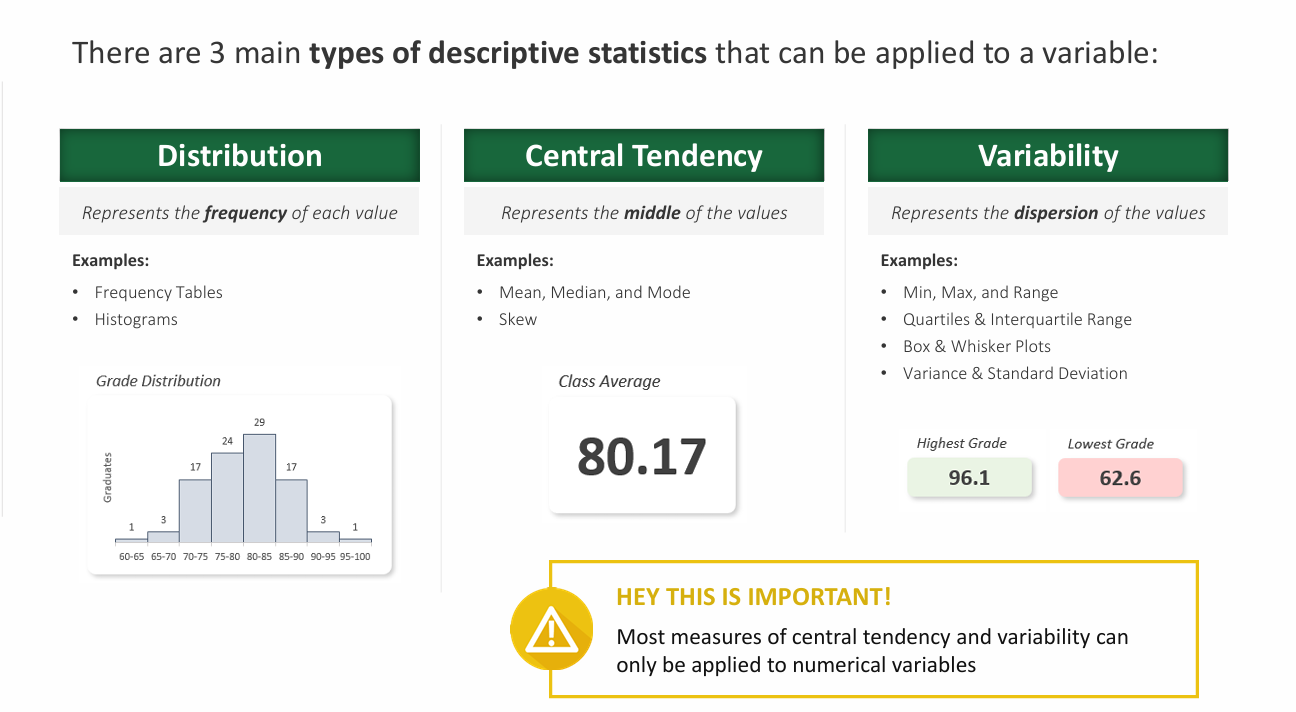
STATISTICA UNIVARIATA

La branca della statistica che studia una sola variabile è la **statistica descrittiva**.

La statistica descrittiva, infatti, riassume le caratteristiche di una variabile.

Essa prende in esame un gran numero di osservazioni e le riassume in un grafico, leggibile anche agli occhi dei meno esperti.

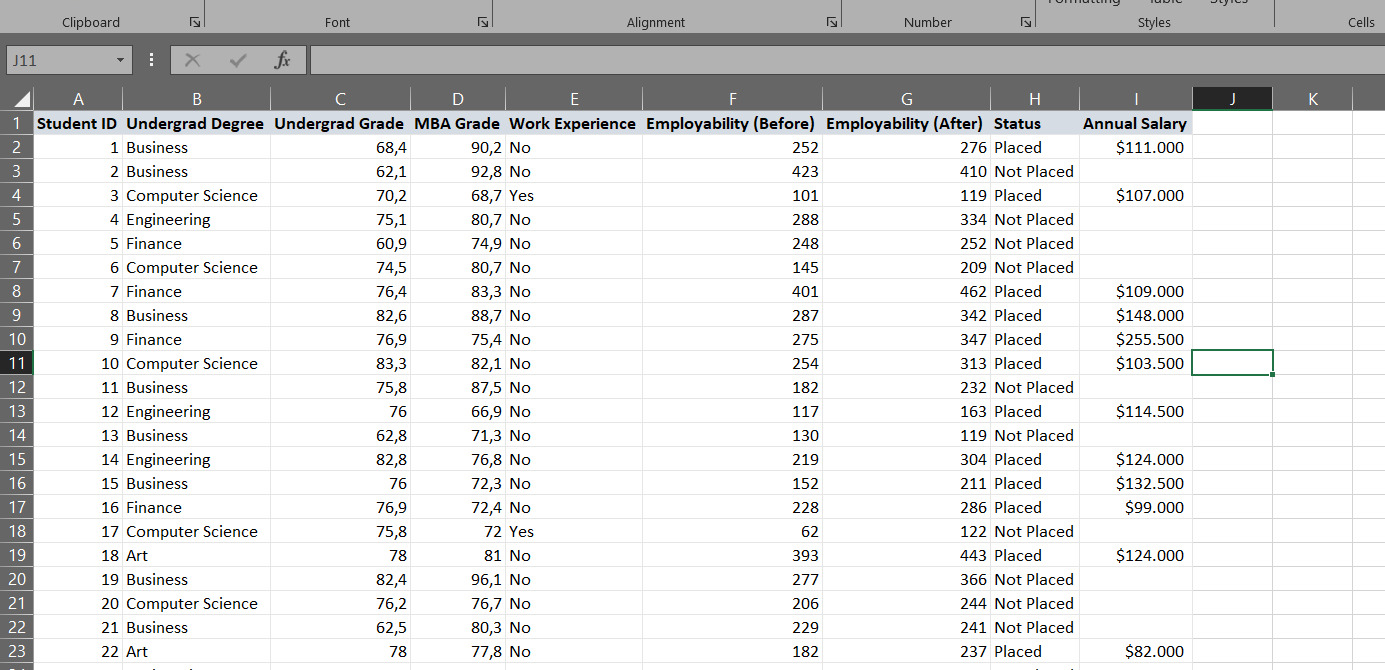
In generale la statistica descrittiva può essere riassunta nel seguente modo:



