Principe des tests statistiques

- La notion d'intervalle de confiance emmène naturellement au principe de **test statistique** : à partir de l'étude d'un échantillon, que peut-on en conclure pour la population ?
  - ➢ On reste dans le domaine de l'inférence
  - ➢ La plupart des tests statistiques sont basés sur la construction d'intervalles de confiance

- Par définition, un test statistique est associé à des probabilités et un risque : on cherche à **prendre une décision en sachant que l'on peut se tromper**
  - ➢ On déterminera un seuil d'erreur *alpha* que l'on qui définit le niveau de risque que l'on est prêt à accepter

- De **nombreux tests statistiques** existent, selon:
  - ➢ L'hypothèse que l'on va tester (comparaison de proportions, de moyennes…)
  - ➢ Le nombre et la nature des variables à tester
  - ➢ La vérification ou non de certaines conditions (normalité, taille d'échantillons, égalité des variances…)

La principe difficulté consiste à choisir le bon test, puis l'interpréter correctement

Etapes d’un test statistique

- Tous les tests statistique reposent sur la même succession d’étapes :

1) Identifier le **paramètre** sur lequel on va faire une hypothèse
2) Définir une **hypothèse de départ** et **l’hypothèse alternative**
3) Définir **le risque d’erreur** que l’on est prêt à accepter
4) Calculer la **statistique de test**
5) **Prendre une décision** : rejet ou non de l’hypothèse de départ avec le risque d’erreur associé
6) **Conclusion** et interprétation des résultats

Définition des hypothèses

- On appelle **hypothèse nulle** ou H0 le scénario par défaut que l'on veut tester

  Exemples d'hypothèses nulles :

  « Il n'y a pas de différence entre les deux moyennes »

  « Il n'y a pas de lien entre les deux variables »

  « la moyenne est égale à 25 »

- **L'hypothèse alternative** ou H1 est par définition son opposé

  « Il y a une différence entre les deux moyennes »

  « Il y a un lien entre les deux variables »

  « La moyenne est différente de 25 »

- L'objectif de tout test statistique est de **tester l'hypothèse H0** :
  - Soit l'on rejette H0 : on conclut (avec risque de se tromper) que cette hypothèse est fausse
  - Soit on ne rejette pas H0 : rien ne permet d'affirmer que cette hypothèse est fausse

P-value et rejet de H0

- Tous les tests statistiques fournissent une p-valeur ou ***pvalue*** : la probabilité de se tromper si l'on rejette H0
  - Elle est par construction toujours entre 0 et 1
  - Elle est déterminée en fonction de la statistique de test (qui à elle-seule, n'est pas interprétable)

- C'est la **comparaison de cette pvalue avec le risque d'erreur** initialement choisi qui va permettre de de **rejetter ou non** l'hypothèse de départ :

  Si la *pvalue* est inférieure au risque d'erreur -> Je rejette H0

  Si la *pvalue* est supérieure au risque d'erreur -> Je ne rejette pas H0

Soit un test statistique basé sur l'hypothèse H0 « *les résultats des femmes sont identiques à ceux des hommes* »

Une pvalue de 0.02 nous indique qu'il n'y a que 2% que cela soit due au hasard :

Au risque 5%, je rejette H0

Cela n'est pas du au hasard, il y a bien une différence entre hommes et femmes

Au risque 1%, je ne rejette pas H0

Je ne peux pas affirmer avec ce niveau de risque qu'il y a une différence entre hommes et femmes

Tests de comparaison à une norme

- **Objectif** : comparer l'estimateur obtenu avec un **échantillon** avec une **valeur pour la population**
  - L'écart entre la valeur obtenu par l'échantillon et la valeur testée est-elle due au hasard / à l'échantillonnage ?

