الفصل الأول

مقدمـــــة

Introduction

يهدف هذا الكتاب إلى إعطاء مقدمه في الاستدلال الإحصائي في مجال الزراعة وعلوم الحياة. وعلى الرغم من أن قارئ هذا الكتاب قد لا يكون إحصائياً إلاّ أنه عند الانتهاء منه سيكون لديه القدرة على فهم مادة الموضوع ومن ثم استخدام معظم طرق التحليل الإحصائي المعتادة مع قدرته على مناقشة المسائل المعقدة بمهارة.

ما هو الإحصاء؟ ربما يكون المفهوم الأكثر شيوعاً للإحصاء بأنه يعبر عن بيانات عددية استخدمت كسجلات للماضي وبذلك فإن الإحصاء كبيانات عددية يمكن استخدامه للتعبير عن عدة مفاهيم سواء في الوقت الحاضر أم المستقبلي بناءً على طبيعة تلك المفاهيم ومدلولاتها بالنسبة للشخص المهتم بها. وكمثال على ذلك، خلال مواسم لعب كرة القدم ما هي المعلومات التي تكتب في الصحف اليومية عن المباريات تحت العمود المسمى إحصاءات وما مدى أهمية تلك الإحصاءات؟ تعتمد أهمية تلك البيانات على من يقرأها فإن كان القارئ لها هو مدرب الفريق المنافس فسيهتم بأعداد التمريرات والأخطاء ، ...إلخ في حين أن القارئ العادي مهتم بمعرفة نتيجة المباراة.

أمّا في حالة البيانات الزراعية فإن تاجر تربية المواشي يهتم بمعرفة أعداد القطيع في المرعى والأعداد التي تم إحلالها الأسبوع الماضي، متوسط سعر علائق التغذية.. إلخ وبذلك يستطيع بناءً على ذلك إدارة مشروعه.

وفي هذا السياق فإن العوامل الأساسية التي تحكم ملاءمة الإحصاءات من عدمها تعتمد على معرفة الشخص بالبيانات ومدى مناسبتها بالنسبة للقرارات المستقبلية. ويعرف الإحصاء الذي يهتم بالأرقام والتعامل معها بالإحصاء الوصفي Descriptive Statistics أما الاستدلال الإحصائي Statistical inference فيشير إلى عمليات صنع القرارات حيث يتضمن استخدام المنطق وكذلك الطرق العلمية في صناعة القرارات، حيث يركز على حل المشكلات وكما نعلم فإن المنطق يتضمن الأسباب من الخصوصية إلى العموم. في حالة الاستدلال الإحصائي فإن المعتاد هو جمع المشاهدات من وحدات صغيرة (عينة) من المجتمع وتحليل تلك البيانات باستخدام طرق التحليل المعتادة ثم الاستدلال بها نظراً لأن النتائج المتحصل عليها من العينة يجب أن تكون هي نفسها المتحصل عليها من جميع الوحدات (المجتمع).

وكمثال على ذلك فإن البث الإعلامي في الولايات المتحدة الأمريكية يتنبأ بمن سيربح الانتخابات العامة للرئاسة قبل إغلاق عملية الاقتراع في الشاطئ الغربي. وتتم هذه العملية بسؤال عينه مختارة بعناية من الناخبين كيف تتم عملية تصويتهم ثم يتم الاستدلال من النتائج المتحصل عليها ؛ نظراً لأن ذلك التصويت سوف يتم بنفس الطريقة من قبل المجتمع ككل وهذا ما يسمى بالمنطق الاستقرائي.

الطريقة العلمية تصف الإجراء المتبع في العلم لعمل التجربة. والذي يشتمل عدة خطوات تتمثل في تعريف المشكلة، وصياغة الفروض، وتحديد البيانات المطلوب جمعها وتحليلها، واختيار طريقة التحليل المناسبة ثم تحديد ملاءمة نتائج التحليل للمشكلة المدروسة. و يمكن التساؤل كيف يمكن للاستدلال الإحصائي أن يساهم في هذا الإجراء ؟

