Lecture 12: Experimental Evaluation

Physics 3719
Spring Semester 2011
Lab 3 Experiments

- Gravitational Constant $G$ Jensen, Blackmon (AM)
- Speed of Light $C$ Ramirez, Thomas (PM)
- Electron Charge-to-Mass Ratio $e/m$ Brown, Blatter (PM)
- Millikan Oil Drop Experiment Baehr, Martineau (AM)
- Photoelectric Effect Peterson, Ingebretsen (PM)
- Hydrogen Balmer Spectrum Gibbs, Baird (PM)
- Franck-Hertz Experiment Rowley, Bernal (AM)
- High-Res Optical Spectroscopy Richards, Topham (AM)
Purpose of Experiments

- Purpose of experiment: test existing model
- “Model” (Webster) – a system of postulates, data, and inferences presented as a mathematical description of an entity or state of affairs
- i.e. a functional relationship between variables
Experimental Evaluation

- Determine significance of measurement
- What does experiment say about the model you're testing?
  - Does it support or refute the model?
  - Can we quantify our confidence in the model?
What can experiments achieve?

- An experiment cannot **prove** a model to be true.
- An experiment can:
  - Demonstrate that a model is **false** with some probability
  - Show that a model is **valid** within some precision
What can experiments achieve?

- An experiment improves the strength of its statements by decreasing uncertainties.
- How do I quantify the extent to which the data to the right is inconsistent with Ohm's law?
Gaussian Distribution

It's usually useful to assume (and often true!) that uncertainties in experimental measurements are distributed according to the Gaussian or Normal error distribution:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\bar{x})^2/2\sigma^2}$$

As discussed last time, in the case of high-statistics counting experiments this becomes

$$P_G(n; \mu) = \frac{1}{\sqrt{2\pi\sigma_n}} e^{-(n-\mu)^2/2\sigma_n^2}$$

Where the standard deviation is given by

$$\sigma_n = \sqrt{\mu}$$

and the error in the mean is given by

$$\sigma_{\bar{n}} = \frac{\sigma_n}{\sqrt{N}}$$
Aside: Measurement Error

- Fit Gaussian to a flat distribution in interval $\Delta x$.
- For a measurement in some interval, but you're not sure exactly where in the interval, best guess is

$$x = x_{mid} \pm \frac{\Delta x}{\sqrt{12}}$$
Examples

- Reading lengths, *et cetera* off a ruler with tiny tick marks
- Reading numbers off of digital voltmeter
  - Answer really between 12.875 V and 12.885 V, but you don't know where.
  - Report 12.880 ± 0.003 V
Its usually useful to assume (and often true!) that uncertainties in experimental measurements are distributed according to the *Gaussian* or *Normal* error distribution:

\[
f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}
\]

As discussed last time, in the case of high-statistics counting experiments this becomes

\[
P_G(n; \mu) = \frac{1}{\sqrt{2\pi\sigma_n}} e^{-\frac{(n-\mu)^2}{2\sigma_n^2}}
\]

Where the standard deviation is given by

\[
\sigma_n = \sqrt{\mu}
\]

and the error in the mean is given by

\[
\sigma_{\bar{n}} = \frac{\sigma_n}{\sqrt{N}}
\]

A “1 sigma” error interval is not all-inclusive!
Use “2/3” Rule as a Rough Guide

- Spread of points, error bars consistent with normal error distribution.
Use “2/3” Rule as a Rough Guide

- What is the problem here?
Use “2/3” Rule as a Rough Guide

- What is the problem here?
Recall from Least-Squares Fitting: $\chi^2$

- Problem: find parameters of straight line (or arbitrary curve) which best fits a set of data
- Mechanical analogue: PE minimization
- Least-squares: $\chi^2$ minimization.

$$\chi^2 = \sum_{i=1}^{n} \left( \frac{f(x_i) - y_i}{\sigma y_i} \right)^2$$
Evaluating Models: $\chi^2$

- Saw before, in performing linear least-squares fits by minimizing $\chi^2$
- Definition of $\chi^2$

$$\chi^2 = \sum_{i=1}^{n} \frac{(f(x_i) - y_i)^2}{\sigma^2 y_i}$$

- Absolute magnitude of $\chi^2$ only has meaning relative to the number of degrees of freedom

