Hierarchy of Statistical Goals

Hierarchy of Statistical Goals

Ideal goal of scientific study: Deterministic results

Determine the exact value of a measurement or population parameter

- Prediction: What will the value of a future observation be?
- Comparing groups: What is the difference between response across two populations?

Problem: In the real world, few patterns are deterministic, so we do not observe the same outcome for all individuals

- Hidden (unmeasured) variables
- Inherent randomness

Hierarchy of Statistical Goals

Second choice: Probability model for response

Determine the tendency for the response

- Prediction: What is the probability that a future observation will be some value?
- Within groups: What is the average response within the group?
- Comparing groups: What is the difference in average response between groups

Hierarchy of Statistical Goals

Second choice: Probability model for response (cont.)

Consider the distribution of outcomes for individuals receiving intervention

- Use a probability model to describe distribution of response
- Usually choose a summary measure of the distribution
  - e.g.. mean, median, etc.
- Scientific questions then expressed for values of summary measure

Use of Probability Model

Often we have many choices for the summary measure to be compared across treatment groups

Example: Treatment of high blood pressure with a primary outcome of systolic blood pressure at end of treatment

Statistical analysis might for example compare

- Average
- Median
- Percent above 160 mm Hg
- Mean or median time until blood pressure below 140 mm Hg

Use of Probability Model

Summary measure for comparison should most often be driven by scientific issues

- Thresholds may be most important clinically
- Means allow estimates of total costs/benefits
- Medians less sensitive to outliers
  - Sometimes clinical importance is not proportional to magnitude of measurements
  - However, sometimes effect of intervention is only on outliers

Use of Probability Model

Sometimes choice of summary measure is more arbitrary

Types of scientific questions

- Existence of an effect on the distribution
- Direction of effect on the distribution
- Linear approximations to effect on summary measure
- Quantifying dose-response on summary measure

Only last two need dictate a choice of summary measure

Use of Probability Model

In any case, in choosing the summary measure used to define treatment effect, we should consider (in order of importance)

Current state of knowledge about treatment effect

Scientific (clinical) relevance of summary measure

Plausibility that treatment would affect the summary measure

Statistical precision of inference about the summary measure

(Semi)Parametric vs Nonparametric Models

In addition to the summary measure, we must also choose a model for the distribution of the data

- Parametric models assume a known shape for the distribution of the data
- Semiparametric models assume that the shape is similar in some way across groups, but do not otherwise make any assumptions about the exact shape of the distribution
- Nonparametric models make no assumption about how the shape of the distribution might be similar (or different) across groups

Use of Probability Model

As a general rule, it is rare that there is any advantage in assuming a parametric model in real life

- **IF** we do not even know whether an intervention affects the mean (or median, etc.) of a distribution (characteristics related to first moments),
- **THEN** why would we ever be willing to base our conclusions on assumptions about how the intervention might affect the shape of the distribution (characteristics that depend on 2nd, 3rd, …, ∞ moments)?

Use of Probability Model

Luckily, there is rarely a need to assume a parametric model

- E.g., methods derived from normal theory are usually distribution free tests in large samples
- It should also be noted that many semiparametric tests are quite sensitive to an unrealistic assumption
  - E.g., proportional hazards models for survival data over long periods of time
- Qualification: Distribution free Bayesian methods are not as well established (but we're working on it: Coarsened Bayes)

Hierarchy of Statistical Goals

Problem: The distribution (or summary measure) for the outcome is not directly observable

- Use a sample to estimate the distribution (or summary measure) of outcomes
- Such an estimate is thus subject to sampling error
- In presence of sampling error, we need an infinite sample size to discriminate between contiguous hypotheses
  - (see later discussion on statistical criteria for evidence)

Hierarchy of Statistical Goals

Third choice: Bayesian methods

Use the sample to estimate the probability that the hypotheses are true

- Probability of hypotheses given the observed data

Such a Bayesian approach is analogous to the problem of diagnosing disease in patients using a diagnostic procedure

Bayesian Inference

Statistical analysis is used to "diagnose" a beneficial treatment

Using a sample, we compute an estimate of treatment effect

- The estimate takes on the role of the diagnostic test result

Using the probability model, we can compute the probability of observing results under various hypotheses

- The hypothesis of a beneficial treatment might be like the "disease"

Bayesian Inference

The probability that the hypothesis is true is then like the predictive value of a positive test result

In order to use Bayes rule, we must have some measure of the "prevalence" of a beneficial treatment

Such a measure is termed the "prior distribution", because it is our estimate of the probability of a beneficial treatment prior to observing any data

The probability of the hypotheses based on the data is then called the posterior distribution

Bayesian Inference

The actual implementation of Bayesian inference is a generalization of the diagnostic testing situation

The estimate of treatment effect is continuous, rather than just positive or negative

The parameter measuring a beneficial treatment is continuous, rather than just healthy or diseased

The quantification of the prior distribution is thus an entire distribution (a probability for every possible value of the treatment effect) rather than a single prevalence.

Bayesian Inference

The criticism of Bayesian inference is that we usually do not know the prior probability of a beneficial treatment

As we have seen, the predictive values are very sensitive to the choice of prior distribution

Possible remedies:

- Use data from previous experiments
- Use subjective opinion or consensus of experts
- Do a sensitivity analysis over many different choices for the prior distribution
- Use frequentist approaches

Hierarchy of Statistical Goals

Fourth choice: Frequentist methods

Calculate the probability of observing data such was obtained in the experiment under the hypotheses

- Not affected by subjective choice of prior distributions
- But not really answering the most important question

Hierarchy of Statistical Goals

Fourth choice: Frequentist methods (cont.)

Frequentist methods consider the “sampling distribution” of statistics over (conceptual) replications of the same study

- If we were to repeat the study a large number of times (under the exact same conditions) what would be the distribution of the statistics computed from the samples obtained

Hierarchy of Statistical Goals

Fourth choice: Frequentist methods (cont.)

We do not usually have enough data to know what would happen if we repeated the study under the true setting, but we can often guess what would happen under specific hypotheses

Hence, frequentists characterize the sampling distribution under specific hypotheses and compare the observed data to what might reasonably have been obtained if that hypothesis were true

Hierarchy of Statistical Goals

Example: When playing poker, I get 4 full houses in a row

Bayesian:

- Knows the probability that I might be a cheater based on information derived prior to observing me play
- Knows the probability that I would get 4 full houses for every level of cheating that I might engage in
- Computes the posterior probability that I was cheating (probability after observing me play)
- If that probability is low, calls me a cheater

Hierarchy of Statistical Goals

Example: When playing poker, I get 4 full houses in a row (cont.)

Frequentist:

- Hypothetically assumes I am not a cheater
- Knows the probability that I would get 4 full houses if I were not a cheater
- If that probability is sufficiently low, calls me a cheater