بعد صياغة المشكلة والفروض المناسبة يمكن اتخاذ القرار المناسب للاختبار- ما

هي البيانات اللازمة، كيف يمكن الحصول عليها وتنظيمها وتحليلها بالطريقة التي تعطي إجابة مناسبة للمشكلة محل الدراسة - ويعتبر الاستدلال الإحصائي إحدى الطرق الإحصائية للقيام بهذه المهمة ولكن يتم اختيار الأدوات اللازمة طبقاً لطبيعة المشكلة وطريقة التحليل الملائمة لها ونوع الخلاصة التي نرغب في الوصول إليها. وبصفة عامة إذا أردنا التعبير عن بعض الخلاصات في صورة صيغ احتمالية أو عندما نحاول التعبير عن درجة التأكد من شيء ما بحس احتمالي، فإنه يجب تطبيق بعض صيغ الاستدلال الإحصائي.

ربما الآن نستطيع أن نعرّف الإحصاء، يشتمل الإحصاء على تطوير وتطبيق الطرق والأساليب لجمع وتحليل وتفسير البيانات العددية، لذا فإن الاعتماد على النتائج المستخلصة من هذه العملية يمكن تقييمها بصورة موضوعية باستخدام التعابير الاحتمالية. والجزء الأخير من التعريف- استعمال الاحتمالات لتقييم موضوعية النتائج والاستدلال بناء على البيانات – هي المساهمة الوحيدة والرئيسة للإحصاء في الاستدلال الاستقرائي.

الاستعمالات الخاطئة الشائعة للإحصاء Common Misuses of Statistics

يقول ديسريلي Disreali بأن هناك الكذب ، الكذب الملعون ، والإحصائيين مما يجعل الإحصائيين مع بعضهم في الشركة غير مرغوب فيه. وقد كتب هوف Huff كتاب صغير مرتب أسماه ، *كيف تكذب باستخدام الإحصاء*، كما فسر ريتشمان Reichmann الاستعمال السيئ للإحصاء. والحقيقة القائلة بأن الآخرين يصلون إلى نفس الملاحظات تقودنا للاعتقاد بأنه ليس كل من يستخدم التحليل الإحصائي سوف يقوم بذلك بنفس الدرجة من الاهتمام أو الانتباه. ففي التعاملات التجارية وكذلك

الحال في حياتنا الشخصية يتم الدفاع باستخدام العروض المدعمة بالإحصاءات. لذا فإن أحد الأسباب لتعلم الإحصاء هو للحماية الذاتية! في الجزء التالي سوف نتطرق لمعظم الطرق الشائعة التي تؤدي للتضليل عند استخدام الإحصاء.

التحيز Bias

في الغالب يكون من السهل اكتشاف التحيز الواضح أو المحاباة في الإعلانات التي تستشهد بالإحصاءات لإثبات تفوق منتج معين، في حين أن إعلانات المنافسين تستشهد بإحصاءات أخرى لإثبات العكس. ولكن التحيز غير المقصود ينسلّ للعمل الإحصائي نتيجة التحيز في البيانات المجموعة باستخدام الاستبيانات أو التصويت الناتج من النقص أو الخطأ في طريقة المعاينة المستخدمة، والضعف في تصميم الاستبانة، والأسئلة الايحائية، أو من الباحث الذي يحاول/تحاول إثبات فكرته/فكرتها. وفي حالة استخدام البيانات في أي دراسة فإن المتوقع أن يشير الباحث لمصدر البيانات وعدم ذكر ذلك يوحي بأنه ربما يكون هناك مشكلة.

التعميمات الخاطئة Faulty Generalizations

نرى أحياناً التعميمات الخاطئة نظراً لكون العينة التي تم اختيارها كانت صغيرة جداً، مما يجعل من غير الممكن تمثيل المجتمع على نحو كافٍ بناءً على النتائج التي تم استخلاصها. فمثلاً، من الصعوبة تمثيل جميع أسماك فصيلة معينة بصورة عادلة في البحيرات العظمى باستخدام عينة مكونة من خمس سمكات. والأكثر سوءاً التعميم باستخدام عينة حجمها واحد والتي تستخدم في الغالب. والتعميمات الخاطئة الأخرى تأتي من العينات غير الممثلة مثل، العينات التي تكون كافية في حجمها ولكنها تختلف

