APEX

8 Octobre 2020

Capabilité des plans d’expériences

Didier POIRAULT
Didier.poirault@biomerieux.com

PIONEERING DIAGNOSTICS

PART I – RESUME EPISODE I

PART II – BAYESIAN FRAMEWORK

… *IN A NUTSHELL*

PART III – APPLICATION AU CAS DE MARION

PART IV – DISCUSSIONS

PART I – RESUME EPISODE I

PART II – BAYESIAN FRAMEWORK

… IN A NUTSHELL

PART III – APPLICATION AU CAS DE MARION

PART IV – DISCUSSIONS

DÉROULÉ DE LA PRÉSENTATION DE MARION

Comparaison de 2 conditions expérimentales

Définition d'un seuil de significativité statistique compte tenu :

- *de la connaissance*
- *du protocole expérimental de comparaison*

Calcul de la capacité (capabilité) à détecter un niveau d'écart, compte tenu du protocole expérimental

Evolution du seuil de significativité et de la capabilité selon différents protocoles expérimentaux

Application à la comparaison de plusieurs conditions dans le cadre d'un plan d'expériences

LA QUESTION

Procédé de référence

Ex: θ=33°

Titre moyen M1

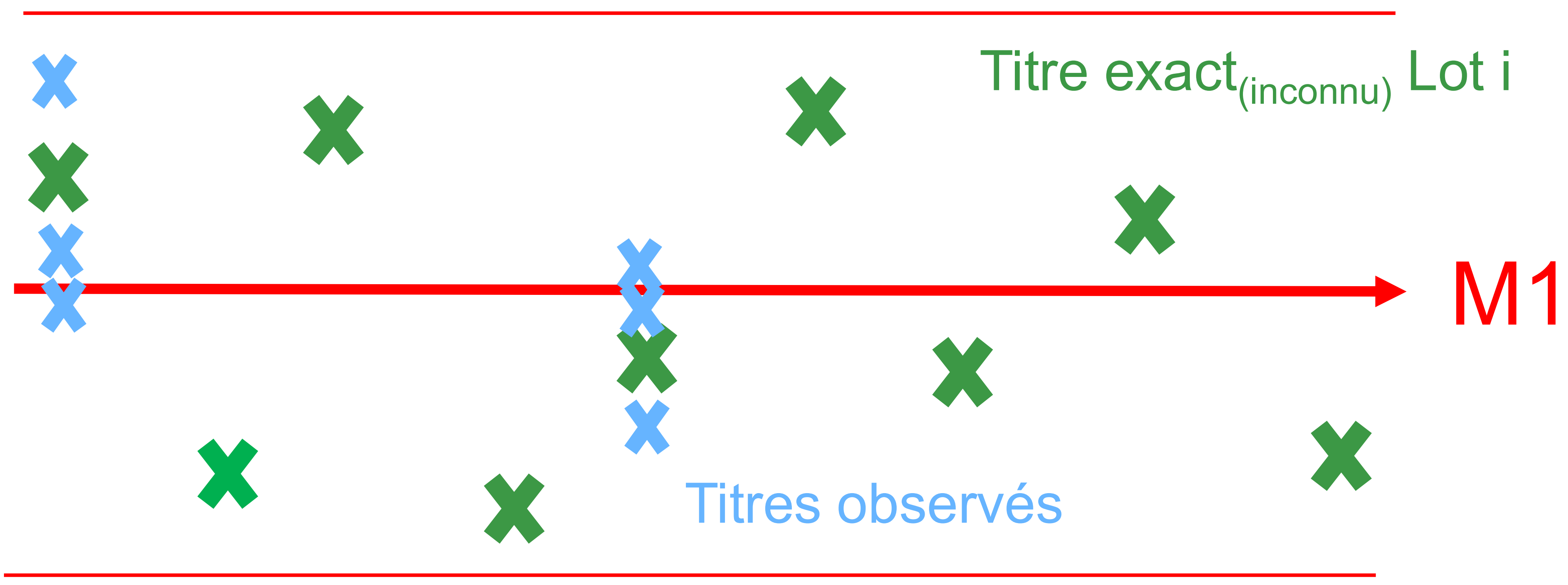
Nouveau procédé

Ex: θ=37°

Titre moyen M2

M1 = M2 ?

PROCÉDÉ DE RÉFÉRENCE



σ procédé + σ analytique

1. CONNAÎTRE NOTRE PROCÉDÉ DE DÉPART

Connaissance

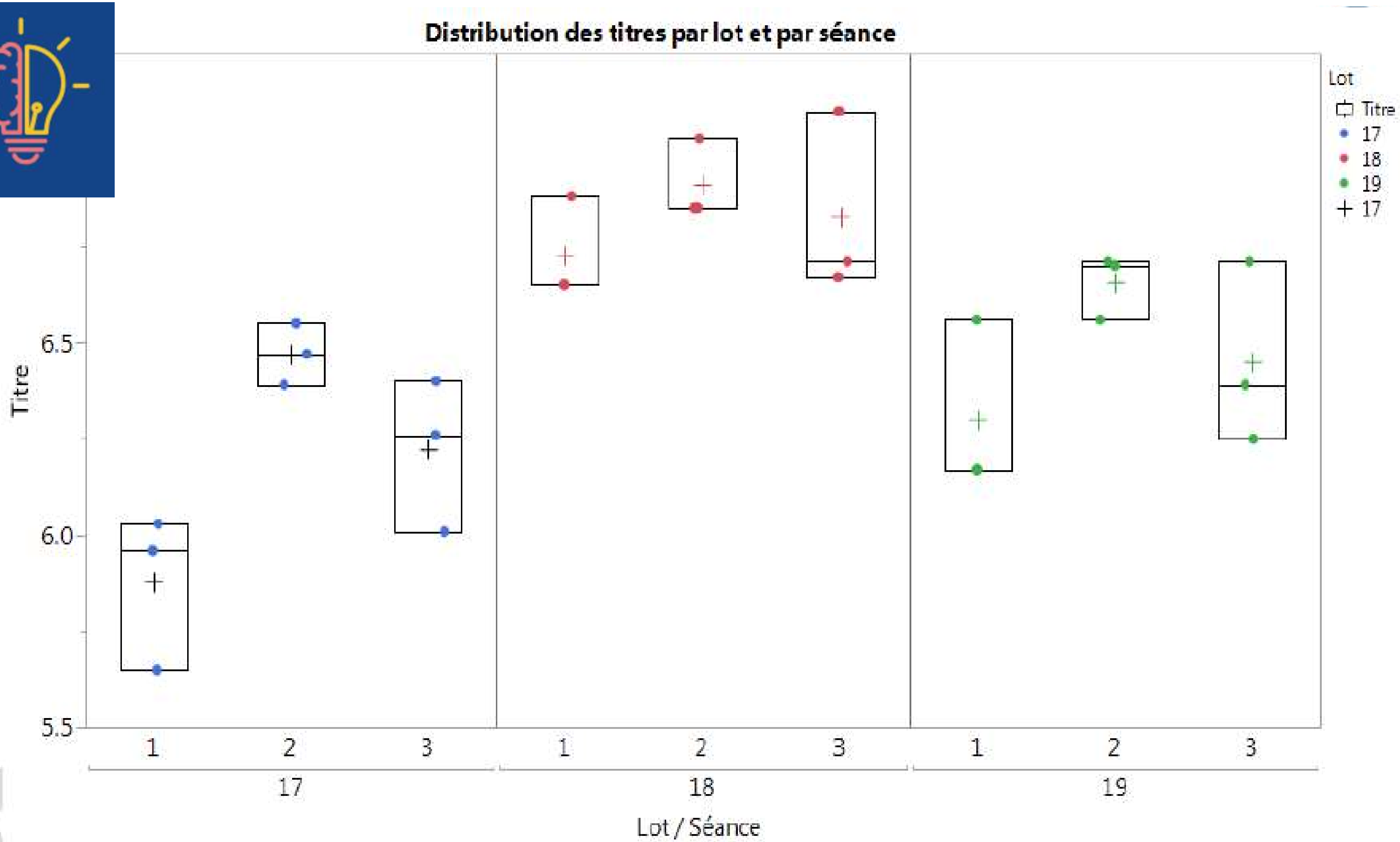
<u>Données</u> :

- 3 lots produits selon le procédé actuel / modalité actuelle
- Répétitions de mesures sur chaque lot :
  - 3 séances de titrages
  - 3 titrages par séance

Lot	Séance 1	Séance 2	Séance 3
17	3 titres	3 titres	3 titres
18	3 titres	3 titres	3 titres
19	3 titres	3 titres	3 titres

Variabilité procédé

Variabilité analytique



$\sigma_{total} = 0.386$

	Variance σ^2	Ecart-type σ
Procédé (Inter-Lots)	0.0854	0.292
Inter-Séances	0.0322	0.180
Répétabilité (Intra-séance)	0.0316	0.178

Variabilité procédé

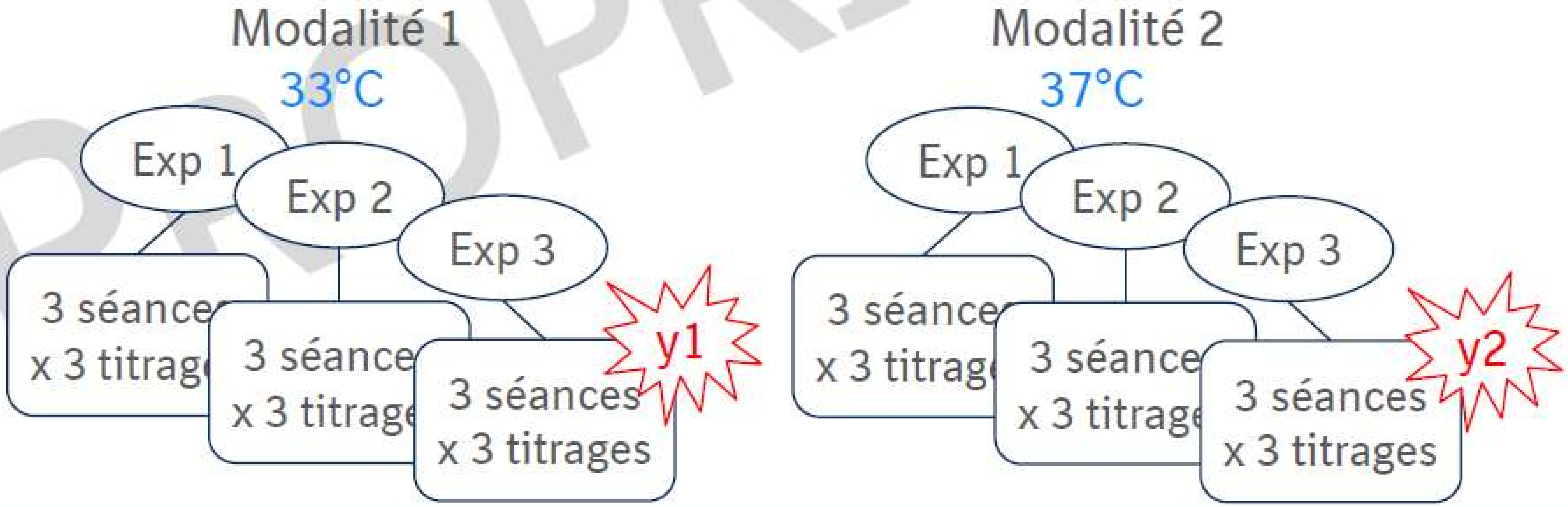
Variabilité analytique

3. PROTOCOLE

- 2 procédés (actuel et nouveau) / modalités.
  => Valeurs centrales = m1 et m2

New DATA

- Protocole (*plan d'expériences*) pour les comparer :
  - Production de lots (*expériences*) avec chaque procédé (*modalité*)
  - Titrage de chaque lot (*expérience*)



4. ERREUR STANDARD

- Précision avec laquelle on estime les valeurs centrales des 2 procédés (actuel et nouveau) / modalités m1 et m2

$$\sigma' = \sqrt{\frac{\sigma^2_{\text{procédé}}}{p} + \frac{\sigma^2_{\text{séance}}}{p \times k} + \frac{\sigma^2_{\text{répétabilité}}}{p \times k \times n}}$$

- p = nombre de répétitions « procédé »

  \# lots produits avec chaque procédé (actuel et nouveau)

  expériences *modalité*

- k = nombre de séances de titrages pour chaque lot / *expérience*
- n = nombre titrages par séance et par lot / *expérience*

