الباب الأول
المجتمع الإحصائي والعينة

1-1 تمهيد

درسنا في الجزء الأول من هذا الكتاب مبادئ الإحصاء الوصفي، فعرفنا كيف تحسب قيمة كل من مقاييس النزعة المركزية (المتوسط، الوسيط، المنوال). ومقاييس التشتت (المدى، المدى الربعي، الانحراف المتوسط، الانحراف المعياري). وكذلك تعلمنا كيف نجد العلاقة بين ظاهرتين معينتين بالاعتماد على البيانات المتجمعة عنهما. وقد تعلمنا أيضاً خصائص كل مقياس من هذه المقاييس. وكذلك أوجه استعمالاته كأداة للوصف بالنسبة إلى الظاهرة مدار البحث.

ولكن استعمال هذه المقاييس لا يقتصر على الوصف فقط، بل إنه يتعدى ذلك إلى مجالات أكثر تنوعاً وتعقيداً. ففي العديد من الدراسات التي نقوم بها، والتي يكون طابعها إحصائياً، نحاول دراسة ظاهرة ما، إما للتحقق من درجة وجودها في المجتمع بأكمله. أو لنجد مدى تأثير مجموعة من العوامل أو المتغيرات المختلفة عليها. فمثلاً قد نحاول دراسة مستوى الذكاء بين طلبة إحدى الجامعات لنصل من ذلك إلى قيمة محددة له. أو لنتأكد من أن قيمته تساوي أو تختلف عن قيمة معينة نفترضها أو نتوقعها له. أو أننا قد نحاول دراسة مستوى التحصيل الأكاديمي عند الأفراد وصلة ذلك بالطبقة الاجتماعية التي ينتمي إليها الفرد الواحد منهم، أو العلاقة بين المستويات الثقافية للأفراد ودرجة مساندتهم لمشروع معين، أو طبيعة اتجاهاتهم نحوه. كما أننا قد ندرس المشكلات التي يصادفها المعلم داخل الصف وعلاقة ذلك بمستوى المرحلة الدراسية التي يعمل بها، أو بعدد السنوات التي قضاها في مهنة التدريس، أو غير ذلك من الدراسات المشابهة. وفي كل دراسة من هذه الدراسات فإننا نسعى إلى اتخاذ قرار أو حكم. أو نعمد إلى تبني موقف بالنسبة إلى الظاهرة أو الظواهر المدروسة بناء على طبيعة المعلومات أو البيانات التي تسير لنا جمعها عنها.

ويساعدنا الاحصاء التحليلي في التوصل إلى قرارات أو الأحكام الخاصة بالظواهر المدروسة وذلك بطرائق مختلفة من بينها دراسة الفروق بين متوسطات هذه الظواهر. أو بين انحرافاتها المعيارية أو معاملات ارتباطاتها، ... إلخ، لتقرير ما إذا كانت احجام هذه الفروق، إن وجدت، كبيرة وملفته للنظر، أم أنها من مستوى منخفض بحيث يمكن إهمالها والتغاضي عنها.

وحين يعالج الاحصاء التحليلي بين المتوسطات أو الانحرافات المعيارية أو غيرها لتقرير ما إذا كانت هذه الفروق كبيرة أم صغيرة، فإنه يلجأ إلى مقارنتها بجداول احصائية أعدت لهذا

الغرض. وسنتعرض لمثل هذه الجداول من حيث طبيعتها وطرائق استخدامها في الأبواب والفصول القادمة.

ومن الجدير بالذكر أن القرارات والأحكام التي نتخذها بصدد الظواهر التي ندرسها لا يمكن أن تؤخذ بشكل قاطع، لأننا في الغالب ندرس هذه الظواهر ليس على كل فرد من الأفراد الذين تتمثل فيهم تلك الظواهر، وإنما على عينات نختارها من بينهم، وأن مثل هذه العينات رغم أننا نختارها بالطرائق العلمية المعروفة. فإنه لا يوجد ما يضمن أن تجيء ممثلة للمجتمعات المأخوذة منها تمثيلاً صادقاً. وبعبارة أخرى. فإن الدراسات المبنية على عينات لا يحتمل أن تعطي نتائج صادقة تماماً وموثوق بها. ولهذا كله، فنحن نحاول أن تأتي قراراتنا وأحكامنا ضمن مستويات معينة من الثقة وتعتمد في قيمتها على أهمية القرارات التي ستتخذ، وعلى طبيعة الظاهرة المدروسة.

