# Definitions

$(x, y, s) =$ data ( x,y ), y-uncertainty $s$

$n =$ number of data points

$N =$ number of parameters in the fit

$m =$ Best fit model curve

$r =$ residual $= y - m$

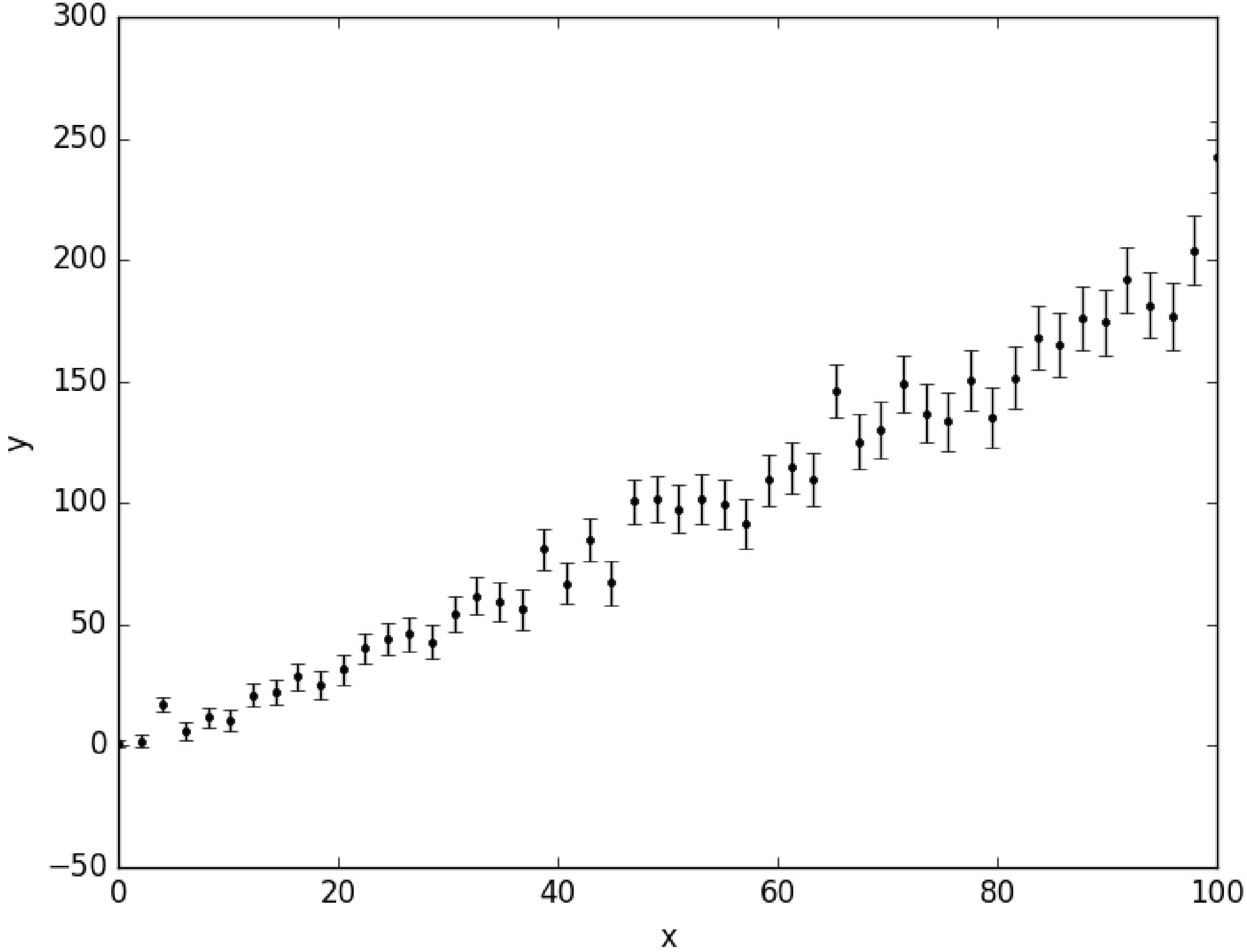
$\tilde{r} =$ normed residual $= r / s$

$\chi^2 = \sum \tilde{r}^2$