- $\#DOF = (#\ of\ data\ points) - (#\ parameters\ determined\ from\ the\ data)$

- $\#DOF = (#\ of\ data\ points) - (#\ constraints)$

- $\chi^2 / \#DOF \sim 1$ for good fit
$\chi^2$ is a quantitative way to evaluate agreement between data and a model:

$$f(x_i) = \frac{(f(x_i) - y_i)^2}{\sigma_{y_i}^2}$$

<table>
<thead>
<tr>
<th>$i$</th>
<th>$x_i$</th>
<th>$2 + x/2$</th>
<th>$y_i$</th>
<th>$\sigma_{y_i}$</th>
<th>$\frac{(f(x_i) - y_i)^2}{\sigma_{y_i}^2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
<td>2.25</td>
<td>2.703</td>
<td>1.0</td>
<td>0.205</td>
</tr>
<tr>
<td>2</td>
<td>1.5</td>
<td>2.75</td>
<td>3.978</td>
<td>1.0</td>
<td>1.508</td>
</tr>
<tr>
<td>3</td>
<td>2.5</td>
<td>3.25</td>
<td>3.108</td>
<td>1.0</td>
<td>0.020</td>
</tr>
<tr>
<td>4</td>
<td>3.5</td>
<td>3.75</td>
<td>3.953</td>
<td>1.0</td>
<td>0.041</td>
</tr>
<tr>
<td>5</td>
<td>4.5</td>
<td>4.25</td>
<td>3.638</td>
<td>1.0</td>
<td>0.374</td>
</tr>
<tr>
<td>6</td>
<td>5.5</td>
<td>4.75</td>
<td>5.232</td>
<td>1.0</td>
<td>0.233</td>
</tr>
<tr>
<td>7</td>
<td>6.5</td>
<td>5.25</td>
<td>5.769</td>
<td>1.0</td>
<td>0.269</td>
</tr>
<tr>
<td>8</td>
<td>7.5</td>
<td>5.75</td>
<td>4.366</td>
<td>1.0</td>
<td>1.914</td>
</tr>
<tr>
<td>9</td>
<td>8.5</td>
<td>6.25</td>
<td>4.732</td>
<td>1.0</td>
<td>2.303</td>
</tr>
<tr>
<td>10</td>
<td>9.5</td>
<td>6.75</td>
<td>7.173</td>
<td>1.0</td>
<td>0.179</td>
</tr>
</tbody>
</table>

$$\chi^2 = 7.050$$

What is #DOF?
• What is the problem here?

\[ \chi^2 = 0.783365 \]
• What is the problem here?
Counting Degrees of Freedom

- # Degrees of freedom in this straight line fit should be easy -> what is it??
- Caveat: the total number of events (normalization for probability distribution) is also obtained from the data.
- An example...
Taylor 12.01

Each member of a class of 50 students is given a piece of metal and told to find its density $\rho$. From the 50 results, the mean and standard deviation are calculated, and the class decides to test whether the results are normally distributed. The measurements are grouped into four bins:

<table>
<thead>
<tr>
<th>Bin</th>
<th>Range of Bin</th>
<th>Observed Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\rho &lt; \bar{\rho} - \sigma_\rho$</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>$\bar{\rho} - \sigma_\rho &lt; \rho &lt; \bar{\rho}$</td>
<td>13</td>
</tr>
<tr>
<td>3</td>
<td>$\bar{\rho} &lt; \rho &lt; \bar{\rho} + \sigma_\rho$</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>$\bar{\rho} + \sigma_\rho &lt; \rho$</td>
<td>14</td>
</tr>
</tbody>
</table>

- Assuming the measurements were normally distributed, calculate the expected number of measurements in each bin.
- Calculate $\chi^2$. Do the measurements seem to be normally distributed?
What are expectations for each bin?

- Use what we know about normal distribution

\[
\frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)
\]
Already, we can see that data does not seem to follow a “bell curve”

What do we need in order to make a more quantitative statement?
Assign Uncertainty to \textit{Expectation}

- $\sqrt{\text{N}_{\text{exp}}}$ error
- Why $\text{N}_{\text{exp}}$ not $\text{N}_{\text{obs}}$?
Now compute $\chi^2$

$$\chi^2 = \sum_{i=1}^{n} \frac{(f(x_i) - y_i)^2}{\sigma_{yi}^2}$$

$$\chi^2 = \sum_{i=1}^{n} \frac{(E_i - O_i)^2}{E_i}$$

- What is #DOF?
- What do we conclude about the data?
Continued on Monday...
HW Problem: *Direct* minimization of $\chi^2$ ...

