تقنيات الاستقصاء

مفاهيم أساسية في الإحصاء

الإحصاء

علم مكون من مجموعة من الأساليب والطرق الرياضية التي تستخدم في جمع وتنظيم وعرض وتحليل البيانات ثم الحصول على استدلالات واستنتاجات تستخدم في اتخاذ القرار.

يصنف إلى:

إحصاء وصفي يتناول طرق جمع، تنظيم، تلخيص، وعرض البيانات في صورة مبسطة.

إحصاء استدلالي ويهتم بطرق الوصول إلى نتائج أو توقعات عن المجتمعات المدروسة من خلال دراسة عينة من هذه المجتمعات

المجتمع

هو جميع العناصر المشتركة في الصفة التي تهم الباحث في دراسته

يقسم المجتمع إلى:

مجتمع محدود عدد محدود من الأشياء أو الأفراد

مجتمع غير محدود عدد من الأشياء غير منته

المجتمع المستهدف (المجتمع النظري)

يتمثل في جميع العناصر التي تحمل الخاصية التي تهم الباحث، بغض النظر عن إمكانية الوصول إليهم أم لا

المجتمع المتاح (مجتمع الدراسة)

جزء من المجتمع النظري الذي يمكن للباحث الوصول إليه فعليا، ويختار منه العينة

إطار المعاينة قائمة تحتوي على جميع عناصر المجتمع تستخدم في اختيار عينة للدراسة

العينة جزء من المجتمع يتم اختيارها بحيث تمثل جميع أفراد المجتمع تمثيلا جيدا

مثال إذا كنا نرغب في دراسة سلوك القراءة بين الطلبة الجامعيين في بلد ما:

المجتمع المستهدف :جميع الطلبة الجامعيين في ذلك البلد

مجتمع الدراسة :الطلبة في جامعة معينة يمكن الوصول إليهم، مثل جامعة بسكرة.

إطار المعاينة :قائمة الطلبة المسجلين في جامعة بسكرة، التي يمكننا استخدامها لاختيار عينة منهم لإجراء البحث.

المعاينة : هي الطريقة التي يتم بها اختيار عينة مناسبة تسمح بتحديد خصائص أو مواصفات معينة عن مجتمع الدراسة .

المجتمع المستهدف

Target population

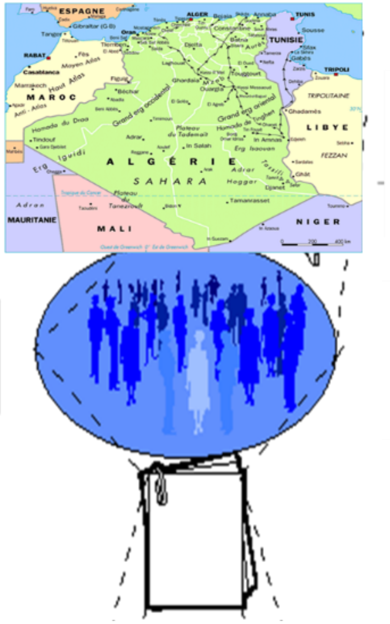
على من تريد التعميم

المجتمع المستهدف
Target population

على من تريد التعميم

المجتمع المتاح
Accessible
Population

ما هو المجتمع الذي يمكنك
الوصول اليه



المجتمع المستهدف
Target population

على من تريد التعميم

المجتمع المتاح
Accessible Population

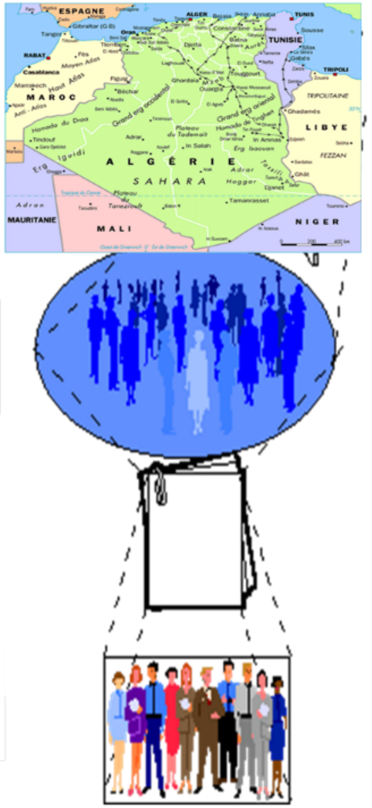
ما هو المجتمع الذي يمكنك الوصول اليه

اطار المعاينة
Sampling Frame

كيف يمكنك الوصول الى ذلك

المجتمع المستهدف
Target population

على من تريد التعميم



المجتمع المتاح
Accessible Population

ما هو المجتمع الذي يمكنك الوصول اليه

اطار المعاينة
Sampling Frame

كيف يمكنك الوصول الى ذلك

العينة
Sample

من سيكون في دراستك

العنصر هو عبارة عن عضو من أعضاء مجتمع الدراسة.

المفردة هي عبارة عن عضو من أعضاء العينة.

الظاهرة هي صفة لعناصر تختلف من عنصر لآخر في الشكل أو النوع أو الكمية

المعلمة قيمة رقمية (المتوسط الحسابي مثلا) لمتغير ما يتم الحصول عليها من دراسة شاملة للمجتمع. في الغالب تكون غير معروفة

الإحصاءة قيمة رقمية (المتوسط الحسابي مثلا) لمتغير ما يتم الحصول عليها من دراسة العينة المختارة.

الصفر المطلق(الحقيقي) هو الصفر الذي يعني الانعدام

الصفر الاعتباري(النسبي) يستخدم كنقطة بدء ، ولا يعني الانعدام

المتغير

- هو الصفة تحت الدراسة
- هو الشيء الذي يمكن أن يأخذ قيما مختلفة في الظروف المختلفة