ولما كانت دراستنا للظواهر المختلفة تقوم كما ذكرنا على دراسة هذه الظواهر في عينات تختار من بين الأفراد الذين تتمثل فيهم مثل هذه الظواهر، وإن ما نتوصل إليه من دراسة العينة نعمل على تعميمه على المجتمع الذي أخذت منه العينة بكاملة، فإنه حري بنا أن نتعرف إلى الصلات التي تربط الاحصائية الخاصة بالعينة (كالمتوسط، والانحراف المعياري، ... إلخ) بالاحصائيات الخاصة بالمجتمع، وهو ما سيكون مدار بحثنا في البنود القادمة من هذا الباب.

2-1 التوزيعات التكرارية للعينات :Sampling Distributions

إذا كان لدينا مجتمع من المشاهدات أو البيانات المأخوذه على ظاهرة ما، وأخذنا منه عينات مختلفة من حجم معين (حجم العينة الواحدة ٥ حالات مثلاً). وحسبنا قيمة المتوسط الحسابي لكل عينة منه، وكذلك تباينها (مربع الانحراف المعياري لها)، لحصلنا من ذلك على مجموعتين من القيم التي يمكن وضع كل منها في جدول تكراري خاص يسمى الأول منهما بالتوزيع التكراري لمتوسطات العينات .(Sampling distribution of the means) بينما يسمى الآخر بالتوزيع التكراري لتباينات هذه العينات .(Sampling distribution of the variances)

ولتوضيح ما سبق نفرض أن المجتمع الاحصائي الذي نحن بصدده يتألف من أعمار ستة أشخاص أعطيت بالسنوات على النحو الآتي:

11، 16، 12، 15، 16، 14

وأننا أخذنا منه كل العينات الممكنة والتي عدد أفراد الواحدة منها شخصين اثنين، مثلاً (وعددها 36 عينة في هذه الحالة – جدول رقم1). والآن، بعد أن تم لنا تحديد الحالات الداخلة

في كل عينة منها على النحو المبين في الجدول رقم (1) نحسب قيمة المتوسط الحسابي وقيمة التباين لكل واحدة منها بالطرق الحسابية المعتادة. وتظهر قيم المتوسطات الحسابية في الأطراف العليا للخانات الدالة على هذه الفئات والتي ينقسم اليها الجدول. بينما تظهر قيم التباينات في الاطراف الدنيا لها. فالعينة التي مفرداتها هي (11، 16) مثلاً يكون متوسطها 13,5 وتباينها 12,5، فالعينة التي مفرداتها هي (15، 12) يكون متوسطها 13,5 وتباينها 4,5، وهكذا، ومن الجدير بالذكر أن التباين في حالة كل عينة قد حسب من القانون:

$$\text{ع}^2 = \frac{\text{مجـ }(\text{س}-\text{م})^2}{\text{ن}-1}$$

وليس من القانون

$$\text{ع}^2 = \frac{\text{مجـ }(\text{س}-\text{م})^2}{\text{ن}-1}$$

الذي اعتدنا عليه في السابق، والخاص بمجتمع كامل من البيانات وليس بعينة منها.

جدول (1)

	11	16	12	15	16	14
11	11 -	13,5 12,5	11,5 0,5	13 8	13,5 12,5	12,5 4,5
16	13,5 12,5	16 -	14 8	15,5 0,5	16 -	15 2
12	11,5 0,5	14 8	12 -	13,5 4,5	14 8	13 2
15	13 8	15,5 0,5	13,5 4,5	15 -	15,5 0,5	14,5 0,5
16	13,5 12,5	16 -	14 8	15,5 0,5	16 -	15 2
14	12,5 4,5	15 2	13 2	14,5 0,5	15 2	14 -

وبعد ترتيب قيم هذه المتوسطات في جدول تكراري، كما هو مبين في جدول (2). نحسب المتوسط العام لها بالطرائق العادية.