عن المجتمع في بعض التقديرات المهمة، فمثلاً دراسة استقصائية عن خريجي الكلية من دفعة ١٩٤٠م أشارت بأن متوسط رواتبهم كانت ٩٥١٠٧ دولارات سنوياً، وهذه الدراسة مبنية على عدد ٢١ استجابة من عينة الدراسة البالغ عددهم ١٤٢. وبذلك تكون هذه النتيجة خاطئة إذا كان أولئك الذين استجابوا للدراسة لهم دخول عالية أكثر من الآخرين أو ربما أنهم قاموا بتجميع كل مكتسباتهم كما يحدث في مثل هذه الحالات. أيضاً إذا كان بعض الخريجين الذين استجابوا للدراسة يتقاضون دخول عالية جداً فإن ذلك سوف يؤدي لتضخيم المتوسط بدرجة كبيرة جداً.

الاستدلال الخاطئ Faulty Deduction

الخطأ الشائع للاستدلال الخاطئ يحدث من القيام بتطبيق قاعدة عامة صالحة كما لو كانت قانون ثابت. لذا فالتقرير الحكومي الختامي المتضمن بأن سعة مراعي الغزلان أكبر من الطلب بـ ١٠,٢٪ ربما تكون صحيحة على مستوى الدولة ولكن ليس من الضروري أن يكون ذلك صحيح لكل منطقة. فقد يكون هناك عجز أو نقص في الغرب وسعة زائدة بنسبة ١٥,٥٪ في الجنوب الشرقي. ولكن إذا أمكن صياغة التقرير كاتجاه عام يسمح بالاختلافات الفردية فإن هذه المشكلة قد لا تكون مزعجة.

البيانات غير المتماثلة Noncomparable data

المقارنات تتم عادة بين وحدات غير متشابهة فعلاً. فمثلاً أشار ممثل نقابة السيارات بأن سعر سيارة العائلة السيدان ذات الأربعة أبواب ارتفع تقريباً ٥٠٪ في آخر عشر سنوات. وقد أهمل الإشارة إلى أنه خلال تلك الفترة قام المصنع بإضافة العديد من الخصائص للسيارات مثل التجهيزات القياسية التي تجعل السيارات الجديدة تختلف

عن القديمة. والخطأ الناتج عن عدم التماثل يؤثر في مؤشر السعر بشكل عام ؛ نظراً لأن التقنية تتغير باستمرار والمواصفات الخاصة تتغير من فترة زمنية لأخرى.

الأخطاء في الدلالات Errors in Semantics

تستخدم أحيانا الكلمات الشاذة أو الملمّعة للتأثير على المستمعين كما في الحملات السياسية. لذا فإن الاحتراس واجب من الأسئلة في الاستبيانات التي ربما تقود أو توحي بإجابة معينة. فمثلاً، سؤال " هل تحب الشرطة" سوف يعطي إجابة تختلف قليلاً فيما لو تمت صياغة السؤال على النحو " ما هو رأيك في قسم الشرطة".

افتراض السببية نتيجة الارتباط Assuming causation from correlation

تعني هذه العبارة أنه نظراً لأن بعض الأشياء تسبق الآخر في الحدوث عبر الزمن فقد نفترض أن الأول يسبب الثاني. فمثلاً، افترض أنك تعرضت للعواصف الرعدية ومن ثم أصبت بنزلة برد. فإنك تفترض أن تعرضك للمطر هو سبب إصابتك بالبرد وليس بالضرورة أن يكون ذلك. بصفة عامة إذا كان العاملان A وB تتذبذب معاً (A وB مرتبطين) فربما أن A تسبب B ولكن أيضا قد يكون B تسبب A ، أو أن A و B تؤثر في بعضها بصفة مستمرة، أو بشكل متقطع، أو ربما A و B تتأثر بعامل آخر C أو ربما يكون الارتباط نتيجة الصدفة. وعليه فقد تكون تعرضت لفيروس برد غير معروف والذي هاجمك أثناء تعرضك للمطر. ولا يمكن استخدام الإحصاء للاستدلال عن السببية الناتج عن الارتباط. وإذا تم عمل مثل ذلك الاستدلال فإنه يجب أن يكون مبني على حكم مهني للباحث أو الباحثة والذي يعرف عن العوامل من واقع مهنته. وكل ما يقوله الإحصاء في هذه الحالة إن العاملين يتحركا معاً بطريقة معينة.