أنواع المتغيرات

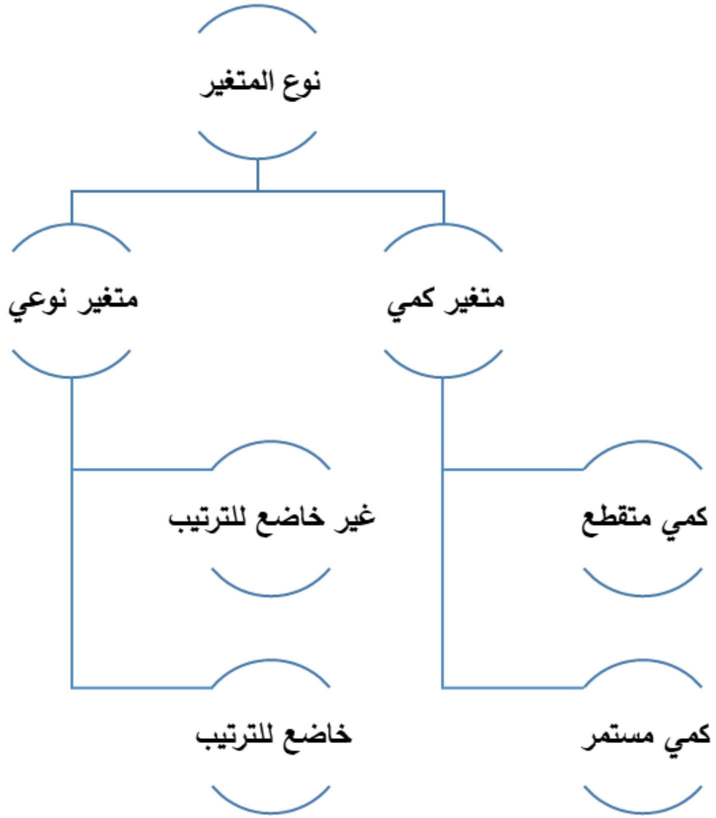
1 – متغيرات نوعية (كيفية) يعبر عنها في شكل صفات

- متغير نوعي خاضع للترتيب
- متغير نوعي غير خاضع للترتيب

2. متغيرات كمية (عددية) يعبر عنها في صورة عددية وتتقسم إلى:

متغير متقطع يأخذ أعداد صحيحة

متغير متصل يمكن أن يأخذ أي قيمة بين قيمتين صحيحتين



مقاييس النزعة المركزية

- عدد من المتوسطات للتعبير عن القيمة المتوسطة للمتغيرات
- تختلف باختلاف الغرض الذي تستخدم فيه، وطبيعة البيانات المحسوبة منها
- تصف الظاهرة المدروسة مثل الجداول الإحصائية، إلا أنها أكثر اختصارا وأكثر فائدة
- تمكننا من المقارنة بين مجموعة من القيم ومجموعة أخرى، أو بين ظاهرة وأخرى.
- أهم هذه المتوسطات وأكثرها استخداما: المتوسط الحسابي، الوسيط، والمنوال.

1- المتوسط الحسابي

$$\overline{X} = \frac{x_1 + x_2 + x_3 + \ldots\ldots + x_n}{n} = \frac{\sum_{i=1}^{n} Xi}{n}$$

المتوسط الحسابي في حالة بيانات مكررة

$$\overline{X} = \frac{n_1x_1 + n_2x_2 + n_3x_3 + \ldots\ldots\ldots\ldots n_nx_n}{n_1 + n_2 + n_3 + n_4 + \ldots\ldots\ldots\ldots\ldots n_n} = \frac{\sum_{i=1}^{n} n_ix_i}{\sum n_i}$$

المتوسط الحسابي في حالة توزيع تكراري

$$\overline{X} = \frac{\sum_{i=1}^{n} n_ix_i}{\sum_{i=1}^{n}}$$

حيث

xi = مراكز الفئات

ni = التكرارات

∑ni = مجموع التكرارات

ملاحظة تستخدم نفس العلاقة لحساب المتوسط الحسابي الموزون حيث

حيث

xi = القيم

ni = الأوزان

∑ni = مجموع الأوزان

الوسيط

- إذا كانت بيانات الظاهرة تحتوي على قيم متطرفة في الصغر أو الكبر فإن النتيجة التي يعطيها المتوسط الحسابي تكون غير واقعية

- الوسيط هو القيمة التي تقسم مجموع البيانات إلى قسمين بحيث يكون نصف عدد البيانات أكبر منه ونصف عدد البيانات أصغر

1– إذا كان عدد البيانات فردي فإن الوسيط هو القيمة التي ترتيبها $(n+1)/2$

2 – إذا كان عدد البيانات زوجي فإن الوسيط هو متوسط القيمتين اللتين ترتيبهما $n/2$ و $n/2+1$

مثال: البيانات التالية تمثل الدرجات التي تحصل عليها 10 طلبة في امتحان معين:
16، 17، 17، 15، 14، 16، 15، 13، 4، 3.

أوجد متوسط ووسيط درجات الطلبة

الحل: أ) متوسط الدرجات = 13

ب) الوسيط: = 15

نلاحظ أن المتوسط الحسابي لم ينصف أغلب الطلبة الذين تحصلوا على درجات أكبر بكثير من هذا المتوسط وأنحاز ناحية نتيجة الطالبين اللذين تحصلا على نتائج سيئة، في حين أن وسيط هذه الدرجات قسم نتائج الطلبة إلى قسمين بحيث نصف عدد الطلبة تحصلوا على درجات أعلى منه ونصف عدد الطلبة تحصلوا على درجة أقل منه وهو هنا أصدق من المتوسط الحسابي

الوسيط في حالة بيانات متكررة

$$M_e = L_1 + \frac{\frac{N}{2} - N_0}{n_e}.K$$

حيث: M_e = الوسيط.

L_1 = الحد الأدنى للفئة الوسيطية.

N = مجموع التكرارات.

N_0 = التكرار المتجمع الصاعد للفئة قبل الفئة الوسيطية.

n_e = تكرار الفئة الوسطية.

المنوال: هو القيمة الأكثر انتشارا

المنوال لبيانات مبوبة في جداول توزيع تكراري

$$M_0 = L_1 + \frac{d_1}{d_1 + d_2}.K$$

حيث

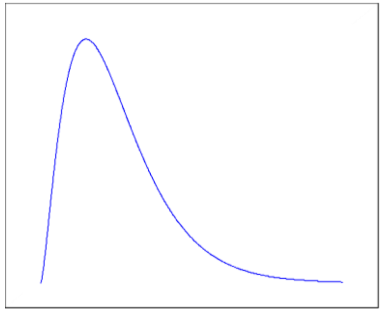
- ✓ d_1 = الفرق بين تكرار الفئة المنوالية وتكرار الفئة التي قبلها
- ✓ d_2 الفرق بين تكرار الفئة المنوالية و التي بعدها
- ✓ K طول الفئة المنوالية
- ✓ L_1 الحد الأدنى للفئة المنوالية

العلاقة بين المتوسط الحسابي والوسيط والمنوال:

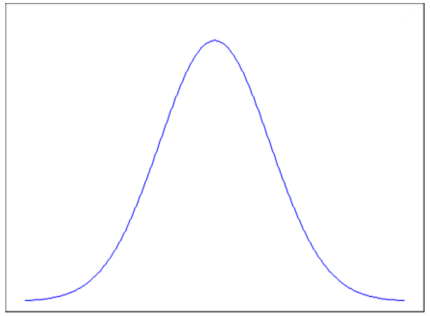
1 – تنطبق هذه المقاييس على بعضها وتتساوى في حالة التوزيع التكراري المعتدل وعندها يكون شكل منحنى التوزيع على شكل جرس.