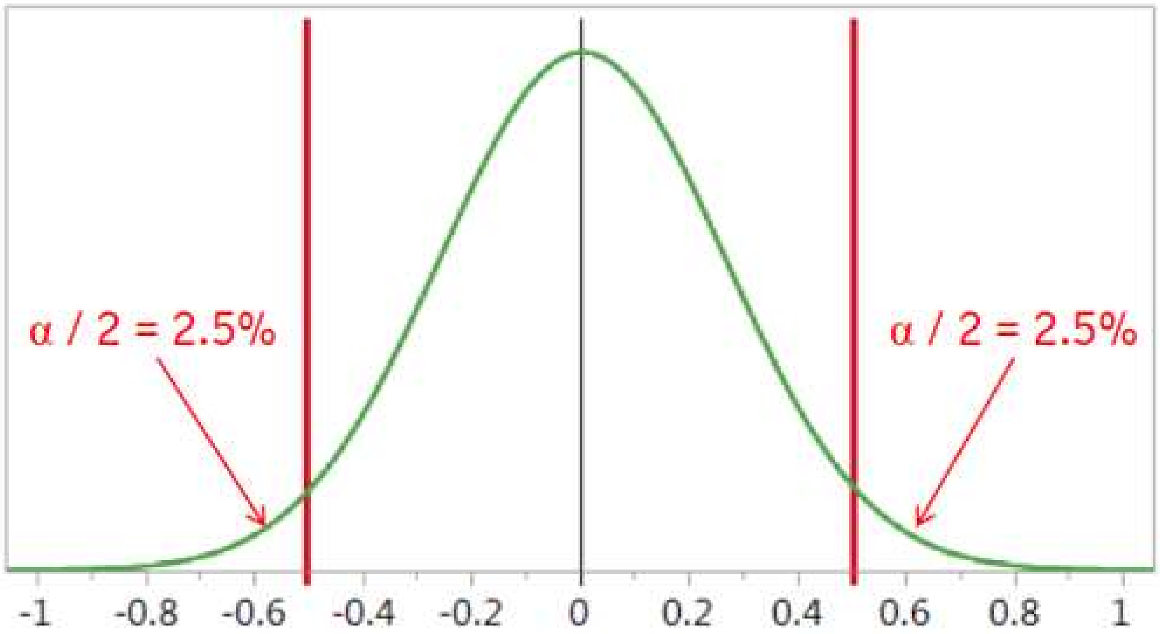
σ'= 0.182 pour p=k=n=3

HYPOTHÈSE M2 = M1

+ Hypothèse $\sigma'_{cond1} = \sigma'_{cond2}$

- **Distribution des différences y2 – y1 :**

➡ Centrée sur **0** et de variance **$2*\sigma^{2\prime}$**

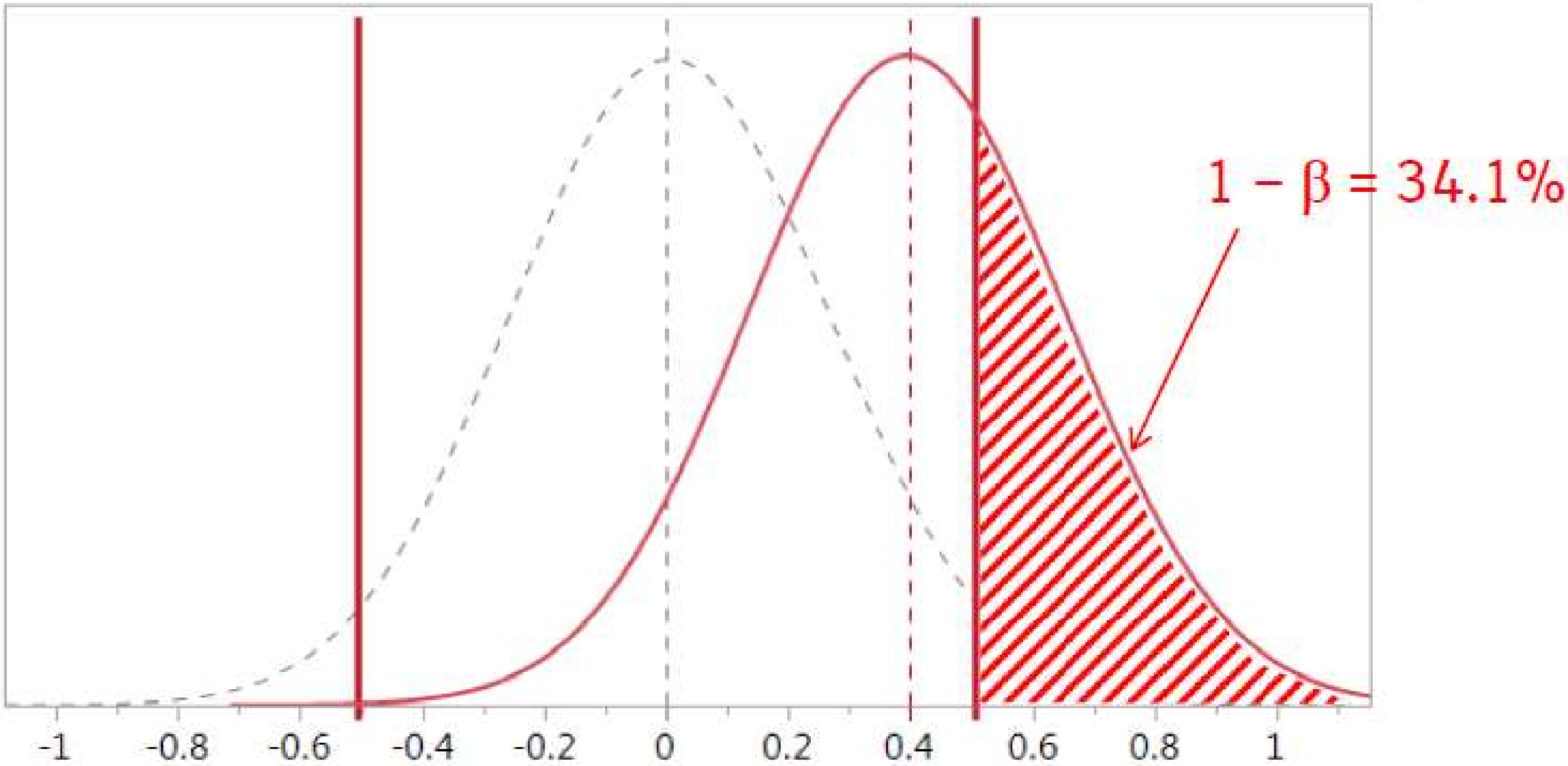


- **Seuils pour une différence y2-y1 significative (α= 5%) :**

➡ $\pm\ 1.96 \times \sqrt{2}\ \sigma' = \pm\ \mathbf{0.505}$

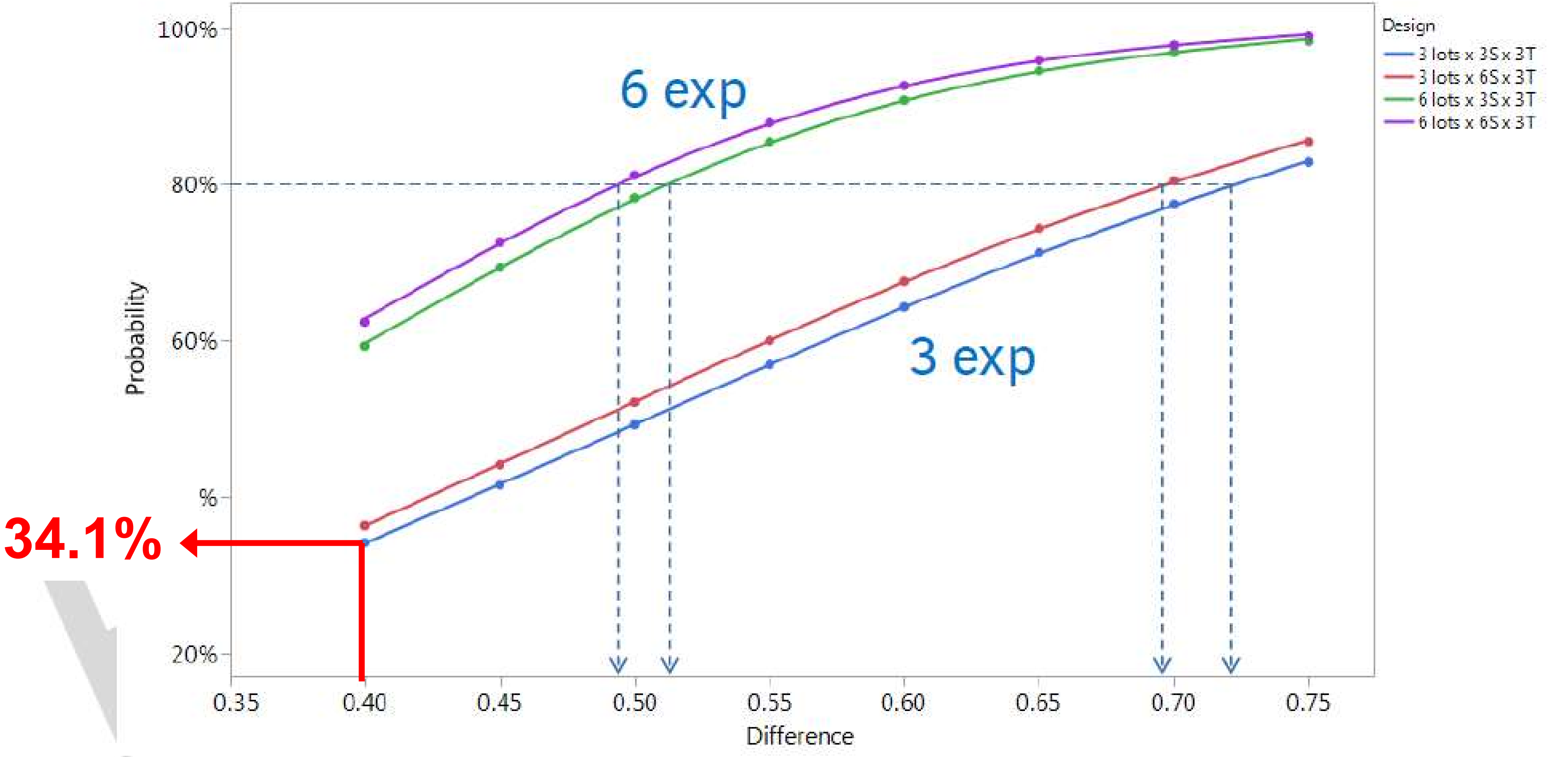
σ'= 0.182 pour p=k=n=3

HYPOTHÈSE M2 = (M1+0.4)



- Avec le protocole envisagé (3 expériences x 3 séances x 3 titres par modalité) :
  Capabilité = 34.1 %
  (probabilité d'identifier que l'on a un effet significatif)

5. COURBES DE PUISSANCE



Décliné aux DoE

- Ce qui entre dans le calcul de la capabilité :
  - L'écart réel entre les 2 procédés / modalités (m1 – m2) que l'on veut pouvoir identifier
  - L'erreur standard σ'
  - Le risque α (fixé à 5%)

➡ + l'incertitude associée

Ce sont des estimations !!

étabilité

	Variance σ^2	Ecart-type σ
Procédé (Inter-Lots)	0.0854	0.292
Inter-Séances	0.0322	0.180
Répétabilité (Intra-séance)	0.0316	0.178

$$\pm\ 1.96 \times \sqrt{2}\,\sigma' = \pm\ 0.505$$

PART I – RESUME EPISODE I

PART II – BAYESIAN FRAMEWORK

… IN A NUTSHELL

PART III – APPLICATION AU CAS DE MARION

PART IV – DISCUSSIONS

DEVINETTE

klaxoon

CHOISISSEZ PARMI LES 3 HYPOTHÈSES SUIVANTES….

- **Hypothèse 1**

  Elles viennent d'apprendre qu'elles ont gagné au LOTO

- **Hypothèse 2**

  Jean-Luc vient de leur annoncer qu'il va devenir RH de la société AZURAD

- **Hypothèse 3**

  Les participants à l'APEX viennent de les féliciter chaleureusement pour la qualité de l'organisation

COMMENT AVEZ-VOUS PROCÉDÉ ?

Hypothèse	**Probabilité H_i** sans aucune observation (faible / élevée) **P_1**		
H1: LOTO	**Faible**		
H2: Jean-Luc	**Faible**		
H3: Félicitations	**Elevée**		

COMMENT AVEZ-VOUS PROCÉDÉ ?

<u>Hypothèse</u>	**Probabilité H_i** sans aucune observation (faible / élevée) **P_1**	**Probabilité** de l'**observation** sachant l'hypothèse **P_2**	
H1: LOTO	**Faible**	**Faible**	
H2: Jean-Luc	**Faible**	**Elevée**	
H3: Félicitations	**Elevée**	**Elevée**	

Réponse fréquentiste $P_{(data/Hi)}$

COMMENT AVEZ-VOUS PROCÉDÉ ?

THE question!
$P_{(Hi/data)}$

<u>**Hypothèse**</u>	**Probabilité H_i** sans aucune observation (faible / élevée) **P_1**	**Probabilité** de l'**observation** sachant l'hypothèse **P_2**	**Probabilité Hi** compte-tenu de l'observation **$f(P_1;P_2)$**
H1: LOTO	**Faible**	**Faible**	**Faible**
H2: Jean-Luc	**Faible**	**Elevée**	**Medium**
H3: Félicitations	**Elevée**	**Elevée**	**Elevée**

Interprétation = f(connaissance ; observations)

BAYES

1702-1761

- **Probabilités conditionnelles :**

$$P(Hypothesis|data) = \frac{P(data|Hypothesis) \times P(Hypothesis)}{P(data)}$$

Interprétation (total data) = Observed data x Available data (knowledge)

Posterior probability — ***Likelihood*** — ***Prior probability***

Posterior probability: Combinaison de l'information collectée avant les data et de celle apportée par les data

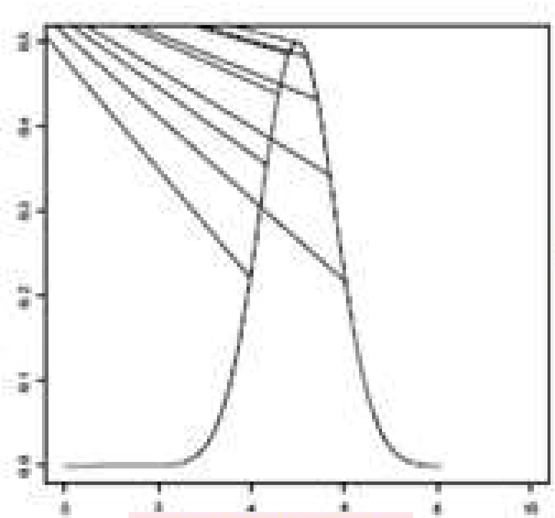
Likelihood: Information apportée par les data de l'expérience (observations)

Prior probability: Connaissance antérieure (études, experts, biblio …)

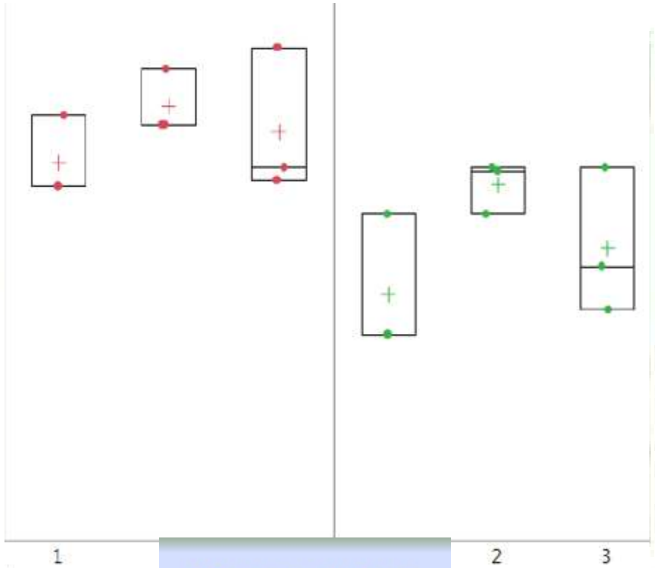
Comment les nouvelles données modifient mon opinion sur l'hypothèse ?

~~POINT ESTIMATES~~ → DISTRIBUTIONS

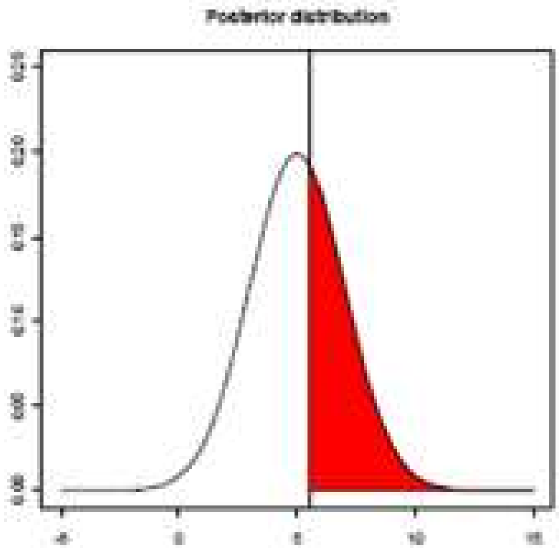
PRIOR distribution + *Exp. data* $\propto$ *POSTERIOR*



Available data
(knowledge)



Observed data



Interpretation
(total data)

- Prior
- Vraisemblance
- Postérieure
- Prédictive

PRIOR DISTRIBUTION: EXEMPLES

Paramètres à estimer

- Ex: stabilité

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

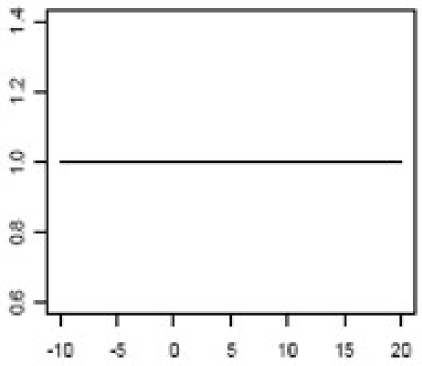
ε_i: Erreur Résiduelle

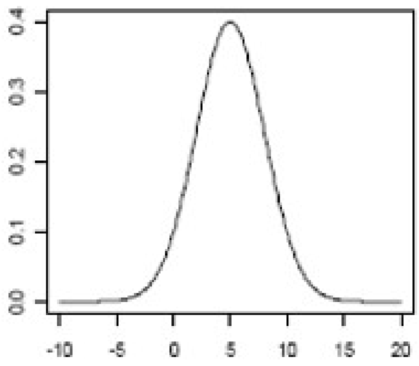
β_0: Intercept

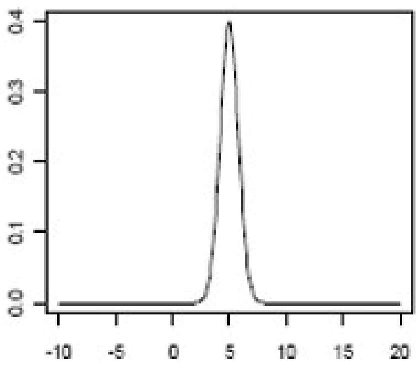
β_1: Slope

Pas d'information *a priori* sur θ **(uniforme)**

<u>Exemple</u> : β_0







β_0 peut varier entre -5 & 15
Distribution Normale centrée sur 5

Forte connaissance *a priori* sur θ

VRAISEMBLANCE, POSTÉRIEURE & PRÉDICTIVE

Vraisemblance

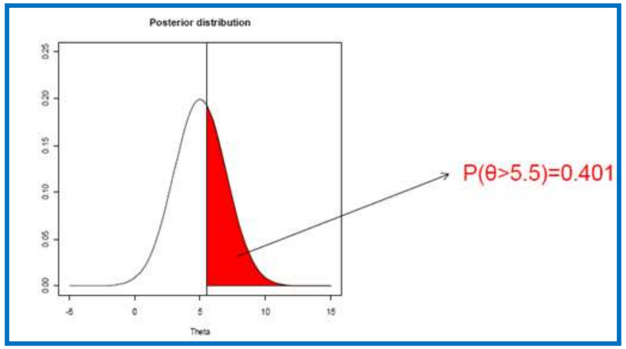
- Probabilité conditionnelle des données sachant β_i : $P(data \mid \beta_i)$

Distribution postérieure

- Distribution de β_i en ajoutant au prior la connaissance apportée par les données $P(\beta_i \mid DATA)$

Distribution prédictive

- Sachant le modèle et la distribution postérieure des β_i , quelles sont les valeurs possibles pour une observation future?



