

# Il t-test per le prove di velocità: quanto è veloce la connessione Internet

---

Andy Cheshire

Ogni tanto testo la velocità della mia connessione internet di casa utilizzando siti come <http://speedtest.comcast.net> o <http://www.att.com/speedtest>. La necessità di eseguire questi test potrebbe derivare dalle interfacce “cool-looking” che vengono impiegate sui siti, oppure per verificare se la velocità è quella promessa dal provider.



Recentemente ho iniziato a pensare alla distribuzione di questa velocità. Quanto i dati di velocità potrebbero essere normalmente distribuiti?

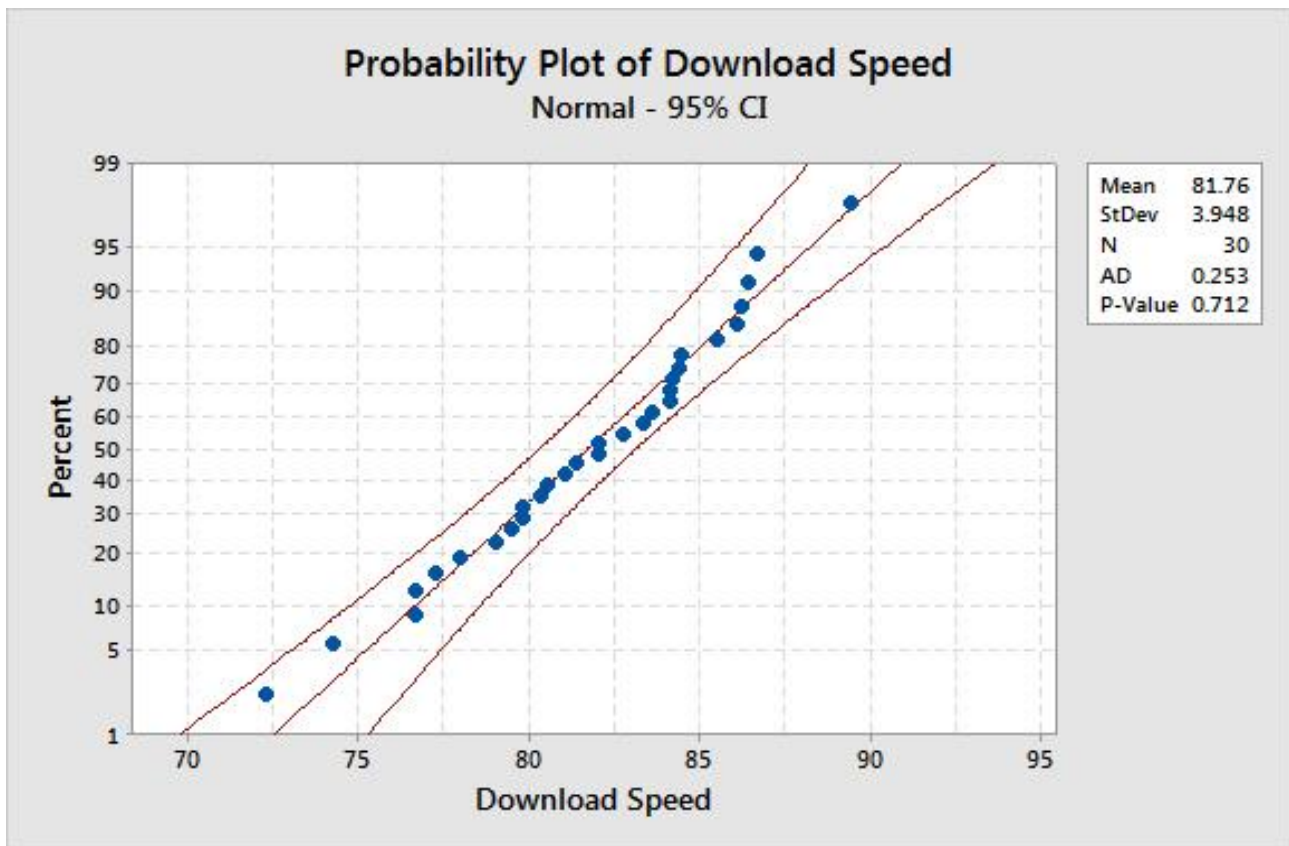
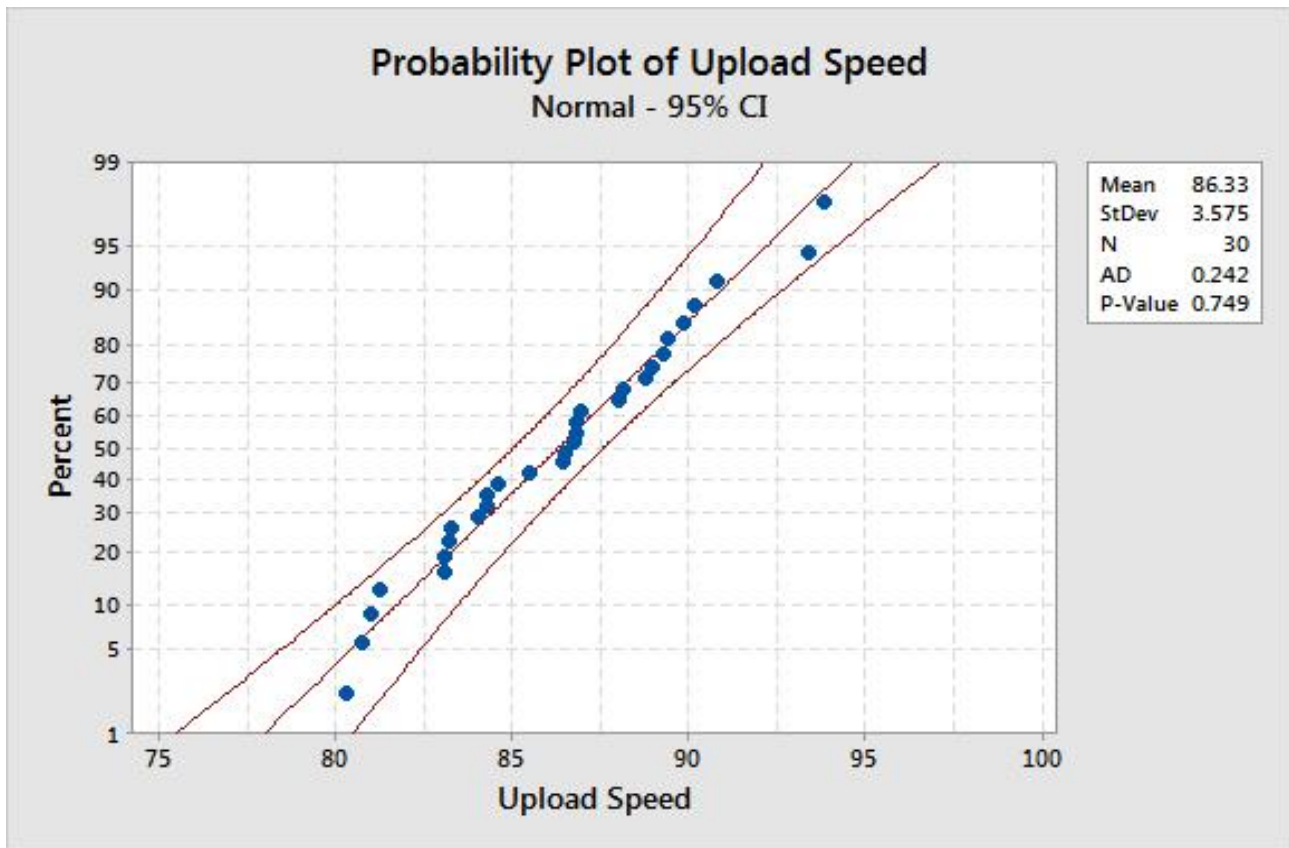
Quando si esegue un test di velocità di Internet, vengono date una velocità di download e una di upload stimate. La velocità di download è la velocità con cui i dati viaggiano da Internet al dispositivo; la velocità di upload è la velocità con cui i dati viaggiano dal dispositivo ad internet. Ero anche curioso di vedere se le velocità medie delle due popolazioni fossero statisticamente differenti.

