المقدّرات الإحصائية وخصائصها :

إن الاســـتنتاجات الإحصـــائية هي التعميمات والقرارات التي يمكن اتخاذها بناءً على معلومات أو بيانات تم جمعها أو توفرت لدينا . فمثلاً إذا كان لدينا بستان فيه عدد كبير من أشجار التفاح وقمنا بتفحص عدد من هذه الأشجار وتم ملاحظة ثمار هذه الأشجار فانه ولاشك سنستنتج أن المحصول لهذه الأشجار جيد وان المحصول لهذه السنة عالي وبالعكس . إن هذا يعني أننا بنينا رأينا عن طبيعة المحصول المنتظر لهذه الســـنة بناءً على محصـــول عينة من الأشـــجار تم ملاحظتها، أي انه وبناءً على نتائج العينة والتي قمنا بتعميمها على المجتمع واعتبرنا أن ما حصـلنا عليه من العينة يصـلح تعميمه للمجتمع ، ولكن يبقى السـؤال هو : إلى أي مدى سـوف تكون هذه التعميمات صـحيحة . وللإجابة على هذا السـؤال سـوف نسـتدل على سؤال آخر وهو (ما هو الأسلوب الأمثل والصحيح للتقدير) .

المعلمة (Parameter) : تعرف المعلمة بأنها ثابت إحصـــائي يصـــف المجتمع وغالباً ما تكون مجهولة مما يتطلب تقدير قيم عددية لها على أســـاس العينة (العينات) المختارة من ذلك المجتمع ويرمز للمعلمة بأحرف كبيرة مثل : $(\overline{\mathbf{X}},\ \sigma^2,\ \mathbf{R},\ldots)$.

المقدّر (Estimator) : وهو كمية عددية محسـوبة من قياسـات العينة . والغرض من هذه الكمية هو توفير معلومة حول قيمة غير معروفة عن المجتمع ، ويرمز لها عادة بأحرف صـغيرة أو أحرف كبيرة توضـع عليها العلامة ($\wedge$) للدلالة على التقدير مثل :

$$[(\bar{x} \text{ or } \hat{X}),(s^2 \text{ or } \hat{\sigma}^2),(r \text{ or } \hat{R}),\ldots.]$$

التقدير (Estimate) : وهو القاعدة الرياضية أو الصيغة الرياضية التي تستخدم نتائج العينة للحصول على مقدّر معلمة المجتمع قيد الدراسة :

فعلى سبيل المثال : $\bar{y} = \dfrac{\sum_{i=1}^{n} y_i}{n}$

هو مقدّر يستخدم لتقدير معلمة وهي متوسط المجتمع وهي : $\overline{Y} = \dfrac{\sum_{i=1}^{N} Y_i}{N}$.

من الناحية التطبيقية فان الاسـتدلال بالقيم التقديرية لمعالم المجتمع من قياسـات العينة يمكن أن يأخذ نوعين من التقديرات وهما :

1- التقدير بنقطة Point Estimate : يمثل التقدير هنا بقيمة أو نقطة واحدة محددة بحيث نحاول جعلها اقرب ما يمكن إلى قيمة المعلمة . كان نقول ($\bar{x}$) هو تقدير بنقطة وكذلك (s^2) ، ...الخ .

2- التقدير بفترة (التقدير بفترة ثقة) Interval Estimate : وهي عملية حســـاب حدود يتوقع أن تقع فيه القيمة الحقيقية للمعلمة باحتمال (مســتوى معنوية) معين وكلما كان الاحتمال عالياً كلما كانت هنالك موثوقية اكبر في حصـــولنا على القيمة الحقيقية للمعلمة ضـــمن المدى . كان نقول نحن متأكدين **(90%)** إن درجة

حرارة الجو تقع مـابين ($35^\circ - 45^\circ$) . أي أننـا نثق بمقـدار (**90%**) بـان درجـة حرارة الجو هي مـابين ($35^\circ - 45^\circ$) . ويعبر عن فترة الثقة رياضياً بالصيغة الآتية :

$$\Pr[c_1 < \theta < c_2] = 1 - \alpha$$

حيث أن :

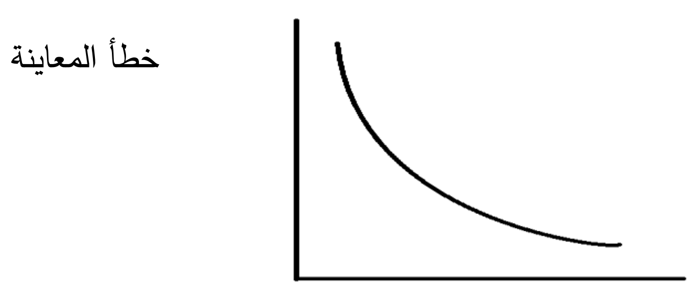
θ : قيمة المعلمة .

c_1, c_2 : الحدين الأدنى والأعلى على التوالي والذين نتوقع أن تقع القيمة الحقيقية (المعلمة) بينهما .