- L'incertitude est exprimée en **probabilités**

Ex : STABILITÉ

Observations

Paramètres à estimer

- Modèle : $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

Y_i: Réponse

β_0: Intercept

β_1: Slope

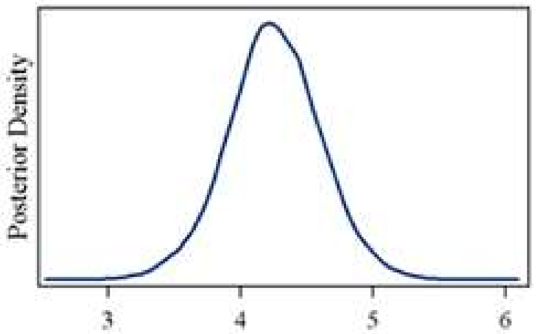
X_i: Durée de conservation

ε_i: Erreur Résiduelle

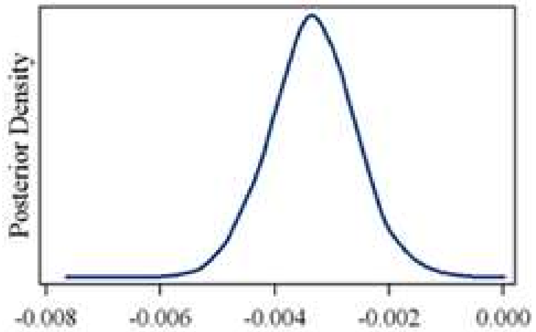
- Connaissance (ou pas) a priori sur les paramètres :

β_0 , β_1 , ε_i => Distributions Prior

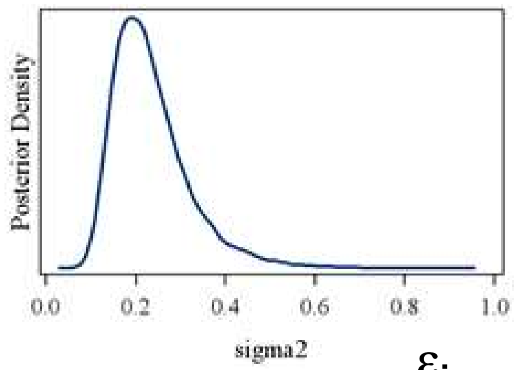
- Mise à jour avec les données => Distributions Postérieures



Intercept



Slope

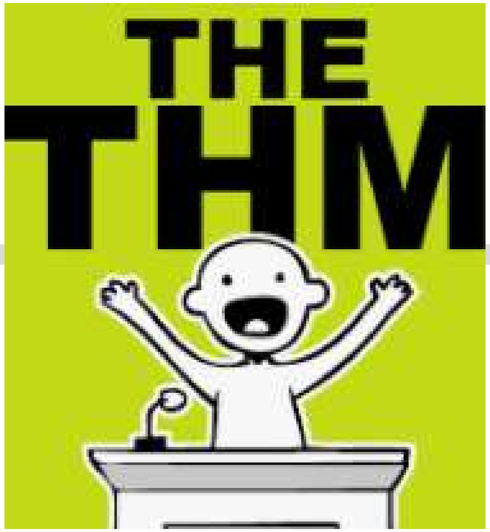


ε_i

EX : STABILITÉ (2)

- A partir des distributions postérieures, on peut alors générer des **distributions prédictives de Y** pour répondre à un ensemble de questions, comme :
  - Quelle est la probabilité que Y soit < à la spécification à péremption ?
  - Quelle doit être la spécification à T0 pour garantir le maintien de la performance (Y > spec.) à péremption ?
  - Quelle est la probabilité que Y > XX unités après T mois de conservation ?
  - …

BAYESIAN: TAKE HOME MESSAGES



 +  = 

Prédictive

Postérieure

~~POINT ESTIMATES~~ → DISTRIBUTIONS

Prior

Vraisemblance

PART I – RESUME EPISODE I

PART II – BAYESIAN FRAMEWORK
… IN A NUTSHELL

PART III – APPLICATION AU CAS DE MARION

PART IV – DISCUSSIONS

PART III – APPLICATION AU CAS DE MARION

PRISE EN COMPTE DE L'INCERTITUDE DES PARAMÈTRES

IMPACT DE LA CONNAISSANCE (PRIOR)

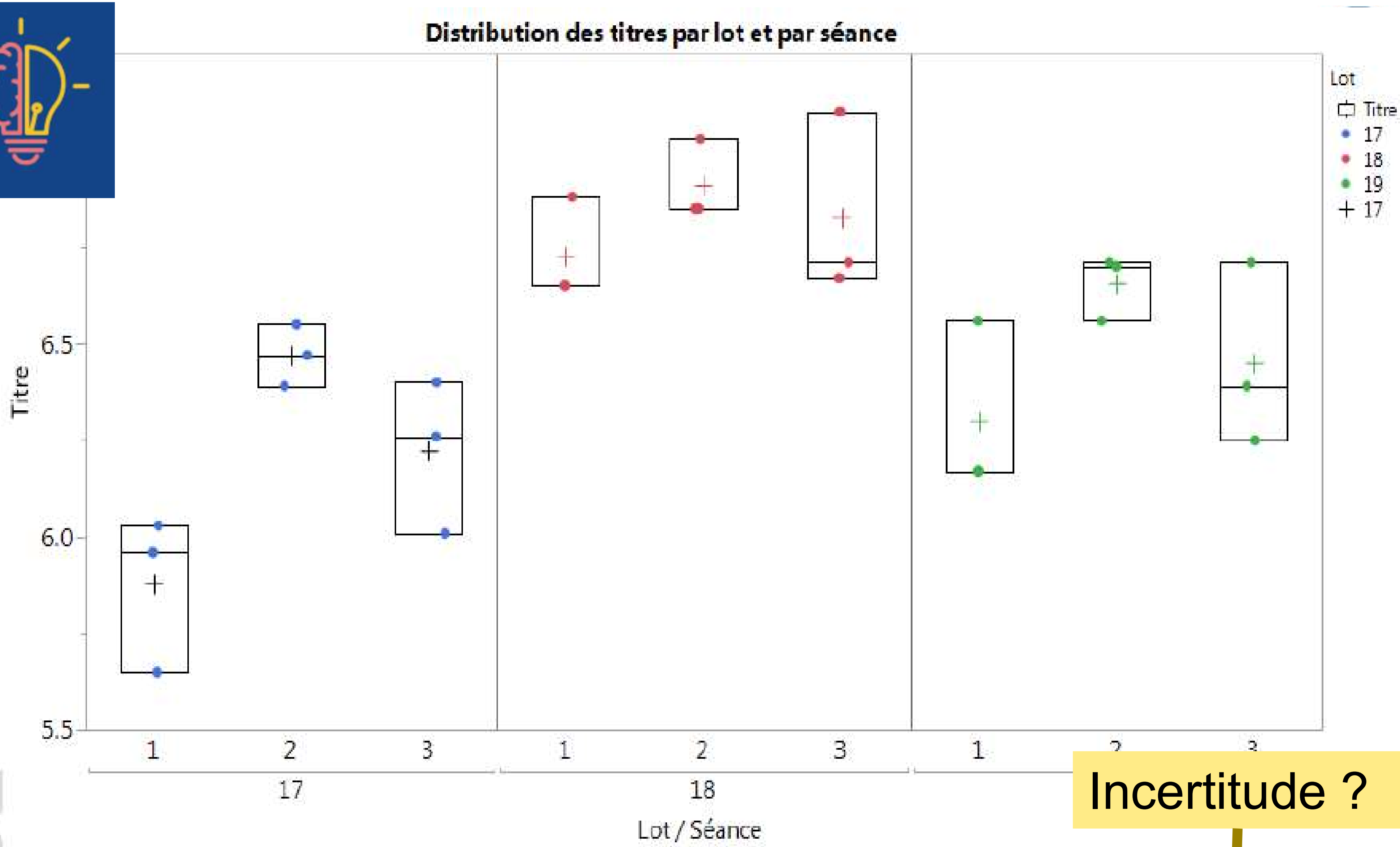
FOOD FOR …

PART III – APPLICATION AU CAS DE MARION

PRISE EN COMPTE DE L'INCERTITUDE DES PARAMÈTRES

IMPACT DE LA CONNAISSANCE (PRIOR)

FOOD FOR …



Incertitude ?

	Variance σ^2	Ecart-type σ
Procédé (Inter-Lots)	0.0854	0.292
Inter-Séances	0.0322	0.180
Répétabilité (Intra-séance)	0.0316	0.178

Variabilité procédé

Variabilité analytique

1- GÉNÉRER DISTRIBUTIONS POSTÉRIEURES

Model: $Titre_{m(ijk)} = \mu_0 + Lot_i + Séance_{j(i)} + error_{k(ij)}$

Paramètres à estimer

mean | σ_{lot} | $\sigma_{séance}$ | $\sigma_{répétabilité}$

```
proc mcmc data=DATA.MARION_25_NEW outpost=Marion_25_NEW seed=1234
nmc=1000000 thin=100 MONITOR=(_parms_) diag=all stats=all;

parms mu0 6.49;
parms S2_Lot 0.0854;
parms S2_Seance 0.0322;
parms S2_error 0.0316;
parms A 1 ;

prior mu0 ~ normal(0, var=1e8);
prior A ~ uniform(0,1000);
prior S2_: ~ cauchy(0, scale=A, lower=0);

RANDOM L ~ normal(mu0, var= S2_Lot) subject = Lot;
RANDOM S ~ normal(L, var= S2_Seance) subject = Day1;

MODEL Titre ~ normal(S, var = S2_error);
run;
```

4- PROC paramètres

3- valeurs initiales

2- Priors

1- Le modèle

SAS 9.4

2- DONNÉES DE SORTIE : CHAINE DE MARKOV

BIOMÉRIEUX

mean | **σ_{lot}** | **$\sigma_{séance}$** | **$\sigma_{répétabilité}$**

	Iteration	mu0	s_lot	s_seance	s_repeat
1	1001	6.6532	0.3990797181	0.1240752092	0.1640690954
2	1101	6.5352	0.1945054049	0.1852791786	0.1570978033
3	1201	6.2303	0.4775682231	0.2134134938	0.1591052591
4	1301	6.8393	0.4912806346	0.1582230222	0.2954878147
5	1401	6.4186	0.171513757	0.1140468156	0.159190667
6	1501	6.4533	0.187686151	0.2823303089	0.1801890789
7	1601	6.6134	0.2446147213	0.3061039876	0.2008923548
8	1701	6.5785	0.3916802342	0.1984442852	0.1465485836
9	1801	6.5683	0.2518248065	0.4085273568	0.1687827917
10	1901	6.2143	0.5358887763	0.1963631856	0.257127908
11	2001	7.0738	0.6386147917	0.5219876882	0.2039584828
12	2101	6.6190	0.5705176676	0.1610307567	0.1769985682
13	2201	6.4879	0.1959115214	0.2562225889	0.2931403136
14	2301	6.5956	0.1091902299	0.2692709157	0.2460540502

...

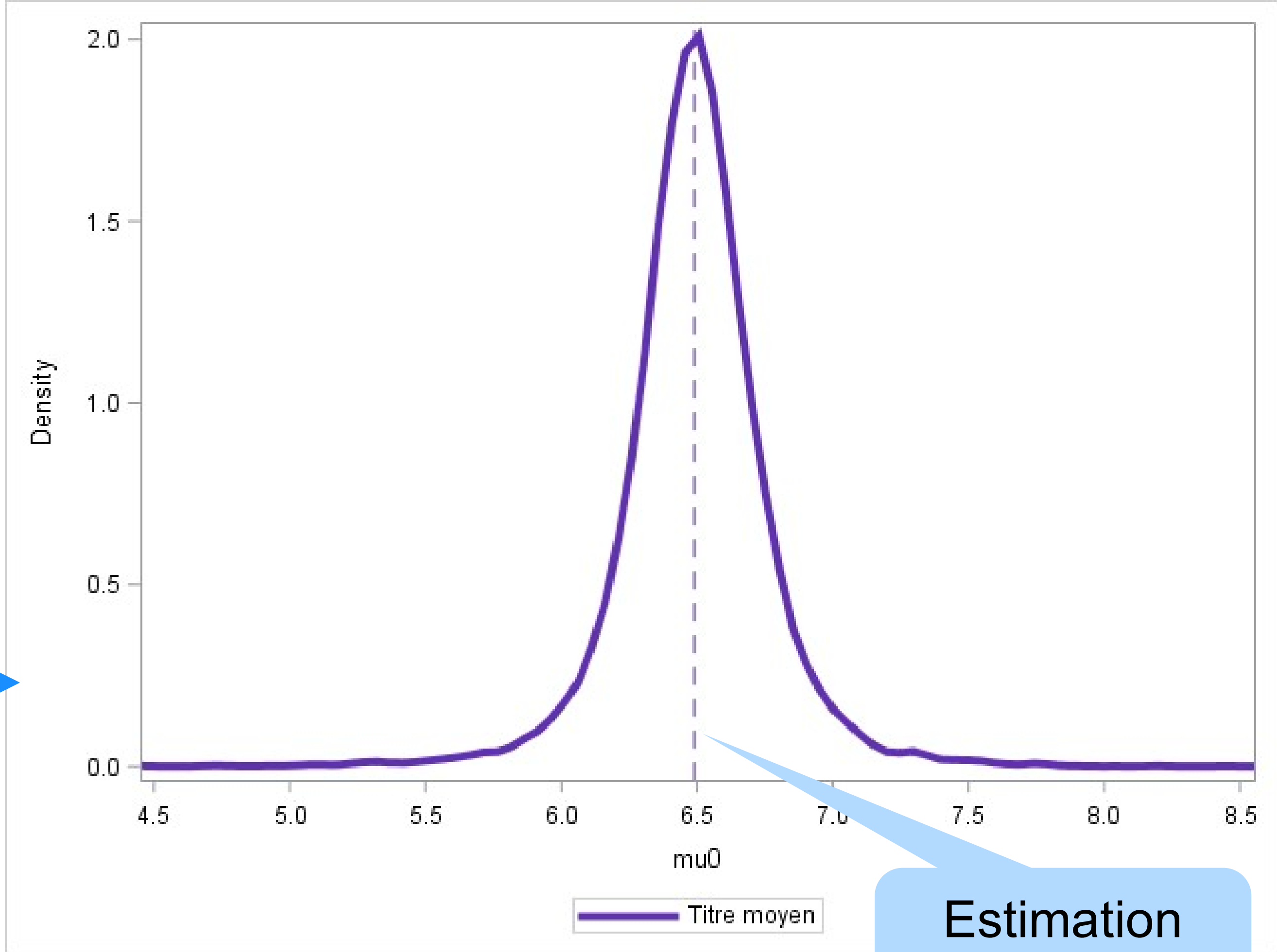
9994	1000301	6.5115	0.1568546515	0.1512094102	0.2051016878
9995	1000401	6.6597	0.12752758	0.3454206521	0.2075730466
9996	1000501	5.6267	0.602579911	0.2347767287	0.1728491937
9997	1000601	6.9091	1.2586401801	0.1777860353	0.2411309557
9998	1000701	6.7589	0.5880779625	0.2892974556	0.206276387
9999	1000801	6.5344	0.1393413588	0.3058158306	0.1637520105
10000	1000901	6.4632	0.4848268942	0.1147327542	0.2203812708

Chaque ligne =
une estimation plausible
(= appartenant à la postérior)
pour chacun des 4
paramètres

DISTRIBUTION POSTÉRIEURE μ_0

mean | σ_{lot} | $\sigma_{séance}$ | $\sigma_{répétabilité}$

	Iteration	mu0	s_lot	s_seance	s_repeat
1	1001	6.6532	0.3990797181	0.1240752092	0.1640690954
2	1101	6.5352	0.1945054049	0.1852791786	0.1570978033
3	1201	6.2303	0.4775692231	0.2134134938	0.1591052591
4	1301	6.8393	0.4912806346	0.1582230222	0.2954878147
5	1401	6.4186	0.171513757	0.1140468156	0.159190667
6	1501	6.4533	0.187686151	0.2823303989	0.1801890789
7	1601	6.6134	0.2446147213	0.3061039876	0.2008923548
8	1701	6.5785	0.3916802342	0.1984442852	0.1465485836
9	1801	6.5683	0.2518248065	0.4085273566	0.1687827917
10	1901	6.2143	0.5358887763	0.1963831856	0.257127908
11	2001	7.0738	0.6386147917	0.5219876882	0.2039584828
12	2101	6.6190	0.5705176676	0.1610307567	0.1769985682
13	2201	6.4879	0.1959115214	0.2562225889	0.2931403136
14	2301	6.5956	0.1091902299	0.2692709157	0.2460540502
...					
9994	1000301	6.5115	0.1568546515	0.1512094102	0.2051016878
9995	1000401	6.6597	0.12752758	0.3454206521	0.2075730466
9996	1000501	5.6267	0.602579911	0.2347767287	0.1728491937
9997	1000601	6.9091	1.2586401801	0.1777860353	0.2411309557
9998	1000701	6.7589	0.5880779625	0.2892974556	0.206276387
9999	1000801	6.5344	0.1393413588	0.3058158306	0.1637520105
10000	1000901	6.4632	0.4848268942	0.1147327542	0.2203812708

