- Estimation and confidence intervals are the first type of statistical inference.
  - Preferred by some journals and researchers
  - Not so dependent on a single value
  - Better idea of the possible values of the effect
- Hypothesis testing is the second type
  - Calculate a test statistic and then determine the probability that it is comparing data from the same distribution
- Research hypotheses, as opposed to statistical hypotheses, are the research questions that drive the research
  - e.g., lowering the fat in a person's diet lowers the blood cholesterol levels
  - e.g., better nutrition in childhood leads to increased adult height
  - e.g., pain control in incubated pre-term neonates leads to better behavioral development than no pain control

Estimation - Rationale

- **Estimator validity -** good, bad, low, high confidence?
- **Need** measurement of statistical properties (variance, bias, distribution, confidence intervals)
- **Bias** $E(\hat{\theta})-\theta$

  where $\hat{\theta}$ is the point estimate and θ the true parameter.

  Can be positive, negative or zero.

  Permits calculation of other properties, such as $MSE = E(\hat{\theta}-\theta)^2$

  where this quantity and variance of estimator only the same if estimator is unbiased. Obtained by both analytical and "bootstrap methods"

$$Bias = \sum_j \hat{\theta}_j f(x) - \theta$$

Similarly, for continuous variables

or for b bootstrap replications,

$$Bias = \frac{1}{b}\sum_{i=1}^{b} \hat{\theta}_i - \theta$$

Estimation Rationale- contd.

- For any, **even unbiased**, estimator $\hat{\theta}$, still a difference between estimator and true parameter = **sampling error**

  Hence the need for probability statements around $\hat{\theta}$

$$P\{T_1 < \hat{\theta} < T_2\} = \gamma$$

  with C.I. for estimator = (T_1, T_2), similarly to before and γ the **confidence coefficient**. If the estimator is unbiased in other words, γ is the probability that the true parameter falls into the interval.

- **In general**, confidence intervals can be determined using parametric and non-parametric approaches, where parametric construction needs a **pivotal quantity** = variable which is a function of parameter and data, but whose distribution does not depend on the parameter.

Significance Levels

- Level of significance, α, is the probability that the test statistic was declared significantly different from zero (typically) when the data are from the same population or distribution
  - Computed value of the test statistic that falls in the "rejection region" is said to be "statistically significant"or just "significant
    - May be different from "clinically significant"
  - Distribution of the test statistic is divided into rejection and acceptance regions
- We want to keep the α error low
  - Typically 0.05 or 0.01
  - In reality, we calculate the exact level of significance for a test statistic
- Guidelines:
  - $p > 0.10$ – not significant
  - $0.05 \le p \le 0.10$ – suggestive
  - $0.01 \le p < 0.05$ – significant
  - $0.001 \le p < 0.01$ – highly significant
  - $p < 0.001$ – very highly significant

Hypothesis testing

- Statistical hypotheses are specific research hypotheses that are stated in such a way that they may be evaluated by appropriate statistical techniques
  - H_0: there is no difference in blood cholesterol levels between those with reduced fat diets and those without
  - "Under the null hypothesis,…"
- Alternative hypothesis is the hypothesis of a significant difference
  - H_1: people on reduced fat diets will have lower blood cholesterol levels than those on regular diets

Steps to Hypothesis Testing

- Understanding research data
- Assumptions about data (i.e., distribution, independence, etc.)
- Development of hypothesis from knowledge base
- Determine test statistic to answer hypothesis
- Develop decision rule
- Generate test statistic
- Statistical decision
- Conclusion

Hypothesis testing- Rationale

- **Starting Point** of scientific research
  e.g. No Genetic Linkage between the genetic markers and genes
  when we design a linkage mapping experiment (Biological)
  $H_0 : \theta = 0.5$ (No Linkage) (2-locus linkage experiment)
  $H_1 : \theta \neq 0.5$ (two loci linked with specified R.F. = 0.2)
- **Critical Region**
  Given a cumulative probability distribution of a test statistic, *F(x)* say, the critical region for the hypothesis test is the region of rejection in the distribution, i.e. the area under the probability curve where the observed test statistics value **is unlikely to be observed** if H_0 true.
  $[1 - F(x)] \leq \alpha$ α **= significance level**

7

Interpreting Results of Hypothesis Testing

- We cannot "prove" hypotheses, only provide support for either the null or alternative
  - i.e., we "accept" or "fail to reject" the null hypothesis if the test statistic indicates that the two groups may be from the same population or we "reject" the null hypothesis if the test statistic indicates that they may be from different populations
- Hypothesis testing results are couched in probability terms since we can never be 100% sure

8

Relation between confidence intervals and significance tests

- If the 95% confidence interval does not contain μ_0, then the null hypothesis would be rejected at the 0.05 level.
- Conversely, if the 95% confidence interval does contain μ_0, then the null hypothesis is accepted at the 0.05 level

9

Power of a test

- Power used to plan a study or to give further insight into a non-significant result
- Important to design a study with a projected difference large enough to be "detected" with a statistical test
  - Otherwise, study is doomed to be a "negative" (non-significant) study
- Determination of difference not just a matter of guessing
  - Study design issues are critical
  - Selection of patients, appropriate control subjects or placebo medication, selection of high risk patients in whom a major difference would be substantial

10

Factors Affecting Power

- Smaller α-error leads to lower power for same sample size
- Bigger difference between means leads to more power for same sample size
- Bigger standard deviations leads to less power for same sample size
- Bigger sample size leads to more power

11

Types of Error and Power of the Test

- Type I error, α, is the probability of rejecting a true null hypothesis
  - Reflected in the level of significance
  - Typical values are 0.05 or 0.01
- Type II error, β, is the probability of accepting a false null hypothesis
  - Only an issue if fail to reject the null hypothesis
  - Typical values are 0.10 or 0.20
- Power of a test, $1-\beta$, is the probability of correctly rejecting a false null hypothesis
  - Typical values are 0.80 or 0.90

12

Types of Error and Power of the Test- contd.

- **Probability of False Positive and False Negative errors**

  e.g. ***false positive*** *if linkage between two genes declared, when really unlinked*

	Hypothesis Test Result	
Fact	**Accept H_0**	**Reject H_0**
H_0 True	*$1-\alpha$*	*False positive = Type I error =α*
H_0 False	*False negative =Type II error=β*	*Power of the Test = $1-\beta$*

- **Power of the Test** or **Statistical Power =** probability of rejecting H_0 when correct to do so. (Related strictly to alternative hypothesis and α)