Statistics and Visualization for Data Analysis: Resampling etc.

Mike Frank & Ed Vul

IAP 2009
Today’s goal

• Classically, statistics is full of equations.
• This is (partly) because computers have not been around for long

• Convey the principles behind frequentist statistics using only numerical methods (i.e., by using the brute force of computers)
now what?

statistics and data analysis

Courtesy of xkcd.org
Does that happen every time?

now what?

Courtesy of xkcd.org
What has happened?
What’s going to happen next?

• We don’t know.

• Let’s assume ‘more of the same’.

• ‘More of the same’:
  – Some process was producing events.
  – Events were
    • Independent
    • Identically distributed

• Assume “more independent, identically distributed events will follow”
Uncertainty

• We don’t know exactly what will happen if we touch the podium again.

• However, we have some data.

• The data allow us to make predictions.

• We can measure our uncertainty about what will happen with probability.
“Probability”? 

• Frequentist:  
  One specific event will happen next.  
  Another specific event will happen after that.  
  All we can say is that over many such events, the frequency of  
  a specific one occurring will match the frequency we  
  observed up to now.  
  — Probability is long-run frequency.  

• Bayesian:  
  I don’t know what will happen next, but I have some beliefs  
  about what it could be. These beliefs follow the laws of  
  probability. (My beliefs will reflect more than just the data.)  
  — Probability is degree of belief.
Frequentist or Bayesian?

• Most statistics you have been exposed to are ‘frequentist’.
  – Interpretations of e.g., ‘confidence intervals’ are rather weird.
  – Prior beliefs (such as theory, or good reason) don’t matter.

• We will be frequentist for most of today, but there are reasonable Bayesian interpretations of what we are doing.

• Let’s not worry about it for now.
Our class trajectory

Model-based

Frequentist

Bayesian

Data only
What will happen next?

• One of the prior events will repeat with a probability matching its previous frequency.

• So... we can just draw samples (with replacement!) from the previous data to predict future data.

• This is resampling
It’s not that simple

now

what?

Courtesy of xkcd.org
What do we want to know?

• The mean font size of a zap?
• Do zaps happen more often in this case than otherwise?
• How much bigger are average font sizes at the podium?
• If we got zapped at the podium or somewhere else, which zap would have a bigger font size?
The mean font size of a zap?

• Great. Wait. We’re not done.

• What we really want to be able to do is predict the average font size of zaps we haven’t yet seen.
Predicting the mean zap in unseen data.

Matlab code:

```matlab
O_zaps = [8 10 10 14 18 18 18 18 18 22 28 36];

hist(O_zaps, 8:2:36);
set(gca, 'FontSize', 16, 'FontWeight', 'bold');
```
Predicting the mean zap in unseen data.

• This is a good start...
• But we know future events will not be exactly the same as past events.
• So, the mean zap will not always be: ZAP

• What else might it be?
• ......
Introducing: The Bootstrap!

Courtesy of Rudolph Erich Raspe. Used with permission.
Bootstrapping: Make more samples, measures

• General reasoning:
  – We will see ‘more of the same’
  – We can produce more of the same to predict the future
  – Compute measure (mean) on more of the same
  – Tabulate the value of the measure.
Bootstrapping, more specifically

• We have a sample X containing n observations
• Generate possible future samples:
  – From X draw n times, producing $B_1$ (another possible sample)
  – Compute measure $f()$ on $B_1 = M_1$
  – Repeat # times.
Predicting the mean zap in unseen data.

```matlab
ntimes = 10000;
n = length(O_zaps);

f = @(x)(mean(x));

for i = [1:ntimes]
    B = randsample(O_zaps, n, true);
    M(i) = f(B);
end

hist(M, 80);
```

- (Don’t use this code – it is really inefficient, consider the Matlab function “bootstrap”)
Predicting the mean zap in unseen data.

• So this represents possible scenarios about what the mean of future data might be.

