AID-521 Mathematics for Data Science

Module: Statistics | Lecture: 5

HYPOTHESIS TESTING & STATISTICAL DECISIONS

hypotheses, power, p-value, testing for a parameter, tests for count data*, **interval estimation**

Hypothesis Testing -- Introduction

→ Make decisions about the population on the basis of sample information (Statistical Decisions)

→ Begin with initial conjectures about the population (Statistical Hypotheses)

→ Compare conjecture with sample observations in a probabilistic manner (Tests of significance / Rules of decision)

Elements of a Statistical Test

- Null hypothesis H_0
  - Usually, the nullification of a claim.
- Alternative hypothesis H_1 or H_a
  - The claim itself.
- Test statistic *TS*
  - Function of the sample measurements used for the statistical decision – to reject H_0 or not.
  - Known distribution under H_0.
- Rejection region (or critical region) *RR*
  - Values of the observed *TS* for which H_0 will be rejected.
  - Such values are usually extreme values of *TS*, or in other words, highly unlikely values of *TS*

Usual Alternative Hypotheses

One may have hypotheses such as

$$H_0 : \mu = \mu_0$$

against one of the following alternatives:

$\rightarrow$ a two-tailed test/alternative

$$H_1 : \mu \neq \mu_0$$

$\rightarrow$ a one-tailed test

$$\begin{cases} H_1 : \mu < \mu_0, & \text{a lower (or left) tailed alternative} \\ H_1 : \mu > \mu_0, & \text{an upper (or right) tailed alternative} \end{cases}$$

The Test Statistic

→ A function of random sample (data), hence is a r.v.
  → Usually, an estimator for the unknown parameter

→ Its prob. distribution is known under null hypothesis H_0
  → Assume population $\sim \mathcal{N}(\mu, \sigma^2 = \sigma_0^2)$
  → Consider a simple hypothesis[†] $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$
  → Then $Z(\mu_0 | X_1, \ldots, X_n) = \sqrt{n}\frac{\bar{X} - \mu_0}{\sigma_0}$ is a *TS* with known distribution $\mathcal{N}(0, 1)$

→ Observed test statistic is its value when substituted with a given sample's values
  → $z(\mu_0 | x_1, \ldots, x_n) = \sqrt{n}\frac{\bar{x} - \mu_0}{\sigma_0}$

[†] A hypothesis that uniquely specifies the distribution from which the sample is taken is called a simple hypothesis.

Interpretation of Statistical Decision

If evidence (sampled data) strongly contradicts H_0 (beyond a reasonable doubt), then we reject H_0 in favor of H_1.

If H_0 is not rejected, then H_1 is automatically rejected.

Failure to reject H_0 does not necessarily mean that H_0 is true.†

† For e.g., “not guilty” does not mean a person “is innocent”. This basically means that there is not enough evidence to reject H_0.

Errors in Statistical Decision

Statistical decision	True state of null hypothesis	
	H_0 true	**H_0 false**
Do not reject H_0	Correct decision	Type II error (β)
Reject H_0	Type I error (α)	Correct decision

Level of significance = $\alpha = P(\text{reject } H_0 \mid H_0 \text{ is true})$

$\beta = P(\text{don't reject } H_0 \mid H_0 \text{ is false})$

$\rightarrow$ For fixed α, as n increases β decreases and vice versa.

Errors in Statistical Decision

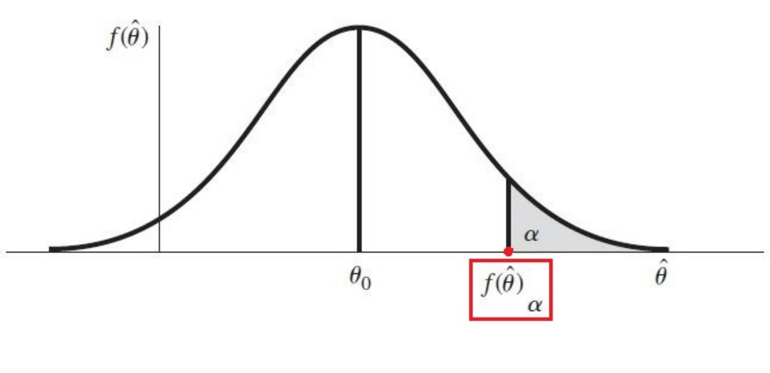
- Consequences of different types of errors are, in general, very different.
  - H_0: Person is innocent, vs. H_1: Person is guilty
  - H_0: Person is healthy, vs. H_1: Person is sick

