

PSY38X2 : Traitement de données en Psychologie - TD N°2

F de Fisher - Analyse de variance à 1 facteur - Corrélation et Régression Linéaires

1) Comparaison de deux variances

1.1 Comparaison de deux variances avec Minitab

On reprend les données de l'exercice 2 de la fiche de TD de statistiques.
Chargez Minitab et ouvrez le projet W:\PSY3\TD-Minitab\TR-lampes.mpj

Observez la feuille de données et remarquez que l'organisation des données ne correspond pas à la disposition choisie dans le polycopié. Ici, les données se trouvent réparties sur deux colonnes :

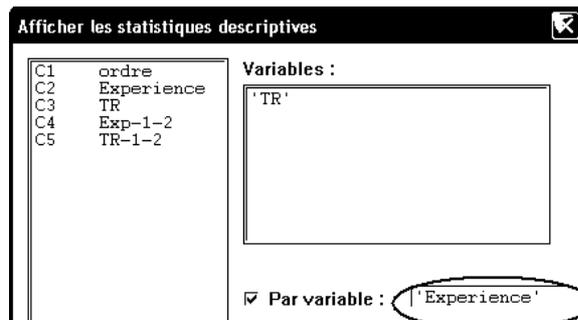
- La colonne C2 (nommée "Expérience") donne le niveau du facteur "condition expérimentale" (1, 2 ou 3)
- La colonne C3 (nommée "TR") donne l'ensemble des valeurs observées de la variable "Temps de réaction".

Quant aux colonnes C4 (Exp-1-2) et C5 (TR-1-2) , elles rassemblent les mêmes données, mais en se limitant aux conditions expérimentales 1 et 2.

| | C1 | C2 | C3 | C4 | C5 |
|----|-------|------------|----|---------|--------|
| ↓ | ordre | Expérience | TR | Exp-1-2 | TR-1-2 |
| 17 | 17 | 1 | 21 | 1 | 21 |
| 18 | 18 | 1 | 19 | 1 | 19 |
| 19 | 19 | 1 | 17 | 1 | 17 |
| 20 | 20 | 1 | 23 | 1 | 23 |
| 21 | 1 | 2 | 32 | 2 | 32 |
| 22 | 2 | 2 | 40 | 2 | 40 |

Utilisez tout d'abord le menu Stat - Statistiques Élémentaires - Afficher les statistiques descriptives... pour retrouver les résultats de statistiques descriptives donnés par l'énoncé.

N.B. On peut obtenir globalement les paramètres correspondant aux trois échantillons en complétant le dialogue comme suit :



Utilisez le menu Stat - ANOVA - Test de comparaison de la variance... appliqué aux variables Exp-1-2 et TR-1-2 pour comparer les variances des temps de réaction *des deux premières expériences*. Vous devriez obtenir :

Test F (loi normale)
Statistique du test: 3,617
Niveau de P: 0,007

Comparez le niveau de significativité affiché (0,7%) avec celui obtenu à l'aide d'Excel (0,37% - cf. 1.3). Minitab fait ici un test bilatéral.

Comparez ensuite les variances des trois expériences ; Minitab utilise dans ce cas un autre test : le **test de Bartlett**. La statistique de test est une expression calculée à partir des variances et des effectifs des échantillons, la loi suivie est approximativement une loi du khi-2. Il est cependant

surprenant que le test de Levene (qui utilise les écarts des observations aux médianes, plutôt qu'aux moyennes) donne un résultat très différent de celui du test de Bartlett.

1.2 Comparaison de deux variances avec Excel, sans Utilitaire d'Analyse

Ouvrez le fichier W:\PSY3\TD-Excel\TR-lampes.xls et examinez les données saisies.

Calculez en B23:D23 les variances des temps de réaction des trois séries de données.

Calculez en B24 la valeur observée de la statistique de test.

Indiquez les degrés de liberté du numérateur et du dénominateur en B25 et B26.

Pour calculer le niveau de significativité (cellule B28), on pourra utiliser la fonction LOI.F. La fonction inverse, qui permet le calcul de F critique, est la fonction INVERSE.LOI.F.

| | A | B | C |
|----|---------------------------|-----------------------------|--------------|
| 23 | Variance | =VAR(B2:B21) | =VAR(C2:C21) |
| 24 | Valeur de F | =C23/B23 | |
| 25 | ddl1 | 19 | |
| 26 | ddl2 | 19 | |
| 27 | Seuil | 0,05 | |
| 28 | Niveau de significativité | =LOI.F(B24;B25;B26) | |
| 29 | F critique | =INVERSE.LOI.F(B27;B25;B26) | |

1.3 Comparaison de deux variances à l'aide de l'utilitaire d'analyse

L'utilitaire d'analyse comporte un item nommé Test d'égalité des variances (F-test) permettant de réaliser le traitement précédent. Le dialogue pourra être complété comme suit :

ce qui permettra d'obtenir les résultats suivants :

Test d'égalité des variances (F-Test)

| | <i>Expérience 1</i> | <i>Expérience 2</i> |
|-------------------------------------|---------------------|---------------------|
| Moyenne | 19,75 | 40,95 |
| Variance | 15,67105263 | 56,68157895 |
| Observations | 20 | 20 |
| Degré de liberté | 19 | 19 |
| F | 0,276475231 | |
| P(F<=f) unilatéral | 0,003689267 | |
| Valeur critique pour F (unilatéral) | 0,461200855 | |

Remarquez qu'Excel fait un test unilatéral "à gauche", alors qu'en cours nous nous sommes limités à l'étude du cas où l'hypothèse H_1 est de la forme $\sigma_1 > \sigma_2$.

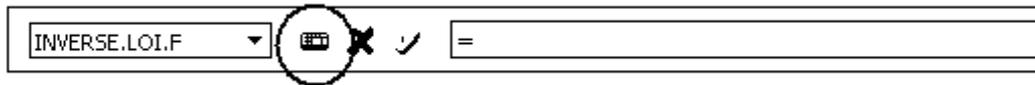
Pour qu'Excel fasse le test en plaçant la variance de la première série d'observations en dénominateur (comme nous le faisons à la main), il suffit d'inverser les rôles des deux plages de données indiquées.

1.4 Comparaison de deux variances, avec OpenOffice Calc

A titre d'exercice, refaites le calcul du paragraphe 1.2 à l'aide d'OpenOffice Calc.

Chargez OpenOffice Calc (le tableur d'OpenOffice) et ouvrez le fichier le fichier W:\PSY3\TD-Calc\TR-lampes.sxc (ou le fichier TR-lampes.xls du répertoire TD-Excel, car OpenOffice Calc ouvre sans problème les documents Excel.).