• Usually we want to say something a bit more concise, like:
  – The mean will be between A and B with confidence P.
Confidence intervals

• An interval [min to max] which will contain the measure with some level of confidence, P.
  – Confidence as probability
    • Probability as frequency of possible outcomes

• Sort all of our outcomes, consider the bounds of the middle P proportion:
Predicting the mean zap in unseen data.

\[ P = 0.95; \text{ confidence level} \]

\[ \text{omit}P = 1-P; \]
\[ \text{lower\_bound\_percentile} = \text{omit}P/2; \]
\[ \text{upper\_bound\_percentile} = 1-\text{omit}P/2; \]
\[ \text{lower\_bound\_index} = \text{round}((\text{lower\_bound\_percentile} \times ntimes)); \]
\[ \text{upper\_bound\_index} = \text{round}((\text{upper\_bound\_percentile} \times ntimes)); \]

\[ M\_\text{sorted} = \text{sort}(M); \]

\[ \text{lower\_bound} = M\_\text{sorted}\text{(lower\_bound\_index)}; \]
\[ \text{upper\_bound} = M\_\text{sorted}\text{(upper\_bound\_index)}; \]

\[ \text{CI} = [\text{lower\_bound} \ \text{upper\_bound}] \]

This can all be done with the “quantile” function.

With 95% Confidence:
mean zap between 13 and 22
Bootstrapping

• Mean here was a measure.
• You can use *any measure* you like, I won’t judge.

• It’s all good*

• * Some measures are more sensitive to the “Black Swan”
What do we want to know?

• The mean font size of a zap?
• Do zaps happen more often in this case than otherwise?
• How much bigger are average font sizes at the podium?
• If we got zapped at the podium or somewhere else, which zap would have a bigger font size?
Do zaps happen more often at the podium?

Podium

Otherwise
Podium zaps more often than otherwise?

<table>
<thead>
<tr>
<th></th>
<th>Zap</th>
<th>No Zap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Podium</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>Otherwise</td>
<td>8</td>
<td>14</td>
</tr>
</tbody>
</table>

• Well... yes... in this set of observations.
• But we might have observed this difference by chance even if they were the same...
Null Hypothesis Significance Testing

• $H_0$(null): The effect is 0
  – These groups have the same mean
  – ...same frequency of $X$
  – No correlation is present

• $H_1$: $H_0$ is not true.

• Basically: Are these observations so improbable under the null hypothesis that we must begrudgingly reject it?
Podium zaps more often than otherwise?

<table>
<thead>
<tr>
<th></th>
<th>Zap</th>
<th>No Zap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Podium</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>Otherwise</td>
<td>8</td>
<td>14</td>
</tr>
</tbody>
</table>

• Well... yes... in this set of observations.
• But we might have observed this difference by chance same...
• How often would a difference at least this big have occurred if these were truly the same? (probability of observing this effect under null hypothesis)
Introducing: Randomization (permutation)

• For most hypothesis tests, null hypothesis is: These things came from the same process.
• So... treat them as such.
• Resample many times from this new combined sample
• Measure the difference of interest in these samples
• See if the difference observed is particularly unlikely
Permutation (simple)

- We have two groups A and B.
- A has \( n \) observations, B has \( m \) observations.
- Assume they are ‘the same’ (IID), so permute assignments into A and B (while maintaining \( n \) and \( m \)).
- Calculate measure of interest on permutation.
- Rinse, repeat.
Podium zaps more often than otherwise?

```matlab
podium = [1 1 1 1 1 1 1 1 1 1 0 0 0 0 0];
other = [1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0];

f_comp = @(a,b)((sum(a==1)./length(a)) - (sum(b==1)./length(b)));

d_p = f_comp(podium, other);

allobs = [podium, other];
nperm = 10000;

for i = 1:nperm
    permall = allobs(randperm(length(allobs)));
    perm_podium = permall(1:length(podium));
    perm_other = permall(length(podium)+1:end);

    P(i) = f_comp(perm_podium, perm_other);
end

p = sum(P >= d_p)./length(P);
```

Probability that a difference at least this big would have been observed if these were really ‘the same’? 0.0139
Permutation

• Proportion was a measure here.
• You can use *any measure* you like, I won’t judge.

• It’s all good*.

• * Some measures are more sensitive to the “Black Swan”*
How unlikely is *too* unlikely?

... it is convenient to draw the line at about the level at which we can say: "Either there is something in the treatment, or a coincidence has occurred such as does not occur more than once in twenty trials."...

