Chapter 1
Introduction to Statistics

Section 1-1
Review and Preview

Preview عرض مسبق

Polls, studies, surveys and other data collecting tools collect data from a small part of a larger group so that we can learn something about the larger group. This is a common and important goal of statistics: Learn about a large group by examining data from some of its members.

استطلاعات الرأي والدراسات والاستقصاءات وغيرها من البيانات جمع أدوات جمع البيانات من جزء صغير من مجموعة أكبر حتى يتسنى لنا أن نتعلم شيئا عن مجموعة أكبر. هذا هو هدف مشترك ومهم من الإحصائيات: تعرف على مجموعة كبيرة من خلال دراسة البيانات من بعض أعضائها.

Preview عرض مسبق

In this context, the terms sample and population have special meaning. Formal definitions for these and other basic terms will be given here.

في هذا السياق، حيث العينة والسكان لها معنى خاص. سيتم تعاريف رسمية لهذه المصطلحات والأساسية الأخرى هنا.

In this section we will look at some of the ways to describe data. في هذا القسم سوف نلقي نظرة على بعض الطرق لوصف البيانات.

Data

❖ **Data البيانات**

collections of observations (such as measurements, genders, survey responses)

مجموعات من الملاحظات (مثل القياسات والأجناس، والاستجابات مسح).

شرح// مثلا لو قلنا بناخذ اطوال مجموعه من الاطفال ناخذ عينه من المجموعة ونقيس الاطوال طول 120 او 130 او 100 الاعداد هذي هي الداتا المقصوده واذا اعمار الاعمار هي الداتا واذا درجات هم الدرجات هي الداتا

WEDAD ALMOHANNA-140027958

Statistics

❖ **Statistics** الإحصاء

is the science of planning studies and experiments, obtaining data, and then organizing, summarizing, presenting, analyzing, interpreting, and drawing conclusions based on the data. هو علم من الدراسات والتجارب التخطيط، والحصول على البيانات، ومن ثم تنظيم، يلخص، عرض وتحليل، وتفسير، واستخلاص الاستنتاجات استنادا إلى البيانات

Population

❖ **Population** المجموعه

The complete collection of all individuals (scores, people, measurements, and so on) to be studied; the collection is complete in the sense that it includes *all* of the individuals to be studied. مجموعة كاملة من جميع الأفراد (عشرات، والناس، والقياسات، وهلم جرا) ليتم دراستها. مجموعة كاملة بمعنى أنه يشمل جميع الأفراد لدراستها.

شرح// يعني الكل اللي بناخذ منهم العينه ونسوي عليها دراسه ركز على كلمة كل.

Census versus Sample تعداد مقابل عينة

❖ **Census** التعداد Collection of data from *every* member of a population.

جمع البيانات من كل عضو من السكان. شرح/ التعداد نفس البيلكيشن يعني الكل فيها وهي اللي ناخذ منها العينه.

❖ **Sample** عينة *Subcollection* of members selected from a population.

المجموعة الفرعية من أعضاء مختارين من السكان. شرح / هي العينه من المجموعه بالعاميه.

Chapter Key Concepts فصل مفتاح المفاهيم

❖ **Sample data must be collected in an appropriate way, such as through a process of *random* selection.**
يجب جمع بيانات العينة بطريقة مناسبة، وذلك من خلال عملية الاختيار العشوائي.

❖ **If sample data are not collected in an appropriate way, the data may be so completely useless that no amount of statistical torturing can salvage them.**

إذا لم يتم جمع بيانات العينة بطريقة مناسبة، قد تكون البيانات حتى عديمة الفائدة تماما أن أي قدر من تعذيب الإحصائي يمكن إنقاذ لهم.

Section 1-2
Statistical Thinking

Key Concept المفهوم الرئيسي

This section introduces basic principles of statistical thinking used throughout this book. Whether conducting statistical analysis of data that we have collected, or analyzing a statistical analysis done by someone else, we should not rely on blind acceptance of mathematical calculation. We should consider these factors:

يقدم هذا القسم المبادئ الأساسية للتفكير الإحصائية المستخدمة في هذا الكتاب. ما إذا كان إجراء التحليل الإحصائي للبيانات التي جمعناها، أو تحليل التحليل الإحصائي الذي قام به شخص آخر، لا ينبغي لنا أن نعتمد على القبول الأعمى من العمليات الحسابية. يجب علينا النظر في هذه العوامل:

Key Concept (continued) مفهوم رئيسي (تابع)

- **Context of the data** سياق(مسار) البيانات
- **Source of the data** مصدر البيانات
- **Sampling method** طريقة أخذ العينات
- **Conclusions** الاستنتاجات
- **Practical implications** نواتج عملية

Context السياق (المسار)

- **What do the values represent?** ماذا تمثل القيم؟
- **Where did the data come from?** من أين تأتي البيانات؟
- **Why were they collected?** لماذا كانوا جمعها؟
- **An understanding of the context will directly affect the statistical procedure used.** وفهم من سياق تؤثر بشكل مباشر على إجراء إحصائي المستخدمة.

Source of data مصدر البيانات

- **Is the source objective?** هل المصدر هدف؟
- **Is the source biased?** هل المصدر متحيز؟
- **Is there some incentive to distort or spin results to support some self-serving position?**

هناك بعض الحوافز لتشويه أو النتائج تدور لدعم بعض موقف لخدمة مصالح ذاتية؟

- **Is there something to gain or lose by distorting results?**

هل هناك شيء لكسب أو فقدان من خلال تشويه النتائج؟

- **Be vigilant and skeptical of studies from sources that may be biased..** توخي الحذر وتشكك في الدراسات من المصادر التي قد تكون متحيزة

Sampling Method طريقة أخذ العينات

- Does the method chosen greatly influence the validity of the conclusion? هل الطريقة المختارة بشكل كبير تؤثر على صحة الاستنتاج؟
- Voluntary response (or self-selected) samples often have bias (those with special interest are more likely to participate). These samples' results are not necessarily valid. استجابة طوعية (أو اختيار الذاتي) عينات غالبا ما يكون التحيز (ذوي المصالح الخاصة أكثر عرضة للمشاركة). نتائج هذه العينات "غير صالحة بالضرورة.
- Other methods are more likely to produce good results.

أساليب أخرى من المرجح أن تسفر عن نتائج جيدة.

Conclusions الاستنتاجات

- Make statements that are clear to those without an understanding of statistics and its terminology.

الإدلاء ببيانات واضحة لأولئك دون فهم الإحصاءات ومصطلحاته. شرح/ يجب اختيار مصطلحات مفهومه للناس في عمل الاحصائيات.

- Avoid making statements not justified by the statistical analysis. تجنب الإدلاء بتصريحات لا تبرره التحليل الإحصائي.

Practical Implications نواتج عملية

- State practical implications of the results. الآثار العملية حالة من النتائج.
- There may exist some *statistical significance* yet there may be NO *practical significance*.

قد تكون هناك بعض الأهمية الإحصائية بعد قد يكون هناك أي مغزى عملي.

- Common sense might suggest that the finding does not make enough of a difference to justify its use or to be practical. الحس السليم قد يوحي بأن النتيجة لا تجعل ما يكفي من الفرق لتبرير استخدامه أو أن يكون عملي.