## I dati sono normalmente distribuiti?

Ho effettuato 30 prove di velocità dal mio ufficio registrandole in un foglio-dati di Minitab; ecco un esempio:

| ↓  | C1             | C2           |
|----|----------------|--------------|
|    | Download Speed | Upload Speed |
| 1  | 82.05          | 83.31        |
| 2  | 77.27          | 84.31        |
| 3  | 79.08          | 86.97        |
| 4  | 74.27          | 89.31        |
| 5  | 76.71          | 81.26        |
| 6  | 85.51          | 84.64        |
| 7  | 84.40          | 88.76        |
| 8  | 84.25          | 84.06        |
| 9  | 79.86          | 83.09        |
| 10 | 78.04          | 86.75        |
| 11 | 72.31          | 83.21        |
| 12 | 79.84          | 86.49        |
| 13 | 81.41          | 83.12        |
| 14 | 83.67          | 88.18        |
| 15 | 84.18          | 86.81        |
| 16 | 79.51          | 80.77        |
| 17 | 76.71          | 86.45        |
| 18 | 84.14          | 86.84        |
|    | ---            | ---          |

Seleziono il comando **Stat > Basic Statistics > Normality Test**. Qui ci sono i grafici di probabilità per le velocità di download e upload.

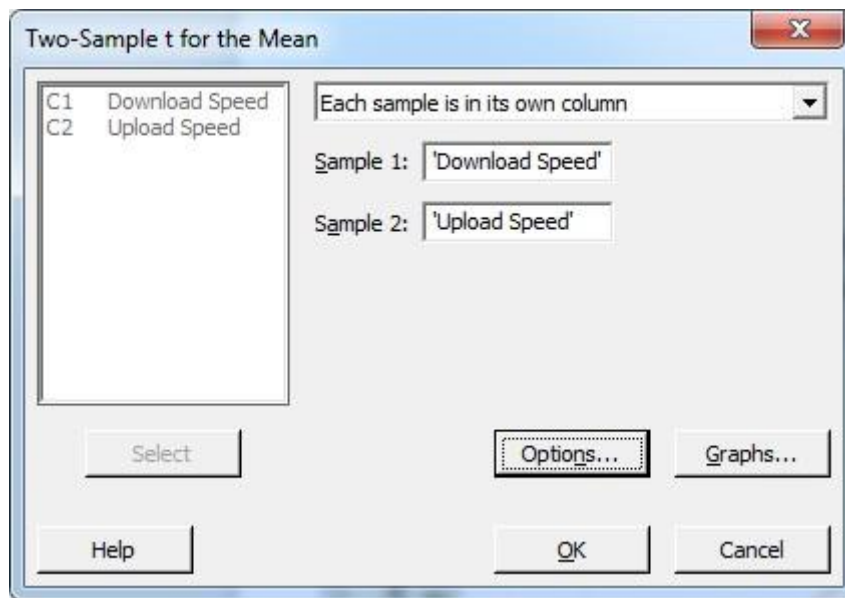


Utilizzo un livello di alfa di 0.05 per confrontare i p-value ottenuti dai test di normalità di *Anderson-Darling*. Entrambi i grafici di probabilità mostrano che i p-value sono maggiori di alfa e quindi non abbiamo prove sufficienti per rifiutare l'ipotesi nulla di normalità dei dati.

# C'è una differenza statisticamente significativa tra la velocità di upload e quella di download?

Scopriamo se vi è una differenza statisticamente significativa tra la velocità media di download e la velocità media di upload.

Seleziono il comando **Stat > Basic Statistics > 2-Sample t**:



Ho scelto **“Each Sample is in its own column”** sotto il menu a discesa, ed ho inserito la colonna **“Download Speed”** in **Sample 1** e la colonna **“Upload Speed”** in **Sample 2**.

Se si fa clic su **Options** viene visualizzata una casella di controllo **“Assume Equal Variances”**. Selezionando questa casella il test inferenziale sarà più potente. Ma come faccio a sapere se le varianze sono uguali o no? Utilizzando un apposito test di Minitab!