α : تمثل احتمال الوقوع في الخطأ ، وفي التطبيقات العملية غالباً ما يتم اختيار قيمة (α) لان تكون مساوية إلى **[0.05 or 0.01]** ويطلق على ($1-\alpha$) مستوى المعنوية ، مستوى الثقة او درجة الثقة . فمثلاً : عندما نقول بأننا نتوقع أن يصــيب احمد الهدف مابين (10−20) إصـــابة بدرجة ثقة (95%) أي أننا نثق (95%) بان احمد سيصيب الهدف ضمن الفترة وانه باحتمال قدره (5%) لن يصيب الهدف

الأخطاء ومصادرها :

عند إجراء دراســـة لظاهرة معينة في مجتمع ما بطريقة العينة تصـــطحبه عدة أخطاء منها أخطاء المعاينة (أخطاء عشوائية) وأخطاء التحيز (أخطاء غير عشوائية) وأخطاء أخرى ، فبالنسبة لأخطاء المعاينة فهي ناتجة عن اختيار جزء من المجتمع فبعد حســـاب المقدّرات نجد انه كثيراً ما تكون غير مســـاوية إلى المعلمة (القيمة الحقيقية للمجتمع) لذا فان الفرق بين القيمة التقديرية و القيمة الحقيقية يســمى خطأ المعاينة وهذا الخطأ يقل كلما كبر حجم العينة وكلما تم اســتخدام الأســلوب الأمثل للمعاينة أيضــاً . والشــكل الآتي يوضح العلاقة بين حجم العينة وخطا المعاينة .



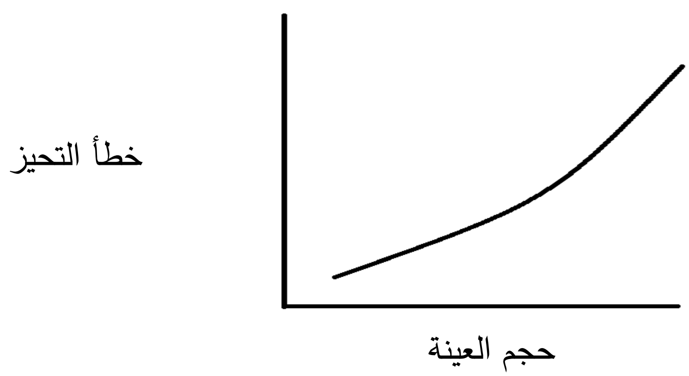
أما بالنسبة إلى أخطاء التحيز فيمكن إجمالها بالاتي :

1− خطأ الملاحظة، وسـببه عدم الدقة في القياس أو عطل بعض الأجهزة المسـتخدمة أو تدوين بعض الأرقام في أماكنها الغير صحيحة . أو عدم تدوين أو قياس مفردة أو مجموعة مفردات..الخ .

2- أخطاء في تصميم استمارة الاستبيان . إذ تشترط الاستمارة الجيدة ، وضوح الأسئلة وشمولها لموضوع الدراسة وعدم التحيز والتسلسل المنطقي للأسئلة ... الخ وان عدم تحقق هذه الشروط فسوف تكون النتائج عرضة لخطا التحيز وعدم الشمول وبالتالي إلى نتائج غير دقيقة .

3- أخطاء في تحليل البيانات وتفسيرها . وهذا النوع من الأخطاء يشتمل على الأخطاء الحسابية في إيجاد المقاييس المختلفة ومن ثم اتخاذ القرارات بناءً على الحسابات الخاطئة مما يؤدي إلى نتائج غير دقيقة .

ويعد خطا التحيز أكثر خطورة من خطا المعاينة وذلك لصعوبة السيطرة عليه وانه بزيادة حجم العينة يزداد هذا النوع من الأخطاء . والشكل الآتي يوضح علاقة حجم العينة بأخطاء التحيز .



وبشكل عام فان كبر الخطأ أياً كان نوعه يجعل نتائج الدراسة مضللة ولا أهمية لها .

خصائص التقديرات الإحصائية :

علمنا مما سبق بأنه يتم تقدير معلمة المجتمع من عينة يتم سحبها من هذا المجتمع إلا انه يمكننا ملاحظة بأنه يمكن الحصول على أكثر من تقدير فمثلا عندما نقول بان (μ) تمثل الوسط الحسابي للمجتمع وأردنا حساب تقدير لهذه المعلمة فانه سوف نقوم بسحب عينة عشوائية من هذا المجتمع إذ من الممكن أن تكون هذه العينة عبارة عن مفردة واحدة ونجد من خلالها أن $\left(\bar{y}_1 = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{1} y_i}{1} = \frac{y_1}{1} = y_1\right)$. أو أن تكون العينة عبارة عن مفردين فيكون لدينا : $\left(\bar{y}_2 = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{2} y_i}{2} = \frac{y_1 + y_2}{2}\right)$ وكذلك نستطيع استخدام (n) مفردة وكالاتي : $\left(\bar{y}_k = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{y_1 + y_2 + y_3 + \dots + y_n}{n}\right)$. من هنا نرى أن هنالك أكثر من تقدير للوسط الحسابي للمجتمع . والسؤال الذي يطرح نفسه أي من هذه التقديرات هي أفضل لتقدير الوسط الحسابي للمجتمع من هنا يمكننا القول بأنه لاختيار التقدير الملائم لمعلمة المجتمع يجب أن يمتلك خصائص معينة منها:

1− عدم التحيز : Unbiasedness

يقال أن المقدّر $(\hat{\theta})$ هو مقدّر غير متحيز للمعلمة (θ) إذا كان: $\mathbf{E(\hat{\theta})=\theta}$ ، وهذه الصيغة تعني : يكون المقدّر غير متحيز إذا كانت القيمة المتوقعة له تساوي القيمة الحقيقية . وبمعنى آخر يكون التقدير جيد وغير متحيز إذا تمركزت قيمة ذلك التقدير حول المعلمة وبخلافه يكون المقدّر متحيزاً . وفي حالة الحصول على أكثر من مقدّر فانه يتم إيجاد متوسط المقدّرات التي يتم الحصول عليها من كافة العينات الممكن سحبها من المجتمع فيما يخص المعلمة (θ). وكمثال على ذلك إذا تم سحب عينة عشوائية مكونة من $(\mathbf{n})$ مفردة من مجتمع حجمه $(\mathbf{N})$ فانه يمكن الحصول على $(\mathbf{C_n^N})$ عينة مختلفة بحيث أن لكل عينة مقدّر خاص بها وهو $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, ..., \hat{\theta}_{C_n^N})$ عند ئذٍ يمكننا القول بان المقدّر $(\hat{\theta})$ هو تقدير غير متحيز للمعلمة (θ) إذا كان :

$$\mathbf{E(\hat{\theta}) = \frac{\sum_{i=1}^{C_n^N} \hat{\theta}_i}{C_n^N} = \theta}$$

وفي غير ذلك يقال أن المقدّر $(\hat{\theta})$ هو مقدّر متحيز وان مقدار التحيز (Bias) ويمكن قياس مدى تحيزه بالصيغة الآتية :

$$\mathbf{Bias(\hat{\theta}) = E(\hat{\theta}-\theta)}$$

ملاحظة : إذا كان المقدّر متحيزاً فان هذا التحيز إما أن يكون موجب أو ســــالب أي انه إذا كانت $(\hat{\theta})$ هو مقدّر متحيز للمعلمة (θ) فان :

$$\mathbf{Bias(\hat{\theta}) = E[\hat{\theta}-\theta] = \begin{cases} >0 & \text{Positive bias} \\ <0 & \text{Negative bias} \end{cases}}$$

ويعرف المقدار : $\mathbf{E[\hat{\theta}-\theta]}$ بأنه متوسط مربعات الخطأ (MSE) للمقدّر $(\hat{\theta})$ بحيث أن :

- إذا كان المقدّر غير متحيز $\Rightarrow$ $\mathbf{MSE(\hat{\theta}) = V(\hat{\theta})}$.
- إذا كان المقدّر متحيز $\Rightarrow$ $\mathbf{MSE(\hat{\theta}) = V(\hat{\theta}) + [Bias(\hat{\theta})]^2 = V(\hat{\theta}) + [E(\hat{\theta}-\theta)]^2}$.

وكلما كانت قيمة (MSE) صـــغيرة فان هذا مؤشـــر على جودة المقدّر عند مقارنته المقدّرات مع مقدّرات أخرى (كما انه مؤشر على قرب قيمة المقدر الى معلمة المجتمع).

مثال : مجتمع حجمه (N) بالقياســات $(\mathbf{Y = 1,3,5,2})$. تم ســحب عينة عشــوائية بحجم $(\mathbf{n=2})$ بين أن : $\mathbf{E[\bar{y}] = \bar{Y}}$.

الحل :

من قياسات المجتمع نجد أن :

$$\mathbf{\bar{Y} = \frac{\sum_{i=1}^{N} Y_i}{N} = \frac{1+3+5+2}{4} = 2.75}$$

إن عدد العينات الممكن سحبها من هذا المجتمع بحجم (n) هي :

$$C_n^N = C_2^4 = \frac{4!}{2!\ 2!} = 6$$

وهي كما في الجدول التالي :

تسلسل العينات	1	2	3	4	5	6	Σ
العينات	(1,3)	(1,5)	(1,2)	(3,5)	(3,2)	(5,2)	
$\bar{y}$	2	3	1.5	4	2.5	3.5	16.5

$$\because E[\hat{\theta}] = \frac{\sum_{i=1}^{C_n^N} \hat{\theta}_i}{C_n^N} \Rightarrow E[\bar{y}] = \frac{\sum_{i=1}^{C_n^N} \bar{y}_i}{C_n^N} = \frac{2+3+1.5+4+2.5+3.5}{6} = \frac{16.5}{6} = 2.75$$

$$\Rightarrow E[\bar{y}] = \bar{Y}$$

وهذه النتيجة تقودنا إلى القول بان ($\bar{y}$) هو مقدّر غير متحيز للمعلمة ($\bar{Y}$) .