توزيع T

توزيع T هو أحد التوزيعات المهمة والمستخدمة خصيصاً للاستدلال الإحصائي حول بعض معالم المجتمع بما في ذلك الاستدلال حول متوسطات المجتمع.

فإذا كان لدينا متغير عشوائي X يتبع التوزيع الطبيعي، أي

$$X \sim N(\mu, \sigma^2)$$

فإذا تم سحب عينة بحجم (n)، فإن المتوسط $\bar{X}$ متغير عشوائي يتبع التوزيع الطبيعي، أي

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

وبذلك فإن الكمية:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

تتبع التوزيع الطبيعي المعياري.

وفي حالة عدم معرفة σ ، فإن يمكن استخدام S باعتبارها تقدير جيد للانحراف المعياري للمجتمع، وبذلك فإن:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}$$

أي أننا نستخدم توزيع T بدلا عن توزيع Z في حالة استخدام الانحراف المعياري للعينة بدلا عن الانحراف المعياري للمجتمع. وتوزيع T هو توزيع مشابه لتوزيع Z، ويقترب التوزيعان من التطابق عندما يكون حجم العينة كبير. ويمكن استخدام توزيع T في هذه الحالة لحساب الاحتمالات المرتبطة بقيمة $\bar{X}$ بنفس الطريقة المتبعة باستخدام توزيع Z لحساب الاحتمالات المرتبطة بقيمة X.

وبناء على ذلك، وباعتبار أن $\bar{X}$ تقدير جيد لمتوسط المجتمع المجهول μ، فإنه يمكن إنشاء فترة ثقة لمتوسط المجتمع μ وبمستوى ثقة $(1 - \alpha)$.

$$P\left(\bar{X} - t_{\left(1-\frac{\alpha}{2};n-1\right)}\ S/\sqrt{n} < \mu < \bar{X} + t_{\left(1-\frac{\alpha}{2};n-1\right)}\ S/\sqrt{n}\right) = 1 - \alpha$$

وتعبر α عن احتمال أن لا تشمل فترة الثقة القيمة الحقيقية لمتوسط المجتمع

اختبارات T حول المتوسطات:

- اختبارات T حول متوسط مجتمع واحد

اختبار ذو طرف واحد $\left.\begin{matrix} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{matrix}\right\}$،
(منطقة الرفض تقع في الطرف الأيمن من توزيع T)

اختبار ذو طرف واحد $\left.\begin{matrix} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{matrix}\right\}$،
(منطقة الرفض تقع في الطرف الأيسر من توزيع T).

اختبار ذو طرفين $\left.\begin{matrix} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{matrix}\right\}$،
(منطقة رفض تقع في الطرف الأيمن وأخرى تقع في الطرف الأيسر).
حيث (μ_0) قيمة عددية مفترضة.

- واختبار T حول متوسطات مجتمعين مستقلين.

اختبار ذو طرف واحد أيمن $\left.\begin{matrix} H_0: \mu_1 \leq \mu_2 \\ H_1: \mu_1 > \mu_2 \end{matrix}\right\}$،

اختبار ذو طرف واحد أيسر $\left.\begin{matrix} H_0: \mu_1 \geq \mu_2 \\ H_1: \mu_1 < \mu_2 \end{matrix}\right\}$،

اختبار ذو طرفين $\left.\begin{matrix} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{matrix}\right\}$،

حيث (μ_1) تمثل المتوسط المفترض للمجتمع الأول و (μ_2) تمثل المتوسط المفترض للمجتمع الثاني.

خطوات اختبار الفرضيات:

1. تحديد فرضية العدم وفرضية الباحث (الفرضية البديلة)
2. تحديد مدى ثقة الباحث في صحة قرار رفض أو عدم رفض فرضية العدم $(1-\alpha)$، او احتمال رفض فرضية عدم صحيحة (α).
3. تحدي إحصائية الاختبار، وتختلف هذه الإحصائية بناء على الاختبار المطلوب.
4. تحديد قاعدة اتخاذ القرار، وهي الشرط أو الشروط الواجب تحققها لرفض فرضية العدم.
5. اتخاذ القرار، وهو رفض أو عدم رفض فرضية العدم بناء على نتائج العينة، وكذلك إيضاح مدلول هذا القرار.

