الإحصاء - المحاضرة الثامنة – هندسة الروبوت

Statistics and probabilities- Lecture 8

Dr Fadi KHALIL

Doctor lecturer in statistics and programing

في هذه المحاضرة سيتم التطرق لـ

- فكرة اختبار الفرضيات Hypothesis testing
- مبدأ تصميم التجارب
- مفهوم محاكاة Monte CARLO
- استخدام محاكاة Monte Carlo لدراسة أداء مؤشر الاختبار

11-3- الاختبار أحادي وثنائي الجانب One tailed and wo tailed test :

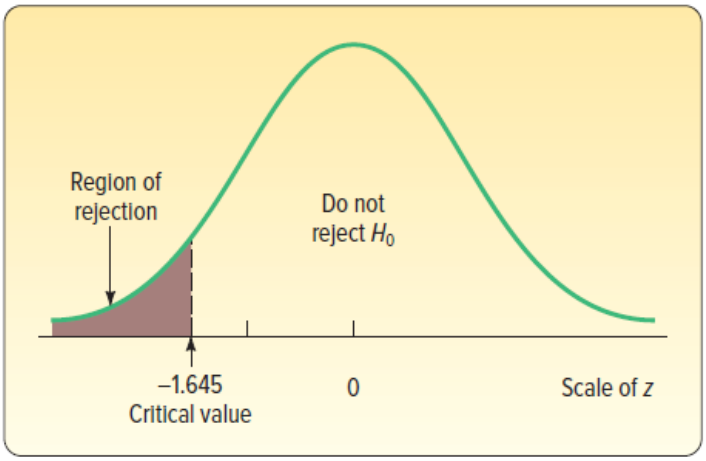
يلاحظ فيما سبق أنّ الفرضية البديلة تأخذ شكل عدم المساواة ولكن في بعض الأحيان تكون الفرضية البديلة من نوع أكبر او أصغر من، ويسمّى الاختبار عندها بالاختبار أحادي الجانب أي أنّ منطقة الرفض تكون من جانب واحد.

مثلاً لدى الشركات المصنّعة للسيارات الثقيلة التي تستخدم في البناء رغبة في شراء الإطارات التي يصل استخدامها العادي إلى 60000 ميل قبل التآكل، بالتالي سيتم رفض شحنة الإطارات التي يكون متوسط عمرها أقل بكثير من 60000 ميل، وبالطبع سيتم قبول شحنة الإطارات التي يكون عمر استخدامها 60000 فأكثر، وبالتالي قلق الشركات المصنعة للسيارات يتعلق باختبار فيما إذا كان عمر استخدام الإطارات أقل من 60000 ألف ميل. بالتالي يمكن صياغة الفرضية العدم والبديلة كالآتي:

$$H_0: \mu \geq 60000$$

$$H_1: \mu < 60000$$

ويمكن توضيح منطقة الرفض والقبول في الاختبار أحادي الجانب وفقاً للشكل الآتي:



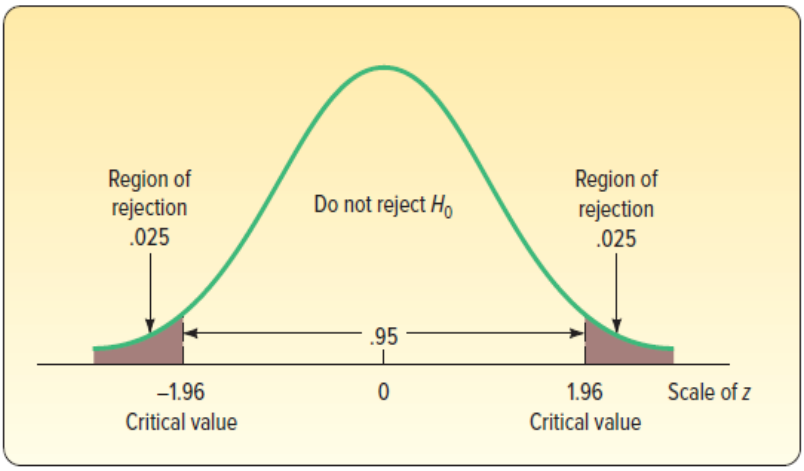
هنا يمكن القول أنّه يمكن رفض فرضية العدم فقط في الحالة التي يكون فيها متوسط عمر الإطارات أصغر من 60000 أي تقع منطقة الرفض في المنطقة المظللة على اليسار كما في الشكل السابق.

بينما لو كان اهتمام الشركة المصنعة للإطارات مثلاً في المثال السابق حول متوسط عمر الاستخدام فيما إذا كان 60000 ميل أم لا، يمكن صياغة الفرضية العدم والبديلة كالآتي:

$$H_0: \mu = 60000$$

$$H_1: \mu \neq 60000$$

هنا يلاحظ أنّه يتم رفض فرضية العدم سواء أكان متوسط عمر الإطارات أكبر أو أصغر بشكل جوهري من 60000، ولاخذ هذين الاتجاهين بعين الاعتبار في رفض فرضية العدم يتم تقسيم منطقة الرفض على قسمين على جانبي المنحنى كما في الشكل الآتي:



مثلاً لو تم الاختبار عند مستوى دلالة 5%، يتم تقسيم منطقة الرفض إلى قسمين متساويين مساحة كل منهما 0.025 ، أما مساحة عدم رفض فرضية العدم فتبلغ 0.95.

ولتوضيح اختبار الفرضيات من كلاالنوعين سيتم تناول الأمثلة الآتية:

مثال :

قام أحد المخابر باتباع خطة جديدة لزيادة الإنتاج الشهري في إحدى الشركات المختصّة بإنتاج محركات الروبوتات، وبفرض أنّ حجم الإنتاج يخضع للتوزيع الطبيعي بمتوسط 200 و تباين 16، وإذا أرادت هذه الشركة معرفة فيما إذا كان هناك تغير في الإنتاجية نتيجة اتباع الخطة جديدة قائمة على توظيف مجموعة من الكوادر فقامت بدراسة عينة لـ 50 شهر من الأعمال المنجزة خلال الأشهر الماضية فكان متوسط حجم الإنتاج 203.5، فما هو القرار المناسب لشركة المقاولات عند مستوى دلالة 1%:

الحل :

