

Evaluation de la performance d'un jury et des juges

François Husson

Laboratoire de mathématiques appliquées

Agrocampus ouest

husson@agrocampus-ouest.fr

Probabilités critiques du modèle :

$$\text{Note} = P + J + S + J * P + J * S + P * S$$

	Product	Panelist	Session	Prod:Panel	Prod:Sess	Panel:Sess	median
MilkF	3.74e-40	7.51e-11	0.0924	0.00682	0.118	0.0154	0.01111
Bitterness	1.88e-29	1.64e-10	0.012	0.0258	0.19	0.0554	0.0189
CocoaF	1.75e-28	2e-11	0.0971	0.00665	0.449	0.0941	0.05038
Crunchy	2.88e-27	2.88e-14	0.0546	0.0238	0.687	0.00448	0.01414
Sweetness	1.11e-24	2.42e-12	0.0618	0.118	0.119	0.00325	0.03253
Caramel	1.18e-22	2.39e-23	0.949	0.00312	0.0732	0.000846	0.001983
Astringency	1.47e-15	1.14e-22	0.94	0.0594	0.0818	0.0585	0.05895
CocoaA	8.85e-14	2.55e-08	0.239	0.993	0.801	0.826	0.52
Melting	4.58e-12	1.2e-17	0.118	6.88e-05	0.304	0.062	0.03103
Acidity	5.26e-11	6.14e-25	0.532	0.000915	0.983	0.00812	0.004518
Vanilla	2.78e-10	9.75e-23	0.227	0.00258	0.318	0.0377	0.02014
MilkA	6.44e-08	1.1e-13	0.289	0.0928	0.829	0.233	0.1629
Granular	9.08e-05	2.32e-21	0.147	0.015	0.283	0.31	0.081
Sticky	0.000531	8.37e-21	0.473	0.0541	0.00996	0.0479	0.02893

Modèle : $Note = P + J + S + J * P + J * S + P * S$

- Effet Produit pour tous les descripteurs

⇒ Discrimination des produits

- Effet Juge pour tous les descripteurs

⇒ Utilisation différente de l'échelle de note

- Effet Séance et des interactions pour certains descripteurs

Étude de la performance d'un juge ou d'un jury

Un juge ou un jury est performant si :

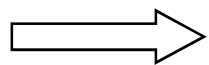
- Il **discrimine** beaucoup les produits (c'est le cas ici car effet Produit significatif)
- Il est **répétable** (i.e. il discrimine les produits de la même façon d'une séance à l'autre)

Étude de la répétabilité du jury

- Évaluation de la répétabilité des juges (dans leur ensemble) par l'interaction Produit - Séance du modèle :

$$\text{Note} = P + J + S + J*P + J*S + P*S$$

- **Si interaction Produit - Séance significative** : les juges dans leur ensemble n'évaluent pas chacun des produits de la même manière d'une séance à l'autre



Problématique car juges pas répétables

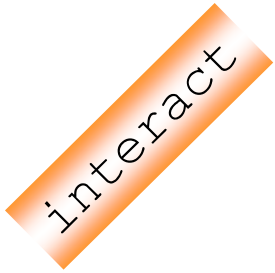
Étude de la répétabilité des juges par l'interaction Produit - Séance

	Product:Session
Sticky	0.009964
Caramel	0.07315
Astringency	0.08183
MilkF	0.1182
Sweetness	0.1188
Bitterness	0.1897
Granular	0.283
Melting	0.3041
Vanilla	0.3183
CocoaF	0.4487
Crunchy	0.6869
CocoaA	0.8011
MilkA	0.8287
Acidity	0.9835

Tous les produits n'ont pas été évalués de la même façon aux 2 séances pour le descripteur collant :

les juges ne sont pas répétables
ou
les produits ont été modifiés
(T° différente, ...)

Pour les autres descripteurs le jury est répétable



Quelles interactions ont contribué à l'interaction Produit - Séance ?