Fisher, 1926

Courtesy of The Barr Smith Library, University of Adelaide. Used with permission.
How unlikely is it that... it is convenient to draw the line at about the level at which we can say: "Either there is something in the treatment, or a coincidence has occurred such as does not occur more than once in twenty trials."…

Fisher, 1926

Table of \( \chi^2 \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>( P = .99 )</th>
<th>( .98 )</th>
<th>( .95 )</th>
<th>( .90 )</th>
<th>( .80 )</th>
<th>( .70 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.000157</td>
<td>.000628</td>
<td>.00393</td>
<td>.0158</td>
<td>.0642</td>
<td>.148</td>
</tr>
<tr>
<td>2</td>
<td>.0201</td>
<td>.0404</td>
<td>.103</td>
<td>.211</td>
<td>.446</td>
<td>.713</td>
</tr>
<tr>
<td>3</td>
<td>.115</td>
<td>.185</td>
<td>.352</td>
<td>.584</td>
<td>1.005</td>
<td>1.424</td>
</tr>
<tr>
<td>4</td>
<td>.297</td>
<td>.429</td>
<td>.711</td>
<td>1.064</td>
<td>1.649</td>
<td>2.195</td>
</tr>
<tr>
<td>5</td>
<td>.554</td>
<td>.752</td>
<td>1.145</td>
<td>1.610</td>
<td>2.343</td>
<td>3.000</td>
</tr>
<tr>
<td>6</td>
<td>.872</td>
<td>1.134</td>
<td>1.635</td>
<td>2.204</td>
<td>3.070</td>
<td>3.828</td>
</tr>
<tr>
<td>7</td>
<td>1.239</td>
<td>1.564</td>
<td>2.167</td>
<td>2.833</td>
<td>3.822</td>
<td>4.671</td>
</tr>
<tr>
<td>8</td>
<td>1.646</td>
<td>2.032</td>
<td>2.733</td>
<td>3.490</td>
<td>4.594</td>
<td>5.527</td>
</tr>
<tr>
<td>9</td>
<td>2.088</td>
<td>2.532</td>
<td>3.325</td>
<td>4.168</td>
<td>5.380</td>
<td>6.393</td>
</tr>
<tr>
<td>10</td>
<td>2.558</td>
<td>3.059</td>
<td>3.940</td>
<td>4.865</td>
<td>6.179</td>
<td>7.267</td>
</tr>
<tr>
<td>11</td>
<td>3.053</td>
<td>3.609</td>
<td>4.575</td>
<td>5.578</td>
<td>6.989</td>
<td>8.148</td>
</tr>
<tr>
<td>15</td>
<td>5.229</td>
<td>5.985</td>
<td>7.261</td>
<td>8.547</td>
<td>10.307</td>
<td>11.721</td>
</tr>
<tr>
<td>29</td>
<td>14.256</td>
<td>15.574</td>
<td>17.708</td>
<td>19.768</td>
<td>22.475</td>
<td>24.577</td>
</tr>
</tbody>
</table>

For larger values of \( n \), the expression \( \sqrt{2\chi^2} - \sqrt{2n-1} \) may be used as a normal deviate with unit standard error.
Podium zaps more often than otherwise?

Probability that a difference at least this big would have been observed if these were really ‘the same’? 0.0139

Yes.

“The difference is significant at p<0.05.”

“Significant at p=x”
This is a little bit weird.

(Talk about tails)
What do we want to know?

• The mean font size of a zap?
• Do zaps happen more often in this case than otherwise?
• How much bigger are average font sizes at the podium?
• If we got zapped at the podium or somewhere else, which zap would have a bigger font size?
How much bigger are font sizes at podium?

Podium

ZAP ZAPZAP ZAP ZAP

Otherwise

ZAP ZAP ZAPZAP ZAP ZAP ZAP
Font sizes observed

Podium

Other
Bootstrapping functions of two samples

- Same thing as bootstrapping one sample.
- Resample each sample
- Compute function of two samples
- Proceed.
Bootstrapping difference of two samples.

\[
P_{\text{zap}} = [8 \ 10 \ 10 \ 10 \ 14 \ 14 \ 18 \ 18 \ 18 \ 18 \ 18 \ 18 \ 22 \ 28 \ 36];
\]

\[
O_{\text{zap}} = [8 \ 8 \ 10 \ 10 \ 10 \ 14 \ 14 \ 18 \ 18 \ 18 \ 28];
\]

\[
f = @(a,b)(\text{mean}(a)-\text{mean}(b));
\]

\[
\text{nsamp} = 10000;
\]

\[
\text{for } i = [1:\text{nsamp}]
\begin{align*}
\text{BP} &= \text{randsample}(P_{\text{zap}}, \text{length}(P_{\text{zap}}), \text{true}); \\
\text{BO} &= \text{randsample}(O_{\text{zap}}, \text{length}(O_{\text{zap}}), \text{true}); \\
M(i) &= f(\text{BP}, \text{BO});
\end{align*}
\]

\[
\text{CI} = \text{quantile}(M, [0.025, 0.975]);
\]

Note:
Confidence interval contains zero
This is another way of testing null hypotheses.
(Arguably a much more useful way)
Bootstrapping two-sample measures

- Mean here was a measure.
- You can use *any measure* you like, I won’t judge.