2 – عندما يكون التوزيع التكراري المدروس غير متناظر من اليمن، أي عندما تكون البيانات الصغيرة كثيرة (أكثر تكرارا) تكون المقاييس بالشكل التالي $\overline{X} > M_e > M_0$ ويكون منحنى التوزيع ملتوي ناحية اليمين.

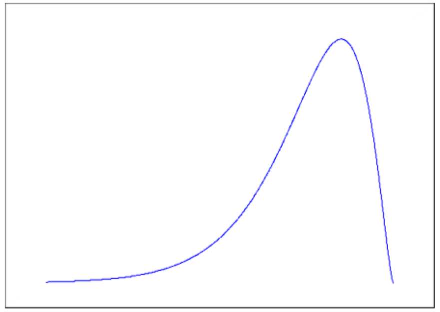
3 – عندما يكون التوزيع التكراري المدروس غير متناظر من اليسار، أي عندما تكون البيانات الكبيرة كثيرة فإن المقاييس تكون بالشكل التالي: $\overline{X} < M_e < M_0$



$\overline{X} > M_e > M_0$



$\overline{X} = M_e = M_0$



$\overline{X} < M_e < M_0$

يعتمد اختيار مقياس النزعة المركزية على نوع البيانات والسياق. المتوسط مفيد للبيانات المتوزعة بشكل طبيعي، بينما الوسيط يكون أفضل في حالة وجود قيم شاذة. المنوال يعتبر مفيدًا عند تحليل البيانات النوعية أو لتحديد القيم الأكثر شيوعًا.

مقاييس التشتت

لا تكفي مقاييس النزعة المركزية لوحدها لمعرفة الصفات الإحصائية اللازمة لوصف الظواهر، لأن الفروق بين قيم الظواهر قد تزداد أو تنقص رغم تساوي المتوسطات لهذه الظواهر

مثال

نفترض أن طالبين تحصلا على النتائج التالية في خمس مواد دراسية:

الطالب (X): 10، 11، 13، 14، 15.

الطالب (Y): 8، 9، 13، 15، 18.

- متوسط درجات الطالب (X) يساوي 12،6 وكذلك متوسط درجات الطالب (Y) يساوي 12،6

- وسيط درجات الطالب (X) يساوي 13 وكذلك وسيط درجات الطالب (Y) يساوي 13

قد يفهم مما سبق أن الطالبين (X)و (Y) لهما نفس المستوى غير أن التمعن الجيد في الدرجات التي تحصل عليها الطالبين تبين أن الطالب (X) ناجح في كل المواد المدروسة في حين أن الطالب (Y) ناجح في ثلاث مواد فقط.

مقاييس النزعة المركزية لا تعطي فكرة وافية عن اختلاف قيم الظواهر لذلك فإن مقاييس النزعة المركزية لا بد أن تكون مصحوبة بمقاييس أخرى لقياس مدى تباعد أو تقارب البيانات من بعضها البعض أو من متوسطها، تسمى هذه المقاييس بمقاييس التشتت.

يقاس تشتت البيانات بعدة مقاييس أهمها:

المدى (المطلق)

المدى= أكبر قيمة – أصغر قيمة.

المدى في التوزيعات التكرارية

المدى = مركز الفئة الأخيرة – مركز الفئة الأولى

= الحد الأعلى لفئة الأخيرة – الحد الأدنى للفئة الأولى

التباين

عبارة عن المتوسط الحسابي لمربعات الفروق بين قيم المتغير الإحصائي ومتوسطها الحسابي

$$V_X = \frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + (X_3 - \overline{X})^2 + \ldots\ldots.. + (X_n - \overline{X})^2}{n}$$

التباين لبيانات متكررة أو مبوبة

$$V_X = \frac{\sum_{i=1}^{n} ni(X_1 - \overline{X})^2}{\sum ni}$$

الانحراف المعياري

الجذر ألتربيعي لمتوسط مجموع مربع انحراف القيم عن متوسطها، أي أنه الجذر التربيعي للتباين.

الانحراف المعياري لبيانات مفردة

$$Sx = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n}}$$

الانحراف المعياري لبيانات متكررة أو مبوبة

$$Sx = \sqrt{\frac{\sum hi(X_i - \overline{X})^2}{\sum hi}}$$

الاستدلال الاحصائي

اختبار الفرضيات

اول خطوة في الدراسة الاستقصائية هي تحديد الهدف من الدراسة

تحديد الاشكالية

صياغة فرضية أو فرضيات البحث بناءا على الاشكالية

اختبار الفرضية أو الفرضيات

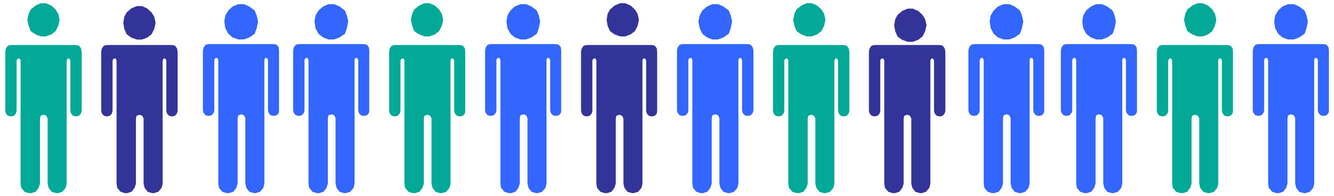
بناء على نتيجة الاختبار يكتب الباحث نتائج دراسته

اختبار الفرضيات يعتبر
من أهم الخطوات في
الدراسة الاستقصائية

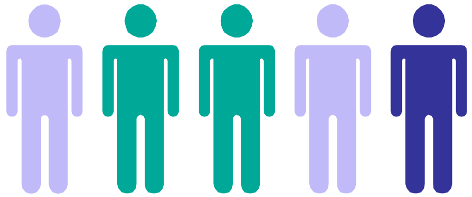
أغلب الدراسات الاستقصائية تتم على أساس عينة وليس على أساس المجتمع وذلك لصعوبة أو استحالة التعامل مع المجتمع ككل

وهذا ما يسمى بالاستدلال الإحصائي

المجتمع



عينة



عن طريق اختبار الفرضيات نستطيع ان نحدد هل الفرق بين الاحصائية المسحوبه من العينة وبين معلمة المجتمع فرقا يرجع الى الصدفة ام فرق حقيقي

