جامعة جيجل

Université de Jijel

كلية العلوم الإنسانية والاجتماعية

قسم علم الاجتماع

مقياس الإحصاء التطبيقي في البحوث الاجتماعية

الإنحدار الخطي المتعدد

Multiple Linear Regression

إعداد الدكتورة: سامية بوكحيل

السنة الجامعية: 2024 - 2025

مقدمة: ما هو الإنحدار الخطي المتعدد؟

الإنحدار الخطي المتعدد (Multiple Linear Regression) هو امتداد طبيعي وتطوير للإنحدار الخطي البسيط، حيث يتضمن أكثر من متغير مستقل واحد في المعادلة التنبؤية. بدلاً من التنبؤ بالمتغير التابع Y بناءً على متغير مستقل واحد X فقط، يتيح الإنحدار المتعدد استخدام عدة متغيرات مستقلة (X1, X2, X3, ... Xk) في آن واحد، مما يُحسّن دقة التنبؤ ويُعطي صورة أكثر واقعية وشاملة عن العلاقات بين المتغيرات.

في الواقع الاجتماعي، نادراً ما يتأثر ظاهرة واحدة بعامل واحد فقط. فالتحصيل الدراسي مثلاً لا يتأثر بساعات الدراسة فحسب، بل بالذكاء والدافعية والمستوى الاقتصادي والبيئة الأسرية وغيرها. الإنحدار البسيط يُتجاهل هذه الحقيقة ويُعطي صورة مُبسّطة جداً، بينما الإنحدار المتعدد يأخذ جميع هذه العوامل بعين الاعتبار ويُقدّم نموذجاً أكثر تمثيلاً للواقع المعقد. يُعد هذا الأسلوب من أكثر الأدوات الإحصائية استخداماً في البحوث الاجتماعية والتربوية والنفسية والاقتصادية.

- **تعريف الإنحدار الخطي المتعدد:** هو أسلوب إحصائي يهدف إلى إيجاد معادلة خطية تُمثّل العلاقة بين متغير تابع واحد (Y) وعدة متغيرات مستقلة (X1, X2, ... Xk)، بحيث تُمكّن هذه المعادلة من التنبؤ بقيم Y بناءً على قيم المتغيرات المستقلة مجتمعة.
- **المعادلة العامة:** $Y = a + b1X1 + b2X2 + b3X3 + ... + bkXk + e$
- **الفرق عن الإنحدار البسيط:** الإنحدار البسيط يتعامل مع متغير مستقل واحد، بينما المتعدد يتعامل مع عدة متغيرات مستقلة في معادلة واحدة، مما يُحسّن القدرة التفسيرية والتنبؤية.

محتويات المحاضرة

1. أهمية الإنحدار المتعدد في البحث العلمي
2. الفرق بين الإنحدار البسيط والمتعدد

3. المعادلة العامة ومكوناتها
4. الفرضيات الأساسية
5. طريقة المربعات الصغرى OLS في الإنحدار المتعدد
6. معامل التحديد R^2 ومعامل التحديد المعدّل Adjusted R^2
7. اختبار F لدلالة النموذج
8. اختبار t لدلالة معاملات الانحدار الفردية
9. تعدّي الخطية (Multicollinearity)
10. تحليل المُتبقيات وفحص الفرضيات
11. تطبيق عملي مفصّل خطوة بخطوة
12. التطبيق باستخدام برنامج SPSS
13. تطبيقات في العلوم الاجتماعية

أولاً: أهمية الإنحدار المتعدد في البحث العلمي

يتميز الإنحدار الخطي المتعدد بأهمية خاصة في البحث العلمي التطبيقي لعدة أسباب جوهرية تجعله الأداة المفضلة لدى كثير من الباحثين في مجالات متنوعة. إن القدرة على دمج عدة متغيرات في نموذج واحد تفتح آفاقاً واسعة للفهم والتفسير والتنبؤ.

1. تمثيل الواقع المعقد

الظواهر الاجتماعية والنفسية والتربوية معقدة بطبيعتها ومتأثرة بعدد كبير من العوامل المتشابكة. الإنحدار المتعدد يُتيح للباحث بناء نماذج أكثر واقعية تأخذ هذا التعقيد بعين الاعتبار. فبدلاً من دراسة تأثير عامل واحد في عزلة، يمكن دراسة تأثير عدة عوامل معاً، مع التحكم في تأثير كل عامل على الآخر. هذا يُعطي صورة أوضح وأكثر دقة عن الواقع المُعقّد.

