جمهورية العراق

وزارة التعليم العالي والبحث العلمي

جامعة الأنبار

كلية الآداب

قسم الجغرافية

المرحلة : الرابع

أستاذ المادة : أ.د. احمد سلمان حمادي

اسم المادة باللغة العربية : الجيوماتكس

اسم المادة باللغة الانكليزية : Geomatics

اسم المحاضرة الحادي عشر باللغة العربية الإحصاء وتحليل البيانات

اسم المحاضرة الحادي عشر باللغة الإنكليزية : Statistics and data analysis

المحاضرة الحادي عشر

الإحصاء وتحليل البيانات

Statistics and data analysis

تمثل البيانات المكانية وأيضا البيانات غير المكانية جوهر نظم المعلومات المكانية، ومن ثم فأن التأكد من جودة هذه البيانات ودقتها ومحاولة اكتشاف أية مصادر للأخطاء بها يعد من أهم أسس تطوير نظام يحقق الأهداف المنشودة منه. وحيث أن القياسات (سواء قياسات المساحة الأرضية أو المواقع المقاسة بتقنيات النظم العالمية لتحديد المواقع وأيضا القياسات من الصور الجوية والمرئيات الفضائية) لا تخلو من الأخطاء مهما صُغرت قيمتها، فأن ضبط وتصحيح هذه القياسات ي صبح مطلبا أساسيا قبل إدخالها في قواعد بيانات نظم المعلومات المكانية. وعلى ص عيد آخر فان تحليل البيانات يمثل خطوة أساس ية في مراحل بناء النظام بغرض اشتقاق معلومات تصف الواقع الحقيقي الدقيق للظاهرات الأرضية الممثلة داخل النظام وبناءا عليه فان ضبط القياسات وتحليل البيانات المكانية وغير المكانية يعد من أسس الجيوماتكس.

٨-٣ التحليل الإحصائي للبيانات

القيمة الأكثر احتمالا : من الصعب – إن لم يكن من المستحيل – معرفة القيمة الحقيقية لأي كمية مقاسه وذلك لوجود أخطاء في القياسمهما كانت قيمة هذه الأخطاء صغيرة جدا. إن كانت الأرصاد مستقلة ولا تعتمد على بعضها البعض وقمنا بتكرار القياس عدة مرات فان قيمة المتوسط الحسابي ستمثل القيمة الأكثر احتمالا أو الأكثر توقعا أو الأكثر قربا للقيمة الحقيقية

الخطأ الحقيقي : هو الفرق بين القيمة المرصودة والقيمة الحقيقية لها. وبما أن القيمة الحقيقية لا يمكن معرفتها ففي معظم الأحيان فان الخطأ الحقيقية أيضا لا يمكن معرفته. لكن في بعض الحالات يمكن معرفة الخطأ الحقيقي من خلال مواصفات أو قواعد هندسية معلومة فمثلا عند قياس الزوايا الثلاثة لمثلث فيجب أن يساوي مجموع الزوايا ١٨٠ درجة، ففي هذه الحالة يكون الخطأ الحقيقي هو ناتج طرح مجموع الزوايا المقاسة من ١٨٠.

الأخطاء المتبقية أو الفروق : الفرق أو الخطأ المتبقي (أو الباقي residual) هو الفرق بين القيمة المرصودة والقيمة الحقيقية لها. لكننا نستعيض عن القيمة الحقيقية بالقيمة الأكثر احتمالا لها وبذلك يكون الخطأ المتبقي:

الفرق = القيمة الأكثر احتمالا – القيمة المرصودة

التباين : التباين هو مؤشر إحصائي يحدد مدي تباين أو انتشار أو تشتت مجموعة من الأرصاد حول القيمة الحقيقية لها أو القيمة الأكثر احتمالا لها، ولذلك يوجد نوعين من التباين:

- **تباين المجتمع** إذا تم قياس كل الأرصاد الممكنة للقيمة المطلوبة فأن تباين المجتمع يساوي مجموع مربعات الأخطاء الحقيقية مقسوما على عدد الأرصاد:
- **تباين العينة :** إذا تم قياس عينة أو مجموعة من الأرصاد للقيمة المطلوبة فأن تباين هذه العينة يساوي مجموع مربعات الأخطاء المتبقية (وليست الأخطاء الحقيقية) مقسوما على عدد الأرصاد ناقص واحد

الخطأ المعياري الخطأ المعياري هو الجذر التربيعي لقيمة تباين

الانحراف المعياري يعبر الانحراف المعياري (يطلق عليه أيضا أسم الخطأ التربيعي المتوسط عن مدي انحراف (ابتعاد أو اقتراب) القيمة المقاسة عن القيمة الأكثر احتمالا لها، وقيمته تساوي الجذر التربيعي لقيمة تباين العينة:

وللانحراف المعياري عدة مميزات تجعله أشهر وأفضل مقاييس التشتت والانتشار. فأولي مميزاته أن وحداته هي نفس وحدات مجموعة الأرقام الأصلية مما يجعله سهل الفهم والتحليل. أما ثاني مميزات الانحراف المعياري وبناءا على خصائص شكل التوزيع الطبيعي أو ما نسميه المنحني الطبيعي فهي أنه يمكن استخلاص معلومات مهمة إذا عرفنا قيمة المتوسط والانحراف المعياري لمجموعة من البيانات

الانحراف المعياري للمتوسط الانحراف المعياري للمتوسط الحسابي هو حاصل قسمة الانحراف المعياري للعينة على الجذر التربيعي لعدد الأرصاد: وتعبر قيمة الانحراف المعياري عن مدي تشتت أو تباعد القياسات عن بعضها البعض وبالتإلى فهي قيمة معبرة عن مدي التوافق بين الأرصاد ومن ثم فأن الانحراف المعياري يؤخذ على أنه مقياس أو مؤشر للصحة. وفي العمل المساحي لا نعبر عن القيمة الأكثر احتمالا بقيمة المتوسط فقط إنما بقيمتي المتوسط والانحراف المعياري معا، فنقول أن المسافة المقاسة – على سبيل المثال – تساوي ٢١.٥٣ ± ٠٣.٠ متر.