Les noms et la syntaxe des fonctions à utiliser sont identiques. Il es par contre utile de savoir que l'autopilote d'Open Office Calc est activé à l'aide du bouton suivant de la barre de formules :



2) Analyse de variance à un facteur

2.1 - Analyse de variance à 1 facteur avec Minitab - Plan S<A>

2.1.1 - Les menus Stat - ANOVA - A un facteur ...

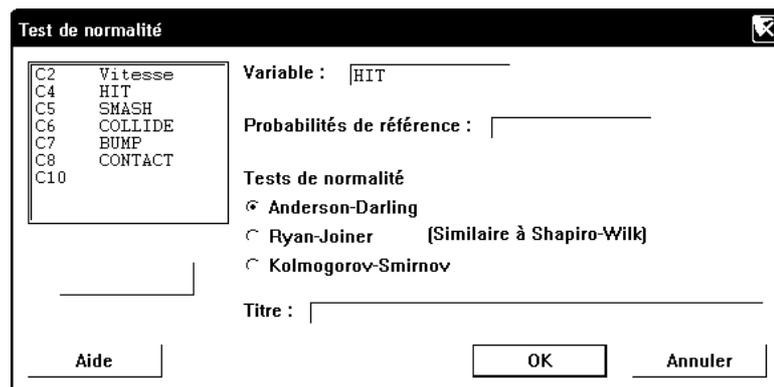
On reprend la situation décrite dans l'exercice 9 de la fiche de TD de statistiques.

Chargez Minitab et ouvrez le projet W:\PSY3\TD-Minitab\Loftus.mpj

On se propose de comparer les 5 groupes des données Loftus et effectuant une analyse de variance.

En toute rigueur, on s'est assuré au préalable et de la normalité des distributions (menu Stat - Statistiques Élémentaires - Test de Normalité...), puis de l'homogénéité des variances.

Par exemple, pour tester la normalité de la variable HIT, on pourra compléter le dialogue comme suit :



Minitab produit alors une représentation graphique. On trouve en annotation :

Test de normalité de Anderson-Darling
 A-Carré : 0,177
 Niveau de P : 0,893

Le niveau de significativité de la statistique de test est de 89%. Autrement dit, c'est l'hypothèse H_0 : Normalité de la distribution parente qui est retenue.

De même, l'égalité des variances dans les populations parentes des cinq groupes pourra être testée à l'aide du menu Stat - ANOVA - Test de comparaison de la variance. Les variables sont alors Vitesse et Verbe. On obtient :

Test de Bartlett(loi normale)

Statistique du test: 6,180

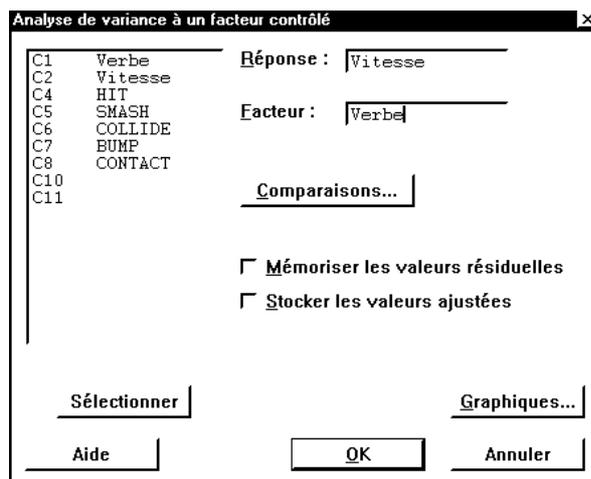
Niveau de P: 0,186

et c'est encore l'hypothèse H_0 (égalité des variances) qui est retenue.

Pour réaliser l'analyse de variance proprement dite, utilisez le menu Stat - ANOVA - A un facteur contrôlé... Complétez la fenêtre de dialogue en indiquant :

- dans le champ "Réponse", la variable dépendante (ici, Vitesse)

- dans le champ "Facteur", la variable indépendante, ou facteur de variation :

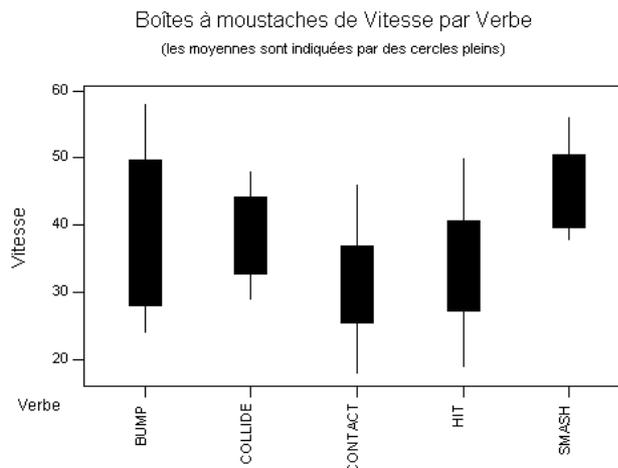


A l'aide du bouton "graphiques", demandez aussi la composition d'un graphique en "boîtes à moustaches" pour les données.

Minitab fournit le tableau d'analyse de variance suivant :

| Analyse de variance pour Vitesse | | | | | |
|----------------------------------|----|--------|-------|------|-------|
| Source | DL | SC | CM | F | P |
| Verbe | 4 | 1256.5 | 314.1 | 4.06 | 0.007 |
| Erreur | 45 | 3481.0 | 77.4 | | |
| Total | 49 | 4737.5 | | | |

Le niveau de significativité de la valeur F_{cal} est ici 0,7%. L'hypothèse H_1 est donc retenue, aux seuils traditionnels : le verbe utilisé dans la question influe sur la réponse du sujet. Le graphique permet de situer les variables les unes par rapport aux autres :



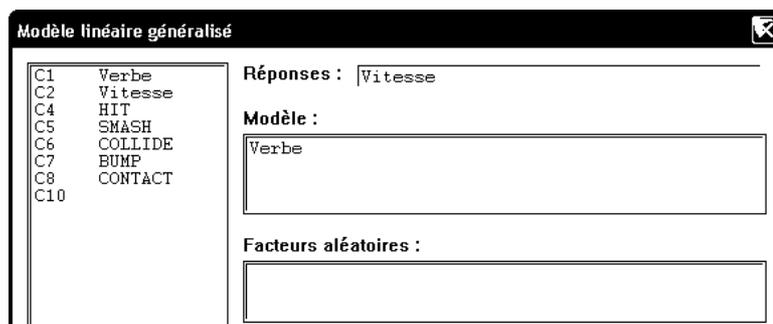
Remarque : L'analyse de variance peut aussi être réalisée à partir des données saisies dans les colonnes C4 à C8. Mais il faut alors sélectionner le menu Stat - ANOVA - A un facteur contrôlé (désempilé)...

