner2} \leq i|X) = \mathbb{P}(Y^* \leq \alpha_{0,i+1}|X) = F(\alpha_{0,i+1} - X'\beta_0) \quad \forall i \in \{0, 1\}$$

et $\mathbb{P}(\text{winner2} \leq 2|X) = 1$ où $F(x) = \frac{1}{1+e^{-x}}$ pour un modèle logit.

Notre modèle est donc un cas particulier du modèle :

$$\mathbb{P}(\text{winner2} \leq i|X) = F(\alpha_{0,i+1} - X'\beta_{0,i}) \quad \forall i \in \{0, 1\}$$

avec $\beta_{0,0} = \beta_{0,1}$ et $F(x) = \frac{1}{1+e^{-x}}$.

L'hypothèse que $\beta_{0,0} = \beta_{0,1}$ peut donc paraître restrictive et on souhaite la tester comme hypothèse nulle.

On va tester cela pour les vecteurs (test "Omnibus") et pour chaque composante par un test de Brant des pentes parallèles qui utilise nottament un test du score.

Le test s'interprète comme l'existence d'une unique variable latente ou pas.

Table 39: Test de Brant pour les pentes parrallèles

	X2	df	probability
Omnibus	3.200	4	0.525
beaute_pond	2.719	1	0.099
skin	0.206	1	0.650
ab	0.473	1	0.492
female	0.606	1	0.436

Hormis pour le score de la beauté pondérée, on voit qu'on n'est pas capable de rejeter l'hypothèse nulle a un taux raisonnable. Pour le score de la beauté pondérée, on est très proche du fait de rejeter H_0 au taux 10 pourcent. Comme le rejet au niveau 10 pourcents est relativement faible, et aux vus des résultats des autres variables et notamment du test "Omnibus" on peut décider de ne pas rejeter l'hypothèse nulle.

Ceci est une bonne indication pour conserver les hypothèses du modèle des questions précédentes et considérer qu'il n'y a qu'une seule variable latente.