الفصل الرابع
توليد أرقام عشوائية

خواص الأرقام العشوائية:

- التوزيع المتساوي
- الاستقلال

كل رقم عشوائي R_i هو عبارة عن عينة مستقلة مسحوبة من توزيع متساوي مستقل بين 0 و 1.

$$f(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & Otherwise \end{cases}$$

القيمة المتوقعة لكل R_i هي

$$E(R)=\int_0^1 x\,dx=\left.\frac{x^2}{2}\right|_0^1=\frac{1}{2}$$

وتباين

$$V(R)=\int_0^1 x^2 dx-[E(R)]^2=\left.\frac{x^3}{3}\right|_0^1-\left(\frac{1}{2}\right)^2=\frac{1}{12}$$

من الخواص المهمة للتوزيع المتساوي والإستقلال هي:

1) إذا قسمت الفترة $(0,1)$ إلى n فئة ذات أطوال متساوية فإن عدد المشاهدات المتوقعة في كل فئة هي N/n حيث N عدد المشاهدات الكلي.

2) إحتمال مشاهدة قيمة في فترة معينة مستقل عن قيمة القيم السابقة.

توليد ارقام شبه عشوائية:
طريقة التطابق الخطى:

هذه الطريقة وتراكيبها تعطي المتتابعة العددية $X_1, X_2, \ldots,$ بين الصفر و $m-1$ حسب الصيغة التكرارية

$$X_{i+1} = \left(aX_i + c\right) \bmod m, \quad i = 0,1,2,\ldots$$

القيمة الأولية X_0 تسمى النواة Seed والثابت a يسمى ثابت التضاعف و الثابت c يسمى الزيادة و m يسمى القياس Modulus . لاحظ أن $a = b \bmod m$ يعني ان $(a-b)$ يقسم m بدون باقي. إذا كانت $c \neq 0$ تسمى المعادلة السابقة طريقة التطابق المختلط وعندما تكون $c = 0$ تسمى طريقة التطابق التضاعفي.

مثال:

سوف نستخدم طريقة التطابق الخطي لتوليد متتابعة من الأرقام العشوائية بأخذ القيم

$$X_0 = 27,\ a = 17,\ c = 43,\ m = 100$$

لاحظ ان الأرقام التي ستولد ستكون موزعة بالتساوي بين 0 و99 لأن ($m-1=99$) لاحظ ايضا ان المتولد أعداد عشوائية وليست أرقام ولكي نولد من هذه المتتابعة أرقام عشوائية لها توزيع متساوي بين 0 و 1 نوجد التالي:

$$R_i = \frac{X_i}{m}, \quad i = 1, 2, \ldots$$

تحسب المتتابعة كالتالي:

$$X_0 = 27$$
$$X_1 = (17 \times 27 + 43) \bmod 100 = 502 \bmod 100 = 2$$
$$R_1 = \frac{2}{100} = 0.02$$
$$X_2 = (17 \times 2 + 43) \bmod 100 = 77 \bmod 100 = 77$$
$$R_2 = \frac{77}{100} = 0.77$$
$$\vdots$$

وهكذا...

على الأرقام شبه العشوائية المولدة بطريقة التطابق الخطى او غيرها ان تحقق بالإضافة إلى خاصيتي التوزيع المتساوي والإستقلال خاصية الكثافة العظمى Maximum Density والتي تعني ان القيم التي تأخذها $R_i, i=1,2,...$ لاتترك ثغرات كبيرة في الفترة $[0,1]$. وخاصية الدورة العظمى Maximum Period or Cycle التي نفسرها بملاحظة ان الأرقام المولدة يمكن ان تأخذ القيم المحدودة $0,\frac{1}{m},\frac{2}{m},...,\frac{m-1}{m}$ فقط فإذا تكرر اي رقم فإن المتتابعة تتكرر من جديد ولهذا فإن أعظم دورة للطريقة السابقة هي m وللحصول على أعظم دورة يجب إختيار قيم X_0 و a و c و m بعناية شديدة ولهذا تؤخذ قيم m على الشكل $m=2^{31}-1$ ومثل هذه القيمة تستخدم في الحاسبات كما يجب ان تكون c و m (عندما $c \neq 0$) اوليين بالنسبة لبعضهما (اي لاتوجد عوامل مشتركة بينهما) وكذلك يجب ان تكون X_0 عدد فردي و a تأخذ قيمة مثل $a=1+4k, a=3+8k, a=5+8k$ للرقم $k=0,1,...$.

اختبارات للأرقام العشوائية:

- ***اختبار التوزيع المتساوي***

1. اختبار كولموجوروف- سمير نوف
2. اختبار مربع كاي

- ***اختبار الاستقلال***

1. اختبار الجري
2. اختبار الترابط الذاتي

1- إختبار كولموجوروف- سميرنوف Kolmogorov-Smirnov Test

سوف نختبر الفرضية

$$H_0 : R_i \quad distributed\ as\ U[0,1]$$
$$H_1 : R_i \quad is\ not\ distributed\ as\ U[0,1]$$

هذا الإختبار يقارن بين دالة التوزيع التراكمي المستمرة $F(x)$ للتوزيع المتساوي ودالة التوزيع التراكمي التجريبي Empirical CDF للعينة $S_N(x)$ لعينة حجمها N.