2.1.2 Le menu Stat - ANOVA - Modèle linéaire généralisé

L'analyse de variance à un facteur peut également être réalisée à l'aide d'un outil beaucoup plus général (et que nous devons utiliser par la suite dans d'autres situations) : le modèle linéaire généralisé.

Utilisez le menu Stat - ANOVA - Modèle linéaire généralisé et complétez le dialogue en indiquant :

- dans le champ "Réponses", la variable dépendante (ici, Vitesse)
- Dans le champ "Modèle", le facteur de variation (ici, le facteur Verbe) :



Vous pouvez ensuite enregistrer votre document Loftus.mpj avant de l'envoyer par e-mail à votre enseignant.

Exercice

Ouvrez sous Minitab le fichier W:\PSY3\TD-Minitab\Bransfor.MPJ.

Ce fichier contient l'énoncé et les données de l'exercice 8 de la fiche de TD de statistiques.

Vérifiez à l'aide de Minitab les différents résultats indiqués sur la fiche de TD (sommes par groupe, tableau d'analyse de variance, etc).

2.2 - Analyse de variance à un facteur avec Excel

2.2.1 Analyse de variance en utilisant l'Utilitaire d'Analyse

On considère les données décrites dans l'énoncé "Données Bransfor" (énoncé 8), qui ont été saisies dans la feuille BRANSFOR du classeur W:\PSY3\TD-Excel\Bransfor.XLS

Nous allons tester l'hypothèse "*dans les populations parentes, les moyennes des quatre groupes sont différentes*" à l'aide d'une analyse de variance.

Dans la panoplie de fonctions d'Excel, il n'existe pas de fonction permettant de faire directement une analyse de variance. Nous allons donc recourir à l'Utilitaire d'Analyse.

Sélectionnez le menu Outils-Utilitaire d'analyse... puis l'item Analyse de variance: un facteur.

Complétez la fenêtre de dialogue comme suit :

| | |
|---|---|
| Paramètres d'entrée | |
| Plage d'entrée: | \$A\$1:\$D\$6 |
| Groupées par: | <input checked="" type="checkbox"/> Colonnes <input type="checkbox"/> Lignes |
| <input checked="" type="checkbox"/> Intitulés en première ligne | |
| Seuil de signification: | 0,05 |

Parmi les résultats calculés par Excel, on trouve alors le tableau d'analyse de variance suivant, que l'on comparera à celui figurant sur la fiche de TD :

ANALYSE DE VARIANCE

| Source des variations | Somme des carrés | Degré de liberté | Moyenne des carrés | F | Probabilité | Valeur critique pour F |
|---------------------------|------------------|------------------|--------------------|------|-------------|------------------------|
| Entre Groupes | 50,95 | 3 | 16,98 | 7,23 | 0,28% | 3,24 |
| A l'intérieur des groupes | 37,6 | 16 | 2,35 | | | |
| Total | 88,55 | 19 | | | | |

2.2.2 Analyse de variance sans utiliser l'Utilitaire d'Analyse

Rappelons que le principe de l'analyse de variance consiste à décomposer la "variation totale" en une "variation entre les groupes" et une "variation à l'intérieur des groupes", et de les comparer. Ici, le terme de "variation" peut être compris comme étant le produit de la variance par le nombre d'observations.

$$\text{Variation} = N \times s^2 = (N - 1) \times s_c^2$$

Pour déterminer la variation inter-groupes, constituez en A10:D14 un tableau de mêmes dimensions que le tableau de données, mais en remplaçant l'observation relative à chaque sujet par la moyenne de son groupe.

Calculez ensuite la variation inter-groupes en cellule G3, par exemple à l'aide de la formule :

=VAR(A10:D14)*19

Pour déterminer la variation intra-groupes, constituez en A18:D22 un tableau de mêmes dimensions que le tableau de données, dont les valeurs sont obtenues comme différences des valeurs correspondantes dans les deux tableaux précédents.

Calculez ensuite la variation intra-groupes en cellule G4.

La variation totale pourra être obtenue soit comme somme des deux précédentes, soit à l'aide d'une formule telle que :

=VAR(A2:D6)*19

Complétez ensuite le tableau d'analyse de variance en indiquant les ddl et en calculant les carrés moyens, et le rapport F.

| Variations Intergroupes | | | | Variations Intragroupes | | | | | | | |
|-------------------------|-----|-----|-----|-------------------------|-----|-----|-----|-----|-----|------|------|
| GR1 | GR2 | GR3 | GR4 | GR1 | GR2 | GR3 | GR4 | GR1 | GR2 | GR3 | GR4 |
| 3 | 5 | 2 | 5 | 3 | 7 | 3,2 | 4,2 | 0 | -2 | -1,2 | 0,8 |
| 3 | 9 | 4 | 4 | 3 | 7 | 3,2 | 4,2 | 0 | 2 | 0,8 | -0,2 |
| 2 | 8 | 5 | 3 | 3 | 7 | 3,2 | 4,2 | -1 | 1 | 1,8 | -1,2 |
| 4 | 4 | 4 | 5 | 3 | 7 | 3,2 | 4,2 | 1 | -3 | 0,8 | 0,8 |
| 3 | 9 | 1 | 4 | 3 | 7 | 3,2 | 4,2 | 0 | 2 | -2,2 | -0,2 |

Enregistrez ensuite votre fichier et joignez-le à un mail adressé à votre enseignant.

Exercice : Reprendre cette dernière partie en utilisant OpenOffice Calc.

Exercice :

Saisissez dans une feuille de calcul Minitab les données de l'exercice 10 de la fiche de TD de statistiques (données Besancon).

Etudiez la normalité des distributions parentes et la pertinence de l'hypothèse d'égalité des moyennes.

Retrouvez les résultats d'ANOVA indiqués dans la fiche de TD.

Transférez les données dans une feuille de calcul Excel et reprendre le tableau d'analyse de variance.

3) Un test non paramétrique permettant de comparer plusieurs groupes : le test de Kruskal-Wallis

Le test de Kruskal-Wallis peut être présenté comme une extension du test des rangs de Wilcoxon Mann Whitney, ou comme l'équivalent non paramétrique de l'analyse de variance, à utiliser notamment lorsque l'hypothèse de normalité des distributions parentes n'est pas pertinente. Ce test est présenté brièvement dans l'exercice 12 de la fiche de TD de statistiques. En voici un exemple d'utilisation.

Les données figurant dans le fichier [W:\PSY3\TD-Minitab\Kruskal-Wallis.mpi](#) représentent le nombre de problèmes arithmétiques simples (sur 85) résolus (de manière correcte ou incorrecte) en une heure par des sujets ayant reçu un médicament dépresseur, un stimulant ou un placebo.