التبسيط الزائد Oversimplification

تحدث الأخطاء المعتادة نتيجة التبسيط الزائد للموضوع وإهمال الشروط الضرورية. وقد تكون الحقيقة المعروضة صحيحة ولكن ربما تضلل الجمهور إذا تم إهمال بعض الحقائق ذات الصلة بها. فمثلاً ، الإعلان للجرار الحالي يوضح بأن مبيعاته زادت بنسبة ١٥٥٪ في ولاية تكساس. ولكن من متى؟ ولذلك فإن المقارنة ليس لها معنى لعدم إعطاء معلومات عن سنة الأساس. وعموماً فإن أصحاب الدعايات والسياسيين وآخرين يستخدمون هذا التكتيك أو هذه الطريقة باستمرار.

الدقة الزائفة Spurious Accuracy

نحن نعيش الآن في عصر الكمبيوتر حيث يمكن إنشاء إجابات بعدد كبير من الفواصل العشرية. ولكن لسوء الحظ فإن الإجابة لن تكون أكثر دقة من البيانات. فلو تم تقريب البيانات لأقرب ١٠٠ دولار، كما في حالة بيانات الدخل، ففي هذه الحالة لن يتم كتابة متوسط الدخل للدراسة على أنه ٤٤٢٠٣,٥٠ دولاراً. وقد يبدو ذلك مثير للإعجاب ولكن تلك دقة زائفة. وحيث إن البيانات في هذه الحالة تم تقريبها لآخر ١٠٠ دولاراً فإنه يجب كتباتها على النحو ٤٤٢٠٠ دولار. وقد يتم تشويه المعنى الحقيقي للحقائق ببساطة. لذا يجب على الباحث الإحصائي أن يحاول تجنب العرض المشوّه للحقيقة ويكتشف الخطأ في استخدام الإحصاء من قبل الآخرين وبذلك فإن موقف النقد يعتبر أساسياً.

نظرة للأمام A Look Ahead

الموضوعات التي ستتم تغطيتها في بقية الكتاب مقسمة لثلاثة أجزاء هي :

١- استخدام الإحصاء في تلخيص البيانات ممّا يمكّن من فهمها ببساطة.

٢- الاحتمالات، والتوزيعات الاحتمالية، والمعاينة، وتوزيعات المعاينة والتي تتضمن الخلفية المطلوبة للموضوعات في الجزء الأخير.

٣- الاستدلال الإحصائي.

نحن نصادف الإحصاء الوصفي يومياً في تعاملاتنا الروتينية. فمثلاً شخص معين يعطينا دائماً المتوسط لشيء معين وآخر قد يخبرنا بمتوسط إنتاجية المحاصيل أو متوسط المحتوى الرطوبي لآخر حمولة للحبوب أخذت من المخزن...إلخ. ولذلك نستطيع ببساطة فهم ماذا تعني عندما يقوم شخص ما بعرضها لنا وكذلك عندما نرغب في استخدامها كرقم أو رقمين لوصف البيانات. الاحتمالات والتوزيعات الاحتمالية...إلخ تستخدم كمعلومات أساسية لفهم الاستدلال الإحصائي؛ لأن معظم التحليل الإحصائي تقريباً مبني على الاحتمال. وسوف يساعد قاموس المصطلحات في آخر الكتاب في فهم المصطلحات غير المألوفة خلال دراسة الموضوعات. ويتركز الاهتمام في هذا الجزء على المفاهيم الأساسية والتي سيتم استخدامها في الجزء الأخير في التقدير واختبارات الفروض.

الموضوعات في الاستدلال الإحصائي توضح طرق تحليل البيانات ولذلك نستطيع تحديد ما إذا كانت إحدى المعالجات تختلف فعلاً عن الأخرى أم لا. ويمكن تحديد ذلك باستخدام تعبير احتمالي حول ذلك الاختلاف إذا كان الاختبار الإحصائي يشير لوجود اختلاف بينهما. بينما في بعض الحالات يمكن النظر للبيانات ونجد أنها نفسها و لا يوجد اختلاف، وفي حالات أخرى فإن ذلك ليس مهماً. وقد تم تطوير الطرق الإحصائية خلال نصف القرن الماضي والتي تتخذ من التخمين محاولة لوجود الاختلاف من عدمه. وسيتم قضاء بعض الوقت لتعلم هذه الطرق. وهناك بعض الطرق المعتادة كل منها مخصص لإجراء اختبار معين وسوف يتم التطرق لها جميعاً.