معامل الاختلاف يستخدم معامل الاختلاف عند مقارنة مدي التشتت في بيانات مجموعتين مختلفتين من البيانات غير المكانية، خاصة في حالة اختلاف نوعي البيانات ذاتها واختلاف وحداتهما. تتم هذه المقارنة عن طريق تخليص الانحراف المعياري من أثر الاختلاف وذلك بنسبته (أو قسمته) إلى المتوسط:
معامل الاختلاف = الانحراف المعياري / المتوسط

الالتواء الالتواء هو بعد التوزيع التكراري لظاهرة عن التماثل أو التوزيع المتماثل. فإذا كان التوزيع متماثلا فأن ٥٠% من القيم ستقع على كل جانب من المنوال (المنوال هو القيمة التي تتكرر أكثر من غيرها من القيم، أو هو القيمة الأكثر تكرارا بين مجموعة الأرقام)، ويكون المنوال = الوسيط = المتوسط الحسابي: **التفلطح** التفلطح هو مدي اختلاف التوزيع التكراري لظاهرة عن التوزيع الطبيعي أو التوزيع العادي. قد يكون التوزيع مدببا إذا كان أكثر تحديا عند قمته أو قيمته المركزية وكانت تلك القيمة اعلى

منها للتوزيع الطبيعي، وقد يكون التوزيع مفرطحا إذا كانت قمته أكثر استقامة وادني من تلك للتوزيع الطبيعي.

الارتباط والانحدار من أهم خطوات تحليل وتعليل الظاهرات المختلفة دراسة أوجه التشابه والاختلاف بينها. وقد تتم هذه الخطوة من خلال المقارنات الوصفية البسيطة أو من خلال المقارنات الاستنتاجية التفسيرية. أما المقارنات الوصفية البحتة لتحليل الظاهرات فتعتمد على تطبيق بعض النظريات الحسابية لمقارنة المؤشرات الإحصائية (مثل المتوسط والانحراف المعياري) بين مجموعتين أو أكثر من البيانات غير المكانية. لكن هناك طرق تعطي صورة أكثر دقة عن مدي التشابه والاختلاف وأيضا العلاقات بين الظاهرات، ومن هذه الطرق الارتباط والانحدار.

الارتباط يقيس الارتباط مدي الترابط بين مجموعتين من البيانات غير المكانية توجد عدة صور من الارتباط بين ظاهرتين وكذلك تختلف قيمة الارتباط من حيث قوتها وضعفها كما نري في الشكل التالي:

وتتراوح قيمة معامل الارتباط بين –١ و +١، فان كانت قيمته موجبه فتدل على وجود ارتباط موجب أو طردي بين مجموعتي البيانات (بمعني أن زيادة قيمة المتغير الأول تؤدي بصفة عامة إلى زيادة قيمة المتغير الثاني أيضا) وان كانت قيمته سالبه فتدل على وجود ارتباط سالب أو عكسي بين المجموعتين أو الظاهرتين (بمعني أن زيادة قيمة المتغير الأول تؤدي عامة إلى تناقص قيمة المتغير الثاني). أما إن كانت قيمة معامل الارتباط تساوي الصفر فهذا يدل على عدم وجود أي ارتباط بين هاتين الظاهرتين.

أما قيم معامل الارتباط ذاتها فيمكن تقسيمها إلى عدة مستويات كالتالي:

من ± ٧.٠ إلى ± ٠.١	درجة ارتباط عالية و قوية
من ± ٤.٠ إلى ±٧.٠	درجة ارتباط جوهرية أو حقيقية
من ± ٢.٠ إلى ±٤.٠	درجة ارتباط منخفضة أو ضعيفة
أقل من ± ٢.٠	درجة ارتباط ضعيفة للغاية أو منعدمة

الانحدار إن الارتباط يعطي درجة العلاقة بين ظاهرتين، فإذا ما وجدت هذه العلاقة فأننا نحاول أن نحددها بصورة رياضية تسمح لنا بتقدير قيمة أحد المتغيرين إذا عرفنا قيمة المتغير الآخر وهذا ما يعرف بالانحدار. الانحدار Regression هو دراسة للتوزي عالم شترك لمتغيرين احدهما يقاس ويسمي المتغير المستقل والآخر متغير غير مستقل أو متغير تابع بهدف تحديد شكل العلاقة بينهما بصورة رياضية. أهم صور الانحدار هي ما يعرف باسم الانحدار الخطي أي محاولة توفيق خط مستقيم يمر بأكبر عدد ممكن من النقاط.

إلا أنه تجدر الإشارة لإمكانية وجود انحدار غير خطي بين متغيرين ومن أهم تطبيقات خط الانحدار ما يسمي بالتنبؤ المستقبلي، أي التنبؤ بقيمة المتغير التابع عند قيمة محددة – مستقبلية – للمتغير المستقل.