- In class, worked out analytic solution to $\chi^2$ minimization for linear fits.
- This can also be done graphically...
HW Problem: *Direct* minimization of $\chi^2$ ... 

- In class, worked out analytic solution to $\chi^2$ minimization for linear fits.
- This can also be done graphically...
HW Problem: *Direct* minimization of $\chi^2$ ...

- In class, worked out analytic solution to $\chi^2$ minimization for linear fits.
- This can also be done graphically...
HW Problem: *Direct* minimization of $\chi^2$ ...  

- In class, worked out analytic solution to $\chi^2$ minimization for linear fits.
- This can also be done graphically...
HW Problem: *Direct* minimization of $\chi^2$ ...

- In class, worked out analytic solution to $\chi^2$ minimization for linear fits.
- This can also be done graphically...
HW Problem: *Direct* minimization of $\chi^2$ ...

- In class, worked out analytic solution to $\chi^2$ minimization for linear fits.
- This can also be done graphically...
HW Problem: *Direct* minimization of $\chi^2$ ...

- In class, worked out analytic solution to $\chi^2$ minimization for linear fits.
- This can also be done graphically...
HW Problem: *Direct* minimization of $\chi^2$ ...

- In class, worked out analytic solution to $\chi^2$ minimization for linear fits.
- This can also be done graphically...
HW Problem: *Direct* minimization of $\chi^2$ ...

- In class, worked out analytic solution to $\chi^2$ minimization for linear fits.
- This can also be done graphically...
HW Problem: *Direct* minimization of $\chi^2$...

- In class, worked out analytic solution to $\chi^2$ minimization for linear fits.
- This can also be done graphically...
HW Problem: *Direct* minimization of $\chi^2$ ...

- In class, worked out analytic solution to $\chi^2$ minimization for linear fits.
- This can also be done graphically...
HW Problem: *Direct* minimization of $\chi^2$ ...

- In class, worked out analytic solution to $\chi^2$ minimization for linear fits.
- This can also be done graphically...
HW Problem: *Direct* minimization of $\chi^2$ ...

- Uncertainty can be obtained from $\chi^2$ parabola as well...
New Concept: Chance Probability or p-value

**p-value**: The probability of obtaining a result *at least as extreme* as a given data point, assuming the data point was the result of *chance alone*. 
Example: Suppose there are 30,000 University of Utah students of which 400 carry concealed weapons permits. If JB is teaching an astronomy class of 120 students, what is the probability that one or more is packing?
Example: Suppose there are 30,000 University of Utah students of which 400 carry concealed weapons permits. If JB is teaching an astronomy class of 120 students, what is the probability that one or more is packing?

Answer: The probability distribution for the number of students with weapons is Poissonian, with a mean value of \((\frac{400}{30,000}) \times 120 = 1.6\)

The probability of observing one or more in a Poisson distribution with mean 1.6 is:

\[
P_p(\geq 1, 1.6) = 1 - P_p(0, 1.6) = 1 - 0.201 = 0.799
\]
Follow up: Suppose a metal detector reveals that there are 6 people in JB's class carrying guns. What is the p-value of this observation? Should JB be worried that they're out to get him?
Follow up: Suppose a metal detector reveals that there are 6 people in JB's class carrying guns. What is the p-value of this observation? Should JB be worried that they're out to get him?

- **Answer:** The p-value is the sum of probabilities for \( n = 6, 7, 8, 9 \ldots \)
- **P-value** = \( \text{PP}(\geq 6, 1.6) \)
  \[ = 0.006 \]
- This is a very small chance probability.
- Perhaps JB should wear a kevlar vest!
p-value: The probability of obtaining a result at least as extreme as a given data point, assuming the data point was the result of chance alone

- The smaller the p-value, the more likely the hypothesis is false.
- Example: I flip a coin 20 times, 14 times it comes up heads. Use this data as a test of the hypothesis that the coin is fair.

\[ P = (\text{Binomial Probability of } \geq 14 \text{ heads}) + (\text{Binomial Probability of } \leq 6 \text{ heads}) = 0.115 \]
p-value: The probability of obtaining a result at least as extreme as a given data point, assuming the data point was the result of chance alone.