المراجع References

Huff, D. *How to lie with statistics*. New York: W.W. Norton, 1954.
Ltd., 1961. Reichmann, W.J. Use and absue of Statistics. London: Methuen

الملحق: المتغيرات ومتغيرات التجميع

Appendix: variables and Summations variables

المتغيرات variables

يعرف المتغير كأي شيء والذي يمكن أن يكون له مدى من القيم. في الإحصاء يتم التعبير عن المتغيرات بالرموز المعتادة للحروف الهجائية مثل X و Y . وهذا من الأشياء الروتينية المتعارف عليها. ويمكن ببساطة إعطائها أسماء أو استخدام الحروف الإغريقية ولكن ذلك يتطلب استخدام نوع معين من الدلالة. ونظراً لأن المتغيرات تأخذ مدى من القيم فإنه يمكن تعريفها بناءً على أنواع القيم المفترضة إلى نوعين.

المتغيرات الوصفية: هي المتغيرات التي لا يمكن قياسها حيث تعبر عن بعض الصفات التي لا يمكن قياسها عددياً مثل الحالة الزوجية، الأجزاء الجيدة أو المعيبة، أنواع الفواكه، فئات العمر ...إلخ. ويتم التعبير عن هذا النوع من المتغيرات باستخدام قيم افتراضية لها مثلاً تكون X=1 إذا كانت الكرة سوداء وصفر عدا ذلك، أو Y=1 للأنثى و ٢ للذكر ...إلخ.

المتغيرات الكمية: تعبر عن الأشياء التي يمكن قياسها أو عدّها. فإذا كان المتغير يمكن عده فقط فإنه يكون منفصل أو متقطع.

المتغيرات المنفصلة: تكون قيمها أعداد صحيحة. وهناك العديد من مثل هذه المتغيرات في الإحصاء مثل أعداد العجول الذكور في حظيرة سعتها ثمانية، عدد مرات الوصول لمنطقة العرض خلال فترة زمنية معينة، أو عدد مرات النجاح في عدد n محاولة لتجربة ما.

المتغيرات المتصلة: هي التي يمكن قياسها ولذلك يمكن أن تأخذ أي قيمة في المدى للمتغير بناء على دقة أدوات القياس المستخدمة. فمثلاً إذا كانت X تعبر عن إنتاج السماد بالطن في اليوم من أحد مصانع الأسمدة فإن قيمة X يمكن أن تتراوح بين الصفر والسعة القصوى للمصنع والتي نفترض أنها تساوي ١٠٠ طن. وإنتاج الأسمدة متغير مستمر ولذلك يمكن استخدام وحدات القياس التي نرغبها: ١٠٠طن، أطنان، عشرات الأطنان، مئات الأطنان، آلاف الأطنان، ...إلخ، وخاصة إذا كنا نستخدم الميزان الإلكتروني.

التجميع Summations

تستخدم الصيغ في الإحصاء كطرق عامة للتعبير عن كيفية حساب المتغيرات محل الاهتمام. فمثلاً ، الصيغة المستخدمة لحساب المتوسط الحسابي لمجموعة من البيانات توضح كيفية الحساب بغض النظر عن حجم البيانات التي تم اختيارها. ولذلك فإنه لا يجب تغيير تلك الصياغة عند تغير حجم أو نوع البيانات موضع الدراسة. وقد تحتوى الصيغ على علامة التجميع والتي تساعد في تحديد كيفية إجراء الحسابات.لذا سيتم عرض جبر المجاميع في الجزء التالي.