فمن التعريف نجد

$$F(x) = x, \quad 0 \le x \le 1$$

ولعينة من الأرقام العشوائية المولدة $R_1, R_2, \ldots, R_N$ فإن دالة التوزيع التراكمي التجريبي للعينة $S_N(x)$ تعرف كالتالي

$$S_N(x) = \frac{number\ of\ R_1, R_2, \ldots, R_N\ which\ are \le x}{N}$$

وإذا رتبت الأرقام العشوائية المولدة $R_1, R_2, ..., R_N$ على الشكل $R_{(1)}, R_{(2)}, ..., R_{(N)}$ بحيث $R_{(1)} \le R_{(2)} \le ... \le R_{(N)}$ فإن

$$S_N(x) = \begin{cases} 0, & x < R_{(1)} \\ \frac{k}{N}, & R_{(k)} \le x < R_{(k+1)}, \quad k = 1, 2, ..., N \\ 1, & x \ge R_{(N)} \end{cases}$$

لاحظ انه كلما تكبر قيمة N فإن $S_N(x)$ تصبح اكثر تقريبا للدالة $F(x)$ على أساس صحة الفرضية الصفرية. دالة التوزيع التراكمي التجريبي للعينة هي دالة درجية Step Function ذات قفزات عند كل قيمة مشاهدة (سوف نشرحها بشكل مفصل في فصل تحليل المدخلات) ولهذا فإن الإختبار يفحص القيمة المطلقة لأكبر إنحراف بين $F(x)$ و $S_N(x)$ أي الإحصائة

$$D = \max \left| F(x) - S_N(x) \right|$$

التوزيع العيني للإحصائة D معروف ومجدول ولكننا سوف نستخدم الصيغة التقاربية Asymptotic Formula للإحصائة D عندما تكون $N \geq 35$ لأن توليد عينة من الارقام العشوائية اقل من 35 لأي مولد غير منصفة له.

سوف نستخدم القيم التالية للإحصائة D

N	$D_{0.10}$	$D_{0.05}$	$D_{0.01}$
≥ 35	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$

لإجراء الإختبار تحت فرضية التوزيع المتساوي نستخدم الخوارزم التالي:

خطوة (1) : رتب الأرقام تصاعديا. فإذا كانت $R_{(i)}$ هي المشاهدة التي رتبتها i في الصغر فإن

الناتج يكون $R_{(1)} \le R_{(2)} \le \cdots \le R_{(N)}$

خطوة (2) : أحسب الكميات

$$D^{+} = \max_{1 \le i \le N} \left\{ \frac{i}{N} - R_{(i)} \right\}$$

$$D^{-} = \max_{1 \le i \le N} \left\{ R_{(i)} - \frac{i-1}{N} \right\}$$

خطوة (3) : أحسب الكمية $D = \max\left(D^{+}, D^{-}\right)$

خطوة (4) : حدد القيمة الحرجة D_α من الجدول السابق لقيم معينة لمستوى المعنوية α وللحجم العينة $N \geq 35$

خطوة (5) : إذا كانت القيمة المحسوبة للإحصائة D اكبر من القيمة المجدولة D_α فإننا نرفض الفرضية الصفرية بأن الأرقام العشوائية المولدة تتبع التوزيع المتساوي وإلا فإننا نستنتج اننا لم نكتشف فرق بين التوزيع الحقيقي للأرقام $\{R_1, R_2, ..., R_N\}$ والتوزيع المتساوي.

مثال:-

إستخدم إختبار كولموجوروف – سميرنوف Kolmogorov-Smirnov لإختبار الفرضية فيما إذا كانت الأرقام العشوائية

0.136 0.513 0.844 0.681 0.154 0.239 0.888 0.090 0.631 0.245
0.394 0.531 0.715 0.276 0.880

لها توزيع منتظم على الفترة $[0,1]$ ام لا؟ خذ $\alpha = 0.05$.

R(i)	i	i/N	(i-1)/N	i/N-R(i)	R(i)-(i-1)/N
0.09	1	0.066667	0	-0.023333	0.09
0.136	2	0.133333	0.066667	-0.002667	0.069333
0.154	3	0.2	0.133333	0.046	0.020667
0.239	4	0.266667	0.2	0.027667	0.039
0.245	5	0.333333	0.266667	0.088333	-0.021667
0.276	6	0.4	0.333333	0.124	-0.057333
0.394	7	0.466667	0.4	0.072667	-0.006
0.513	8	0.533333	0.466667	0.020333	0.046333
0.531	9	0.6	0.533333	0.069	-0.002333
0.631	10	0.666667	0.6	0.035667	0.031
0.681	11	0.733333	0.666667	0.052333	0.014333
0.715	12	0.8	0.733333	0.085	-0.018333
0.844	13	0.866667	0.8	0.022667	0.044
0.88	14	0.933333	0.866667	0.053333	0.013333
0.888	15	1	0.933333	0.112	-0.045333

