VE401 Probabilistic Methods in Eng.
Final Review Part 1

CHEN Xiwen

UM-SJTU Joint Institute

May 9, 2020

Table of contents

Fisher's Null Hypothesis Test

Overview.

1. Set up a ***null hypothesis*** H_0 that compares a population parameter θ to a given null value θ_0.
   - $H_0 : \theta = \theta_0$,
   - $H_0 : \theta \leq \theta_0$,
   - $H_0 : \theta \geq \theta_0$.
2. Try to reject the null hypothesis by finding ***P-value*** for the test.
   - One-tailed: upper bound of probability of obtaining the data or more extreme data (based on the null hypothesis), given that the null hypothesis is true.

     $$P[D|H_0] \leq P-\text{value}.$$

   - Two-tailed: twice of p-value for one-tailed test.
3. We either
   - fail to reject H_0 or
   - reject H_0 at the [p-value] level of significance.

Understanding P-values

One-tailed tests. For one-tailed test, the p-value is an upper bound of probability of obtaining the data or more extreme data based on the null hypothesis, given that the null hypothesis is true.

- $H_0 : \mu \geq \mu_0$, extreme $\Leftrightarrow$ too small $\widehat{\mu}$.
- $H_0 : \mu \leq \mu_0$, extreme $\Leftrightarrow$ too large $\widehat{\mu}$.

Note: Refer to an estimator of the parameter in the null hypothesis (μ in this case).

Understanding P-values

One-tailed tests. For one-tailed test, the p-value is an upper bound of probability of obtaining the data or more extreme data based on the null hypothesis, given that the null hypothesis is true.

- $H_0 : \mu \geq \mu_0$, extreme $\Leftrightarrow$ too small $\widehat{\mu}$.
- $H_0 : \mu \leq \mu_0$, extreme $\Leftrightarrow$ too large $\widehat{\mu}$.

Note: Refer to an estimator of the parameter in the null hypothesis (μ in this case).

Assignment 5.2. $X_1, \ldots, X_n$ are i.i.d. exponential random variables with parameter β, then the sum Y

$$f_Y(y) = \frac{\beta^n}{\Gamma(n)} y^{n-1} e^{-\beta y} \quad \Rightarrow \quad 2n\beta\overline{X} \sim \chi^2_{2n}.$$

The maximum-likelihood estimator for β is $\widehat{\beta} = 1/\overline{X}$. Then we reject $H_0 : \beta \leq \beta_0$ with test statistic $2n\beta_0\overline{X}$ if $P[2n\beta_0\overline{X} < 2n\beta_0\overline{x} | \beta = \beta_0]$ is too small.

Understanding P-values

Two-tailed tests. For two-tailed test, the p-value is twice of p-value for one-tailed test. Namely, suppose for one-tailed hypotheses $H_{0,u} : \mu \geq \mu_0$ and $H_{0,l} : \mu \leq \mu_0$, the corresponding p-values are p_u and p_l, then for the two-tailed H_0,

$$\text{pvalue} = 2 \min \left(p_u, p_l \right).$$

The intention of doubling is to consider “more extreme data” in both directions, while taking the minimum corresponds to taking the one-tailed hypothesis such that the data seems more “extreme”.

In some cases (where the distribution of test statistic is not symmetric), the p-value for a two-tailed test is not understood in terms of “probability of more extreme data”, but in terms of “double of p-value for one-tailed test”.

P-values for Representative Parametric Tests

T-test for mean. With sample $X_1, \ldots, X_n$ from normal populations with mean μ and unknown variance σ^2, we have test statistic

$$T_{n-1} = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

and p-values: (F is the c.d.f. of T_{n-1}.)

- $\underline{H_0 : \mu \leq \mu_0}$.
$$\text{pvalue} = 1 - F(t_{n-1});$$
- $\underline{H_0 : \mu \geq \mu_0}$.
$$\text{pvalue} = F(t_{n-1});$$
- $\underline{H_0 : \mu = \mu_0}$.
$$\text{pvalue} = 2 \min \left(F(t_{n-1}), 1 - F_{T_{n-1}}(t_{n-1}) \right).$$

P-values for Representative Parametric Tests

F-test for comparing variances. Let S_1^2 and S_2^2 be sample variances from independent samples of sizes n_1, n_2 from normal populations with means μ_1, μ_2 and variances σ_1^2, σ_2^2, we have test statistic

$$F_{n_1-1,n_2-1} = \frac{S_1^2}{S_2^2}.$$

and p-values: (F is the c.d.f. of F_{n_1-1,n_2-1}.)