Sticky

	Séance1	Séance2
choc1	0.325	-0.325
choc2	-0.261	0.261
choc3	-0.193	0.193
choc4	0.704	-0.704
choc5	-0.417	0.417
choc6	-0.158	0.158

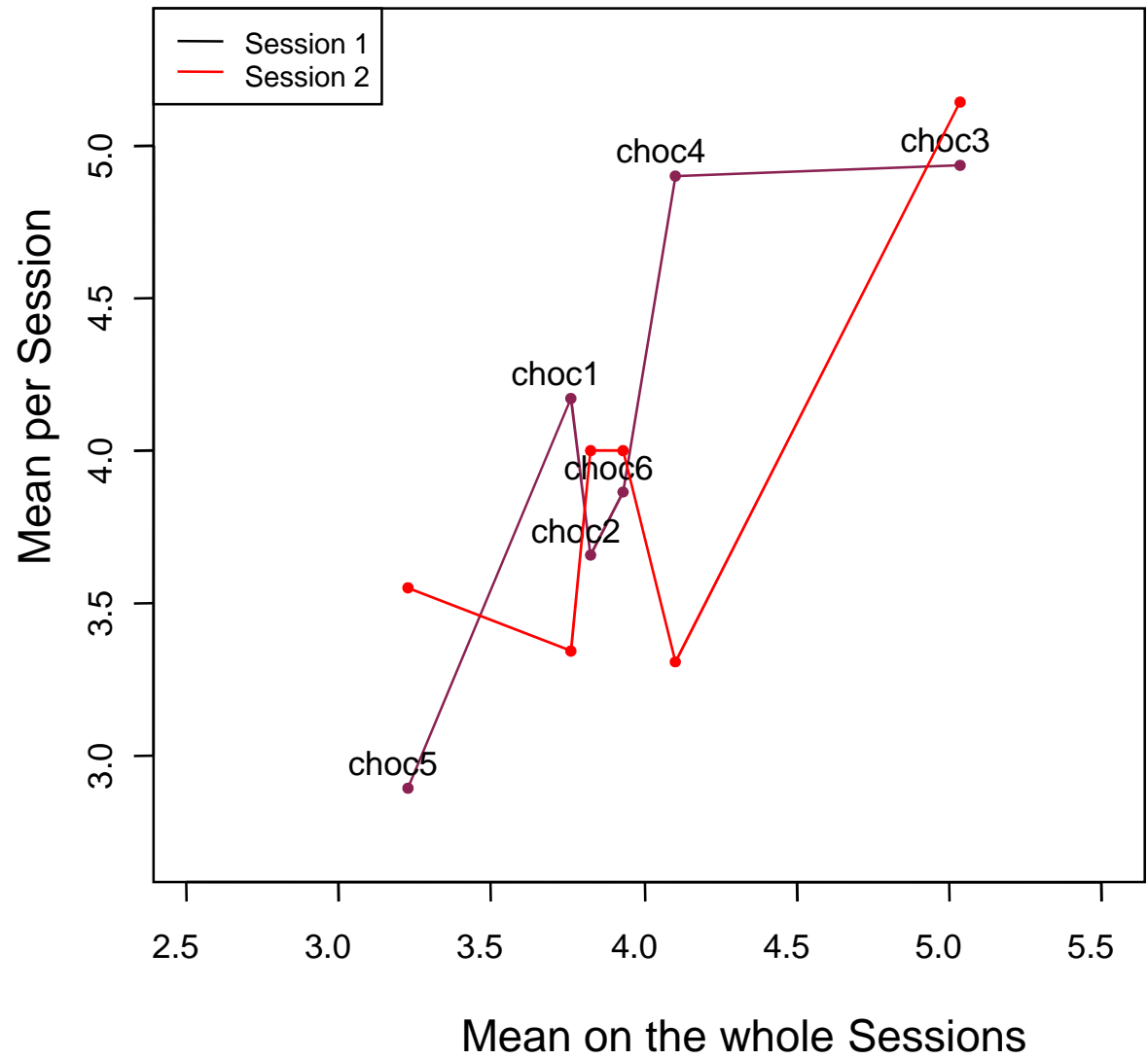
choc4 moins collant à la séance 2 qu'à la séance 1

Notation de choc6 homogène aux 2 séances

Coefficients de l'interaction Produit - Séance

Étude de l'interaction Produit - Séance

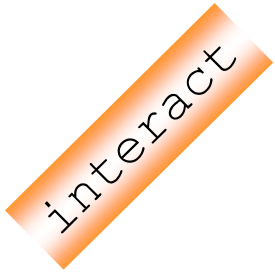
Sticky



Produits triés par moyenne croissante.