جدول رقم (2)

م	16,0	15,5	15	14,5	14	13,5	13	12,5	12	11,5	11
ك	4	4	5	2	5	6	4	2	1	2	1
م × ك	64	62	75	29	70	81	52	25	12	23	11

مجـ ك = 36

مجـ ك م= 504

إذن المتوسط العام لمتوسطات العينات يساوي:

مـ = $\frac{504}{36}$ = 14

والآن، ترتيب تبـاينات العـينات في جـدول تكراري، كـمـا هو مـبين أدناه، ونحـسب لهـا متوسطها بالطرائق المعتادة:

وبذلك يكون متوسط تباينات العينات:

$ع^2$	ك	ك × $ع^2$
12,5	4	50
8,0	6	48
4,5	4	18
2,0	6	4
0,5	8	4
∴	8	∴
	36	132

وبذلك يكون متوسط تباينات العينات

$ع^2$ = $\frac{132}{36}$ = $\frac{11}{3}$

وأخيراً نحسب قيمة متوسط المجتمع المؤلف من القيم الست (11، 16، 12، 15، 16، 14)، وكذلك قيمة تباينه من أجل مقارنتها مع القيم المناظرة لهما والمحسوبة من التوزيعين التكرارين

السابقين، أي التوزيع الخاص بمتوسطات العينات، وكذلك التوزيع الخاص- بتبايناتها.

متوسط المجتمع (هـ)

$$= \frac{11 + 16 + 12 + 15 + 16 + 14}{6} = 14$$

تباين المجتمع (ع^2)

$$= \frac{\{(11-14)^2 + (16-14)^2 + (12-14)^2 + (15-14)^2 + (16-14)^2 + (14-14)^2\}}{6}$$

$$= \frac{22}{6} = \frac{11}{3}$$

يتضح من مقارنة القيم المحسوبة في حالة المجتمع وحالة التوزيعات التكرارية الخاصة بمتوسطات العينات وتبايناتها ما يأتي:

(1) المتوسط الحسابي لمتوسطات العينات يساوي متوسط المجتمع.

(2) المتوسط الحسابي لتباينات العينات يساوي تباين المجتمع.

ومع أن هاتين النتيجتين تأكدت صحتهما بالنسبة إلى الحسابات التي أجريت على المثال السابق (أعمار ستة أشخاص)، إلا أنها في الواقع صحيحة ويمكن تعميمها على كل الحالات المشابهة.

ومن الجدير بالذكر أن النتيجة الثانية لا تصح إلا إذا حسب تباين المجتمع من القانون:

$$\frac{\text{مجـ (س−م)}^2}{\text{ن}}$$

وتباين العينة من القانون

$$\frac{\text{مجـ (س−م)}^2}{\text{ن}}$$

لذلك علينا أن نفرق عند حساب التباين لمجموعة من البيانات فيما إذا كانت تخص عينة أو مجتمعاً بأكمله.

3-1 الخطأ المعياري :Standard Error

لو كانت كل العينات المأخوذة من مجتمع ما ممثلة له تمثيلاً صادقاً لكانت متوسطاتها جميعاً متساوية وتساوي في قيمتها لمتوسط ذلك المجتمع. ولكن مثل هذه الحالة المثالية يندر تحقيقها في واقع الأمور العملية. فقد رأينا في المثال السابق (أعمار ستة أشخاص) أنه بينما كانت متوسطات بعض العينات تساوي في قيمتها لمتوسط المجتمع، كانت متوسطات البعض الآخر أما أكبر منه أو أقل. ولقياس مدى تذبذب متوسطات العينات حول متوسط المجتمع المأخوذة منه، نستعمل لذلك ما يسمى بالخطأ المعياري. فالخطأ المعياري هو الانحراف المعياري لمتوسطات العينات بالنسبة إلى متوسطها العام، الذي هو متوسط المجتمع المأخوذة منه هذه العينات. ويرمز للخطأ المعياري بالرمز ع م تمييزاً له عن الانحرافات المعيارية الأخرى.