Statistical Significance دلالة إحصائية

- Consider the likelihood of getting the results by chance.

النظر في احتمال الحصول على النتائج عن طريق الصدفة.

- If results could easily occur by chance, then they are *not statistically significant*.

إذا كانت النتائج يمكن أن تحدث بسهولة عن طريق الصدفة، ثم أنها ليست ذات دلالة إحصائية.

- If the likelihood of getting the results is so small, then the results are *statistically significant*.

إذا كان احتمال الحصول على نتائج صغيرة جدا، ثم كانت النتائج ذات دلالة إحصائية.

Section 1-3
Types of Data

Key Concept المفهوم الرئيسي

The subject of statistics is largely about using sample data to make inferences (or generalizations) about an entire population. It is essential to know and understand the definitions that follow. موضوع الاحصاءات إلى حد كبير حول استخدام بيانات العينة لجعل الاستدلالات (أو التعميمات) عن مجموعه بأكملها. ومن الضروري معرفة وفهم التعاريف التالية.

Parameter معامل

❖ **Parameter** معامل

a numerical measurement describing some characteristic of a population. قياس العددي واصفا بعض الخصائص للمجموعه.

Population المجموعة

↕

Parameter معامل

Statistic احصائيه

❖ **Statistic** إحصائية

a numerical measurement describing some characteristic of a sample. قياس العددي واصفا بعض من سمات عينة.

Sample عينه

↕

Statistic احصائية

شرح/ اللي ناخذه من البوبيوليشن اسمه باراميتر واللي ناخذه من السيمبل اسمه ستاتيستيك يعني لو شغال احصاء بنكون نشتغل على العينات و لو شغالين على المجموعات البوبيوليكيشن يعني بيكون شغلنا على الباراميتر .
ملاحظه مهمه// تجي بالأختبارات صح و خطأ و اختياري يجب فهم المعنى جيدا .

Quantitative Data البيانات الكمية

❖ **Quantitative (or numerical) data** البيانات الكمي (أو العددي) **consists of *numbers* representing counts or measurements.** يتكون من أرقام تمثل التهم أو القياسات.

Example:The weights of supermodels. على سبيل المثال: الأوزان من عارضات الأزياء

Example: The ages of respondents. على سبيل المثال: أعمار المشاركين

Categorical Data بيانات تسلسلية

❖ **Categorical (or qualitative or attribute) data** القاطع (أو نوعي أو سمة) البيانات

consists of names or labels (representing categories)

يتكون من الأسماء أو التسميات (تمثل فئات)

Example: The genders (male/female) of professional athletes. على سبيل المثال: الجنسين (ذكور / إناث) من الرياضيين المحترفين

Example: Shirt numbers on professional athletes uniforms - substitutes for names.. على سبيل المثال: أرقام قميص على الرياضيين الزي المهنية - بدائل للأسماء

Working with Quantitative Data العمل مع البيانات الكمية

Quantitative data can further be described by distinguishing between discrete and continuous types.

يمكن أن تزيد من البيانات الكمية وصفها عن طريق التمييز بين أنواع المنفصلة والمتصلة.

شرح/ تنقسم الداتا الى كوانتيتيف يعني الي كمية وفيها ارقام والى كواليتيف(كاتوجري)يعني نوعيه اللي فيها غير الارقام جنس (ذكر-انثى) وله ارقام بس مثل الرقم الجامعي.

Discrete Data بيانات متقطعه

❖ **Discrete data** بيانات متقطعه

result when the number of possible values is either a finite number or a 'countable' number (i.e. the number of possible values is 0, 1, 2, 3, . . .) تنتج عند عدد من القيم الممكنة إما أن يكون عدد محدود أو عدد "معدود"(أي عدد من القيم الممكنة هي 0، 1، 2، 3، . . .)

Example: The number of eggs that a hen lays. مثال: عدد البويضات التي تضع الدجاجة

شرح/بيانات متقطعة لوكنت في مزرعة فراخ و ابي اراقب انتاجها كل فرخه تبض بيضه وحده فقط مستحيل تجيب بيضه ونص عشان كذا يسمونها متقطعة .(باختصار اللي فهمته انها تبدأ من الصفر و تكون اعداد صحيحه)

Continuous Data بيانات المستمر

❖ **Continuous (numerical) data** مستمر (العددية) البيانات

result from infinitely many possible values that correspond to some continuous scale that covers a range of values without gaps, interruptions, or jumps. ينتج من عدد لانهائي من القيم الممكنة التي تتوافق مع بعض على نطاق والمستمر الذي يشمل مجموعة من القيم دون ثغرات، انقطاع، أو يقفز

Example: The amount of milk that a cow produces; e.g. 2.343115 gallons per day. مثال: كمية الحليب التي تنتجها البقرة. مثلا 2.343115 جالون يوميا

شرح/بيانات مستمره عكس البيانات المتقطعه الداتا مستمره فيها يكون في صورة كسور مثل كمية الحليب اللي تنتجه بقره 2,434جالون في كل يوم .(باختصار اللي فهمته انها لا تبدأ بالصفر ابداً و تكون اعداد عشريه)

Levels of Measurement مستويات القياس

Another way to classify data is to use levels of measurement. Four of these levels are discussed in the following slides.

وهناك طريقة أخرى لتصنيف البيانات لاستخدام مستويات القياس. أربعة من هذه المستويات تمت مناقشتها في الشرائح التالية. (اربع مستويات للقياس)

Nominal Level المستوى الاسمي

1- Nominal level of measurement المستوى الاسمي للقياس

characterized by data that consist of names, labels, or categories only, and the data cannot be arranged in an ordering scheme (such as low to high) تتميز البيانات التي تتكون من أسماء، والعلامات، أو فئات فقط، ولا يمكن أن يتم ترتيب البيانات في مخطط ترتيب (مثل الأقل إلى الأعلى)

Example: Survey responses yes, no, undecided مثال: الردود مسح نعم، لا، غير محدد

شرح/ مستوى النورمل مقسم لى فئات نعم/لا/ الى حدما لا نستطيع ترتيبها تصاعديا او تنازلياً لانها نعم /لا .

Ordinal Level المستوى الترتيبي

2- Ordinal level of measurement المستوى الترتيبي القياس

involves data that can be arranged in some order, but differences between data values either cannot be determined or are meaningless. تتضمن البيانات التي يمكن ترتيبها في بعض الأمر، ولكن الخلافات بين قيم البيانات إما لا يمكن تحديده أو لا معنى لها

Example: Course grades A, B, C, D, or F على سبيل المثال: الدرجات الدورة A، B، C، D، أو F

شرح/ الاوردينال يستخدم الفئات نفس اللي قبل بس نقدر نقسمه تصاعدي و تنازلي.

Interval Level المستوى الفاصل

3- Interval level of measurement مستوى الفاصل القياس

like the ordinal level, with the additional property that the difference between any two data values is meaningful, however, there is no natural zero starting point (where none of the quantity is present) مثل المستوى الترتيبي، مع الخاصية الإضافية التي الفرق بين أي قيم البيانات اثنين هو مفيد، ومع ذلك، لا يوجد أي نقطة انطلاق طبيعية صفر (حيث لا أحد من كمية موجود)

Example: Years 1000, 2000, 1776, and 1492 مثال: سنوات 1000، 2000، 1776، و 1492

شرح/ انترافيل هي نومريكال عدديه يعني بس لا تبدأ بالصفر أبداااا مثل السنوات بالتواريخ لا توجد سنه صفر.