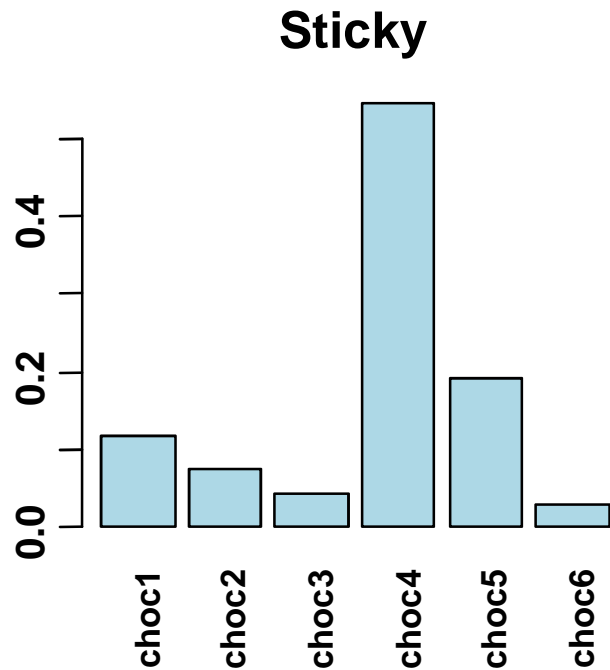
Une ligne brisée par séance.

Non parallélisme = interaction



Étude de l'interaction Produit – Séance

Visualisation de la contribution des produits



Le chocolat 4 contribue a plus de 50% à l'interaction produit – séance calculé de la façon suivante :

$$Prod_i = \frac{\sum_{j=1}^J \alpha\beta_{ij}^2}{\sum_{i=1}^I \sum_{j=1}^J \alpha\beta_{ij}^2}$$

Étude de la répétabilité des juges

- Évaluation par l'interaction P*J*S du modèle :
Note = P + J + S + J*P + J*S + P*S + P*J*S
Impossible car pas de répétitions
- Évaluation (**par juge**) de l'interaction P*S du modèle :
Note = P + S + P:S (par juge)
Impossible car pas de répétitions
- Évaluation de la variance résiduelle (globale ou par juge)
Problème : pas de test possible (comparaison qualitative entre juges ou entre descripteur ou par rapport à d'autres études)

Étude de la répétabilité des juges

Par l'écart-type résiduel du modèle :

$$\text{Note} = P + J + S + J*P + J*S + P*S$$

	Ecart-type résiduel
Cacao	1,41
Vanille	1,44
Lait	1,47
Croquant	1,62
Caramel	1,63
Fondant	1,65
Acide	1,66
Sucre	1,67
Amer	1,69
Astringent	1,70
Collant	1,82
O.Cacao	1,87
Granuleux	1,89
O.lait	1,89

Étude de la performance d'un juge

- Objectif : sélectionner les juges les plus performants pour constituer un jury d'experts
- Un juge est performant si :
 - Il **discrimine** beaucoup les produits
 - Il est **répétable**
 - Il est **en accord** avec ce que dit le jury

Performance individuelle des juges capacité à discriminer

- Évaluation par juge de l'effet Produit du modèle :
Note = Produit (+ Séance)
- Effet Produit de ce modèle mesuré par :

$$F_{obs} = \frac{SC_{\text{produit}} / ddl_{\text{produit}}}{SC_{\text{résiduelle}} / ddl_{\text{résiduelle}}}$$