- **Hypothèses** pour une comparaison de moyenne à une norme:

  H0 : la moyenne de la population **est égale** à la valeur *mu*

  H1 : la moyenne de la population **est différente** de cette valeur *mu*

  ou inférieur / supérieur à dans le cas d'un test unilatéral

- **Principe** : Il s'agit en fait de mesurer l'**écart entre l'estimateur et la distribution qu'on aurait si l'hypothèse nulle était vérifiée** : cet écart est-il dû au hasard ou est-il le reflet d'une véritable différence?
  - Il s'agit de voir si l'estimateur tombe ou non dans l'intervalle de confiance autour de la valeur *mu*

- **Conditions** : l'estimateur (moyenne ou proportion) doit suivre une loi normale
  - On considèrera que c'est le cas dès qu'il y a plus de 30 individus
  - Dans le cas contraire, il faut tester la normalité de l'échantillon

Tests de comparaison à une norme sous R

- On utilisera les mêmes fonctions que pour la construction d'un intervalle de confiance autour d'une moyenne / d'une proportion
  - On précise en plus la valeur à laquelle on se compare

- Test de comparaison de moyenne avec une valeur de référence avec la fonction ***t.test(v, mu)***

  Soit v le vecteur numérique des valeurs de l'échantillon

  Soit **mu la valeur de référence** sur laquelle on effectue le test

  Paramètre *conf.level* pour fixer le seuil de significativité (0.95 par défaut)

  Paramètre *alternative* pour préciser si c'est un un test bilatéral (« *two.sided* », par défaut) ou unilatéral *(« greater* » ou « *less* »*)*

```
> t.test(age, mu = 30, alternative = "two.sided")

        One Sample t-test

data:  age
t = 0.98915, df = 99, p-value = 0.325
alternative hypothesis: true mean is not equal to 30
95 percent confidence interval:
 29.89560 30.31195
sample estimates:
mean of x
 30.10378
```

- Test de comparaison de proportion avec une valeur de référence avec la fonction ***prop.test(x, n, p)***

  Soit *x* le nombre de fois où l'évènement a lieu

  Soit *n* le nombre total d'évènements

  Soit **p la proportion de référence** sur laquelle on effectue le test

  Paramètre *conf.level*

  Paramètre *alternative*

```
> prop.test(20, 100, p = 0.21)

        1-sample proportions test with continuity correction

data:  20 out of 100, null probability 0.21
X-squared = 0.015069, df = 1, p-value = 0.9023
alternative hypothesis: true p is not equal to 0.21
95 percent confidence interval:
 0.1292482 0.2943230
sample estimates:
  p
0.2
```

Tests de comparaison de moyennes

- **Objectif** : comparer les **moyennes de deux échantillons indépendants** entre elles
  - L'écart entre les deux moyennes est-elle due au hasard / à l'échantillonnage ?

- **Hypothèses** pour une comparaison de moyennes de deux échantillons :

  H0 : les **moyennes** des deux sous-populations sont **égales**

  H1 : les **moyennes** des deux sous-populations sont **différentes**

  ou inférieur / supérieur à dans le cas d'un test unilatéral

- **Principe** : Il s'agit en fait de calculer l'intervalle de confiance de **la différence des deux moyennes observées** : est-ce que 0 appartient à cet intervalle (ce qui implique des moyennes similaires) ?
  - On parle de t.test ou test de student, car cette différence suit une loi de student (approximation de la loi normale)

