Test Statistiques Inférentielles - Test 1

Sujet: Enquêtes et Tests d'Hypothèses
Niveau: Intermédiaire
Nombre de questions: 25

Questions et Réponses

Q1. Quelle est la différence entre statistique descriptive et statistique inférentielle?

R1. | Descriptive | Inférentielle | |————-|—————| | **Décrit** les données | **Généralise** à une population | | Résume un échantillon | Tire des conclusions | | Pas d'incertitude | Quantifie l'incertitude | | Moyenne, écart-type | Tests, intervalles de confiance |

Q2. Qu'est-ce qu'une hypothèse nulle (H_0) et alternative (H_1)?

R2. - **H_0 (Hypothèse nulle):** Affirmation de "pas d'effet" ou "pas de différence". C'est ce qu'on cherche à rejeter. - **H_1 (Hypothèse alternative):** Ce qu'on veut démontrer.

Exemple: - H_0: Le taux de défaut = 5% (pas de changement) - H_1: Le taux de défaut ≠ 5% (changement)

Q3. Qu'est-ce que la p-value et comment l'interpréter?

R3. La **p-value** est la probabilité d'obtenir un résultat aussi extrême (ou plus) que celui observé, **si H_0 est vraie**.

Interprétation: - $p < \alpha$ → Rejeter H_0 (résultat statistiquement significatif) - $p \geq \alpha$ → Ne pas rejeter H_0

Attention: - p-value N'EST PAS la probabilité que H_0 soit vraie - Seuil classique: $\alpha = 0.05$ (5%)

Q4. Quelle est la différence entre erreur de Type I et Type II?

R4. | | H_0 vraie | H_0 fausse | |–|———-|———–| | **Rejeter H_0** | **Type I (α)** Faux positif | Correct ✓ | | **Ne pas rejeter H_0** | Correct ✓ | **Type II (β)** Faux négatif |

Exemples bancaires: - Type I: Refuser un bon client (faux positif de risque) - Type II: Accepter un mauvais client (faux négatif de risque)

Q5. Qu'est-ce que la puissance d'un test?

R5. Puissance = 1 - β

La probabilité de rejeter H_0 quand elle est effectivement fausse (détecter un vrai effet).

Facteurs qui augmentent la puissance: - Augmenter la taille d'échantillon (n) - Augmenter α (mais augmente aussi Type I) - Effet plus grand - Réduire la variabilité

Standard: Puissance ≥ 80%

Q6. Comment calculer un intervalle de confiance pour une moyenne?

R6. Formule (n > 30 ou σ connu):

```
IC = x̄ ± z × (s/√n)
```

Pour 95%: z = 1.96 **Pour 99%:** z = 2.576

Exemple: x̄ = 100, s = 20, n = 100 IC 95% = 100 ± 1.96 × (20/10) = 100 ± 3.92 = [96.08, 103.92]

```python
from scipy import stats
confidence = 0.95
mean = df['montant'].mean()
se = stats.sem(df['montant'])
ci = stats.t.interval(confidence, len(df)-1, loc=mean, scale=se)
```

Q7. Quand utiliser un test z vs un test t?

R7. | Test z | Test t | |——-|——-| | σ population **connu** | σ population **inconnu** | | n **grand** (>30) | n **petit** (<30) | | Distribution normale | Distribution normale |

En pratique: Test t presque toujours (σ rarement connu).

Q8. Expliquez le test t pour comparer deux moyennes indépendantes.

R8. Hypothèses: - H_0: $\mu_1 = \mu_2$ (pas de différence) - H_1: $\mu_1 \neq \mu_2$ (différence)

```python
from scipy.stats import ttest_ind

# Test t (variances égales assumées)
t_stat, p_value = ttest_ind(group1, group2)

# Test Welch (variances inégales)
t_stat, p_value = ttest_ind(group1, group2, equal_var=False)

if p_value < 0.05:
    print("Différence significative")
```

Q9. Qu'est-ce qu'un test apparié (paired t-test)?

R9. Utilisé quand les **mêmes sujets** sont mesurés deux fois (avant/après).