2. التحكم الإحصائي في المتغيرات

يُمكّن الإنحدار المتعدد الباحث من التحكم إحصائياً في تأثير المتغيرات المُربكة (Confounding Variables) دون الحاجة إلى تصميم تجريبي. فمن خلال إدخال هذه المتغيرات في المعادلة كمتغيرات مستقلة، يتم التحكم في تأثيرها تلقائياً، وتظهر معاملات المتغيرات الأخرى كأنها خالية من هذا التأثير. هذا يُعادل تقريباً عمل الارتباط الجزئي لكن بشكل أكثر شمولاً وتكاملاً.

2. مقارنة مساهمات المتغيرات

يُتيح الإنحدار المتعدد مقارنة أثر كل متغير مستقل في المتغير التابع عندما تكون المتغيرات الأخرى ثابتة. تُسمى هذه المعاملات بمعاملات الانحدار الجزئية (Partial Regression Coefficients)، وكل منها يقيس مقدار التغيّر في Y لكل وحدة تغيّر في Xi عندما تكون جميع المتغيرات المستقلة الأخرى ثابتة. هذا يُساعد في ترتيب المتغيرات حسب أهميتها وتأثيرها.

4. تحسين القدرة التنبؤية

كلما زاد عدد المتغيرات المستقلة المُفسِّرة في النموذج، زادت القدرة على تفسير التباين في المتغير التابع وبالتالي تحسنت دقة التنبؤ. فإذا كان الإنحدار البسيط يُفسّر مثلاً 40% من التباين، فقد يصل الإنحدار المتعدد إلى 70% أو أكثر بإضافة متغيرات مستقلة ذات صلة. بالطبع هناك حدود لهذا التحسن وتحذيرات من الإفراط في إضافة المتغيرات.

ثانياً: الفرق بين الإنحدار البسيط والمتعدد

لفهم الإنحدار المتعدد بشكل سليم، من المفيد مقارنته بالإنحدار البسيط الذي سبق دراسته. يوضح الجدول التالي أوجه الشبه والاختلاف بينهما بشكل مفصل:

وجه المقارنة	الإنحدار البسيط	الإنحدار المتعدد
عدد المتغيرات المستقلة	متغير واحد (X)	متغيران أو أكثر (X1, X2, ... Xk)

المعادلة	Y = a + bX	Y = a + b1X1 + b2X2 + … + bkXk
معامل الميل	بسيط: مقدار تغيّر Y لكل وحدة X	جزئي: تغيّر Y لكل وحدة Xi (ثبات الباقي)
معامل التحديد R^2	يساوي r^2 مباشرة	لا يساوي مجموع r^2 (للتبادل المشترك)
اختبار الدلالة	اختبار t للميل فقط	اختبار F للنموذج + اختبار t لكل معامل
تعدّي الخطية	لا ينطبق	مشكلة محتملة يجب فحصها
التعقيد	بسيط - يمكن حسابه يدوياً	معقد - يحتاج برامج إحصائية
التمثيل البصري	خط مستقيم في مستوى ثنائي الأبعاد	مستوي في فضاء متعدد الأبعاد
التطبيق	دراسات استكشافية أولية	نماذج تفسيرية متقدمة

● **ملاحظة جوهرية:** في الإنحدار البسيط، المعامل b يقيس التأثير الإجمالي لـ X على Y. أما في الإنحدار المتعدد، المعامل b1 يقيس التأثير الفريد (الصافي) لـ X1 على Y بعد إزالة تأثير جميع المتغيرات المستقلة الأخرى. هذا الفرق بين التأثير الإجمالي والتأثير الصافي هو جوهر الإنحدار المتعدد.

ثالثاً: المعادلة العامة ومكوناتها

تأخذ معادلة الإنحدار الخطي المتعدد شكلاً موسعاً عن الإنحدار البسيط، حيث تتضمن عدة متغيرات مستقلة بمعاملاتها الخاصة. المعادلة العامة هي:

$$Y = a + b1X1 + b2X2 + b3X3 + ... + bkXk + e$$

المكوّنات بالتفصيل

1. المتغير التابع (Y)

هو المتغير الذي نريد تفسير تباينه أو التنبؤ بقيمه. يجب أن يكون متغيراً كمياً مستمراً (فئوي أو نسبي). أمثلة: التحصيل الدراسي، مستوى الرضا، الدخل، الأداء الوظيفي، درجة القلق. لا يمكن استخدام متغيرات تابعة فئوية (اسمية) في الإنحدار الخطي العادي، بل يُستخدم الإنحدار اللوجستي (Logistic Regression) بدلاً منه.

2. التقاطع (a - Constant)

هو القيمة المتوقعة لـ Y عندما تكون جميع المتغيرات المستقلة مساوية للصفر (X1 = X2 = ... = Xk = 0). في كثير من الحالات قد لا يكون لهذا التقاطع تفسير منطقي، لكنه ضروري رياضياً لتحديد موضع خط (أو مستوى) الانحدار في الفضاء.