Grand si Produits discriminés

Petit si faible résiduelle (= interaction Produit - Séance) ➡ si juge répétable

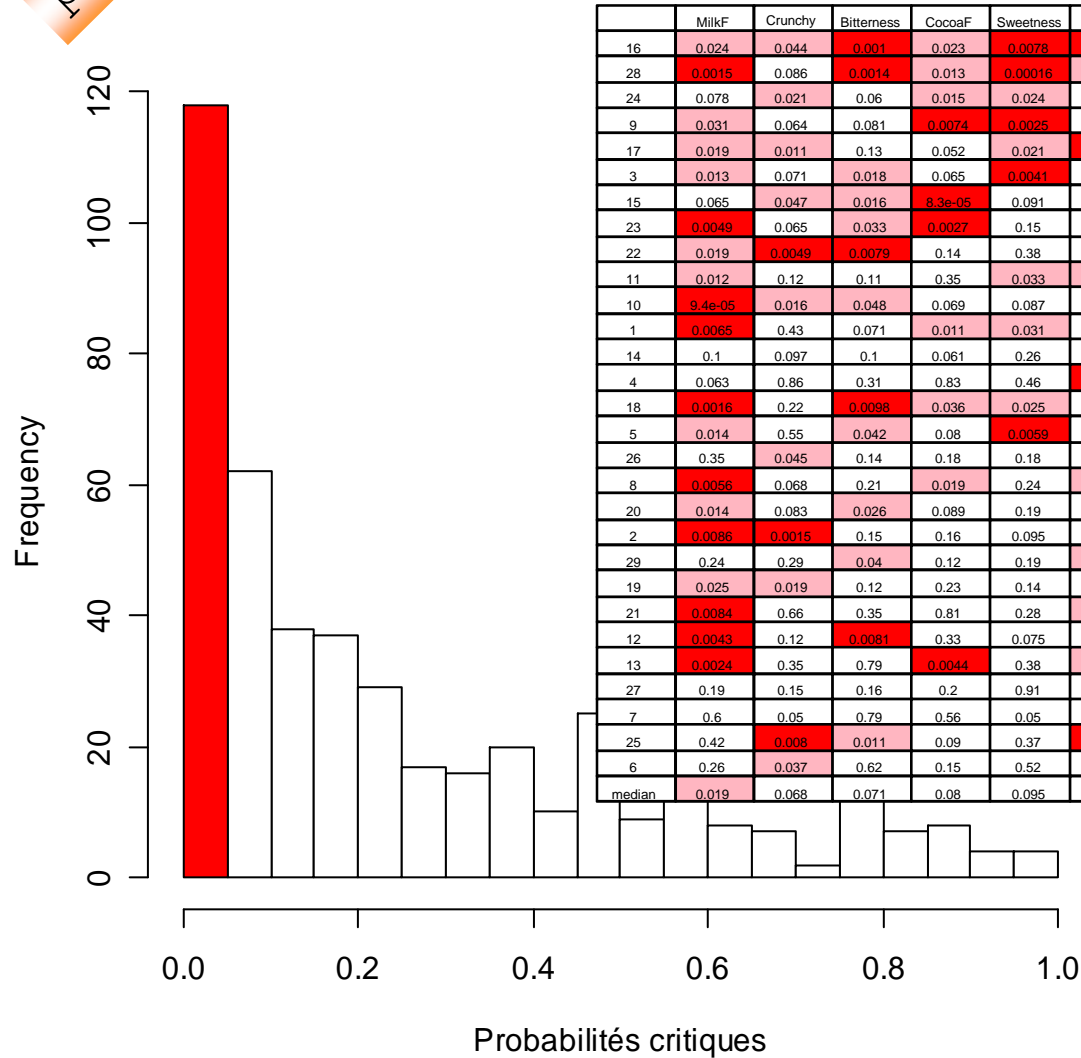
➡ Ce test évalue bien la performance

Performance individuelle des juges consensus avec le panel

- Évaluation par juge des moyennes ajustées de l'effet Produit (modèle : Note = Produit)
- Évaluation des moyennes ajustées de l'effet produit pour le jury (modèle (Note = P + J + S + PJ + PS + JS))
- Consensus entre un juge et le jury évalué par le coefficient de corrélation entre :

$\left(\begin{array}{c} \hat{\mu} + \hat{\alpha}_1 \\ \hat{\mu} + \hat{\alpha}_2 \\ \vdots \\ \hat{\mu} + \hat{\alpha}_I \end{array} \right)$;	$\left(\begin{array}{c} \hat{\mu} + \hat{\alpha}_1 \\ \hat{\mu} + \hat{\alpha}_2 \\ \vdots \\ \hat{\mu} + \hat{\alpha}_I \end{array} \right)$
Juge j		Panel