ففي حالة المثال السابق (أعمار ستة أشخاص) يكون الخطأ المعياري كالآتي:

$$ع_{م}^{2} = \{4\,(14-16)^2 + (14-5,5)^2 + 5\,(14-15)^2$$
$$+2\,(14-14,5)^2 + 5\,(14-14)^2 + 6\,(14-13,5)^2$$
$$+4\,(14-13)^2 + 2\,(14-12,5)^2 + 1\,(14-12)^2$$
$$+2\,(14-11,5)^2 + 1\,(14-11)^2\} / 36$$
$$= \frac{66}{36} = \frac{11}{6}$$
$$ع_{م} = \sqrt{\frac{11}{6}} = 1,36$$

ومن الجدير بالذكر أن تشير إلى وجود علاقة احصائية هامة بين الخطأ المعياري لمتوسطات العينات والانحراف المعياري للمجتمع، وهي التي يعبر عنها عادة بالصورة الآتية:

$$ع_{م} = \frac{عـ}{\sqrt{ن}}$$

حيث إن: ع م = الخطأ المعياري

، عـ = الانحراف المعياري للمجتمع

، ن = عدد الأفراد العينة.

ويمكن البرهنة على صحة هذه العلاقة بطريقة جبرية لا مجال لذكرها هنا، وسنكتفي بدلاً

من ذلك بالتأكيد من صحتها عن طريق الأمثلة العددية. ففي المثال السابق (عمر ستة أشخاص) وجدنا أن:

$$ع_{م} = \sqrt{\frac{11}{6}} \quad ، \quad عـ = \sqrt{\frac{11}{6}} \quad ، \quad ن = 2$$

وهذه القيم تحقق العلاقة المذكورة، شأنها في ذلك شأن أية مجموعة من القيم الأخرى المشابهة.

وبما أن قيمة الانحراف المعياري للمجتمع (عـ) تكون ثابتة. فإن العلاقة:

$$ع_{م} = \sqrt{\frac{عـ}{ن}}$$

ترينا أنه كلما زاد عدد أفراد العينة كلما قلت قيمة الخطأ المعياري لمتوسطات العينات، والعكس. ويعني هذا الأمر أن العينات تكون أصدق تمثيلاً للمجتمع المأخوذة منه كلما كبر حجمها. وهذا بلا شك أمر منطقي، إذ أننا نتوقع أن تكون العينات الصغيرة أقل قدرة من العينات الكبيرة على تمثيل المجتمع المأخوذة منه، هذا بالطبع إذا أحسن اختيارها، وكان وفق الأسس العلمية لذلك.

ولا يخفى أنه لاعتبارات عملية لا يمكن أن نختار حجم العينة بحيث يكون كبيراً جداً لأنه في مثل هذه الحالة قد يصعب الاحاطة بأفرادها من ناحية، وقد يجعل هذا كمية البيانات المجتمعة، والتي تتطلب تحليلاً وتفسيراً كبيرة من ناحية ثانية، بحيث يتطلب انجازها جهداً ووقتاً طويلاً. ويقترح بعض الإحصائيين أن يكون حجم العينة في حدود 5% من حجم المجتمع الأصلي، ولكن هذه النسبة قد تعطي عينة كبيرة أحياناً وصغيرة أحياناً أخرى، ولذلك اعتماداً على حجم المجتمع المأخوذة منه. ولذا يفضل الكثيرون عدم إعطاء قوانين لتحديد حجم العينة وترك الأمر لحكمة الباحث ليختار عينته بحيث يتمكن من معالجة البيانات الخاصة بها ضمن الزمن المحدد له، وفي ضوء الطاقات الفنية المتوفرة لديه. ومن الجدير بالذكر أن عدداً كبيراً من طلبة الماجستير أو الدكتوراه يختارون أحياناً عينات من حم كبير، نظراً لقلة خبرتهم، فيجدون أنفسهم في مأزق لعدم تمكنهم من تحليل البيانات التي تتجمع لديهم عن الظواهر التي يدرسونها في أزمان تتناسب مع أوضاعهم الدراسية.

4-1 الصعوبات التي تواجهنا في حساب الخطأ المعياري

قلنا أن الخطأ المعياري هو مقياس لتباعدات متوسطات العينات عن متوسط المجتمع. فكلما صغر حجم الخطأ المعياري كلما اقترب متوسط العينة من متوسط المجتمع، والعكس بالعكس. وبعبارة أخرى، فإن حجم الخطأ المعياري يمكن أن يتخذ دليلاً على درجة تمثيل العينة لمجتمعها.