من العمودين قبل الأخير نجد $D^+ = 0.124$ و $D^- = 0.09$ ومن العامود الأخير $D = 0.124$

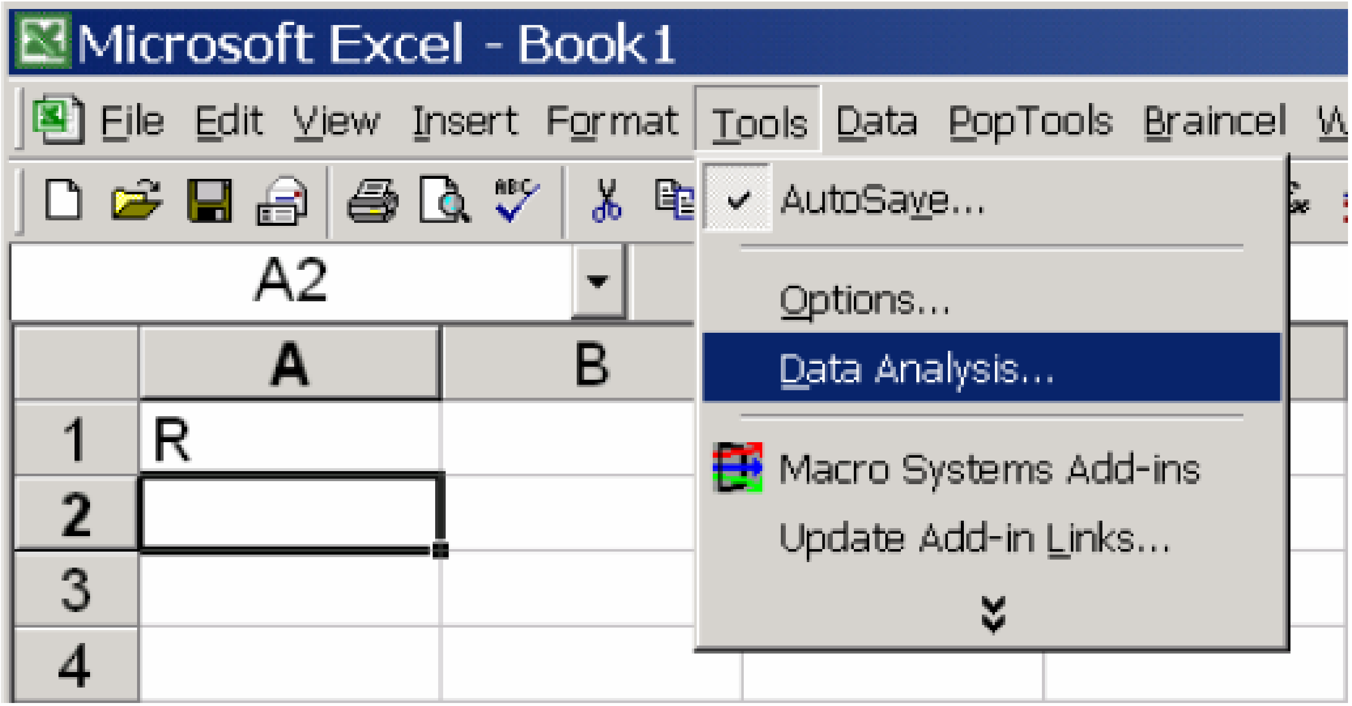
من جدول القيم الحرجة لكولموجوروف- سميرنوف وتحت العمود $D_{0.05}$ والسطر $N = 15$ نجد