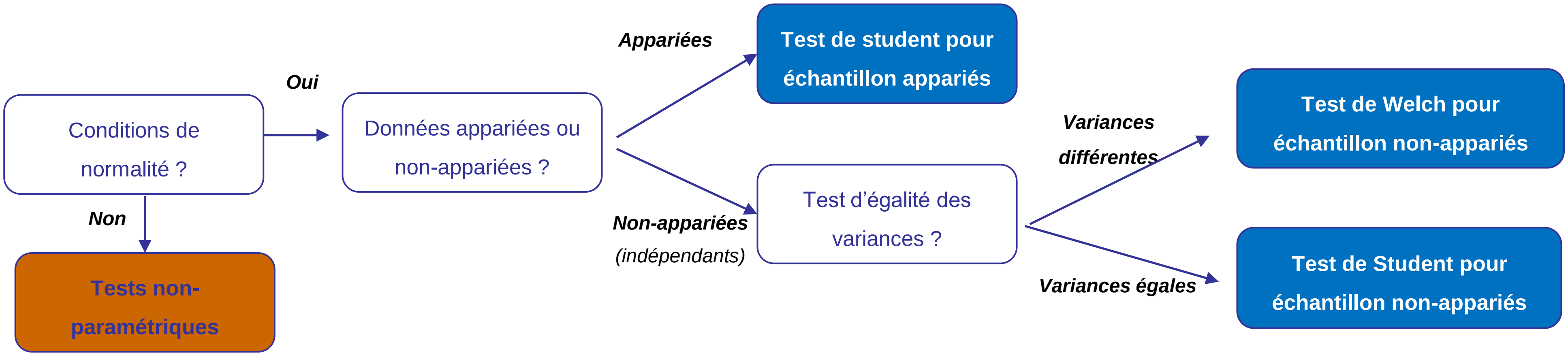
- **Conditions** : les deux estimateurs doivent suivre une loi normale
  - On considèrera que c'est le cas dès qu'il y a plus de 30 individus dans chaque échantillon
  - Dans le cas contraire, il faut tester la normalité de chaque échantillon

Tests de comparaison de moyennes

- Le test de student ne fonctionne *en principe* que sous la **condition d'égalité des variances** des deux populations.
  - Un test de student implique donc en premier lieu un **test d'égalité des variances**
  - Si cette condition n'est pas respectée, c'est **le test de Welch** qui est utilisé à la place (au fonctionnement similaire)

- **Variante** du **test de student pour échantillons appariés** : les deux échantillons ne sont pas indépendants
  - Dans cette variante, le paramètre étudié est la moyenne des différences entre les deux valeurs

- **Décision** : si la **p.value** donnée par le test est **inférieure au seuil** de risque choisi, on rejette H0 et on peut donc conclure que les **moyennes ne sont pas égales**.



Tests de comparaison de moyennes sous R

- Fonction t.test(x, y) pour le test de student et ses variantes

  Soit x et y les deux vecteurs numériques correspondant aux deux échantillons

  Paramètre *paired* pour préciser si les échantillons sont appariés (FALSE par défaut)

  Paramètre *conf.level* pour fixer le seuil de significativité (0.95 par défaut)

  Paramètre *var.equal* pour préciser si les variances des populations sont égales ou non (FALSE par défaut)

  Paramètre *alternative* pour préciser si c'est un un test bilatéral (« *two.sided* », par défaut) ou unilatéral *(« greater »* ou *« less »*

  Alternative : t.test (x ~ y) si y est une variable qualitative à deux niveaux

- Fonction ***var.test****(x, y)* pour tester au préalable **l'égalité des variances**

```
# Egalité des variances ?
var.test(longueur_petales_setosa, longueur_petales_virginica)
```

```
        F test to compare two variances

data:  longueur_petales_setosa and longueur_petales_virginica
F = 0.30729, num df = 49, denom df = 49, p-value = 6.366e-05
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1743776 0.5414962
sample estimates:
ratio of variances
         0.3072862
```

```
# Test de Welch à 90%, échantillons indépendants
t.test(longueur_petales_setosa, longueur_petales_virginica,
       paired  = FALSE, var.equal = FALSE, conf.level = 0.9)
```

```
        Welch Two Sample t-test

data:  longueur_petales_setosa and longueur_petales_virginica
t = -15.386, df = 76.516, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
90 percent confidence interval:
 -1.753196 -1.410804
sample estimates:
mean of x mean of y
    5.006     6.588
```

Test d'indépendance du khi2

- **Objectif** : tester **l'indépendance** entre **deux variables qualitatives**
  - ➢ Existe-t-il un lien entre les deux variables, la connaissance de l'une influençant les valeurs de l'autre?