بأسلوب احصائي هل هو فرق معنوي Significant او فرق غير معنوي؟

فما هي الفرضية؟ وكيف يتم اختبارها؟

الفرضية هي إجابة مؤقتة لسؤال بحثي تتم صياغتها في شكل علاقة بين متغيرين أو أكثر

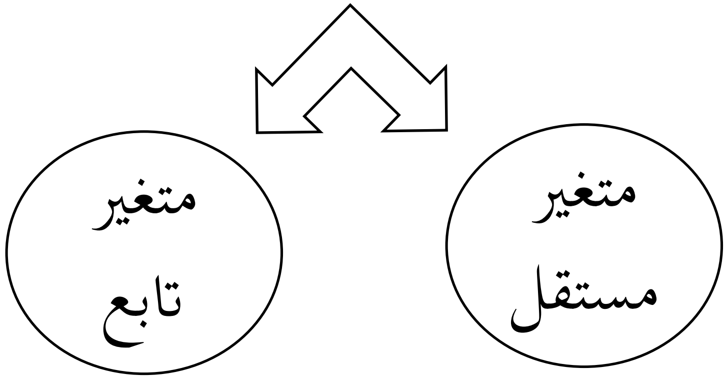
هناك أمرين يجب التأكيد عليهما:

- أن كتابة فرضية البحث تقوم على مجموعة **افتراضات** علمية قائمة على مجموعة من المعطيات والمعلومات المتاحة

- افتراضات البحث هي مسلمات البحث التي يجب التسليم بصحتها

- تصاغ الفرضية بغرض اختبارها، وبالتالي فهي **مرشحة للقبول أو للرفض**

مما تتكون الفرضية؟



المتغير الذي يرغب الباحث التعرف على أثره في متغير آخر.

هو الذي يتأثر بالمتغير المستقل، ويتغير بتغيره.

الفرضية هي بمثابة اقتراح عن معالم المجتمع موضوع الدراسة، والتي ما زالت غير معلومة للباحث، فهي حلول ممكنة لمشكلة البحث

الفرضية الصفرية (فرضية العدم) H_O (Null Hypothesis) :

تقوم على نفي وجود علاقة أو علاقات سببية بين متغيرين أو أكثر

أمثلة :

- حرق نفايات المستشفيات وسط المدن لا يسبب تلوث هواء المدينة.
- لا تؤدى الأخطاء الإدارية في مؤسسة معينة إلى اختلاف معدل العائد الفعلي عن العائد المتوقع من الاستثمار

H0

المتهم برئ

- تتعلق دائما بمعلمة المجتمع، وليس بإحصائية العينة

$H_0 : \mu = 30$ $\cancel{H_0 : \bar{X} = 30}$

الفرضية البديلة Alternative Hypothesis H_1 :

تشير إلى وجود علاقة (عكسية أو طردية) بين المتغيرات المدروسة

امثلة :

- وجود برامج تدريبية داخل المنظمة يؤثر على إنتاجية الافراد.
- المتدربون ذوى المؤهلات العلمية المنخفضة يحصلون على درجات أقل من المتدربين ذوى المؤهلات العلمية العالية في البرامج التدريبية

H1

المتهم مذنب

الفرضية البديلة هي الفرضية التي يحاول الباحث إثباتها

هي عكس الفرضية الصفرية على سبيل المثال

$H0: \mu = 30$ $H1: \mu \neq 30$

الفرضية البديلة لها ثلاث حالات

- ❖ $H1: \neq$
- ❖ $H1: >$
- ❖ $H1: <$

لا تحتوي أبدًا على العلامة "="، أو "≤"، أو "≤".

ما الذي يحدد شكل الفرض البديل ؟

هو مدى اقتناع الباحث بذلك أو مدى توفر المعلومات الأولية

اذا كانت الفرضية الصفرية: متوسط أعمار الطلبة يساوي 30 سنة

$$H0: \mu = 30$$

- إذا كانت وجهة نظر الباحث أن متوسط عمر الطالب موضع الدراسة لا يمكن أن يقل عن 30 سنة فإنه يختار الفرضية البديلة "أكبر من"

❖ $H1: >$

- إذا كان يعتقد أن متوسط عمر الطالب موضع الدراسة لا يزيد عن 30 سنة فإنه يختار الفرضية البديلة "أقل من"

❖ $H1: <$

- إذا لم يكن لديه أي تصور أو أي معلومات فإنه يختار الفرضية البديلة "لا يساوي"

❖ $H1: \neq$

اختبار الفرضية

- يهدف اختبار الفرضية إحصائيا إلى اتخاذ قرار حول ما إذا كانت الفرضية الصفرية مرفوضة أم غير مرفوضة
- عدم رفض الفرضية الصفرية لا يعني بالضرورة أنها صحيحة وإنما لا يوجد أدلة كافية من بيانات العينة لرفضها
- كما أن رفضها لا يعني بالضرورة أنها خاطئة بل يعني أن الإحصائية المحسوبة من العينة كانت بعيدة عن المعلمة المناظرة لها في المجتمع

رفض (H0) تعني ادانة المتهم أي أن هناك دليل كافي على أن المتهم مذنب

عدم رفض (H0) يعني انه لا يوجد دليل كافي لإدانة المتهم

خطوات إجراء الاختبار الإحصائي

الاختبار الاحصائي قد يكون

قد يكون متعلقا بـ:

- عينة واحدة
- عينتين
- اكثر

وقد يكون:

- معلمياً
- غير معلميا

الاختبارات الإحصائية قد تدور حول معالم المجتمع المجهولة مثل:
الوسط الحسابي، النسبة، التباين، معامل الارتباط،...

وفى هذه الحالة يطلق على هذه الاختبارات اسم
الاختبارات المعلميه Parametric Tests

وقد تتعلق بأشياء اخرى قد تكون وصفية مثل: العلاقة بين
التعليم والتدخين، العلاقة بين لون العينين ولون الشعر،....