Performance individuelle des juges capacité à discriminer



	MilkF	Crunchy	Bitterness	CocoaF	Sweetness	Melting	Caramel	Vanilla	Acidity	Astringency	Granular	MilkA	Sticky	CocoaA	median
16	0.024	0.044	0.001	0.023	0.0078	0.0015	0.002	0.16	0.033	0.0053	0.77	0.17	0.37	0.59	0.028
28	0.0015	0.086	0.0014	0.013	0.00016	0.04	0.019	0.18	0.68	0.052	0.76	0.0032	0.045	0.049	0.043
24	0.078	0.021	0.06	0.015	0.024	0.37	0.022	0.032	0.047	0.32	0.78	0.035	0.091	0.096	0.053
9	0.031	0.064	0.081	0.0074	0.0025	0.31	0.022	0.34	0.054	0.28	0.27	0.008	0.25	0.014	0.059
17	0.019	0.011	0.13	0.052	0.021	0.0097	0.004	0.008	0.19	0.37	0.068	0.56	0.45	0.54	0.06
3	0.013	0.071	0.018	0.065	0.0041	0.79	0.28	0.54	0.14	0.0085	0.58	0.038	0.69	0.022	0.068
15	0.065	0.047	0.016	8.3e-05	0.091	0.24	0.005	0.15	0.41	0.064	0.69	0.79	0.26	0.082	0.087
23	0.0049	0.065	0.033	0.0027	0.15	0.24	0.013	0.13	0.024	0.21	0.014	0.95	0.92	0.99	0.096
22	0.019	0.0049	0.0079	0.14	0.38	0.057	0.003	NaN	0.056	0.5	0.62	0.11	0.47	0.84	0.11
11	0.012	0.12	0.11	0.35	0.033	0.027	0.0064	0.041	0.026	0.94	0.11	0.16	0.2	0.5	0.11
10	9.4e-05	0.016	0.048	0.069	0.087	0.74	0.068	0.14	0.5	0.18	0.014	0.41	0.25	0.79	0.11
1	0.0065	0.43	0.071	0.011	0.031	0.072	0.13	0.066	0.21	0.17	0.89	0.48	0.47	0.095	0.11
14	0.1	0.097	0.1	0.061	0.26	0.11	0.025	0.83	0.38	0.18	0.062	0.31	0.12	0.24	0.12
4	0.063	0.86	0.31	0.83	0.46	0.0018	0.21	0.11	0.14	0.15	0.64	0.14	0.1	0.089	0.14
18	0.0016	0.22	0.0098	0.036	0.025	0.43	0.21	0.5	0.022	0.18	0.071	0.32	0.12	0.59	0.15
5	0.014	0.55	0.042	0.08	0.0059	0.82	0.059	0.22	0.34	0.23	0.16	0.77	0.16	0.14	0.16
26	0.35	0.045	0.14	0.18	0.18	0.65	0.0065	0.11	0.067	0.55	0.088	0.79	0.56	0.41	0.18
8	0.0056	0.068	0.21	0.019	0.24	0.018	0.64	0.2	0.18	0.32	0.04	0.23	0.094	0.39	0.19
20	0.014	0.083	0.026	0.089	0.19	0.37	0.2	0.69	0.24	0.092	0.096	0.84	0.53	0.57	0.2
2	0.0086	0.0015	0.15	0.16	0.095	0.18	0.5	0.5	0.5	NaN	0.35	0.57	0.21	0.84	0.21
29	0.24	0.29	0.04	0.12	0.19	0.024	0.13	0.093	0.47	0.48	0.15	0.23	0.56	1	0.21
19	0.025	0.019	0.12	0.23	0.14	0.055	0.19	0.27	0.38	0.89	0.64	0.22	0.68	0.39	0.23
21	0.0084	0.66	0.35	0.81	0.28	0.045	0.00044	NaN	0.00091	0.021	0.25	0.0028	0.79	0.78	0.25
12	0.0043	0.12	0.0081	0.33	0.075	0.63	0.25	0.036	0.54	0.5	0.5	0.24	0.32	0.38	0.28
13	0.0024	0.35	0.79	0.0044	0.38	0.024	0.37	0.034	0.59	0.29	0.85	0.049	0.33	0.035	0.31
27	0.19	0.15	0.16	0.2	0.91	0.42	0.37	0.5	0.47	0.4	0.26	0.35	0.87	0.48	0.39
7	0.6	0.05	0.79	0.56	0.05	0.078	0.49	0.68	0.047	0.37	0.25	0.88	0.28	0.9	0.43
25	0.42	0.008	0.011	0.09	0.37	6.9e-73	0.5	0.5	0.5	0.19	0.5	0.97	0.9	0.74	0.46
6	0.26	0.037	0.62	0.15	0.52	0.75	0.53	0.65	0.15	0.075	0.056	0.57	0.52	0.92	0.52
median	0.019	0.068	0.071	0.08	0.095	0.11	0.13	0.18	0.19	0.22	0.26	0.31	0.33	0.48	0.15

Note = Produit + Séance

panelperf

Performance individuelle des juges


consensus avec le panel

Consensus = corrélation

$$\begin{pmatrix} \hat{\mu} + \hat{\alpha}_1 & \hat{\mu} + \hat{\alpha}_1 \\ \hat{\mu} + \hat{\alpha}_2 & \hat{\mu} + \hat{\alpha}_2 \\ \vdots & \vdots \\ \hat{\mu} + \hat{\alpha}_I & \hat{\mu} + \hat{\alpha}_I \end{pmatrix}$$