- It’s all good*.

- * Some measures are more sensitive to the “Black Swan”
What do we want to know?

- The mean font size of a zap?
- Do zaps happen more often in this case than otherwise?
- How much bigger are average font sizes at the podium?
- If we got zapped at the podium or somewhere else, which zap would have a bigger font size?
- Are font sizes more variable at the podium?

Courtesy of xkcd.org
Which zap is more likely to be bigger?

- So far we have asked what we might expect of reasonably large samples. If our samples were bigger, we could probably ‘detect’ even smaller changes.
- We don’t care about being able to detect small differences. We often want to know, how much of a difference will it make. Period.
- This is a measure of **effect size**
Dominance (a simple measure of effect size)

- What is the probability that an observation of A will be bigger than an observation of B?
- Choose an A, a B
- Compare
- Repeat
Which zap is more likely to be bigger?

```matlab
f = @(a, b)(a-b);
nsamp = 10000;
for i = [1:nsamp]
    BP = randsample(P_zap, 1, true);
    BO = randsample(O_zap, 1, true);
    M(i) = f(BP, BO);
end

PdO = sum(M>0)./length(M)
OdP = sum(M<0)./length(M)
T = sum(M==0)./length(M)
d = PdO - OdP

Podium is bigger	Tie	Other is bigger
dominance
0.58	0.19	0.23	0.35
Podium wins.
What do we want to know?

• The mean font size of a zap?
• Do zaps happen more often in this case than otherwise?
• How much bigger are average font sizes at the podium?
• If we got zapped at the podium or somewhere else, which zap would have a bigger font size?

Courtesy of xkcd.org
What we need

• An assumption of IID observations
• And a computer

What we get

• Predictive distributions of any measure of our choosing:
  • Confidence intervals
  • Significance
  • Effect sizes
What more could we want?

• Ability to deal with “factors”
  – Generally complicated, can do simple cases.
    • Permute within factors
    • (Later) resample residuals (requires more assumptions)
      (won’t get into dealing with multiple factors)

• Work with *really* big datasets.
  – Wrong class, we are doing stuff numerically.
Does the location alter font-size? (one factor)
Analysis of within-factor variation

• (I made up this name – there may be an official name out there)

• Define some measure over all three groups, that answers the question: “Does this factor alter the observations?”

• Here is an example: standard deviation of the mean font-size across different ‘levels’ of the ‘factor’ (can choose something different, e.g., the range of squared font-sizes across levels)
Permute within factors!

- Null hypothesis: levels of this factor don’t matter.
- Permute observations across levels
- Build null-hypothesis distribution of this measure.
Does the location alter font-size?

\[
\begin{align*}
Z_{1,1} &= [8 \ 10 \ 10 \ 10 \ 14 \ 14 \ 18 \ 18 \ 18 \ 18 \ 22 \ 28 \ 36]; \\
Z_{1,2} &= [8 \ 8 \ 10 \ 10 \ 10 \ 14 \ 14 \ 18 \ 18 \ 18 \ 28]; \\
Z_{1,3} &= [8 \ 8 \ 8 \ 10 \ 10 \ 10 \ 10 \ 10 \ 10 \ 18 \ 18 \ 36]; \\
\end{align*}
\]

\begin{verbatim}
figure();
for i = [1:3]
    subplot(1,3,i);
    hist(Z{1,i}, [8:2:36]);
end
\end{verbatim}

Podium  Kitchen  Bathroom
Does the location alter font-size?

\[
\text{f\_meas} = @(a,b,c)\left(\text{std}([\text{mean}(a), \text{mean}(b), \text{mean}(c)])\right); \\
\text{Omeas} = \text{f\_meas}(Z\{1,1\}, Z\{1,2\}, Z\{1,3\}); \\
\text{nsamp} = 10000; \\
\text{alldata} = [Z\{1,1\}, Z\{1,2\}, Z\{1,3\}]; \\
n1 = \text{length}(Z\{1,1\}); \\
n2 = \text{length}(Z\{1,2\}); \\
n3 = \text{length}(Z\{1,3\}); \\
\text{for} \ i = [1:\text{nsamp}] \\
\quad P = \text{alldata}(\text{randperm}(n1+n2+n3)); \\
\quad P1 = P(1:n1); \\