- **Hypothèses** pour le test d'indépendance du khi2:

  H0 : il y a indépendance / il n'y a pas de lien entre les deux variables qualitatives x et y

  H1 : il n'y a pas indépendance / il existe un lien entre les deux variables qualitatives x et y

- **Principe** :
  - ➢ **Comparer les effectifs réels observés et les effectifs dit théoriques** (cas où les deux variables sont indépendantes)
  - ➢ Calculer la **statistique du Khi2** (basée sur la différence entre effectifs théoriques et effectifs réels) et la comparer à une loi du khi2
  - ➢ Conclure sur le **lien d'indépendance ou non** entre les deux variables
    - ✓ La pvalue nous indique la probabilité que l'hypothèse d'indépendance soit vraie, c'est-à-dire que les deux variables soient indépendantes (si elle est très faible, alors les deux variables sont liées)

Effectifs théoriques	Originaux	Reprise
Couleur & N&B	7612	297
Couleur	175	7
N&B	253	10

Ecart?

Effectifs observés	Originaux	Reprise
Couleur & N&B	7154	179
Couleur	157	12
N&B	140	88

Test d'indépendance du khi2

- **Condition** : les **effectifs théoriques** de chaque croisement doivent être au moins de taille **5**
  - Si ce n'est pas le cas, il faut regrouper des modalités pour que ce seuil soit atteint

- **Décision** : si la **p.value** donnée par le test est **inférieure au seuil de risque** choisi, on rejette H0 et on peut donc conclure que les **deux variables ne sont pas indépendantes**.

- **Variante** du **test du khi 2 d'homogénéité** : similaire au test d'indépendance, si ce n'est qu'il s'agit de comparer la distribution d'**une** variable qualitative d'un échantillon par rapport à une **norme**
  - Il s'agit en fait d'une extension du test de proportions par rapport à une norme, mais pour un variable ayant (potentiellement) plus de deux modalités
  - Exemple : déterminer si la répartition des profils par tranche d'âge d'un échantillon de clients interrogés par sondage est la même que celle en base de donnée

Tests du khi2 sous R

- Rappel : ***table****(x, y)* pour créer au préalable un tableau de contingence

  prop.table*(table(x,y))* pour un tableau de contingence un pourcentages (préciser *margin = 1* ou *margin = 2 pour pourcentages lignes / colonnes*)

- Fonction ***chisq.test(x, y)*** pour le **test d'indépendance** du **khi2**

  Soit x et y les deux variables qualitatives de l'échantillon testé

  Alternative : utiliser la fonction sur le tableau de contingences plutôt que les données d'origine

  Récupération des **effectifs théoriques** avec la composante **$expected** du test

  Un message de type « *approximation du khi2 utilisée : le test peut être faux* » est affiché si les conditions du test ne sont pas respectées (effectifs théoriques > 5)

```
> chisq.test(data$couleur, data$reprise)

        Pearson's Chi-squared test

data:  data$couleur and data$reprise
X-squared = 977.7, df = 2, p-value < 2.2e-16
```

```
> chi2$expected
                 data$reprise
data$couleur           FALSE       TRUE
  Couleur          7611.7261 297.273881
  Couleur et N&B    175.1592   6.840795
  N&B               253.1147   9.885324
> chi2$observed
                 data$reprise
data$couleur      FALSE  TRUE
  Couleur          7714   195
  Couleur et N&B    167    15
  N&B               159   104
```

- *chisq.test(x, p)* pour le test **d'homogénéité**

  Avec x le vecteur numérique des fréquences sur l'échantillon et p les probabilités théoriques pour comparaison

```
> tab_couleurs <- table(data$couleur)
> chisq.test(tab_couleurs, p = c(0.95,0.02,0.03))

        Chi-squared test for given probabilities

data:  tab_couleurs
X-squared = 2.0378, df = 2, p-value = 0.361
```

Test d’adéquation à une loi normale

- **Objectif** : tester si un échantillon de données suit une loi normale
  - Dans le cas d’échantillon réduit (n < 30), conclure sur la normalité de l’échantillon permettra de savoir si un test paramétrique peut ou non être envisagé
  - Complète de façon plus robuste une analyse visuelle comme un *qqplot*, qui compare les quantiles d’une loi normale théorique à ceux observés sur l’échantillon (alignement des points si normalité parfaite)

- **Hypothèses** pour le test :