3. معاملات الانحدار الجزئية (b1, b2, ... bk)

كل معامل bi يُقيس مقدار التغيّر في Y لكل وحدة تغيّر واحدة في المتغير Xi عندما تكون جميع المتغيرات المستقلة الأخرى ثابتة (لا تتغير). تُسمى هذه المعاملات "جزئية" لأنها تقيس التأثير الصافي لكل متغير مستقل. فمثلاً: b2 = 3 تعني أن Y يزيد بمقدار 3 وحدات لكل وحدة زيادة في X2، بغض النظر عن قيم X1 و X3 وغيرها.

4. خطأ المصادفة (e - Error Term)

يُمثّل الفرق بين القيمة الفعلية لـ Y والقيمة المُتنبأ بها من المعادلة. يشمل كل العوامل غير المُدرجة في النموذج التي تؤثر في Y. يُفترض أن يتبع توزيعاً طبيعياً بمتوسط صفري وتباين ثابت.

مثال تطبيقي

معادلة تنبؤ التحصيل الدراسي (Y) من ساعات الدراسة (X1) والذكاء (X2) والدافعية (X3):

$$Y = 2 + 1.5X1 + 0.8X2 + 2.3X3$$

- **b1 = 1.5:** كل ساعة دراسة إضافية تزيد التحصيل بمقدار 1.5 درجة (عند ثبات الذكاء والدافعية).
- **b2 = 0.8:** كل وحدة زيادة في الذكاء تزيد التحصيل بمقدار 0.8 درجة (عند ثبات الدراسة والدافعية).
- **b3 = 2.3:** كل وحدة زيادة في الدافعية تزيد التحصيل بمقدار 2.3 درجة (عند ثبات الدراسة والذكاء).
- **التأثير الأقوى:** الدافعية (b3 = 2.3) هي المتغير الأكثر تأثيراً.

رابعاً: الفرضيات الأساسية

يعتمد الإنحدار الخطي المتعدد على مجموعة من الفرضيات الإضافية بالإضافة إلى فرضيات الإنحدار البسيط. انتهاك هذه الفرضيات يؤثر على دقة التقديرات وموثوقية الاختبارات الإحصائية.

1. الخطية (Linearity)

يجب أن تكون العلاقة بين كل متغير مستقل والمتغير التابع خطية. كما يجب أن تكون العلاقة مجتمعة (الجمع الخطي للمتغيرات المستقلة) طريقة مناسبة لوصف تأثيرها المشترك. يمكن فحص ذلك من خلال مخططات الانتشار الجزئية (Partial Regression Plots) أو بإضافة متغيرات تفاعلية ومتغيرات تربيعية للنموذج.

2. الاستقلالية (Independence)

يجب أن تكون المُتبقيات مستقلة عن بعضها البعض، أي عدم وجود ارتباط تسلسلي (Autocorrelation). يُفحص ذلك باختبار دوربين-واتسون (Durbin-Watson) الذي يجب أن تكون قيمته قريبة من 2 (المجال المقبول عادة 1.5 - 2.5).

3. تجانس التباين (Homoscedasticity)

يجب أن يكون تباين المُتبقيات ثابتاً عند جميع مستويات المتغيرات المستقلة المُتنبأ بها. يُفحص ذلك بمخطط المُتبقيات مقابل القيم المُتنبأ بها (يجب أن يكون التوزيع عشوائياً بشكل مربع). يُعالج عدم التجانس باستخدام الانحدار الموزون أو تحويلات البيانات.

4. التوزيع الطبيعي (Normality)

يجب أن تتبع المُتبقيات توزيعاً طبيعياً بمتوسط صفري. يُفحص ذلك بمخطط Q-Q أو اختبار شابيرو-ويلك. عند حجم العينات الكبيرة (أكبر من 100)، يكون الانحراف عن الطبيعية أقل تأثيراً بسبب مبرهنة النهاية المركزية.

5. عدم وجود تعدّي خطية عالٍ (No Multicollinearity)

وهي فرضية خاصة بالإنحدار المتعدد وتعني أن المتغيرات المستقلة يجب ألا تكون مرتبطة ببعضها ارتباطاً عالٍ جداً. إذا كان الارتباط بين متغيرين مستقلين أعلى من 0.80، يُصبح تمييز تأثير كل منهما صعباً وتصبح المعاملات غير مستقرة. يُفحص ذلك بعامل تضخم التباين VIF.

المؤشر	القيمة المقبولة	الدلالة
VIF	أقل من 5 (ويفضل أقل من 10)	كلما كان VIF أكبر، زادت مشكلة التعدّي الخطي
Tolerance	أكبر من 0.1 (ويفضل أكبر من 0.2)	القيمة العكسية لـ VIF (Tolerance = 1/VIF)
Condition Index	أقل من 30	أكبر من 30 يدل على تعدّي خطي شديد

⚠ **ملاحظة مهمة عن حجم العينة:** كقاعدة عامة، يُنصح بأن يكون حجم العينة على الأقل 10 مرات عدد المتغيرات المستقلة ($n \geq 10k$)، وأفضل 20 مرة ($n \geq 20k$). فإذا كان النموذج يضم 5 متغيرات مستقلة، فالحد الأدنى هو 50 مشاهدة، والأفضل 100 مشاهدة.

