今日やること
1. 度数分布・累積度数
2. 回帰

学生番号　1234567　…ここに学生番号を入力する。
氏名　　　　　　　…ここに氏名を入力する。

度数分布表
度数分布表は、データをいくつかの階級ごとに集計したものです。
例として、次のようなデータ（試験の結果、10人分）の度数分布を作ります。
0　　5　　15　　20　　31　　51　　65　　77　　80　　90 ← 試験の結果（10人分）

人数を数えるのはExcelがやってくれるのですが、まず自分で数えてみましょう。

下限値	階級値	上限値	度数（頻度）、この場合は人数
0		10	
11		20	
21		30	
31		40	
41		50	
51		60	
61		70	
71		80	
81		90	
91		100	

「の値を「度数分布表」といいます。度数分布表の3行の「階級」といい、階級を代表する値を「階級値」といいます。

度数分布表のヒストグラムにしたものを「ヒストグラム」といいます。
ヒストグラムでは、棒を並べて見やすくつけて棒くのが一般です。
それのためには、棒をクリック一右のオプション・要素の間隔を「0%」にします。（操作...）

同じデータから、少し幅う度数分布表を作ってみましょう。

下限値	階級値	上限値	度数（頻度）、この場合は人数
		10	
		20	
		30	
		40	
		50	
		60	
		70	
		80	
		90	
		100	

できあがりのR図を参考に、ヒストグラムを書いてください。

そろそろ人数を数えるのは行にちゃってでしょうから、Excelに数えてもらいましょう。
1 データ
0　　5　　15　　20　　31　　51　　65　　77　　80　　90 ← 試験の結果（10人分）

上限値	度数	
10		度数のついた部分を選択する。
20		「=」に入力する。
30		「FREQUENCY」が候補に出てきたら、TABキーを押す。
40		データの配列とは（データを、キーボードで入力してもいいし、マウスで選択してもいい。
50		区間配列はA1:A10である。
60		「=FREQUENCY(A1:A10,A1:A10)」まで入力できたら、Ctrl+Shift+Enter
70		
80		
90		
100		

次の度数分布表を完成させ、ヒストグラムを書く。

下限値	階級値	上限値	度数（頻度）、この場合は人数
0	10.0		20
21	30.0		40
41	50.0		60
61	70.0		80
81	90.0		100

FREQUENCYを使う方法の利点は、データが変わると、度数も自動的に変わることです。
データが変わるとヒストグラムも自動的に書き換えられます。
最初にやったように、手作業で人数を数えて入力していると、データが変わったときには自分でやり直しです。
そういうことにならないように、度数はFREQUENCYで求めることを習慣にしましょう。

ヒストグラムはデータの分布の様子を見るためのものです。
便利な方法ですが、これまでの扱うので、次のようなことがわかるでしょう。
・階級幅をちょっと変えるだけでも結果がガラッと変わることがある。（上の例では、上限を10〜9のように変えたら、ヒストグラムはガラッと変わった。）
・階級数が変わる結果がガラッと変わることがある。

ですから、階級を適切に決めるかが大事ということのですが、そういうのはありません。
階級数の一応、スタージェスの公式という方があります。
データの数はN、階級数を1+log₂Nとするのです（小数の四捨五入）。
たとえば、データが1000件の場合の階級数は → 10.9673

ヒストグラムを作り慣れるために、試験結果1000人分も作ります。

乱数を使う以上は値が変わりますが、
乱数配列を生成し、1000行を表にしたので、
上に並んだ数字だけがずべてです。（最後にある階級値のサイズが操作をしています。）

度数分布表を完成させ、ヒストグラムを書いてください（データはA2:A1001）。（度数を使っているので、できあがりと同じになりません。）

試験の点数のヒストグラムは、階級を10のきざみにするとわかりやすいのですが、それがベストというわけではありません。
たとえば、仮に切りのよいステージェスの公式を使うと、1000人分の階級数は約11です。
10きざみなら階級数は10になり、あまり変わらないかもしれませんが、階級数が11の場合の度数分布表とヒストグラムを作ってみましょう。
データの最小値と最大値を求め、その差を11で割って、階級の幅を求めます。
　最小値＝最小値 T=min(a1:a10,B10) で求めます。
　最大値＝最大値 T=max(a1:a10,B10) で求めます。
8.272727 ＝階級の幅は T=（最大値−最小値）/11 です。

度数分布表とヒストグラムを作ります。（乱数を使っているので、できあがりと同じにはなりません。）

下限値	上限値	度数
0	8.272727	
8.272727	16.54546	
16.54545	24.81818	
24.81818	33.09091	
33.09091	41.36364	
41.36364	49.63636	
49.63636	57.90909	
57.90909	66.18182	
66.18182	74.45455	
74.45455	82.72727	
82.72727	91	

「この式は」1000になることを確認する。

度数分布表を作るときに、累積度数を計算することがあります。
累積度数というのは、全部の値の、度数が度数なのですが、例を見た方が理解は早いでしょう。
最初に使ったデータを再利用します。

上限値	度数	累積度数	
10	2	2	←一番目は同じ
20	2	4	←ここは2+2=4が2+2？
30	0	4	←あとはコピペ
40	1	5	
50	0	5	
60	1	6	
70	1	7	
80	2	9	
90	1	10	
100	0	10	←最後に元の人数になることを確認する。
	10		

さらに、累積度数をデータ数で割ると、累積相対度数と計算します。

上限値	度数	累積度数	累積相対度数	
10	2	2	0.2	←累積度数をデータ数で割ったもの
20	2	4	0.4	←あとはコピペ
30	0	4	0.4	
40	1	5	0.5	
50	0	5	0.5	
60	1	6	0.6	
70	1	7	0.7	
80	2	9	0.9	
90	1	10	0.9	
100	0	10	1	←最後に1になることを確認する。
	10			

ヒストグラム（度数分布グラフ）上、累積相対度数をまとめて可視化します。

A1:A1130のデータのヒストグラムと累積相対度数をまとめて可視化します。（乱数を使っているので、できあがりと同じにはなりません。）

おまけ：分析ツールの「ヒストグラム」を使うとは以上の作業が簡単にできますが、動画解説は → https://xxx.xx.xxxxxxxx
でも、今回は勉強なので、これで終わらないようにしてください。

データ区間	頻度	累積 %
	1	0.20%
6.677419	1	0.02%
12.35484	2	0.02%
15.03226	6	1.02%
17.70968		1.00%
20.3871		
23.06452		
25.74194		
28.41935		
31.09677		
33.77419		
36.45161		
39.12903		
41.80645		
44.48387		
47.16129		
49.83871		
52.51613		
55.19355		
57.87097		
60.54839		
63.22581		
65.90323	33	86.80%
68.58065	20	90.40%
71.25806	15	93.60%
73.93548	17	95.30%
76.6129	18	97.10%
79.29032	12	98.10%
81.96774	8	98.40%
84.64516	3	96.40%
87.32258	4	99.80%
次の値	2	100.00%

ヒストグラム