سيتم اتباع خطوات اختبار الفرضيات وأولها صياغة الفرضيات، وهنا يتضح أنّ غاية الشركة هو معرفة فيما إذا متوسط الإنتاجية يختلف عن 200، أي:

$$H_0: \mu = 200$$

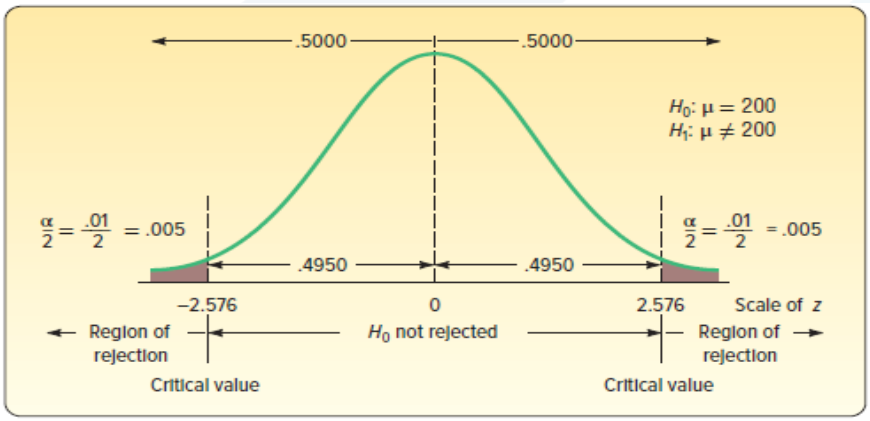
$$H_1: \mu \neq 200$$

الخطوة الثانية تتمثل ببناء مؤشر الاختبار :

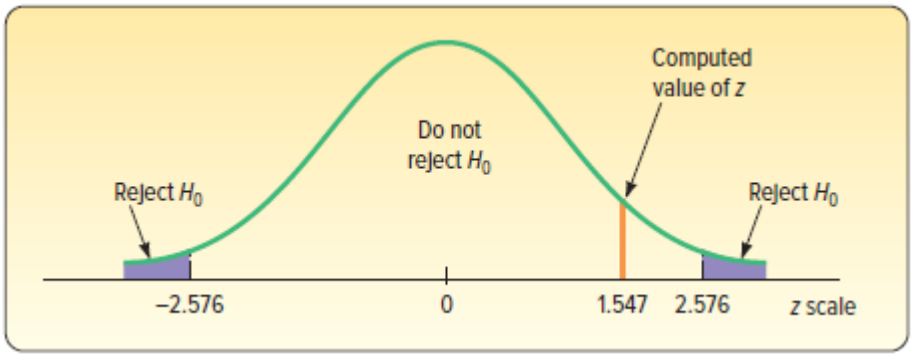
$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{203.5 - 200}{\frac{4}{\sqrt{50}}} = 1.547$$

وهي قيمة مؤشر الاختبار المحسوب.

الخطوة الثالثة تتمثل بمقارنة قيمة مؤشر الاختبار المحسوب مع القيمة المعيارية عند مستوى دلالة 1% والتي تبلغ 2.576، و 2.576- والشكل الآتي يساعد في استخراجها.



وبمقارنة أنّ القيمة المحسوبة Z 1.547 أصغر من القيمة المعيارية بالقيمة المطلقة 2.576 أو انّ القيمة المحسوبة 1.547 تقع ضمن المجال $[-2.576, +2.576]$ بالتالي لا يمكن رفض فرضية العدم (do not reject) .



وكنتيجة للاختبار السابق إنّ متوسط الإنتاجية لم يتغير نتيجة تطبيق الخطة الجديدة، وأنّ الاختلاف بين قيمة العينة 203.5 و قيمة المجتمع 200 هو اختلاف ناتج عن الصدفة.

مثال :

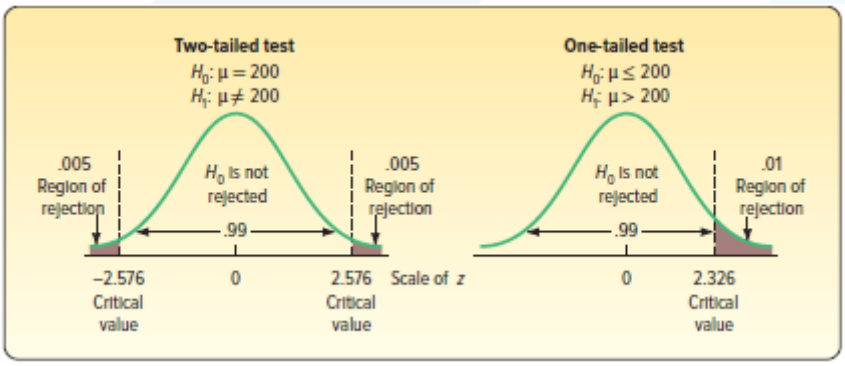
ليتم العودة للمثال السابق وفي حال أرادت الشركة معرفة فيما إذا كان هناك تحسن في إنتاجية الشركة نتيجة اتباع الخطة الجديدة. أي يمكن الاستنتاج أن تبعاً لتطبيق الخطة الجديدة تمّ زيادة إنتاجية الشركة عن 200؟ وهذا يختلف عن الحالة السابقة حيث أرادت الشركة معرفة فيما إذا كان هناك **تغير** في الإنتاجية نتيجة تطبيق الخطة الجديدة.

بالتالي الاختبار هنا هو اختبار أحادي الجانب one tailed test، يمكن صياغة الفرضية العدم والبديلة في هذه الحالة كما يأتي:

$$H_0: \mu \leq 200$$

$$H_1: \mu > 200$$

مع الإشارة إلى أنّ القيمة المعيارية critical value في الأختبار أحادي الجانب تختلف عن القيمة المعيارية في الاختبار ثنائي الجانب. ففي الحالة السابقة تمّ تقسيم منطقة الرفض إلى قسمين متساويين، أمّا في هذه الحالة فإنّ منطقة الرفض تتركز في جانب واحد على اليمين:



حيث يلاحظ أنّ القيمة المعيارية في هذه الحالة هي 2.326، والقرار هنا أيضاً يتمثل بعدم إمكانية رفض فرضية العدم لأنّ القيمة المحسوبة Z 1.547 أصغر من القيمة المعيارية 2.326.