أمثلة على اختبار T حول متوسط مجتمع واحد: (تطبيق على بيانات الموظفين)

مثال (1): اختبر فرضية الباحث القائلة بأن متوسط الدخل الحالي للموظفات ذواتي الوظائف المكتبية أكبر من 24500 دولار.

الخطوة الأولى: تحديد فرضية الباحث وفرضية العدم.

$$H_0: \mu_f \leq 24500$$
$$H_1: \mu_f > 24500$$

حيث μ_f تمثل متوسط الدخل الحالي للموظفات ذواتي الوظائف المكتبية.

وبناء على هذه الفرضية، فإنه يمكن رفض فرضية العدم وقبول فرضية الباحث عندما يكون متوسط العينة أكبر بكثير من 24500، أما إذا كانت قيمة متوسط العينة قريب من 24500 أو اقل من 24500 فإنه لا يمكن رفض فرضية العدم.

الخطوة الثانية: تحديد قيمة (α)

$$1 - \alpha = 0.95 \implies \alpha = 0.05$$

الخطوة الثالثة: تحديد إحصائية الاختبار

$$T^* = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

وهذه الإحصائية خاصة فقط اختبار T حول متوسط مجتمع واحد.

الخطوة الرابعة: تحديد قاعدة اتخاذ القرار (متى نرفض فرضية العدم؟)

نظرا لأن الاختبار ذو طرف واحد (منطقة رفض فرضية العدم تقع في الطرف الأيمن من توزيع T)، لذا فإنه يوجد شرطين لرفض فرضية العدم هما:

1. قيمة (T^*) أكبر من الصفر (أي موجبة).
2. قيمة P-value أقل من قيمة (α)، أي أقل من 0.05.

الخطوة الخامسة: اتخاذ القرار بناء على نتائج.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Current Salary	206	$25,003.69	$5,812.838	$405.000

One-Sample Test

	Test Value = 24500					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Current Salary	1.244	205	.215	$503.689	$-294.81-	$1,302.19

يمكن ملاحظة أن قيمة $(T^* = 1.244)$ وهي قيمة موجبة، وهذا يعني تحقق الشرط الأول في الخطوة الرابعة، وبالنظر إلى الشرط الثاني، فإن:

$$P - value = \frac{0.215}{2} = 0.1075$$

أي أن قيمة $(P - value > 0.05)$، وبذلك فإن الشرط الثاني لم يتحقق. لذلك فإنه لا يمكن رفض فرضية العدم. وهذا يعني أن متوسط الدخل الحالي للموظفات اقل من أو يساوي 24500. الملاحظ أن قيمة $(\bar{X} = 25003.69)$ وهي أكبر بقليل من 24500. لذلك فإن نتائج العينة لا تؤيد فرضية الباحث. وحتى نتمكن من رفض فرضية العدم وتأييد فرضية الباحث، فإنه يتوجب الحصول على قيمة $\bar{X}$ تكون أكبر بكثير من القيمة المفترضة وهي 24500.

ملاحظة: تمثل P-value قيمة احتمالية تحسب من توزيع T وباستخدام العلاقة التالية:

$$P - value = P\left(T_{(n-1)} > |T^*|\right)$$

مثال (2): اختبر فرضية الباحث القائلة بأن متوسط الدخل الابتدائي لجميع الموظفين غير الأقلية يختلف عن 16500 دولار.

الخطوة الأولى: تحديد فرضية الباحث وفرضية العدم.

$$H_0: \mu_{no} = 16500$$
$$H_1: \mu_{no} \neq 16500$$

حيث μ_{no} تمثل متوسط الدخل الابتدائي لجميع الموظفين غير الأقلية.

وبناء على هذه الفرضية، فإنه يمكن رفض فرضية العدم وقبول فرضية الباحث عندما يكون متوسط العينة $(\bar{X})$ أكبر بكثير من 16500، أو إذا كانت قيمة $(\bar{X})$ اقل بكثير من 16500، أما إذا كانت قيمة $(\bar{X})$ قريبة من القيمة المفترضة وهي 16500، فإنه لا يمكن رفض فرضية العدم.

الخطوة الثانية: تحديد قيمة (α)

$$1-\alpha = 0.95 \implies \alpha = 0.05$$

الخطوة الثالثة: تحديد إحصائية الاختبار

$$T^* = \frac{\bar{X}-\mu}{S/\sqrt{n}}$$

وهذه الإحصائية خاصة فقط اختبار T حول متوسط مجتمع واحد.

