احصاء حيوي

الكورس الاول

(موضوع المحاضرة)

اختبارات تتعلق بالمتوسطات

t-Distribution (توزيع– t)

Dr.Safwan Nathem Rashed

اختبار الفرضيات باستخدام توزيع – t
Test of Hypothesis by t-Distribution

تكلمنا عن كيفية استخدام التوزيع الطبيعي القياسي في اختبار الفرضيات بافتراض ان التباين المجتمع المدروس معلوماً وحجم العينة صغير او كبير.

اما عندما يكون تباين المجتمع غير معلوم وحجم العينة صغير عندئذٍ اختبار الفرضيات نستخدم توزيع – t.

ففي كثير من الاحيان لا يمكن معرفة تباين المجتمع الذي سحبت منه العينة اي تباينه غير معلوم وكانت حجم العينة $(n \geq 30)$

فيمكن استخدام تباين العينة (S^2) عوضاً عن تباين (σ^2) الغير معلومة علماً ان (S^2) هو مقدر جيد لـ (σ^2) لأنه لا يتغير كثيراً من عينة الى اخرى مادام **حجم العينة كبير** وعندئذٍ المختبر (Z) لا يزال يتوزع تقريباً توزيعاً قياسياً.

ولكن اذا كان **حجم العينة صغير (n<30)** فان قيمة (S^2) سوف تستخرج ايضاً من العينة ولكن المختبر الاحصائي سوف يتبع **توزيع – t** الذي يأخذ الصيغة الرياضية للاختبار :

$$t_{cal} = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \sim t(\alpha, v)$$

والشكل ادناه يوضح منحنى توزيع –t وتقاربه من التوزيع الطبيعي القياسي:

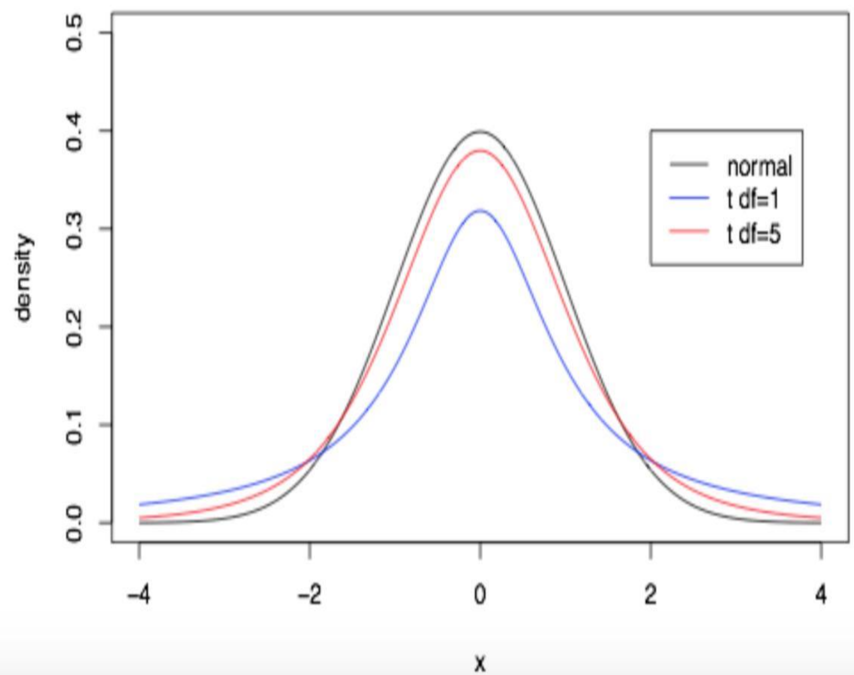
Population standard deviation

$$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim \ normal\ distribution\ N(0,1)$$