11-4- اختبار الفرق بين متوسطي عينتين مستقلتين :

غالبًا ما يهتم المهندسون والعلماء بمقارنة طريقتين أو أسلوبين مختلفين لتحديد ما إذا كان أي منهما ينتج تأثيراً معنويًا على الظاهرة التي يتم ملاحظتها (أو ما يُسمّى الاستجابة). فمثلاً إذ تم تجريب تركيبتين للخلطة المعدنية لتصنيع الروبوتات، بحيث تمثل الظاهرة الملاحظة (أو الاستجابة) هنا هي وقت السرعة في تنفيذ المهام. الغرض من الدراسة هو تحديد ما إذا كانت التركيبة الجديدة للخلطة االمعدنية لها تأثير معنوي في تقليل وقت التنفيذ. في هذه الحالة، قام مطور المنتج (أو القائم على التجربة) باختيار 10 عينات اختبار بشكل عشوائي لتركيبة معينة و 10 عينات اختبار للتركيبة الأخرى. تم تنفيذ الأعمال على عينات الاختبار بترتيب عشوائي حتى تم تطبيق جميع العينات العشرين. وبالتالي إذا لوحظ فرق ذو دلالة إحصائية في التجربة عشوائية، يمكن أن يكون المجرب واثقًا في الاستنتاج بأن الاختلاف في معالجات تركيبة الخلطة المعدنية أدّى إلى اختلاف في أوقات التنفيذ.

هذه الحالة يتم الاستدلال الإحصائي على الفرق بين متوسطي مجتمعين عن طريق دراسة الفرق بين متوسطي عينتين مسحوبتين من هذين المجتمعين ومن ثمّ اختبار وجود فرق بينهما. ففي المثال السابق تمّ الاستدلال عن تأثير طريقة معالجة التركيبة المعدنية في سؤعة إنجاز المهام وذلك عن طريق اختيار عينتين عشوائيين ودراسة معنوية الفرق بينهما في وقت التنفيذ.

بفرض أن متوسطي مجتمعين μ_1، μ_2، يمكن تقديرهما عن طريق سحب عينتين بحجم n_1، n_2 من كل مجتمع على حدا $\bar{x}_1$، و $\bar{x}_2$ ، بالتالي، فإنّ تقدير الفرق بين متوسطي المجتمعين يمكن تقديره $\mu_1 - \mu_2$ بالفرق بين متوسطي العينتين $\bar{x}_1 - \bar{x}_2$

بناءً عليه، ولاختبار الفرق[1] بين متوسطي المجتمعين μ_1، μ_2 يمكن صياغة الفرضية العدم الآتية:

$$H_0 = \mu_1 - \mu_2 = \Delta_0$$

والفرضية البديلة :

$$H_1 = \mu_1 - \mu_2 \neq \Delta_0$$

بعد ذلك يمكن استخدام صيغة مؤشر الاختبار السابقة لتعميمها على حالة الفرق بين المتوسطين:

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

بحيث يتم استبدال μ_0 بـ $(\mu_1 - \mu_2)$، و قيمة مؤشر العينة $\bar{x}$ بـ $\bar{x}_1 - \bar{x}_2$ ، و الخطأ المعياري للمعاينة $\frac{\sigma}{\sqrt{n}}$ بـ $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ ، وذلك على افتراض أنّ العينتين مستقلتين بالتالي يصبح مؤشر الاختبار المستخدم لاختبار الفرق بين متوسطي مجتمعين:

$$Z_c = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

بحيث يخضع هذا المؤشر للتوزيع الطبيعي المعياري $N(0,1)$.

وكما تمّ عرضه في اختبار الفرضيات لمتوسط عينة واحدة *one sample test* فمن الممكن أن يكون الاختبار ثنائي الجانب أو أحادي الجانب والجدول الآتي يلخص هذه الحالات مع القيمة المعيارية لاتخاذ القرار:

القيمة المعيارية *critical values*	الفرضيات البديلة *alternative*	الاختبار
$Z_{\frac{\alpha}{2}}$	$H_1 = \mu_1 - \mu_2 \neq \Delta_0$	ثنائي الجانب *two sided*
Z_α	$H_1 = \mu_1 - \mu_2 > \Delta_0$	أحادي الجانب *one sided*
$-Z_\alpha$	$H_1 = \mu_1 - \mu_2 < \Delta_0$	أحادي الجانب *one sided*

- يتم رفض فرضية العدم عموماً إذا كانت قيمة مؤشر الاختبار المحسوب Z_c أكبر من القيمة المطلقة للقيمة المعيارية.

[1] من الجدير للذكر أنّه يمكن اختبار ان يكون الفرق بين متوسطي المجتمعين يساوي الصفر، أي يتم اختبار جوهرية أو معنوية الفرق بين متوسطي المجتمعين وتصبح قيمة Δ_0 مساوية للصفر ، $H_0 = \mu_1 - \mu_2 = 0 \ or \ \mu_1 = \mu_2$

مثال :

بالعودة للمثال الأول المتعلق بتطوير الخلطة المعدنية لتحسين وقت التنفيذ، حيث تمّ اختبار تركيبتين، الأولى وهي التركيبة المعيارية،أمّا الثانية فهي التركيبة الجديدة والتي يجب أن تزيد سرعة التنفيذ (أو تقلل وقت التنفيذ). من المعروف من خلال التجربة أنّ الانحراف المعياري لوقت التنفيذ هو 8 دقائق مهما كان نوع الخلطة المستخدمة في تصنيع الروبوت فإذا تمّ اختيار 10 عينات من التركيبة الاولى المعيارية و 10 عينات من التركيبة الجديدة، وكان متوسط وقت التنفيذ للعينات من التركيبة الأولى هو 36 دقيقة بينما كان متوسط وقت التنفيذ للعينات من التركيبة الجديدة هو 27 ساعة، فما هي النتيجة التي يمكن استخلاصها من قبل مطوّر التركيبة الجديدة، وذلك عند مستوى دلالة 0.05؟

الحل :

لمعرفة أثر تركيبة الخلطة الجديدة في تحسين وقت التنفيذ يجب مقارنة الوقت عند استخدام كلا التركيبتين المعيارية والجديدة، لذلك تمّ سحب عينتين ومقارنة وقت التنفيذ في كلا العينتين.