الخطوة الرابعة: تحديد قاعدة اتخاذ القرار (متى نرفض فرضية العدم؟)

نظرا لأن الاختبار ذو طرفين (توجد منطقتين لرفض فرضية العدم، الأولى تقع في الطرف الأيمن من توزيع T، والثانية تقع في الطرف الأيسر من توزيع T)، لذا فإنه يوجد شرط واحد لرفض فرضية العدم وهو أن قيمة P-value أقل من قيمة $\left(\frac{\alpha}{2}\right)$، أي أقل من 0.025. ومن المناسب التأكيد بأن كون قيمة (T^*) سالبة أو موجبة لا تؤثر في اتخاذ القرار وذلك بسبب وجود منطقتي رفض في الطرف الأيمن والطرف الايسر. كذلك فإنه تم تقسيم قيمة (α) إلى قسمين متساويين لوجود منطقتي رفض في طرفي توزيع T.

الخطوة الخامسة: اتخاذ القرار بناء على نتائج العينة.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Beginning Salary	370	$17,673.01	$8,392.419	$436.301

One-Sample Test

	Test Value = 16500					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Beginning Salary	2.689	369	.008	$1,173.014	$315.07	$2,030.96

يمكن ملاحظة أن $\left(P-value = \frac{0.008}{2} = 0.004\right)$، وهي أقل من 0.025، لذا فإنه يمكن رفض فرضية العدم وقبول فرضية الباحث، وهذا يعني أن متوسط الدخل الابتدائي لجميع الموظفين غير الأقلية يختلف عن 16500 دولار.

أمثلة على اختبار T حول متوسط مجتمع مستقليين: (تطبيق على بيانات الموظفين)

مثال (3): (بالنظر فقط إلى الموظفين في الوظائف المكتبية)،اختبر الفرضية القائلة بأن متوسط الدخل الابتدائي للأقلية أقل من متوسط الدخل الابتدائي لغير الأقلية.

الخطوة الأولى:

$$H_0: \mu_1 \geq \mu_0$$
$$H_1: \mu_1 < \mu_0$$

حيث μ_1 تمثل متوسط الدخل الابتدائي للموظفين الأقلية ذوي الوظائف المكتبية،

و μ_0 تمثل متوسط الدخل الابتدائي للموظفين غير الأقلية ذوي الوظائف المكتبية.

في هذه الفرضية، نقارن بين مجتمعين مستقلين، المجتمع الأول هو الموظفين الأقلية والمجتمع الثاني هو الموظفين غير الأقلية. ويدعي الباحث أن متوسط الدخل الابتدائي للموظفين الأقلية ذوي الوظائف المكتبية أقل من متوسط الدخل الابتدائي للموظفين غير الأقلية ذوي الوظائف المكتبية. ونلاحظ أنه لم يتم تحديد قيمة لمتوسط الدخل الابتدائي لأي من المجتمعين، بل يتم اختبار أن متوسط مجتمع أقل من متوسط مجتمع آخر بدون تحديد لقيم المتوسطين في المجتمعين. ويتم رفض فرضية العدم إذا كان متوسط المجتمع الأول أقل بكثير من متوسط المجتمع الثاني. ولذا فإنه إذا كان متوسط الدخل الابتدائي للعينة من المجتمع الأول أقل بكثير من متوسط الدخل الابتدائي للعينة من المجتمع الثاني، فهذه النتيجة تؤيد فرضية الباحث، لذلك نرفض فرضية العدم. أما إذا كان متوسط الدخل الابتدائي للعينة من المجتمع الأول قريب من أو (أكبر من) من متوسط الدخل الابتدائي للعينة من المجتمع، فإن هذه النتيجة لا تؤيد فرضية الباحث بل تؤيد فرضية العدم، لذا لا يمكن رفض فرضية العدم.