- A statement about the probability that an “effect” is really a statistical fluctuation.
- Replaces discrete True/False (or 1/0) choice with a continuum ranging from “likely true” to “likely false”.
in the parlance of our time...

\[ p\text{-value} = \text{truthiness} \]
in the parlance of our time...

\[(1 - \text{p-value}) = \text{“wronginess”}\]
**p-value**: The probability of obtaining a result **at least as extreme** as a given data point, assuming the data point was the result of chance alone

- The smaller the p-value, the more likely the hypothesis is false.

*Example*: I flip a coin 20 times, 14 times it comes up heads. Use this data as a test of the hypothesis that the coin is fair.

- \( P = (\text{Binomial Probability of } \geq 14 \text{ heads}) + (\text{Binomial Probability of } \leq 6 \text{ heads}) \) = 0.115

- While the “fair coin” hypothesis does not have a high level of truthiness (p-value), it's not so low that we can say for sure that the coin isn't fair.
The $\chi^2$ and number of degrees of freedom can be directly translated into a p-value:

Source: pdg.lbl.gov
Table D. The percentage probability $\text{Prob}_d(\chi^2 \geq \chi_o^2)$ of obtaining a value of $\chi^2 \geq \chi_o^2$ in an experiment with $d$ degrees of freedom, as a function of $d$ and $\chi_o^2$. (Blanks indicate probabilities less than 0.05%.)

<table>
<thead>
<tr>
<th>$\bar{\chi}_o^2$</th>
<th>0.5</th>
<th>1.0</th>
<th>1.5</th>
<th>2.0</th>
<th>2.5</th>
<th>3.0</th>
<th>3.5</th>
<th>4.0</th>
<th>4.5</th>
<th>5.0</th>
<th>5.5</th>
<th>6.0</th>
<th>8.0</th>
<th>10.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>48</td>
<td>32</td>
<td>22</td>
<td>16</td>
<td>11</td>
<td>8.3</td>
<td>6.1</td>
<td>4.6</td>
<td>3.4</td>
<td>2.5</td>
<td>1.9</td>
<td>1.4</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>61</td>
<td>37</td>
<td>22</td>
<td>14</td>
<td>8.2</td>
<td>5.0</td>
<td>3.0</td>
<td>1.8</td>
<td>1.1</td>
<td>0.7</td>
<td>0.4</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>68</td>
<td>39</td>
<td>21</td>
<td>11</td>
<td>5.8</td>
<td>2.9</td>
<td>1.5</td>
<td>0.7</td>
<td>0.4</td>
<td>0.2</td>
<td>0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>74</td>
<td>41</td>
<td>20</td>
<td>9.2</td>
<td>4.0</td>
<td>1.7</td>
<td>0.7</td>
<td>0.3</td>
<td>0.1</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>78</td>
<td>42</td>
<td>19</td>
<td>7.5</td>
<td>2.9</td>
<td>1.0</td>
<td>0.4</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2</td>
<td>0.4</td>
<td>0.6</td>
<td>0.8</td>
<td>1.0</td>
<td>1.2</td>
<td>1.4</td>
<td>1.6</td>
<td>2.0</td>
<td>2.2</td>
<td>2.4</td>
<td>2.6</td>
<td>2.8</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>65</td>
<td>53</td>
<td>44</td>
<td>37</td>
<td>32</td>
<td>27</td>
<td>24</td>
<td>21</td>
<td>18</td>
<td>16</td>
<td>14</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>82</td>
<td>67</td>
<td>55</td>
<td>45</td>
<td>37</td>
<td>30</td>
<td>25</td>
<td>20</td>
<td>17</td>
<td>14</td>
<td>11</td>
<td>9.1</td>
<td>7.4</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>90</td>
<td>75</td>
<td>61</td>
<td>49</td>
<td>39</td>
<td>31</td>
<td>24</td>
<td>19</td>
<td>14</td>
<td>11</td>
<td>8.6</td>
<td>6.6</td>
<td>5.0</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>94</td>
<td>81</td>
<td>66</td>
<td>52</td>
<td>41</td>
<td>31</td>
<td>23</td>
<td>17</td>
<td>13</td>
<td>9.2</td>
<td>6.6</td>
<td>4.8</td>
<td>3.4</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>96</td>
<td>85</td>
<td>70</td>
<td>55</td>
<td>42</td>
<td>31</td>
<td>22</td>
<td>16</td>
<td>11</td>
<td>7.5</td>
<td>5.1</td>
<td>3.5</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>100</td>
<td>98</td>
<td>88</td>
<td>73</td>
<td>57</td>
<td>42</td>
<td>30</td>
<td>21</td>
<td>14</td>
<td>9.5</td>
<td>6.2</td>
<td>4.0</td>
</tr>
<tr>
<td>7</td>
<td>100</td>
<td>99</td>
<td>90</td>
<td>76</td>
<td>59</td>
<td>43</td>
<td>30</td>
<td>20</td>
<td>13</td>
<td>8.2</td>
<td>5.1</td>
<td>3.1</td>
<td>1.9</td>
<td>1.1</td>
</tr>
<tr>
<td>8</td>
<td>100</td>
<td>99</td>
<td>92</td>
<td>78</td>
<td>60</td>
<td>43</td>
<td>29</td>
<td>19</td>
<td>12</td>
<td>7.2</td>
<td>4.2</td>
<td>2.4</td>
<td>1.4</td>
<td>0.8</td>
</tr>
<tr>
<td>9</td>
<td>100</td>
<td>99</td>
<td>94</td>
<td>80</td>
<td>62</td>
<td>44</td>
<td>29</td>
<td>18</td>
<td>11</td>
<td>6.3</td>
<td>3.5</td>
<td>1.9</td>
<td>1.0</td>
<td>0.5</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>99</td>
<td>95</td>
<td>82</td>
<td>63</td>
<td>44</td>
<td>29</td>
<td>17</td>
<td>10</td>
<td>5.5</td>
<td>2.9</td>
<td>1.5</td>
<td>0.8</td>
<td>0.4</td>
</tr>
</tbody>
</table>