IMPORTANTE.

Il foglio excel nel quale possiamo trovare i dati presenti nelle lezioni del corso si chiama File\_Del\_Corso, presente in questa cartella.

La tabella di base è la seguente:



Questa tabella si trova nel foglio Tabella presente nel foglio excel presente nel file File\_Del\_Corso.

**DISTRIBUZIONE DI FREQUENZA**

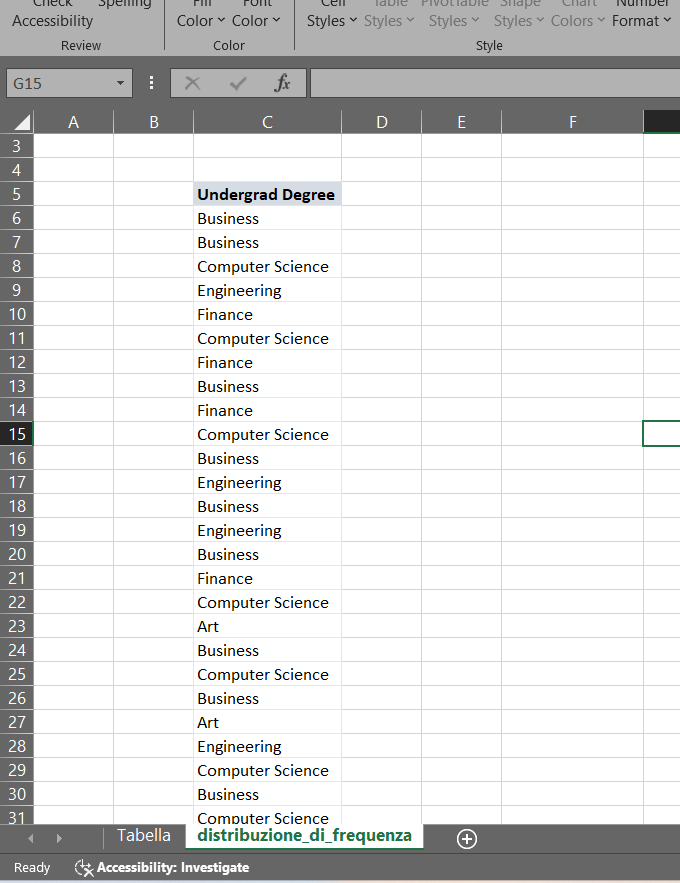
La distribuzione di frequenza è un conteggio. Può essere creata sia per variabili numeriche sia per variabili categoriche.

Il foglio excel nel quale lavoreremo si chiama distribuzione\_di\_frequenza.