وفى هذه الحالة يسمى الاختبار باسم الاختبار اللامعلمى
Non Parametric Test

توجد طريقتين لاتخاذ القرار في الاختبارات الاحصائية وهي:

I. **حساب احصاء الاختبار ومقارنته بقيمة جدوليه** وتحدد القيمة الجدولية بناء على نوع الاختبار ذو طرف واحد One Tail Test أو ذو طرفين Two Tail Test

I. **حساب ما يسمى بالقيمة الاحتمالية P–value** ويرمز لها في بعض البرامج الإحصائية بالرمز .Sig فاذا كان الاختبار ذو طرف واحد تقارن .Sig بالقيمة α لكن اذا كان الاختبار ذو طرفين تقارن بالقيمة $\alpha/2$

يمر الاختبار أياً كان نوعه بعدة خطوات يمكن إيجازها في التالي:

1. صياغة الفرضية الصفرية H_0

تأخذ – عادة – شكل " يساوي"

إذا كان المطلوب هو اختبار ما إذا كان متوسط عمر الطلبة هو 20 سنة فإن هذه الفرضية تصاغ كما يلي:

$$Ho: \mu = 20$$

صياغة الفرضية البديلة H1

2. تحديد مستوى المعنوية α

- واحدا من أهم المصطلحات المستخدمة في اختبار الفرضيات.
- القرار الذي نتخذه بناء على الاختبار الإحصائي لا يمكن اعتباره صحيح % 100 فهناك مقدار من الخطأ
- لأن المعلومات التي نتخذ قرارنا بناء عليها بيانات مأخوذة من عينة وليس من المجتمع الأصلي
- هو "احتمال حدوث الخطأ من النوع الأول". أي احتمال رفض الفرضية الصفرية بينما هي صحيحة".

أشهر قيمتين لمستوى المعنوية α هما (0.01، 0.05)

3. تحديد الاختبار الاحصائي المناسب

4. تحديد القيم الحرجة

القيم الحرجة لاختبار Z

مستوى المعنوية	الاختبار ثنائي الطرف	الاختبار أحادي الطرف
0,01	2,56	2,32
0,05	1,96	1,64
0,1	1,64	1,28

5. حساب إحصائية الاختبار:

هي الإحصائية التي يتم حسابها من بيانات العينة.

يتوقف حساب الإحصائية على العوامل التالية:

أ- توزيع المجتمع، وهل هو طبيعي أم لا، وهل تباينه معروف أم لا.

ب- حجم العينة، وهل هو كبير أم صغير.

ج- الفرضية الصفرية المراد اختبارها وهل هي عن المتوسط أو النسبة أو التباين أو الارتباط... الخ.

الفكرة الأساسية في تحديد إحصائية الاختبار هي: حساب الفرق بين قيمة المعلمة التي نفترضها للمجتمع والقيمة المقابلة لها في العينة

عندما يكون المتوسط الحسابي للعينة قريبا من المتوسط المفترض للمجتمع ⇦ لا ترفض H0

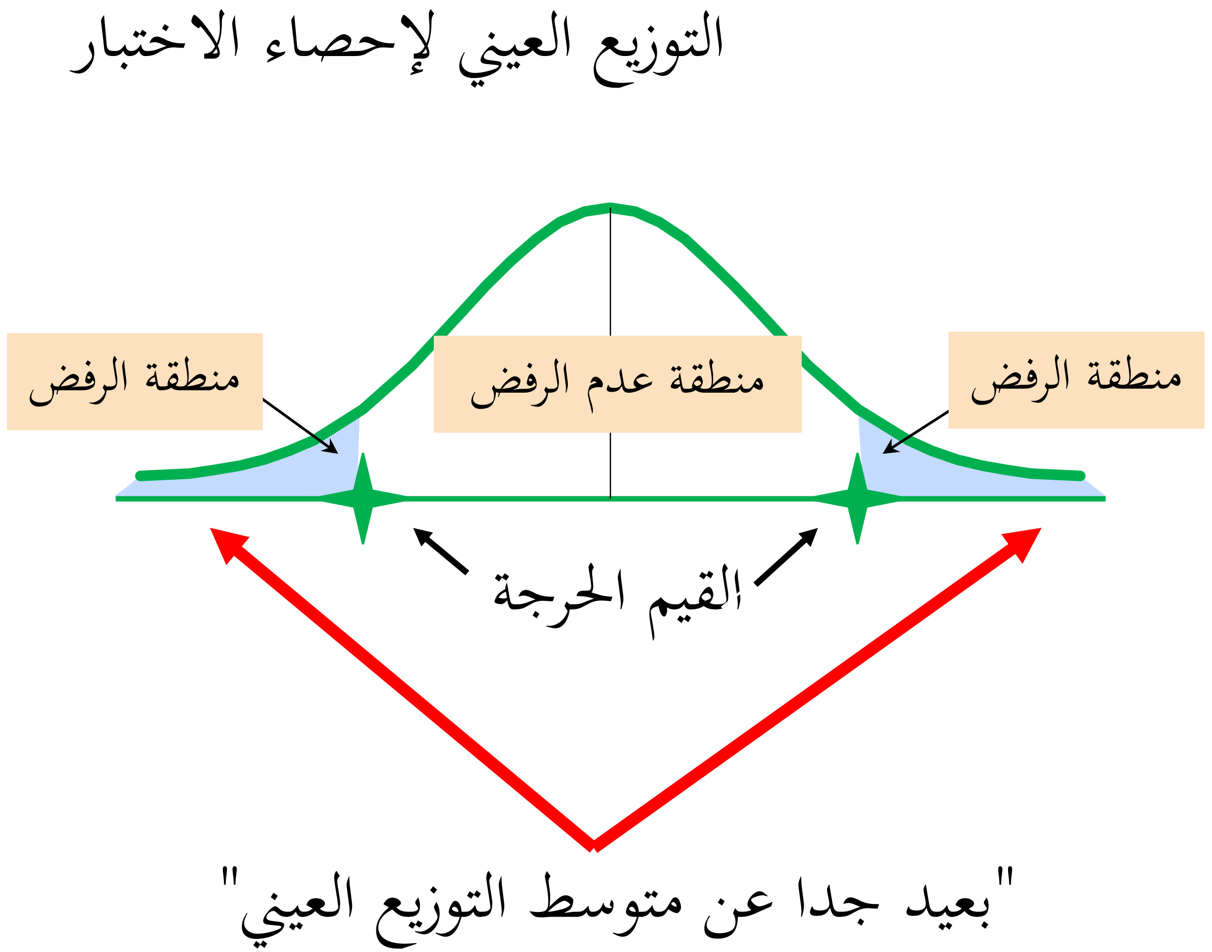
إذا ابتعد المتوسط الحسابي للعينة بعدا كافيا عن المتوسط المفترض للمجتمع ⇦ ترفض H0

المشكلة تكمن في تحديد متى يكون هذا الابتعاد كافيا لرفض H0

⇩

تكمن الإجابة في **القيمة الحرجة لإحصاء الاختبار**، إذ تمثل الحد الفاصل لاتخاذ القرار، وتحدد ما إذا كان الفارق بين المتوسطين يعد دالا إحصائيا أم لا.