Exemple: Satisfaction client avant et après une formation.

```python
from scipy.stats import ttest_rel

# Test apparié
t_stat, p_value = ttest_rel(before, after)
```

Hypothèses: - H_0: $\mu\_\text{différence} = 0$ - H_1: $\mu\_\text{différence} \neq 0$

Q10. Quand utiliser l'ANOVA (Analysis of Variance)?

R10. Pour comparer les moyennes de **3 groupes ou plus**.

Exemple: Comparer les montants moyens entre 4 agences.

```python
from scipy.stats import f_oneway

# ANOVA
f_stat, p_value = f_oneway(agence1, agence2, agence3, agence4)
```

Si p < 0.05: Au moins une moyenne diffère. **Ensuite:** Tests post-hoc (Tukey) pour identifier lesquelles.

Q11. Qu'est-ce que le test du Chi-carré et quand l'utiliser?

R11. Test pour **variables catégorielles**.

Types: 1. **Test d'indépendance:** Deux variables sont-elles liées? 2. **Test d'ajustement:** La distribution correspond-elle à un modèle?

```python
from scipy.stats import chi2_contingency

# Tableau de contingence
contingency = pd.crosstab(df['secteur'], df['defaut'])

# Test Chi-carré
chi2, p_value, dof, expected = chi2_contingency(contingency)
```

Q12. Comment interpréter le résultat d'un test Chi-carré?

R12. Exemple de résultat: - $Chi^2 = 15.3$ - df = 4 - p-value = 0.004

Interprétation: - $p < 0.05$ → Les variables sont **significativement associées** - Le secteur et le défaut ne sont pas indépendants - Certains secteurs ont des taux de défaut différents

---

Q13. Quelle est la différence entre test unilatéral et bilatéral?

R13. | Bilatéral | Unilatéral | |———-|———-| | H_1: $\mu \neq \mu_0$ | H_1: $\mu > \mu_0$ ou $\mu < \mu_0$ | | Différence dans n'importe quel sens | Direction spécifique | | p-value complète | p-value / 2 | | Plus conservateur | Plus puissant si direction connue |

Règle: Utiliser bilatéral sauf si direction clairement justifiée.

Q14. Comment calculer la taille d'échantillon nécessaire?

R14. Pour estimer une proportion:

```
n = (z² × p × (1-p)) / e²
```

Exemple: Estimer taux de défaut avec précision 2% - z = 1.96 (95% confiance) - p = 0.05 (estimation) - e = 0.02 (marge d'erreur)

$n = (1.96^2 \times 0.05 \times 0.95) / 0.02^2 =$ **457 clients minimum**

```python
from statsmodels.stats.power import zt_ind_solve_power
n = zt_ind_solve_power(effect_size=0.5, alpha=0.05, power=0.8)
```

Q15. Qu'est-ce que le théorème central limite (TCL)?

R15. Énoncé: Quelle que soit la distribution de la population, la distribution des moyennes d'échantillons tend vers une normale quand n est grand.

Implications: - Permet d'utiliser les tests z/t même si les données ne sont pas normales - Seuil pratique: $n \geq 30$ - Plus l'échantillon est grand, meilleure est l'approximation

---

Q16. Comment vérifier la normalité des données?

R16. Tests statistiques:

```python
from scipy.stats import shapiro, normaltest

# Shapiro-Wilk (n < 5000)
stat, p = shapiro(data)

# D'Agostino-Pearson
stat, p = normaltest(data)
```

Méthodes visuelles: - QQ-plot - Histogramme avec courbe normale

Si p > 0.05: Distribution approximativement normale.

Q17. Quand utiliser les tests non-paramétriques?

R17. Utiliser quand: - Données non-normales - Échantillon petit - Données ordinales - Présence d'outliers

Test Paramétrique	Équivalent Non-Paramétrique
t-test indépendant	**Mann-Whitney U**
t-test apparié	**Wilcoxon signé**
ANOVA	**Kruskal-Wallis**

Q18. Comment interpréter un intervalle de confiance?

R18. IC 95% = [45%, 55%] pour un taux de défaut.

Interprétation correcte: "Si on répétait l'échantillonnage 100 fois, 95 des intervalles contiendraient la vraie valeur."