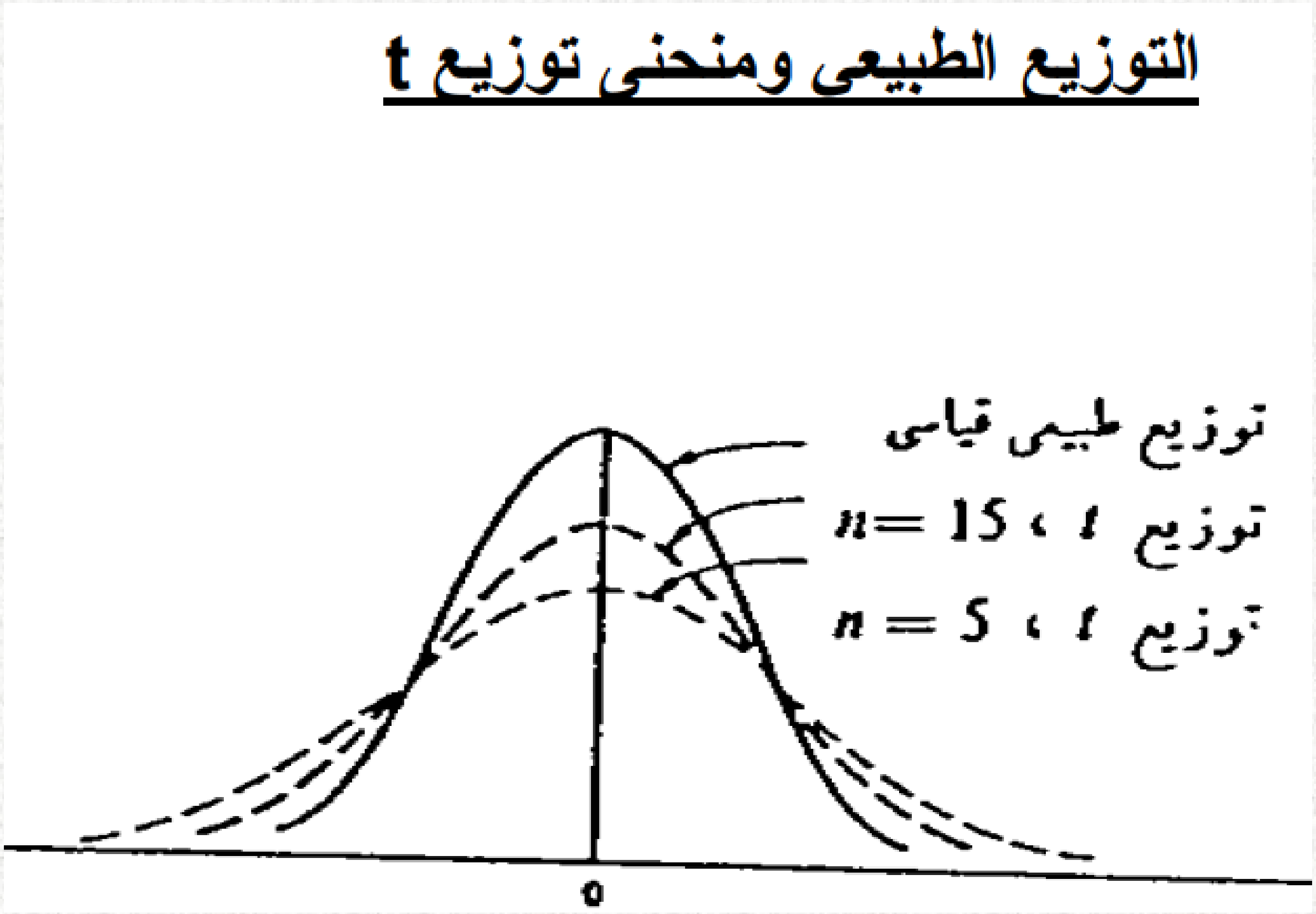
$$\frac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}} \sim \ t\ distribution\ dependent\ on\ (n-1)$$

Sample standard deviation

degree of freedom



والشكل الثاني ايضاً يوضح منحنى توزيع –t وتقاربه من التوزيع الطبيعي القياسي:



وهذا الاختبار خاص بالعينات الصغيرة، مع الاحتفاظ بصيغ بناء الفرضيات التي تم دراستها سابقاً والمتمثلة بفرضية العدم (H0) والفرضية البديلة (H1) والموضحة لدراسة الوسط الحسابي وهي:

$$H_0 : \mu = \mu_0 \quad ; \quad \text{v.s.} \begin{cases} H_1 : \mu \neq \mu_0 \\ H_1 : \mu > \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$

وان المختبر الاحصائي لتوزيع – t المعتمد فيه على المتوسط ($\bar{X}$) والانحراف المعياري (S) فضلاً عن حجم العينة (n<30) تمثل قيم المتغيرات العشوائية المسحوبة.

بعد حساب المختبر الاحصائي-t يتم مقارنته مع القيمة الجدولية لتوزيع – t والتي يتم ايجادها من القيمة الجدولية بالاعتماد على مستوى المعنوية (α) ودرجة الحرية (v= n - k)، k تمثل عدد المقدرات المدروسة وفي دراستنا هناك مقدر واحد لتكون قيمة (v=n-1) مع مراعاة الفرضية التي تم بنائها اي مناطق الرفض إن كانت من طرف واحد من المنحنى او من طرفين او جانبين من خلال اعتمادها على الفرضية البديلة.

الجدول الخاص بتوزيع- t

PERCENTAGE POINTS OF THE t DISTRIBUTION

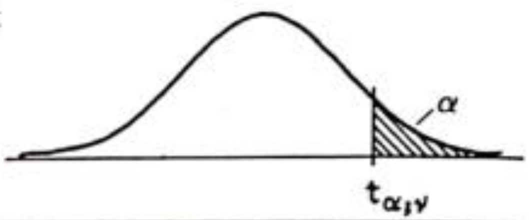
The table gives the value of $t_{\alpha;\nu}$ — the 100α percentage point of the t distribution for ν degrees of freedom.

$(\alpha, n-1)$

The values of t are obtained by solution of the equation:-

$$\alpha = \Gamma\{\tfrac{1}{2}(\nu+1)\}\{\Gamma(\tfrac{1}{2}\nu)\}^{-1}(\nu\pi)^{-1/2}\int_t^{\infty}(1+x^2/\nu)^{-(\nu+1)/2}dx$$

Note. The tabulation is for one tail only i.e. for positive values of t. For $|t|$ the column headings for α must be doubled.



ν =	α = 0.10	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221

كيفية ايجاد القيمة الجدولية او القيمة الحرجة لتوزيع - t

ادناه مجموعة من القيم الحرجة تم استنتاجها من جداول توزيع-t معتمدين فيها على الفرضية البديلة لتحديد مناطق الرفض من جانب واحد او جانبين حسب الصيغ الاتية:

جانب واحد: $\boxed{t_{(\alpha,v=n-1)}}$

جانبين: $\boxed{t_{\left(\frac{\alpha}{2},v=n-1\right)}}$

$$t_{(0.05,v=5)}=2.015\ \ ;\ \ t_{(0.01,v=20)}=2.528$$

$$t_{(0.025,v=30)}=2.042\ \ ;\ \ t_{(0.005,v=34)}=\frac{2.750+2.704}{2}=2.727$$

كيف يتم استخراج القيمة الحرجة لتوزيع- t

$\alpha =$	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
$\nu =$ 1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850

ν \ α =	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

مثال:

ادعت احدى شركات انتاج السكائر بان نسبة النيكوتين في انتاجها من السكائر لا يتجاوز (17.5) ملغم. وعليه تم اخذ عينة عشوائية من المنتج مؤلفة من (9) سكائر وتم قياس نسبة النيكوتين فيها فكانت النتائج على النحو الاتي:

Xi	18	16	20	19	18	19	18	18	17

فهل ادعاء الشركة صحيح عند مستوى معنوية ($\alpha = 0.10$).