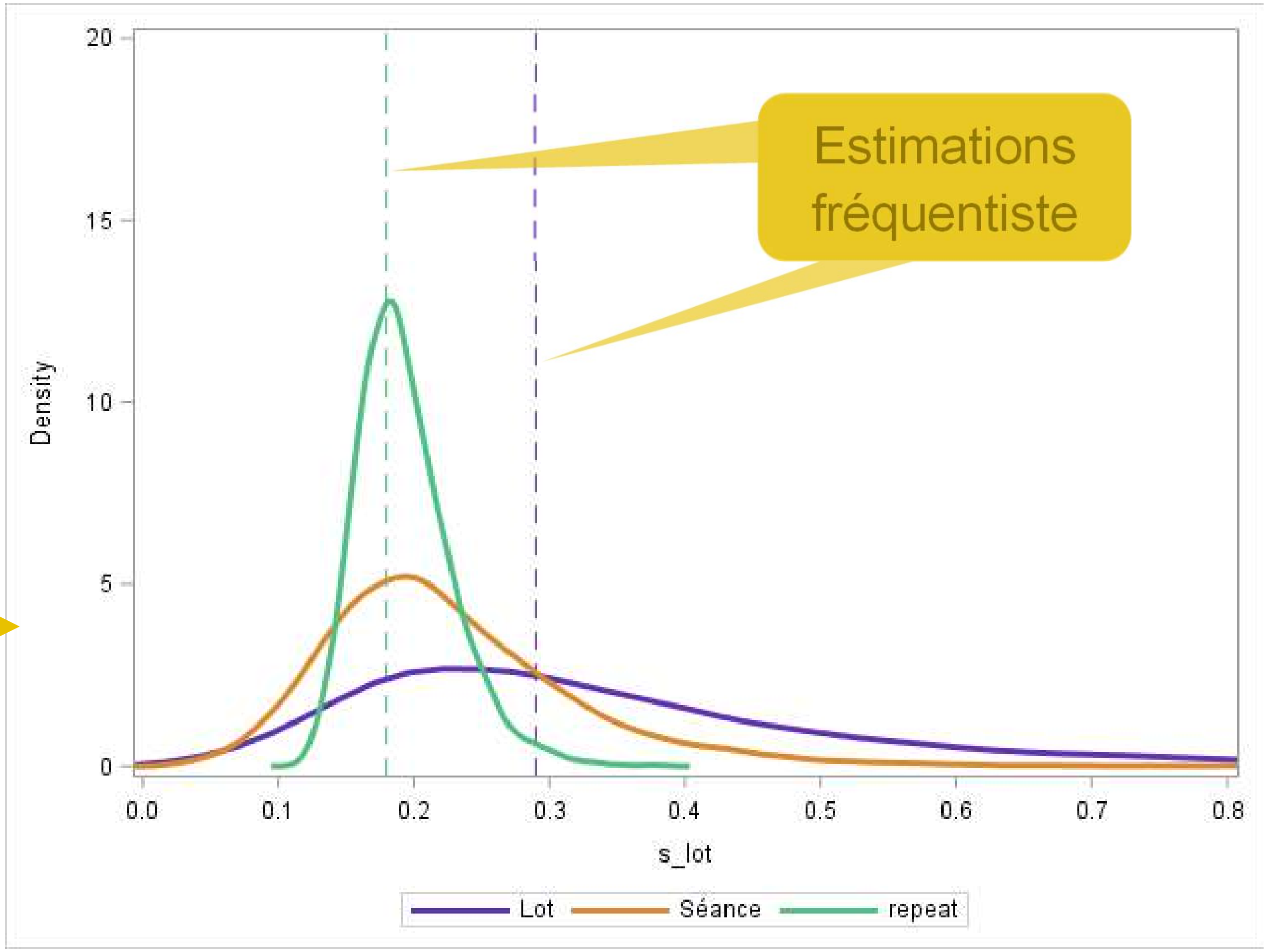


Estimation fréquentiste

DISTRIBUTIONS POSTÉRIEURES σ_i

mean | σ_{lot} | $\sigma_{séance}$ | $\sigma_{répétabilité}$

	Iteration	mu0	s_lot	s_seance	s_repeat
1	1001	6.6532	0.3990797181	0.1240752092	0.1640690954
2	1101	6.5352	0.1945054049	0.1852791786	0.1570978033
3	1201	6.2303	0.4775692231	0.2134134938	0.1591052591
4	1301	6.8393	0.4912806346	0.1582230222	0.2954878147
5	1401	6.4186	0.171513757	0.1140468156	0.159190667
6	1501	6.4533	0.187686151	0.2823303989	0.1801890789
7	1601	6.6134	0.2446147213	0.3061039876	0.2008923548
8	1701	6.5785	0.3916802342	0.1984442852	0.1465485836
9	1801	6.5683	0.2518248065	0.4085273566	0.1687827917
10	1901	6.2143	0.5358887763	0.1963831856	0.257127908
11	2001	7.0738	0.6386147917	0.5219876882	0.2039584828
12	2101	6.6190	0.5705176676	0.1610307567	0.1769985682
13	2201	6.4879	0.1959115214	0.2562225889	0.2931403136
14	2301	6.5956	0.1091902299	0.2692709157	0.2460540502
…					
9994	1000301	6.5115	0.1568546515	0.1512094102	0.2051016878
9995	1000401	6.6597	0.12752758	0.3454206521	0.2075730466
9996	1000501	5.6267	0.602579911	0.2347767287	0.1728491937
9997	1000601	6.9091	1.2586401801	0.1777860353	0.2411309557
9998	1000701	6.7589	0.5880779625	0.2892974556	0.206276387
9999	1000801	6.5344	0.1393413588	0.3058158306	0.1637520105
10000	1000901	6.4632	0.4848268942	0.1147327542	0.2203812708



DISTRIBUTIONS POSTÉRIEURES

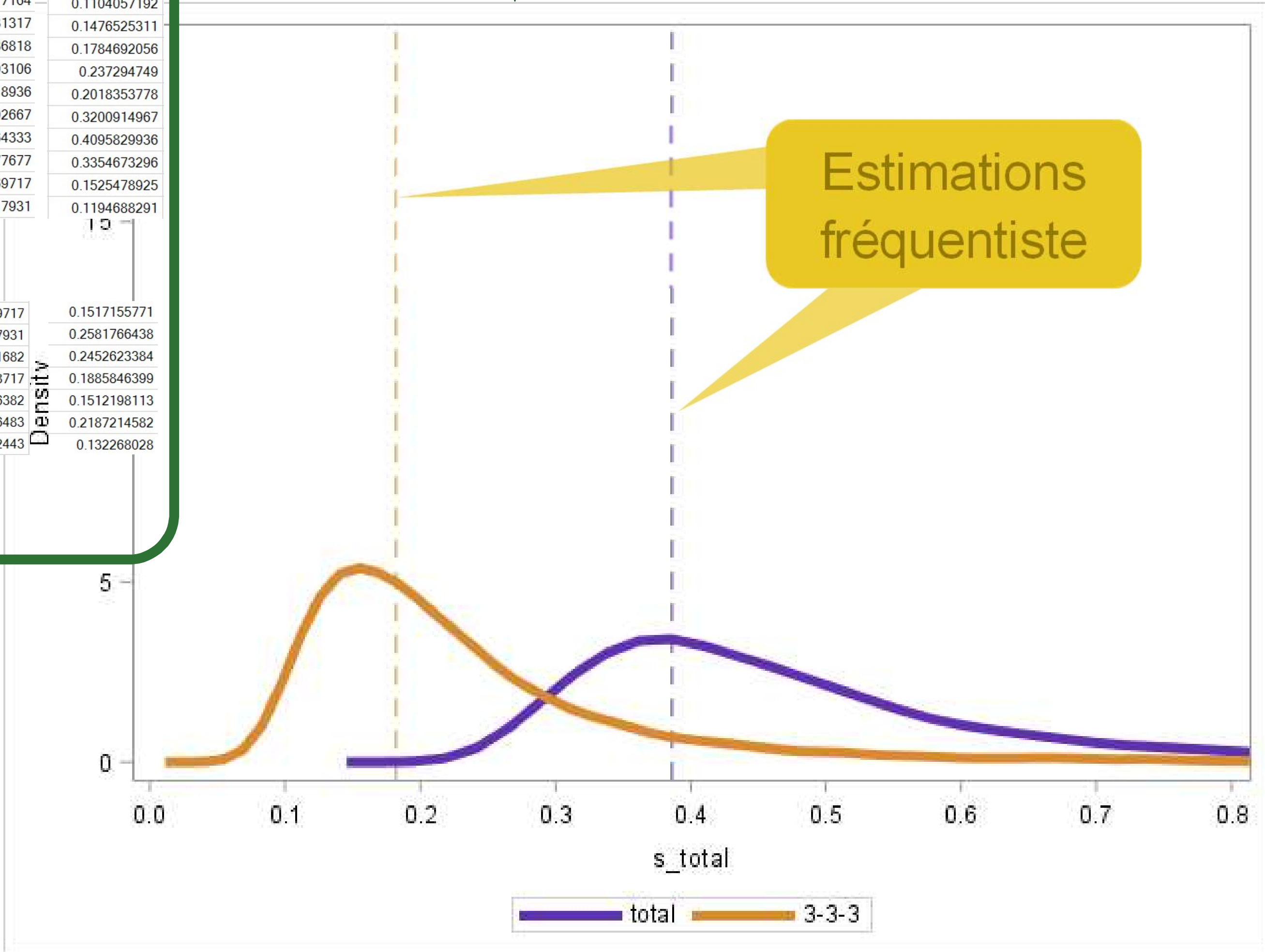
σ_{total}

mean | σ_{lot} | $\sigma_{séance}$ | $\sigma_{répétabilité}$

σ_{total} N=1

σ_{total} N=27 (3,3,3)

	Iteration	mu0	s_lot	s_seance	s_repeat	s_total	s_scenario1
1	1001	6.6532	0.3990797181	0.1240752092	0.1640690954	0.4489743278	0.2362111605
2	1101	6.5352	0.1945054049	0.1852791786	0.1570978033	0.3111919767	0.1316780606
3	1201	6.2303	0.4775692231	0.2134134938	0.1591052591	0.5467468936	0.2863953307
4	1301	6.8393	0.4912806346	0.1582230222	0.2954878147	0.5947303888	0.2940538235
5	1401	6.4186	0.171513757	0.1140468156	0.159190667	0.2603177164	0.1104057192
6	1501	6.4533	0.187686151	0.2823303989	0.1801890789	0.3839331317	0.1476525311
7	1601	6.6134	0.2446147213	0.3061039876	0.2008923548	0.4403336818	0.1784692056
8	1701	6.5785	0.3916802342	0.1984442852	0.1465485836	0.462893106	0.237294749
9	1801	6.5683	0.2518248065	0.4085273566	0.1687827917	0.5087218936	0.2018353778
10	1901	6.2143	0.5358887763	0.1963831856	0.257127908	0.6259792667	0.3200914967
11	2001	7.0738	0.6386147917	0.5219876882	0.2039584828	0.8496464333	0.4095829936
12	2101	6.6190	0.5705176676	0.1610307567	0.1769985682	0.6186677677	0.3354673296
13	2201	6.4879	0.1959115214	0.2562225889	0.2931403136	0.4358469717	0.1525478925
14	2301	6.5956	0.1091902299	0.2692709157	0.2460540502	0.3807517931	0.1194688291
...							
9994	1000301	6.5115	0.1568546515	0.1512094102	0.2051016878	0.4358469717	0.1517155771
9995	1000401	6.6597	0.12752758	0.3454206521	0.2075730466	0.3807517931	0.2581766438
9996	1000501	5.6267	0.602579911	0.2347767287	0.1728491937	0.3460291682	0.2452623384
9997	1000601	6.9091	1.2586401801	0.1777860353	0.2411309557	0.5081618717	0.1885846399
9998	1000701	6.7589	0.5880779625	0.2892974556	0.206276387	0.4796436382	0.1512198113
9999	1000801	6.5344	0.1393413588	0.3058158306	0.1637520105	0.3905306483	0.2187214582
10000	1000901	6.4632	0.4848268942	0.1147327542	0.2203812708	0.3762442443	0.132268028



BIOMÉRIEUX

QUELLE EST LA QUESTION ?

Procédé de référence
Ex: θ=33°
Titre moyen M1

Nouveau procédé
Ex: θ=37°
Titre moyen M2

- **Sous l'hypothèse M2 = M1** **et** $\sigma_{process1} = \sigma_{process2}$
- **Avec un protocole de comparaison N= 27 par condition**

3 lots, 3 séances/lot, 3 répétitions/séance

Question 1

- Quelle est la **distribution des valeurs possibles** pour **y2 – y1 ?**

DISTRIBUTION PRÉDICTIVE

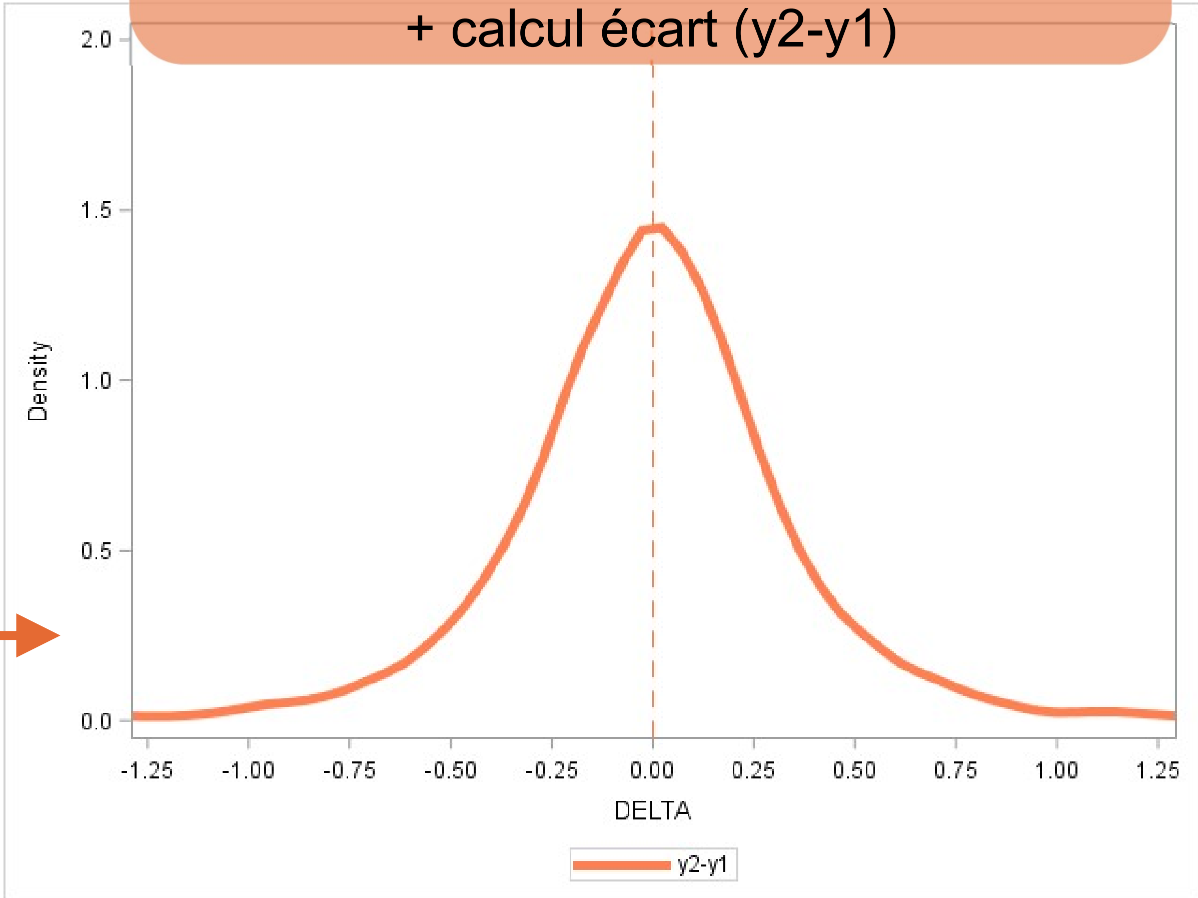
Question 1

y1-y2

mean | σ_{lot} | $\sigma_{séance}$ | $\sigma_{répétabilité}$

	Iteration	mu0	s_lot	s_seance	s_repeat	DELTA
1	1001	6.6532	0.3990797181	0.1240752092	0.1640690954	0.0808736846
2	1101	6.5352	0.1945054049	0.1852791786	0.1570978033	-0.402231732
3	1201	6.2303	0.4775692231	0.2134134938	0.1591052591	0.0465058963
4	1301	6.8393	0.4912806346	0.1582230222	0.2954878147	-0.04422647
5	1401	6.4186	0.171513757	0.1140468156	0.159190667	-0.038926298
6	1501	6.4533	0.187686151	0.2823303989	0.1801890789	-0.240210826
7	1601	6.6134	0.2446147213	0.3061039876	0.2008923548	0.0891165209
8	1701	6.5785	0.3916802342	0.1984442852	0.1465485836	0.2773856183
9	1801	6.5683	0.2518248065	0.4085273566	0.1687827917	0.3581432856
10	1901	6.2143	0.5358887763	0.1963831856	0.257127908	-0.298977499
11	2001	7.0738	0.6386147917	0.5219876882	0.2039584828	1.0552713392
12	2101	6.6190	0.5705176676	0.1610307567	0.1769985682	0.3732887521
13	2201	6.4879	0.1959115214	0.2562225889	0.2931403136	0.1334114669
14	2301	6.5956	0.1091902299	0.2692709157	0.2460540502	0.2676634326
...						
9994	1000301	6.5115	0.1568546515	0.1512094102	0.2051016878	-0.281040431
9995	1000401	6.6597	0.12752758	0.3454206521	0.2075730466	-0.353678377
9996	1000501	5.6267	0.602579911	0.2347767287	0.1728491937	-0.212580808
9997	1000601	6.9091	1.2586401801	0.1777860353	0.2411309557	0.1326335038
9998	1000701	6.7589	0.5880779625	0.2892974556	0.206276387	0.0966489387
9999	1000801	6.5344	0.1393413588	0.3058158306	0.1637520105	0.3982779062
10000	1000901	6.4632	0.4848268942	0.1147327542	0.2203812708	0.1121588784