إحصائية الاختبار والقيم الحرجة



تقسم المساحة تحت المنحنى إلى منطقتين:

الأولى تسمى "منطقة القبول" والثانية تسمى "منطقة الرفض"

تحديد هاتين المنطقتين يعتمد على:

أ- توزيع المعاينة (وهل هو طبيعي أو t أو...)

ب- والفرضية البديلة (أي هل يستخدم اختبار الطرفين أو الطرف الأيمن أو الأيسر).

ج- ومستوى المعنوية (وهل هو 1% أو 5% أو غير ذلك).

اختبار الفرضيات من طرف واحد:

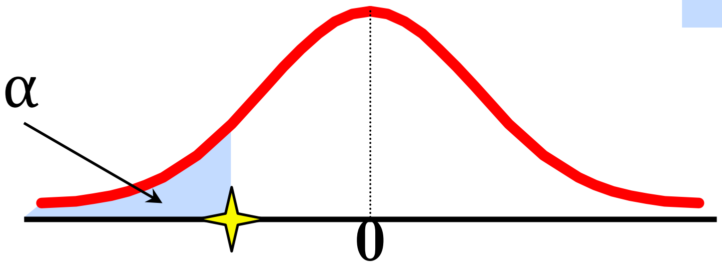
هو الاختبار الذي تبين فيه الفرضية البديلة بأن معلمة المجتمع اكبر أو اصغر من معلمة المجتمع المفترضة.

مستوى المعنوي = α

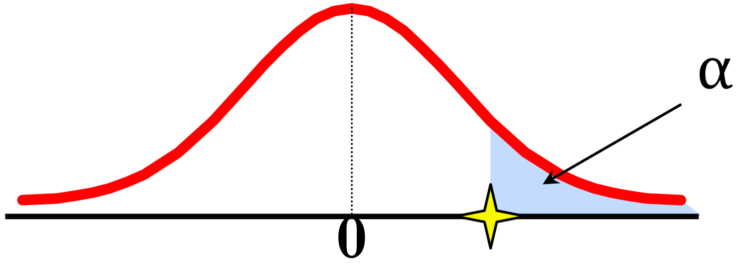
تظهر القيمة الحرجة

منطقة الرفض هي الجزء المضلل

$$H_0: \mu = 3$$
$$H_1: \mu < 3$$



$$H_0: \mu = 3$$
$$H_1: \mu > 3$$



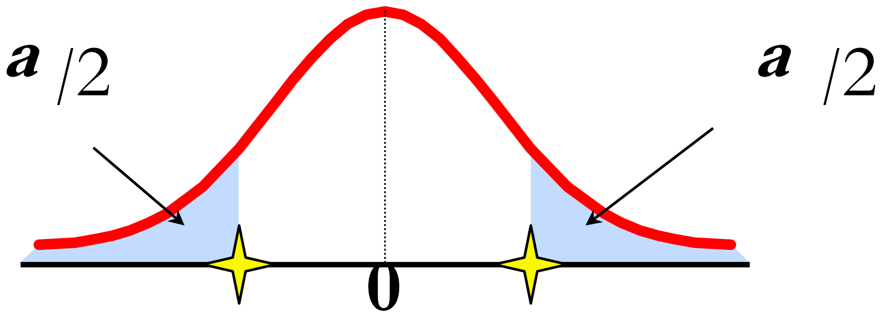
اختبار الفرضيات من طرفين:

هو الاختبار الذي لا تبين فيه الفرضية البديلة بأن معلمة المجتمع اكبر أو اصغر من معلمة المجتمع المفترضة بل مجرد أنها تختلف.

مستوى الدلالة = α

$H_0: \mu = 3$

$H_1: \mu \neq 3$



6. المقارنة والقرار:

أي أن نقارن قيمة الإحصائية (المحسوبة من الخطوة الثالثة) بحدود منطقتي القبول والرفض (والتي حددناها في الخطوة الرابعة).

- اذا وقعت قيمة الإحصائية داخل منطقة القبول فإن القرار هو: عدم رفض H_0
- إذا وقعت قيمة الإحصائية في منطقة الرفض فإن القرار هو رفض H_0 وقبول H_1

اختبارات الفرضيات للمتوسط

اختبارات الفرضيات للمتوسط μ

- σ معلوم (Z test)
- σ غير معلوم (t test)

اختبار ثنائي الطرف–اختبار Z (σ معلوم)

تحويل إحصائية العينة $(\bar{X})$ الى إحصاء الاختبار Z_{STAT}

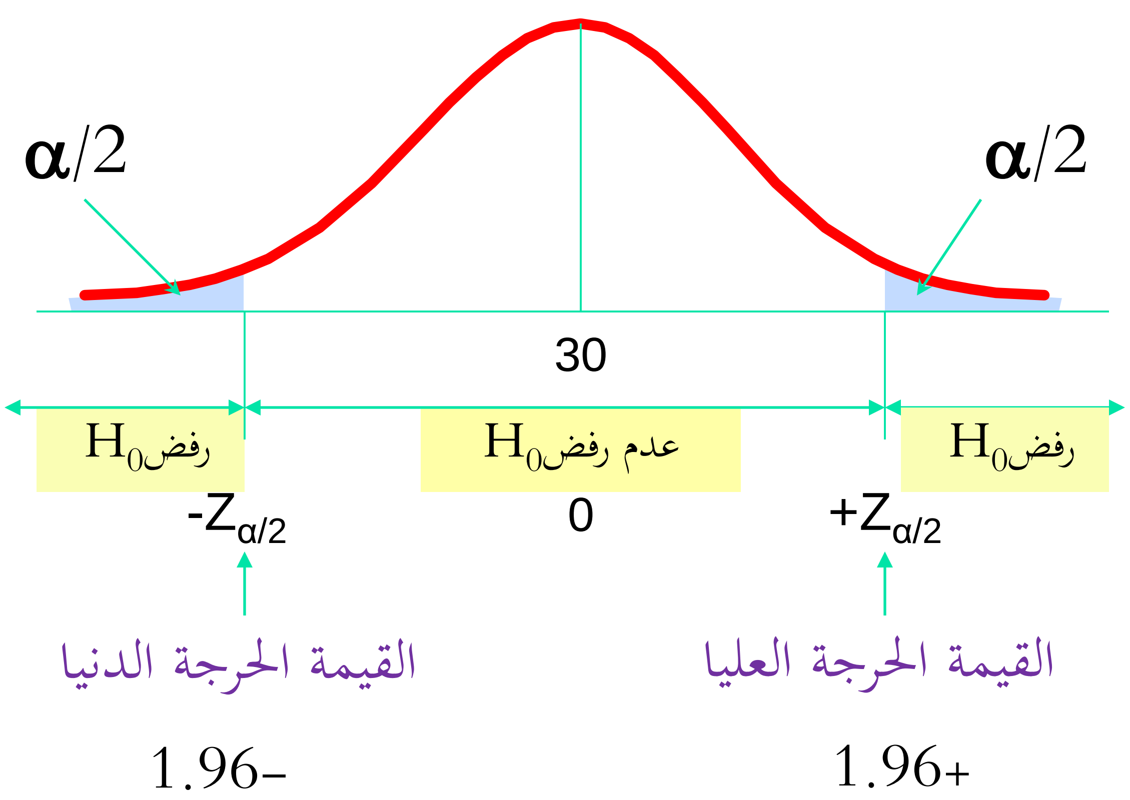
$$Z_{STAT} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