The values in Table D were calculated from the integral

$$\text{Prob}_d(\chi^2 \geq \chi_o^2) = \frac{2}{2^{d/2} \Gamma(d/2)} \int_{\chi_o}^{\infty} x^{d-1} e^{-x/2} \, dx.$$
Now compute $\chi^2$

- #DOF = 1
- P-value = 0.002
- “Normal Distribution Hypothesis has low truthiness (p-value)”
The coefficient of linear correlation $r$ can be directly translated into a p-value:

**Table 9.4.** The probability $Prob_N(|r| \geq r_0)$ that $N$ measurements of two uncorrelated variables $x$ and $y$ would produce a correlation coefficient with $|r| \geq r_0$. Values given are percentage probabilities, and blanks indicate values less than 0.05%.

<table>
<thead>
<tr>
<th>$N$</th>
<th>0</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>100</td>
<td>94</td>
<td>87</td>
<td>81</td>
<td>74</td>
<td>67</td>
<td>59</td>
<td>51</td>
<td>41</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>85</td>
<td>70</td>
<td>56</td>
<td>43</td>
<td>31</td>
<td>21</td>
<td>12</td>
<td>6</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>78</td>
<td>58</td>
<td>40</td>
<td>25</td>
<td>14</td>
<td>7</td>
<td>2</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>67</td>
<td>40</td>
<td>20</td>
<td>8</td>
<td>2</td>
<td>0.5</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>49</td>
<td>16</td>
<td>3</td>
<td>0.4</td>
<td>0.5</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
1. A professor observes that the 10 students in his class obtain the following grades in their homework and exam scores:

<table>
<thead>
<tr>
<th>Student</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework Average</td>
<td>90</td>
<td>60</td>
<td>45</td>
<td>100</td>
<td>15</td>
<td>23</td>
<td>52</td>
<td>30</td>
<td>71</td>
<td>88</td>
</tr>
<tr>
<td>Exam Score</td>
<td>90</td>
<td>71</td>
<td>65</td>
<td>100</td>
<td>45</td>
<td>60</td>
<td>75</td>
<td>85</td>
<td>100</td>
<td>80</td>
</tr>
</tbody>
</table>

Does the data support the professors claim that the exam scores and homework are positively correlated, and that therefore the students should do their homework? Give a quantitative argument.
Linear Correlation Example

Coefficient of linear correlation:

\[ r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \]

where

\[ \sigma_{xy} \equiv \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) \]

and

\[ \sigma_x = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2} \]

\[ \sigma_y = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \bar{y})^2} \]

- \( r \sim 0.8, 10 \) data points
- \( p\text{-value} = 0.005 \)
- The “random correlation” hypothesis has low truthiness, the students should do their homework!