بفرض أنّ μ_1 متوسط وقت التنفيذ عند استخدام التركيبة المعيارية، وأنّ μ_2 متوسط وقت التنفيذ عند استخدام التركيبة الجديدة، وفي حال التحسن نتيجة لاستخدام التركيبة الجديدة يجب أن يصبح $\mu_1 - \mu_2 > 0$،

بالتالي يمكن صياغة الفرضية العدم والبديلة كالآتي :

$$H_0 = \mu_1 - \mu_2 = 0$$
$$H_1 = \mu_1 - \mu_2 > 0$$

وبحساب مؤشر الاختبار تحت شرط ففرضية العدم :

$$Z_c = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(36 - 27) - (0)}{\sqrt{\frac{(8)^2}{10} + \frac{(8)^2}{10}}} = 2.52$$

عند مستوى دلالة 0.05، $Z_{0.05}$ تساوي 1.96 وبالتالي $Z_c > Z_{0.05}$ أي يتم رفض الفرضية العدم والاستنتاج بأنّ التركيبة الجديدة ساعدت بشكل جوهري في تحسين وقت التنفيذ.

12 – مفهوم تصميم التجارب *Experiment*

التجارب هي جزء طبيعي من عملية صنع القرار الهندسي والعلمي. لنفترض ، على سبيل المثال ، أن مهندس ميكاترونيكس يحقق في تأثيرات طرق المعالجة المختلفة على متوسط استهلاك الوقود في محركات الدرجات النارية.

ستتألف التجربة من تكوين عدة عينات اختبار باستخدام كل من طرق المعالجة المقترحة ثم اختبار استهلاك الوقود. يمكن استخدام البيانات من هذه التجربة لتحديد طريقة المعالجة التي يجب استخدامها لتوفير أقصى متوسط لقوة الانضغاط.

إذا كانت هناك طريقتان فقط من طرق المعالجة ذات الأهمية، فيمكن تصميم هذه التجربة وتحليلها باستخدام طرق الفرضية الإحصائية لعينتين كما تمّ تقديمه في المبحث السابق. أي أن المجرب لديه عامل واحد مثير للاهتمام (نوع تركيبة الطلاء) وهناك طريقتان فقط (أي هنا تجربة أحادية العامل *one factor* ومستويين فقط). إذا كان المجرب مهتمًا بتحديد طريقة المعالجة التي تنتج أقصى استهلاك للوقود، فيمكن تحديد عدد العينات المراد اختبارها، وي ثمّ استخدام اختبار Z أو t لعينتين *two samples* وذلك لتحديد ما إذا كانت الطريقتين مختلفتان . ولكن العديد من التجارب أحادية العامل (*one factor*) لا تفي بغرض الدراسات التي تتضمن أكثر من مستويين من العامل.

على سبيل المثال ، قد يرغب المهندس في التحقق من خمس طرق معالجة مختلفة. وفيما يأتي سيتم توضيح كيف يمكن استخدام مفهوم تحليل التباين (*Analysis of Variance, ANOVA*) لمقارنة الطرق عندما يكون هناك أكثر من مستويين لعامل واحد.

1-12- تصميم التجارب الهندسية *Designing Engineering Experiments*:

تعد تقنيات التصميم التجريبي القائمة على الإحصاء مفيدة بشكل خاص في عالم الهندسة لحل العديد من المشكلات المهمة:، واكتشاف ظواهر أساسية جديدة يمكن أن تؤدي إلى تسويق وتطوير تكنولوجيا ومنتجات جديدة. على سبيل المثال ، في حال تطوير عملية ما *process* جديدة، يمكن تحديد المتغيرات *variables* التي يمكن التحكم فيها والتي تؤثر في هذه العملية، مثل درجة الحرارة والضغط ومعدل التغذية. وبالتالي باستخدام تصميم التجارب *design expirement*، يمكن للمهندسين المختصين تحديد أي مجموعة من هذه المتغيرات لها التأثير الأكبر على أداء هذه العملية. حيث يمكن أن تؤدي نتائج هذه التجربة إلى تحسين العائد، تخفيف التقلبات *variability*، تقليص وقت التصميم والتطوير، وتقليص تكلفة الإنتاج.

في هذا السياق ، تعد طرق التصميم التجريبية *expirement design* مفيدة أيضًا في أنشطة التصميم الهندسي التي يتم خلالها تطوير المنتجات الجديدة وتحسين المنتجات الحالية. حيث تتضمن بعض التجارب المصممة إحصائياً في في الإطار الهندسي:

- تقييم ومقارنة مكونات التصميم الأساسية
- اختيار معايير التصميم التي تؤثر في أداء المنتج، والتي تؤدي في عمل التصميم بشكل جيّد.

عادة ما يتسم تصميم التجارب بصفة التتابع. أي أن التجربة الأولى مع نظام معقد (مثلاً عملية تصنيع) تحتوي على العديد من المتغيرات التي يمكن التحكم فيها، وغالبًا ما تكون تجربة فحص مبدئية مصممة لتحديد تلك المتغيرات الأكثر أهمية. ومن ثمّ تُستخدم التجارب اللاحقة لتنقيح هذه المعلومات وتحديد التعديلات المطلوبة لهذه المتغيرات

الحرجة لتحسين العملية. أخيراً، الهدف من المجرب هو التحسين ، أي تحديد مستويات المتغيرات الحرجة التي تؤدي إلى أفضل أداء للعملية.

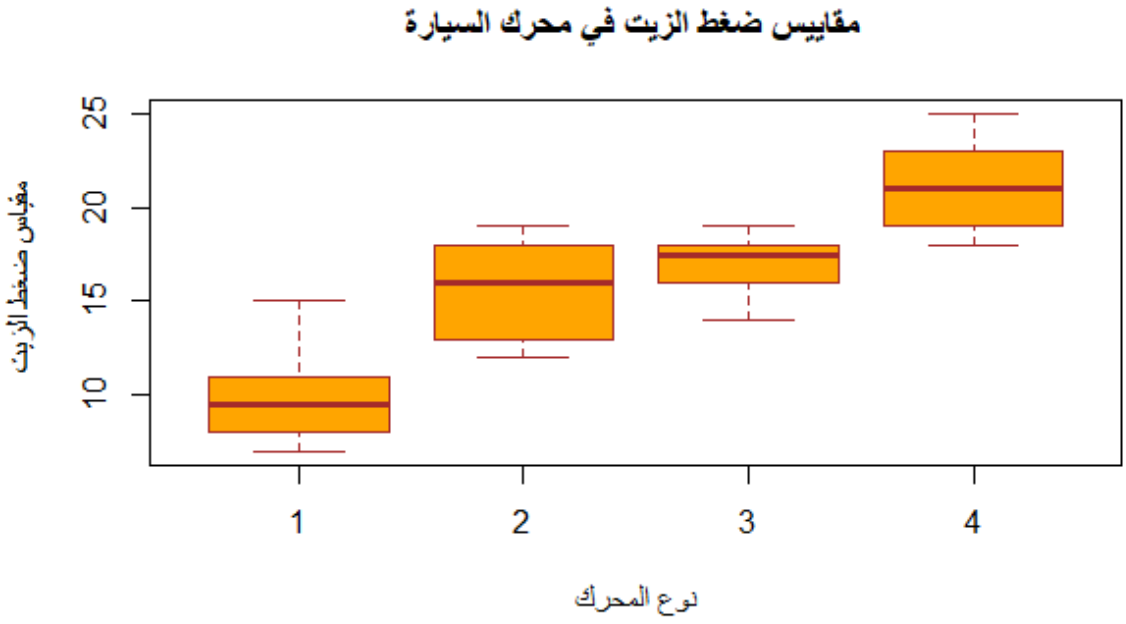
ولكي لا يبقى شرح مفهوم التجارب في الإطار النظري فقط، ربما أفضل طريقة هو تقديم مثال عملي:

مثال:

بفرض أنّ شركة لتصنيع السيارات قامت بقياس ضغط الزيت في 24 محرك للسيارات موزعة ضمن أربع أنواع أو فئات بحيث تمّ من أجل كل نوع أخذ 6 قياسات بأوقات مختلفة. وكانت المعطيات كالآتي:

نوع المحرك	قياسات الضغط (كيلو باسكال) بالعشرات						الإجمالي	المتوسط
	1	2	3	4	5	6		
1	7	8	15	11	9	10	60	**10**
2	12	17	13	18	19	15	94	**15.67**
3	14	18	19	17	16	18	102	**17**
4	19	25	22	23	18	20	127	**21.17**
المجموع							383	**15.96**

من المفيد مبدئياً عرض الإحصاءات الخاصة بضغط الزيت وذلك بالنسبة لكل فئة من فئات التصدع ، وهذا ما يعبر عنه بأسلوب الرسم البياني *Box plot* حيث يعبر الخط المتواجد داخل الصندوق عن قيمة المتوسط الحسابي لكل فئة، وكذلك تعبر الخطوط الافقية العلوية والسفلية الممتدة على جانبي الصندوق عن أكبر قيمة واصغرقيمة على الترتيب، أمّا حدود الصندوق العلوية والسفلية تعبر عن قيمة الربيع الثالث والأول على الترتيب.



ومن الواضح في الرسم البياني أنّ الفئة الأخيرة المقابلة للنوع 4 تشكل أعلى نسبة ضغط زيت حيث يتوضع الصندوق المرافق لها في أعلى مكان مقارنة بالصناديق الأخرى. على أيّة حال هناك الكثير من المعلومات يمكن استخراجها من الرسم البياني، ولكن السؤال الأهم يتعلق حول العلاقة بين نوع المحرك مع ضغط الزيت، بمعنى آخر هل تتوزع قيم ضغط زيت المحرك بشكل عشوائي وفقاً لنوع المحرك، أم هناك توزيع لقيم ضغط االزيت يتوافق مع نوع المحرك.

ملاحظة: الحالة هنا تشبه المبحث السابق من ناحية اختبار وجود فرق في وقت التجفف تبعاً لتركيبة الطلاء، حيث تم حينها استخدام اختبار Z لاختبار الفرق بين عينتين مستقلتين، نظراص لأنّ العامل *factor* هنا يتضمن مستويين أو تركيبتين فقط، أمّا في هذا المبحث فالعامل يتضمن أكثر من مستوى (4 مستويات للتصدع)، وهنا تقتضي الحالة اتسخدام ما يُسمّى تحليل التباين *Analysis of variance ANOVA* لاختبار معنوية أو جوهرية اختلاف قيم ضغط الزيت وفقاً لنوع المحرك. ونظراً لوجود عامل واحد فقط يسمى تحليل التباين ***One way ANOVA***.

قبل البدء بتحليل التباين لا بدّ من الإشارة إلى ضرورة توفر شرطين وهما:

- أن تكون العينات مستقلة عن بعضها
- أن يكون التوزيع الإحتمالي لقيم المتغير التابع متجانس بين فئات المتغير المستقل.

12-2- تحليل التباين *Analysis of Variance ANOVA*:

بفرض وجود a مستوى لعامل ما وحيد تم تحديدها مسبقاً من قبل المجرب ضمن تجربة معينة وكان هدف هذه التجربة هو مقارنة تأثير مستويات هذا العامل على متغير آخر عشوائي يُسمى *response* . حيث تمّ مشاهدة n قيمة

لكل مستوى من مستويات العامل المدروس وذلك بترتيب عشوائي *completely randomized design (CRD)*. وليكن y_{ij} المشاهدة j، للمستوى i من مستويات العامل. ويمكن إظهار ذلك في الجدول الآتي:

مستويات العامل	المشاهدات				الإجمالي	المتوسط
	1	2	$\cdots$	n		
1	y_{11}	y_{12}	$\cdots$	y_{1n}	$y_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	$\cdots$	y_{2n}	$y_{2.}$	$\bar{y}_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
α	$y_{\alpha 1}$	$y_{\alpha 2}$	$\cdots$	$y_{\alpha n}$	$y_{\alpha.}$	$\bar{y}_{\square.}$
					$y_{..}$	$\bar{y}_{..}$

من توزيع قيم المتغير y_{ij} في الجدول السابق يمكن القول انّ كل قيمة من قيمة المتغير هي عبارة عن ثلاثة أنواع من العوامل، **النوع الأول** هي عوامل مشتركة بين جميع مستويات العامل المدروس ويرمز له بالرمز μ، **النوع الثاني** خاص بكل مستوى من مستويات العامل المدروس ويرمز له بالرمز τ_i ، أمّا **النوع الثالث** فهو يعبر عن العوامل العشوائية التي لا يمكن حصرها والتي تؤثر في قيم y_{ij} ويمز لها بالرمز ε_{ij}. بمعنى آخر يمكن صياغة المعادلة الآتية:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

بحيث أنّ $i = 1,2, \dots . a$ و $j = 1,2, \dots . n$

ويمكن كتابة المعادلة السابقة بشكل آخر كما يأتي:

$$y_{ij} = \mu_i + \varepsilon_{ij}$$

حيث: $\mu_i = \mu + \tau_i$

إذا الهدف من هذه القياسات هو دراسة تأثير مستويات العامل المدورس المحدد مسبقاً من قبل المجرب على المتغير العشوائي y_{ij} وهذا ما يسمى بـ*fixed effect model*.