  H0 : la distribution de l’échantillon est normale

  H1 : la distribution de l’échantillon n’est pas normale

- **Principe** : comparer la distribution des données de l’échantillon avec ce que donnerait une distribution normale
  - Le test habituellement considéré comme le plus robuste est celui de Shapiro-Wilk
  - Si la pvalue est inférieure au seuil fixé, on rejette l’hypothèse de normalité de l’échantillon

- **Conditions** : aucune condition particulière (si ce n’est que les données doivent être quantitatives)

Test d’adéquation à une loi normale sous R

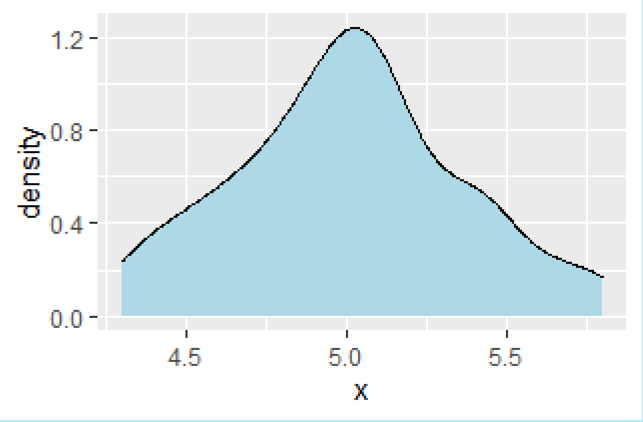
- Fonction ***shapiro.test*(x)** pour **tester la normalité des données** d’un échantillon avec le test de Shapiro-Wilk

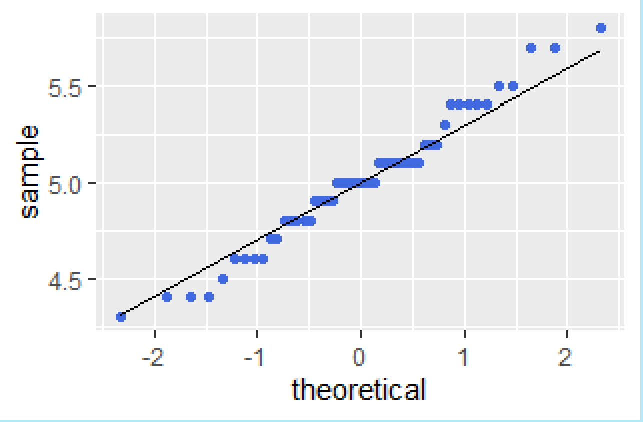
  Soit x le vecteur numérique des données de l’échantillon

```
> shapiro.test(longueur_petales_setosa)

        Shapiro-Wilk normality test

data:  longueur_petales_setosa
W = 0.9777, p-value = 0.4595
```

- Analyses visuelles complémentaire avec ggplot de type **courbe de densité** ou ***qqplot***

```
ggplot(data, aes(sample = x)) + geom_qq(col = "royal blue") + geom_qq_line()
```






```
ggplot(data) + geom_density(aes(x), fill = "light blue")
```

Les tests non-paramétriques

- Tous les tests vus jusqu'à présent reposent sur un ajustement de la statistique de test par rapport à une loi (de student, de khi2...) : c'est-ce qu'on appelle les tests paramétriques.
  - Ils supposent certaines conditions dans les données (la normalité pour les tests de Student)

- A l'opposé, les tests non-paramétriques ne reposent sur **aucune hypothèse sur la distribution des données** : en ce sens ils n'ont **aucune condition** sur les données pour pouvoir être utilisés.
  - Ils sont donc adaptés pour des situations pour des petits échantillons, ou pour des distributions très atypiques.