خامساً: طريقة المربعات الصغرى OLS في الإنحدار المتعدد

تُستخدم نفس طريقة المربعات الصغرى العادية (OLS) في تقدير معاملات الإنحدار المتعدد، لكنها أصبحت أكثر تعقيداً لأننا نبحث الآن عن أفضل مستوي (أو مستوى فرعي) وليس أفضل خط. الهدف هو إيجاد المعاملات (a, b1, b2, ... bk) التي تُقلّص مجموع مربعات المُتبقيات:

$$\text{Minimize } \Sigma [Y(i) - (a + b1X1(i) + b2X2(i) + ... + bkXk(i))]^2$$

التمثيل المصفوفي

يمكن كتابة المعادلة في صورة مصفوفية، وهو الأسلوب المستخدم في البرامج الإحصائية:

$$Y = X\beta + e$$

حيث Y هو متجه القيم الفعلية (n × 1)، و X هي مصفوفة المتغيرات المستقلة (n × k+1)، و β هو متجه المعاملات (k+1 × 1)، و e هو متجه الأخطاء (n × 1). حل OLS يُعطى بالمعادلة:

$$\beta = (X'X)^{-1} X'Y$$

هذه المعادلة المصفوفية هي الأساس الرياضي لحساب جميع معاملات الانحدار المتعدد. البرامج الإحصائية مثل SPSS و R و Stata تقوم بهذه الحسابات تلقائياً ولا يحتاج الباحث لإجرائها يدوياً. إن الفهم النظري لهذه العملية يُساعد الباحث على تفسير النتائج وفهم حدود النموذج.

المعادلة الطبيعية لحساب المعاملات

في حالة متغيرين مستقلين (k = 2)، يمكن حساب المعاملات بالمعادلات التالية:

$$b1 = [r(y1) - r(y2)\times r(12)] / (1 - r^2(12)) \times (Sy / Sx1)$$

$$b2 = [r(y2) - r(y1)\times r(12)] / (1 - r^2(12)) \times (Sy / Sx2)$$

حيث r(y1) و r(y2) هما معاملات ارتباط Y بكل من X1 و X2، و r(12) هو ارتباط X1 بـ X2. يُلاحظ أن المقام (1 - r²(12)) يأخذ بعين الاعتبار الارتباط بين المتغيرين المستقلين. كلما زاد هذا الارتباط، زاد المقام في المعادلة الأصلية، مما يؤدي إلى معاملات غير مستقرة - وهذا هو بالضبط ما يُسمى مشكلة تعدّي الخطية.

سادساً: معامل التحديد R^2 ومعامل التحديد المعدّل

يُعد معامل التحديد من أهم المؤشرات لتقييم جودة نموذج الإنحدار المتعدد. لكن في الإنحدار المتعدد يظهر فرق جوهري بين R^2 العادي و R^2 المعدّل (Adjusted R^2).

معامل التحديد R^2

$$R^2 = SSR / SST = 1 - (SSE / SST)$$

يُقيس النسبة المئوية من التباين الكلي في Y التي يُفسّرها النموذج (جميع المتغيرات المستقلة مجتمعة). يتراوح بين 0 و 1. لكن في الإنحدار المتعدد، يواجه R^2 مشكلة أساسية: **إنه لا ينقص أبداً عند إضافة متغير مستقل جديد**، حتى لو كان هذا المتغير عديم الفائدة. بل إنه يزيد قليلاً بسبب الصدفة البحتة، مما قد يُعطي انطباعاً مُضللاً بتحسن النموذج.

معامل التحديد المعدّل Adjusted R^2

$$R^2(adj) = 1 - [(SSE/(n-k-1)) / (SST/(n-1))] = 1 - [(1-R^2)$$

```
(n-1)/(n-k-1)]
```

يُعدّل هذا المعامل R^2 الأخذ بعين الاعتبار عدد المتغيرات المستقلة وحجم العينة. على عكس R^2 العادي، يمكن لـ R^2 المعدّل أن ينقص عند إضافة متغير مستقل لا يُساهم بشكل حقيقي في تفسير التباين. لذلك يُعد R^2 المعدّل المؤشر الأكثر موثوقية في الإنحدار المتعدد ويُستخدم للمقارنة بين نماذج مختلفة.