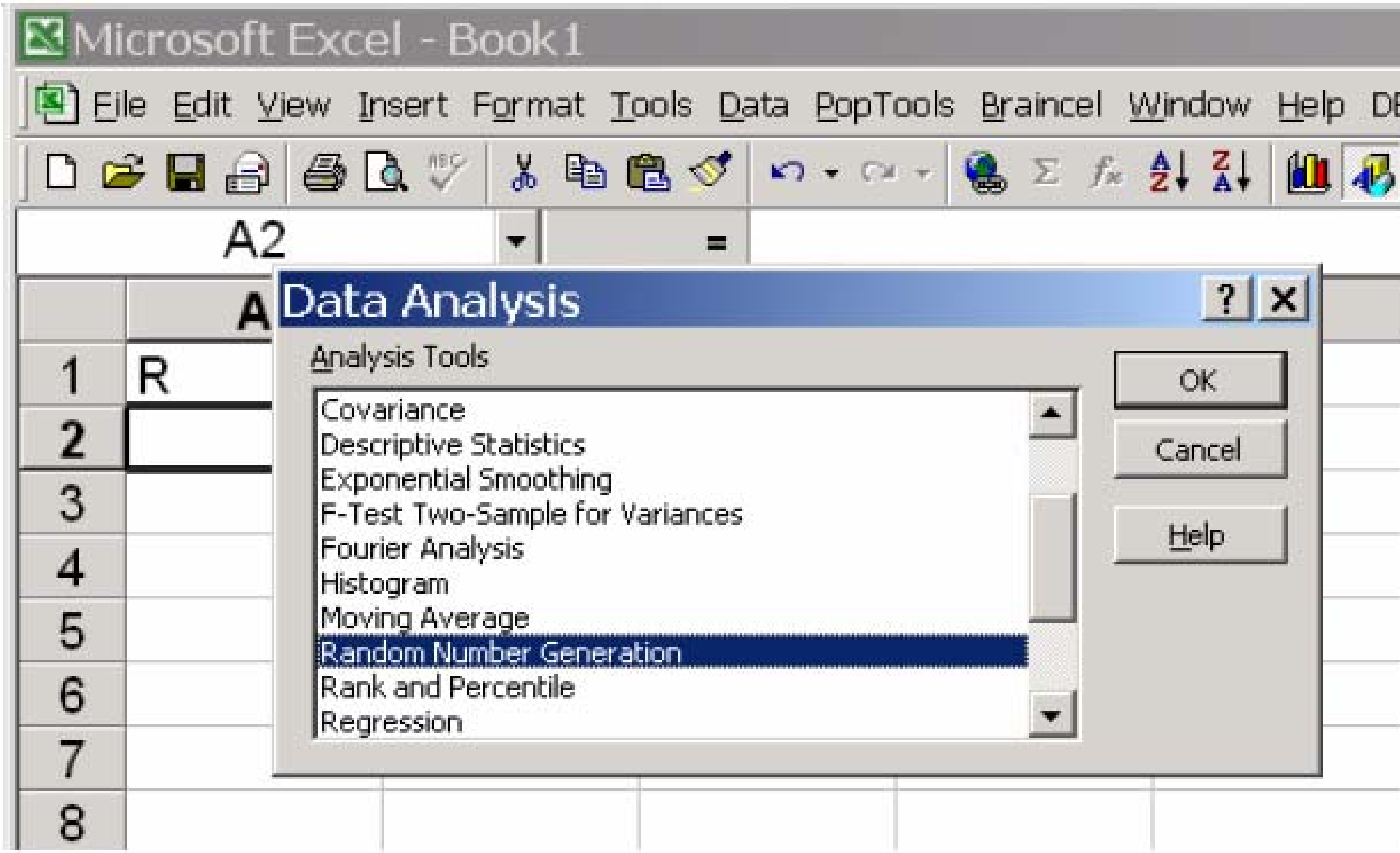
ان القيمة الحرجة هي $D_{0.05,15} = 0.338$

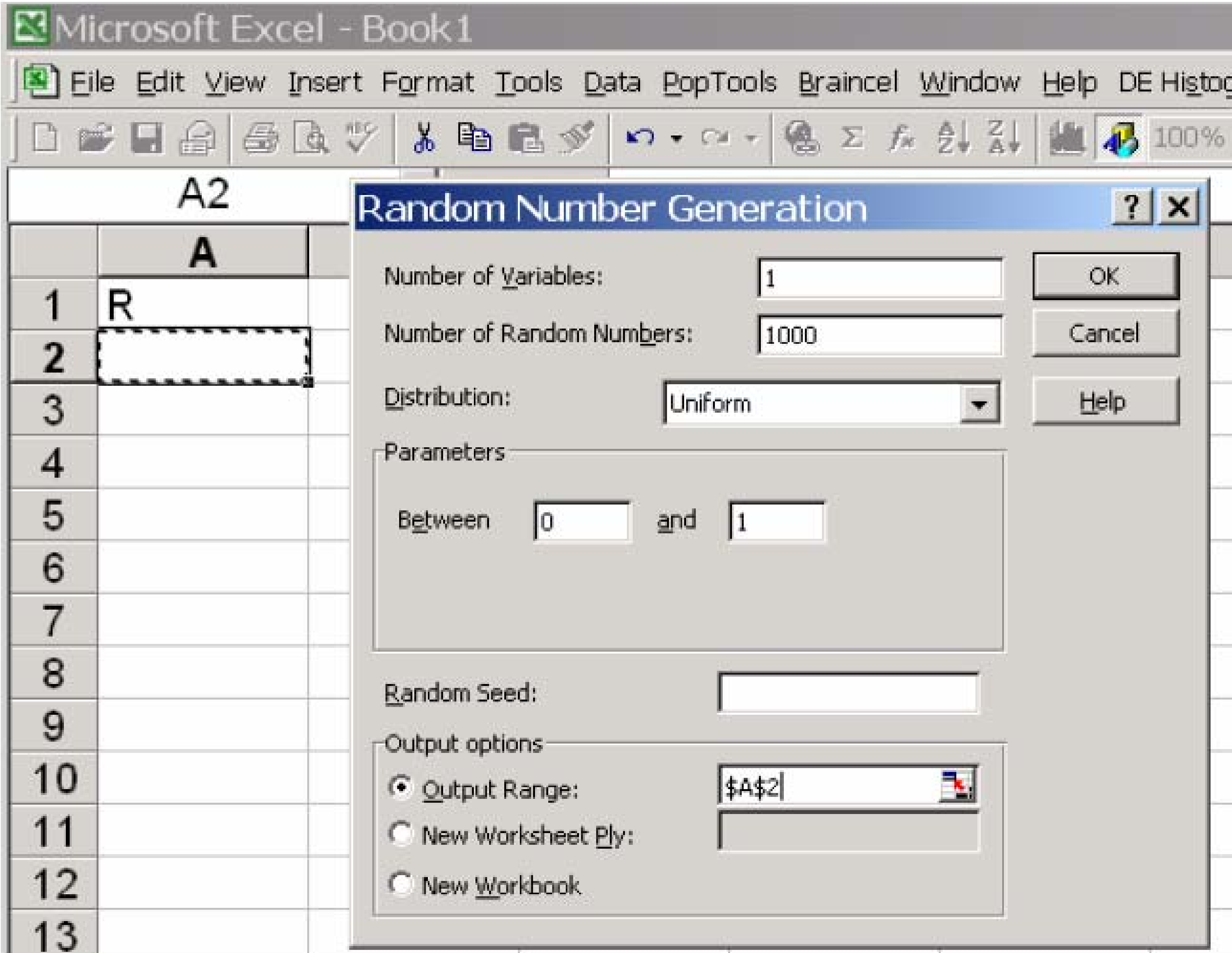
القرار: بما أن $D = 0.124 < D_{0.05,15} = 0.338$ فإننا لا نرفض الفرضية الصفرية أن المشاهدات