مؤشر التجميع Index of Summation

رمز التجميع Σ يعني إضافة القيم للمتغير بدءاً بأول قيمة في المؤشر مروراً بكل القيم ثم الانتهاء بالقيمة الأخيرة في المؤشر.وتكتب أول قيمة في المؤشر أسفل الرمز Σ والقيمة الأخيرة تكتب أعلى الرمز Σ . فمثلاً يتم تقدير المصطلح في المعادلة رقم (1.1) بإضافة القيم العشر لـ X_i .

$$\sum_{1}^{10} X_i \qquad (1.1)$$

ابتداء بـ X_1 ثم الاستمرار بالقيم X_2 ثم X_3 وهكذا... حتى ننتهي بالقيمة X_{10}. وحيث إن لدينا عشر قيم في مؤشر التجميع قمنا بتجميع القيم العشر الأولى لـ X . وهذا يعتمد على مجموعة البيانات التي تحتوي على المتغير X بجميع قيمه ولكن الصياغة تختص بتجميع القيم العشر الأولى. ولجمع كل قيم المتغير X_i يتم إعادة كتابة الصياغة بحيث تحتوي على الرمز n (عموماً) بدلاً من الرقم الأخير في المؤشر، ١٠. أيضا عند عدم احتواء إشارة التجميع على أرقام فإن ذلك يعني أنها من رقم ١ إلى n ويتم جمع جميع القيم للمتغير.

تجميع الثابت Summation of a Constant

إذا كان الحرف k يعبر عن رقم ثابت، فإن مجموع k يقدر بحاصل ضرب عدد مرات التجميع في k. لذا يتم تقدير التعبير الوارد في المعادلة رقم (1.2) بضرب k بالعدد ٥، عدد الحدود في التجميع، وبالتالي نحصل على الجواب 5k.

$$\sum_{1}^{5} k \qquad (1.2)$$

تجميع حاصل ضرب المتغير بثابت Summation of a Constant Times a Variable

في حالة جمع تعبير رياضي مثل $2X_i$ حيث 2 عبارة عن ثابت و X_i تعبر عن المتغير والتي يمكن كتابتها على النحو التالي $\sum 2X_i$ ، ولذلك يمكن تحويل الرقم الثابت خارج عملية التجميع وإعادة كتابتها لتصبح $2\sum X_i$. والتعبير الأخير يتطلب تجميع قيم المتغير ثم ضرب المجموع بالثابت 2.

جبر التجاميع An Algebra of Summation

تتوزع علامة التجميع على التعبير الرياضي بين الأقواس. وبذلك يمكن إعادة

كتابة التعبير الرياضي $\sum(3X_i+7)$ بحيث يكون $\sum 7+\sum 3X_i$ ، والذي يمكن اختصاره ليكون $3\sum X_i$ +7n ، حيث n تمثل عدد الحدود في مؤشر التجميع. ويمكن كتابة التجميع لمربعات المتغيرات على النحو $\sum X_i^2$ وتسميتها مجموع المربعات. ويتم تقديرها بتربيع كل قيمة للمتغير X ثم جمعها، كما في المعادلة رقم (1.3).

$$\sum X_i^2 = X_1^2 + X_2^2 + \ldots\ldots\ldots + X_n^2. \qquad (1.3)$$

من جهة أخرى فإنه يمكن كتابة مربع المجموع على النحو $\left(\sum X_i\right)^2$ وحسابه عن طريق إيجاد المجموع ثم تربيعه. ومجموع المربعات ومربع المجموع مفهومين مختلفين تماماً ويتم استخدام كلا التعبيرين باستمرار في الصيغ الرياضية.

تمارين Exercises

١- اكتب بالتفصيل المجاميع الممثلة بالتعابير الرياضية التالية:

أ) $\sum X_i$ ب) $\sum X_i^2$

ج) $\sum(Y_i+4)$ د) $\sum cX_i$

هـ) $\sum f_i X_i^2$ و) $\left(\sum X_i\right)^2$

٢- اكتب التعابير الرياضية التالية باستخدام رمز التجميع والنهايات المناسبة:

أ) $X_1+X_2+X_3+X_4+X_5+X_6$

ب) $X_1^2+X_2^2+X_3^2+X_4^2+X_5^2$

ج) $[(X_3+3)+(X_4+3)+(X_5+3)]^2$

د) $(X_1-a)^2+(X_2-a)^2+(X_3-a)^2$

٣- إذا كانت قيمة $X_1 = 2$، و $X_2 = 7$، $X_3 = -8$، $X_4 = -1$، $X_5 = 6$، $X_6 = 3$ فأوجد القيم الحسابية للفقرات أ ، ب، ج في السؤال الثاني.