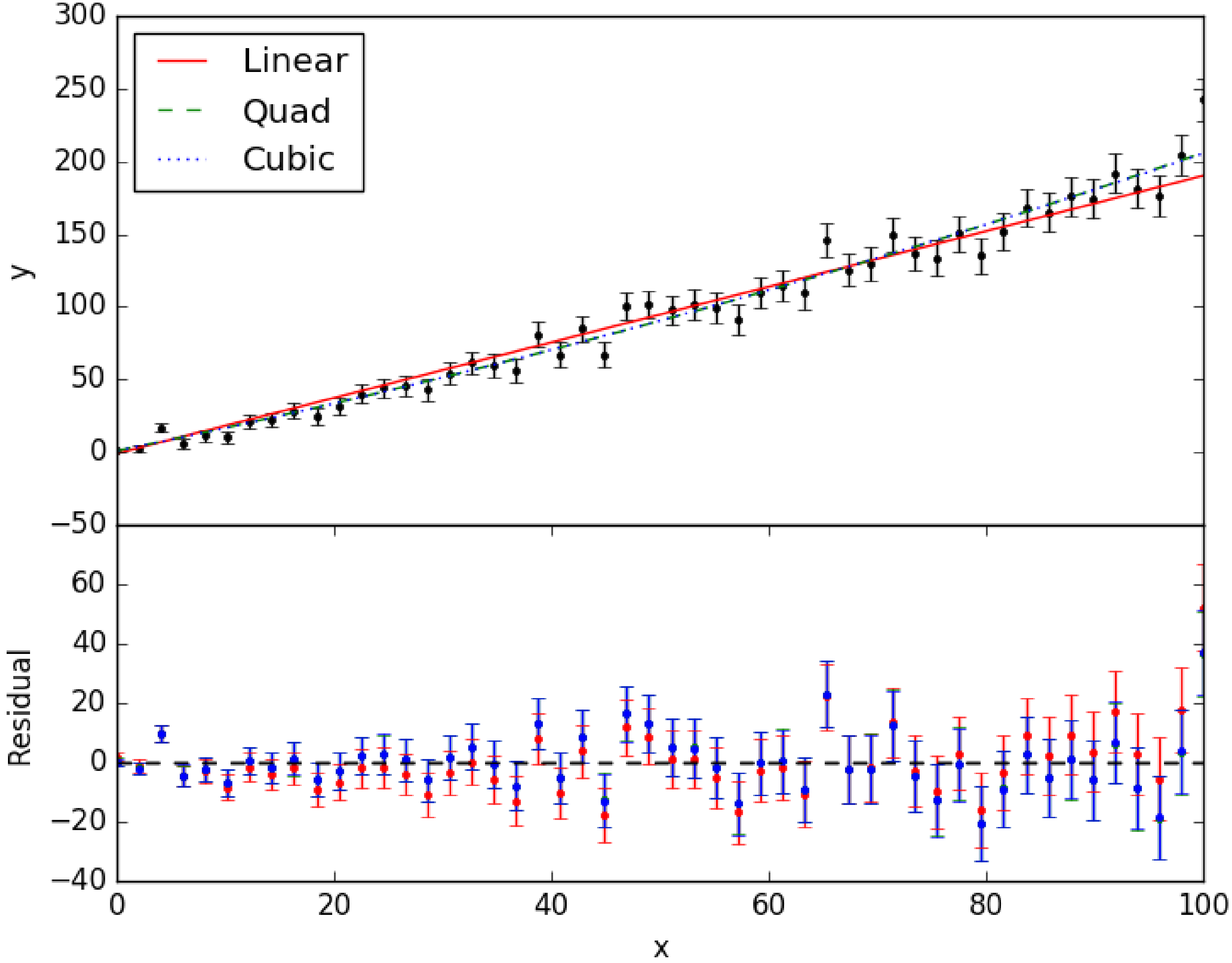
dof $=$ Degrees of freedom $= n - N$

$\tilde{\chi}^2 =$ Reduced $\chi^2 = \chi^2/\mathrm{dof}$

# Linear(?) Data

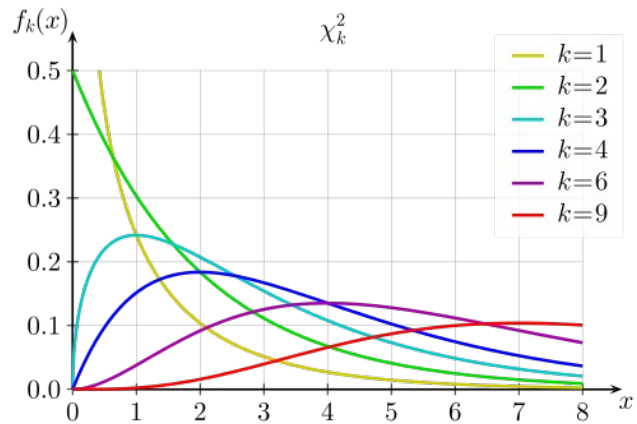


# Occam's Razor:
# Which Model is Correct?



# χ² PDF

$$\mathrm{PDF}(\chi^2;\mathrm{dof}) = \frac{(\chi^2)^{\left(\frac{\mathrm{dof}}