

Université Paris 1, UFR 02, Licence de Sciences Economiques  
STATISTIQUE, cours de Mme PRADEL  
Partiel 17 janvier 2005

Sont autorisées, outre une calculatrice et les tables statistiques : deux feuilles (recto verso) manuscrites, remplies par l'étudiant (à sa convenance et de sa main) des formules qu'il a choisies.

*Il sera tenu compte de la cohérence et de la pertinence des arguments avancés autant que de l'exactitude des résultats.*

*Les exercices sont indépendants les uns des autres*

### Exercice 1 (6 pts)

Soit  $X$  une variable aléatoire de loi Normale d'espérance  $m$  et de variance  $\sigma^2$ . Nous disposons d'un échantillon de  $n$  tirages de  $X$  :

$$X_1, X_2, \dots, X_n \approx i.i.d.N [m; \sigma^2]$$

Nous désirons estimer  $m$  et prévoir le prochain tirage  $X_{n+1}$ .

1. La variance est connue, égale à 4.
  - (a) Quelle est le meilleur estimateur de  $m$  ? Quelle est la loi de cet estimateur ?
  - (b) Le prochain tirage  $X_{n+1}$  est supposé indépendant des  $X_1, X_2, \dots, X_n$  et de même loi  $N [m; \sigma^2]$ . Quelle est la meilleure prévision de  $X_{n+1}$  ?
  - (c) Quelle est la loi de la variable  $Z = X_{n+1} - \bar{X}_n$ , où  $\bar{X}_n$  désigne la moyenne empirique de l'échantillon ?
  - (d) En déduire un intervalle de prévision à 80% à partir d'un échantillon de taille  $n = 24$
2. Comment les réponses sont-elles modifiées si la variance n'est pas connue ?
3. Les observations faites conduisent à

$$\bar{x}_n = \frac{1}{24} \sum_{i=1}^{24} x_i = 143.82 \quad \text{et} \quad s^2 = \frac{1}{23} \sum_{i=1}^{24} (x_i - \bar{x}_n)^2 = 4.624$$

Donner l'estimation de  $m$  et les intervalles de prévision à 80% pour  $X_{25}$  lorsque  $\sigma^2 = 4$  et lorsque  $\sigma^2$  est inconnue

### Exercice 2 (6 pts)

1. Une responsable marketing désire évaluer la popularité d'un nouveau produit en Ile de France : si pas plus d'un client sur deux n'apprécie le produit, son développement sera interrompu (elle redoute principalement des investissements inutiles en Recherche et Développement). Dans un échantillon de 500 clients elle observe que 287 clients apprécient le nouveau produit tandis que 213 ne l'aiment pas ou bien n'ont aucune opinion.
2. Décrire le modèle statistique : observations, lois des observations, valeurs du paramètre et question posée.
3. Elle décide de faire le test de ( $p \leq 0.5$ ) contre ( $p > 0.5$ ). Justifiez son choix de ( $p \leq 0.5$ ) comme hypothèse de base.
4. Soit  $p_1$  choisi tel que :  $p_1 > 0.5$ . Construire le test de seuil 5% de ( $p = 0.5$ ) contre ( $p = p_1$ ). Est-ce que ce test dépend de  $p_1$ ? Quelle conclusion en tirez-vous ?

- Pour tout  $p$  tel que  $0.04 < p < 0.96$ , calculer en fonction de  $p$  la probabilité de refuser ( $p = 0.5$ ) .Montrez que le test construit est bien un test de seuil 5% pour tester ( $p \leq 0.5$ ) contre ( $p > 0.5$ ) .En déduire que ce test est uniformément le plus puissant des tests de seuil 5% pour tester ( $p \leq 0.5$ ) contre ( $p > 0.5$ ) .
- Quelle décision la responsable marketing prend-elle au vu des observations faites ?

### Exercice 3 ( 8 pts)

Nous examinons dans cet exercice comment l'effet "génération" de la taille se manifeste chez les hommes et les femmes nés après 1945. Nous disposons d'un échantillon de 70 individus (35 hommes et 35 femmes) pour lesquels nous avons observé les variables suivantes :

- $TAILLE_i$  = la taille de  $i$  à 25 ans
- $NAISSANCE_i$  = l'année de naissance de  $i$