متى نستخدم كل منهما؟

المؤشر	الاستخدام	الميزة	القصور
R^2	وصف نسبة التباين المُفسَّر	سهل الفهم والتفسير	يزداد دائماً مع إضافة متغيرات
R^2 المعدّل	مقارنة نماذج مختلفة	يعاقب الإضافة غير المفيدة	قد يكون سالباً في النماذج الضعيفة

● **قاعدة ذهبية:** إذا كان R^2 أعلى بكثير من R^2 المعدّل، فهذا مؤشر على وجود متغيرات غير مفيدة في النموذج. أما إذا تقاربت القيمتان، فهذا يدل على أن جميع المتغيرات في النموذج مُفيدة ومُساهمة بشكل حقيقي.

سابعاً: اختبار F لدلالة النموذج

يُستخدم اختبار F لتحديد ما إذا كان النموذج ككل (جميع المتغيرات المستقلة مجتمعة) دالاً إحصائياً في تفسير التباين في المتغير التابع. هذا اختبار شامل يختلف عن اختبار t الذي يختبر كل متغير على حدة.

الفرضيات

- **H_0 (الفرضية العدمية):** جميع معاملات الانحدار تساوي صفراً (B1 = B2 = ... = Bk = 0)، أي أن النموذج لا يُفسّر أي تباين في Y.
- **H_1 (الفرضية البديلة):** على الأقل واحد من المعاملات لا يساوي صفراً، أي أن النموذج يُفسّر جزءاً من التباين.

إحصاء الاختبار

```
F = (SSR / k) / (SSE / (n - k - 1)) = MSR / MSE
```

حيث k هو عدد المتغيرات المستقلة، و n هو حجم العينة. درجات الحرية هي: df1 = k (للبسط) و df2 = n - k - 1 (للمقام).

التفسير

قيمة Sig.	القرار	المعنى
Sig. < 0.05	نرفض H_0	النموذج دال إحصائياً - المتغيرات مجتمعة تُفسّر جزءاً معنوياً من Y
Sig. ≥ 0.05	نقبل H_0	النموذج غير دال - لا توجد علاقة خطية بين المتغيرات المستقلة و Y

⚠ **تنبيه:** دلالة اختبار F لا تعني بالضرورة أن جميع المتغيرات المستقلة مُهمة. فقد يكون النموذج دالاً بسبب متغير واحد فقط بين عدة متغيرات. لذلك يجب دائماً إجراء اختبار t لكل معامل على حدة لتحديد أي المتغيرات هي المُساهمة فعلياً.

ثامناً: اختبار t لدلالة معاملات الانحدار الفردية

بعد التأكد من دلالة النموذج ككل (اختبار F)، يأتي دور اختبار t لفحص دلالة كل معامل انحدار على حدة. هذا الاختبار يُحدد ما إذا كان كل متغير مستقل يُساهم بشكل مستقل ومعنوي في تفسير التباين في Y.

إحصاء الاختبار

$$t = bi / Sbi$$

حيث bi هو معامل الانحدار (الميل الجزئي)، و Sbi هو الخطأ المعياري لمعامل الانحدار. درجات الحرية: df = n - k - 1.

التفسير

قيمة t و Sig.	المعنى	الإجراء
t كبير، Sig. < 0.05	المتغير المُستقل يُساهم بشكل معنوي	الاحتفاظ بالمتغير في النموذج
t صغير، Sig. ≥ 0.05	المتغير لا يُساهم بشكل معنوي	دراسة إزالته من النموذج
t كبير، Sig. < 0.001	إسهام عالي الدلالة	متغير بالغ الأهمية

● **ملاحظة مهمة:** قد يكون المتغير مرتبطاً بقوة بالمتغير التابع (معامل ارتباط عالٍ)، لكنه يظهر غير دال في الإنحدار المتعدد. هذا يحدث عادة عندما يكون هذا المتغير مرتبطاً أيضاً بمتغير مستقل آخر في النموذج (تعدّي خطي). في هذه الحالة، تأثيره مُستوعَب بالفعل من قبل المتغير الآخر.

معاملات بيتا الموحدة (Standardized Beta)

للمقارنة النسبية بين مساهمات المتغيرات المستقلة المختلفة (التي قد تكون بمقاييس مختلفة)، تُستخدم معاملات بيتا الموحدة (Beta). هذه المعاملات تُحوّل جميع المتغيرات إلى مقياس موحد (Z-scores)، مما يُتيح مقارنة أثر كل متغير مباشرة.

$$Beta(i) = bi \times (Sxi / Sy)$$

المتغير ذو القيمة المطلقة الأكبر لـ Beta هو الأكثر تأثيراً في المتغير التابع. على سبيل المثال، إذا كانت Beta(X1) = 0.45 و Beta(X2) = 0.30، فهذا يعني أن X1 أكثر تأثيراً بـ 50% من X2 تقريباً.