Sol/1 μ_0: متوسط معلوم من المجتمع يساوي ١٧،٥

فرضية العدم

$$1-\quad H_0 : \mu < 17.5$$

فرضية بديلة

$$2-\quad H_1 : \mu > 17.5$$

مستوى المعنوية

$$3-\quad \alpha = 0.10 \rightarrow v = n - 1 = 9 - 1 = 8$$

اذا منطقة القبول تكون

$$t_{(0.10,8)} = 1.397$$

المختبر الاحصائي

يتم ايجاد الوسط الحسابي والتباين او الانحراف المعياري للعينة المسحوبة.

$$4-\quad t_{cal} = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{163}{9} = 18.1\ ; \sum_{i=1}^{n} X_i^2 = 18^2 + 16^2 + \cdots + 17^2 = 2963$$

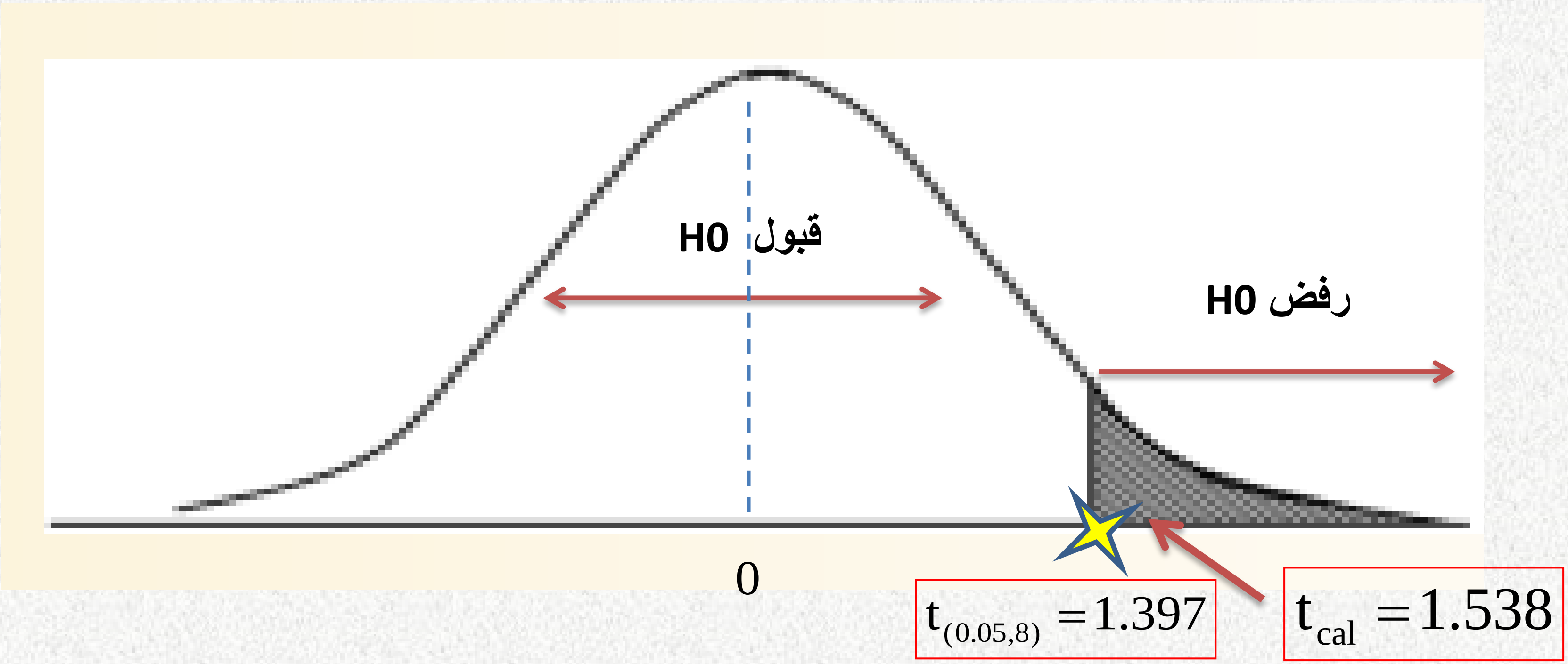
$$S^2 = \frac{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}{n-1} = \frac{2963 - \frac{(163)^2}{9}}{8} = 1.363$$