الخطوة الثانية: تحديد قيمة (α)

$$1 - \alpha = 0.95 \implies \alpha = 0.05$$

الخطوة الثالثة: تحديد إحصائية الاختبار (هذه الاحصائية خاصة بهذا الاختبار)

$$T^* = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

حيث:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

و $(\bar{X}_1)$ هي متوسط العينة من المجتمع الأول و $(\bar{X}_2)$ هي متوسط العينة من المجتمع الثاني، وفي هذا الاختبار فإن المجتمع الأول هو الأقلية والمجتمع الثاني هو غير الأقلية. كذلك فإن (n_1) و (S_1^2)هما حجم العينة وتباين العينة من المجتمع الأول و (n_2) و (S_2^2) هما حجم العينة وتباين العينة من المجتمع الثاني. وتسمى (S_p^2) بالتباين المجمع من العينتين.

الخطوة الرابعة: تحديد قاعدة اتخاذ القرار (متى نرفض فرضية العدم؟)

نظراً لأن الاختبار ذو طرف واحد (منطقة رفض فرضية العدم تقع في الطرف الأيسر من توزيع T)، لذا فإنه يوجد شرطين لرفض فرضية العدم هما:

1. قيمة (T^*) أقل من الصفر (أي سالبة).
2. قيمة P-value أقل من قيمة (α)، أي أقل من 0.05.

الخطوة الخامسة: اتخاذ القرار بناء على نتائج العينة.

Group Statistics

	Minority Classification	N	Mean	Std. Deviation	Std. Error Mean
Beginning Salary	Yes	87	$13,695.86	$2,686.119	$287.982
	No	276	$14,222.19	$2,967.275	$178.609

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t	df	Sig. (2-tailed)
Beginning Salary	Equal variances assumed	.451	.502	-1.475-	361	.141
	Equal variances not assumed			-1.553-	157.594	.122

يوضح الجدول الأول إحصائيات العينتين، أما الجدول الثاني فيوضح نتائج اختبار T حول متوسطات مجتمعين مستقلين. وتحتوي النتائج في الجدول الثاني على اختبارين، الأول يعنى باختبار اختلاف تباين المجتمعين والثاني يعنى باختبار T حول متوسطات المجتمعين. ويمكن صياغة فرضية الاختبار الأول على النحو التالي:

$$H_0: \sigma_1^2 = \sigma_0^2$$
$$H_1: \sigma_1^2 \neq \sigma_0^2$$

وبناء على نتائج اختبار ليفين، فإنه لا يمكن رفض فرضية العدم والتي تشير إلى تساوي تباين المجموعتين وذلك لأن $(P - value > 0.05)$ لاختبار F، وهذا يعني أنه لا يوجد اختلاف

جوهري بين تباين المجتمعين. وفي هذه الحالة، يتم استخدام نتائج اختبار T الناتجة من افتراض تساوي تباين المجتمعين (السطر الأول من النتائج) لاتخاذ القرار في اختبار T حول متوسطات مجتمعين مستقلين.

ومن نتائج العينة فإن $(T^{*} = -1.475)$ و $\left(P - value = \frac{0.141}{2} = 0.0705\right)$.

وبذلك فإن الشرط الأول تحقق، أما الشرط الثاني فإنه لم يتحقق. لذلك فإنه لا يمكن رفض فرضية العدم، وهذا يعني أن متوسط الدخل الابتدائي للأقلية لا يقل عن متوسط الدخل الابتدائي لغير الأقلية.

ملاحظة:

في الحالة رفض فرضية العدم الخاصة باختلاف التباين في المجتمعين (اختبار ليفين)، فإنه يتم استخدام نتائج اختبار T التقريبي والذي يفترض عدم تساوي التباين (السطر الثاني من النتائج) لاتخاذ قرار في اختبار T حول متوسطات مجتمعين مستقلين. ويمكن ملاحظة أن نتائج اختبار T بافتراض تساوي التباين ونتائج اختبار T التقريبي بافتراض عدم تساوي التباين متقاربة جداً، لذلك قد يكون من المناسب دائماً استخدام اختبار T التقريبي حيث أن النتائج ستكون متقاربة بالمقارنة باختبار T وذلك في حالة تساوي تباين، وسيكون هو الاختبار المناسب في حالة عدم تساوي التباين.

مثال (4): اختبر الفرضية القائلة بأن متوسط ضغط الدم الانبساطي للمدخنين الحاليين يختلف عن متوسط ضغط الدم الانبساطي للمدخنين السابقين. (البيانات الطبية). ويمكن التعرف على البيانات بالرجوع إلى الملف الخاص بوصف البيانات.