قاعدة القرار (اختبار ثنائي الطرف):

- نرفض H0 إذا كانت $Z < -Z\alpha/2$ أو $Z > +Z\alpha/2$
- لا نرفض H0 إذا كانت $-Z\alpha/2 \le Z \le +Z\alpha/2$

$H_0: \mu = 30$

$H_1: \mu \neq 30$



مثال

المطلوب: اختبار الادعاء أن متوسط قطر أنبوب مصنع هو $\mu_0 = 30$ بافتراض أن الانحراف المعياري للمجتمع يساوي $\sigma=0.8$

الحل

1. نكتب الفرضية الصفرية والبديلة

- $H_1: \mu \neq 30$ $H_0: \mu = 30$ (هذا اختبار ثنائي الطرف)

2. نحدد مستوى المعنوية وحجم العينة

- نفترض أن $\alpha = 0.05$ و $n = 100$ تم اختيارهما لهذا الاختبار

3. نحدد الاختبار الاحصائي المناسب

- افترضنا في البداية ان الانحراف المعياري للمجتمع معلوم وبالتالي فان الاختبار الاحصائي المناسب هو Z

4. نحدد القيمة الحرجة

- عند $\alpha = 0.05$ القيم الحرجة لـ Z هي ± 1.96

5. نجمع البيانات ونحسب الاحصائية

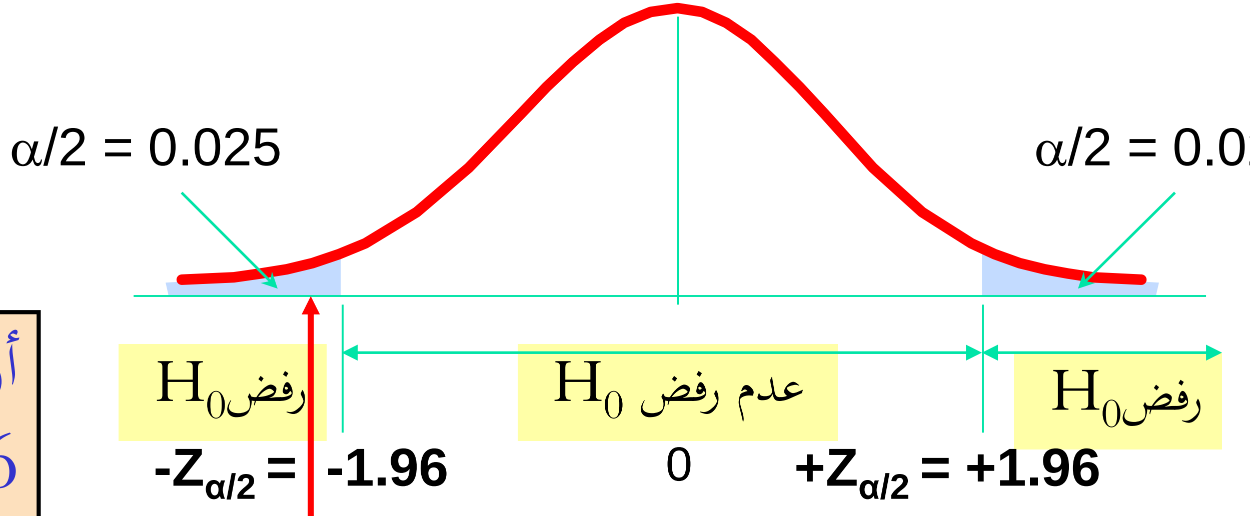
- نفترض أن نتائج العينة كانت

$n = 100$, $\bar{X} = 29.84$ ($\sigma = 0.8$ معروف)

ومنه اختبار الافتراض:

$$Z_{STAT} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{29.84 - 30}{\frac{0.8}{\sqrt{100}}} = \frac{-0.16}{0.08} = -2.0$$

6. هل إحصاء الاختبار في منطقة الرفض



أرفض H_0 اذا كان

$Z_{STAT} < -1.96$

$Z_{STAT} > 1.96$;

والا لا ترفض H_0

هنا $Z_{STAT} = -2.0 < -1.96$,

وبالتالي فان إحصائية الاختبار موجودة في منطقة الرفض

نرفض الفرضية العدمية ونستنتج أن هناك أدلة كافية على أن متوسط قطر الأنبوب المصنع لا يساوي 30.

طريقة القيمة الاحتمالية (p–Value) في الاختبار الإحصائي

القيمة الاحتمالية هي احتمال الحصول على إحصائية اختبار مساوية أو أكثر تطرفا من القيمة المحسوبة، بافتراض أن الفرضية الصفرية صحيحة