ولما كان الخطأ المعياري يمكن أن تحسب قيمته بالاعتماد على قيمة الانحراف المعياري للمجتمع (ع/$\sqrt{ن}$)، أو بمعالجة انحرافات المتوسطات عن المتوسط العام لها، فإن هناك عدة صعوبات يمكن أن تقف حائلاً دون إمكانية حساب قيمته باستخدام أي من هاتين الطريقتين. فالانحراف المعياري للمجتمع قلما تكون قيمته معروفة، ولذا يتعذر في مثل هذه الحالة الاستفادة من القانون (ع/ $\sqrt{ن}$).. كما أن حجم المجتمع قد يكون كبيراً، أو أنه يصعب الاحاطة به. مما يجعل كل العينات الممكنة منه أمراً صعباً أو غير ممكن. وللتدليل على ذلك يكفي أن نذكر بأنه إذا كان حجم المجتمع (100) فرد فقط، وأردنا أن نختار من بينهم كل العينات الممكنة التي عدد أفراد الواحدة منها خمسة فقط، لبلغ عدد هذه العينات في حدود (30) مليون عينة. ولا يخفى أن حساب متوسطات هذه العينات ثم إيجاد قيمة الخطأ المعياري لها أمر صعب جداً، أن لم يكن مستحيلاً، فكيف يكون عليه الحال إذا كبر حجم المجتمع، وكبر كذلك حجم العينة عن هذه الحدود!!

ولكن رغم كل الصعوبات التي قد تواجهنا في إيجاد قيمة الخطأ المعياري. فإن إيجاد قيمة له. ولو تقريبية، تظل ذات فائدة، وبخاصة إذا كان اختيار العينة أو العينات التي ستستخدم في حسابه قد تم على أسس عملية صحيحة. وسنرى في البند القادم بعض الفوائد العملية للخطأ المعياري، حتى ولو كانت قيمته مبنية على البيانات المأخوذة من عينة واحدة فقط، حيث أن ذلك هو واقع الحال بالنسبة إلى غالبية المسائل الاحصائية التي تواجهنا. وف حالة عدم معرفتنا لقيمة الانحراف المعياري للمجتمع. فإننا نستعيض عنها بقمية الانحراف المعياري للعينة عند قيامنا بحساب قيمة الخطأ المعياري.

5-1 الفوائد العملية للخطأ المعياري

إذا علم متوسط المجتع (مـ) وكذلك الخطأ المعياري لمتوسطات العينات المأخوذة منه (ع م)، أمكن بالاستعانة بجدول التوزيع السوي في تحديد نسبة العينات التي تقع متوسطاتها بين أية قيمتين معنيتين من قيمه المختلفة. فمثلاً من المفروض أن يقع 68%من هذه المتوسطات بين القيمتين:

مـ + ع$_{م}$ ، مـ − ع$_{م}$

وأن يقع 95% من المتوسطات بين القيمتين:

مـ + 2ع$_{م}$ ، مـ − 2ع$_{م}$

ويقع 99,9% من هذه المتوسطات بين القيمتين:

مـ + 3ع$_{م}$ ، مـ − 3ع$_{م}$

وهكذا ....

وبعبارة أخرى، يمكن الحكم على مدى تمثيل عينة ما للمجتمع المأخوذة منه من خلال معرفتنا بمقدار بعد متوسطها أوقربه من متوسط المجتمع مقدراً بوحدات الخطأ المعياري. فكلما كان هذا البعد كبيراً كلما كان تمثيل العينة لمجتمعها قليلاً، والعكس بالعكس فإذا أصبح هذا البعد في حدود ث أخطاء معيارية أو أكثر أمكننا القول بأن هذه العينة لم تعد ممثلة لمجتمعنها . وبالتالي لا يمكن الاعتماد على نتائجها أو تعميمها .

ولكن في معظم الدراسات الاحصائية التي تجريها . فإن متوسط المجتمع لا يكون معروفاً لدينا، وبالتالي لا توجد لدينا أية وسيلة للحكم بموجبها على مدى تمثيل العينة للمجتمع المأخوذة منه، وبذلك تكون غايتنا في هذه الدراسات ليس التأكد من مدى تمثيل العينة للمجتمع الماخوذة منه. حيث أن هذا الأمر نأخذ بصحته ضمناً، وذلك إذا تم لنا اختيار العينة حسب الأصول العلمية الصحيحة، وإنما تكون غايتنا بدلاً من ذلك التعرف إلى طبيعة المجتمع وخواصه من خلال معرفتنا بطبيعة العينة المأخوذة منه وخواصها، فإذا تمت لنا معرفة متوسط عينة ما مثلاً، فكيف نستطيع أن نتعرف إلى متوسط مجتمعها في ضوء ذلك!! ونفس التساؤل ينطبق على بقية الاحصائيات الأخرى كالانحراف المعياري أو معامل الارتباط ... إلخ.