$$S = \sqrt{1.363} = 1.17$$

$$\therefore t_{cal} = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{18.1 - 17.5}{\frac{1.17}{\sqrt{9}}} = 1.538$$

$$\therefore t_{cal} = 1.538$$

5- الرسم البياني لتوضيح مناطق الرفض والقبول



6-

القرار: نلاحظ ان القيمة المحسوبة لـ $|t_{cal}|=1.538$ اكبر من القيمة الجدولية لـ $t_{table}=1.379$ اي $(|tcal|>t_{table})$ (1.397<1.538) فضلاً عن ان القيمة المحسوبة تقع في منطقة الرفض لـ H0 تحت المنحنى وهذا يدل على رفض فرضية العدم والقبول بالفرضية البديلة، (هذا الجزء مهم) وهذا اكد على ان ادعاء الشركة غير صحيح بنسبة النيكوتين في السكائر اكبر من (17.5) وسوف يؤدي الى زيادة في الخطورة اضافة الى الخطورة في تعاطيها.

مثال:

قام احد الباحثين بدراسة متوسط اعمار مجموعة من المرضى المصابين بمرض ترهل البطين الايمن من القلب من خلال اختيار عينة من هؤلاء المرضى قوامها $(n=20)$ مريض، حيث تم اختيارهم من مجتمع متوسط اعمارهم (60) سنة فكان الوسط الحسابي للعينة (55) سنة وبتباين قدره (64) ، اختبر هل يوجد اختلاف بمتوسط اعمار العينة عن المجتمع عند مستوى معنوية $(\alpha = 0.05)$.

Sol/1 μ_0 : متوسط معلوم من المجتمع يساوي ٦٠

1- $H_0 : \mu = 60$ فرضية العدم

2- $H_1 : \mu \neq 60$ فرضية بديلة

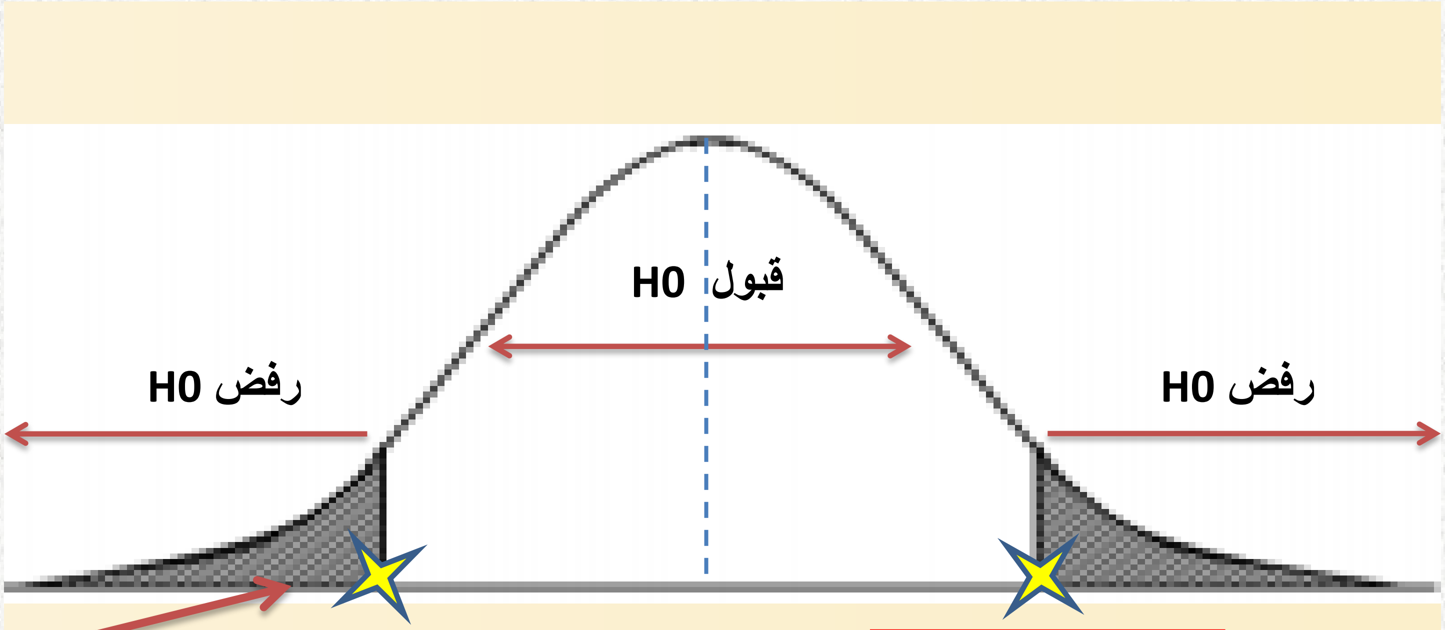
3- $\alpha = \dfrac{0.05}{2} = 0.025 \rightarrow v = 20 - 1 = 19$ مستوى المعنوية