تاسعاً: تعدّي الخطية (Multicollinearity)

تُعد مشكلة تعدّي الخطية من أخطر المشكلات في الإنحدار المتعدد. تحدث عندما يكون هناك ارتباط عالٍ بين متغيرين أو أكثر من المتغيرات المستقلة، مما يجعل من الصعب عزل التأثير الفريد لكل متغير ويؤدي إلى معاملات غير مستقرة وغير موثوقة.

أنواع التعدّي الخطي

1. التعدّي الخطي الكامل (Perfect Multicollinearity)

يحدث عندما يكون أحد المتغيرات المستقلة مزيجاً خطياً دقيقاً لمتغيرات مستقلة أخرى (مثلاً X3 = X1 + X2). في هذه الحالة لا يمكن تقدير المعاملات أصلاً لأن المصفوفة (X'X) تكون غير قابلة للانعكاس. هذه الحالة نادرة في الممارسة العملية.

2. التعدّي الخطي المرتفع (High Multicollinearity)

يحدث عندما يكون الارتباط بين بعض المتغيرات المستقلة مرتفعاً (أعلى من 0.80) لكنه ليس كاملاً. في هذه الحالة يمكن تقدير المعاملات لكنها تكون غير مستقرة: أي تغيير بسيط في البيانات يُغيّر المعاملات بشكل كبير، وتكون الأخطاء المعيارية للمعاملات كبيرة مما يُصعّب الوصول إلى دلالة إحصائية.

مؤشرات الكشف عن التعدّي الخطي

المؤشر	الطريقة	القيمة الآمنة	قيمة الخطر
مصفوفة الارتباطات	فحص ارتباطات المتغيرات المستقلة	$r < 0.80$	$r \geq 0.80$
VIF	عامل تضخم التباين	VIF < 5	VIF ≥ 10
Tolerance	1 / VIF	Tol > 0.2	Tol ≤ 0.1

Condition Index	من التحليل البصري للأبعاد	CI < 15	CI ≥ 30

طرق المعالجة

1. **إزالة أحد المتغيرات:** أبسط وأكثر الطرق فعالية. إذا كان متغيران مرتبطين بقوة، نُزيل الأقل أهمية نظرياً.
2. **دمج المتغيرات:** إنشاء متغير مركب (مؤشر) من المتغيرات المتشابهة بدلاً من إدخالها منفصلة.
3. **زيادة حجم العينة:** العينات الكبيرة تُقلّل من تأثير التعدّي الخطي.
4. **الانحدار الحزمي (Ridge Regression):** أسلوب إحصائي متقدم يُقلّل من تأثير التعدّي الخطي.
5. **تحليل المكونات الرئيسية (PCA):** تحويل المتغيرات المرتبطة إلى مكونات مستقلة.

عاشراً: تحليل المُتبقيات وفحص الفرضيات

تحليل المُتبقيات (Residual Analysis) هو خطوة أساسية في تقييم أي نموذج انحدار. المُتبقيات هي الفروق بين القيم الفعلية والمُتنبأ بها، وتخضع لفرضيات معينة يجب التحقق من تحققها.

1. التحقق من الخطية

يُرسم مخطط المُتبقيات مقابل القيم المُتنبأ بها (Residuals vs Fitted). إذا كانت العلاقة خطية، يجب أن يكون التوزيع عشوائياً حول الصفر بدون أي نمط واضح (مثل منحنى أو شكل قمع). إذا ظهر نمط، فهذا يدل على عدم خطية العلاقة ويجب تحويل البيانات أو إضافة متغيرات تربيعية.

2. التحقق من تجانس التباين

في نفس المخطط (Residuals vs Fitted)، إذا اتسع نطاق انتشار المُتبقيات أو ضاق بشكل منتظم مع تغيّر القيم المُتنبأ بها، فهذا يدل على عدم تجانس التباين. المعالجة تكون بتحويل المتغير التابع (مثل Log Y) أو استخدام الانحدار الموزون.

3. التحقق من التوزيع الطبيعي

يُرسم مخطط Q-Q (Quantile-Quantile Plot) للمُتبقيات. إذا كانت المُتبقيات تتبع توزيعاً طبيعياً، يجب أن تقع النقاط على الخط القطري تقريباً. الانحرافات الطفيفة عند الأطراف تكون مقبولة خاصة مع العينات الكبيرة.

4. الكشف عن القيم الشاذة والقيم المؤثرة

يجب فحص وجود قيم شاذة تؤثر بشكل كبير على النموذج. تُستخدم عدة مقاييس:

- **Residuals الموحدة (Standardized Residuals):** القيم المطلقة أكبر من 3 تعتبر شاذة.
- **Cook's Distance:** القيم أكبر من $4/n$ تدل على نقاط مؤثرة جداً.
- **Leverage (قوة النقطة):** القيم أكبر من $2(k+1)/n$ تدل على نقاط ذات تأثير عالٍ.
- **DFBETAS:** يقيس تأثير كل نقطة على كل معامل على حدة.