- $\underline{H_0 : \sigma_1 \leq \sigma_2}$.

$$\text{pvalue} = 1 - F(f_{n_1-1,n_2-1});$$

- $\underline{H_0 : \sigma_1 \geq \sigma_2}$.

$$\text{pvalue} = F(f_{n_1-1,n_2-1});$$

- $\underline{H_0 : \sigma_1 = \sigma_2}$.

$$\text{pvalue} = 2\min\left(F(f_{n_1-1,n_2-1}), 1 - F(f_{n_1-1,n_2-1})\right).$$

P-values for Representative Non-parametric Tests

Signed test for median. Let $X_1, \ldots, X_n$ be a random sample, we have test statistic

$$Q_- = \#\{X_k : X_k - M_0 < 0\}.$$

and p-values:

- $\underline{H_0 : M \leq M_0}$. pvalue $= F[q_-]$;
- $\underline{H_0 : M \geq M_0}$. pvalue $= 1 - F[q_-]$;
- $\underline{H_0 : M = M_0}$. pvalue $= 2\min\left(F(q_-), 1 - F(q_-)\right)$,

where with a binomial random variable Y and $n' = q_+ + q_-$,

$$F(k) = P[Y \leq k | M = M_0] = \sum_{y=0}^{k} \binom{n'}{y} \frac{1}{2^{n'}}.$$

Note. In lecture slides, we use $\min(q_-, q_+)$ for two-tailed test, which is equivalent since

$$\text{pvalue} = \min\left(F(q_-), F(q_+)\right) = F\left(\min(q_-, q_+)\right).$$

P-values for Representative Non-parametric Tests

Wilcoxon signed rank test. Let $X_1, \ldots, X_n$ be a random sample of size n from a symmetric distribution. We have test statistic

$$|W_-| = \sum_{R_i<0} |R_i|, \quad \mathrm{E}[|W_-|] = \frac{n(n+1)}{4}, \quad \mathrm{Var}|W_-| = \frac{n(n+1)(2n+1)}{24},$$

and for large sample size, we have p-values (F is the c.d.f for normal distribution with mean and variance above.):

- $\underline{H_0 : M \leq M_0}$.
$$\text{pvalue} = F(|w_-|);$$
- $\underline{H_0 : M \geq M_0}$.
$$\text{pvalue} = 1 - F(|w_-|);$$
- $\underline{H_0 : M = M_0}$.
$$\text{pvalue} = 2\min\left(F(|w_-|), 1 - F(|w_-|)\right).$$

Hypothesis Tests

- Fisher's Null Hypothesis Test
- Neyman-Pearson Decision Theory
- Null Hypothesis Significance Testing

Neyman-Pearson Decision Theory

Overview.

1. Set up a ***null hypothesis*** H_0 and an ***alternative hypothesis*** H_1.
2. Determine a desirable α and β, where
   - $\alpha := P[\text{reject } H_0 | H_0 \text{ true}]$,
   - $\beta := P[\text{accept } H_0 | H_1 \text{ true}]$, and
   - power $:= 1 - \beta = P[\text{reject } H_0 | H_1 \text{ true}]$.
3. Use α and β to determine the appropriate sample size n.
4. Use α and n to determine the ***critical region***.
5. Obtain sample statistics, and reject H_0 at significance level α and accept H_1 if the test statistic falls into critical region. Otherwise, accept H_0.

Type-I Error: α

Type-I error. It is

- only related to H_0:

$$P[\text{reject } H_0 | H_0 \text{ true}];$$

- used to determine the critical region: this means that the critical region (whether H_0 is rejected or not, whether H_1 is accepted or not), is only determined by H_0.

△ *If H_0 is true, the probability of rejecting H_0 is no more than α.*

Type-II Error: β

Type-II error. It is

- related to both H_0 and H_1:

$$P[\text{fail to reject } H_0 | H_1 \text{ true}];$$

- used to analyze the power:

$$\text{power} = 1 - \beta;$$

- either calculated analytically or read from OC curves.