{2}-1\right)}\exp(-\chi^2/2)}{2^{\mathrm{dof}/2}\,\Gamma(\mathrm{dof}/2)}$$




Quality-of-fit: What are the odds of getting a poorer value of $\chi^2$ by chance?

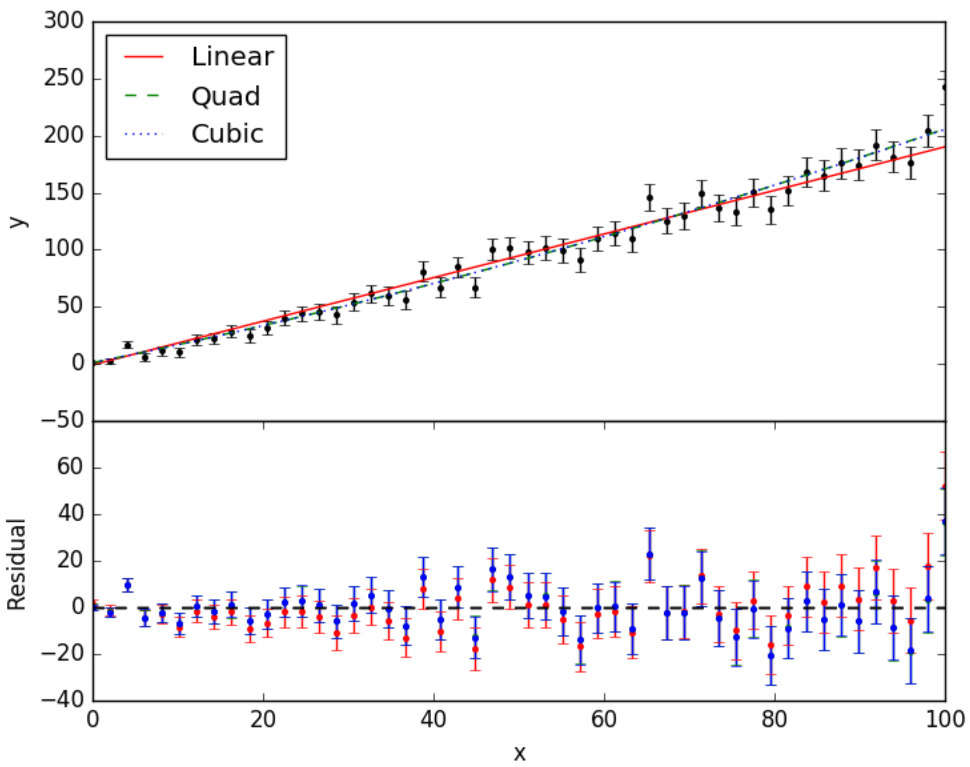
$$\text{Probability of poorer fit by chance} = \int_{\chi^2}^{\infty} \mathrm{PDF}(x^2;\mathrm{dof})\,dx^2$$