$t_{(0.025,19)} = 2.093$ اذا منطقة القبول تكون

4- $t_{cal} = \dfrac{\bar{X} - \mu_0}{\dfrac{S}{\sqrt{n}}} = \dfrac{55 - 60}{\sqrt{\dfrac{64}{20}}} = -2.795$ المختبر الاحصائي بتعويض الوسط الحسابي والتباين للعينة المسحوبة.

$$\boxed{t_{cal} = -2.795}$$

5- الرسم البياني لتوضيح مناطق الرفض والقبول



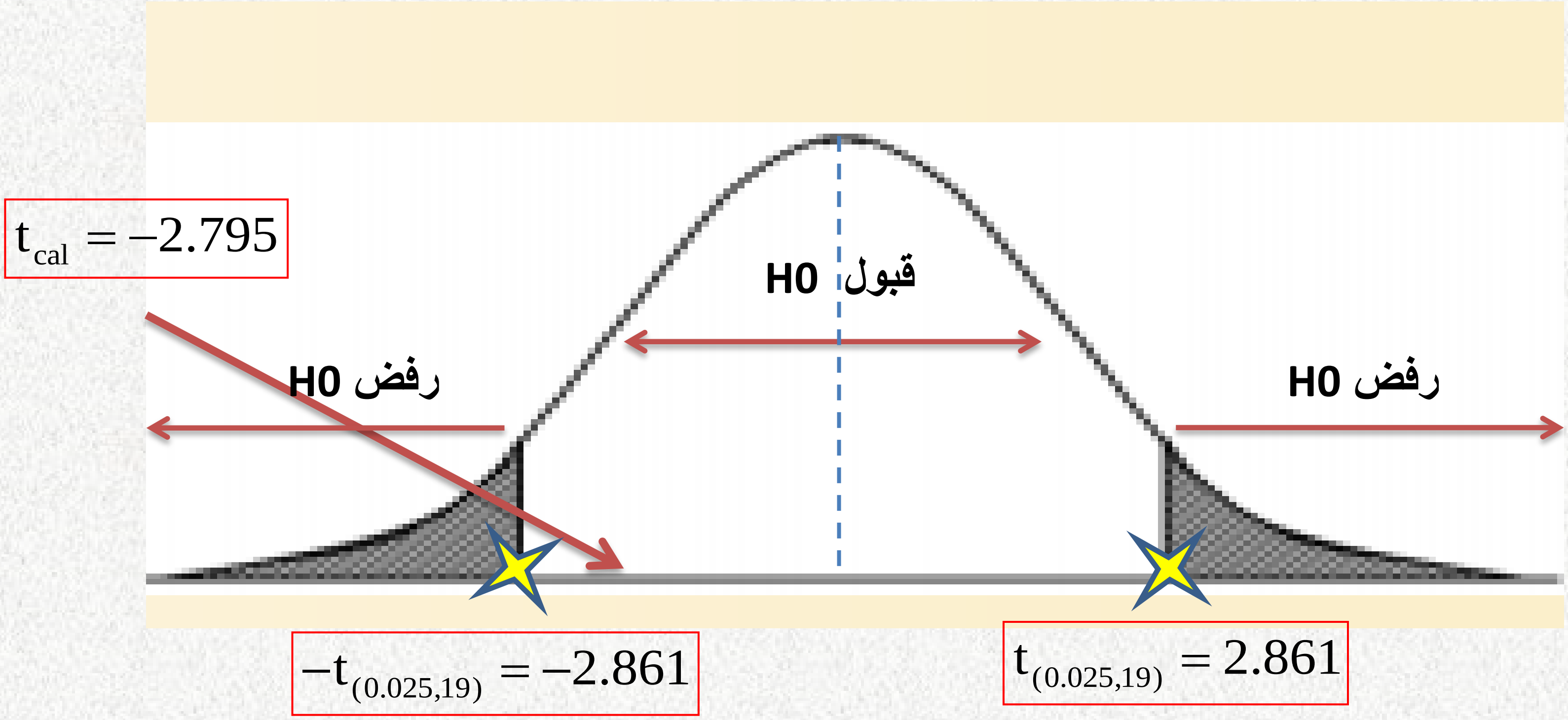
$\boxed{t_{cal} = -2.795}$ $\boxed{-t_{(0.025,19)} = -2.093}$ $\boxed{t_{(0.025,19)} = 2.093}$