Ratio Level نسبة مستوى

4- Ratio level of measurement مستوى نسبة قياس

the interval level with the additional property that there is also a natural zero starting point (where zero indicates that none of the quantity is present); for values at this level, differences and ratios are meaningful. مستوى الفاصلة مع الخاصية الإضافية أن هناك أيضا نقطة انطلاق طبيعية صفر (حيث الصفر يشير إلى أن أيا من الكمية غير موجودة)؛ للقيم على هذا المستوى، والاختلافات والنسب ذات مغزى

Example: Prices of college textbooks ($0 represents no cost, a $100 book costs twice as much as a $50 book) مثال: أسعار الكتب الدراسية (0 $ تمثل أي تكلفة، وتكاليف كتاب 100 $ ضعفي كتاب 50 $)

شرح/ الراتشيو او الراتيو نفس اللي قبل بس هنا يوجد بها صفر مثل لما يتكلم عن كوست قيمه لشيء ممكن مايكون له تكلفه وتكون قيمته صفر .

Summary - Levels of Measurement ملخص - مستويات القياس

- **Nominal** - categories only الاسمية - فئات فقط
- **Ordinal** - categories with some order. ترتيبي - فئات مع بعض النظام
- **Interval** - differences **but no natural** starting point.

فاصل - الاختلافات ولكن لا نقطة البداية الطبيعية

- **Ratio** - differences **and a natural** starting point

نسبة - الخلافات ونقطة انطلاق طبيعية

شرح/ اول ثنين تتكلم عن الكاتوجري الفئات, و ثاني اثنين عن الأرقام.

ملاحظة مهمه/ نذاكر الملخص لهذه النقاط والشرح السابق لهم فقط للفهم وليس الحفظ .

Recap خلاصة

In this section we have looked at:في هذا القسم لقد بحثنا في:

- **Basic definitions and terms describing data**التعاريف الأساسية والمصطلحات التي تصف البيانات
- **Parameters versus statistics.**المعاملات مقابل الإحصاءات
- **Types of data (quantitative and qualitative).**أنواع البيانات (الكمية والنوعية)
- **Levels of measurement.**مستويات القياس

ملاحظه / هالشي مو للحفظ طبعاً .

Quiz اختبار خاص بالجزئية السابقة

1- Voluntary response is a bad sample

True False

2- Data are classified due to level of measurement to these categories

- Nominal-ordinal-interval-ratio
- Quantitative data-categorical data-discrete data-continous data
- Nominal-ordinal
- Discrete-continous

3- Voluntary response sample (or self-selected sample) is one in which the respondents themselves decide whether to be included

True False

4- The term "population" in statistics means the complete collection of all individuals to be included

True False

5- the population is a collection of all individuals, objects, or measurements of interest

True False

6- The collection of all cars registered in Saudi Arabia is an example of population

True False

Section 1-4
Critical Thinking

Key Concepts المفاهيم الرئيسية

- Success in the introductory statistics course typically requires more **common sense** than mathematical expertise..النجاح في الإحصاءات دورة تمهيدية يتطلب عادة بمعنى أكثر شيوعا من الخبرات الرياضية
- Improve skills in interpreting information based on data. تحسين المهارات في تفسير المعلومات استنادا إلى البيانات.
- This section is designed to illustrate how common sense is used when we think critically about data and statistics. تم تصميم هذا القسم لتوضيح كيفية شعور شائع يستخدم عندما نفكر بشكل نقدي حول البيانات والإحصاءات.
- Think carefully about the context, source, method, conclusions and practical implications.
فكر مليا في السياق، مصدر، طريقة والاستنتاجات والآثار العملية.

Misuses of Statistics سوء استخدام للإحصاء

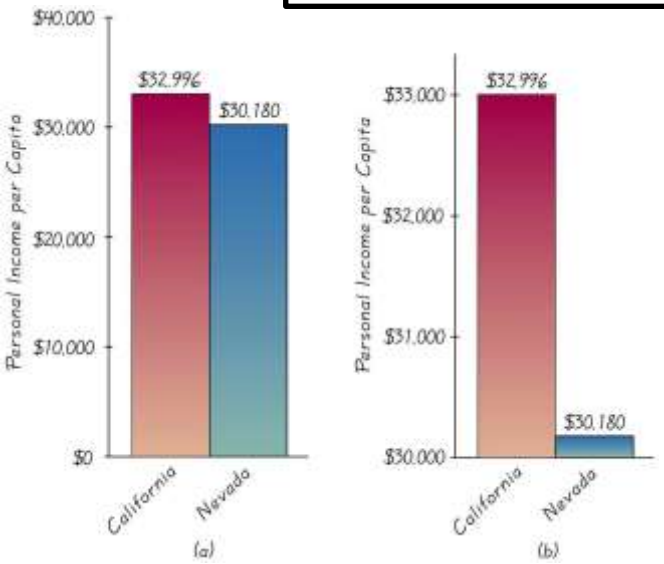
1. **Evil intent on the part of dishonest people.** نية الشر من جانب الناس غير شريفة.
2. **Unintentional errors on the part of people who don't know any better.** أخطاء غير مقصودة من جانب الناس الذين لا يعرفون أي على نحو أفضل.

We should learn to distinguish between statistical conclusions that are likely to be valid and those that are seriously flawed.

علينا أن نتعلم كيف نميز بين الاستنتاجات الإحصائية التي من المحتمل أن تكون صالحة وتلك التي معيبة على محمل الجد.

الشرح// باختصار النوع الأول ممكن يكون له نوايا سيئة و بالجزء الثاني نية سليمه لكن يعطي نتائج غير دقيقه لانه يكون مش عارف الجواب اصلاً.

Graphs الرسوم البيانية

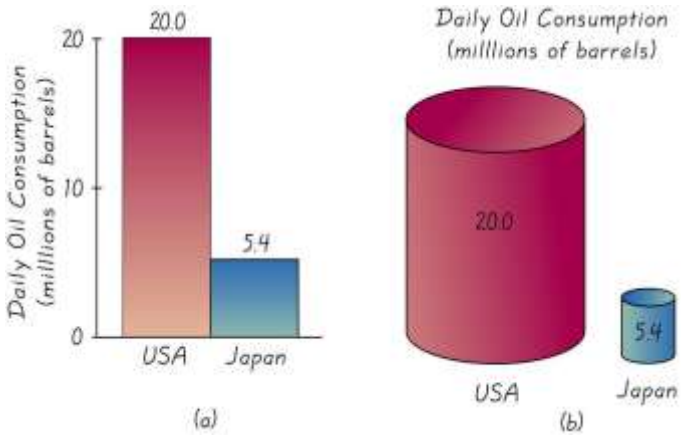


الشرح :(معنى كلامه شوف الصورتين كأنها مختلفة بس اقرأ البيانات تلاقيها متطابقة اللهم احداثيات العامود اختلف تقسيمها
اول شي مقسمه ب10 ألاف بالرسمه ع اليسار و بالرسم اللي ع اليمين ابتدأ ب 30 الف و زاد ألف الف بكل خطوه بالنتيجه الفرق بين كالفورنيا ونيفادا بالرسمتين نفس الشي)
باختصار الرسم عباره عن خدعه بصريه يجب الانتباه لها

To correctly interpret a graph, you must analyze the numerical information given in the graph, so as not to be misled by the graph's shape. READ labels and units on the axes!