Small probability = unlikely to occur by chance; poor fit.

Illustration from Wikipedia

# Occam's Razor: Reduced $\chi^2$

| Model | $\chi^2$ | dof | $\chi^2$/dof | Prob (%) |
|---|---|---|---|---|
| Linear | 68.8 | 48 | 1.43 | 2.6 |
| Quadratic | 52.5 | 47 | **1.12** | **26.9** |
| Cubic | 52.5 | 46 | 1.14 | 23.6 |




```
from scipy.stats import chi2
print 1.0 - chi2.cdf(68.8, 48)
print 1.0 - chi2.cdf(52.5, 47)
print 1.0 - chi2.cdf(52.5, 46)
```

# Occam's Razor: Information Criteria

**Akaike Information Criterion (AIC)**

$$\mathrm{AIC} = \chi^2 + 2N$$

Lowest AIC chooses the model that maximizes information entropy.

**Bayesian Information Criterion (BIC)**

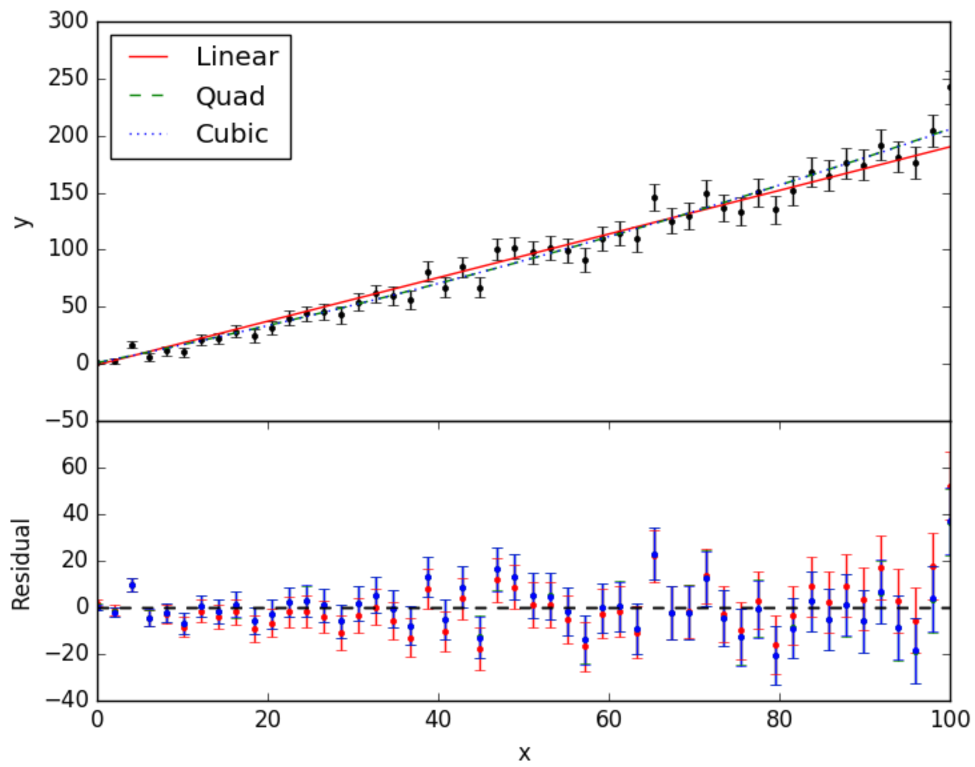
$$\mathrm{BIC} = \chi^2 + N \ln(n)$$

Lowest BIC chooses the best fitting model by integrating the conditional probability of measurements given the model and parameters.