هي أصغر قيمة لـ α يمكن عندها رفض الفرضية الصفرية H0

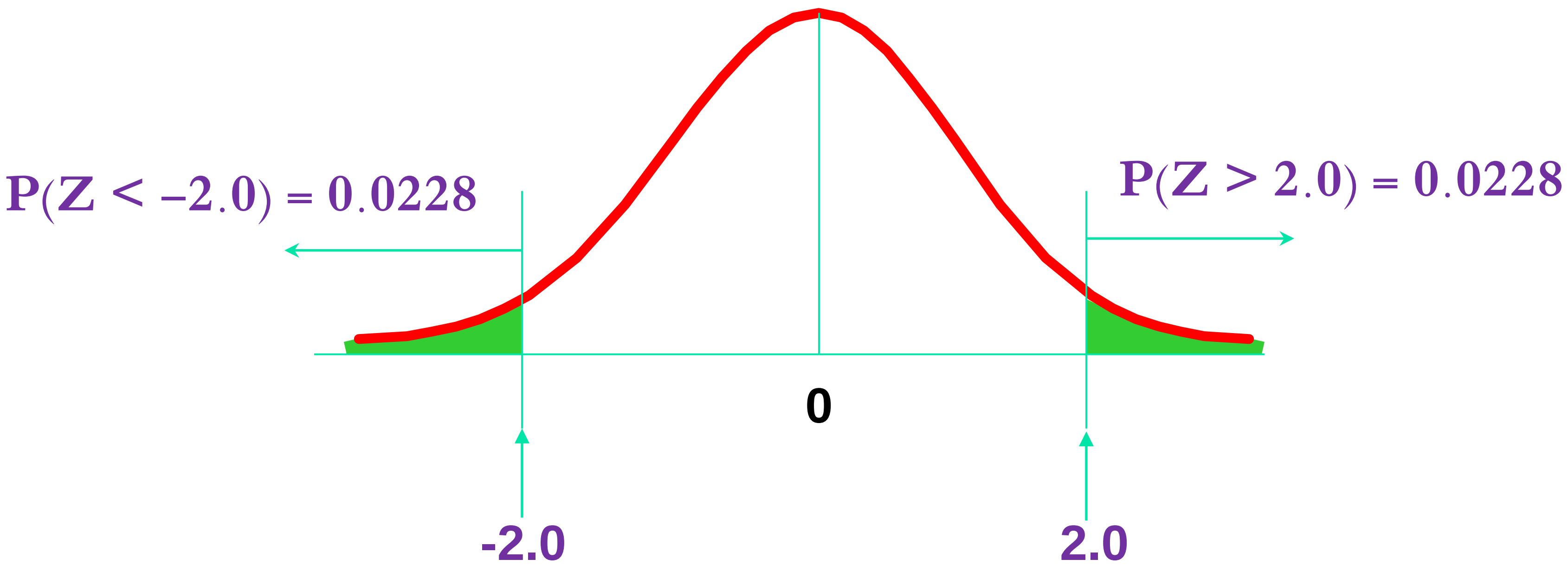
- مقارنة p-value مع α

  - إذا كانت القيمة الاحتمالية $(\text{p-value}) < \alpha$، نرفض الفرضية الصفرية $H0$.
  - إذا كانت القيمة الاحتمالية $(\text{p-value}) \geq \alpha$، لا نرفض الفرضية الصفرية $H0$.

مثال على اختبار الفرضيات باستخدام القيمة الاحتمالية: نطبق على بيانات المثال السابق

نحسب القيمة الاحتمالية p–value

ما مدى احتمال الحصول على قيمة Z_{STAT} تساوي 2 (أو أبعد) من المتوسط 0 في أي من الاتجاهين إذا كانت الفرضية الصفرية H0 صحيحة؟



p–value = 0.0228 + 0.0228 = 0.0456

- هل $p\text{-value} < \alpha?$

بما أن $0.05 = \alpha > 0.0456 = p\text{-value}$ نرفض H_0

الاستنتاج: هناك أدلة كافية لنحكم على أن متوسط قطر الأنبوب لا يساوي 30 مم

اختبارات الفرضيات للمتوسط

اختبارات الفرضيات للمتوسط μ

- σ غير معلوم (t test)
- σ معلوم (Z test)

اختبار الفرضية عندما يكون σ غير معلوم

عندما يكون الانحراف المعياري للمجتمع غير معلوم فإننا نطبق **t Test**

نحول إحصائية العينة $(\bar{X})$ الى إحصاء الاختبار t_{STAT}

$$\boxed{t_{STAT} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}}$$

مثال: يقال إن متوسط تكلفة غرفة فندق في نيويورك هو 168$ لليلة. للتحقق من صحة هذا القول، أُخذت عَيِّنة عشوائية مكوّنة من 25 فندقًا فكان متوسط العينة $\bar{X}$=$172.50 والانحراف المعياري للعينة = s=$15.40

اختبر الفرضيات المناسبة عند مستوى المعنوية .$\alpha=0.05$

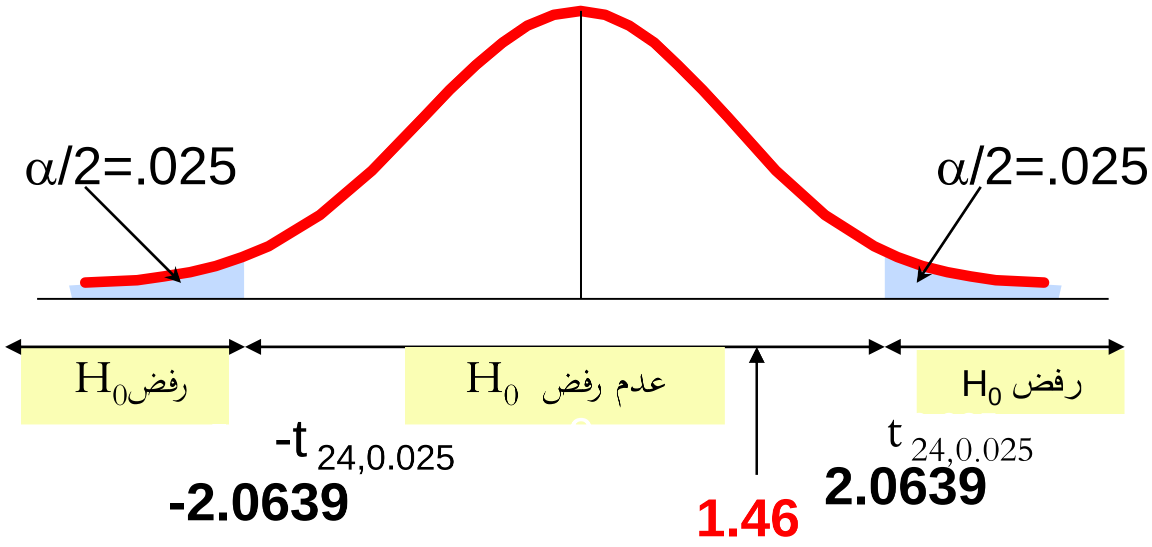
(افترض أن توزيع المجتمع طبيعي.)

الحل: $H_1: \mu \neq 168 \qquad H_0: \mu = 168$

بما أن الانحراف المعياري للمجتمع (σ) غير معروف،

فإننا نستخدم **اختبار t** بدلا من إحصاء Z.

$$\boxed{t_{STAT} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{172.50 - 168}{\frac{15.40}{\sqrt{25}}} = 1.46}$$



- $a = 0.05$
- $n = 25$, $df = 25-1=24$
- σ غير معروف
- القيمة الحرجة:
- $\pm t_{24,0.025} = \pm 2.0639$

لا نرفض الفرض الصفري H0

لا توجد أدلة كافية تشير إلى أن متوسط التكلفة الحقيقي يختلف عن **168** **دولارًا**.