La régression des Moindres Carrés Ordinaires de la taille sur  $NAISSANCE$  avec constante fournit les résultats suivants :

|               |          |                    |                     |  |
|---------------|----------|--------------------|---------------------|--|
| AJUSTEMENT    | 0        | DEPENDENT VARIABLE | $TAILLE$ (à 25 ans) |  |
| SMPL          | 1        | TOTAL OBSERVATIONS | 70                  |  |
| FISHER global | 4.560235 | Prob(F-statistic)  | 0.036327            |  |
| DURBIN-WATSON | 0.934160 | SCR                | 0.580360            |  |

| $VARIABLE$  | $COEFFICIENT$ | $STAND.ERROR$ | $T - STATISTIC$ | $PROB$ |
|-------------|---------------|---------------|-----------------|--------|
| $constante$ | 1.570257      | 0.050734      | 30.95106        | 0.0000 |
| $NAISSANCE$ | 0.164819      | 0.077181      | 2.135471        | 0.0363 |

- Rappeler les hypothèses du modèle statistique linéaire standard. Les résultats obtenus ici sont-ils compatibles avec ces hypothèses ?
- Est-il possible d'affirmer qu'au seuil de 5% la taille à 25 ans dépend de la date de naissance ?
- Les régressions séparées sur les 35 hommes puis sur les 35 femmes de l'échantillon fournissent les résultats suivants, dans lesquels les nombres figurant entre crochets sont les  $PROB$  associées aux coefficients sous lesquels ils sont placés :

$$Ajustement (F) \quad TAILLE_i = 1.548 + 0.0943NAISSANCE_i + \hat{u}_i \quad i = 1, \dots, 35$$

$[0.000]$ 
 $[0.0919]$

$$SCR_F = 0.064597 \quad DW_F = 1.820$$

et

$$Ajustement (H) \quad TAILLE_i = 1.599 + 0.2248NAISSANCE_i + \hat{u}_i \quad i = 36, \dots, 70$$

$[0.000]$ 
 $[0.0137]$

$$SCR_H = 0.188948 \quad DW_H = 2.066$$

- Vérifier que les hypothèses du modèle statistique linéaire standard sont satisfaites sur chaque sous population.
- Décider si l'effet "génération" sur la taille est significatif au seuil de 5% :
  - pour les femmes ?

(b) pour les hommes ?

6. En faisant intervenir les variables :

$$F_i = \begin{cases} 1 & \text{si } i \text{ est une femme} \\ 0 & \text{sinon} \end{cases} \quad \text{et} \quad H_i = \begin{cases} 1 & \text{si } i \text{ est un homme} \\ 0 & \text{sinon} \end{cases}$$

$$NAISSANCEF_i = NAISSANCE * F_i \quad \text{et} \quad NAISSANCEH_i = NAISSANCE * H_i$$

montrer que les deux équations ( $F$ ) et ( $H$ ) peuvent s'écrire sous la forme d'une seule équation valable pour toutes les observations,  $i = 1, \dots, 70$  :

$$TAILLE_I = aF_i + bH_i + cNAISSANCEF_i + dNAISSANCEH_i + res_i$$

En déduire les coefficients ( $a, b, c, d$ ) trouvés dans l'équation de régression de la taille sur les nouvelles variables

7. Montrer que le modèle global ne peut être accepté comme modèle linéaire standard (coup de pouce : il reste un test à faire qui montre que l'hypothèse H2 n'est pas vérifiée : trouvez le)

8. Les PROB associées aux différents coefficients par le calcul standard sont [0.2196] pour  $c$  et [0.0022] pour  $d$ .

(a) Expliquez pourquoi elles sont contradictoires avec les régressions séparées ( $F$ ) et ( $H$ ) .

(b) Finalement, la taille des femmes a-t-elle augmenté depuis 1945, ou non ?