Assumes *n* is large.

# Occam's Razor - Results

| Model | $\chi^2$ | dof | $\chi^2$/dof | Prob (%) | AIC | BIC |
|---|---|---|---|---|---|---|
| Linear | 68.8 | 48 | 1.43 | 2.6 | 72.8 | 76.6 |
| Quadratic | 52.5 | 47 | **1.12** | **26.9** | **58.5** | **64.2** |
| Cubic | 52.5 | 46 | 1.14 | 23.6 | 60.5 | 68.1 |




```
print 'AIC = ', chisq + \
    2. * (order+1)
print 'BIC = ', chisq + \
    (order+1) * log(1.0*len(x))
```

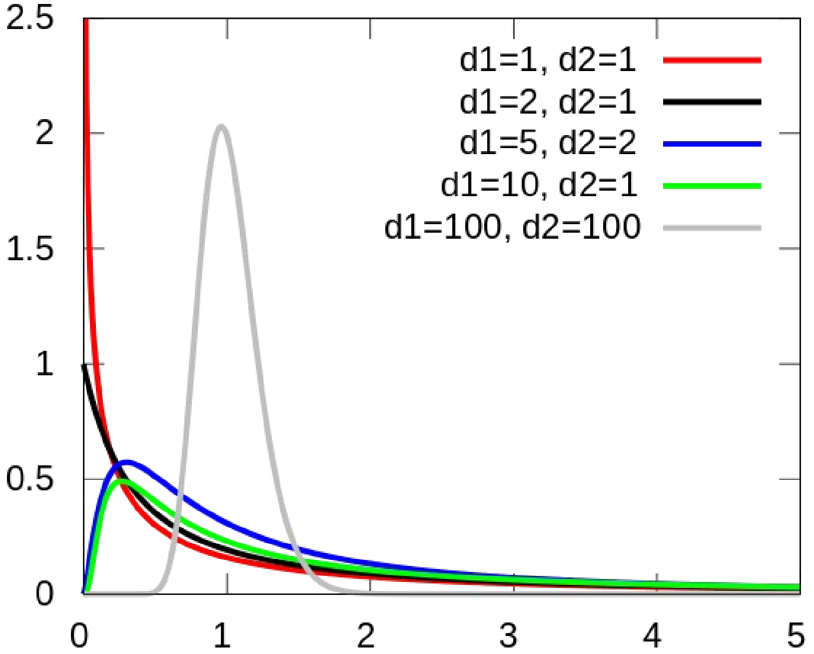
# Occam's Razor: F-test

Consider two models, where model 1 is **nested** within model 2.

**Example:** A linear model is nested within a quadratic model.

The (Fisher's) F-statistic is given by,

$$F = \left( \frac{\chi_1^2 - \chi_2^2}{\chi_2^2} \right) \left( \frac{n - N_2}{N_2 - N_1} \right).$$