الخطوة الأولى:

$$H_0 : \mu_s = \mu_{ns}$$
$$H_1 : \mu_s \neq \mu_{ns}$$

حيث μ_s تمثل متوسط ضغط الدم الانبساطي للمدخنين الحاليين،

و μ_{ns} تمثل متوسط ضغط الدم الانبساطي للمدخنين السابقين.

في هذه الفرضية، تتم المقارنة بين متوسطين مجتمعين مستقلين، المجتمع الأول هو المدخنين الحاليين والمجتمع الثاني هو المدخنين السابقين. ويدعي الباحث أن متوسط ضغط الدم الانبساطي للمدخنين الحاليين يختلف عن متوسط ضغط الدم الانبساطي للمدخنين السابقين. وأيضاً يمكن ملاحظة أن الباحث لم يحدد قيمة لأي من المتوسطين في المجتمعين. ويتم رفض فرضية العدم إذا كان متوسط ضغط الدم الانبساطي للعينة من المجتمع الأول أقل بكثير او أكبر بكثير من متوسط ضغط الدم الانبساطي للعينة من المجتمع الثاني، وهذه النتيجة تؤيد فرضية الباحث. ولا يمكن رفض فرضية العدم إذا كان متوسط ضغط الدم الانبساطي للعينة من المجتمع الأول قريب من متوسط ضغط الدم الانبساطي للعينة من المجتمع، وهي النتيجة لا تؤيد فرضية الباحث بل تؤيد فرضية العدم.

الخطوة الثانية: تحديد قيمة (α)

$$1 - \alpha = 0.95 \implies \alpha = 0.05$$

الخطوة الثالثة: تحديد إحصائية الاختبار

$$T^* = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

الخطوة الرابعة: تحديد قاعدة اتخاذ القرار (متى نرفض فرضية العدم؟)

نظراً لأن الاختبار ذو طرفين (منطقة رفض لفرضية العدم تقع في الطرف الأيسر من توزيع T ومنطقة رفض أخرى لفرضية العدم تقع في الطرف الأيمن من توزيع T)، لذا فإنه يوجد شرط وحيد لرفض فرضية العدم هو:

قيمة P-value أقل من قيمة $\left(\frac{\alpha}{2}\right)$، أي أقل من 0.025.

الخطوة الخامسة: اتخاذ القرار بناء على نتائج العينة.

Group Statistics

	Somking status	N	Mean	Std. Deviation	Std. Error Mean
diastolic blood pressure	Former	359	72.20	10.615	.560
	Current	142	70.30	10.309	.865

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t	df	Sig. (2-tailed)
diastolic blood pressure	Equal variances assumed	.163	.687	1.823	499	.069
	Equal variances not assumed			1.847	265.680	.066

كما ذكرنا سابقا فإن الجدول الثاني يحتوي على نتائج اختبارين، الأول يعنى باختبار اختلاف تباين المجتمعين والثاني يعنى باختبار T لاختلاف متوسطات المجموعتين. ويمكن صياغة فرضية الاختبار الأول على النحو التالي:

$$H_0: \sigma_s^2 = \sigma_{ns}^2$$
$$H_1: \sigma_s^2 \neq \sigma_{ns}^2$$

وبناء على نتائج اختبار ليفين، فإنه لا يمكن رفض فرضية العدم والتي تشير إلى تساوي تباين المجموعتين وذلك لأن $(P - value > 0.05)$ لاختبار F، وهذا يعني أنه لا يوجد اختلاف جوهري بين تباين المجتمعين. وفي هذه الحالة، يتم استخدام نتائج اختبار T الناتجة من افتراض تساوي تباين المجتمعين (السطر الأول من النتائج) لاتخاذ القرار في اختبار T حول متوسطات مجتمعين مستقلين.

ومن نتائج العينة فإن $(T^* = 1.823)$ و $\left(P - value = \frac{0.069}{2} = 0.0345\right)$.

وحيث أن $(P - value > 0.025)$، فإن الشرط الوحيد لم يتحقق. لذلك فإنه لا يمكن رفض فرضية العدم، وهذا يعني أن متوسط ضغط الدم الانبساطي للمدخنين الحاليين لا يختلف عن متوسط ضغط الدم الانبساطي للمدخنين السابقين.