تأتي من توزيع منتظم.

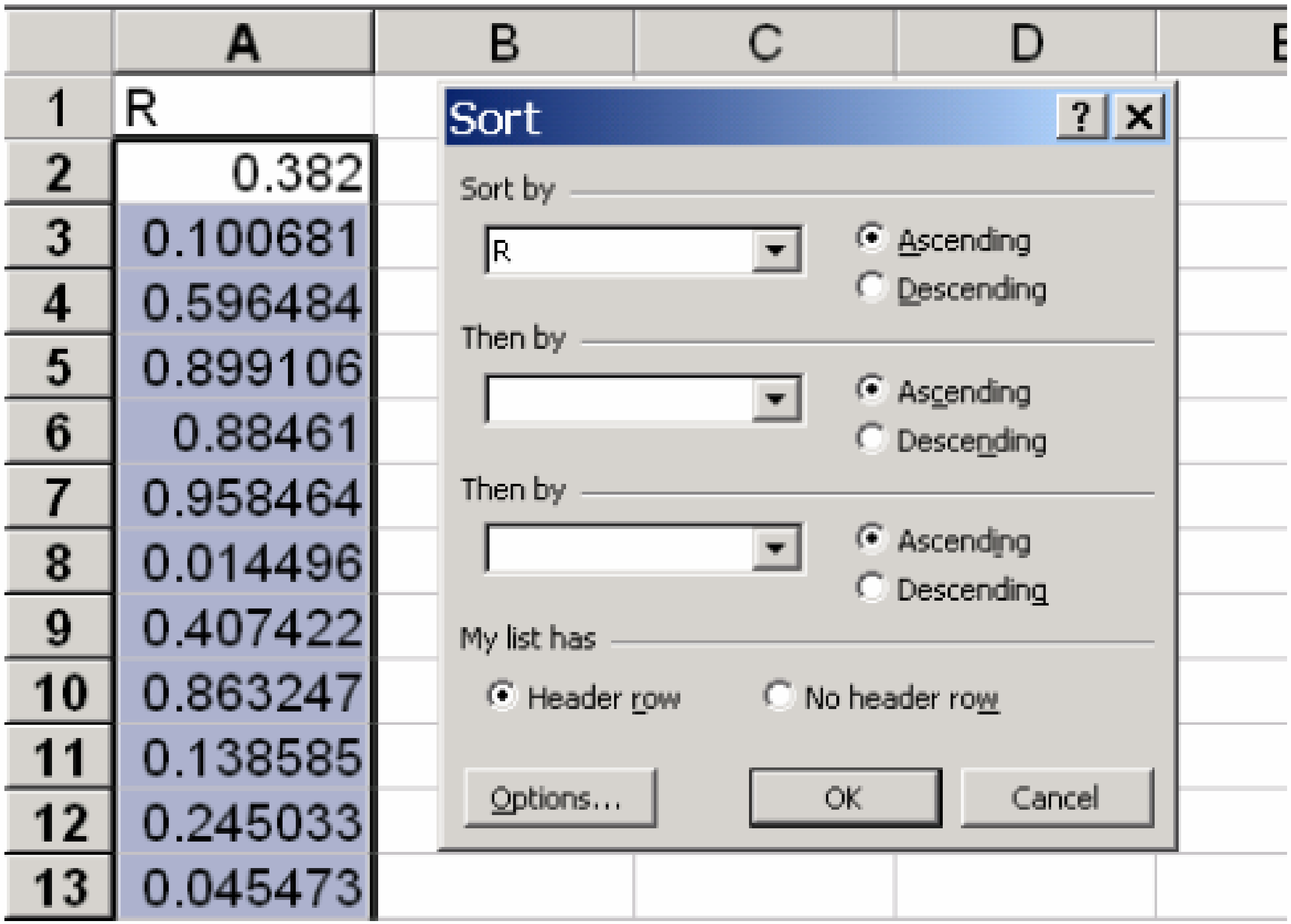
اختبار كولموجوروف- سمير نوف باستخدام Excel




