Week 12

Introduction to Data Analysis for Physics

Overview

- Z-Scores
- P-Values
- Confidence Intervals
- Model Fits

Z-Scores

- Back to the Normal distribution $N_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Re-scale by $z(x) = \frac{x - \mu_x}{\sigma_x}$
- Recall with *linear* functions, the mean is $\mu_f = f(\mu_x)$
- And variance is $\sigma_f^2 = \left(\frac{\partial f}{\partial x}\bigg|_{x=\mu_x} \right)^2 \sigma_x^2$

So what are μ_z, σ_z in terms of μ_x, σ_x

Z-Scores

- This is a *normalizing* transformation
  - Puts things in a standard form:
  - $P(|z|<1)=.68$
  - $P(|z|<2)=.95$
  - $P(|z|<3)=.997$
- So 99.7% of data should be within 3 standard deviations
- Obtained from integral: $P(|z| < z^*) = \int_{-z^*}^{z^*} N_{\mu,\sigma}(z)dz$

P-Value

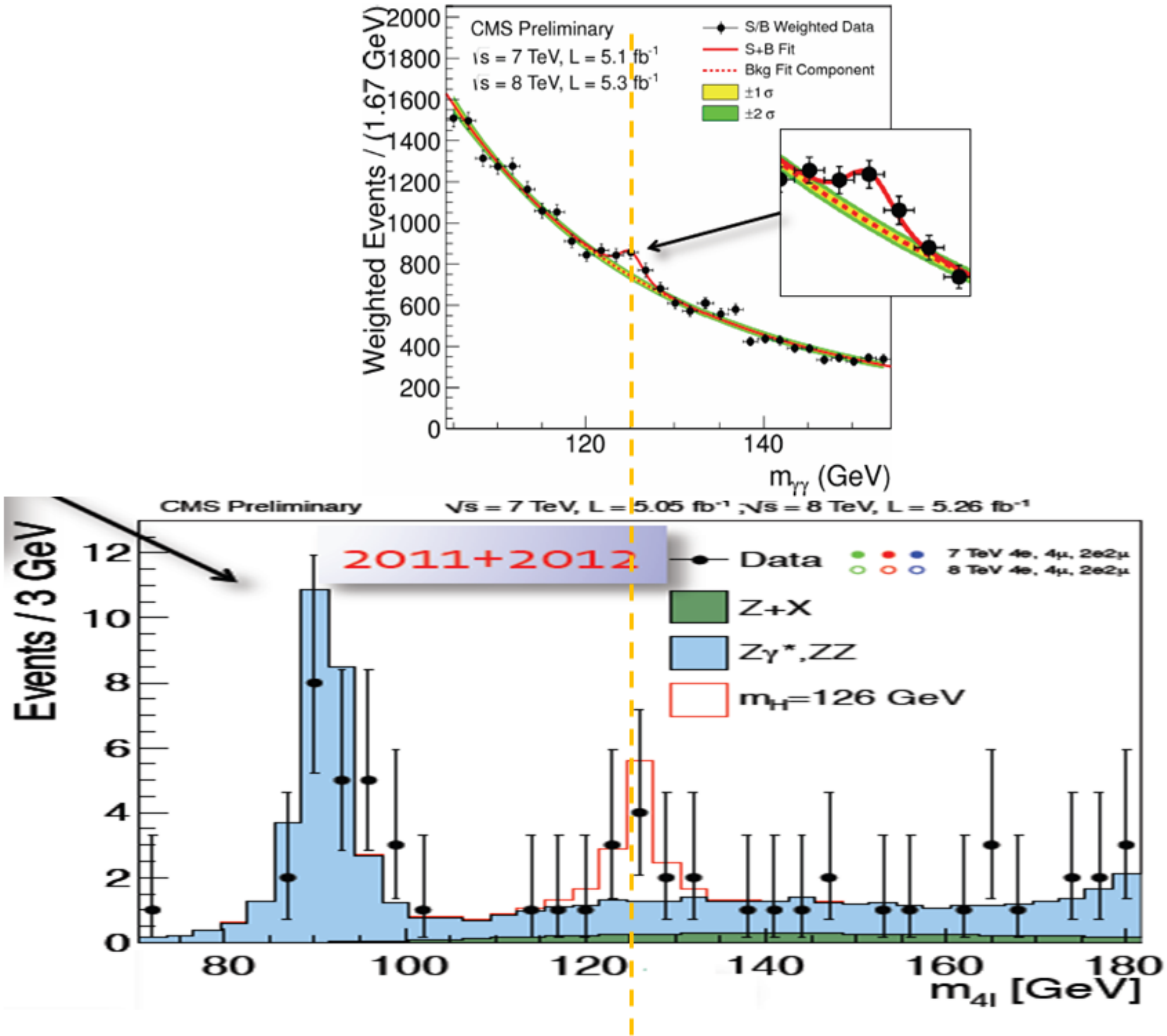
- Related to probabilities in last slide:

$$p = \frac{1 - P(|z| < z^*)}{2}$$

- This is the probability *under the assumption of a "null hypothesis,"* of obtaining a result as strange as we did.
- So if $p < .05$, only 1 in 20 trials should give a result this strange

P-Value

- Example: Higgs discovery
- Higgsdependence Day 2012, CMS and ATLAS release results
- Individually, find $p<0.000001$ relative to background expectation (the null hypothesis)
- Graph on next page


CMS Preliminary
√s = 7 TeV, L = 5.1 fb⁻¹
√s = 8 TeV, L = 5.3 fb⁻¹
S/B Weighted Data
S+B Fit
Bkg Fit Component
±1 σ
±2 σ
Weighted Events / (1.67 GeV)
2000
1800
1600
1400
1200
1000
800
600
400
200
0
120
140
m_γγ (GeV)
CMS Preliminary
√s = 7 TeV, L = 5.05 fb⁻¹ ; √s = 8 TeV, L = 5.26 fb⁻¹
2011+2012
Data
7 TeV 4e, 4μ, 2e2μ
8 TeV 4e, 4μ, 2e2μ
Z+X
Zγ*,ZZ
m_H=126 GeV
Events / 3 GeV
12
10
8
6
4
2
0
80
100
120
140
160
180
m_4l [GeV]

Confidence Interval

- Another way to report uncertainty
- Familiar with $\mu \pm \sigma$ notation, such as $125 \pm 0.4 \text{ GeV}$
- Have a "confidence level" of 0<C<1 such that

$$C = P(|z| < z_C)$$

- Solve for z_C and use to find min and max values

$$x_\pm = \mu_x \pm \sigma_x z_C$$

- Useful to say whether 0 is included at 99% confidence

Confidence Interval

- If 0 included, then cannot rule it out due to statistical error
- If 0 excluded, can rule out *at that confidence level*
- With Normal distribution, cannot have 100% confidence. Ever.

Model Fits

- Back to *Mathematica*!
- NonlinearModelFit[data, model, params, vars]

```
data = Table[{i, 2 i^3 + Random[]*.02}, {i, 10}];
fit = NonlinearModelFit[data, a x^b, {a, b}, x]
fit["ParameterTable"]
```

Model Fits

- FittedModel provides statistics on how good the fit was
- Careful with interpreting values (P-Value is not probability this result is right, or anything of the sort)
- Effectively useful stats are estimate and error for modeling in error propagation