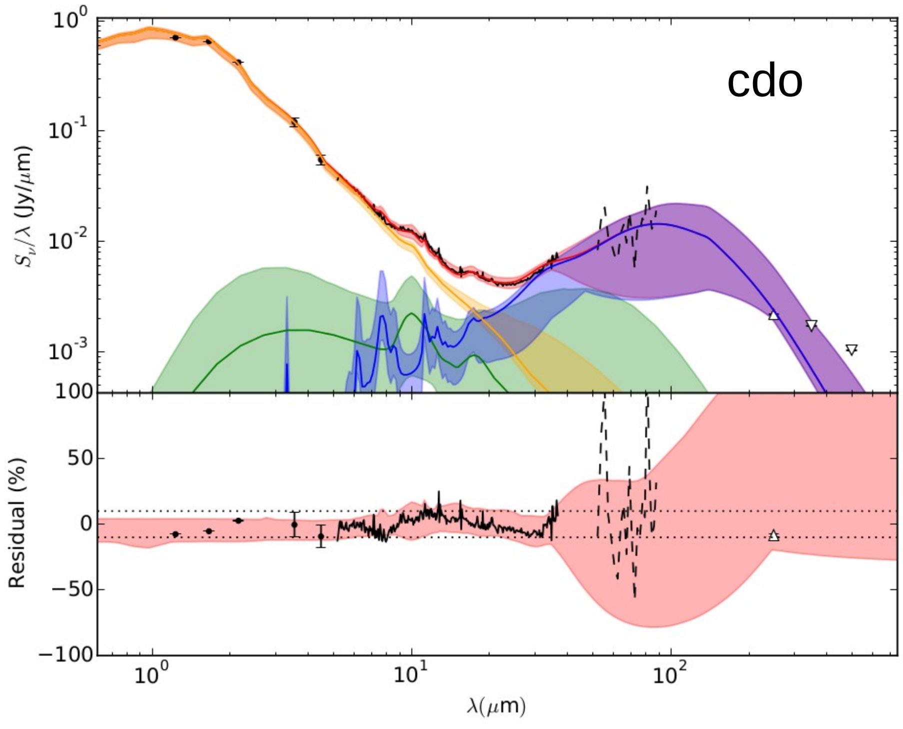


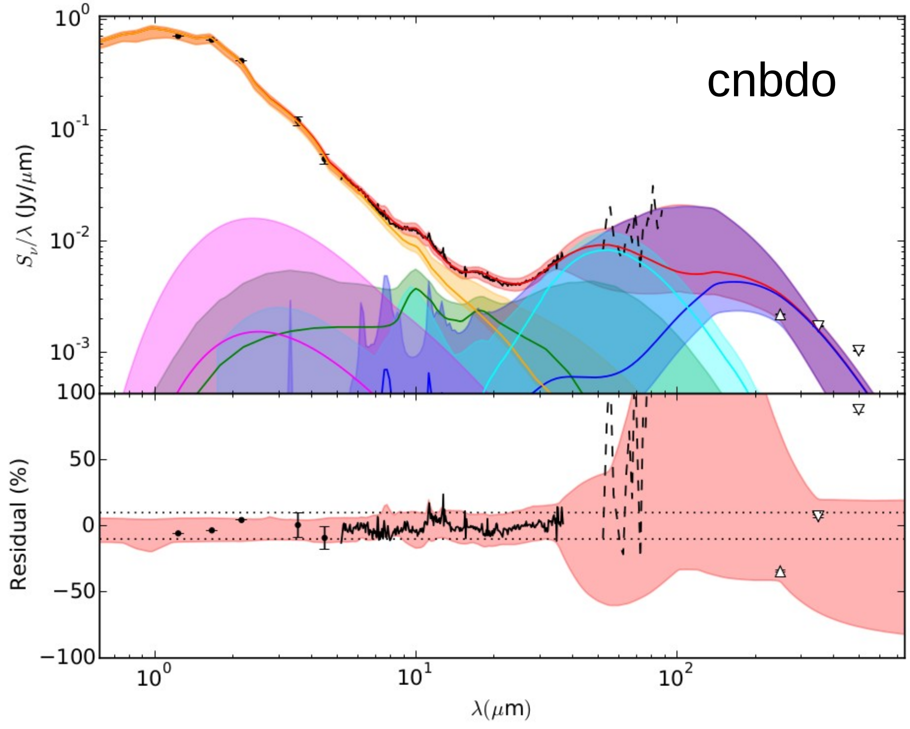

The F-statistic follows the F-distribution, $F(N_2 - N_1, n - N_2)$. Large values of the statistic suggests significantly better fit at higher order.