لتفسير الرسم البياني بشكل صحيح، يجب تحليل المعلومات العددية التي وردت في الرسم البياني، حتى لا تنخدع شكل الرسم البياني. قراءة العلامات وحدات على المحاور!

Pictographs التصويرية



شرح // رسم الجراف عشان يكون مثل ثلاثي الابعاد يعطي حجم اكبر مبالغ فيه باختصار الرسم عباره عن خدعه بصريه يجب الانتباه لها.

Part (b) is designed to exaggerate the difference by increasing each dimension in proportion to the actual amounts of oil consumption.

تم تصميم الجزء (ب) للمبالغة الفرق عن طريق زيادة كل البعد بما يتناسب مع المبالغ الفعلية للاستهلاك النفط.

Bad Samples عينات سيئة

❖ **Voluntary response sample (or self-selected sample)**

عينة استجابة طوعية (أو عينة مختارة النفس)

one in which the respondents themselves decide whether to be included. واحد الذي المستطلعين أنفسهم تقرر ما إذا كانت ستدرج

In this case, valid conclusions can be made only about the specific group of people who agree to participate and not about the population. في هذه الحالة، يمكن أن تقدم استنتاجات صحيحة فقط عن مجموعة معينة من الناس الذين يوافقون على المشاركة وليس عن السكان.

شرح // تعطي الحق للشخص يكون من ضمن اللي عليهم الاحصائية أو لا وهالشي طبعا خطأ ويعطي نتايج خطأ.

Correlation and Causality الارتباط والسببية

❖ **Concluding that one variable *causes* the other variable when in fact the variables are linked.**

وخلصت إلى أن متغير واحد يسبب متغير آخر في حين أن ترتبط المتغيرات

Two variables may seemed linked, smoking and pulse rate, this relationship is called correlation. Cannot conclude the one causes the other. *Correlation does not imply causality*.

قد يبدو اثنين من المتغيرات المرتبطة، والتدخين ومعدل النبض، وهذا ما يسمى علاقة الارتباط. لا يمكن استنتاج واحد يسبب الآخر. ؟ لا يعني ارتباط السببية.

شرح //الاثنين مرتبطين ببعض ولكن غير مسببين لبعض هذه هي الجمله المهمه في هالجزء وزبدة هالجزء (ان الارتباط لا يعني السببيه)

Small Samples عينات صغيرة

Conclusions should not be based on samples that are far too small.

لا ينبغي أن تستند استنتاجات بشأن العينات التي هي أبعد ما تكون صغيرة جدا.

Example: Basing a school suspension rate on a sample of only three students.

على سبيل المثال: مستندة معدل تعليق الدراسة على عينة من الطلاب الثلاثة فقط

شرح // باختصار يجب ان تكون النسبه معقوله و منطقيه و ما تكون صغيره جدا مثل ناخذ 5 طلاب من ألف طالب.

Percentages النسب المئوية

Misleading or unclear percentages are sometimes used. For example, if you take 100% of a quantity, you take it all. If you have improved 100%, then are you perfect?! 110% of an effort does not make sense.

وتستخدم في بعض الأحيان نسبة مضللة أو غير واضحة. على سبيل المثال، إذا كنت تأخذ 100٪ من الكمية، وكنت أعتبر كل شيء. إذا كنت قد تحسنت بنسبة 100٪، ثم أنت الكمال ؟! 110٪ من جهد لا معنى له.

شرح // لما اقول سويت دراسة على طلاب النتيجة كانت 100% بيكون شغلي بيرفكت وتمام بس ما يصير نقول نتائجنا كانت 110% لانها غير منطقية والاحصائية ايضا غير منطقيه.

Loaded Questions أسئلة تحميل

If survey questions are not worded carefully, the results of a study can be misleading. Survey questions can be “loaded” or intentionally worded to elicit a desired response.

إذا لم يتم اللهجة أسئلة الاستطلاع بعناية، ونتائج الدراسة يمكن أن تكون مضللة. أسئلة الاستطلاع يمكن "تحميل" أو عمدا اللهجة للحصول على الاستجابة المطلوبة.

Too little money is being spent on “welfare” versus too little money is being spent on “assistance to the poor.” Results: 19% versus 63%.

تنفق القليل من المال أيضا على "الرعاية الاجتماعية" في مقابل القليل من المال أيضا ينفق على نتائج "المساعدة للفقراء.": 19٪ مقابل 63٪.

شرح // استخدام ترتيب معين في اسئلة الدراسة ممكن يؤثر على نتائج الاحصائية.

Order of Questions ترتيب الأسئلة

Questions are unintentionally loaded by such factors as the order of the items being considered.

يتم تحميل الأسئلة عن غير قصد من قبل عوامل مثل ترتيب العناصر التي يجري النظر فيها.

Would you say traffic contributes more or less to air pollution than industry? Results: traffic - 45%; industry - 27%.

لن أقول لكم المرور ويساهم بشكل أو بآخر في تلوث الهواء من الصناعة؟ النتائج: حركة - 45٪. صناعة - 27٪.

When order reversed. Results: industry - 57%; traffic - 24%

عندما أمر عكسه. النتائج: صناعة - 57٪. حركة المرور - 24٪

شرح // النسب تختلف اذا غيرنا الترتيب يجي بالاختبار سؤال /هل تغيير ترتيب الاسئله يؤثر على نتائج الاحصائية او التحليل؟ طبعاً نجاوب بنعم يؤثر.

Nonresponse عدم الاستجابة

Occurs when someone either refuses to respond to a survey question or is unavailable.. يحدث عندما يكون أي شخص يرفض الرد على سؤال الاستطلاع أو غير متوفر

People who refuse to talk to pollsters have a view of the world around them that is markedly different than those who will let poll-takers into their homes. الناس الذين يرفضون التحدث الى استطلاعات الرأي لديهم رؤية للعالم من حولهم التي تختلف بشكل ملحوظ من أولئك الذين سوف تسمح محتجزي الاقتراع إلى منازلهم.

Missing Data بيانات مفقودة

Can dramatically affect results. يمكن أن تؤثر على النتائج بشكل كبير.

Subjects may drop out for reasons unrelated to the study.

مواضيع قد تنخفض لأسباب لا علاقة لها بالدراسة.

People with low incomes are less likely to report their incomes.

ذوي الدخل المنخفض هم أقل عرضة للمعاناة من دخولهم.

US Census suffers from missing people (tend to be homeless or low income). . الإحصاء الأمريكي يعاني من المفقودين (تميل إلى أن تكون بلا مأوى الدخل أو منخفضة)

Self-Interest Study دراسة مصالحهم الذاتية

Some parties with interest to promote will sponsor studies.
بعض الأطراف مع الاهتمام لتعزيز ورعاية الدراسات.
Be wary of a survey in which the sponsor can enjoy monetary gain from the results.. أن نكون حذرين من مسح الذي الراعي يمكن أن يتمتع تحقيق مكاسب مالية من نتائج
When assessing validity of a study, always consider whether the sponsor might influence the results. عند تقييم صحة دراسة، والنظر دائما ما إذا كان الراعي قد تؤثر على النتائج.