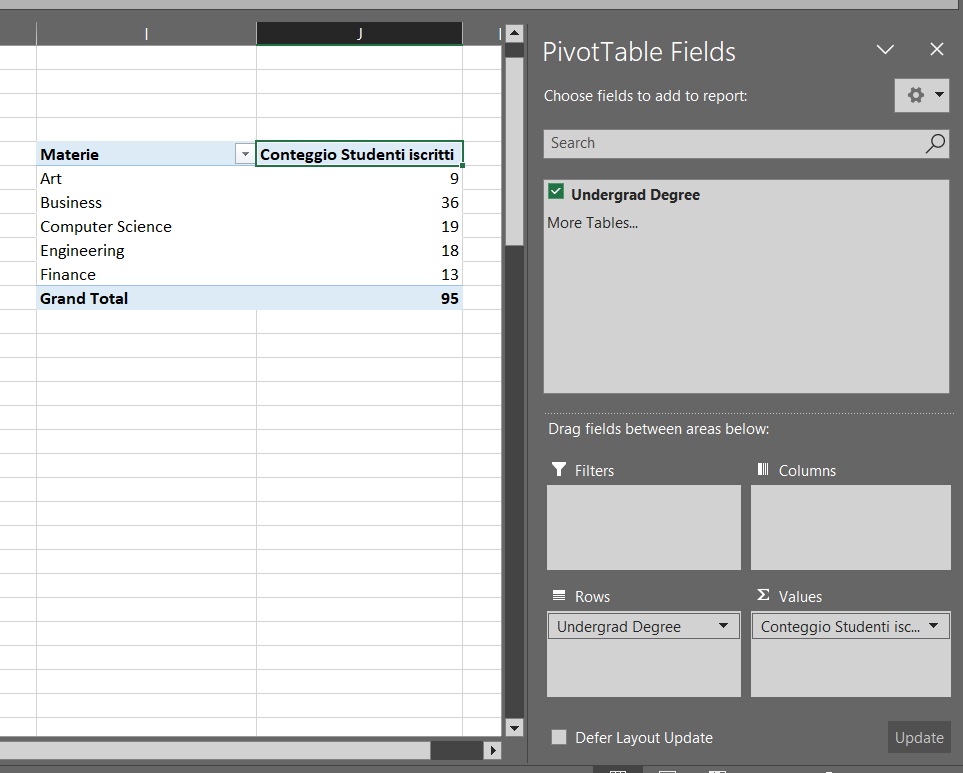
**DISTRIBUZIONE DI FREQUENZA PER VARIABILI CATEGORICHE**

Dal foglio Tabella prendiamo in considerazione la variabile Undergrad Degree.

Consideriamo la variabile Undergrad Degree, in quanto su di essa è molto semplice costruire la tabella di frequenza.



Creiamo la tabella di frequenza utilizzando una tabella pivot.

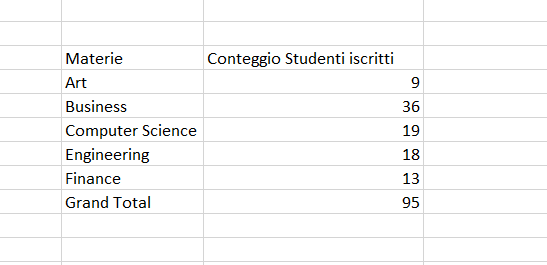


La tabella pivot mostra le frequenze assolute: 9 studenti si sono iscritti al corso Art, 36 studenti si sono iscritti al corso business etc.

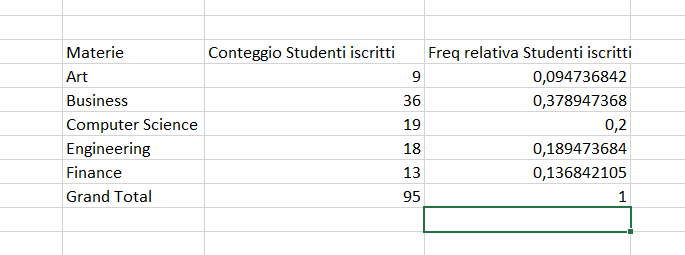
Gli studenti in totale sono 95.

A livello statistico, le frequenze assolute sono utili, ma conviene anche calcolare le frequenze relative.

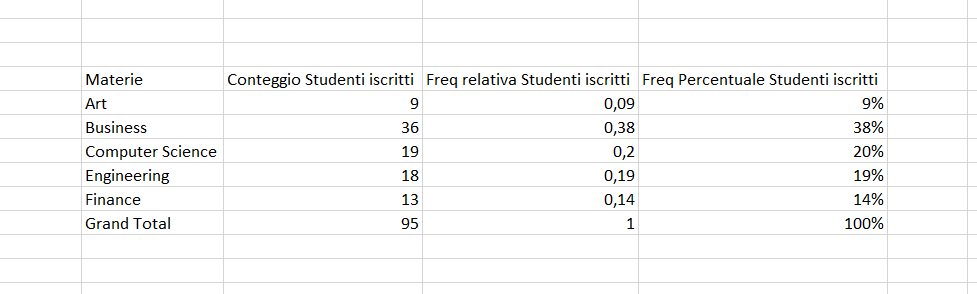
Copiamo quindi solo i valori della nostra tabella pivot.



Dividiamo il conteggio degli studenti iscritti ai vari corsi per la loro somma.

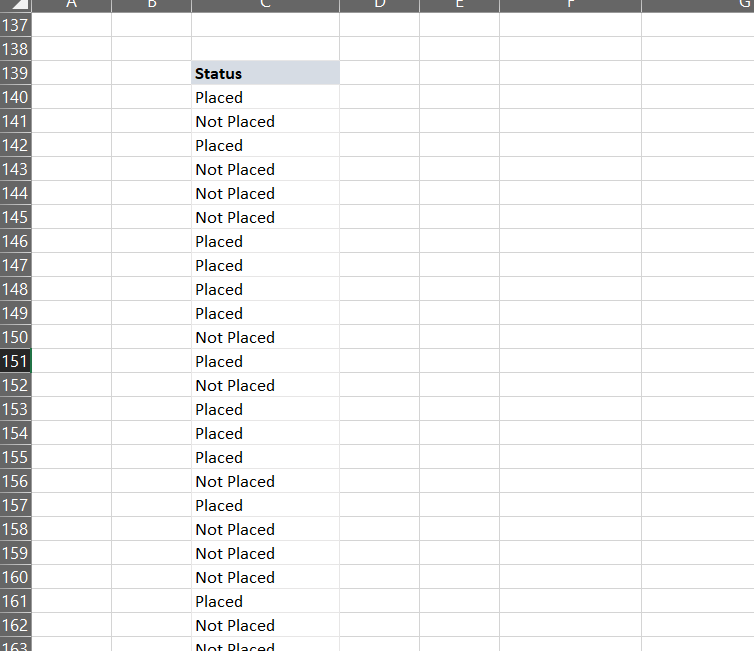


La tabella finale diventerà:

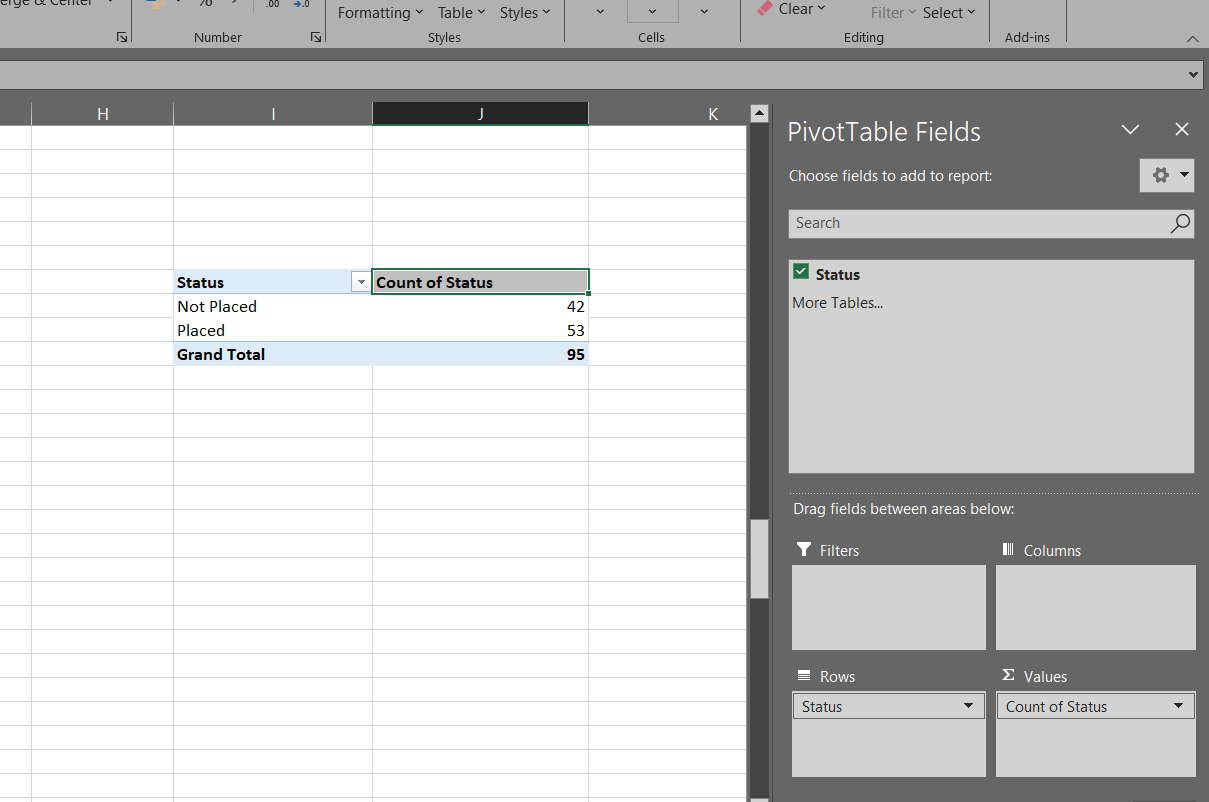


Vediamo un altro esempio con la variabile status. Questa è una variabile dicotomica: assume solo due valori. Nel nostro caso, i valori saranno placed e not placed.

Come al solito ci riferiamo al foglio distribuzione\_di\_frequenza del file excel\_per\_studio

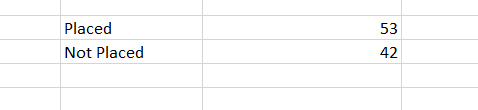


Vediamo quindi quanti sono i placed ed i not placed con la nostra tabella pivot.



Abbiamo lavorato con le tabelle pivot.

Possiamo svolgere lo stesso compito con la funzione countifs. Vediamo quello che esce fuori