- In many situations it is possible to determine which of the two errors is more serious.
  - Choose null hypothesis such that its rejection should be considered to be more serious.

Rejection Regions in Statistical Decision

- Given the probability distribution of a *TS* under H_0, the rejection region consists of those values of *TS* that are "extremely unlikely".
- The statistical analyst decides what values of *TS* are "extreme".
- The rejection region *RR* is pre-determined using the analyst's tolerance for error in decision.
  - Usually, the level of significance α is used to specify the level of error tolerance, and hence the *RR*
  - Each value of α corresponds to corresponding critical value(s) of *TS*

Visualizing α



Plotting $f(\hat{\theta})$ requires knowledge of an estimator $\hat{\theta}$ for parameter θ, and the sampling distribution of $\hat{\theta}$ under null hypothesis H_0.

Sample Size

...

Power

The power of a test is the probability that the test rejects H_0 when the alternative H_1 is true.

- If $H_0 : \theta = \theta_0$, and $H_1 : \theta = \theta_0$, then the power of the test at some $\theta = \theta_1 \neq \theta_0$ is

$$\pi(\theta_1) = \mathit{Power}(\theta_1) = P(\text{ reject } H_0 \mid \theta = \theta_1)$$

- A good test will have high power.

Likelihood Ratio Tests

...

p-Value

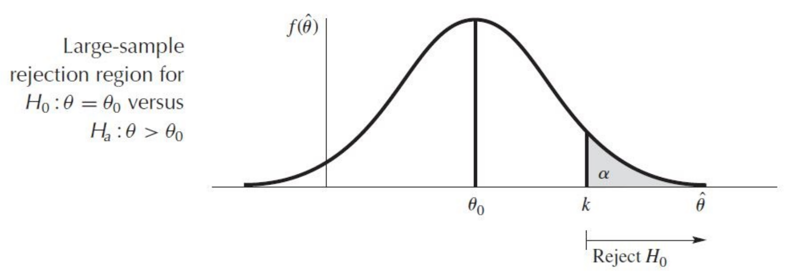
Corresponding to an observed value of a test statistic, the p-value (or attained significance level) is the lowest level of significance at which the null hypothesis would have been rejected.

- $\rightarrow$ The maximum value of α, willing to tolerate, is chosen.
- $\rightarrow$ If the *p*-value of the test is less than the maximum value of α, reject H_0.

The lower the p-value, the stronger the evidence.

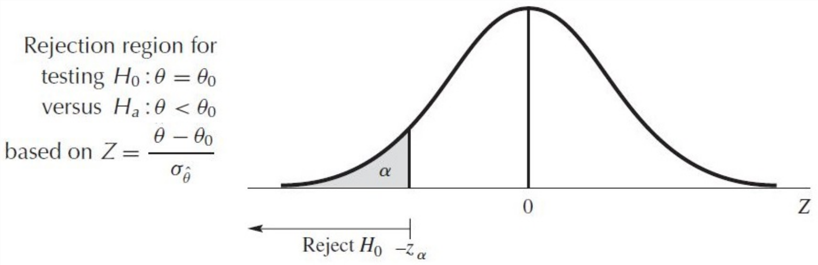
p-Value and Rejection Regions

- → Large sample
- → One-tailed test (right/upper)



