

## \* Inferential statistics:-

① باستخدام بيانات العينة يمكننا استخلاص استنتاجات أو عمل تنبؤات حول السكان

probability

② يتم استخدام نظرية الاحتمالات لتقدير واختبار نسبة صحة الاستنتاجات (Confidence Intervals & Margins of Errors)

## \* There are two types of Data:-

Numbers ← Quantitative	بيانات كمية	Words ← Qualitative	بيانات وصفية
تتضمن أرقاماً حقيقية فقط	يشتمل أي عدد	تتضمن أي شيء غير عددي	يشتمل أي شيء غير عددي
0, 1, 2, 3	8.23 3.14	الألوان (أحمر - أزرق - ...)	الألوان (أحمر - أزرق - ...)
1 2 3 4	1 2 3 4	تقديرات (ممتاز - جيد جداً - ...)	تقديرات (ممتاز - جيد جداً - ...)
8 cats ✓ 3.5 cats ✗	5 m 7.9 Kg	صفات (Soft/hard) (Low/High)	صفات (Soft/hard) (Low/High)
Counted أي شيء يُعد	Measured أي شيء يُقاس		
	مسافة - سرعة - وزن		

## \* Scales of Measurements

### ① Nominal Scale Data

	الترتيب	الردود	النسبة
1. Qualitative / Categorical	↓	↓	↓
2. Names, Colors, Labels, etc. $\Rightarrow$ مثالاً		5	50%
		4	40%
3. Order doesn't matter		1	10%

لا يمكن استخدامه في الحسابات لكن يمكن استخدامه في التفسير

### ② Ordinal Scale Data

	الترتيب
1. Ranking / Placement	1 <sup>st</sup> 4:53
2. The Order does matter $\Rightarrow$	2 <sup>nd</sup> 4:54
3. Differences cannot be measured	3 <sup>rd</sup> 5:01

لا يمكن حساب الفرق لكن يمكن حساب فرق الترتيب بينهم

### ③ Interval Scale Data

	temp
1. The order matters	30°C
2. Differences can be measured (except ratios) $\Rightarrow$	36°C $\leftarrow +6^{\circ}\text{C}$
3. No true "0" starting point	39°C $\leftarrow +4^{\circ}\text{C}$

$$\frac{36}{30} = 1.2$$

لا يمكن قول أن درجة الحرارة 36°C أحر من 30°C بمقدار 1.2 الضعف  
الصفر ليس نقطة بداية لوجود درجات حرارة تحت الصفر (-30°C, -45°C)  
الترتيب مهم بحيث (39°C > 36°C > 30°C)



4

## Ratio Scale Ratio

Grades

1. The Order matters  $\Rightarrow 70, 30, 56, 82, 90, 0$
2. Differences are measurable  $\Rightarrow 0, 30, 56, 70, 82, 90$   
(including ratios)
3. Contains "0" starting point

الترتيب مهم لأن  $90 > 70$

يمكن حساب الفرق بينهم  $(56 - 30 = 26)$   $(70 - 56 = 14)$

يمكن حساب النسب ratios  $(\frac{90}{30} = 3)$  بحيث أن 90 ثلاث أضعاف 30

والهم مفر لهم نقطة بداية بحيث إذا لم يجيب طالب على أي سؤال فالنتيجة 0

ولا يوجد درجات أقل من الصفر "إلا في الهندسة"