ليكون القرار هو : ان القيمة المحسوبة لـ $|tcal|=2.795$ اكبر من القيمة الجدولية لـ $t_{table}=2.093$ اي $(|t_{cal}|>t_{table})$ $(2.795>2.093)$ فضلاً عن ان القيمة المحسوبة تقع في منطقة الرفض لـ $H0$ تحت المنحنى وهذا يدل على رفض فرضية العدم والقبول بالفرضية البديلة، وهذا يدل على ان متوسط اعمار المرضى المصابين بمرض ترهل البطين الايمن من القلب اقل من (60) وعليه لابد من متابعة الحالة ايضا دون عمر (60) سنة لغرض تجنب المرض او للعلاج.

ملاحظة: ما هو قرارك حول المرض عند مستوى معنوية (0.01).

$$\boxed{t_{cal} = -2.795}$$

5- الرسم البياني لتوضيح مناطق الرفض والقبول



مثال/واجب:

ادعى احد الطلبة ان اطوال المرحلة الثالثة في قسم الاحصاء والمعلوماتية يساوي (155) سم. ولاختبار هذا الادعاء تم اختيار عينة عشوائية من الطلبة في نفس المرحلة وتم اخذ قياس اطوالهم وكانت نتائج القياس موضحة في الجدول ادناه، وقد تم وضع احتمال خطأ عند مستوى معنوية ($\alpha = 0.05$) فهل كان ادعاء الطالب صحيح حسب الحالات الاتية.

١ – دراسة اطوال الطلبة ذكور واناث

٢ –دراسة اطوال الطلبة للذكور

٣ –دراسة اطوال الطلبة للإناث

جدول بقياس اطوال الطلبة

Xi	ذ	ذ	ا	ا	ذ	ذ	ذ	ا	ذ
	150	152	142	180	178	165	145	140	166
	ا	ذ	ذ	ذ	ا	ا	ا	ا	ذ
	150	180	179	178	170	150	140	144	140
	ا	ا	ذ	ذ	ا	ذ	ذ	ا	ا
	146	146	145	150	145	145	155	160	155
	ا	ا	ا	ذ	ذ	ا	ذ	ذ	ذ
	168	167	169	166	179	168	168	166	150

مخطط يوضح اختيار الاختبار الاحصائي