	MilkF	Crunchy	Bitterness	CocoaF	Sweetness	Melting	Caramel	Vanilla	Acidity	Astringency	Granular	MilkA	Sticky	CocoaA	median
16	0.86	0.91	0.69	0.66	0.7	0.3	0.75	0.85	0.33	0.83	0.2	0.93	0.053	0.7	0.7
28	0.96	0.67	0.96	0.9	0.95	0.9	0.91	0.88	0.68	0.85	0.87	0.99	0.85	0.95	0.9
24	0.95	0.73	0.91	0.95	0.9	0.77	0.93	0.91	0.95	0.91	0.37	0.82	0.85	0.95	0.91
9	0.88	0.95	0.86	0.86	0.83	0.86	0.91	0.88	0.95	0.9	0.49	0.81	0.28	0.72	0.86
17	0.98	0.82	0.89	0.93	0.96	0.83	0.96	0.86	0.83	0.71	0.31	0.81	0.36	0.82	0.83
3	0.96	0.72	0.87	0.98	0.87	0.67	0.66	0.86	0.97	0.92	0.53	0.62	0.9	0.84	0.86
15	0.91	0.92	0.79	0.91	0.94	0.76	0.96	0.95	-0.051	0.82	-0.81	0.41	0.84	0.84	0.84
23	0.95	0.9	0.93	0.96	0.74	0.78	0.95	0.9	0.81	0.91	0.24	-0.12	0.76	-0.12	0.86
22	0.71	0.84	0.79	0.77	0.55	0.92	0.92		0.7	0.79	0.76	0.72	0.84	0.7	0.77
11	0.94	0.71	0.91	0.59	0.95	0.88	0.98	0.73	0.38	0.51	0.5	-0.34	0.45	-0.012	0.65
10	0.96	0.93	0.68	0.82	0.8	0.81	0.78	0.41	-0.046	0.84	0.45	0.81	0.82	0.82	0.81
1	0.95	0.91	0.92	0.89	0.97	0.85	0.94	0.9	0.74	0.34	-0.26	0.41	0.5	0.73	0.87
14	0.98	0.65	0.95	0.95	0.87	0.75	0.88	0.6	-0.46	0.67	0.35	0.53	-0.46	0.55	0.66
4	0.97	0.87	0.86	0.59	0.57	0.75	0.52	0.8	0.81	0.95	0.5	0.96	-0.068	0.89	0.81
18	0.98	0.81	0.66	0.87	0.72	0.34	0.96	0.75	0.88	0.65	0.42	0.5	0.58	0.9	0.73
5	0.73	0.44	0.83	0.87	0.84	0.42	0.77	0.53	-0.54	0.63	0.63	0.093	0.37	0.28	0.58
26	0.71	0.97	0.91	0.9	0.91	0.97	0.61	0.68	0.031	0.36	0.69	-0.13	0.69	0.15	0.69
8	0.97	0.95	0.85	0.94	0.76	0.65	0.73	0.33	0.67	0.51	0.6	0.66	0.86	0.88	0.75
20	0.98	0.83	0.9	0.97	0.93	0.55	0.91	0.66	0.6	0.85	0.65	0.38	0.62	0.73	0.78
2	0.95	0.6	0.92	0.96	0.94	0.88	0.9	0.82	0.68		0.16	0.35	-0.087	0.84	0.84
29	0.87	0.79	0.76	0.85	0.93	0.43	0.69	0.89	0.79	0.64	0.43	-0.24	0.046	0.54	0.72
19	0.96	0.96	0.96	0.92	0.9	0.61	0.47	-0.073	0.82	-0.51	0.17	0.8	0.48	0.75	0.77
21	0.88	0.88	0.72	0.12	0.92	0.63	0.95		0.66	0.84	0.82	0.52	0.26	0.75	0.75
12	0.94	0.98	0.93	0.95	0.68	0.75	0.76	0.76	0.51	0.011	0.5	0.86	0.79	0.82	0.78
13	0.99	0.74	0.86	0.97	0.85	0.47	0.96	0.92	0.9	-0.00053	0.48	0.95	0.16	0.92	0.88
27	0.36	0.82	0.52	0.99	0.041	-0.34	0.61	0.13	0.31	0.9	0.12	0.48	0.7	0.67	0.5
7	0.63	0.94	0.57	0.28	0.51	0.68	0.085	-0.39	0.75	0.83	0.9	-0.37	-0.63	0.058	0.54
25	0.92	0.93	0.91	0.79	0.84	-0.53	0.044	-0.16	-0.18	0.89	-0.39	-0.45	-0.096	0.36	0.2
6	0.71	0.082	0.66	0.61	-0.36	-0.079	0.89	-0.29	0.57	0.78	0.1	0.23	0.16	0.34	0.29
median	0.95	0.84	0.86	0.9	0.85	0.75	0.89	0.76	0.68	0.81	0.45	0.52	0.48	0.73	0.77