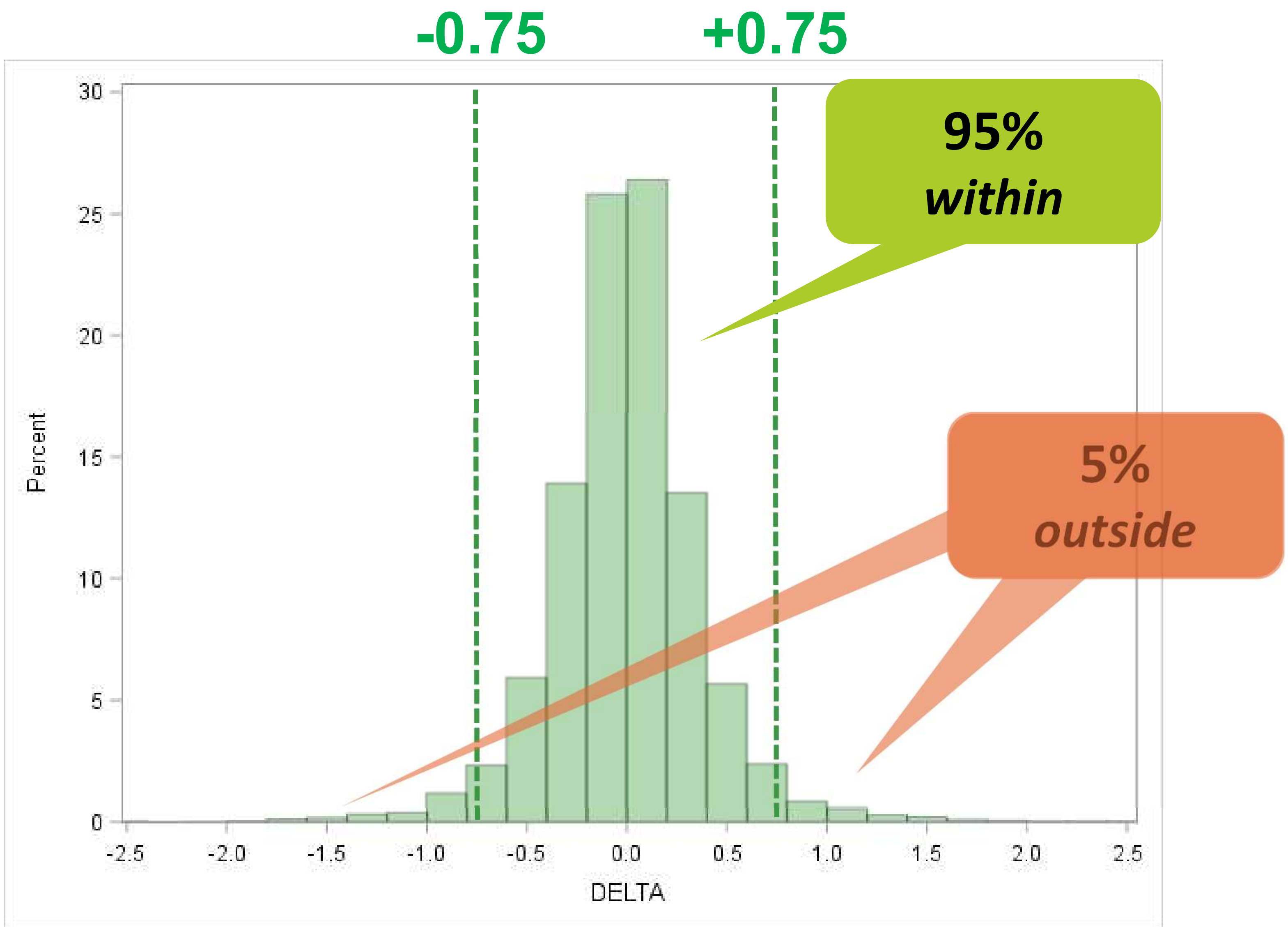
Pour chaque ligne, simulations une étude future (N= 27) pour condition 1 et une étude future (N= 27) pour condition 2 (**avec hypothèse M2 = M1**)
+ calcul écart (y2-y1)



INFÉRENCE

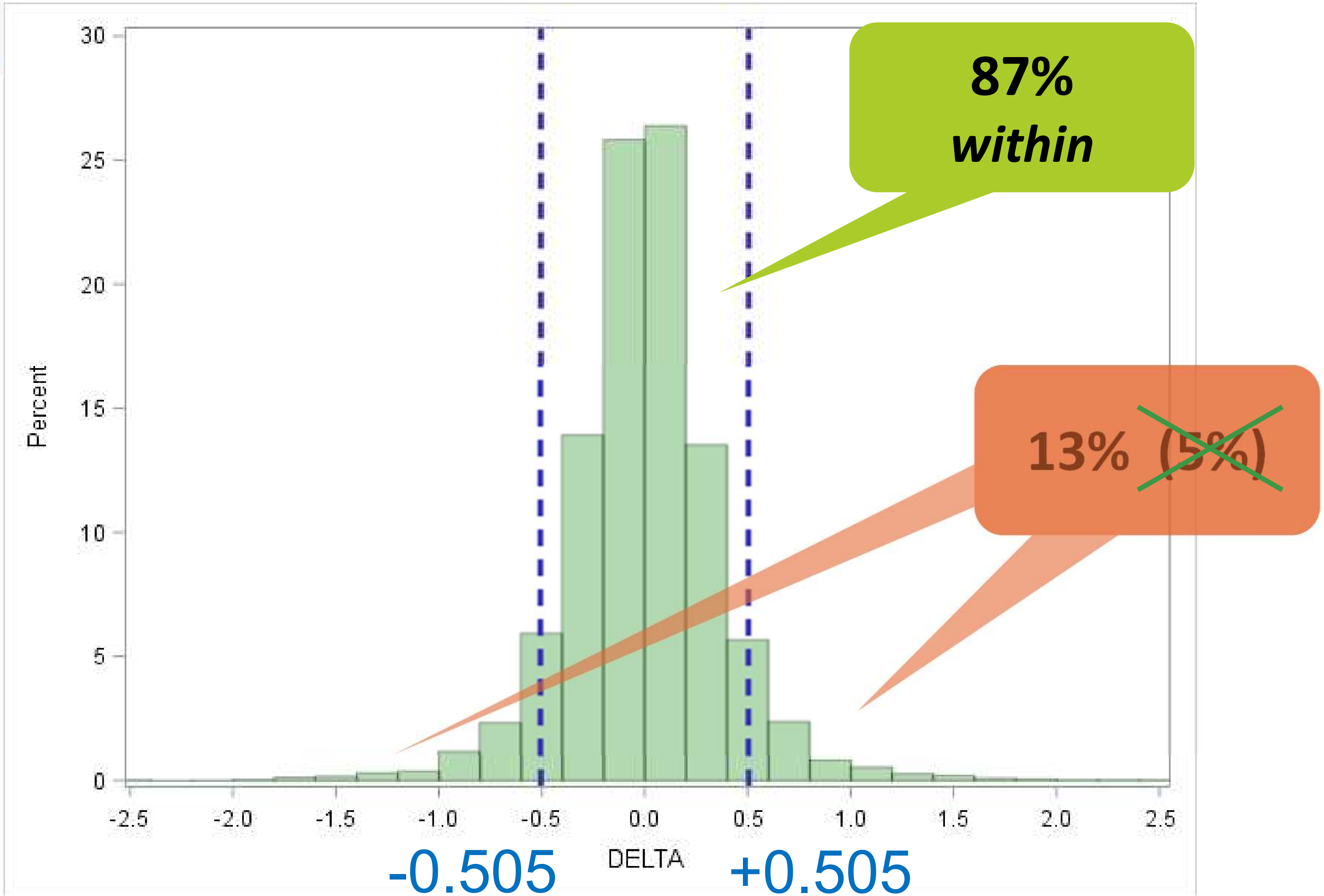
Question 1

Percentiles 2.5% et 97.5% ?   *intervalle de crédibilité*



RAPPELS CALCULS
SANS CONSIDÉRER INCERTITUDE

Avec les seuils ± 0.505, le risque α n'est pas de 5% mais ≈ 13%



DISTRIBUTIONS PRÉDICTIVES

QUELLE EST LA QUESTION ?

Procédé de référence
Ex: θ=33°
Titre moyen M1

Nouveau procédé
Ex: θ=37°
Titre moyen M2

- **Sous l'hypothèse M2 = (M1 +0.4) et** $\sigma_{process1} = \sigma_{process2}$
- **Avec un protocole de comparaison N= 27 par condition**

3 lots, 3 séances/lot, 3 répétitions/séance

Question 2

- Quelle est la **distribution des valeurs possibles** pour **y2 – y1 ?**

DISTRIBUTION PRÉDICTIVE

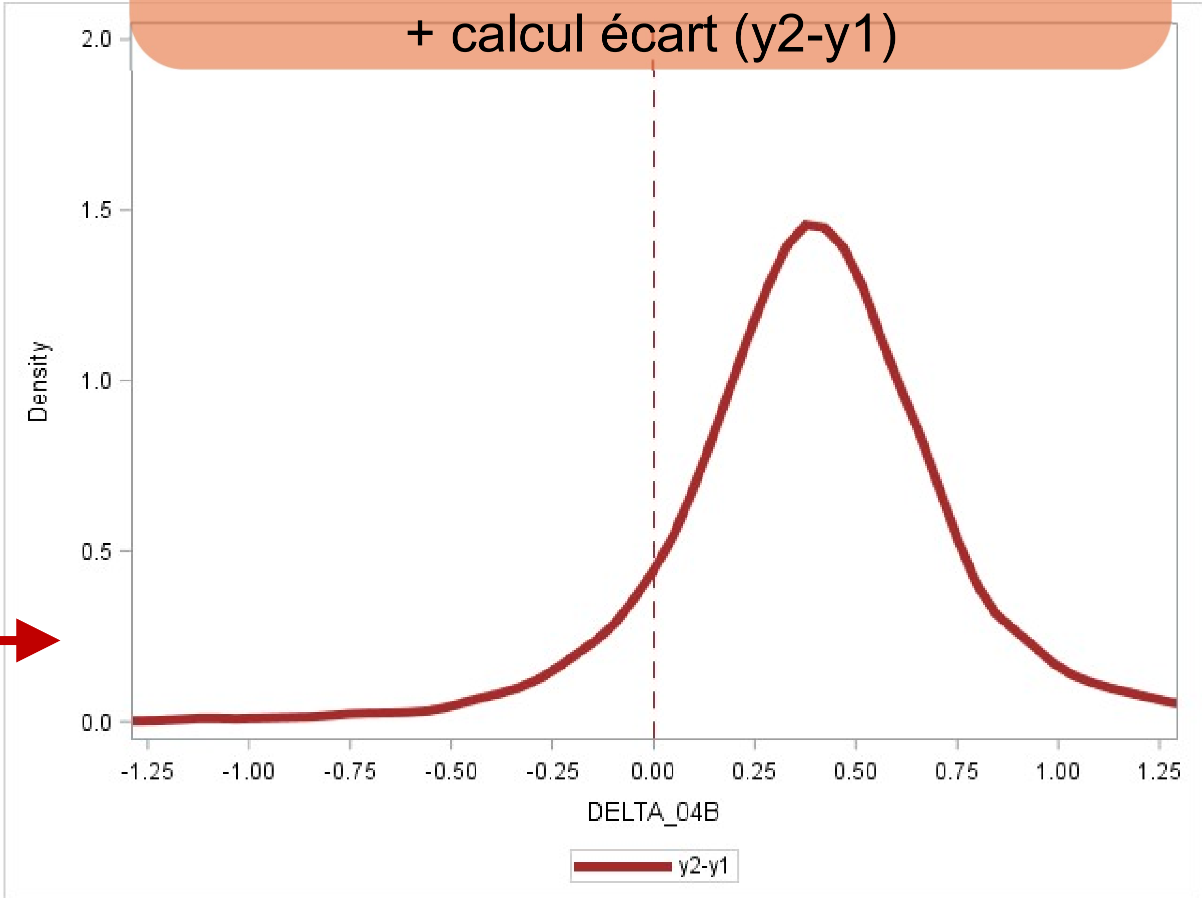
Question 2

y1-y2

	mean	σ_{lot}	$\sigma_{séance}$	$\sigma_{répétabilité}$	y1-y2

	Iteration	mu0	s_lot	s_seance	s_repeat	DELTA_04B
1	1001	6.6532	0.3990797181	0.1240752092	0.1640690954	-0.251398556
2	1101	6.5352	0.1945054049	0.1852791786	0.1570978033	0.1481620581
3	1201	6.2303	0.4775692231	0.2134134938	0.1591052591	0.304214423
4	1301	6.8393	0.4912806346	0.1582230222	0.2954878147	0.4057496639
5	1401	6.4186	0.171513757	0.1140468156	0.159190667	0.4236259851
6	1501	6.4533	0.187686151	0.2823303989	0.1801890789	0.2631477351
7	1601	6.6134	0.2446147213	0.3061039876	0.2008923548	-0.105938608
8	1701	6.5785	0.3916802342	0.1984442852	0.1465485836	0.4315266131
9	1801	6.5683	0.2518248065	0.4085273566	0.1687827917	0.429242011
10	1901	6.2143	0.5358887763	0.1963831856	0.257127908	0.6206906768
11	2001	7.0738	0.6386147917	0.5219876882	0.2039584828	1.7368979633
12	2101	6.6190	0.5705176676	0.1610307567	0.1769985682	0.3220126733
13	2201	6.4879	0.1959115214	0.2562225889	0.2931403136	0.3815994367
14	2301	6.5956	0.1091902299	0.2692709157	0.2460540502	0.5279214391
...						
9994	1000301	6.5115	0.1568546515	0.1512094102	0.2051016878	0.3220126733
9995	1000401	6.6597	0.12752758	0.3454206521	0.2075730466	0.3815994367
9996	1000501	5.6267	0.602579911	0.2347767287	0.1728491937	0.5279214391
9997	1000601	6.9091	1.2586401801	0.1777860353	0.2411309557	0.2885711474
9998	1000701	6.7589	0.5880779625	0.2892974556	0.206276387	-0.153560051
9999	1000801	6.5344	0.1393413588	0.3058158306	0.1637520105	0.2483963686
10000	1000901	6.4632	0.4848268942	0.1147327542	0.2203812708	0.2354917692

Pour chaque ligne, simulations une étude future (N= 27) pour condition 1 et une étude future (N= 27) pour condition 2 (**avec hypothèse M2 = M1+0.4**)
+ calcul écart (y2-y1)

