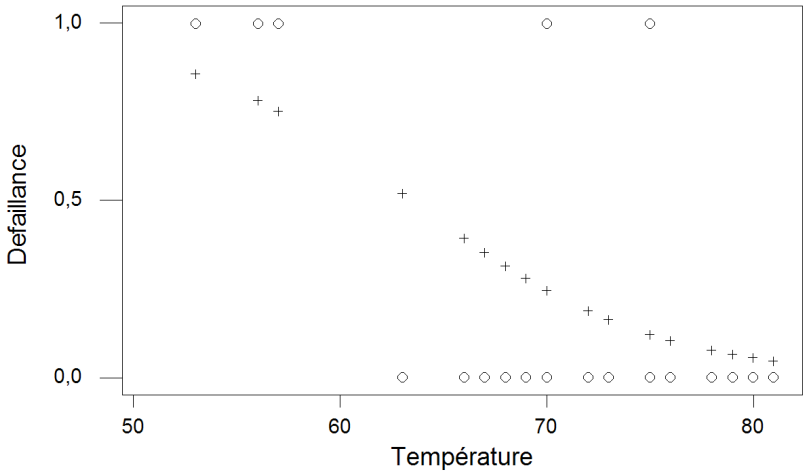
Feuille de Travaux Dirigés n° 1
Régression logistique

Exercice I.1. Défaillance d'un élément mécanique
On s'intéresse ici à l'étude de la pièce ayant causé l'explosion de la navette spatiale Challenger en 1986. (Données de Dalal et *al.*, 1986). L'étanchéité du moteur de la navette spatiale est assurée par six pièces identiques appelées « O-ring ». L'explosion de la navette Challenger est due à la défaillance d'au moins l'une de ces pièces. On a reporté dans le tableau ci-dessous des données collectées au cours des 24 vols précédents d'une navette spatiale : la variable *Température* est la température au moment du lancement, exprimée en °F, et la variable *Défaillance* est une variable qui vaut 0 si aucun des « O-ring » n'a été endommagé au cours du lancement et 1 si au moins l'un d'entre eux a été endommagé.

Température	*Défaillance*	*Température*	*Défaillance*
53	1	70	1
56	1	70	1
57	1	72	0
63	0	73	0
66	0	75	0
67	0	75	1
67	0	76	0
67	0	76	0
68	0	78	0
69	0	79	0
70	0	80	0
70	1	81	0

1. Préciser la nature de la réponse observée ainsi que celle du facteur explicatif considéré. Quelle modélisation peut-on utiliser ?
2. Représenter graphiquement à l'aide d'un nuage de points les données du tableau ci-dessus. Que conclure quant à une éventuelle relation entre la température et l'apparition d'une défaillance ?
3. En utilisant une régression logistique, déterminer si le facteur *Température* a une influence sur l'apparition d'une *Défaillance*. Représenter les probabilités prédites par le modèle sur le nuage de points construit à la question **2.**. On comparera le résultat obtenu avec le graphique au verso.
4. Ces données étaient à la disposition des ingénieurs avant le lancement de la navette Challenger. Sachant que le lancement de la navette Challenger a eu lieu à une température de 31 °F, quelle était la probabilité prédite à l'aide du modèle de l'apparition d'au moins une défaillance ? Bien que les valeurs avec lesquelles on a établi le modèle soient éloignées de la valeur de 31 °F, et que de ce fait

la probabilité de l'apparition d'au moins une défaillance est modélisée avec une grande incertitude, auriez-vous tout de même autorisé le décollage ?



Exercice I.2. Survie de parasites

On reproduit ici les données de l'exemple du cours portant sur la léthalité de la cypermethrine sur des parasites.

Total	N morts	Niveau de dose	Sexe
20	1	1	M
20	4	2	M
20	9	4	M
20	13	8	M
20	18	16	M
20	20	32	M
20	0	1	F
20	2	2	F
20	6	4	F
20	10	8	F
20	12	16	F
20	16	32	F

1. Expliquer pourquoi il est pertinent d'étudier ces données avec une régression logistique.
2. Quels sont les différents modèles que l'on peut utiliser ? Combien il y en a-t-il ?
3. Procéder à l'étude de ces modèles.
4. Comparer ces différents modèles entre eux.