Étude de la reproductibilité des juges par l'interaction produit - juge



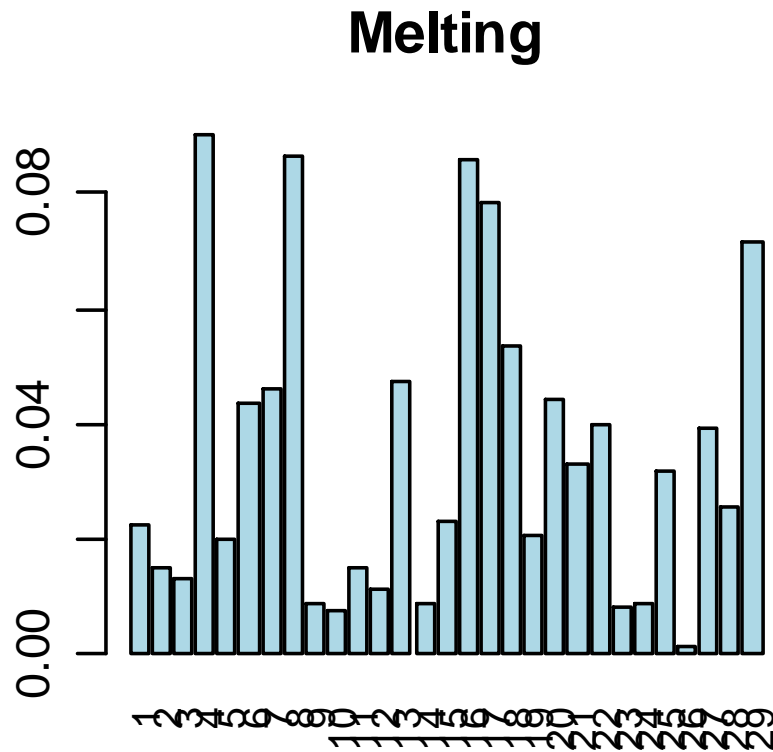
Melting	6.88e-05
Acidity	0.0009151
Vanilla	0.002577
Caramel	0.00312
CocoaF	0.00665
MilkF	0.006824
Granular	0.01502
Crunchy	0.02381
Bitterness	0.0258
Sticky	0.05412
Astringen	0.05936
MilkA	0.09279
Sweetness	0.1177
CocoaA	0.9929

Descripteurs difficiles (pas de consensus entre juges)

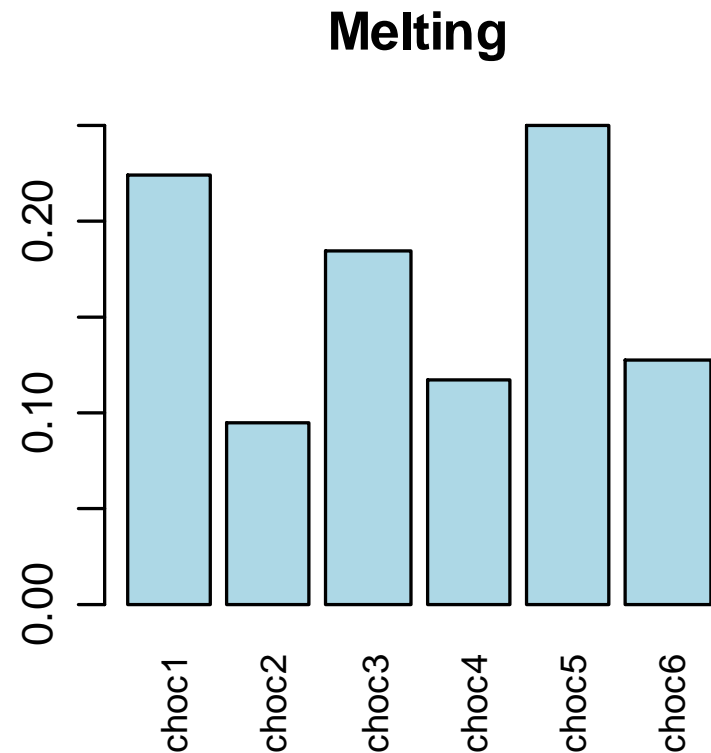
Descripteurs avec un fort consensus entre juges

interact

Étude de la reproductibilité des juges par l'interaction produit – juge



Contribution des juges
à l'interaction



Contribution des produits
à l'interaction 39

Test sur l'effet Rang

- Y a-t-il un effet Rang ?