> ⚠ **قاعدة مهمة:** لا تحذف القيم الشاذة تلقائياً! يجب أولاً التحقق من أنها أخطاء في إدخال البيانات. إذا كانت صحيحة، فحلّلها بعناية - فقد تكون هذه النقاط هي الأكثر إثارة للاهتمام علمياً.

حادي عشر: تطبيق عملي مفصّل خطوة بخطوة

للتوضيح العملي، لنأخذ مثالاً يفحص العلاقة بين ثلاثة متغيرات مستقلة والتحصيل الدراسي (Y) لدى عينة من 7 طلاب:

- **X1:** ساعات الدراسة الأسبوعية
- **X2:** درجة الذكاء (من 100)
- **Y:** التحصيل الدراسي (من 20)

البيانات

الطالب	X1 (الدراسة)	X2 (الذكاء)	Y (التحصيل)
1	3	100	7
2	5	110	11
3	2	95	5
4	8	120	15
5	6	105	12
6	4	115	10
7	7	125	14

الخطوة 1: حساب معاملات الارتباط البسيطة

	Y	X1	X2
Y	1.00	0.99	0.97
X1	0.99	1.00	0.91
X2	0.97	0.91	1.00

الخطوة 2: المعادلة المُقدّرة

باستخدام طريقة المربعات الصغرى (أو برنامج SPSS)، نحصل على:

$$Y = -8.17 + 1.30X1 + 0.07X2$$

الخطوة 3: تفسير المعاملات

- **التقاطع (a = -8.17):** قيمة افتراضية عندما تكون ساعات الدراسة والذكاء صفراً.

- **b1 = 1.30:** كل ساعة دراسة إضافية تزيد التحصيل بمقدار 1.30 درجة (عند ثبات الذكاء).
- **b2 = 0.07:** كل درجة ذكاء إضافية تزيد التحصيل بمقدار 0.07 درجة (عند ثبات ساعات الدراسة).
- **الأكثر تأثيراً:** ساعات الدراسة (b1 = 1.30) هي الأكثر تأثيراً.

الخطوة 4: تقييم النموذج

المؤشر	القيمة	التفسير
R^2	0.995	النموذج يُفسّر 99.5% من التباين
Adjusted R^2	0.993	بعد تعديل عدد المتغيرات - مازال ممتازاً
F	392.5	النموذج دال إحصائياً (Sig. < 0.001)
VIF (X1)	5.8	عند الحد المقبول - يُنصح بالمراقبة
VIF (X2)	5.8	عند الحد المقبول - يُنصح بالمراقبة

✓ **التنبؤ:** طالب يدرس 5 ساعات وذكاؤه 110:

Y = -8.17 + (1.30 × 5) + (0.07 × 110) = -8.17 + 6.50 + 7.70 = 6.03 درجات

ملاحظة: التقاطع السالب لا يعني مشكلة رياضية، بل يعني أن المعادلة مُقدّرة بناءً على نطاق محدود من البيانات.

ثاني عشر: التطبيق باستخدام برنامج SPSS

يُعد برنامج SPSS من أسهل البرامج لتطبيق الإنحدار الخطي المتعدد، حيث يُوفر مخرجات شاملة تشمل جميع المؤشرات الإحصائية اللازمة.

1 **إدخال البيانات:** نُدخل المتغير التابع Y في عمود والمتغيرات المستقلة X1, X2, ... في أعمدة منفصلة.

2 **فتح التحليل:** Analyze → Regression → Linear

3 **تحديد المتغيرات:** ننقل Y إلى مربع Dependent، و X1, X2, ... إلى مربع Independent(s).

4 **اختيار طريقة الإدخال (Method):**

- **Enter:** إدخال جميع المتغيرات دفعة واحدة (الأكثر استخداماً).
- **Stepwise:** إدخال المتغيرات تدريجياً حسب أهميتها الإحصائية.
- **Forward:** إضافة المتغيرات الأكثر أهمية تباعاً.
- **Backward:** حذف المتغيرات الأقل أهمية تباعاً.

5 **الخيارات:** من Statistics نختار: Estimates, Confidence Intervals, Model fit, Descriptives, Collinearity diagnostics.

6 **مخططات:** من Plots نختار: Histogram, Normal probability plot, و *ZRESID مقابل *ZPRED.