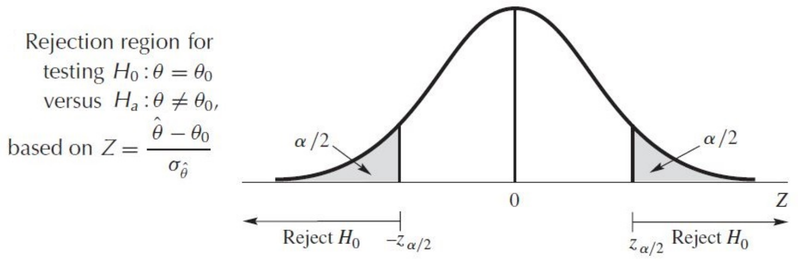
p-Value and Rejection Regions

- → Large sample
- → One-tailed test (left/lower)



p-Value and Rejection Regions

- → Large sample
- → Two-tailed test



Hypothesis Test for Parameter μ

SUMMARY OF HYPOTHESIS TESTS FOR μ

Large Sample ($n \geq 30$)

To test

$H_0 : \mu = \mu_0$

versus

$$H_a : \begin{array}{l} \mu > \mu_0, \text{ upper tail test} \\ \mu < \mu_0, \text{ lower tail test} \\ \mu \neq \mu_0, \text{ two-tailed test} \end{array}$$

Test statistic: $Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$

Replace σ by S, if σ is unknown.

$$\text{Rejection region} : \begin{cases} z > z_{\alpha}, & \text{upper tail RR} \\ z < -z_{\alpha}, & \text{lower tail RR} \\ |z| > z_{\alpha/2}, & \text{two tail RR} \end{cases}$$

Assumption: $n \geq 30$

Small Sample ($n < 30$)

To test

$H_0 : \mu = \mu_0$

versus

$$H_a : \begin{array}{l} \mu > \mu_0, \text{ upper tail test} \\ \mu < \mu_0, \text{ lower tail test} \\ \mu \neq \mu_0, \text{ two-tailed test} \end{array}$$

Test statistic: $T = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$

$$\text{RR} : \begin{cases} t > t_{\alpha,n-1}, & \text{upper tail RR} \\ t < -t_{\alpha,n-1}, & \text{lower tail RR} \\ |t| > t_{\alpha/2,n-1}, & \text{two tail RR} \end{cases}$$

Assumption: Random sample comes from a normal population

Decision: Reject H_0, if the observed test statistic falls in the RR and conclude that H_a is true with $(1 - \alpha)100\%$ confidence. Otherwise, keep H_0 so that there is not enough evidence to conclude that H_a is true for the given α and more experiments may be needed.

Hypothesis Test for Parameter σ^2

If $X_1, \ldots, X_n$ is a random sample from a normal population with the mean μ and variance σ^2, then

$$\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2} = \frac{(n-1)\,S^2}{\sigma^2}$$

has a chi-square distribution with $(n-1)$ degrees of freedom.

We know from Theorem 4.2.7 that $(1/\sigma^2)\sum_{i=1}^{n}(X_i - \mu)^2$ has a chi-square distribution with n degrees of freedom. Thus,

$$\begin{aligned}
\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X} + \overline{X} - \mu)^2 \\
&= \frac{1}{\sigma^2}\left[\sum_{i=1}^{n}(X_i - \overline{X})^2 + \sum_{i=1}^{n}(\overline{X} - \mu)^2\right] \\
&\quad \left(\text{Since } 2\sum_{i=1}^{n}(X_i - \overline{X})(\overline{X} - \mu) = 0\right) \\
&= \boxed{\frac{(n-1)\,S^2}{\sigma^2}} + \left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right)^2.
\end{aligned}$$

The left-hand side of this equation has a chi-square distribution with n degrees of freedom. Also, since $(\overline{X} - \mu)/(\sigma/\sqrt{n}) \sim N(0, 1)$ by Theorem 4.2.6 we have $\left[(\overline{X} - \mu)/(\sigma/\sqrt{n})\right]^2 \sim \chi^2(1)$. Now from Theorem 4.2.4, $(n-1)\,S^2/\sigma^2 \sim \chi^2(n-1)$.