Interprétation pratique: "On est confiant à 95% que le vrai taux est entre 45% et 55%."

INCORRECT: "Il y a 95% de chances que le vrai taux soit dans cet intervalle."

Q19. Qu'est-ce que la correction de Bonferroni pour les tests multiples?

R19. Quand on fait plusieurs tests, le risque d'erreur Type I augmente.

Correction:

```
 _ajusté =   / k
```

Exemple: 5 tests avec $\alpha = 0.05$ $\alpha_{ajusté} = 0.05 / 5 = 0.01$

Utiliser pour chaque comparaison individuelle.

Q20. Comment conduire un test de proportion?

R20. Hypothèses: - H_0: $p = p_0$ - H_1: $p \neq p_0$

```python
from statsmodels.stats.proportion import proportions_ztest

# Données
successes = 48  # Ex: 48 défauts
n = 1000        # sur 1000 prêts
p0 = 0.05       # proportion attendue

# Test
z_stat, p_value = proportions_ztest(successes, n, p0)
```

Q21. Comment comparer deux proportions?

R21. Exemple: Comparer taux de défaut entre deux agences.

```python
from statsmodels.stats.proportion import test_proportions_2indep

# Données
count1, n1 = 45, 1000  # Agence A: 45 défauts sur 1000
count2, n2 = 60, 1200  # Agence B: 60 défauts sur 1200

# Test
z_stat, p_value = test_proportions_2indep(
    count1, n1, count2, n2, compare='diff'
)
```

Q22. Qu'est-ce que la taille d'effet et pourquoi est-elle importante?

R22. La **taille d'effet** mesure l'importance pratique d'un résultat (vs significativité statistique).

Cohen's d (différence de moyennes):

```
d = (  -  ) / s_pooled
```

Interprétation: - d < 0.2: Effet faible - 0.2 < d < 0.8: Effet modéré - d > 0.8: Effet fort

Importance: Un résultat peut être statistiquement significatif mais pratiquement négligeable.

Q23. Comment conduire un A/B test pour une campagne marketing bancaire?

R23. Étapes: 1. **Hypothèse:** H_0: Conversion_A = Conversion_B 2. **Randomisation:** Assigner aléatoirement les clients 3. **Taille échantillon:** Calculer n nécessaire 4. **Exécution:** Collecter les données 5. **Analyse:**

```python
# Données A/B test
conversions_A, n_A = 120, 1000
conversions_B, n_B = 150, 1000

# Test de proportion
from statsmodels.stats.proportion import test_proportions_2indep
z, p = test_proportions_2indep(conversions_A, n_A, conversions_B, n_B)

# Intervalle de confiance pour la différence
diff = conversions_B/n_B - conversions_A/n_A
```

Q24. Quelles sont les conditions d'application d'un test t?

R24. Conditions: 1. **Indépendance:** Observations indépendantes 2. **Normalité:** Distribution normale (ou n > 30) 3. **Homogénéité des variances:** Pour test à deux échantillons (sinon Welch)

Vérification:

```python
# Normalité
from scipy.stats import shapiro
shapiro(data)

# Homogénéité des variances
from scipy.stats import levene
levene(group1, group2)
```

Q25. Interprétez le résultat suivant: - Test: t-test comparant le score crédit moyen des clients en défaut vs non-défaut - t = -8.45 - p-value = 0.0000001 - Moyenne défaut: 520, Moyenne non-défaut: 680 - Cohen's d = 1.2

R25. Interprétation:

1. **Significativité:** $p < 0.001$ → Différence **hautement significative**
2. **Direction:** t négatif → Score moyen des défauts < Score non-défauts
3. **Amplitude:**
   - Différence = 160 points (680 - 520)
   - Cohen's d = 1.2 → **Effet très fort**
4. **Conclusion business:**
   - Le score de crédit est un **excellent discriminant** du risque
   - Les clients en défaut ont en moyenne 160 points de moins
   - Justifie l'utilisation du score dans le processus de décision

---

Scoring

Score	Niveau
0-10	À améliorer
11-17	Intermédiaire
18-22	Avancé
23-25	Expert