- Il existe des équivalents "non-paramétriques" pour la plupart des tests statistiques. Ils sont **moins puissants que leurs équivalents paramétriques**, mais peuvent être utilisés dans un plus grand nombre de situations.
  - Moins puissants : la probabilité de rejeter H0 sera en général plus faible pour un test non-paramétrique

Quelques tests non-paramétriques

- La plupart des tests non-paramétriques se basent sur les **rangs** des variables plutôt que leur mesure quantitative.
  - La notion de rang évite de se rapporter à la comparaison de distributions, et neutralise l'effet des valeurs extrêmes

- Le **test de Wilcoxon – Mann - Whitney** permet de deux **comparer les distributions** de **deux échantillons indépendants** (non-appariés)
  - C'est une alternative non-paramétrique au test de student de comparaison de moyennes sur échantillons indépendants

- Le **test de Wilcoxon** permet également de **comparer les distributions de deux échantillon**s, mais sur **données appariées**
  - C'est une alternative non-paramétrique au test de student de comparaison de moyennes sur échantillons appariés

- Le **test de Krustall-Wallis** permet la comparaison de **distributions de plus de deux échantillons**
  - C'est une alternative non-paramétrique à l'analyse de la variance...

Tests non-paramétriques sous R

- Fonction wilcox.test(x, y) pour les tests de Wilcoxon ou Mann-Whitney

  Soit x et y les deux vecteurs numériques correspondant aux deux échantillons

  Paramètre *paired* = *FALSE* si échantillons indépendants / test de Mann-Whitney

  Paramètre *paired* = *TRU*E si échantillons appariés / test de Wilcoxon

  Même fonctionnement que pour le test de student (arguments *conf.leve*l et *alternative*)

  Alternative : *wilcox.test(x ~ y)* si y est une variable qualitative à deux niveaux

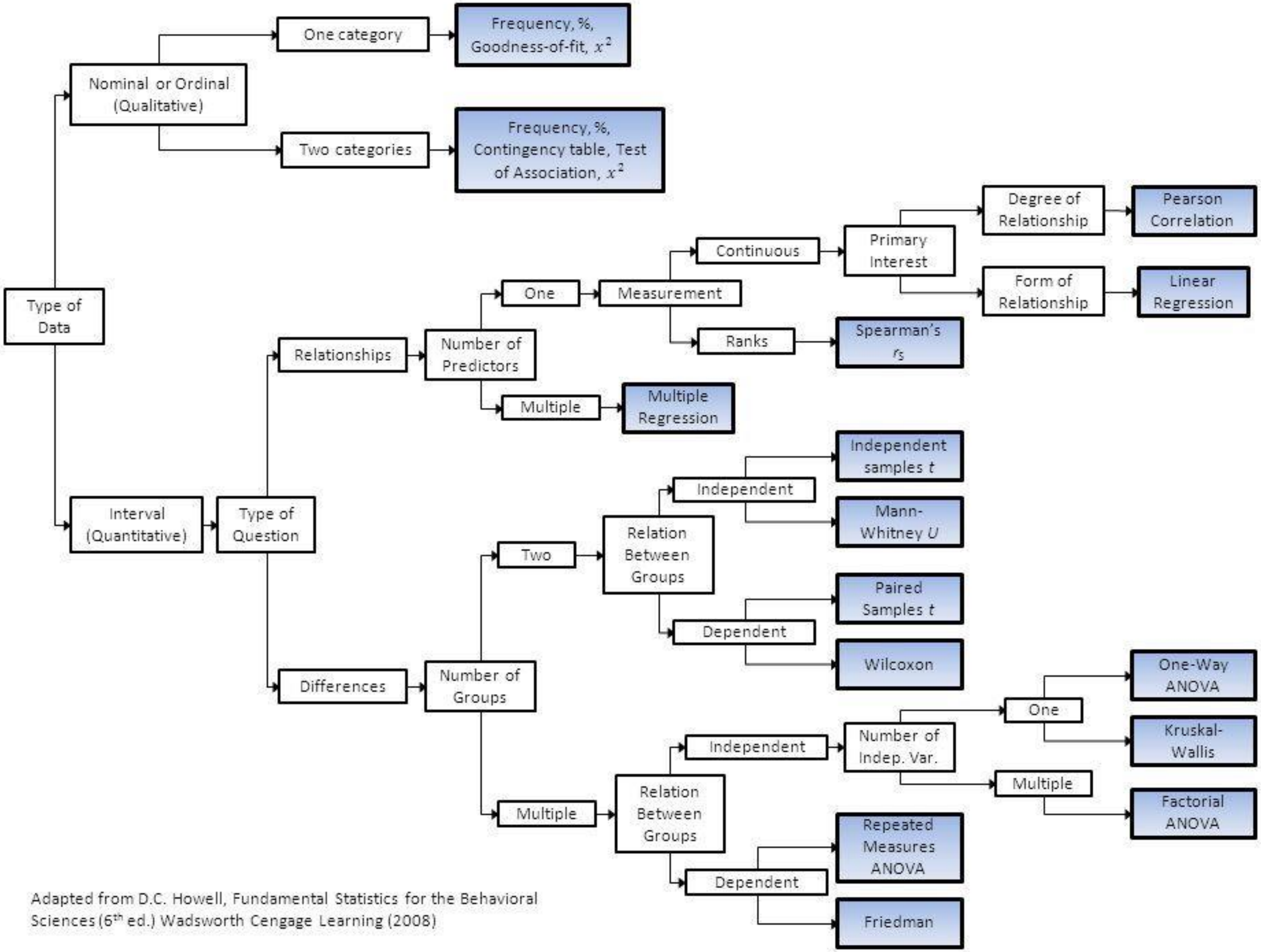
```
age_hommes <- c(18,24,54,32,50,33, 40, 23, 28, 34, 39, 12)
age_femmes <- c(25, 14, 35, 19, 21, 26, 56, 30, 20, 9)

wilcox.test(age_hommes, age_femmes)
```

```
        Wilcoxon rank sum exact test

data:  age_hommes and age_femmes
W = 80, p-value = 0.203
alternative hypothesis: true location shift is not equal to 0
```