وبما إنّ العامل من النوع الثاني τ_i خاص بكل مستوى من مستويات العامل المدروس ، فإنّ السؤال المطروح هنا أنّه لو كانت قيم τ_i متساوية بين جميع مستويات a، أي لو كانت قيم μ_i متساوية بين جميع المتسويات، فهل يبقى هناك تأثير لمستويات هذا العامل على قيم y_{ij}؟ الجواب هو قطعا لا .....

إذا المشكلة هنا هي اختبار للفرضيات بمعنى آخر هو اختبار الفرضية العدم الآتية:

$$H_0: \mu_1 = \mu_2 = \cdots \mu_a$$

بينما الفرضية البديلة هي :

$$H_1: \mu_1 \neq \mu_2 \neq \cdots \mu_a$$

وللتذكير هنا أن هذا الاختبار يشبه اختبار تساوي متوسطي عينتين والذي تمّ تقديمه في المبحث السابق، ولكن هنا يتم اختبار التساوي بين أكثر من متوسطين.

لبناء مؤشر الاختبار سيتم اتباع الخطوات الآتية :

أولاً : حساب التباين بين المستويات، *Between*، MS_B :

$$MS_B = \frac{SS_B}{a-1}$$

حيث يعبر a عن عدد مستويات العامل المدروس، و $SS_{Treatments}$ عن مجموع مربعات الفروقات بين المستويات Sum of squares of treatments ويحسب كما يأتي :

$$SS_B = n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2$$

$$\bar{y}_{i.} = \frac{y_{i.}}{N}, \qquad y_{i.} = \sum_{j=1}^{n} y_{ij}$$

$$\bar{y}_{..} = \frac{y_{..}}{N}, \quad y_{..} = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}$$

حيث، N هي عدد المشاهدات الإجمالي ويساوي $n * a$

ثانياً: حساب تباين ضمن المجموعات *Within* ، MS_W

$$MS_W = \frac{SS_W}{a(n-1)}$$

حيث يعبر SS_W هن مجموع مربعات الفرق بين القيم y_{ij} ومتوسط كل مستوى من مستويات العامل المدروس

$$SS_W = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2$$

مع الإشارة إلى انّ مجموع مربعات الفرق بين القيم والمتوسط العام $\bar{y}_{..}$ يسمى بـ SS_T.

$$SS_T = \sum_{i=1}^{\alpha}\sum_{j=1}^{n}\left(y_{ij} - \bar{y}_{..}\right)^2$$

وهو يساوي:

$$SS_T = SS_B + SS_W$$

ثالثاً: حساب مؤشر الاختبار :

$$F_0 = \frac{SS_B/a-1}{SS_W/a(n-1)} = \frac{MS_B}{MS_W}$$

حيث يخضع مؤشر الاختبار F_0 لتوزيع فيشر الاحتمال بدرجتي حرية $\alpha - 1$، و $\alpha(n-1)$، وبالتالي لاتخاذ القرار برفض أو عدم رفض فرضية العدم يتم مقارنة القيمة المحسوبة F_0 بالقيمة النظرية أو المعيارية $F_{\alpha,a-1,a(n-1)}$:

فإذا كان: $F_0 > F_{\alpha,a-1,a(n-1)}$ يتم رفض فرضية العدم وبالتالي يوجد أثر لمستويات العامل في قيم المتغير y_{ij} ، امّا في حال كان $F_0 < F_{\alpha,a-1,a(n-1)}$ حينها لا يمكن رفض فرضية العدم وبالتالي مستويات العامل لا تختلف جوهرياً في تأثيرها على قيم المتغير y_{ij}.

يمكن تلخيص ما سبق في الجدول الآتي :

مصدر التباين *variation*	مجموع المربعات *sum of squares*	درجات الحرية	متوسط المربعات	F_0
بين المجموعات *Between*	SS_B	$a-1$	$MS_B = \frac{SS_B}{a-1}$	$\frac{MS_B}{MS_W}$
ضمن المجموعات *Within*	SS_W	$a(n-1)$	$MS_W = \frac{SS_W}{a(n-1)}$	
الكلي *Total*	SS_T	$an-1$		

وكمثال تطبيقي على تحليل التباين سيتم العودة لبيانات المثال السابق عن دراسة العلاقة بين ضغط االزيت و نوع المحرك في 24 مشاهدة، ولاختبار وجود اختلاف جوهري في توزع قيم ضغط الزيت وذلك تبعاً لنوع المحرك وذلك عند مستوى دلالة 0.01:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

بينما الفرضية البديلة هي :

$$H_1: \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$$

بحساب SS_T :

$$SS_T = \sum_{i=1}^{\alpha}\sum_{j=1}^{n}\left(y_{ij} - \bar{y}_{..}\right)^2 = 512.96$$

وكذلك SS_B:

$$SS_B = n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2 = 382.79$$

وبالتالي يمكن حساب مجموع مربعات الفروقات ضمن *within* المجموعات SS_W:

$$SS_W = SS_T - SS_B = 512.96 - 382.79 = 130.17$$

بالتالي يمكن بناء جدول تحليل التباين كما يأتي:

مصدر التباين *variation*	مجموع المربعات *sum of squares*	درجات الحرية	متوسط المربعات	F_0
بين المجموعات *Between*	382.79	$4 - 1 = 3$	$MS_B = \frac{382.79}{3} = 127.60$	$\frac{\mathbf{127.6}}{\mathbf{6.51}} = \mathbf{19.6}$
ضمن المجموعات *Within*	130.17	$4(6-1) = 20$	$MS_W = \frac{130.17}{20} = 6.51$	
الكلي *Total*	512.96	$4 * 6 - 1 = 23$		

بمقارنة القيمة المحسوبة $F_0 = \mathbf{19.6}$ مع القيمة الجدولية عن مستوى دلالة 0.01 $F_{\alpha,a-1,a(n-1)} = 4.94$ $F_{0.01,3,20}$ يلاحظ أنّ $F_0 > F_{0.01,3,20}$ وبالتالي يمكن رفض فرضية العدم والتفسير بانّ هناك اتجاه قوي للقول بأنّ ضغط الزيت يختلف اختلافاص جوهريا بين أنواع المحركات المختلفة.