Modèle utilisé :

$$\text{Note} = P + J + S + R + J*P + J*S + P*S$$

Probabilité critique de l'effet rang

Sucre	7,59E-04
Vanille	1,34E-02
O.Cacao	3,38E-02
Astringent	0,041227
Granuleux	4,59E-02
Lait	1,13E-01
Croquant	2,27E-01
O.lait	2,47E-01
Acide	0,289937
Cacao	4,34E-01
Caramel	0,451585
Fondant	0,470673
Collant	0,729917
Amer	8,17E-01



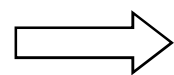
Test sur l'effet Rang

Quels rangs contribuent à cet effet Rang ?

Rang	Sucre	Vanille	O.Cacao	Astringent	Granuleux
1	-1,11 *	-0,71 *	0,92 *	-0,70 *	-1,03 *
2	0,45	-0,33	-0,69 *	-0,46	0,27
3	0,11	0,42	0,35	0,40	-0,50
4	-0,44	0,18	-0,25	0,60	0,46
5	0,43	0,40	-0,40	-0,03	0,40
6	0,57 *	0,04	0,07	0,19	0,40

* Coefficient significativement différent de 0

Le rang 1 contribue beaucoup à l'effet Rang



Faire des tests avec l'effet Premier

Test sur l'effet Rang

Heureusement
que le plan
d'expériences
était équilibré
pour les rangs !!!



Test sur l'effet Précédent

- Y a-t-il un effet Précédent ?

Modèle utilisé :

$$\text{Note} = P + J + S + \text{Prec} + J*P + J*S + P*S$$

Probabilité
critique de
l'effet précédent

Cacao	3,15E-04
Lait	2,34E-03
Sucre	4,99E-03
O.Cacao	1,24E-02
Amer	5,12E-02
Vanille	6,06E-02
Astringent	0,075644
Granuleux	1,30E-01
Acide	2,32E-01
Croquant	4,42E-01
Caramel	0,472994
Collant	0,644745
O.lait	8,22E-01
Fondant	9,03E-01



Test sur l'effet Précédent

Quels produits contribuent à cet effet précédent ?

Précédent	O.Cacao	Sucre	Lait	Cacao
0	0,92 *	-1,24 *	-0,68 *	0,61 *
Excellence	-0,77 *	0,60 *	0,67 *	-0,57 *
Amère	-0,34	0,42	-0,04	-0,61 *
Mi-doux	0,84 *	-0,20	-0,77 *	0,88 *
Amazonie	-0,32	0,03	0,02	0,28
Pâtissier	-0,15	-0,07	0,37	-0,43
Supérieur	-0,18	0,46	0,43	-0,15

* Coefficient significativement différent de 0

Précédent = 0 pour 1^{er} produit déguster : effet premier fort

Produit déguster après Excellence : note sucre augmentée de 0.6014
explication : « Excellence » très peu sucré (cf. decat, transp 22) ₄₄

Récapitulatif

- Effet **Produit** (toujours significatif) : produits bien discriminés (**très intéressant**)
- Effet **Juge** (toujours significatif) : juges utilisent l'échelle de notes différemment (**pas grave**)
- Effet **Séance** (parfois significatif) : utilisation différente de l'échelle d'une séance à l'autre (**pas grave**)
- Effet **Rang** et **Précédent** (parfois significatif) : heureusement, le plan d'exp. était équilibré pour les rangs et les arrières-effets d'ordre 1 (sinon **grave**)

Récapitulatif

- Interaction **Produit - Séance** (1 fois significative) : pour ce descripteur (collant), l'ensemble des juges (i.e. le jury) n'est pas répétable (**grave**)
- Interaction **Produit - Juge** (parfois significative) : pour ces descripteurs, pas de consensus entre juges (**grave**)
- Interaction **Juge - Séance** (parfois significative) : pour certains juges, utilisation différente de l'échelle d'une séance à l'autre (**pas grave**)