Examinez ces données et indiquez pourquoi l'hypothèse de normalité des distributions parentes paraît peu pertinente.

Utilisez ensuite le menu Stat - Tests non paramétriques - Kruskal-Wallis... Vous devriez obtenir le résultat suivant :

Test de Kruskal-Wallis sur Score

| Groupe | N | Médiane | Rang Moyen | Z |
|----------|---|---------|------------|-------|
| Dépresse | 7 | 44,00 | 5,0 | -2,96 |
| Placébo | 4 | 57,50 | 10,0 | 0,00 |

| | | | | |
|----------|----|-------|------|------|
| Stimulan | 8 | 71,00 | 14,4 | 2,89 |
| Global | 19 | | 10,0 | |

H = 10,36 DL = 2 P = 0,006

H = 10,41 DL = 2 P = 0,005 (ajusté pour les nombres de même grandeur)

Les hypothèses H_0 et H_1 portent ici sur les médianes des distributions parentes. La statistique de test, H, vérifie : $H_{cal} = 10,36$. La loi suivie est une loi du khi-2, avec 2 ddl (nombre de groupes -1). Le niveau de significativité du résultat (0,6%) indique que H_1 sera retenue aux seuils traditionnels.

4) Corrélacion et régression

4.1 - Covariance et coefficient de corrélation avec Minitab

4.1.1. Enoncé d'un cas :

Dans une expérience de perception, on étudie l'évaluation des longueurs de figures géométriques. Le sujet est invité à évaluer les longueurs des figures, en s'aidant d'une figure de référence dont il connaît la longueur (9 cm).

Dans la condition 1, les figures sont 11 bâtonnets. Les données recueillies pour un sujet sont les suivantes :

| | | | | | | | | | | | |
|-------------|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|
| Longueur | 2.5 | 4.6 | 6.3 | 7.6 | 8.5 | 9.0 | 9.5 | 10.4 | 11.7 | 13.4 | 15.5 |
| Eval. long. | 2.8 | 4.4 | 6.2 | 7.8 | 8.2 | 9.0 | 9.6 | 10.6 | 12.0 | 13.6 | 15.2 |

Dans la condition 2, les figures sont des cercles de périmètres égaux aux longueurs des bâtonnets de la condition 1. L'évaluation du périmètre par un sujet est alors la suivante :

| | | | | | | | | | | | |
|--------------|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|
| Périmètre | 2.5 | 4.6 | 6.3 | 7.6 | 8.5 | 9.0 | 9.5 | 10.4 | 11.7 | 13.4 | 15.5 |
| Eval. périm. | 1.8 | 3.6 | 5.8 | 7.2 | 8.4 | 9.0 | 9.8 | 11.0 | 13.2 | 16.1 | 21 |

Ouvrir un nouveau projet Minitab et saisir les données dans quatre variables : Longueur, Eval. Longueur, Périmètre, Eval. Périm. La troisième pourra évidemment être recopiée à partir de la première.

4.1.2 Covariance et coefficient de corrélation

Rappel : les notions de covariance et de coefficient de corrélation ont été introduites en 2^e année. Rappelons les formules de calcul :

Soit deux variables numériques X et Y définies sur la même population statistique, d'effectif total n. La covariance des variables X et Y est définie par :

$$Cov(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \left(\frac{1}{n} \sum_{i=1}^n x_i y_i \right) - \bar{x} \bar{y}$$

où \bar{x} et \bar{y} désignent les moyennes des variables X et Y.

Le coefficient de corrélation est défini par :

$$\rho = \frac{Cov(X, Y)}{\sigma(X)\sigma(Y)}$$

Le coefficient de corrélation prend ses valeurs entre -1 et 1, une valeur proche de 0 signifiant que les deux variables sont peu corrélées entre elles, une valeur (positive ou négative) éloignée de 0 signifiant une forte corrélation entre les deux variables.

Notez qu'un logiciel tel que Minitab calcule en fait la covariance corrigée, obtenue en remplaçant n par n-1 dans la première formule donnant la covariance. Comme pour la variance, la covariance corrigée est l'estimation de la covariance de la population, faite à partir des données considérées comme relatives à un échantillon.

Les méthodes de corrélation et de régression linéaires sont bien adaptées à l'étude de la situation proposée. Calculer les covariances et les coefficients de corrélation des couples (Longueur, Eval. Long.) d'une part, (Périmètre, Eval. Périmètre) d'autre part. Utiliser pour cela les menus Stat - Statistiques Élémentaires - Covariance... et Stat - Statistiques Élémentaires - Corrélation. Remarque que, pour la covariance, Minitab fournit un résultat du type suivant :

Covariances

| | | |
|-----------------|-----------------|-----------------|
| | Longueur | Eval Lon |
| Longueur | 14.2220 | |
| Eval Lon | 14.1360 | 14.1025 |

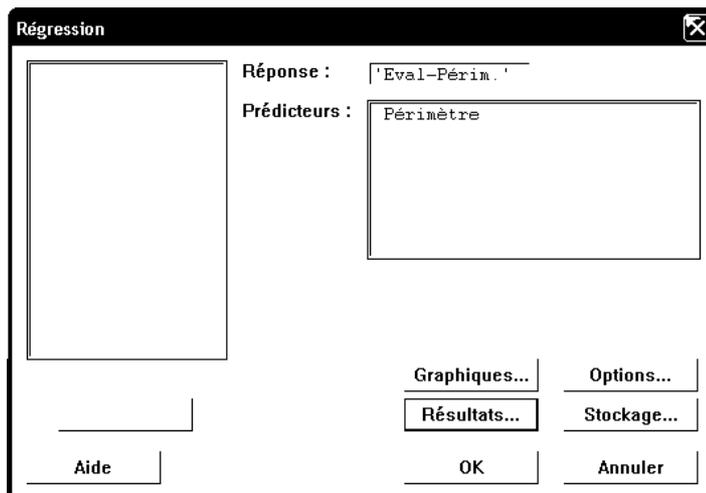
Autrement dit, Minitab calcule les covariances de tous les couples de variables formés à partir des variables fournies, la covariance du couple (Longueur, Longueur), par exemple étant, bien entendu, la variance de la variable Longueur.