شرح //مثلا مؤسسة او شركة تتشارك في الاحصائية حقتنا. لازم نكون حذرين لا تأخذ النتائج لمصلحة شخصية لازم نحافظ على الاستفتاء حقنا ويكون صحيح.

Precise Numbers أرقام دقيقة

Because as a figure is precise, many people incorrectly assume that it is also *accurate*. لأن كشخصية غير دقيق، كثير من الناس بشكل غير صحيح يفترض أنه هو أيضا دقيقة.

A precise number can be an estimate, and it should be referred to that way. والعدد الدقيق يمكن أن يكون تقدير، وكان ينبغي أن تحال إلى هذا الطريق.

شرح / برسايز يعني في رنج معين مثلا بين 19,20,21يعني نقول عنها برسايز بس مش يعني أكورس دقيقه لا تقديريه فقط.

Deliberate Distortion تشويه متعمد

Some studies or surveys are distorted on purpose. The distortion can occur within the context of the data, the source of the data, the sampling method, or the conclusions.

يتم تشويه بعض الدراسات أو الاستقصاءات عن قصد. يمكن أن يحدث تشوه في إطار البيانات، ومصدر البيانات، وطريقة أخذ العينات، أو استنتاجات.

Recap خلاصة

In this section we have: في هذا القسم لدينا:

- **Reviewed misuses of statistics.** إساءة استعرض الإحصاء
- **Illustrated how common sense can play a big role in interpreting data and statistics.** يتضح المعنى كيف مشترك يمكن أن تلعب دورا كبيرا في تفسير البيانات والإحصاءات

Section 1-5
Collecting Sample Data

Key Concept المفهوم الرئيسي

- **If sample data are not collected in an appropriate way, the data may be so completely useless that no amount of statistical torturing can salvage them.** إذا لم يتم جمع بيانات العينة بطريقة مناسبة، قد تكون البيانات حتى عديمة الفائدة تماما أن أي قدر من تعذيب الإحصائي يمكن إنقاذ لهم.
- **Method used to collect sample data influences the quality of the statistical analysis..** الطريقة المستخدمة لجمع البيانات عينة يؤثر على نوعية التحليل الإحصائي
- **Of particular importance is *simple random sample*.** أهمية خاصة هو عينة عشوائية بسيطة.

Basics of Collecting Data أساسيات جمع البيانات

Statistical methods are driven by the data that we collect. We typically obtain data from two distinct sources: *observational studies* and *experiment*. هي التي تحرك الأساليب الإحصائية التي كتبها البيانات التي نجمعها. ونحن عادة الحصول على بيانات من مصدرين مختلفين: الدراسات الرصدية والتجربة.

Observational Study دراسة مبنية على الملاحظة

- **Observational study** دراسة مبنية على المشاهدة

observing and measuring specific characteristics without attempting to modify the subjects being studied.

مراقبة وقياس خصائص محددة دون محاولة لتعديل الموضوعات التي تجري دراستها

شرح / نجمع الداتا عن طريق الملاحظات فقط ولا نعمل أي تعديل على الداتا اللي شغالين عليها .

Experiment تجربة

❖ Experiment تجربة

apply some treatment and then observe its effects on the subjects; (subjects in experiments are called experimental units).

تطبيق بعض العلاج ومن ثم مراقبة آثاره على المواضيع؛ (وتسمى المواد في التجارب وحدات تجريبية)

شرح / هنا عكس الملاحظات قبل هنا نغير او نجرب ونشوف ايش راح يحصل مع كل تغيير .
ملاظه مهمه/ ممكن يجي سؤال بالاختبار عن الفرق بين الملاحظة و التجربة .

Random Sample عينة عشوائية

❖ Simple Random Sample عينة عشوائية بسيطة

of *n* subjects selected in such a way that every possible sample of the same size *n* has the same chance of being chosen.

من مواضيع ن مختارة في مثل هذه الطريقة أن كل عينة ممكنة من نفس الحجم ن لديه نفس الفرصة من اختياره.

شرح / نختار العينه أي واحد بطريقة عشوائية كل واحد بيكون عنده نفس فرص احتمالية اختياره مثل غيره.

Random & Probability Samples عشوائية

❖ Random Sample عينة عشوائية

members from the population are selected in such a way that each individual member in the population has an equal chance of being selected.

ويتم اختيار أعضاء من السكان في مثل هذه الطريقة أن كل فرد في السكان ديه فرصة متساوية لاختياره

شرح / نختار العينه أي واحد بطريقة عشوائية كل واحد بيكون عنده نفس فرص احتمالية اختياره مثل غيره.

❖ Probability Sample احتمال عينة

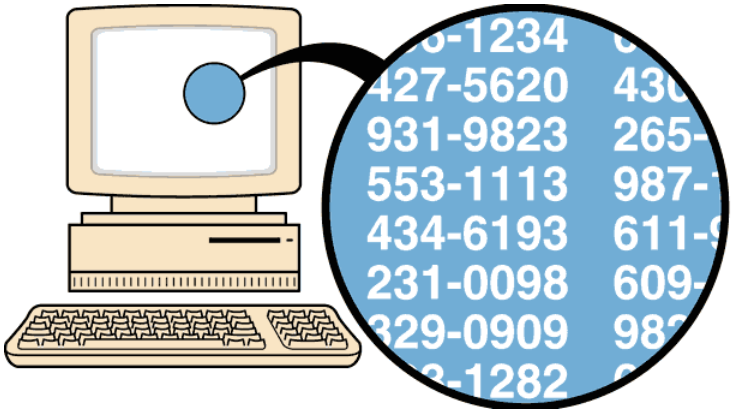
selecting members from a population in such a way that each member of the population has a known (but not necessarily the same) chance of being selected.

اختيار أعضاء من السكان في مثل هذه الطريقة أن كل عضو من السكان لديه معروفة (ولكن ليس بالضرورة نفس)

شرح / كل عينه عندها فرصة مختلفة في اختياره عن الثاني حسب الطريقة الي ماشي عليها باختيار العينات .

Random Sampling أخذ العينات العشوائية

selection so that each individual member has an equal chance of being selected. التحديد بحيث كل فرد لديه فرصة متساوية لاختياره



شرح / نختار العينه أي واحد بطريقة عشوائية كل واحد بيكون عنده نفس فرص احتمالية اختياره مثل غيره.

Systematic Sampling أخذ العينات المنتظمة

Select some starting point and then select every *k*th element in the population. في عدد السكان KTHحدد بعض نقطة الانطلاق ثم حدد كل عنصر

شرح / هذي طريقة من البروبريتيال سيمبل السابقه مثلا ابدأ بنقطة رقم 2 وامشي بمضاعفتها 4ثم 8 وهكذا طريقة منظمه.

Convenience Sampling الراحة أخذ العينات

use results that are easy to get النتائج الاستخدام التي هي سهلة للحصول على

شرح / هي النتائج التي ستطيع بسهولة الحصول عليها بدون حسابات او أي شيء.
هذي مالها علاقة بالطرق السابقه طريقة اضافية.