قراءة مخرجات SPSS

الجدول	المؤشرات الرئيسية	الأهمية

Descriptive Statistics	المتوسطات والانحرافات المعيارية	فهم البيانات الأساسية
Correlations	مصفوفة الارتباطات بين جميع المتغيرات	الكشف المبدئي عن تعدّي الخطية
Model Summary	R, R^2, Adjusted R^2, Std. Error, Durbin-Watson	تقييم جودة النموذج
ANOVA	.F, Sig	دلالة النموذج ككل
Coefficients	B, Std. Error, Beta, t, Sig., VIF	دلالة وأهمية كل متغير
Collinearity Diagnostics	Eigenvalues, Condition Index	فحص متقدم لتعدّي الخطية

ثالث عشر: تطبيقات في العلوم الاجتماعية

يُستخدم الإنحدار الخطي المتعدد على نطاق واسع في مختلف فروع العلوم الاجتماعية. قدرته على التعامل مع عدة متغيرات في آن واحد تجعله الأداة المثالية لتحليل الظواهر الاجتماعية المعقدة.

1. في علم الاجتماع

- **الحراك الاجتماعي:** التحصيل = a + b1(التعليم) + b2(الدخل الأسري) + b3(الجنس) + b4(الموقع الجغرافي). يُمكّن من تحديد أي العوامل أكثر تأثيراً في الحراك الاجتماعي.
- **التمييز الوظيفي:** الأجر = a + b1(الخبرة) + b2(المستوى التعليمي) + b3(الجنس) + b4(القطاع). يُساعد في قياس الفجوة في الأجور بعد التحكم في العوامل المشروعة.
- **الجريمة:** معدل الجريمة = a + b1(البطالة) + b2(الكثافة السكانية) + b3(معدل الفقر) + b4(الإنفاق على التعليم).

2. في علم النفس

- **الصحة النفسية:** الاكتئاب = a + b1(الضغط) + b2(الدعم الاجتماعي) + b3(مرونة الشخصية) + b4(التاريخ العائلي).
- **الأداء الوظيفي:** الأداء = a + b1(الدافعية) + b2(الخبرة) + b3(الرضا الوظيفي) + b4(الذكاء العاطفي).
- **الإدمان:** شدة الإدمان = a + b1(الاكتئاب) + b2(القلق) + b3(الدعم الاجتماعي) + b4(العمر عند البدء).

3. في التربية

- **التحصيل:** التحصيل = a + b1(الذكاء) + b2(الدراسة) + b3(الدافعية) + b4(بيئة الأسرة) + b5(جودة المدرسة).
- **تسرب الطلاب:** التسرب = a + b1(الفقر) + b2(التحصيل السابق) + b3(الدعم الأسري) + b4(المشاركة الأنشطة).
- **فعالية التعليم:** تقييم أثر متغيرات متعددة على نتائج الطلاب.

4. في الاقتصاد

- **الاستهلاك:** C = a + b1(الدخل) + b2(الثروة) + b3(معدل الفائدة).
- **النمو الاقتصادي:** الناتج = a + b1(الاستثمار) + b2(القوى العاملة) + b3(التكنولوجيا).
- **سوق العمل:** الأجر = a + b1(الخبرة) + b2(التعليم) + b3(الجنس) + b4(القطاع) + b5(المنطقة).

الخلاصة: يُعد الإنحدار الخطي المتعدد من أكثر الأساليب الإحصائية فائدة وقوة في البحث العلمي التطبيقي. إن قدرته على دمج عدة متغيرات في نموذج تفسيري واحد يجعله الأداة المفضلة لتحليل الظواهر الاجتماعية والنفسية والتربوية المعقدة. فهم المفاهيم الأساسية للإنحدار المتعدد - من معادلات ومعاملات تحديد واختبارات دلالة ومشاكل تعدّي الخطية - يُمكّن الباحث من بناء نماذج علمية دقيقة وتفسير النتائج بشكل صحيح. يجب دائماً التحقق من الفرضيات وتحليل المُتبقيات وعدم المبالغة في تفسير النتائج خارج نطاق البيانات والإطار النظري.

المراجع والمصادر

1. Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (3rd ed.). Lawrence Erlbaum Associates.

2. Field, A. (2018). Discovering Statistics Using IBM SPSS Statistics (5th ed.). SAGE Publications.

3. Tabachnick, B. G., & Fidell, L. S. (2019). Using Multivariate Statistics (7th ed.). Pearson Education.

4. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to Linear Regression Analysis (5th ed.). Wiley.

5. Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). Applied Linear Statistical Models (5th ed.). McGraw-Hill.

6. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). Multivariate Data Analysis (8th ed.). Cengage Learning.

7. Pallant, J. (2020). SPSS Survival Manual (7th ed.). Routledge.

8. Stevens, J. P. (2009). Applied Multivariate Statistics for the Social Sciences (5th ed.). Routledge.

9. أمل سعيد، عبد الرحمن عدس (2010). مقدمة في علم الإحصاء الوصفي والاستدلالي. دار المناهج، عمّان.

10. فيديو: الإنحدار الخطي المتعدد - قناة إحصاء (Iori Kader) على YouTube.