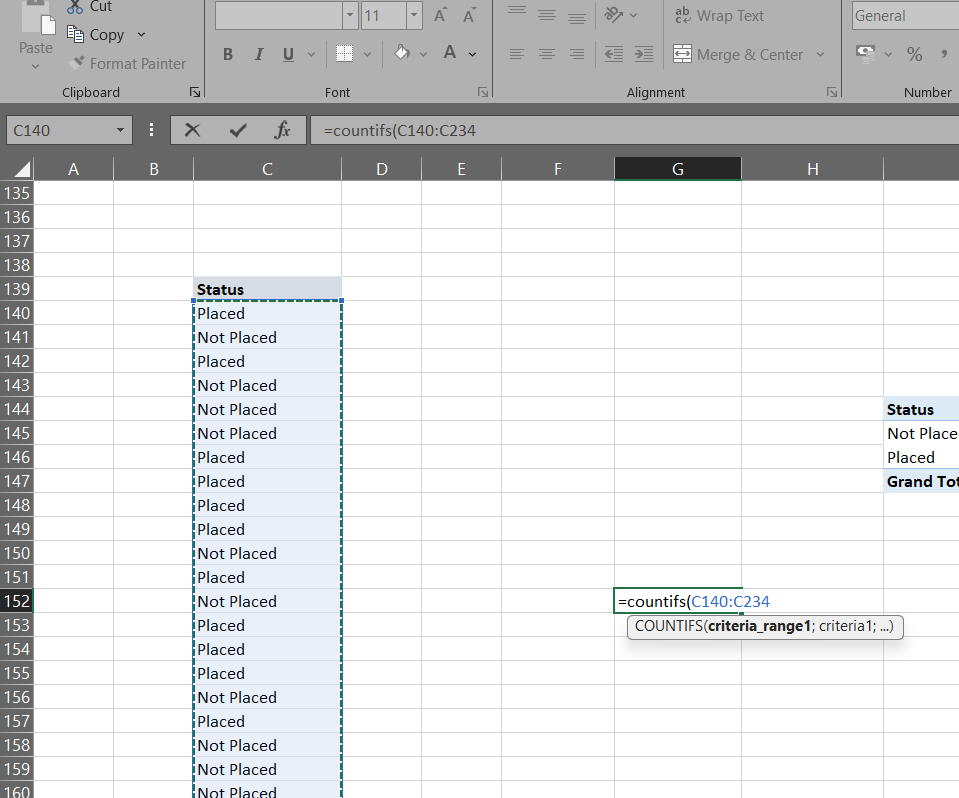


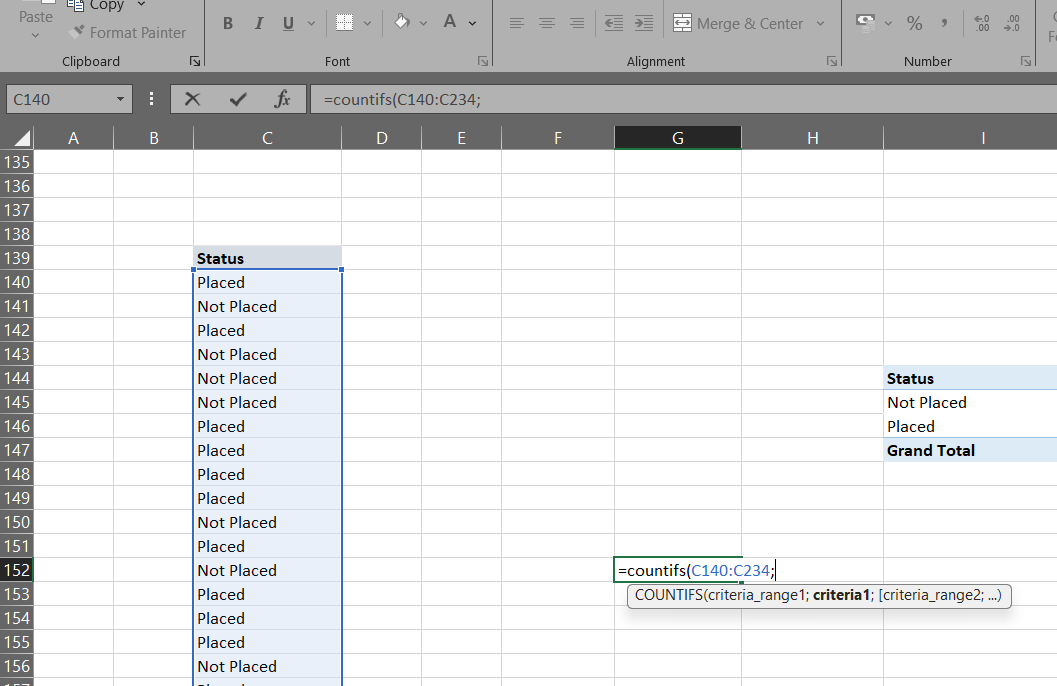
Come siamo arrivati a questo risultato?

La fuzione countifs funziona così.

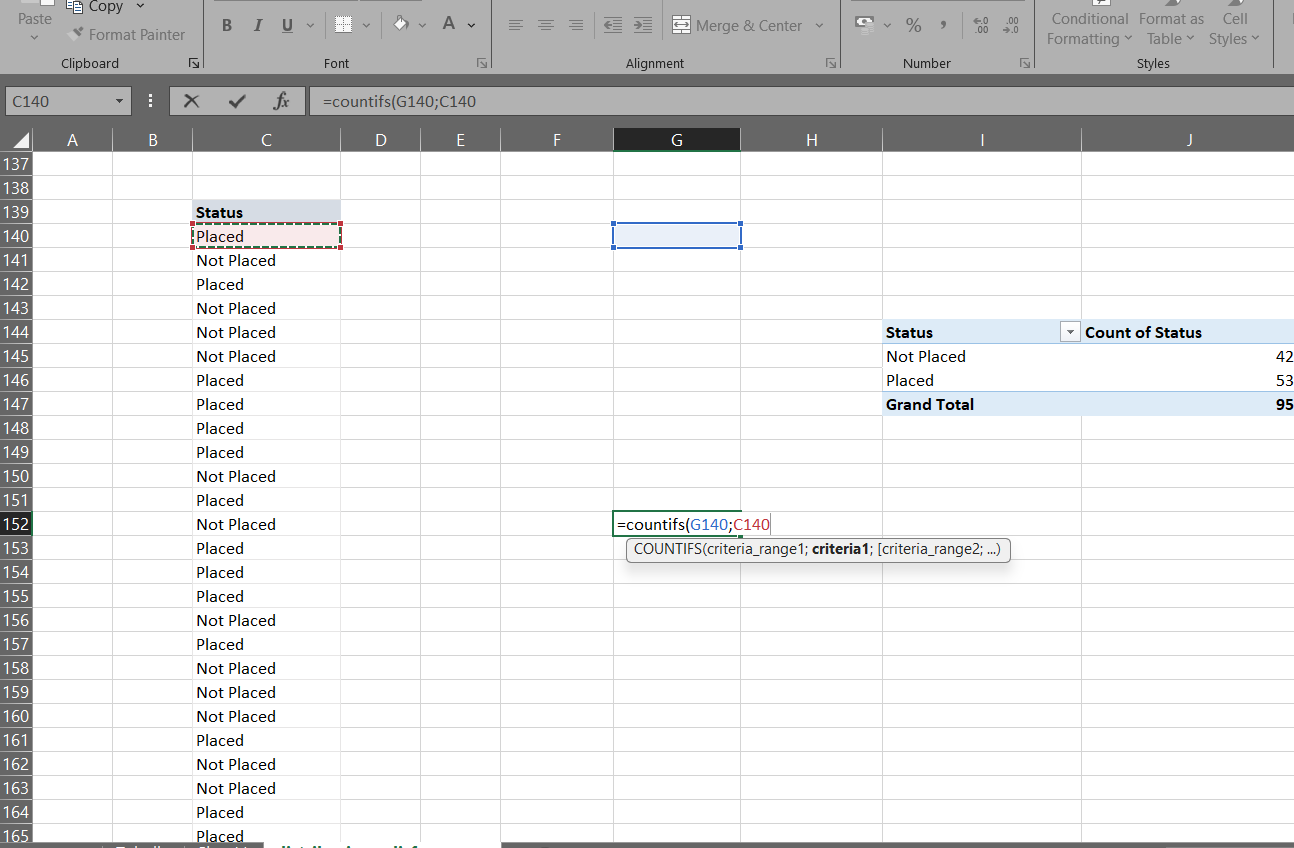
countifs(range;valore\_cella).