\quad P2 = P((n1+1):(n1+n2)); \\
\quad P3 = P((n1+n2+1):\text{end}); \\
\quad M(i) = \text{f\_meas}(P1, P2, P3); \\
\text{End} \\
\text{p} = \text{sum}(M \geq \text{Omeas}) ./ \text{length}(M) \\
\]

\[P = 0.1654\]

No

Or “we can’t reject null hypothesis at p<0.05”
Permuting within factors

- St.Dev. of Mean across levels was our measure.
- You can use *any measure* you like, I won’t judge.

- It’s all good*.

- * Some measures are more sensitive to the “Black Swan”
Really Big Limitation

• “Black swan”
  – A general limitation of having incomplete data

• In case of extreme frequentism, even “dirty swans” go ignored.

• We can deal with this (to varying degrees) by specifying beliefs about our ignorance
What more could we want?

• Prettier histograms (more with less)
  – Getting a little Bayesian

• Respect dependencies in data
  – Generally complicated, can do simple cases.

• Make inferences about the world, rather than predicting the outcomes of more samples
“Yo’ histograms are ugly”

• “I don’t think the real future difference will have those spikes”

• Bayesian!

• New assumption:
  Future data will be “more of the same plus noise” (kernel density at each data point)
Additional assumptions of ignorance

• Protect against the “black swan” to some extent

• Increase uncertainty
  – Increase range of confidence intervals
  – Decrease the level of significance

• (Note: additional beliefs about underlying distributions [tomorrow] do not just increase uncertainty, and can have worrying effects)
Our class trajectory

- Model-based
- Bayesian
- Data only
- Frequentist
Smoothed bootstrap

• Bootstrap, just as before, but to each draw, add some noise, reflecting our new assumption that future data will be “more of the same plus noise”
Smoothed Bootstrap

\[
\text{for } i = [1:nsamp] \\
\quad \text{BP} = \text{randsample}(P_{\text{zap}}, \text{length}(P_{\text{zap}}), \text{true}) + \text{randn}(1, \text{length}(P_{\text{zap}})); \\
\quad \text{BO} = \text{randsample}(O_{\text{zap}}, \text{length}(O_{\text{zap}}), \text{true}) + \text{randn}(1, \text{length}(P_{\text{zap}})); \\
\quad \text{M}(i) = f(\text{BP}, \text{BO}); \\
\text{end}
\]
Hey, this is pretty neat

• I like this Bayesian business.
• What else do I believe about my data that will allow me to get more from less?
  – Smoothed bootstrap
  – Resampling residuals
  – Pivoted bootstrap
  – Scaled, pivoted, smoothed bootstrap of residuals...
  – I think there is a distribution in the world...
Residuals are IID; Maybe also Symmetry

- “More of the same deviations from the mean”
- “More of the same magnitude of deviations from the mean”
- Pivoted bootstrap
Pivoted bootstrap

- Compute some measure of central tendency
- Compute deviations from this measure of all observed data
- Bootstrap deviations, and randomly flip sign.
- Add central measure back in to obtain bootstrapped sample
- Compute the bootstrapped measure
Pivoted Bootstrap

\[ P_{\text{zap}} = [8 \ 10 \ 10 \ 10 \ 14 \ 14 \ 18 \ 18 \ 18 \ 18 \ 18 \ 22 \ 28 \ 36]; \]

\[ f = @(a,b)(\text{mean}(a)); \]
\[ \text{meanP} = f(P_{\text{zap}}); \]

\[ P_{\text{zap\_dev}} = P_{\text{zap}} - \text{meanP}; \]

\[ \text{for } i = [1:10000]; \]
\[ \quad B_{\text{dev}} = \text{randsample}(P_{\text{zap\_dev}}, \text{length}(P_{\text{zap}}), \text{true}); \]
\[ \quad \text{randSign} = \text{round}(\text{rand}(1,\text{length}(B_{\text{dev}}))) \times 2 - 1; \]
\[ \quad B_{\text{dev\_pivot}} = B_{\text{dev}} \times \text{randSign}; \]
\[ \quad B = \text{meanP} + B_{\text{dev\_pivot}}; \]
\[ \quad M(i) = f(B); \]
\[ \text{end} \]
Does number of Zs predict font size?

ZZZZZAP
ZZZZZZZZZAP ZZZZAP

ZAP ZzAP zZAP ZzzAP
Does number of Zs predict font size?