## Extraits de tables statistiques

LOI NORMALE :  $U \approx N(0, 1)$  : table de  $F(u) = P(U \leq u)$

| U   | 0,00   | 0,01   | 0,02   | 0,03   | 0,04   | 0,05   | 0,06   | 0,07   | 0,08   | 0,09   |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 0.2 | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6103 | 0.6141 |
| 0.3 | 0.6179 | 0.6217 | 0.6255 | 0.6293 | 0.6331 | 0.6368 | 0.6406 | 0.6443 | 0.6480 | 0.6517 |
| 0.4 | 0.6554 | 0.6591 | 0.6628 | 0.6664 | 0.6700 | 0.6736 | 0.6772 | 0.6808 | 0.6844 | 0.6879 |
| 0.5 | 0.6915 | 0.6950 | 0.6985 | 0.7019 | 0.7054 | 0.7088 | 0.7123 | 0.7157 | 0.7190 | 0.7224 |
| 0.6 | 0.7257 | 0.7291 | 0.7324 | 0.7357 | 0.7389 | 0.7422 | 0.7454 | 0.7486 | 0.7517 | 0.7549 |
| 0.7 | 0.7580 | 0.7611 | 0.7642 | 0.7673 | 0.7704 | 0.7734 | 0.7764 | 0.7794 | 0.7823 | 0.7852 |
| 0.8 | 0.7881 | 0.7910 | 0.7939 | 0.7967 | 0.7995 | 0.8023 | 0.8051 | 0.8078 | 0.8106 | 0.8133 |
| 0.9 | 0.8159 | 0.8186 | 0.8212 | 0.8238 | 0.8264 | 0.8289 | 0.8315 | 0.8340 | 0.8365 | 0.8389 |
| 1.0 | 0.8413 | 0.8438 | 0.8461 | 0.8485 | 0.8508 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8643 | 0.8665 | 0.8686 | 0.8708 | 0.8729 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8849 | 0.8869 | 0.8888 | 0.8907 | 0.8925 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| 1.3 | 0.9032 | 0.9049 | 0.9066 | 0.9082 | 0.9099 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |
| 1.4 | 0.9192 | 0.9207 | 0.9222 | 0.9236 | 0.9251 | 0.9265 | 0.9279 | 0.9292 | 0.9306 | 0.9319 |
| 1.5 | 0.9332 | 0.9345 | 0.9357 | 0.9370 | 0.9382 | 0.9394 | 0.9406 | 0.9418 | 0.9429 | 0.9441 |
| 1.6 | 0.9452 | 0.9463 | 0.9474 | 0.9484 | 0.9495 | 0.9505 | 0.9515 | 0.9525 | 0.9535 | 0.9545 |
| 1.7 | 0.9554 | 0.9564 | 0.9573 | 0.9582 | 0.9591 | 0.9599 | 0.9608 | 0.9616 | 0.9625 | 0.9633 |

LOI DE STUDENT à  $\nu$  degrés de liberté :

TABLE de  $t$  en fonction du degré de liberté et de la probabilité  $p$ , tels que  $P(|T| > t) = p$

| p        | 0,90    | 0,70    | 0,50    | 0,40    | 0,30    | 0,20    | 0,10    | 0,05    | 0,02    | 0,01    |
|----------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 22       | 0,127   | 0,390   | 0,686   | 0,858   | 1,061   | 1,321   | 1,717   | 2,074   | 2,508   | 2,819   |
| 23       | 0,127   | 0,390   | 0,685   | 0,858   | 1,060   | 1,319   | 1,714   | 2,069   | 2,500   | 2,807   |
| 24       | 0,127   | 0,390   | 0,685   | 0,857   | 1,059   | 1,318   | 1,711   | 2,064   | 2,492   | 2,797   |
| 25       | 0,127   | 0,390   | 0,684   | 0,856   | 1,058   | 1,316   | 1,708   | 2,060   | 2,485   | 2,787   |
| $\infty$ | 0,12566 | 0,38532 | 0,67449 | 0,84162 | 1,03643 | 1,28155 | 1,64485 | 1,95996 | 2,32634 | 2,57582 |

LOI DE FISHER : Table du quantile  $F_{0,95}$  en fonction de  $r_1$  et  $r_2$  :  $P(F(r_1, r_2) > F_{0,95}) = 5\%$

| $r_2 \setminus r_1$ | 30   | 31   | 32   | 33   | 34   | 35   |
|---------------------|------|------|------|------|------|------|
| 30                  | 1.84 | 1.83 | 1.83 | 1.82 | 1.82 | 1.81 |
| 31                  | 1.83 | 1.82 | 1.82 | 1.81 | 1.81 | 1.80 |
| 32                  | 1.82 | 1.81 | 1.80 | 1.80 | 1.79 | 1.79 |
| 33                  | 1.81 | 1.80 | 1.79 | 1.79 | 1.78 | 1.78 |
| 34                  | 1.80 | 1.79 | 1.78 | 1.78 | 1.77 | 1.77 |
| 35                  | 1.79 | 1.78 | 1.77 | 1.77 | 1.76 | 1.76 |

TABLE de DURBIN-WATSON :

Test unilatéral de  $\rho = 0$  contre  $\rho > 0$ , au seuil de 5% (test bilatéral : seuil de 10%)

| n  | $k' = 1$ |      | $k' = 2$ |      | $k' = 3$ |      | $k' = 4$ |      | $k' = 5$ |      |
|----|----------|------|----------|------|----------|------|----------|------|----------|------|
|    | dL       | du   | dL       | du   | dL       | du   | dL       | du   | dL       | du   |
| 35 | 1,40     | 1,52 | 1,34     | 1,58 | 1,28     | 1,65 | 1,22     | 1,73 | 1,16     | 1,80 |
| 70 | 1,58     | 1,64 | 1,55     | 1,67 | 1,52     | 1,70 | 1,49     | 1,74 | 1,46     | 1,77 |