Nel nostro caso, quindi:

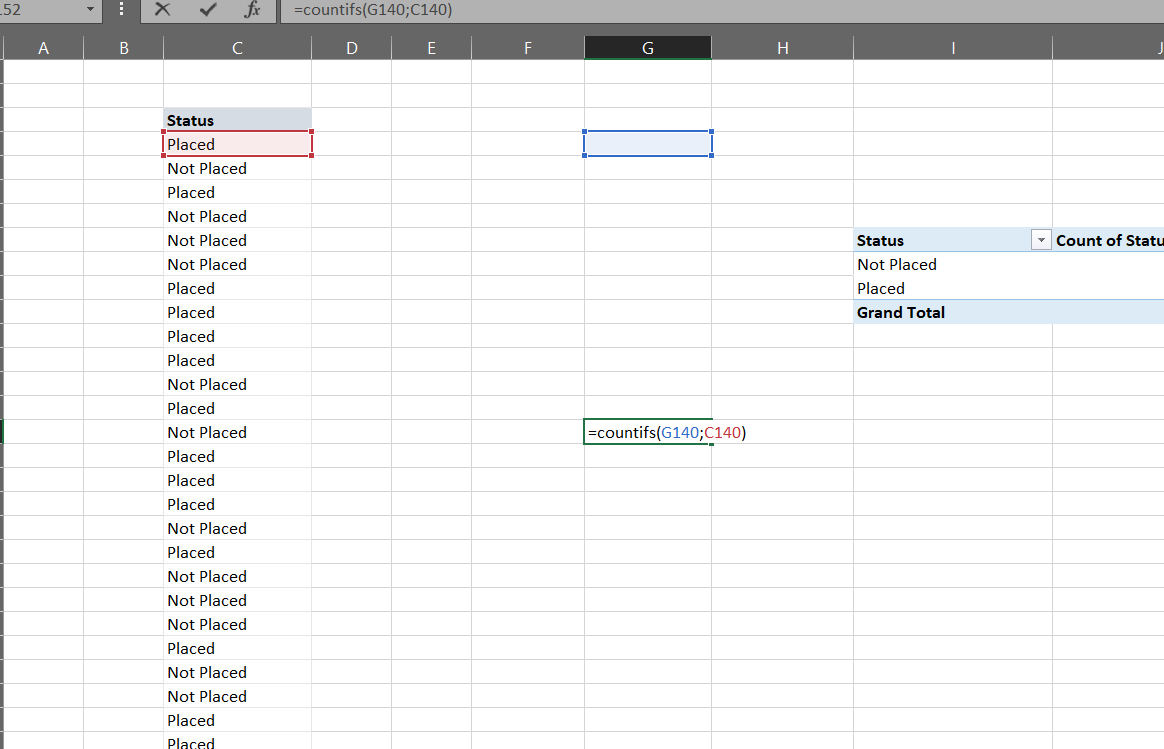
1. Scriviamo countif.
2. Il primo valore è il range che ci serve. Selezioniamo quindi tutta la colonna status. 
3. Dopodiché si scrive il punto e virgola:



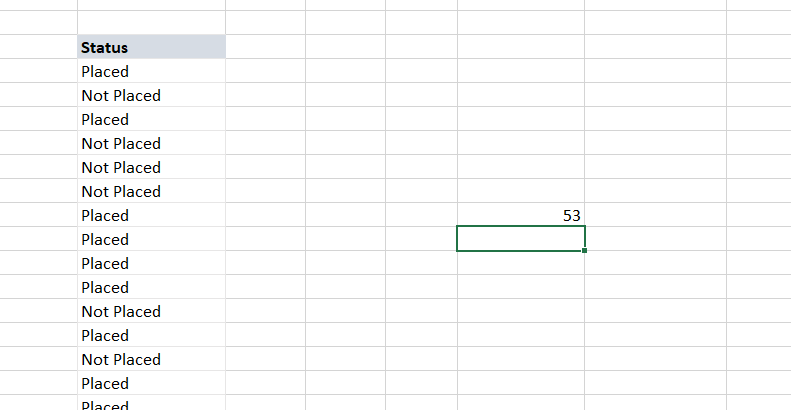
1. Dopodichè si clicca sul valore della colonna del quale si vogliono contare le occorrenze.