4.1.3 Régression linéaire

Lorsque deux variables X et Y sont corrélées, et que la variable X peut s'interpréter comme la "cause" ou simplement comme un prédicteur de Y, on peut rechercher une équation de la droite de régression de Y par rapport à X. Les formules de calcul sont données par :

$$y = ax + b \quad \text{avec} \quad a = \frac{Cov(X,Y)}{\sigma^2(X)} \quad \text{et} \quad b = \bar{y} - a\bar{x}$$

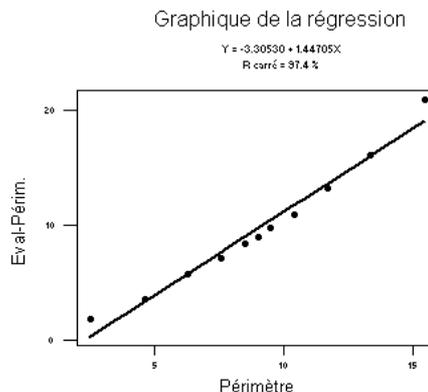
On souhaite étudier si la variable Périmètre est un bon prédicteur de la variable Eval-Périm. Pour cela, on va rechercher une équation de la droite de régression de Eval-Périm. par rapport à Périmètre : Menu Stat - Régression - Régression ...



On obtient ainsi comme équation de régression :

L'équation de régression est
Eval-Périm. = - 3.31 + 1.45 Périmètre

Le menu Stat - Régression - Droite d'ajustement... permet d'obtenir le nuage de point et la droite d'ajustement sur un graphique

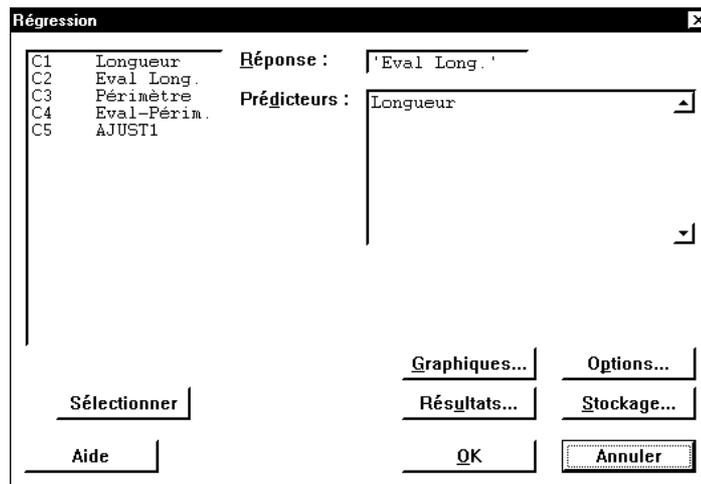


Pour les variables Longueur et Eval. Long., on souhaite étudier si le modèle

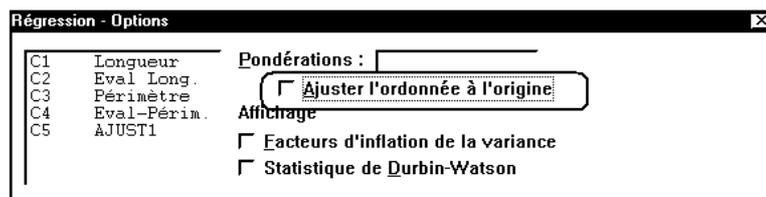
$$\text{Eval. Long} = \text{Longueur}$$

constitue un bon modèle des données observées. Pour cela, on va rechercher une équation de la droite de régression de Eval. Long. par rapport à Longueur **en imposant à cette droite de passer par l'origine** :

Menu Stat - Régression - Régression ...



Ouvrir le dialogue Options... et désactiver la case "Ajuster l'ordonnée à l'origine".

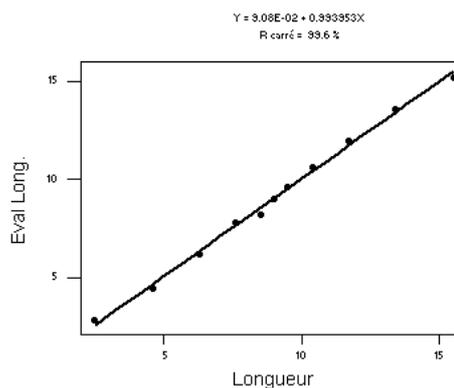


On obtient ainsi comme équation de régression :

$$\text{Eval. Long.} = 1.00 \text{ Longueur.}$$

Le menu Stat - Régression - Droite d'ajustement... permet d'obtenir le nuage de point et la droite d'ajustement sur un graphique (mais il ne s'agit plus exactement de la droite "passant par l'origine") :

Graphique de la régression



4.2 - Corrélation et régression linéaire avec Excel

4.2.1 - Utilisation des fonctions d'Excel

Enoncé du cas

Wagner, Compas et Howell (1988) ont étudié la relation entre le stress et la santé mentale chez des universitaires de première année. A l'aide d'une échelle qu'ils ont mise au point, ils mesurent le stress

perçu par le sujet dans son cadre social et son environnement. Ils ont également demandé aux étudiants de remplir la liste de contrôle d'Hopkins qui évalue la présence ou l'absence de 57 symptômes psychologiques

Le fichier W:\PSY3\TD-Excel\Wagner.xls contient des données numériques relatives à 107 sujets.

Calcul de la covariance et du coefficient de corrélation

Les fonctions utiles ici s'appellent COVARIANCE et COEFFICIENT.CORRELATION. Elles utilisent deux paramètres, les plages désignant les deux variables.

|| Entrez en G2 puis G3 les formules :

=COVARIANCE(B2:B108;C2:C108)

=COEFFICIENT.CORRELATION(B2:B108;C2:C108)

Calcul des coefficients de la droite de régression

Excel dispose d'une fonction DROITEREG. Comme la fonction FREQUENCE vue dans le TD N°5 du premier semestre, cette fonction renvoie plusieurs résultats sous forme d'une matrice. Il s'agit ici des deux coefficients de la droite de régression. Pour calculer ces coefficients :

|| Sélectionnez la plage de cellules G6:H6

|| Entrez la formule :

=DROITEREG(C2:C108;B2:B108)

|| Validez en appuyant sur Maj+Ctrl+Entrée

Calcul des valeurs ajustées

On veut calculer en colonne D les valeurs de la variable Symptômes estimées à partir la régression linéaire qui vient d'être faite.

|| Entrez en cellule D2 la formule :

=B2*\$G\$6+\$H\$6

|| et attardez-vous quelques instants à comprendre pourquoi elle a été rédigée ainsi...

|| Recopiez ensuite cette formule jusqu'en D108.

Nuage de points avec ou sans droite de régression

Excel possède également de nombreuses commandes permettant de réaliser différents graphiques. Il est par exemple facile de construire un nuage de points ; un effort supplémentaire permet d'obtenir la droite de régression sur le même dessin.