Stratified Sampling اخذ العينات الطبقية

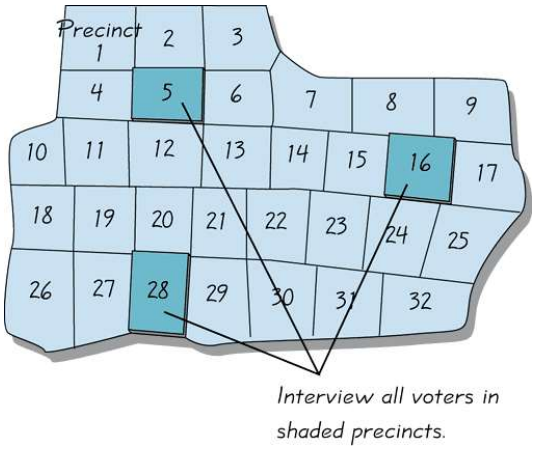
subdivide the population into at least two different subgroups that share the same characteristics, then draw a sample from each subgroup (or stratum). تقسيم السكان إلى اثنين على الأقل من المجموعات الفرعية المختلفة التي تشترك في نفس الخصائص، ثم رسم عينة من كل مجموعة فرعية (أو طبقة)

شرح / تقسم المجموعة الى قروبين و كل قروب يكون له نفس الخصائص و نختار العينات من كل قروب تكون العينه تمثل القروب بخصائصه.

Cluster Sampling أخذ العينات العنقودية

divide the population area into sections (or clusters); randomly select some of those clusters; choose all members from selected clusters.

تقسيم مجال السكان إلى أقسام (أو مجموعات)؛ اختيار عشوائي بعض من تلك المجموعات. اختيار كل أعضاء من مجموعات مختارة.



Interview all voters in shaded precincts.

شرح / كل عينة نقسمها الى اقسام ونحدد بشكل عشوائي أي قسم من الاقسام و ناخذ القسم وندرس كل الاعضاء اللي فيه .
مثلا/ بلد نقسمها الى مناطق وبعدين نختار منطقة بشكل عشوائي و ندرس السكان اللي فيها .

Multistage Sampling أخذ العينات متعددة المراحل

Collect data by using some combination of the basic sampling methods. جمع البيانات باستخدام مزيج من أساليب أخذ العينات الأساسية

In a multistage sample design, pollsters select a sample in different stages, and each stage might use different methods of sampling.

في تصميم عينة متعددة المراحل، استطلاعات الرأي حدد عينة في مراحل مختلفة، ولكل مرحلة قد تستخدم أساليب مختلفة لأخذ العينات.

شرح /اختيار العينه يمر بعدة مراحلا مثلاً نختار عينة بطريقة عشوائية بأول مرحلة وبالمرحلة الثانية نستخدم طريقة اخره لاكمال دراسة العينة مثل العنقودية او المنتظمة لما نخلص الدراسة نكون استخدمنا اكثر من طريقة بدراسة واحدة.

Methods of Sampling – Summary
طرق أخذ العينات - ملخص

- **Random** عشوائية
- **Systematic** منهجي
- **Convenience** السهولة أو الراحة
- **Stratified** الطبقية
- **Cluster** العنقودية
- **Multistage** متعدد المراحل

شرح/ اختصار لكل الطرق السابقه تعدادها

Beyond the Basics of Collecting Data
ما وراء أساسيات جمع البيانات

Different types of observational studies and experiment design.

أنواع مختلفة من الدراسات الرصدية وتصميم التجربة

Types of Studies أنواع الدراسات

❖ **Cross sectional study دراسة مقطع عرضي**

data are observed, measured, and collected at one point in time.

ويلاحظ البيانات، تقاس، والتي تم جمعها في نقطة واحدة في الوقت المناسب

شرح/ نجمع بيانات من الوقت الحالي تكون في نفس الوقت لا بالماضي و لا بالمستقبل .

❖ **Retrospective (or case control) study دراسة (أو مراقبة الحالة) بأثر رجعي**

data are collected from the past by going back in time (examine records, interviews, …) يتم جمع البيانات من الماضي من خلال العودة في الوقت المناسب (فحص السجلات والمقابلات، ...)

شرح/ نجمع بيانات من الماضي .

❖ **Prospective (or longitudinal or cohort) study دراسة (أو طولية أو الجماعة)**

data are collected in the future from groups sharing common factors (called cohorts)

يتم جمع البيانات في المستقبل من الجماعات تقاسم العوامل المشتركة (وتسمى الأفواج)

شرح/ نجمع بيانات من المستقبل .

Randomization التوزيع العشوائي

❖ **Randomization التوزيع العشوائي**

is used when subjects are assigned to different groups through a process of random selection. The logic is to use chance as a way to create two groups that are similar. يستخدم عندما يتم تعيين موضوعات لمجموعات مختلفة من خلال عملية الاختيار العشوائي. المنطق هو استخدام فرصة كوسيلة لإنشاء مجموعتين متشابهة.

شرح/ هي الاختيار العشوائي و لك عضو له نفس فرصة الاختيار.

Replication تكرار

❖ **Replication تكرار**

is the repetition of an experiment on more than one subject. Samples should be large enough so that the erratic behavior that is characteristic of very small samples will not disguise the true effects of different treatments. It is used effectively when there are enough subjects to recognize the differences from different treatments.

هو تكرار تجربة على أكثر من موضوع واحد. يجب أن تكون العينات كبيرة بما فيه الكفاية بحيث فإن سلوك خاطئ الذي هو سمة من عينات صغيرة جدا لا تخفي آثار حقيقية من العلاجات المختلفة. يتم استخدامه بشكل فعال عندما تكون هناك موضوعات ما يكفي للاعتراف الخلافات من العلاجات المختلفة.

Use a sample size that is large enough to let us see the true nature of any effects, and obtain the sample using an appropriate method, such as one based on *randomness*. استخدام حجم العينة كبير بما فيه الكفاية للسماح لنا رؤية الطبيعة الحقيقية لأية آثار، والحصول على عينة باستخدام الطريقة المناسبة، مثل واحد على أساس العشوائية.

شرح/ باختصار هي عينات و نمشيهم كلهم على نفس التجربة و المهم فيها نستخدم عدد كبير او حجم كبير من العينات عشان نسمح بتكرارها عليهم كذا مره على نفس التجربة .

Blinding (مأخوذه من أعمى) تعمية

❖ Blinding تعمية

is a technique in which the subject doesn't know whether he or she is receiving a treatment or a placebo. Blinding allows us to determine whether the treatment effect is significantly different from a placebo effect, which occurs when an untreated subject reports improvement in symptoms. هي تقنية في هذا الموضوع الذي لا يعرف ما إذا كان هو أو هي تلقي العلاج أو دواء وهميا. المسببة للعمى يسمح لنا لتحديد ما إذا كان تأثير العلاج يختلف كثيرا عن تأثير العلاج الوهمي، والذي يحدث عندما تقارير موضوعا غير المعالجة تحسن في الأعراض.