1. Infine si chiude la parentesi tonda:

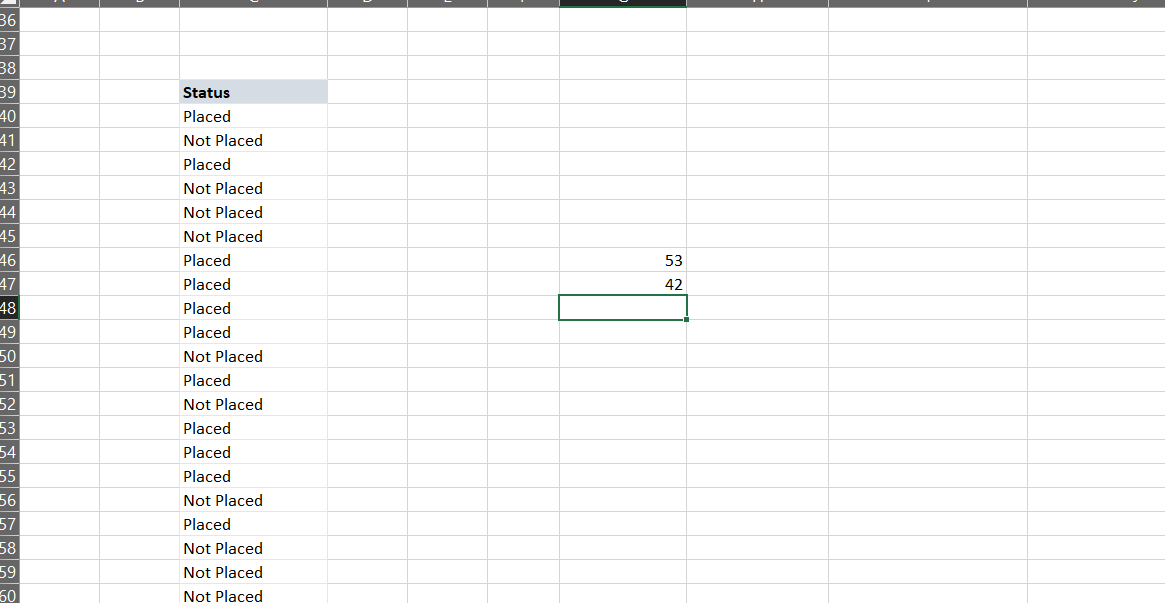


1. E si preme invio

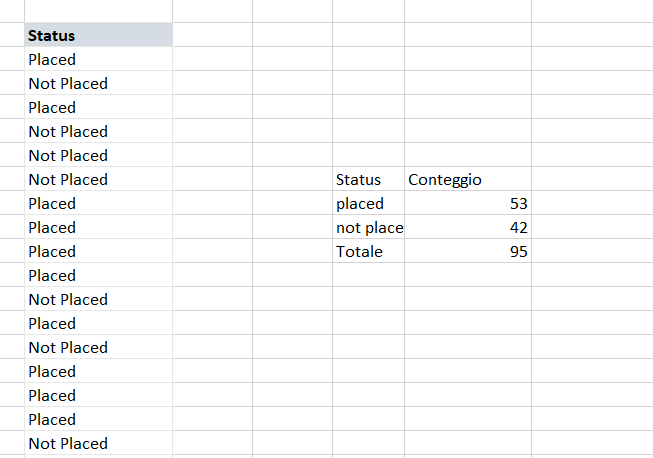


Infatti, placed si ripete 53 volte nella nostra variabile.

Eseguiamo lo stesso procedimento per Not placed



Perfetto.

Rendiamo un po più esplicativa la nostra tabella 

E poi si continua con le frequenze relative e percentuali.

**DISTRIBUZIONE DI FREQUENZA PER VARIABILI NUMERICHE**

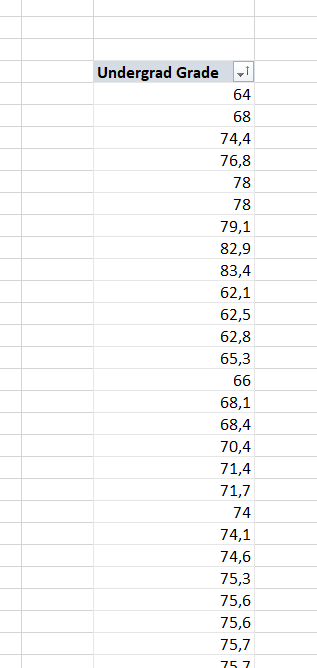
Per quanto riguarda le variabili numeriche, solitamente si calcola il numero di occorrenze che cadono in un determinato range. Il numero di range viene scelto dal data analyst, e viene chiamato bean.