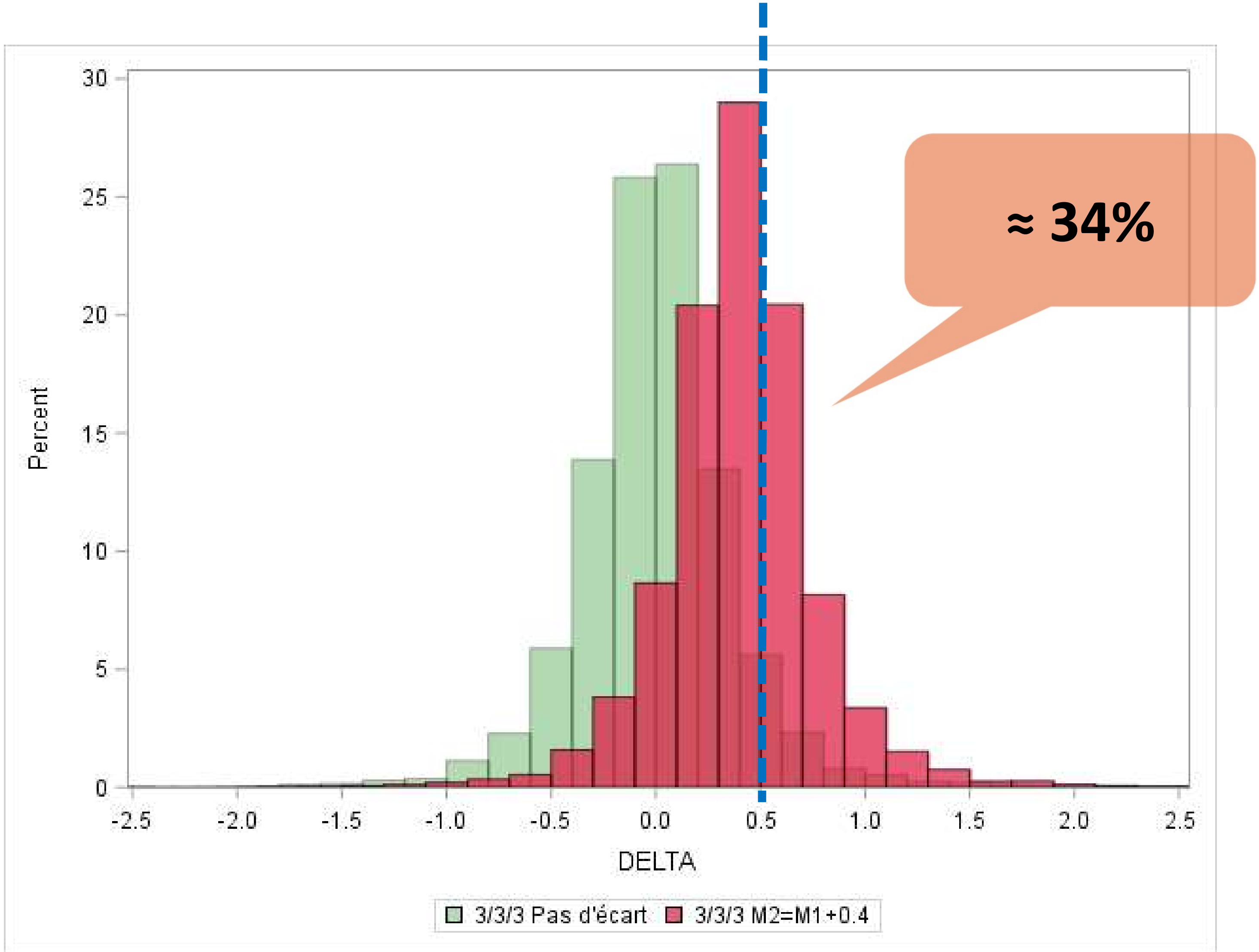


DISTRIBUTION PRÉDICTIVE DES DIFFÉRENCES

Y2 – Y1

Hypothèse M2 = M1+0.4

Seuil fréquentiste : +0.505

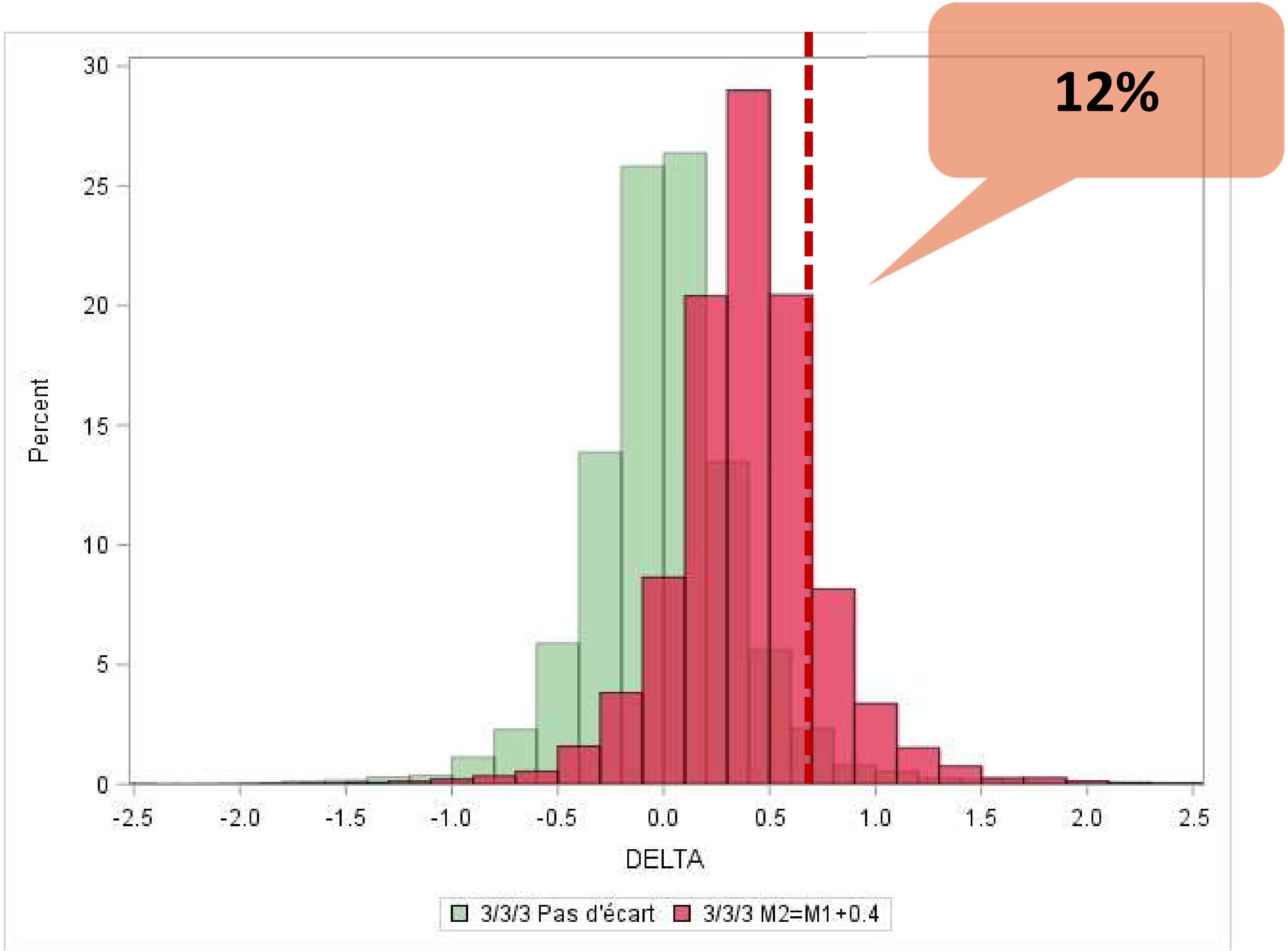


DISTRIBUTION PRÉDICTIVE DES DIFFÉRENCES

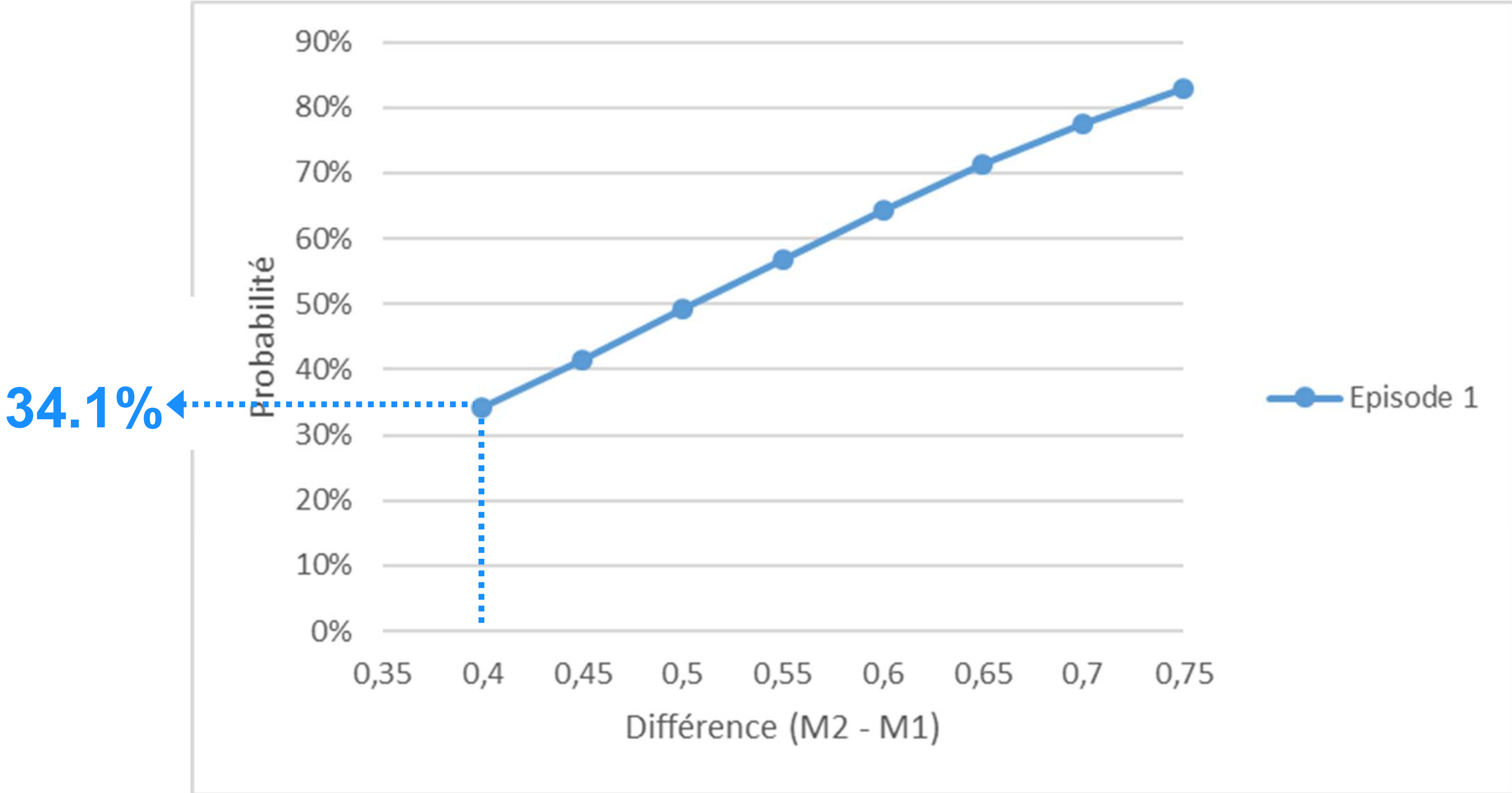
Y2 – Y1 Hypothèse M2 = M1+0.4

Intervalle de crédibilité

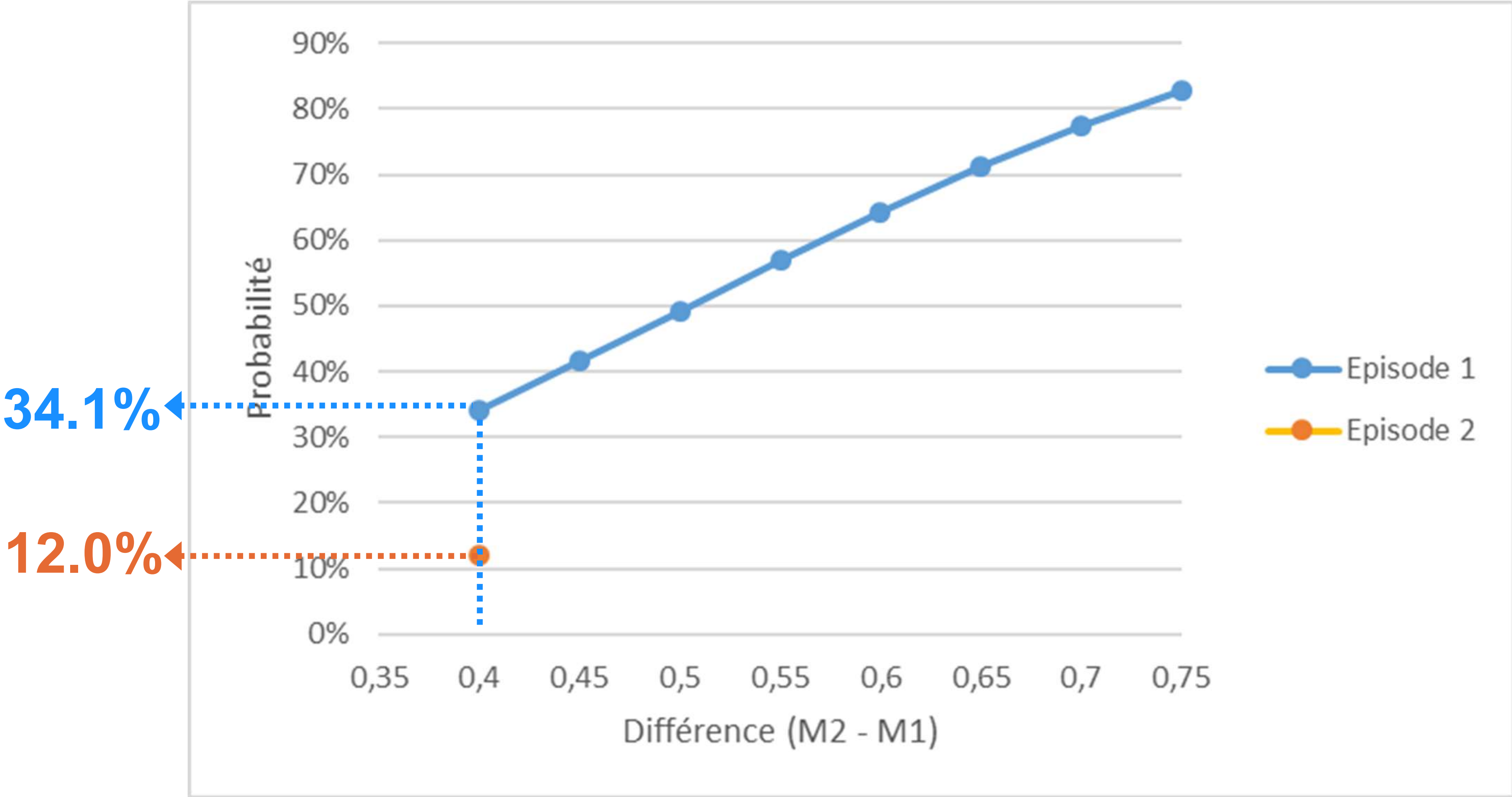
Percentile 97.5% : +0.75



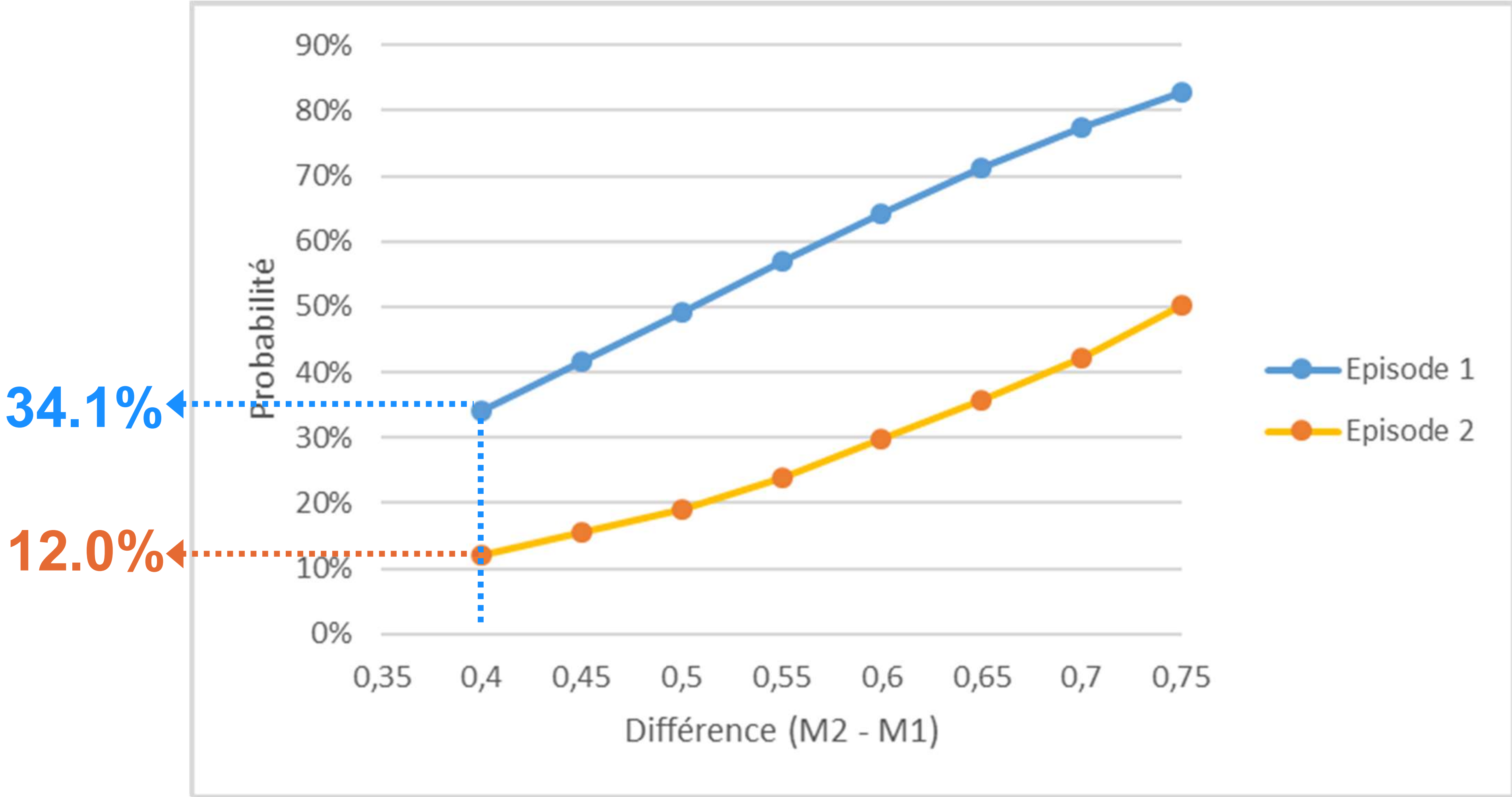
COURBES DE PUISSANCE



COURBES DE PUISSANCE



COURBES DE PUISSANCE



PART III – APPLICATION AU CAS DE MARION

PRISE EN COMPTE DE L'INCERTITUDE DES PARAMÈTRES

IMPACT DE LA CONNAISSANCE (PRIOR)

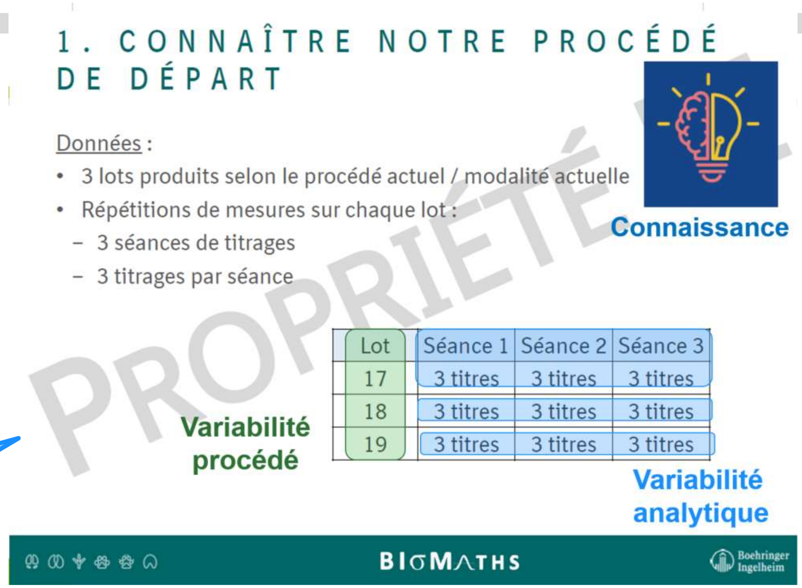
FOOD FOR …

INFLUENCE DE LA CONNAISSANCE

M= 6.49

	Variance σ^2	Ecart-type σ
Procédé (Inter-Lots)	0.0854	0.292
Inter-Séances	0.0322	0.180
Répétabilité (Intra-séance)	0.0316	0.178

Et s'il y avait plus de data
(lots, séances) pour estimer σ_i ?