Tests statistiques : pour résumer…

- Etudier les données
- Poser les **hypothèses, *nulles*** et alternatives
- Choisir le **test approprié**
  - Mettre en place d'éventuels tests de normalité des donnés / égalité des variances
- **Mettre en œuvre** le test
  - Facile sous R, si on a choisi le bon test
- **Conclure avec la *pvalue*** : si elle est suffisamment faible, on rejette H0
  - Sinon, on ne *rejette pas* H0

Cartographie des tests



Adapted from D.C. Howell, Fundamental Statistics for the Behavioral Sciences (6th ed.) Wadsworth Cengage Learning (2008)

Exercice : tests statistiques

- **Objectif : mettre en œuvre différents types de test statistique, selon l'objectif voulu**
  - La phase d'analyse descriptive a pu mettre en évidence certains phénomènes dans les données de l'échantillon ; il s'agit désormais de mettre en place des tests pour valider ou non ces hypothèses sur la population.

- **Hypothèses à tester** :
  1. D'après les informations en bases de données sur les ventes réelles, on sait que le magazine contient en réalité 95% de lecteurs et 5% de lectrice. Peut-on considérer que le **ratio de lecteurs / lectrices dans l'échantillon soit conforme** à cette réalité ?
  2. On s'intéresse à la satisfaction des lecteurs par rapport à celle des lectrices. Peut-on considérer qu'il y a une **différence de satisfaction en fonction du genre** ?
  3. Existe t'il un **lien entre l'ancienneté d'un lecteur et le type de lecteurs** ? Dit autrement, est-ce que parmi les lecteurs abonnés Web ou non-abonnés par exemple, il semble y avoir plus de lecteurs avec de l'ancienneté ?
  4. On sait d'après la base de données clients qu'il y a en réalité 50% d'abonnés papier, 20% d'abonnés Web et 30% de non-abonnés. Peut-on considérer que la **répartition des types de lecteurs dans l'échantillon soit conforme à cette répartition** ?
  5. La rédaction du magazine a un intérêt particulier concernant les **lecteurs très récents, de moins d'un an**. Peut-on dire que sur ces lecteurs-là, les **non-abonnés aient un niveau de satisfaction différent des abonnés ?** (Web ou papier)

- **Marche à suivre**:
  1. Pour chacune de ces hypothèses à tester, mettre en place le ou les test(s) approprié(s).
  2. Conclure sur chacune d'entre elles avec un risque d'erreur fixé à 5%.