Solitamente, gli intervalli hanno la stessa lunghezza, anche se si possono creare intervalli con lunghezze differenti.

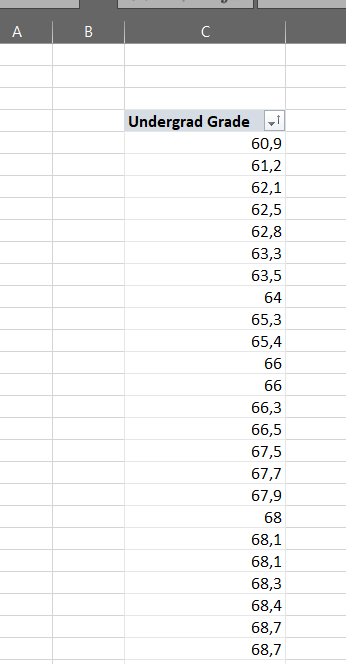
La rappresentazione grafica delle variabili numeriche è solitamente un istogramma.

Riprendiamo quindi il nostro foglio excel distribuzione\_di\_frequenza nella nostra cartella di lavoro excel\_per\_studio.

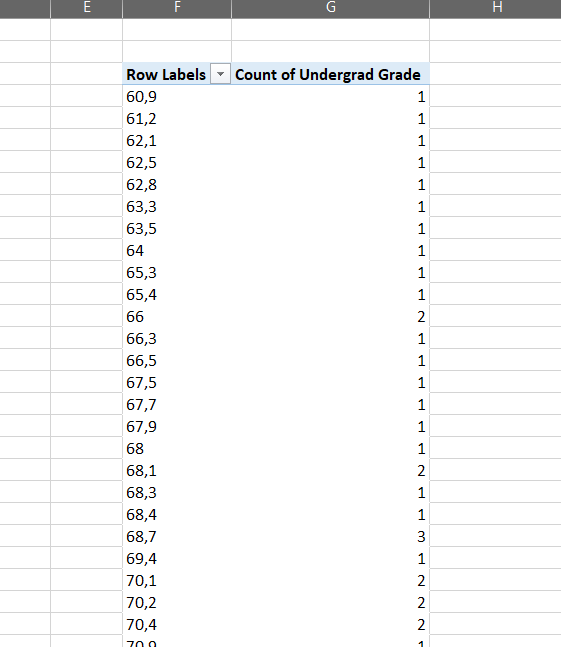
Prendiamo in considerazione la colonna Undergrad Grade, che indica i voti di n studenti presi ai vari esami.

****

**Ordiniamoli dal più piccolo al più grande.**

****

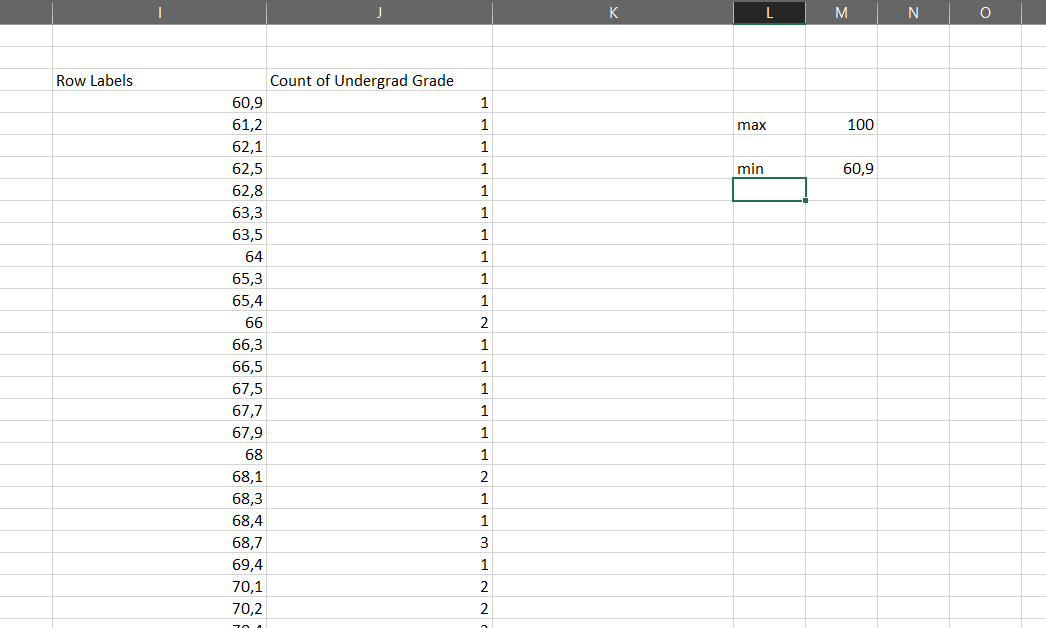
**Contiamo ora le occorrenze tramite tabella pivot.**

****

**Ovviamente abbiamo una o massimo due occorrenze. Non si può andare avanti così.**

**Dobbiamo suddividere le nostre osservazioni in intervalli.**

**Per prima cosa calcoliamo il massimo e il minimo delle nostre osservazioni, utilizzando le funzioni max e min.**

****

**Ora scegliamo il numero di intevalli. Supponiamo di voler suddividere le nostre osservazioni in 6 intervalli.**

**Per prima cosa calcoliamo**

**Facciamo right click su una qualsiasi delle osservazioni.**