- **Simulation d'une base de connaissance centrée sur les mêmes estimations (M et σ^2_i) avec :**
  - 25 LOTS
  - 10 séances de titrages par lot
  - 3 répétitions par séance

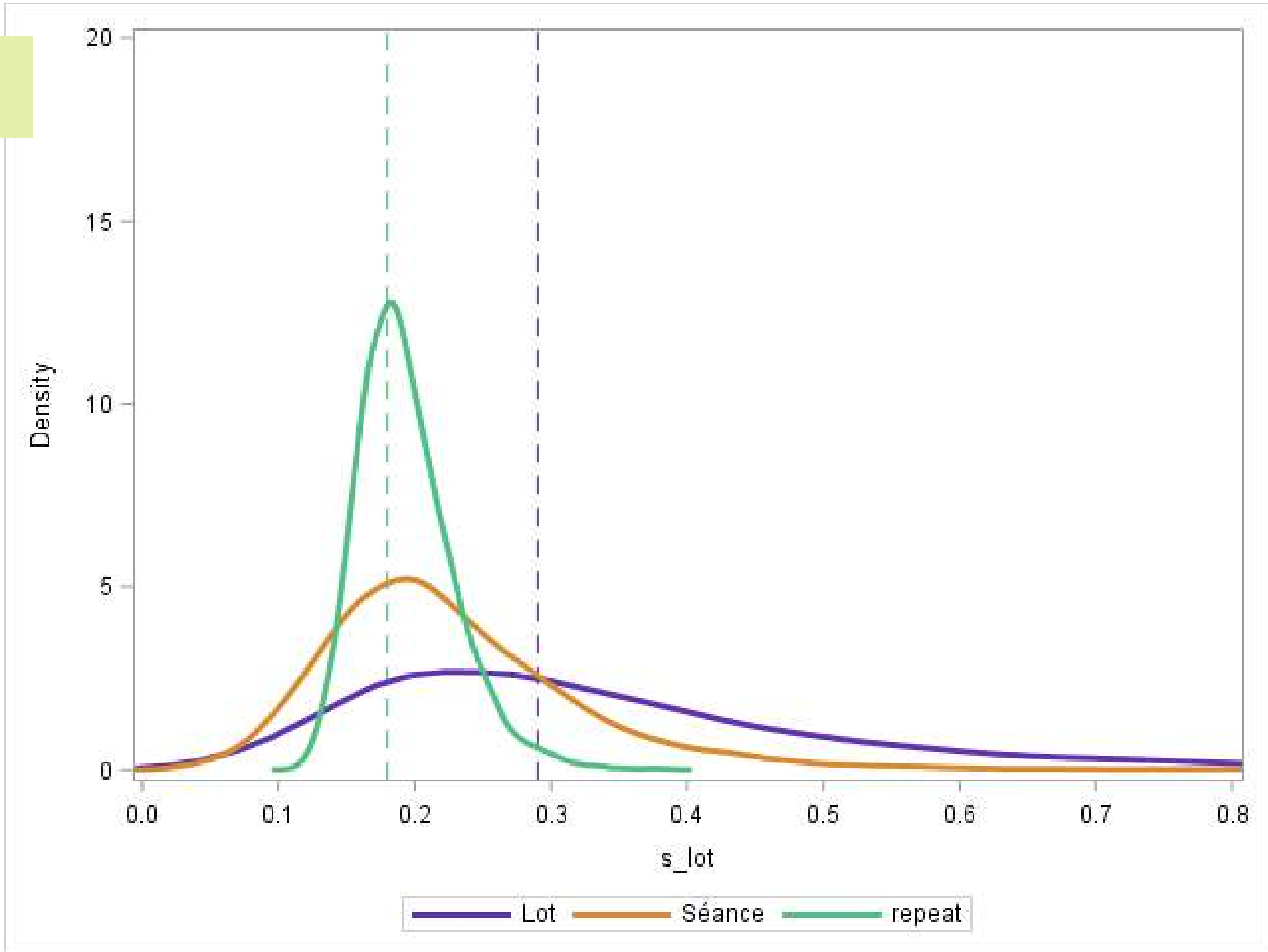
N= 750

Génération de postérieures et prédictives avec cette base de connaissance

DISTRIBUTIONS POSTÉRIEURES

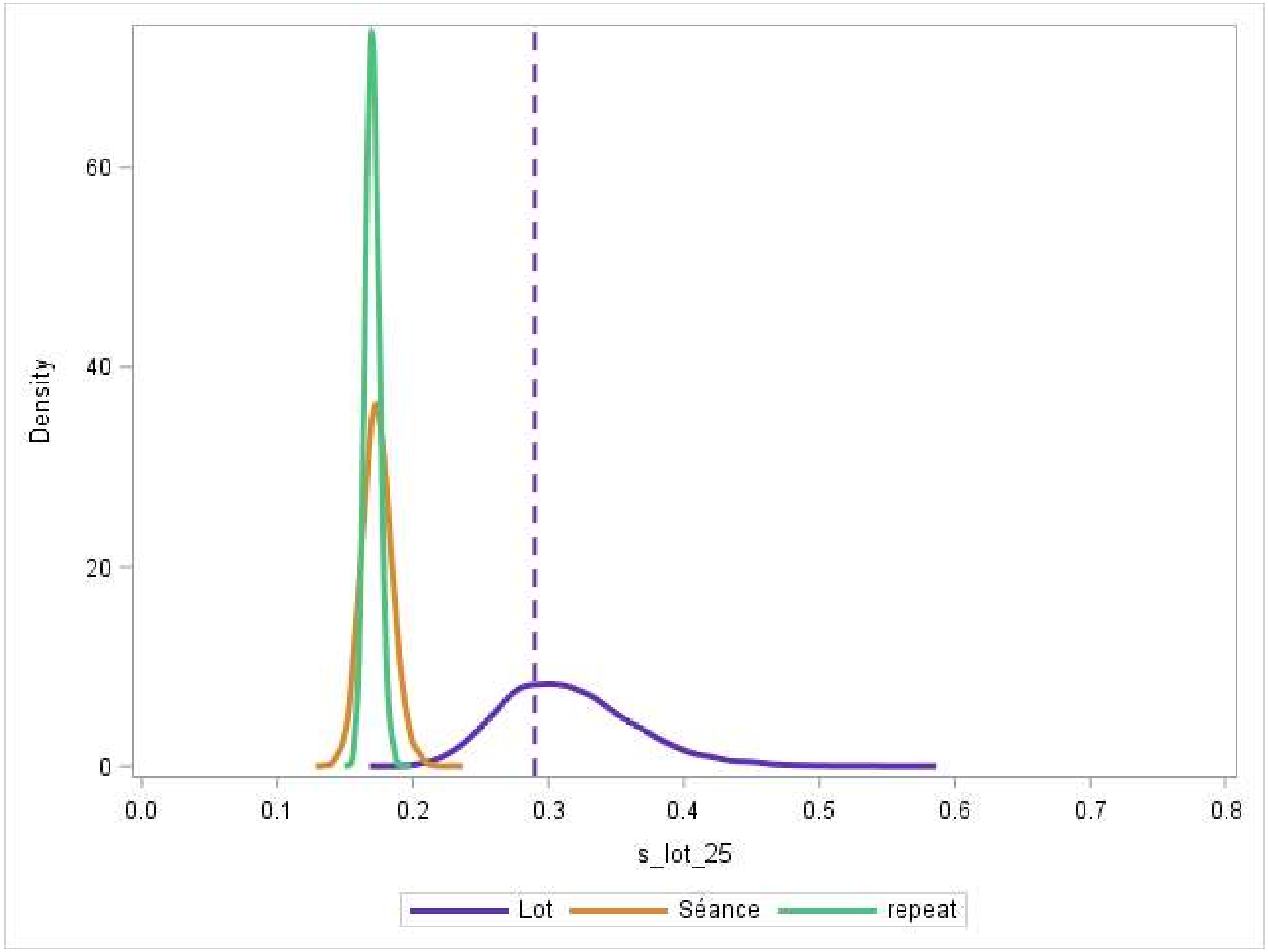
σ_i

N= 27



DISTRIBUTIONS POSTÉRIEURES

σ_i

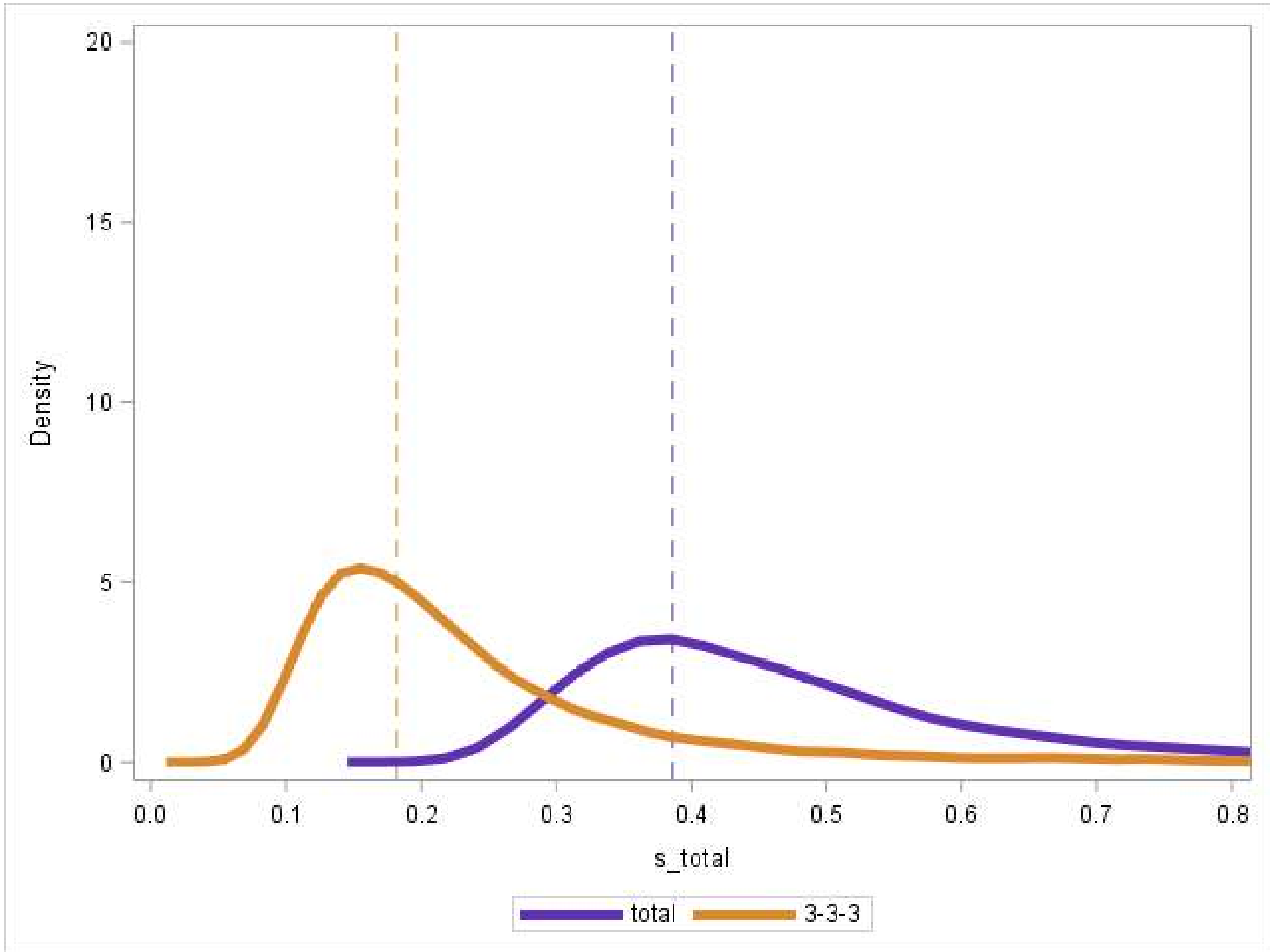


N= 750

DISTRIBUTIONS POSTÉRIEURES

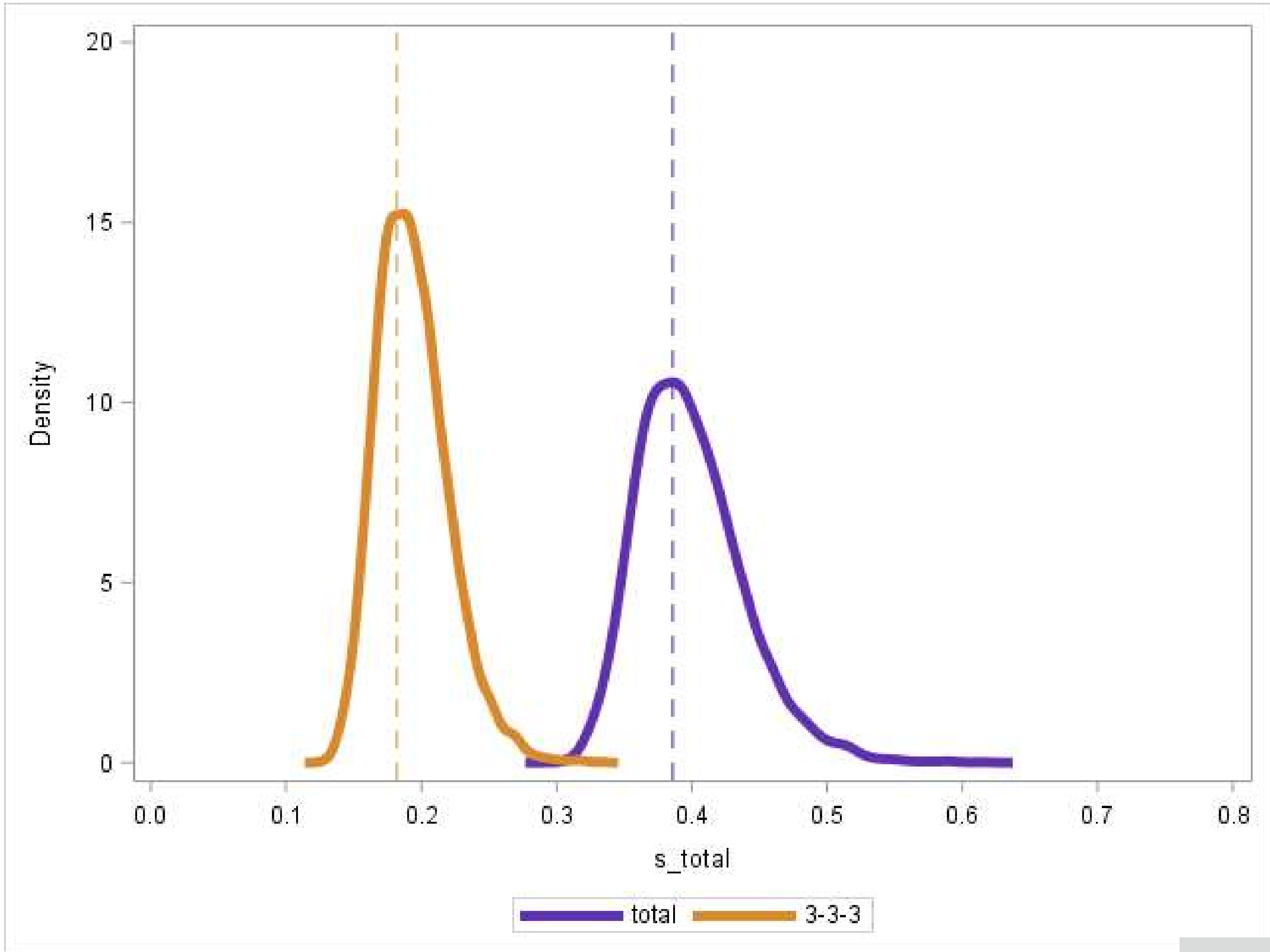
σ_{total}

N= 27



DISTRIBUTIONS POSTÉRIEURES

σ_{total}



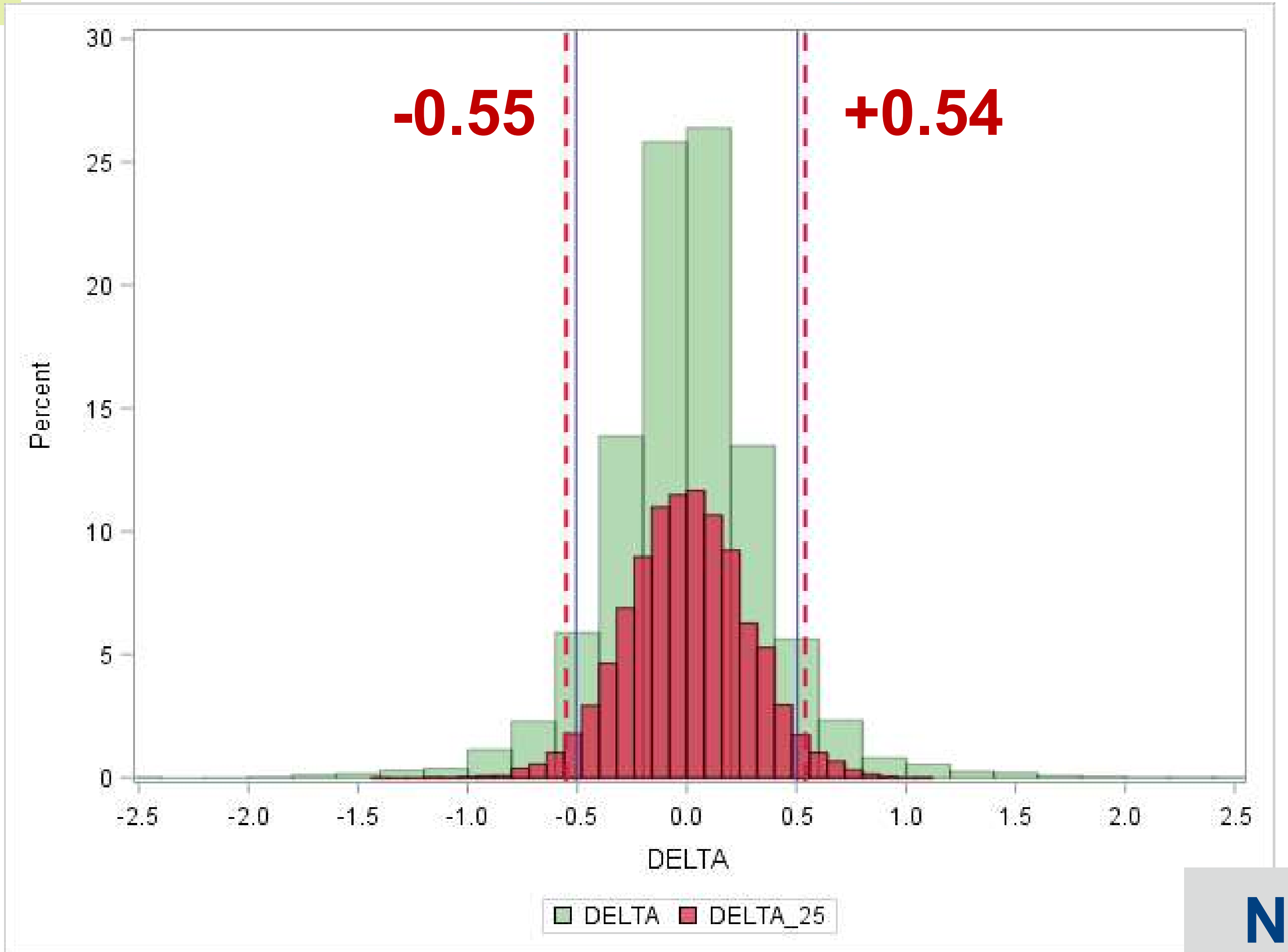
N= 750

INTERVALLE DE CRÉDIBILITÉ

Question 1

N= 27

-0.75 +0.75



N= 750

CAPITALISER SUR LA CONNAISSANCE

Model: $Titre_{m(ijk)} = \mu_0 + Lot_i + Séance_{j(i)} + error_{k(ij)}$

Paramètres à estimer

mean | **σ_{lot}** | **$\sigma_{séance}$** | **$\sigma_{répétabilité}$**

```
proc mcmc data=DATA.MARION_25_NEW
nmc=1000000 thin=100 MONITOR=(_pa

parms mu0 6.49;
parms S2_Lot 0.0854;
parms S2_Seance 0.0322;
parms S2_error 0.0316;
parms A 1 ;

prior mu0 ~ normal(0, var=1e8);
prior A ~ uniform(0,1000);
prior S2_: ~ cauchy(0, scale=A, lower=0);

RANDOM L ~ normal(mu0, var= S2_Lot) subject = Lot;
RANDOM S ~ normal(L, var= S2_Seance) subject = Da

MODEL Titre ~ normal(S, var = S2_error);
run;
```

DATA historique CQ / Prod

$\sigma_{process\ (lot)}$

2- Priors

Données de validation

$\sigma_{analytique\ (séance,\ répétabilité)}$

PART III – APPLICATION AU CAS DE MARION

PRISE EN COMPTE DE L'INCERTITUDE DES PARAMÈTRES

IMPACT DE LA CONNAISSANCE (PRIOR)

QUELQUES REFLEXIONS

PRISE EN COMPTE DE L'INCERTITUDE
AVANTAGES DU BAYÉSIEN

Frequentist / Sattherwhaite's method

$$\pm\ 1.96 \times \sqrt{2}\ \sigma' = \pm\ \mathbf{0.505}$$

$\sigma' = 0.182$ pour p=k=n=3