Construction du nuage de points seul :

|| Menu Insertion - Graphique. Un assistant démarre.

|| *Premier dialogue* : Nuage de points, puis l'icône "points non reliés"

|| *Second dialogue* : Plage de données : =Observations!\$B\$2:\$C\$108. Série en colonnes

|| *Troisième dialogue* : Pas de légende. Titre de l'axe des x : Stress, titre de l'axe des y : symptômes.

|| *Quatrième dialogue* : Insérer le graphique sur une nouvelle feuille nommée Nuage de points.

Pour modifier un élément du graphique après avoir terminé la phase de création avec l'assistant, on peut cliquer sur cet élément avec le bouton droit de la souris. Le menu local qui apparaît alors contient des items tels que Format de la série de données... ou Format de l'axe...

Construction du nuage de points et de la droite de régression :

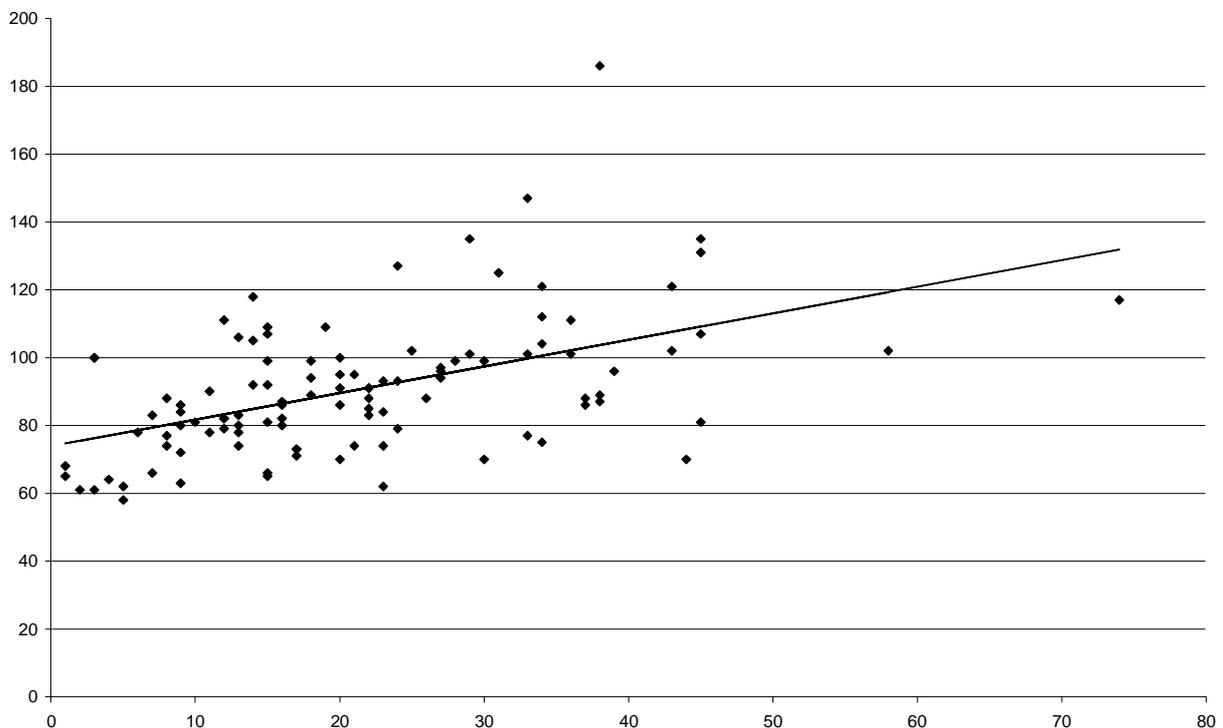
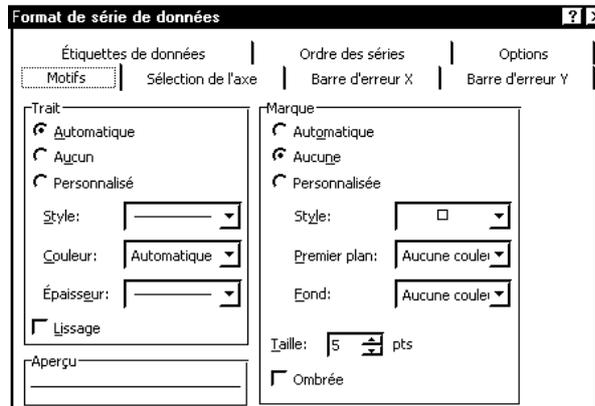
L'idée est de construire sur un même graphique deux nuages de points. Le premier, avec des points non reliés, représente les valeurs observées, le second, dont les points seront reliés, représente les valeurs estimées.

|| Procédez comme dans le cas du graphique précédent, mais spécifiez

=Observations!\$B\$2:\$D\$108

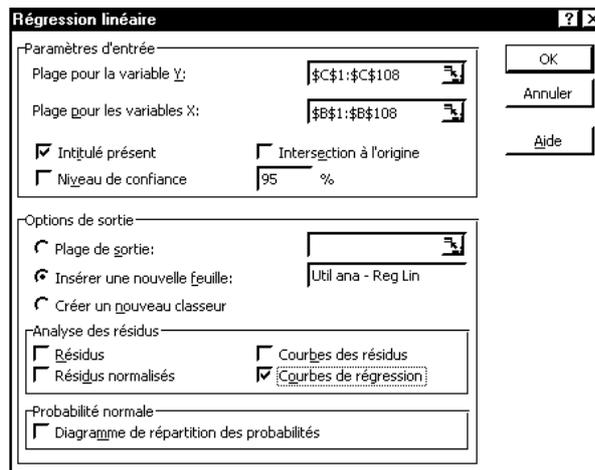
|| comme plage de valeurs. Insérez le graphique dans une nouvelle feuille nommée Nuage et droite.

La droite de régression est alors représentée à l'aide des 107 points correspondant à la colonne D du tableau. Cliquez sur l'un d'entre eux à l'aide du bouton droit, et utilisez le menu Format de la série de données... Dans l'onglet Motifs, modifiez le dialogue en spécifiant : Trait : automatique et Marque : aucune.



4.2.2 - Régression linéaire avec l'utilitaire d'analyse

La situation précédente peut également être traitée par l'utilitaire d'analyse. Pour cela :
 Sélectionnez le menu Outils-Utilitaire d'analyse... puis l'item Régression linéaire.
 Compléter la fenêtre de dialogue comme suit :



Parmi les (trop) nombreux résultats donnés par Excel, recherchez les coefficients de l'équation de la droite de régression de Y par rapport à X ($Y=0.78X+73.89$). Consultez également le graphique représentant les nuages de points expérimental et ajusté.

Enregistrez votre fichier et joignez-le à un mail adressé à votre enseignant.