Illustration from Wikipedia.

# Practical Example: SED Fitting

| Model | Chi-squared | BIC |
|---|---|---|
| cdo | 381.1 | **453** |
| cbdo (add hot dust) | 379.9 | 464 |
| cndo (add ionized gas) | 373.6 | 457 |
| cnbdo (add both) | **372.0** | 468 |




Kharb et al. (2016) MNRAS

# Occam's Razor - Results

| Model | $\chi^2$ | dof | F-stat | Prob (%) |
|---|---|---|---|---|
| Linear | 68.8 | 48 | | |
| Quadratic | 52.5 | 47 | **14.6** | **0.040** |
| Cubic | 52.5 | 46 | 1.14 | 92.3 |

Prob (%) = odds of getting such a large value of the F-statistic by chance.

```
from scipy.stats import f
F = (68.75668121 – 52.491355)/(3. - 2.) * (50. - 3)/ 52.491355
print 1.0 – f.cdf(F, 3.-2., 50.-3.) # 1 - CDF
```

# Does my "G" Agree with Accepted "G?"

Use the "z" statistic
(assumes normally distributed uncertainties).

$$\text{measured} = G_m \pm \delta G_m$$

$$\text{accepted} = G_a \pm \delta G_A$$

$$\text{z statistic} = \frac{\left(G_m - G_a\right)}{\sqrt{\delta G_m^2 + \delta G_a^2}}$$

Under the "null hypothesis," z follows a Gaussian distribution, mean = 0, stdev = 1.0.

Two sided test: the odds of getting such a large *z* is given by,

$$\text{Probability} = 2\int_{|z|}^{\infty} \text{PDF}(z)\,dz = 2\left(1 - \text{CDF}(z)\right)$$

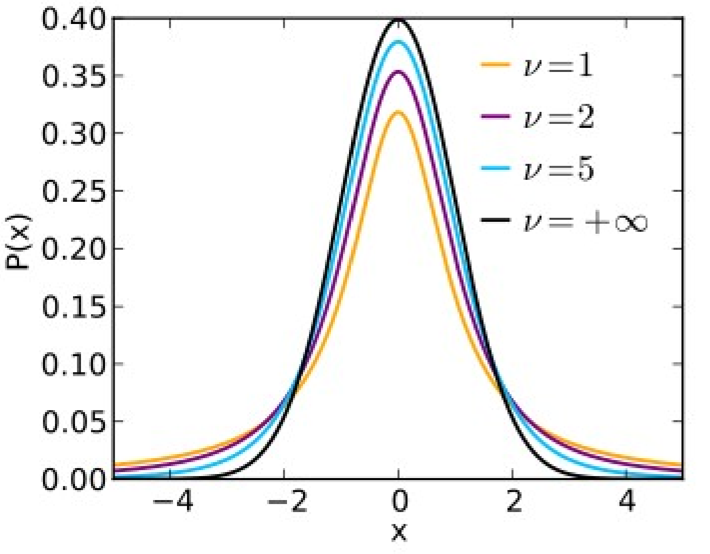
# Repeated Measurements: Student's t-test

measured = $\bar{x} = \sum (x_i)/n$

uncertainty = $s/\sqrt{n}$

accepted = $\mu$

$$t = \frac{(\bar{x} - \mu)}{s/\sqrt{n}} = (\text{ mean - accepted }) / (\text{ SEM })$$




The t-statistic follows a Student's t-distribution with $\nu = n - 1$.

The probability of getting a large value of $t$ by chance is (two-sided),

$$\text{Probability} = 2\int_{|z|}^{\infty} \text{PDF}(z; n-1)\, dz = 2\left(1 - \text{CDF}(z; n-1)\right)$$

Illustration from Wikipedia.