- A partir de la structure des données de connaissance (3 lots * 3 séances * 3 répétitions) + composantes de variances estimées, le nombre de ddl associés à l'estimation de σ_{total} est de **4 ddl**.
- t(0.975 ; 4ddl) = 2.776

$$\pm\ 2.776 \times \sqrt{2}\ \sigma' = \pm\ \mathbf{0.72}$$

$\sigma' = 0.182$ pour p=k=n=3

- Rappel : intervalle de crédibilité Bayesien : ± 0.75

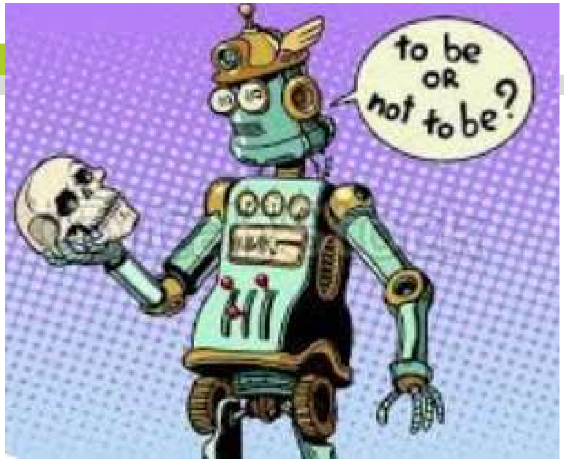
Apport du Bayesien : calcul facilité

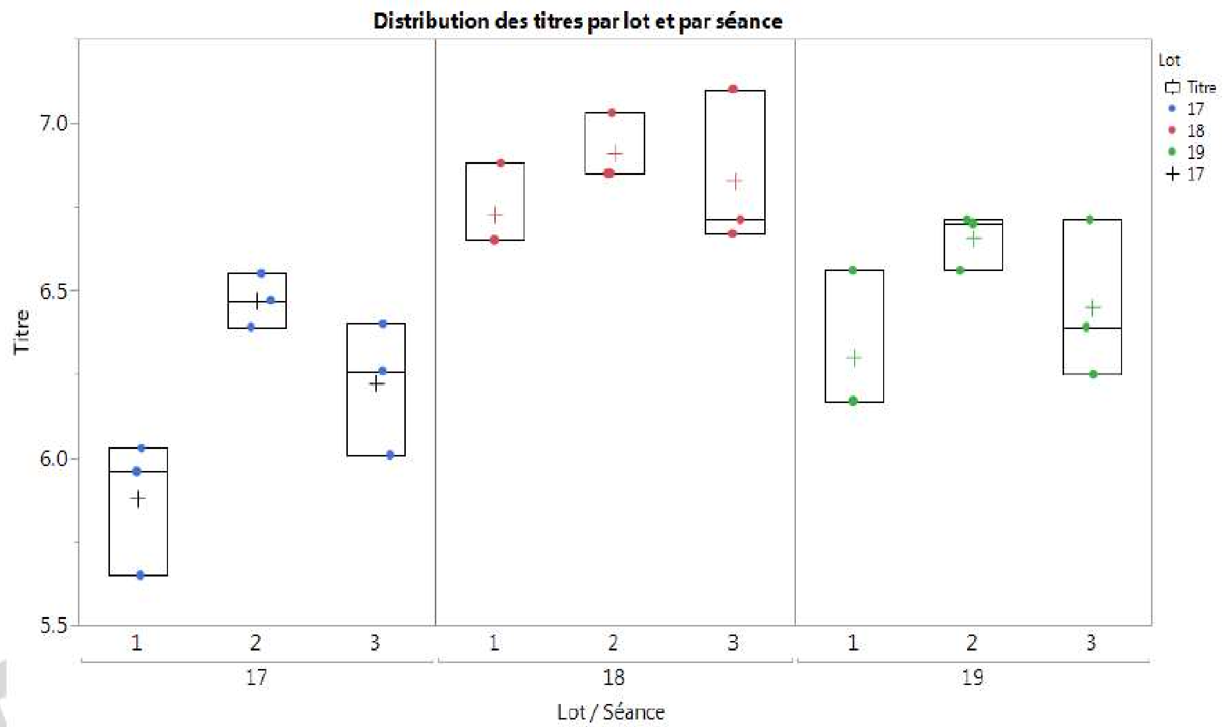
- Et si les données expérimentales entre conditions 1 et 2 sont déséquilibrées ? *Données manquante, 1 séance ratée*

Que pensez vous du calcul fréquentiste ? Et du process Bayésien ?

Avec le process Bayésien, il est aisé de générer, à partir des postérieures, des résultats d'études futures avec des données déséquilibrées.

TO BE OR NOT TO BE INDEPENDENT ?





Indépendance entre les séances de titrage ?

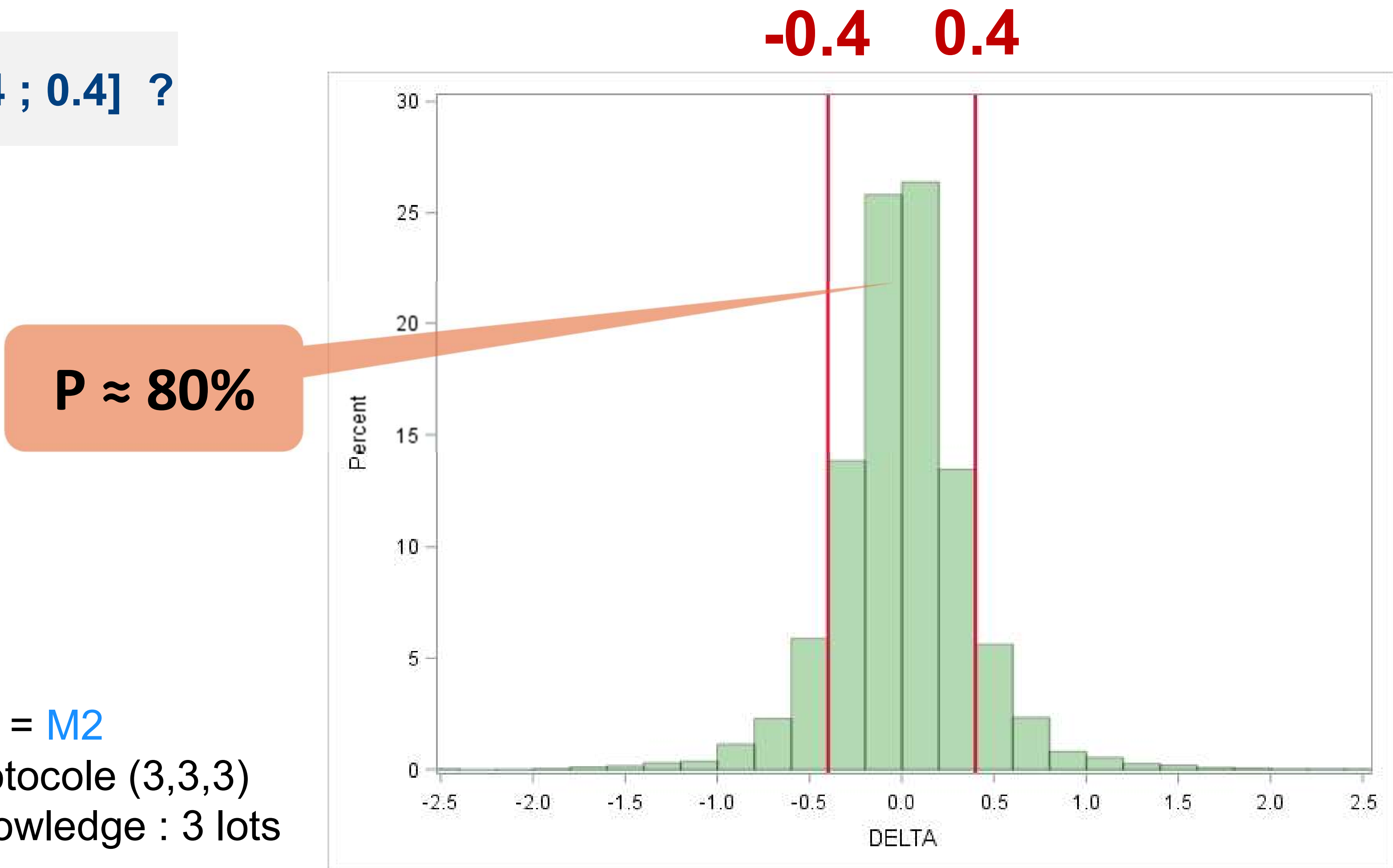
- Et si les séances 1, 2, 3 pour le titrage des lots étaient les mêmes pour les 2 conditions ? Quel serait l'impact sur le calcul de l'incertitude associée à (y2 – y1) ?

➡ Le process Bayésien peut aussi aider à quantifier le bénéfice lié au design expérimental

SEUILS DE DÉCISION

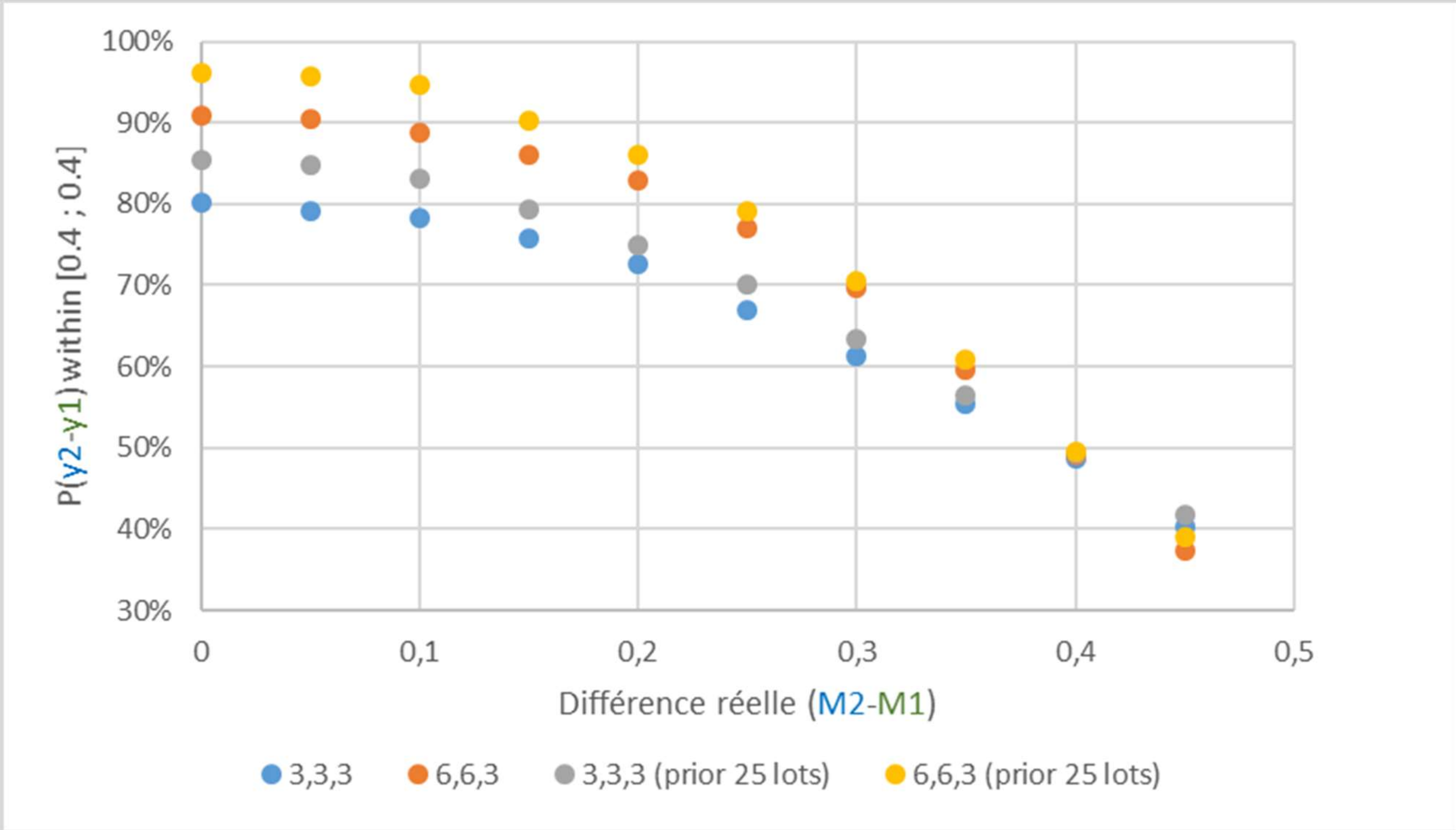
- **Les seuils *(ex: ± 0.505)* varient en fonction de la connaissance, du design, des risques choisis (α, β), des données …**
- **Et si ils étaient fixes ?**

P (y2-y1) ⊂ [-0.4 ; 0.4] ?



M1 = M2
Protocole (3,3,3)
Knowledge : 3 lots

COURBES DE PERFORMANCES



PART I – RESUME EPISODE I

PART II – BAYESIAN FRAMEWORK

… IN A NUTSHELL

PART III – APPLICATION AU CAS DE MARION

PART IV – DISCUSSIONS

ET L'APPORT DU BAYESIEN EN PLANIFICATION EXPERIMENTALE DANS TOUT CELA ?


Merci

BIOMÉRIEUX

PIONEERING DIAGNOSTICS