Microsoft Excel - Book1
File Edit View Insert Format Tools Data PopTools Braincel Window Help DE
A2
=
A
1 R
2
3
4
5
6
7
8
Data Analysis
Analysis Tools
Covariance
Descriptive Statistics
Exponential Smoothing
F-Test Two-Sample for Variances
Fourier Analysis
Histogram
Moving Average
Random Number Generation
Rank and Percentile
Regression
OK
Cancel
Help


Microsoft Excel - Book1
File Edit View Insert Format Tools Data PopTools Braincel Window Help DE Histogr
100%
A2
A
1 R
2
3
4
5
6
7
8
9
10
11
12
13
Random Number Generation
Number of Variables:
1
Number of Random Numbers:
1000
Distribution:
Uniform
OK
Cancel
Help
Parameters
Between
0
and
1
Random Seed:
Output options
Output Range:
$A$2
New Worksheet Ply:
New Workbook

	A	B
1	R	
2	0.382	
3	0.100681	
4	0.596484	
5	0.899106	
6	0.88461	
7	0.958464	
8	0.014496	


A
B
C
D
R
0.382
0.100681
0.596484
0.899106
0.88461
0.958464
0.014496
0.407422
0.863247
0.138585
0.245033
0.045473
Sort
Sort by
R
Ascending
Descending
Then by
Ascending
Descending
Then by
Ascending
Descending
My list has
Header row
No header row
Options...
OK
Cancel

	A	B	C	D	E	F	G	H
1	Ri	i	i/N	(i-1)/N	(i/N)-Ri	Ri-(i-1)/N	D+ =	=MAX(E2:E1001)
2	0.00280770287179174	1	=B2/1000	=(B2-1)/1000	=C2-A2	=A2-D2	D- =	=MAX(F2:F1001)
3	0.00497451704458754	2	=B3/1000	=(B3-1)/1000	=C3-A3	=A3-D3	D =	=MAX(H1,H2)
4	0.00515762810144353	3	=B4/1000	=(B4-1)/1000	=C4-A4	=A4-D4	D(alpha)=	=1.36/SQRT(1000)
5	0.00592059083834346	4	=B5/1000	=(B5-1)/1000	=C5-A5	=A5-D5	Do not reject H0 at alpha=0.05	
6	0.00817896053956725	5	=B6/1000	=(B6-1)/1000	=C6-A6	=A6-D6		

	A	B	C	D	E	F	G	H
1	Ri	i	i/N	(i-1)/N	(i/N)-Ri	Ri-(i-1)/N	D+ =	0.01985519
2	0.002807703	1	0.001	0	-0.001807703	0.002807703	D- =	0.029671468
3	0.004974517	2	0.002	0.001	-0.002974517	0.003974517	D =	0.029671468
4	0.005157628	3	0.003	0.002	-0.002157628	0.003157628	D(alpha)=	0.043006976
5	0.005920591	4	0.004	0.003	-0.001920591	0.002920591	Do not reject H0 at alpha=0.05	
6	0.008178961	5	0.005	0.004	-0.003178961	0.004178961		

لاحظ ان نتيجة الإختبار موجبة أي اننا لم نكتشف فرق بين التوزيع الحقيقي للأرقام $\{R_1, R_2, ..., R_{1000}\}$ والتوزيع المتساوي.

2- اختبار مربع كاي Chi-Square Test

إختبار مربع كاي يستخدم الإحصائة

$$\chi_0^2 = \sum_{i=1}^{n} \frac{\left(O_i - E_i\right)^2}{E_i}$$

O_i هو عدد البيانات المشاهدة في الفئة i و E_i هو عدد البيانات المتوقعة في الفئة i و

n عدد الفئآت. للتوزيع المتساوي العدد المتوقع من البيانات التي تقع في الفئة i هو

$E_i = N/n$ حيث N حجم العينة. الإحصائة χ_0^2 توزع تقريبا كتوزيع مربع كاي

بدرجات حرية $n-1$.