13- محاكاة مونت كارلو Monte Carlo simulations:

أغلب طرق التقدير واختبار الفرضيات تملك خصائص معروفة فقط بشكل تقريبي، و من الناحية العملية تطبيق التوزيع الاحتمالي النظري الدقيق قد لا يكون موثوق في العينات الصغيرة لتطبيق طرق التقدير أو اختبار الفرضيات.

حيث كلما كانت العينة أقل حجماً، كلّما كان من الصعب معرفة فيما إذا كانت الطرق التقريبية كافية لاعتمادها في التقدير واختبار الفرضيات.

لذلك من الضروري دراسة خصائص المقدّرات (أو مؤشرات الاختبار) في العينات الصغيرة أو ما يسمى (Finite sample properties) وذلك عن طريق أسلوب محاكاة *Monte Carlo*.

و تعرّف محاكاة *Monte Carlo* *بأنّها* عبارة عن تقنية رياضية مؤتمتة تسمح بتحليل نسبة الخطأ في الدراسات الكمية وفي اتخاذ القرار. حيث تقوم بتوليد عدد كبير من البيانات العشوائية انطلاقاً من مولّد بيانات ما Data generator process (DGP). عموماً، تستخدم هذه التقنية في نواحي متعددة، على سبيل المثال لا الحصر، في المالية، إدارة المشاريع، الطاقة، التصنيع، الهندسة، دراسات التنمية، التأمين، النقل و البيئة. حيث إن الخاصية الرئيسة لتقنية محاكاة *Monte Carlo* هي بأنها تظهر احتمال الحصول على نتائج معينة انطلاقاً من سيناريوهات مختلفة، الأمر الذي يسمح بتقدير نسبة الخطر أو عدم التأكد في النموذج المدروس. في هذا الإطار كثيراً ما تُستخدم محاكاة *Monte Carlo* لدراسة أداء و خصائص مختلف الاختبارات الإحصائية المقترحة و لا سيّما في العينات المحدودة أو الصغيرة. كذلك تستخدم هذه التقنية بدراسة جودة تقريب التوزيع الاحتمالي النظري.

13-1- استخدام محاكاة مونت كارلو في دراسة أداء مؤشر الاختبار (size and power test) :

كما ذكرنا سابقاً يمكن دراسة أداء مؤشر الاختبار في العينات الصغيرة عن طريق مقياسين، مستوى دلالة الاختبار (Size)، وقوة الاختبار (Power).

مستوى دلالة الاختبار (Size):

في الواقع يكون النموذج الحقيقي أو مولّد البيانات الحقيقي (DGP) غير معروف، وكذلك إنّ التوزيع الاحتمالي الحقيقي لمؤشر الاختبار غير معروف أيضاً. و عليه، يُسمّى التوزيع التقريبي F_{asy} لمؤشر الاختبار بالتوزيع الإحتمالي الاسمي. اعتماداً على هذا التوزيع نقوم بتحديد قيمة معيارية (c).

$$\alpha = 1 - F_{asy}(c)$$

المقارنة بين قيمة مؤشر الاختبار المحسوبة Z_c انطلاقاً من بيانات العينة مع القيمة الجدولية تسمح إما برفض أو عدم رفض فرضية العدم.

في سياق المحاكاة يكون النموذج الحقيقي أو (DGP) الحقيقي معلوم، و طالما أنّه يتم توليد عدد كبير من البيانات أكثر من مرة، وفي كل مرة يتم حساب مؤشر الاختبار فإنّه يتم الحصول في النهاية على قيم كثيرة لمؤشر الاختبار المدروس. و بالتالي يتم معرفة التوزيع الحقيقي لمؤشر الاختبار F_{μ_0} (كلما كان عدد المرات كبيراً كلما كان التوزيع الاحتمالي التجريبي دقيق). وهكذا يمكن حساب احتمال الخطأ من النوع الأول بشكل دقيق ويرمز له

$$RP = 1 - F_{\mu_0}(c)$$

الفرق بين α و RP يطلق عليها اختلال الاختبار أو (Test distorsion) :

$$ERP = RP - \alpha$$

قوة الاختبار(*Power*):

وهو يعبر عن قدرة او احتمال أن يقوم مؤشر الاختبار برفض فرضية العدم عندما تكون غير صحيحة. وهو يساوي واحد ناقص احتمال الخطأ من النوع الثاني. يفيد هذا المفهوم في التفضيل بين مؤشري اختبار يملكان مستوى دلالة متساوي، وبالتالي يتم اختبار مؤشر الاختبار الأكثر قوة. وتجدر الإشارة إلى أن قوة الاختبار تزداد مع ازدياد عدد المشاهدات.

ملاحظة 1:

نعود ونذكر الغاية من محاكاة *Monte Carlo* هو معرفة فيما إذا كان مؤشر الاختبار يؤدي بشكل جيّد في العينات الصغيرة أو معرفة فيما إذا كان التوزيع النظري (المفترض) يمكن أن يطبق في العينات الصغيرة. بالتالي، يُستفاد من حساب ERP لمعرفة:

إذا كان $ERP = 0$، فهذا يدل على أن الاختبار جيّد في العينات الصغيرة، وإن التوزيع التقريبي يمكن أن يُطبّق بشكل جيّد في العينات الصغيرة.

أمّا إذا كان ERP أكبر أو أصغر من الصفر فهذا يدل على أنّ مستوى الدلالة الاسمي يختلف عن مستوى الدلالة التجريبي، أي أنّ الاحتمال الاسمي للخطأ من النوع الأول يختلف عن الاحتمال التجريبي للخطأ من النوع الأول (بعبارة أخرى، إنّ الاحتمال النظري للحصول على قيمة أكبر من قيمة مؤشر الاختبار النظرية، يختلف عن الاحتمال التجريبي للحصول على قيمة أكبر من قيمة مؤشر الاختبار النظرية).

ملاحظة 2: و هنا لا بد من التنويه أنه لا يجب الخلط بين مقياس P-value الذي يعتمد على التوزيع النظري لحساب احتمال الحصول على قيمة أكبر من مؤشر الاختبار المحسوب، وبالتالي تتم مقارنته مع مستوى الدلالة الاسمي (1%، 5%، أو 10%) لرفض أو عدم رفض فرضية العدم، و بين قيمة *RP* (*real probability*) التي تعتمد على التوزيع التجريبي (المقدّر عن طريق المحاكاة والمكوّن من عدد كبير جداً من قيم مؤشر الاختبار المحسوب) خلال مقارنتها مع مستوى الدلالة الاسمي (1%، 5%، أو 10%) من أجل دراسة جودة الاختبار في العينات الصغيرة، أو معرفة فيما إذا كان هناك تطابق بين التوزيع النظري أو التجريبي.