6-1 تقدير قيمة متوسط المجتمع إذا عرف متوسط عينة مأخوذة منه

ذكرنا في البند السابق أننا في كثير من الأحيان نحتاج إلى تقدير قيمة متوسط المجتمع من خلال معرفتنا لقيمة متوسط إحدى العينات المأخوذة منه. وقد ذكرنا أيضاً أن الدافع لمثل هذا الإجراء (تقدير متوسط المجتمع اعتماداً على متوسط عينة تؤخذ منه) يرجع لعدم تمكننا من الاحاطة بكامل أفراد المجتمع، ولا حتى بجزء كاف منهم. وتتم عملية التقدير هذه بإحدى طريقتين هما:

أ- **التقدير بقيمة واحدة** :Point Estimation

وفي هذه الحالة نعتبر متوسط المجمتع مساوياً لمتوسط العينة بالضبط ولا يخفى أن هذا

النوع من التقدير عرضة للأخطاء التي قد تنشأ نتيجة للطريقة التي يتم بها اختيار العينة، سيما وأن الاختيار العشوائي لا يضمن بأي حال من الاحوال أن تجيء العينة ممثلة لمجتمعها تمثيلاً صادقاً. ومن هنا كان استعمال هذا الأسلوب في التقدير غير مرغوب فيه. ولا يضمن إعطاء نتائج صادقة يعتمد عليها.

ب- التقدير ضمن فئة Interval Estimation:

وهنا يتم تقدير متوسط المجتمع ضمن فئة من القيم بحيث يقع متوسط العينة في منتصفها. وبعبارة أخرى، فإن فئة التقدير هذه تترامى على طرفي متوسط العينة. ويبتعد طرفاها عنه بعدين متساويين. ولا يخفى أن هذا النوع من التقدير تعطي نتائج أكثر صدقاً وواقعية، حيث أنه يعطينا سلسلة من القيم بدلاً من قيمة واحدة فقط. فعند تقديرنا لعمر طالب في السنة الجامعية الأولى مثلاً، وبدلاً من أن نقول بأن عمره (19) سنة (اعتماداً على أن سن الدخول للمدرسة الابتدائية هو (6) سنوات، وأن سنوات الدراسة قبل الجامعية هي 12 سنة). فقد يكون من الأفضل أن نقول بأن عمره يتراوح ما بين 18 أو 20 سنة أو بين 17 و 21 أو بين 16 و 22 سنة، وهكذا ولا يخفى أن هذا الاسلوب في التقدير يأخذ بعين الاعتبار كون مثل هذا الطالب قد دخل المدرسة مبكراً، أو أنه قد دخلها متأخراً، أو أنه قد رسب في بعض السنوات، إلى غير ذلك من العوامل التي قد تعجعل من التقدير بقيمة واحدة عملية عرضة للخطأ وعدم الدقة.

ولما كانت العلاقة العامة التي تربط بين كل من متوسط المجتع (مـ)، ومتوسط أي من العينات المأخوذة منه (م)، والخطأ المعياري (ع م) يمكن أن يعبر عنها على النحو الآتي:

$$\text{ز} = \frac{\text{م} - \text{مـ}}{\text{ع}_{\text{م}}}$$

حيث إن (ز) هي العلامة المعيارية، ولما كانت قيمة (ز) من المتوقع أن تكون موجبة أو سالبة بحسب كون الفرق (م –هـ) موجباً أو سالباً، فإن قيمة متوسط المجتمع (مـ) يمكن أن تنشق من هذه العلاقة لتصبح على النحو الآتي:

$$\text{مـ} = \text{م} \pm \text{ز} \times \text{ع}_{\text{م}}$$

وهذا يعني أن حدي فئة التقدير يعتمدان في قيمتهما على عاملين اثنين هما:

أ- قمية الخطأ المعياري (ع م).