4.3 Corrélation et régression linéaire avec OpenOffice Calc

Enoncé du cas

Wagner, Compas et Howell (1988) ont étudié la relation entre le stress et la santé mentale chez des universitaires de première année. A l'aide d'une échelle qu'ils ont mise au point, ils mesurent le stress perçu par le sujet dans son cadre social et son environnement. Ils ont également demandé aux étudiants de remplir la liste de contrôle d'Hopkins qui évalue la présence ou l'absence de 57 symptômes psychologiques

Le fichier <W:\PSY3\TD-Calc\Wagner.sxc> contient des données numériques relatives à 107 sujets.

Calcul de la covariance et du coefficient de corrélation

Les fonctions utiles ici s'appellent COVARIANCE et COEFFICIENT.CORRELATION. Elles utilisent deux paramètres, les plages désignant les deux variables.

|| Entrez en G2 puis G3 les formules :

=COVARIANCE(B2:B108;C2:C108)

=COEFFICIENT.CORRELATION(B2:B108;C2:C108)

Calcul des coefficients de la droite de régression

OOCalc dispose d'une fonction DROITEREG. Comme la fonction FREQUENCE d'Excel vue au premier semestre, cette fonction renvoie plusieurs résultats sous forme d'une matrice. Il s'agit ici des deux coefficients de la droite de régression. Pour calculer ces coefficients :

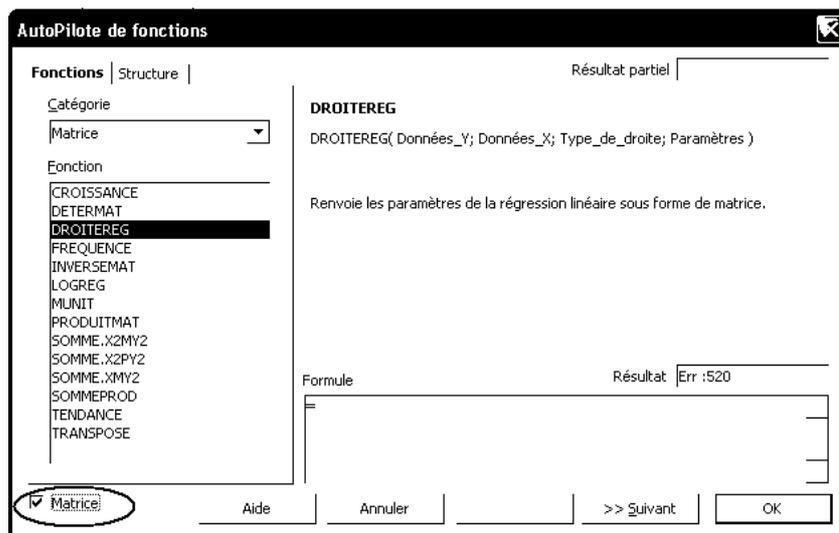
|| Sélectionnez la plage de cellules G6:H6

|| Entrez la formule :

=DROITEREG(C2:C108;B2:B108)

|| Validez en appuyant sur Maj+Ctrl+Entrée

Notez que l'on peut aussi utiliser l'auto-pilote de fonctions, à condition de cocher la case Matrice dans la fenêtre de dialogue de l'auto-pilote, et de chercher la fonction dans la catégorie Matrice.



Remarque : Les coefficients a et b de l'équation $Y=aX+b$ de la droite de régression peuvent aussi être obtenus respectivement à l'aide des fonctions PENTE et ORDONNEE.ORIGINE. Pour calculer ces coefficients :

Placez-vous en cellule G7

Entrez la formule :

=PENTE (C2 : C108 ; B2 : B108)

Placez-vous en cellule H7

Entrez la formule :

=ORDONNEE . ORIGINE (C2 : C108 ; B2 : B108)

Calcul des valeurs ajustées

On veut calculer en colonne D les valeurs de la variable Symptômes estimées à partir la régression linéaire qui vient d'être faite.

Entrez en cellule D2 la formule :

=B2*\$G\$6+\$H\$6

et attardez-vous quelques instants à comprendre pourquoi elle a été rédigée ainsi...

Recopiez ensuite cette formule jusqu'en D108.

Nuage de points avec ou sans droite de régression

OOCalc possède également de nombreuses commandes permettant de réaliser différents graphiques. Il est par exemple facile de construire un nuage de points avec ou sans la droite de régression

Construction du nuage de points seul :

Menu Insertion - Diagramme. Un assistant démarre.

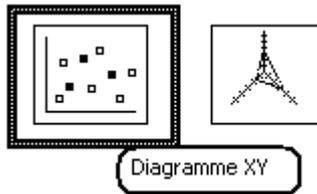
Premier dialogue : Supprimer les croix dans les boîtes à cocher

"Première ligne comme étiquettes", "Première colonne comme étiquette".

Afficher dans la feuille : -nouvelle feuille-

Plage : **\$Observations.\$B\$2:\$C\$108**

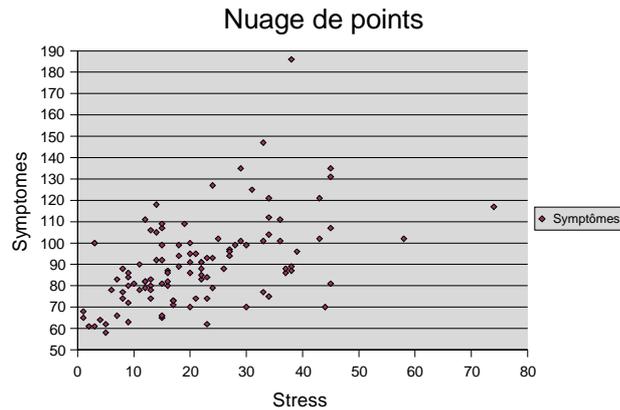
Second dialogue : Sélectionnez le type "Diagramme XY"



Troisième dialogue : Variante "Uniquement symboles"

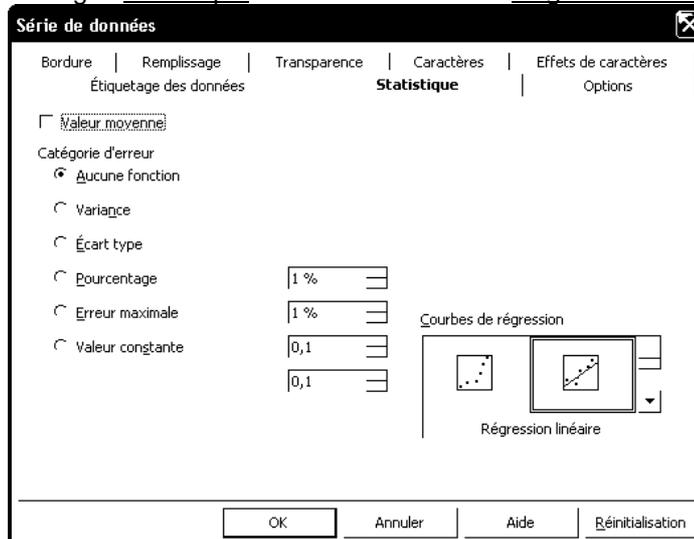
Quatrième dialogue : Titre de l'axe des x : Stress, titre de l'axe des y : symptômes.