مثال:-
للأعداد العشوائية التالية:

0.488	0.226	0.221	0.043	0.055	0.743	0.081	0.685	0.364	0.012
0.372	0.543	0.483	0.050	0.628	0.966	0.750	0.697	0.764	0.040
0.404	0.549	0.203	0.990	0.155	0.079	0.789	0.462	0.795	0.190

مستخدما إختبار مربع كاي أختبر الفرضية التالية:

$$H_0: \quad R_i \sim U[0,1]$$

$$H_1: \quad R_i \nsim U[0,1]$$

مستخدمين إختبار مربع كاي
نقسم الفترة [0,1] إلى 10 فترات متساوية ونسجل عدد الأعداد العشوائية التي تقع في كل فترة

Interval	Observed	Expected	Chi Sq.
0.00 - 0.10	7	3	5.33
0.11 - 0.20	2	3	0.33
0.21 - 0.30	3	3	0
0.31 - 0.40	2	3	0.33
0.41 - 0.50	4	3	0.33
0.51 - 0.60	2	3	0.33
0.61 - 0.70	3	3	0
0.71 - 0.80	5	3	1.33
0.81 - 0.90	0	3	3
0.91 - 1.00	2	3	0.33
			11.33

قيمة كاي تربيع المحسوبة هي 11.33 وبدرجات حرية 9
قيمة كاي تربيع الجدولية عند $\alpha = 0.05$ ودرجات حرية 9 هي 16.9 وهي أكبر من القيمة المحسوبة.
القرار: لانرفض الفرضية الصفرية.

مثال : اختبار مربع كاي باستخدام Excel

R

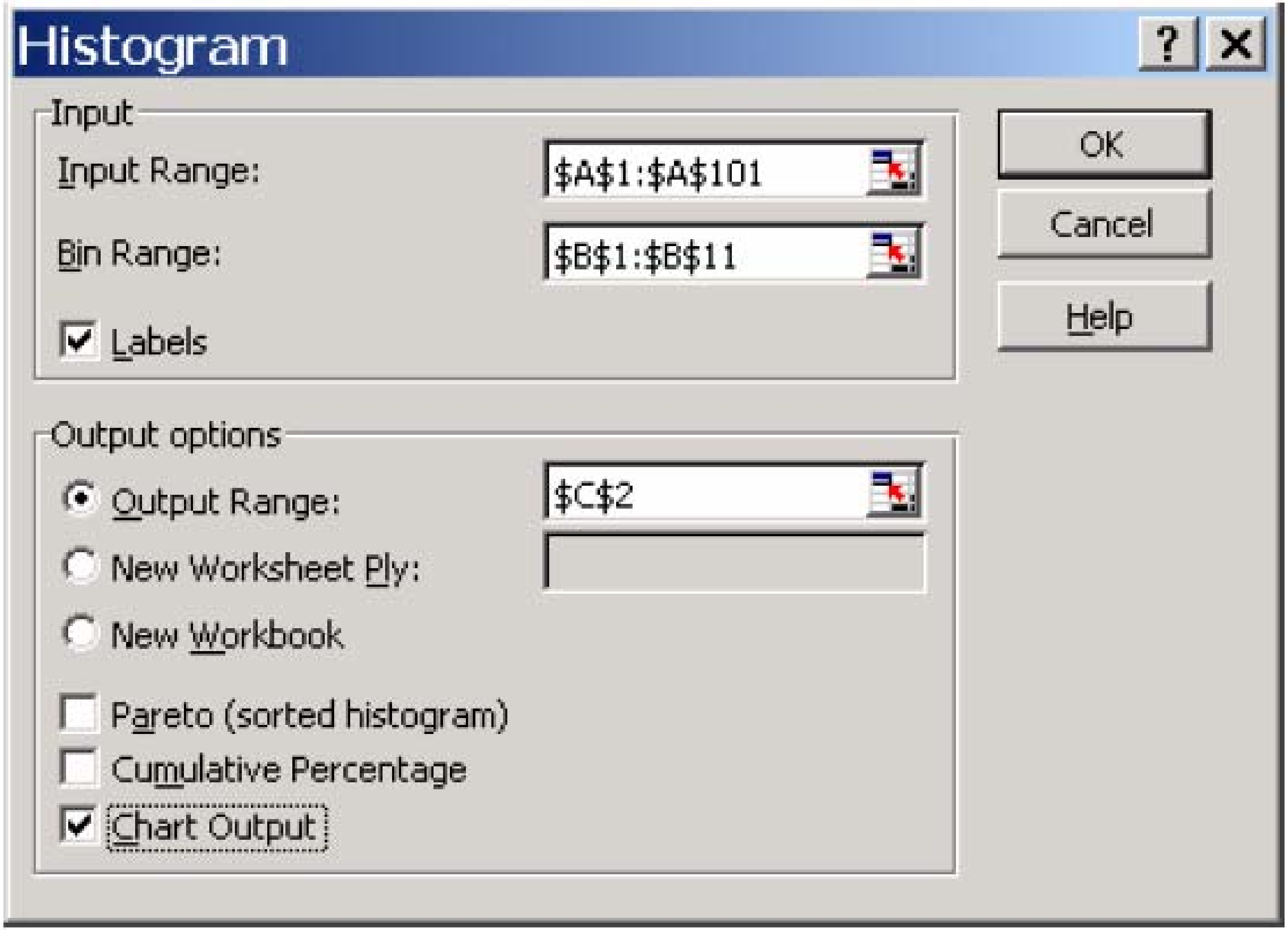
0.34	0.83	0.96	0.47	0.79	0.99	0.37	0.72	0.06	0.18
0.9	0.76	0.99	0.3	0.71	0.17	0.51	0.43	0.39	0.26
0.25	0.79	0.77	0.17	0.23	0.99	0.54	0.56	0.84	0.97
0.89	0.64	0.67	0.82	0.19	0.46	0.01	0.97	0.24	0.88
0.87	0.7	0.56	0.56	0.82	0.05	0.81	0.3	0.4	0.64
0.44	0.81	0.41	0.05	0.93	0.66	0.28	0.94	0.64	0.47
0.12	0.94	0.52	0.45	0.65	0.1	0.69	0.96	0.4	0.6
0.21	0.74	0.73	0.31	0.37	0.42	0.34	0.58	0.19	0.11
0.46	0.22	0.99	0.78	0.39	0.18	0.75	0.73	0.79	0.29
0.67	0.74	0.02	0.05	0.42	0.49	0.49	0.05	0.62	0.78