شرح/ مثلا في الدراسات الطبيه نستخدم دواء اسمه البلاسيبو ايفكت دواء وهمي يكون ماده كيميائية مافيها أي تأثير بالجسم سلبي و لا ايجابي. عشان لا يتأثر نفسيا وتخترب نتائج الدراسة.

Double Blind المزدوج للمكفوفين

❖ Double-Blind المزدوج للمكفوفين

Blinding occurs at two levels: يحدث المسببة للعمى على مستويين:

(1) The subject doesn't know whether he or she is receiving the treatment or a placebo. لا موضوع يعرفون ما إذا كان هو أو هي تلقي العلاج أو دواء وهميا

(2) The experimenter does not know whether he or she is administering the treatment or placebo.

لا المجرب يعرف ما إذا كان هو أو هي إدارة العلاج أو الدواء الوهمي.

شرح/ الشخصيين اللي بنسوي عليه التجربة و اللي بيقوم بالتجربة الاثنين بيكونوا مش عارفين هل هذا دواء وهمي او لا.

Confounding مربك

❖ Confounding مربك

occurs in an experiment when the experimenter is not able to distinguish between the effects of different factors.

يحدث في تجربة عندما المجرب ليس قادرا على التمييز بين آثار العوامل المختلفة.

Try to plan the experiment so that confounding does not occur. محاولة لتخطيط التجربة حتى لا يحدث هذا الخلط.

شرح/ هو شي مش كويس لان الشخص اللي بيسوي التجربة يكون مش عارف يفرق ما بين الحاجات يسوي عليها التجربة و اثارها.
لازم يكون عارف الادوات و الدراسات و اثر كل اداة و تحليلها , لذا يتطلب نسوي الدراسه و التحليل حقنا بسيستيم او بطريقة معينه و واضحه عشان نفهم النتائج والاثار من أي اداة كانت .

Controlling Effects of Variables السيطرة على آثار المتغيرات

- **Completely Randomized Experimental Design** تصميم التجارب العشوائية تماما
  assign subjects to different treatment groups through a process of random selection..تعيين موضوعات لمجموعات العلاج المختلفة من خلال عملية الاختيار العشوائي

شرح/ تصميم يستخدم عمليات عشوائية للاختيار.

- **Randomized Block Design** العشوائية تصميم بلوك
  a block is a group of subjects that are similar, but blocks differ in ways that might affect the outcome of the experiment.

كتلة عبارة عن مجموعة من المواضيع التي هي مماثلة، ولكن الكتل تختلف في الطرق التي قد تؤثر على نتائج التجربة

شرح/ تقسيم المواضيع الى بلوك.

- **Rigorously Controlled Design** تصميم خاضعة لرقابة صارمة
  carefully assign subjects to different treatment groups, so that those given each treatment are similar in ways that are important to the experiment. تعيين بعناية موضوعات لمجموعات العلاج المختلفة، بحيث أن يعطى كل معاملة متشابهة في الطرق التي تعتبر مهمة للتجربة

شرح/ نشتغل فيه بحذر شوية.

- **Matched Pairs Design** يقابل تصميم أزواج
  compare exactly two treatment groups using subjects matched in pairs that are somehow related or have similar characteristics. مقارنة بالضبط مجموعتين العلاج باستخدام مواد مطابقة في أزواج التي ترتبط بطريقة أو بأخرى أو لها خصائص مماثلة

شرح/ تقسم المواضيع الى قطعتين متشابهتين بالخصائص و نبدأ نشتغل على كل قطعة.

ملاحظه / الجمل اللي تميز كل تعريف مظلله بالازرق وهذه التعريفات 4 تجي بالصح و الخطأ و الاختياراي ما عم ها جات عرفي

Summary ملخص

Three very important considerations in the design of experiments are the following: ثلاثة اعتبارات هامة جدا في تصميم التجارب هي ما يلي:

1. **Use *randomization* to assign subjects to different groups.**
   1. استخدام العشوائي لتعيين موضوعات لمجموعات مختلفة.

2. **Use replication by repeating the experiment on enough subjects so that effects of treatment or other factors can be clearly seen.**
   2. استخدام تكرار بتكرار التجربة على المواضيع بما فيه الكفاية بحيث آثار العلاج أو عوامل أخرى يمكن رؤيتها بوضوح.

3. ***Control the effects of variables* by using such techniques as blinding and a completely randomized experimental design.**
   3. السيطرة على آثار المتغيرات باستخدام تقنيات مثل عمى وتصميم التجارب العشوائية تماما

Errors أخطاء

No matter how well you plan and execute the sample collection process, there is likely to be some error in the results.

مهما كان جيدا على تخطيط وتنفيذ عملية جمع العينات، فمن المرجح أن يكون هناك بعض الخطأ في النتائج.

- **Sampling error خطأ المعاينه**

  the difference between a sample result and the true population result; such an error results from chance sample fluctuations.

  الفرق بين نتيجة العينة ونتيجة سكان الحقيقية. هذه النتائج خطأ من تقلبات عينة فرصة
  شرح/ خطأناتج من العينه نفسها .

- **Non-sampling error خطأ غير أخذ العينات**

  sample data incorrectly collected, recorded, or analyzed (such as by selecting a biased sample, using a defective instrument, or copying the data incorrectly).

  بيانات العينة التي تم جمعها بشكل غير صحيح، مسجلة، أو تحليلها (مثل اختيار عينة متحيزة، وذلك باستخدام أداة معيبة، أو نسخ البيانات بشكل غير صحيح)

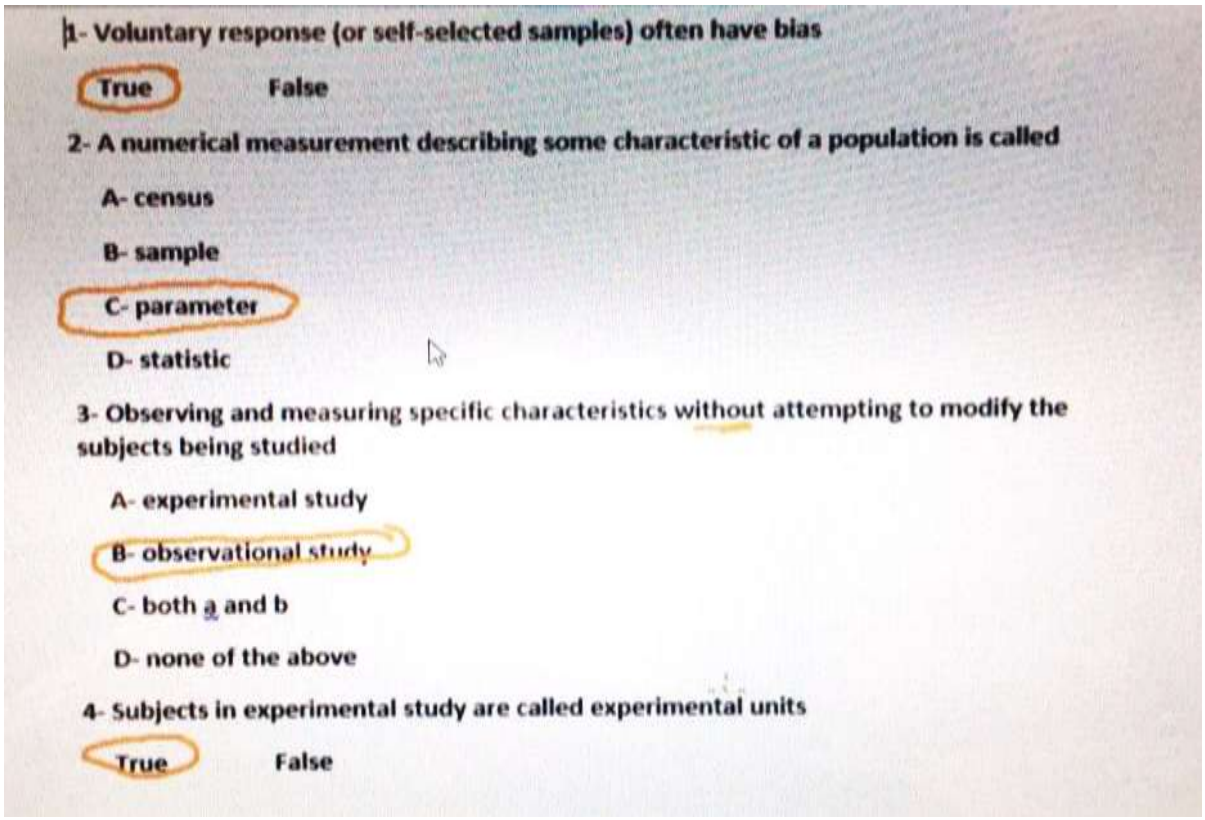
  شرح/ الأخطاء الناتجه عن الاشياء غير العينه. كجمعنا الداتا من ناس غير محايده او محايده او لها مصلحه او استخدام بيانات معيوبه او حصل خطأ اثناء النسخ او فيها مشاكل.