```
Ozap = [1 8;
      1 10
      1 8
      1 12
      2 8
      2 12
      2 16
      2 18
      3 12
      3 18
      3 26
      3 28
      4 24
      4 32
      4 20
      5 38
      5 32];
```
Does number of Zs predict font size?

• Measure on the sample of pairs?
• Slope of least-squares regression
  – Why? (Right now, no good reason, but we think it captures something about ‘predicting X from Y’)
  – We could have used some measure on rank orders, etc.
Does number of Zs predict font size?

- Null hypothesis:
  Two dimensions are independent.
- Procedure: resample from them independently to construct new paired sample
- Obtain measure on new sample
- Repeat, build null-hypothesis distribution, etc.
Does number of Zs predict font size?

- Confidence intervals are more useful.
- How do we bootstrap confidence intervals on measures of dependency?
- We often only have one observation at each level of a variable...

- Resample residuals!
Estimating dependencies in data

• Correlation, regression
• We have a set of paired observations.

• Least squares regression parameters
Does number of Zs predict font size?

```matlab
regression_params = regress(Ozap(:,2), [Ozap(:,1), ones(length(Ozap),1)]);
m = regression_params(1);
b = regression_params(2);
hold on;
plot([1:5], b+m.*[1:5], 'b-', 'LineWidth', 2)
```
Smoothed, pivoted bootstrap of residuals

\[
\text{res}_z = \text{Ozap}(;2) - (b+m.*\text{Ozap}(;1));
\]

for \(i = [1:10000]\)

\[
\text{nz} = \text{Ozap}(;1);
\]

\[
\text{B}\_\text{res} = \text{randsample}(\text{res}_z, \text{length(nz)}, \text{true});
\]

\[
\text{randSign} = \text{round}(\text{rand(}\text{length(B\_\text{res}),1)).*2-1;
\]

\[
\text{B}\_\text{res}\_\text{piv} = \text{B}\_\text{res}.*\text{randSign};
\]

\[
\text{B}\_\text{res}\_\text{piv}\_\text{smoothed} = \text{B}\_\text{res}\_\text{piv} + \text{randn(}\text{length(nz),1});
\]

\[
\text{B}\_\text{fs} = b + m.*\text{nz} + \text{B}\_\text{res}\_\text{piv}\_\text{smoothed};
\]

\[
\text{regression}\_\text{params} = \text{regress}(\text{B}\_\text{fs}, [\text{nz}, \text{ones(}\text{length(nz),1})]);
\]

\[
\text{Mm}(i) = \text{regression}\_\text{params}(1);
\]

\[
\text{Mb}(i) = \text{regression}\_\text{params}(2);
\]

end

Correlation coefficient (not shown)
Data

Measure something about it?

Measure on Data

But these data could have been different...

And I have these other data...

Predictive distribution on measure

Predictive distribution on between-group difference of measure.

Predictive distribution if these data were “the same”

Confidence intervals

Effect sizes

Null Hypothesis Significance Testing
What we have learned

• Resampling ("more of the same")
• Permutation ("condition assignment is random")
• Bootstrapping ("more of the same" + measure)
  Null Hypothesis Significance Testing
  Confidence Intervals
  – Smoothed ("more of the same + noise")
  – Residuals ("more of the same deviations")
  – Pivoted ("more of the same symmetric deviations")
• Dominance to measure effect size
• Watch out for the black swan!
Our class trajectory

- Model-based
- Frequentist
- Bayesian
- Data only
- "More of the same"
- "More of the same + noise"
- "More of the same + noise + pivoting residuals"
- "More of the same + noise + pivoting + scaling"
- "More of the same + noise + pivoting + scaling + residuals"
- More of the same + noise + pivoting + scaling + residuals
- More of the same + noise + pivoting + scaling + residuals + hierarchical model
- More of the same + noise + pivoting + scaling + residuals + hierarchical model + distribution
- More of the same + noise + pivoting + scaling + residuals + hierarchical model + distribution + generative process
- More of the same + noise + pivoting + scaling + residuals + hierarchical model + distribution + generative process
- More of the same + noise + pivoting + scaling + residuals + hierarchical model + distribution + generative process + data only