	A	B	C	D	E	F	G	H	I
1	R	Interval	Upper Interval	*Upper Interval*	*Observed*	Expected	Chi (9)		(O-E)^2/E
2	0.34	1	0.1	0.1	8	=100/10	=1-CHITEST(E2:E11,F2:F11)	>0.05	=(E2-F2)*(E2-F2)/F2
3	0.83	2	0.2	0.2	8	=100/10	Chi (9,0.05)		=(E3-F3)*(E3-F3)/F3
4	0.96	3	0.3	0.3	10	=100/10	=CHIINV(0.05,9)	>3.4	=(E4-F4)*(E4-F4)/F4
5	0.47	4	0.4	0.4	9	=100/10			=(E5-F5)*(E5-F5)/F5
6	0.79	5	0.5	0.5	12	=100/10			=(E6-F6)*(E6-F6)/F6
7	0.99	6	0.6	0.6	8	=100/10			=(E7-F7)*(E7-F7)/F7
8	0.37	7	0.7	0.7	10	=100/10			=(E8-F8)*(E8-F8)/F8
9	0.72	8	0.8	0.8	14	=100/10			=(E9-F9)*(E9-F9)/F9
10	0.06	9	0.9	0.9	10	=100/10			=(E10-F10)*(E10-F10)/F10
11	0.18	10	1	1	11	=100/10			=(E11-F11)*(E11-F11)/F11
12	0.9			More	0				
13	0.76							SUM=	=SUM(I2:I11)

تعبأ C من تقسيم الفترة [0,1) إلى 10 فترات كالتالي [0,0.1), [0.1,0.2), ..., [0.9,1.0)

وتدخل القيم العليا للفئات 0.1, 0.2, ..., 1.0 في المجال C2:C11

أختار Tools => Data Analysis => Histogram وعبئ الفراغات كالتالي:


Histogram
Input
Input Range:
$A$1:$A$101
Bin Range:
$B$1:$B$11
Labels
OK
Cancel
Help
Output options
Output Range:
$C$2
New Worksheet Ply:
New Workbook
Pareto (sorted histogram)
Cumulative Percentage
Chart Output

	A	B	C	D	E	F	G	H	I
1	R	Interval	Upper Interval	*Upper Interval*	*Observed*	Expected	Chi (9)		(O-E)^2/E
2	0.34	1	0.1	0.1	8	10	0.053692325	>0.05	0.4
3	0.83	2	0.2	0.2	8	10	Chi (9,0.05)		0.4
4	0.96	3	0.3	0.3	10	10	16.91896016	>3.4	0
5	0.47	4	0.4	0.4	9	10			0.1
6	0.79	5	0.5	0.5	12	10			0.4
7	0.99	6	0.6	0.6	8	10			0.4
8	0.37	7	0.7	0.7	10	10			0
9	0.72	8	0.8	0.8	14	10			1.6
10	0.06	9	0.9	0.9	10	10			0
11	0.18	10	1	1	11	10			0.1
12	0.9			More	0				
13	0.76							SUM=	3.4
14	0.99								

نلاحظ ان القيمة المحسوبة لمربع كاي 3.4 بدرجات حرية 9 اقل من القيمة الجدولية 16.92 عند مستوى معنوية $\alpha = 0.05$ وبهذا فإن نتيجة الإختبار موجبة أي اننا لم نكتشف فرق بين التوزيع الحقيقي للأرقام $\{R_1, R_2, ..., R_{100}\}$ والتوزيع المتساوي.