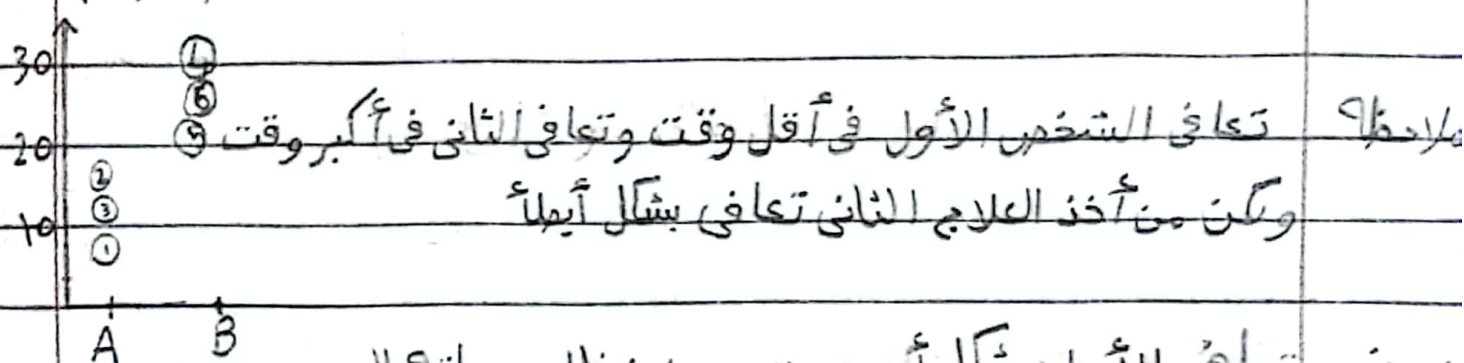
Data	Nominal	Ordinal	Interval	Ratio
Labeled	✓	✓	✓	✓
Meaningful Order	X	✓	✓	✓
Measurable Difference	X	X	✓	✓
True Zero Starting point	X	X	X	✓

## \* Hypothesis Testing (اختبار الفروض)

← هو اختبار الفروض والتحقق من الأفكار عن العالم  
مثلاً: مصنع ( زجاجات مياه معدنية - أدوية - زيت سيارات )  
تطبيقاً مصنع أدوية.

عند وجود نوعين من دواء لعلاج نفس المرض، يتم تجربة كل منهما  
على ٥ يبتاً مكوناً من ثلاث أشخاص ونقيس الوقت المطلوب  
لتعافي الثلاث أشخاص

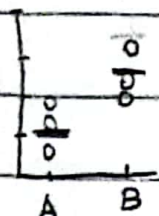
Time to Recover  
(in hours)



فرومن تعافى الأول بشكل أسرع بسبب نخلام حياته الصبي  
أما الثاني فظروفه أسوء قليلاً أما الثالث فظروفه المصيبة هي الأسوء

يستحتاج مع أن كل الأشخاص أخذوا نفس العلاج وعند لهم نفس المرض،  
لكن لم يتعافوا في نفس المدة بسبب أمور خارجية مثل التمارين،  
ضغط العمل، الصحة وأمر أخرى لا يمكن التحكم بها  
والعلاج A أسرع من B

افتراض الأشخاص الذين أخذوا العلاج A تعافوا في وقت أقل من الذين أخذوا B  
وعند حساب المتوسط (mean/average) للدواء A و B نجد فرق  
15 ساعة بينهم





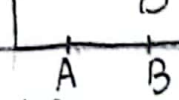
← والآن يوجد افتراض يمكن اختباره للتأكد من صحتها



← عند إعارة التجربة وجدنا أن الذين أخذوا الدواء A



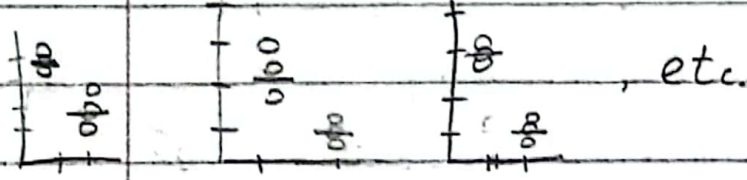
يحتاجوا 35 ساعة إضافية ليتعافوا وعند المقارنة



ولهذا عكس التجربة الأولى (وربما كان بسبب الظروف

المحيطة بعيننا الدواء A (أو) بسبب الظروف المثالية المحيطة بعيننا الدواء B (أو)

ربما لأننا تم عكس اسم الدواءين في أول تجربة) لذا أعدنا التجربة عدة مرات وكلهم عكس أول تجربة