Ho cancellato la finestra di dialogo del **2-Sample t** e ho subito implemento il test **Equal Variances (Stat > Basic Statistics > 2 Variances)** ottenendo i seguenti risultati:

Method

Null hypothesis  $\sigma(\text{Download Speed}) / \sigma(\text{Upload Speed}) = 1$   
Alternative hypothesis  $\sigma(\text{Download Speed}) / \sigma(\text{Upload Speed}) \neq 1$   
Significance level  $\alpha = 0.05$

F method was used. This method is accurate for normal data only.

Statistics

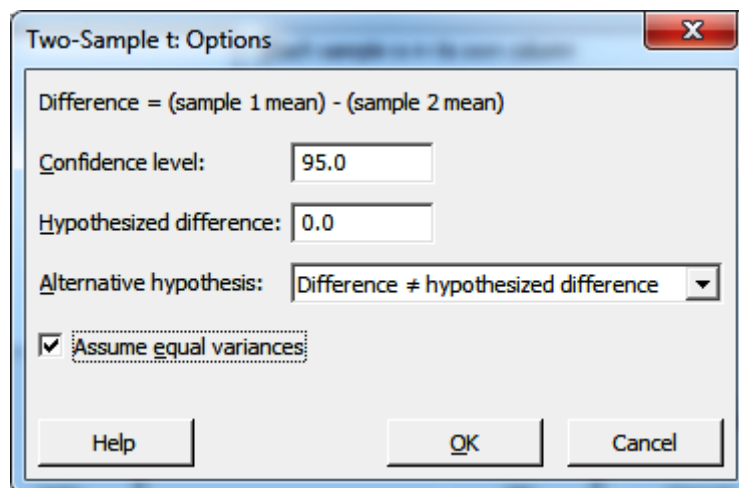
| Variable       | N  | StDev | Variance | 95% CI for StDevs |
|----------------|----|-------|----------|-------------------|
| Download Speed | 30 | 3.948 | 15.586   | (3.144, 5.307)    |
| Upload Speed   | 30 | 3.575 | 12.779   | (2.847, 4.806)    |

Ratio of standard deviations = 1.104  
Ratio of variances = 1.220

Tests

| Method | DF1 | DF2 | Test Statistic | P-Value |
|--------|-----|-----|----------------|---------|
| F      | 29  | 29  | 1.22           | 0.596   |

Considerato che il p-value è maggiore di alfa di 0.05, non abbiamo prove sufficienti per dire che le due varianze siano statisticamente differenti. Pertanto, siamo in grado di tornare al **2-Sample t test** e selezionare la casella **"Assume Equal Variances"**.



Ecco l'output dal mio **2-Sample t-test**:

```
Two-sample T for Download Speed vs Upload Speed

      N   Mean  StDev  SE Mean
Download Speed  30  81.76   3.95    0.72
Upload Speed   30  86.33   3.57    0.65
|

Difference =  $\mu$  (Download Speed) -  $\mu$  (Upload Speed)
Estimate for difference: -4.571
95% CI for difference: (-6.517, -2.625)
T-Test of difference = 0 (vs  $\neq$ ): T-Value = -4.70  P-Value = 0.000  DF = 58
Both use Pooled StDev = 3.7659
```

Dato che il p-value è inferiore a 0.05, possiamo rifiutare l'ipotesi nulla (che entrambe le medie siano le stesse) e dire che le medie delle popolazioni per la velocità di download e della velocità di upload sono statisticamente differenti.

## Vrrrrroooooooooom!

Ero curioso di sapere perché le velocità di upload fosse superiore alla velocità di download durante la mia prova. Ogni volta che ho provato la velocità della connessione internet di casa mia, ho sempre visto il contrario.

Ho chiesto a qualcuno qui a Minitab che è esperto nella configurazione di rete e mi ha detto che avrebbe potuto esserci un utilizzo della banda maggiore del normale, al momento della raccolta dei dati. Questo utilizzo extra può spingere la velocità di download ad essere inferiore delle velocità di upload. Mi è stato anche detto che come viene configurata la connessione internet può contribuire a questo risultato.

Se viene fornita dal provider anche la velocità attesa di download, si potrebbe aggiungere a questo esperimento anche un **1-Sample t-test**. La velocità di download prevista potrebbe essere utilizzata come media target. Si sarebbe quindi in grado di eseguire il test di ipotesi per vedere se la media campionaria è statisticamente diversa dalla media ipotizzata.