Pour modifier un élément du graphique après avoir terminé la phase de création avec l'assistant, on peut double-cliquer sur cet élément.

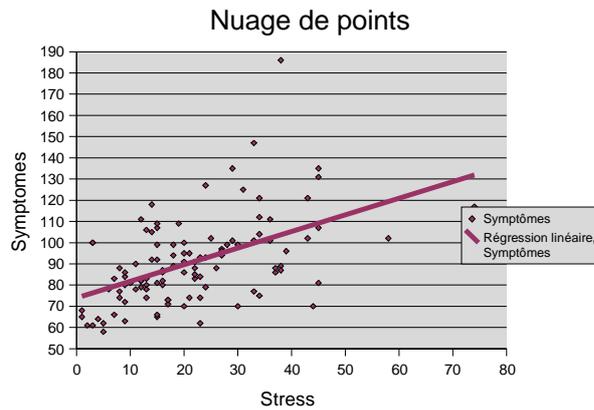


Construction du nuage de points et de la droite de régression :

Reprendre le graphique précédent. Double-cliquez sur l'un des points. Le dialogue "Série de données" s'affiche alors. Affichez l'onglet Statistique et sélectionnez l'item Régression linéaire.



On obtient ainsi la figure suivante :



5) Corrélations et régressions linéaires à plusieurs variables

Les tableurs ne sont guère adaptés aux traitements de la régression linéaire multiple. Nous reviendrons donc à Minitab pour traiter cette partie.

5.1 - Coefficient de corrélation multiple, hyperplan de régression

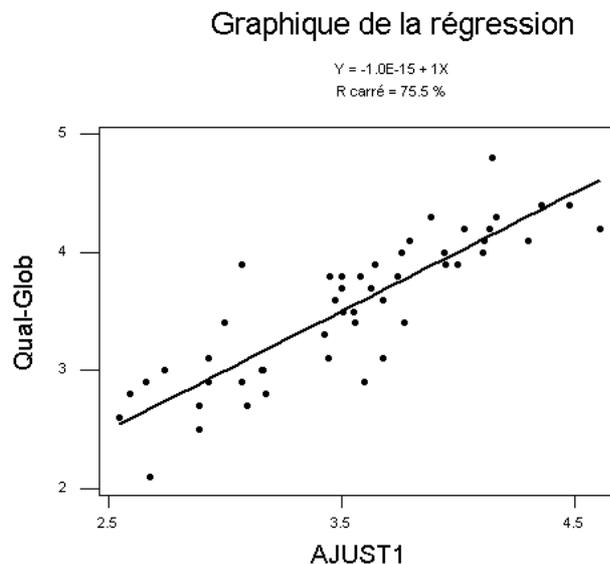
Ouvrir le fichier W:\PSY3\TD-Excel\Eval-Cours.MPJ. Il rassemble les données figurant dans un exercice des fiches de TD de Statistiques. L'énoncé accompagnant ces données est rappelé au début de la fenêtre Session.

- Calculer les paramètres descriptifs des six variables.
 - Calculer les coefficients de corrélation des variables prises deux à deux.
- Quels sont les couples de variables pour lesquels la corrélation est significative, au seuil de 5% ?
- On veut estimer la variable Qual-Glob en utilisant comme prédicteurs les 5 autres variables.
- Déterminer l'équation de régression et le coefficient de corrélation (menu Stat - Régression - Régression). Vous devriez obtenir :

L'équation de régression est
**Qual-Glob = - 1.19 + 0.763 Pédagogie + 0.132 Examen +
 0.489 Connaissance - 0.184 Résultat + 0.000525 Inscription**

s = 0.3202 R-carré = 75.5% R-carré (ajus) = 72.8%

Utilisez le bouton "Stockage" pour créer une nouvelle colonne donnant les valeurs ajustées de la variable Qual-Glob. On peut alors visualiser la qualité de l'ajustement en construisant une droite d'ajustement entre les variables AJUST et Qual-Glob :



5.2 - Calcul des coefficients de corrélation partielle

On veut calculer le coefficient de corrélation partielle entre la variable Qual-Glob et la variable Pédagogie. Nous allons procéder en trois étapes :

- Déterminez les résidus de la régression de la variable Qual-Glob par rapport aux 4 autres variables (Examen, Connaissance, Résultat, Inscription). Utilisez pour cela le menu Stats-Régression-Régression et utilisez le bouton Stockage et cochez l'item Valeurs résiduelles. Les résidus de cette régression sont stockés dans la première colonne libre (variable RESI1).

- Déterminez de même les résidus de la régression de la variable Pédagogie par rapport aux 4 autres variables, et stockez les dans une colonne RESI2.

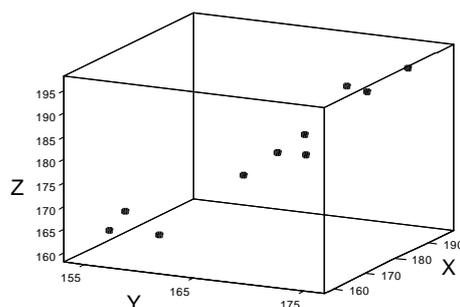
- Calculez enfin le coefficient de corrélation entre les variables RESI1 et RESI2 (menu Stats - Stats Élémentaires - Corrélation...). Vous devriez retrouver le résultat du polycopié, à savoir : $r=0,65$.

Enregistrez votre fichier et joignez-le à un mail adressé à votre enseignant.

6) Exercice de Monitorat

Le fichier W:\PSY3\TD-Minitab\Tailles.MPJ rassemble les données d'un exercice de la fiche de TD de statistiques.

Utilisez le menu Graphique - Graphique 3D... pour représenter le nuage de points dans l'espace :



Calculer les paramètres de statistiques descriptives de chacune des trois séries de données, puis les covariances des variables prises deux à deux.

Déterminer l'équation du plan de régression de la variable Z par rapport aux variables X et Y.

Calculer le coefficient de corrélation global.

Donner la prédiction de la taille Z lorsque $X=188$ et $Y=171$.

A l'aide des items de menu Copier et Coller, transférez les résultats obtenus dans Word et rédigez un rapport de l'étude qui a été faite.

N.B. Les résultats obtenus pourront être comparés à ceux fournis en réponse dans la fiche de TD de statistiques.