← بعد إعارة التجربة عدة مرات تم نفي أول افتراض ووضع فرض جديد

← عند حدوث فرق في حساب المدة المتطلبة للتعافي بعد استخدام واثين تأخذ متوسط

الفرق لعدة تجارب (13، 12.1، 12، 13.5)

← اختلار النتيجة لا يعني أن الفرض خطأ

\* Null Hypothesis :- فرض حسابي يفترض أنه لا يوجد علاقة أو فرق بين عينتين أو متغيرين

إذا وجد فرق حتمي بين تجربتين لهوائين مع وجود عوامل خارجية قد تؤثر على الناتج أي أن الدواء الأول A كان أفضل من B بمقدار نصف ساعة وفي الثانية B أفضل من A بمقدار ربع ساعة. نستخدم ال Null Hypothesis لعدم اكتشاف الفرق بين نتائج التجارب وعدم وجود دليل يثبت فرض أن لا يوجد فرق بين الدواءين

إذا أعدنا التجارب عدة مرات ووجد فرق في النتيجة ولن يؤثر عليها ظروف خارجية بشكل ملحوظ فيتم رفض ال Null Hypothesis

الخلاصة :-

بدل من إعادة التجربة عدة مرات وفرض عدة فروض أن الدواء الأول مختلف عن الثاني ، نستخدم Null Hypothesis لتأكيد إن وجد فرق

لأننا وجدنا أن نتيجة A أفضل دائماً من B نفس وجود عوامل قد تؤثر على اختلاف الناتج ونرفض ال Null Hypothesis ونتيقن من وجود فرق بينهم

إذا غيرت العوامل الخارجية أفضلية الناتج بين A و B إذا لا يوجد فرق وهذا هو ال Null Hypothesis

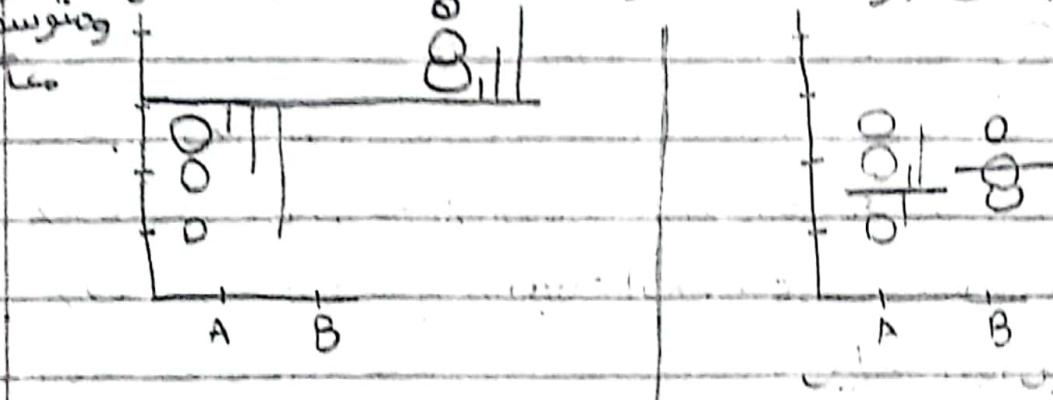


## \* Alternative Hypothesis:

← هي مآس ال Null Hypothesis

← هي فرض أن متغير (شخص) واحد قد يغير من الناتج

لحسابه في تجربة من مجموعتين: نحدد المتوسط لكل من التجريبتين ومتوسط التجريبتين معاً