Recap خلاصة

In this section we have looked at: في هذا القسم لقد بحثنا في:

- **Types of studies and experiments** أنواع من الدراسات والتجارب
- **Controlling the effects of variables.** السيطرة على آثار المتغيرات
- **Randomization.** التوزيع العشوائي
- **Types of sampling.** أنواع العينات
- **Sampling errors.** أخطاء المعاينة

اختبار خاص بالجزئية السابقة Quiz



1- Voluntary response (or self-selected samples) often have bias

True False

2- A numerical measurement describing some characteristic of a population is called

A- census

B- sample

C- parameter

D- statistic

3- Observing and measuring specific characteristics without attempting to modify the subjects being studied

A- experimental study

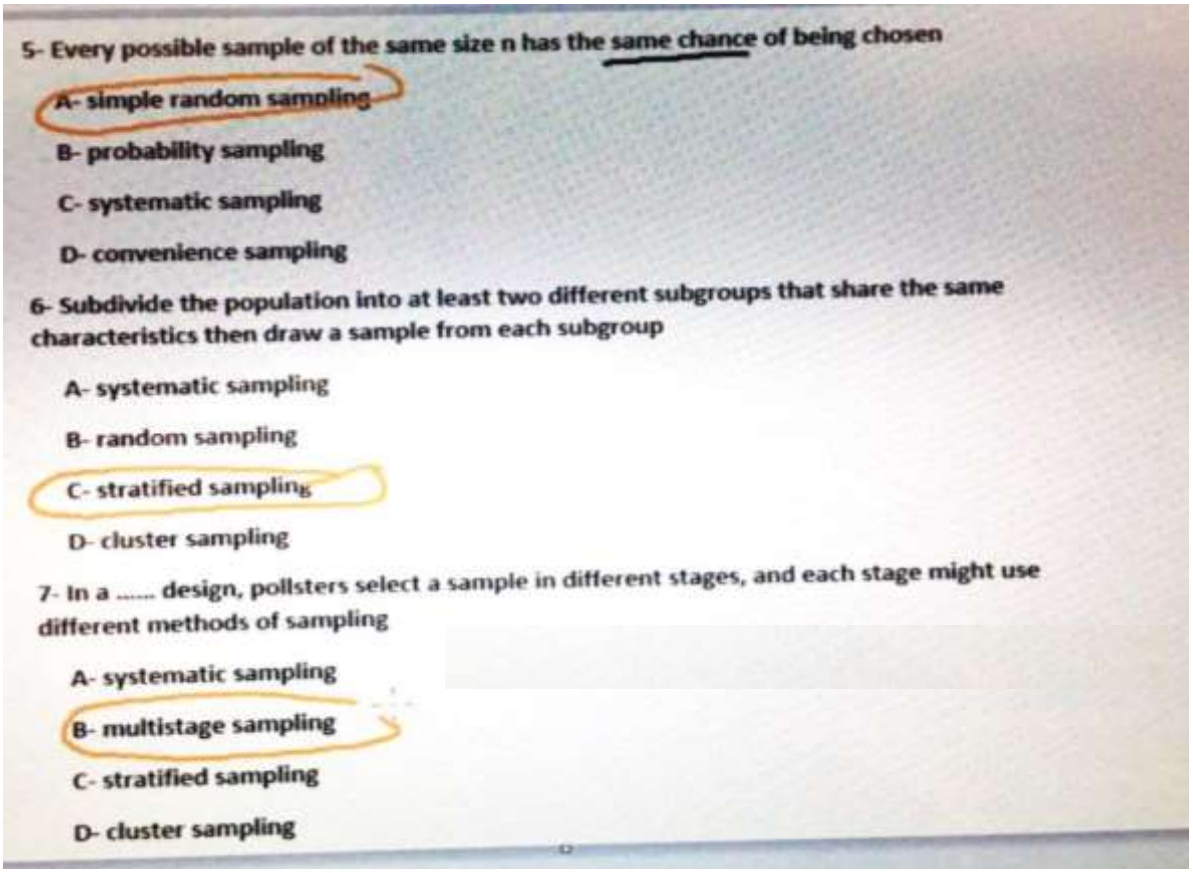
B- observational study

C- both a and b

D- none of the above

4- Subjects in experimental study are called experimental units

True False



5- Every possible sample of the same size n has the same chance of being chosen

A- simple random sampling

B- probability sampling

C- systematic sampling

D- convenience sampling

6- Subdivide the population into at least two different subgroups that share the same characteristics then draw a sample from each subgroup

A- systematic sampling

B- random sampling

C- stratified sampling

D- cluster sampling

7- In a ..... design, pollsters select a sample in different stages, and each stage might use different methods of sampling

A- systematic sampling

B- multistage sampling

C- stratified sampling

D- cluster sampling

8- Methods of sampling are:

A- observational and experiment

B- random and systematic

C- simple random sampling

D- random/ systematic/ convenience/ stratified/ cluster/ multistage

9- Types of observational study are:

A- cross sectional study

B- retrospective (or case control) study

C- prospective (or longitudinal) study

D- all of the above

10- Data which are collected from the past by going back in time (examine records, interviews):

A- cross sectional study

B- retrospective (or case control) study

C- prospective (or longitudinal) study

D- none of the above

11- In cross sectional study, data are observed, measured, and collected at one point in time

True False

True False

12- Examples of experiment designs are:

A- randomization

B- replication

C- blinding

D- all of the above

13- non-sampling error is the difference between a sample result and the true population result; such an error results from chance sample fluctuations

True False