

Biostatistics

Lecture 6

Hypothesis Testing: One-Sample Inference

Hypothesis testing

- 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

Hypothesis testing

- **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 intubated pre-term neonates leads to better behavioral development than no pain control

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**

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

Types of Error

- 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

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

Significance Levels

- 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
 - $0.05 \leq p \leq 0.10$ – suggestive
 - $0.01 \leq p < 0.05$ – significant
 - $0.001 \leq p < 0.01$ – highly significant
 - $p < 0.001$ – very highly significant

Testing a Single Mean – One Sided Alternatives

- Test to compare the mean of a normal distribution against a prespecified value, such as a population mean

- **Test statistic** is
$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

- for $H_0: \mu = \mu_0$ vs $H_1: \mu < \mu_0$ or $\mu > \mu_0$
 - with σ unknown
- Reject if $t < t(n-1, \alpha)$, accept otherwise
 - t is called a **test statistic**
 - $t(n-1, \alpha)$ is called a **critical value**
- Alternatively, use p-values directly

Testing a Single Mean – Two Sided Alternatives

- In some cases, may not be sure of which direction the difference may be going
- In this case, we are testing
 - $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$
 - Test statistic is the same
 - tested against high and low critical values
 - More conservative, since it is harder to reject the null hypothesis
 - As with confidence intervals, two sided tests have higher critical values
 - e.g., for $\alpha = 0.05$ (95% C.I.), the critical z-value is 1.645 one sided, but 1.96 two-sided

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

Power of a test

- Equation for determining power is given by:

$$\Phi\left(z_{\alpha} + \left|\mu_0 - \mu_1\right| \sqrt{n} / \sigma\right)$$

- where Φ is the probability of the z-value from the standard normal distribution

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

Power of a test – two-sided alternative

- Have to take into account both rejection regions

$$\Phi\left(-z_{1-\alpha/2} + \left|\mu_0 - \mu_1\right| \sqrt{n}/\sigma\right) +$$
$$\Phi\left(-z_{1-\alpha/2} + \left|\mu_1 - \mu_0\right| \sqrt{n}/\sigma\right)$$

Sample size determination

- Sample size (and power) depend on test statistic
 - Sample size (and power) can be derived from the test statistic
 - All are interconnected – different expressions of the same underlying relationship

- One-side alternatives for test of means

$$n = \frac{\sigma^2 (z_{1-\beta} + z_{1-\alpha})^2}{(\mu_0 - \mu_1)^2}$$

- Two-sided alternatives

$$n = \frac{\sigma^2 (z_{1-\beta} + z_{1-\alpha/2})^2}{(\mu_0 - \mu_1)^2}$$

Factors Affecting Sample Sizes

- As variance increases, sample size increases
- As significance level decreases, sample size increases
- As power increases, sample size increases
- As difference between means increase, sample size decreases

Sample sizes for confidence intervals

- Suppose we want to estimate the sample size required to construct a 100% x (1- α) C.I. for μ of width L

$$n = 4z_{1-\alpha/2}^2 s^2 / L^2$$

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

One Sample Chi-Square Test for the Variance of a Normal Distribution

- Test statistic is:
 - $\chi^2 = (n-1) s^2 / \sigma_0^2$
- Reject if
 - $\chi^2 < \chi^2_{n-1, \alpha/2}$ or $\chi^2 < \chi^2_{n-1, 1-\alpha/2}$

One-Sample Test for Binomial Proportion

- Extension of the approaches for normal mean
- Assume $np_0q_0 > 5$ so that we can use normal theory results
- Test statistic becomes

$$z = \frac{(\hat{p} - p_0)}{\sqrt{p_0q_0/n}}$$

- Power and sample size have similar extensions

One-Sample Inference for Poisson Distribution

- Usually assume $\mu_0 \geq 10$ to apply normal theory results
- Standard mortality ratio (SMR)
 - $SMR = 100\% \times O/E$
 - Where O is the observed deaths in the sample and E is the expected number based on the μ_0 in the general population