يمثل ال Alternative Hypothesis يمثل ال Null Hypothesis

← إذا كانت المسافات بين كل شخص والمتوسط الواحد أقل من المسافات بين

كل متوسط: إذا فبذل استخدم متوسط واحد مما يعني أننا <sup>Null Hypothesis</sup>

وإذا كان العكس فنستخدم متوسطين ويكون ال Alternative Hypothesis

وأن الفرق بين المتوسطين مجرد ظروف لا يمكن التحكم بها

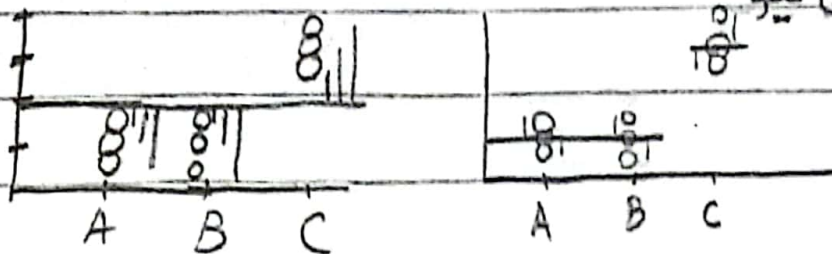
عند وجود ٣ مجموعات أو أكثر: يوجد اختيارات لل Alternative Hypothesis بحيث

١- لا يوجد فرق بين A و B و C ← يوجد فرق بينهم

٢- لا يوجد فرق بين A و B لكن يوجد فرق بينهم وبين C

← نقوم بنفس الخطوات ونرفض أو لا نرفض ال Null Hypothesis فقط

لأننا يوجد احتمالات كثيرة



\* P-values : (Probability value) وهو رقم بين صفر وواحد

التي احتمالية الحصول على النتائج المرغوبة (أو أكبر) من خلال تجارب  
بفرق عدم وجود اختلاف بينهم

مثال: إذا تم إعطاء عينة دواء وهمي وعينة أخرى دواء تجريبي  
والدواء هو حبوب أخضر للوزن وفي نهاية التجربة نجد أن عينة الدواء  
الوهمي لم تزل في الوزن أما عينة الدواء التجريبي نزلوا كيلو في الوزن

الـ Null Hypothesis هو عدم وجود فرق بين الدوائين

الـ P-value هي احتمالية وجود فرق كيلو أو أكثر

كلما قلَّت الـ P-value زادت الأدلة ضد الـ Null Hypothesis

قيمة الـ P-value تقلب بما يقتضيه 0.02 أو 2%

Two population

إذا كانت الـ Null Hypothesis (فرضية) متطابقتين (لا يمكن بنسبة 2%

ملاحظة اختلاف بنفس المقدار (أو أكبر) المرصود من العينة

إلا كان

الفرق بين الذين أخذوا الدواء الوهمي والتجريبي متساوي

بنسبة 2% يوجد نقصان وزن بقدر كيلو أو أكثر بين العينات

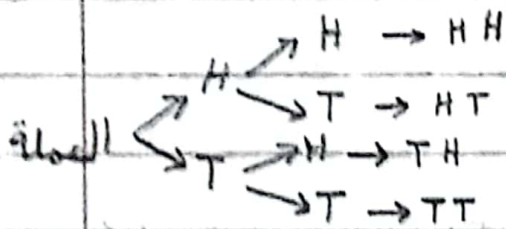
عند المقارنة بين دوائين لا يجد الـ P-value الفرق بينهم إن كان صغيراً  
أو كبيراً



\* How to calculate p-values:-

يوجد نوعان : One-Sided and Two-Sided وهنركز على Two-sided ومن التمارين استخدام الأخرى

H: Heads T: Tails



مثال احتمالية أن هذه العملة خاصة وعند رميها تسقط على H مرتين متتاليتين

$$\rightarrow P(H-H) = \frac{\text{عدد مرات الحصول على HH}}{\text{عدد التجارب}} = \frac{1}{4} = 0.25$$

$$\rightarrow P(T-T) = \frac{1}{4} = 0.25$$

$$\rightarrow P(H-T) = \frac{2}{4} = 0.5$$

الـ P-value تحتوي على 3 أجزاء :-

1- احتمالية حدوث الحدث العشوائي  $P(H-H) = 0.25$

2- احتمالية حدوث حدث آخر بنفس الندر  $P(T-T) = 0.25$

3- احتمالية حدوث حدث أكثر ندرة  $P( ) < P(H-H) = 0$

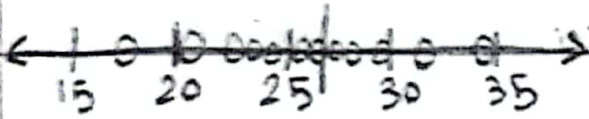
$$P = 0.25 + 0.25 + 0 = 0.5$$

إذا كان كانت قيمة الـ P-value أقل من 0.05 فنستبعد أنها أمانا فتنساوي 0.5 مما يعني أن هذه العملة مثل أي عملة