Hypothesis Test for Parameter σ^2

SUMMARY OF HYPOTHESIS TEST FOR THE VARIANCE σ^2

To test

$$H_0 : \sigma^2 = \sigma_0^2$$

versus

$$H_a : \begin{cases} \sigma^2 > \sigma_0^2, & \text{upper tail test} \\ \sigma^2 < \sigma_0^2, & \text{lower tail test} \\ \sigma^2 \neq \sigma_0^2, & \text{two-tailed test.} \end{cases}$$

Test statistic:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

where S^2 is the sample variance.

Observed value of test statistic:

$$\frac{(n-1)s^2}{\sigma_0^2}$$

$$\text{Rejection region} : \begin{cases} \chi^2 > \chi^2_{\alpha,n-1}, & \text{upper tail RR} \\ \chi^2 < \chi^2_{1-\alpha,n-1}, & \text{lower tail RR} \\ \chi^2 > \chi^2_{\alpha/2,n-1} \text{ or } \chi^2 < \chi^2_{1-\alpha/2,n-1}, & \text{two tail RR} \end{cases}$$

where $\chi^2_{\alpha,n-1}$ is such that the area under the chi-square distribution with $(n-1)$ degrees of freedom to its right is equal to α.

Assumption: Sample comes from a normal population.

Decision: Reject H_0, if the observed test statistic falls in the RR and conclude that H_a is true with $(1-\alpha)100\%$ confidence. Otherwise, do not reject H_0 because there is not enough evidence to conclude that H_a is true for given α and more data are needed.

Hypothesis Test for Count Data

(Optional) Refer Ramachandran & Tsokos Section 7.6

Interval Estimation

- We know how to obtain point estimate(s) of a (vector of) population parameter(s).
- Here, we will discuss interval estimation.
  - Intuition closely related to acceptance region of a statistical test
- Given a sample, we will find L and U such that $L < \theta < U$ with some probability.
  - The higher the probability $P(L < \theta < U)$, the better.
  - The lower the length of interval (L, U), the better.
- Interval estimators are called confidence intervals.

Confidence Intervals

- Probability that a confidence interval (CI) will contain the true parameter θ is called the confidence coefficient.
  - Fraction of the time that the constructed interval will contain the true parameter, under repeated sampling
- For a confidence coefficient of $1-\alpha$, the interval $(L,\ U)$ such that

$$P(L < \theta < U) = 1 - \alpha$$

is called $(1-\alpha)100\%$ confidence interval.
  - For $\alpha = 0.05$, we will obtain a 95% confidence interval.

Large Sample CI for μ

For large sample, using CLT, we know that the z-transform of $\hat{\mu}$, the estimator of population mean μ,

$$Z = \frac{\hat{\mu} - \mu}{\sigma_{\hat{\mu}}}$$

has an approximately standard normal distribution, where $\sigma_{\hat{\mu}}$ is the standard deviation of $\hat{\mu}$.

Therefore

$$\hat{\mu} - z_{\frac{\alpha}{2}}\sigma_{\hat{\mu}} < \mu < \hat{\mu} + z_{\frac{\alpha}{2}}\sigma_{\hat{\mu}}$$

is a $(1-\alpha)100\%$ confidence interval for the population mean μ.

Small Sample CI for μ

- Let $X_1, \ldots, X_n$ be a **small sample** from a **normal population**. σ^2 is unknown.
- We know that

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim T_{(n-1)}$$

- Therefore

$$\bar{X} - t_{\frac{\alpha}{2},\, n-1} \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\frac{\alpha}{2},\, n-1} \frac{S}{\sqrt{n}}$$

is a $(1-\alpha)100\%$ confidence interval for the population mean μ.

CI for σ^2

If $\overline{X}$ *and* S *are the mean and standard deviation of a random sample of size* n *from a normal population, then*

$$P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right) = 1-\alpha$$

where the χ^2*-distribution has* $(n-1)$ *degrees of freedom.*

That is, we are $(1-\alpha)100\%$ *confident that the population variance* σ^2 *falls in the interval*

$((n-1)S^2/\chi^2_{\alpha/2},\ (n-1)S^2/\chi^2_{1-\alpha/2})$.

CI for σ^2