ملاحظة 3: إن مستوى دلالة الاختبار (Size) يقيس أداء مؤشر الاختبار عندما تكون فرضية العدم صحيحة، ولكن لا يقيس أداءه عندما تكون هذه الفرضية خاطئة. و بالتالي، إنّ مستوى دلالة الاختبار (Size) هو خاصية لازمة أو ضرورية للدلالة على أداء مؤشر الاختبار، ولكنّها غير كافية. حيث من الضروري أيضاً قياس قوة الاختبار (Power) أو احتمال رفض فرضية العدم وهي خاطئة.

في هذا السياق، تسمح محاكاة *Monte Carlo* كذلك بتوليد عدد كبير من قيم مؤشر الاختبار ولكن هذه المرة مولّد بيانات (DGP) يقوم على فرضية ابتدائية خاطئة أو فرضية بديلة صحيحة. و بالتالي يتم حساب عدد مرات رفض فرضية العدم، والتي يجب أن يبلغ احتمالها $1-\beta$. حيث β ترمز إلى احتمال الخطأ من النوع الثاني (أي احتمال الحصول على قيمة مؤشر اختبار أصغر من القيمة الجدولية لمؤشر الاختبار).

مثال :

لدراسة أداء مؤشر اختبار حول متوسط العينة واحدة Z-test، بمعنى آخر لقياس كل من Size و Power لمؤشر الاختبار، يفترض أنه تمّ توليد عينة خاضعة للتوزيع الطبيعي ، وفقا لسيناريوهين:

1. فرضية العدم صحيحة، سيتم توليد متغير x ، بمتوسط $\mu = 500$، و انحراف معياري $\sigma = 100$، أي $x \sim N(500, 100)$، والهدف هنا هو دراسة *size* لمؤشر الاختبار *Z-test*.
2. فرضية العدم خاطئة ، سيتم توليد متغير x ، بمتوسط $\mu \neq 500$، و انحراف معياري $\sigma = 100$، أي $x \sim N(500, 100)$، والهدف هنا هو دراسة *Power* لمؤشر الاختبار *Z-test*.

في كلتا الحالتين سيتم اختبار الفرضية العدم:

$$H_0: \mu = 500$$

مقابل الفرضية البديلة:

$$H_1: \mu \neq 500$$

سيتم الاعتماد على لغة برنامج *R* للحسابات الإحصائية والرسوم، وهو برنامج مفتوح المصدر يتمتع بالمرونة من ناحية بناء وتجميع أكواد مفتوحة المصدر وفقا لنظام *GNU Unix*.

حيث سيتم استخدام البرنامج *R* لتوليد البيانات وفقاً للسيناريوهات الموضوعة، ولتنفيذ محاكاة *Monte Carlo* لحساب *Size*، و *Power* لمؤشر اختبار موسط عينة واحدة *One sample Z-test*.

كذلك سيتم تنفيذ المحاكاة وفقاً لحجوم عينات مختلفة ومرات متعدة من المحاكاة:

```
n <- c(20, 100, 150, 200, 500, 1000, 2000); # size of samples

sim <- c(50, 500, 1000, 2000, 5000, 10000,20000); # number of replicates
```

وفيما يأتي البرنامج المستخدم :

```
# function for Monte Carlo simulation
MC_sim <- function (alpha, mu0, sigma, size, test_type)
{
library(BSDA)
row<- 0
col <- 0
p.hat <- matrix(NA,length(n), length(sim))
for (i in n){
 row<-row+1
  for (j in sim) {
  col<-col+1
  p <- numeric (j) # storage for p-values
  for (rep in 1:j){
   if (size==0) {x <- rnorm (n, rnorm (1, 800, sigma), sigma)}
   else if (size==1) {x <- rnorm (n, mu0, sigma)}
 ztest <- z.test (x, alternative= test_type, mu = mu0, sigma.x=sigma)
 p[rep] <- ztest$p.value
  }
 p.hat[row,col] <- mean(p < alpha )
}
 col<- 0
}
return(p.hat)
}
n <- c(20, 100, 150, 200, 500, 1000, 2000); # size of samples;
sim <- c(50, 500, 1000, 2000, 5000, 10000,20000); # number of replicates
# Excuting th function
sim_results <- MC_sim(0.05, 500, 100, 1, "two.sided")
colnames(sim_results)<-sim
rownames (sim_results) <- n
print (sim_results)
# plotting 3D graph
persp(n, sim, sim_results, main="Monte carlo simulation",  zlab = "sim_results",

    theta = 30, phi = 15,    col = "blue", shade = 0.4)
```

بعد تنفيذ البرنامج السابق تمّ الحصول على النتائج الآتية :

Size output							
	50	***500***	***1000***	***2000***	***5000***	***10000***	***20000***
20	*0.06*	*0.07*	*0.057*	*0.046*	*0.0468*	*0.0502*	*0.05005*
100	*0.08*	*0.054*	*0.059*	*0.053*	*0.0484*	*0.0512*	*0.04945*
150	*0.08*	*0.06*	*0.071*	*0.056*	*0.049*	*0.0516*	*0.04975*
200	*0.02*	*0.038*	*0.051*	*0.045*	*0.0518*	*0.0523*	*0.04765*
500	*0.1*	*0.068*	*0.049*	*0.0535*	*0.0486*	*0.0524*	*0.05045*
1000	*0.06*	*0.034*	*0.053*	*0.045*	*0.0504*	*0.0491*	*0.04895*
2000	*0.04*	*0.036*	*0.051*	*0.059*	*0.054*	*0.051*	*0.0504*
2000	*0.04*	*0.036*	*0.051*	*0.059*	*0.054*	*0.051*	*0.0504*

Power output							
	50	**500**	**1000**	**2000**	**5000**	**10000**	**20000**
20	1	0.99	0.983	0.982	0.9854	0.9813	0.98335
100	1	0.986	0.98	0.9785	0.9814	0.9847	0.9842
150	0.92	0.984	0.979	0.983	0.9842	0.981	0.9835
200	1	0.984	0.991	0.981	0.983	0.9839	0.98335
500	0.98	0.978	0.984	0.987	0.9808	0.9825	0.98395
1000	1	0.988	0.981	0.9825	0.9832	0.9833	0.9822
2000	0.96	0.97	0.989	0.9835	0.9836	0.9817	0.98255