## Confidence Intervals:-

parameter

لدينا عالم عن المجتمع يتم أخذ عينه منه



مثال ١٠ تم وضع ١٢ وزن وقياسهم ونحسب ال mean



١١ تم أخذ عينه أخرى مكونة من ١٢ وزن ويسمح بالتكرار ونحسب المتوسط

95% Confidence Intervals



١٢ نعيد تكرار الخطوات مرة أخرى

لدينا 95% من المتوسطات

بإستخدام ال Confidence Interval نستطيع أن نرى أن الجوهر ليس

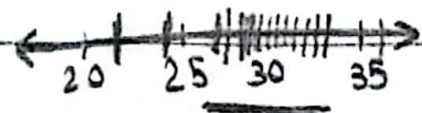
من أول مرة خارج ال Confidence Interval

mean بها أقل من 5% وبهذا ال p-value أقل من 0.05 وهذا غير

مرجع وبنا نقول أن أي قيمة باء ال Confidence Interval مختلف عن ما هو أقل من 5% متوسط

عند القيام بالخطوات مرة أخرى وتظهر هكذا فيوضع أن

لهذان المجتمعان مختلفان





## \* Regression analysis

تفاعل المتغيرات مع أسلوب يساعد على تحليل العلاقات بين المتغيرات

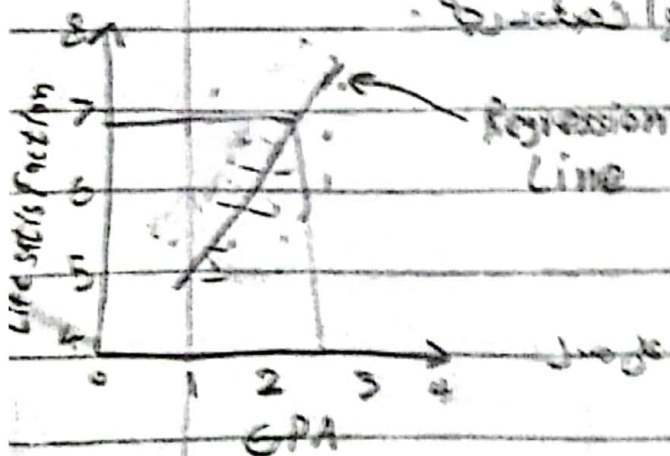
الـ variables التي عندهم تسمى متغيراً أو خاصية لكائن

يمكن أن تسمى مقياس يتم وضعهم في مجموعتين Discrete أو Continuous

هناك تطبيقات كثيرة في Regression analysis فمن ضمن الدراسات التحليلية  
يمكننا فهم كيفية ارتباط شئ أو خصائص الحياة المعيشية بالمتغير في ملاحظة التباين  
أو مستوى العمر

نار يكون كلاً من المتغيرين واحد من الأداة الأكاديمية الطالب ورضاهم عن حياتهم  
في منتصف العمر

بإستخدام Scattered Diagram يمكننا رسم درجة الرضا عن الحياة للأفراد  
مقابل متوسط تحصيلهم الأكاديمي و من درجة رضاهم إننا نحصل على  
متوسط درجات أعلى أو ما يحدث إذا قلوا نصف درجة



بإستخدام مقياس ذلك عن طريق الرسم البياني  
يتم رسم الرضا و متوسط الدرجات

و إذا حددوا أن من حصل على GPA 2.5 حصل على معدل  
رضا 6.8

الخط مجرد ليس دقيق

يمكن العمل ووضع بيانات أخرى على نفس الرسم