△ *If H_1 is true, the probability of accepting H_0 is no more than β.*

Type-II Error: β

F-test for comparing variances. Let S_1^2 and S_2^2 be sample variances based on independent random samples of sizes n_1 and n_2 drawn from normal populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively. The test statistic is given by

$$F_{n_1-1,n_2-1} = \frac{S_1^2}{S_2^2}.$$

We reject at significance level α

- $H_0 : \sigma_1 \leq \sigma_2$ if $S_1^2/S_2^2 > f_{\alpha,n_1-1,n_2-1}$,
- $H_0 : \sigma_1 \geq \sigma_2$ if $S_2^2/S_1^2 > f_{\alpha,n_2-1,n_1-1}$,
- $H_0 : \sigma_1 = \sigma_2$ if $S_1^2/S_2^2 > f_{\alpha/2,n_1-1,n_2-1}$ or $S_2^2/S_1^2 > f_{\alpha/2,n_2-1,n_1-1}$.

OC curve. The abscissa is defined by

$$\lambda = \frac{\sigma_1}{\sigma_2}.$$

Type-II Error: β

F-test for comparing variances. Consider the following cases (suppose $\delta > 1$).

- <u>One-tailed test</u>.

$$H_0 : \sigma_1 \leq \sigma_2, \qquad H_1 : \frac{\sigma_1}{\sigma_2} \geq \delta.$$

- <u>One-tailed test</u>.

$$H_0 : \sigma_1 \geq \sigma_2, \qquad H_1 : \frac{\sigma_1}{\sigma_2} \leq \delta.$$

- <u>Two-tailed test</u>.

$$H_0 : \sigma_1 = \sigma_2, \qquad H_1 : \max\left(\frac{\sigma_1}{\sigma_2}, \frac{\sigma_2}{\sigma_1}\right) \geq \delta.$$

Recall that $\sigma_2^2 S_1^2/(\sigma_1^2 S_2^2) \sim F_{n_1-1,n_2-1}$. We only discuss the last case here.

Type-II Error: β

F-test for comparing variances. Based on the distribution, we calculate

$$P[\text{accept } H_0|H_1 \text{ true}] = P\left[\frac{\sigma_2^2}{\sigma_1^2}f_1 \le F_{n_1-1,n_2-1} \le \frac{\sigma_2^2}{\sigma_1^2}f_2 \middle| \max\left(\frac{\sigma_1}{\sigma_2}, \frac{\sigma_2}{\sigma_1}\right) \ge \delta\right]$$
$$= P\left[\frac{\sigma_1}{\sigma_2} \ge \delta\right] \times P_1 + P\left[\frac{\sigma_1}{\sigma_2} \le \frac{1}{\delta}\right] \times P_2 \le \max\left(P_1, P_2\right),$$

where with $f_1 = f_{1-\alpha/2,n_1-1,n_2-1}, f_2 = f_{\alpha/2,n_1-1,n_2-1}$,

$$P_1 \le \left[\frac{\sigma_2^2}{\sigma_1^2}f_1 \le F_{n_1-1,n_2-1} \le \frac{\sigma_2^2}{\sigma_1^2}f_2 \middle| \frac{\sigma_1}{\sigma_2} = \delta\right] = F\left(\frac{1}{\delta^2}f_2\right) - F\left(\frac{1}{\delta^2}f_1\right),$$
$$P_2 \le \left[\frac{\sigma_2^2}{\sigma_1^2}f_1 \le F_{n_1-1,n_2-1} \le \frac{\sigma_2^2}{\sigma_1^2}f_2 \middle| \frac{\sigma_1}{\sigma_2} = \frac{1}{\delta}\right] = F\left(\delta^2 f_2\right) - F\left(\delta^2 f_1\right),$$

where F is the c.d.f. for F_{n_1-1,n_2-1}. Then we have the upper bound $\beta = \max(P_1, P_2)$.

Determining Sample Size using β

Normal case. Suppose the sample mean $\overline{X}$ follows a normal distribution with unknown mean μ and known variance σ^2, and we have hypothesis

$$H_0 : \mu = \mu_0, \qquad H_1 : |\mu - \mu_0| \geq \delta_0.$$

Relation between α, β δ, σ and n. With true mean $\mu = \mu_0 + \delta$, the test statistic $Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathrm{N}(\delta\sqrt{n}/\sigma, 1)$.

$$\begin{aligned} P[\text{fail to reject } H_0 | \mu = \mu_0 + \delta] &= \frac{1}{\sqrt{2\pi}} \int_{-z_{\alpha/2}}^{z_{\alpha/2}} e^{-(t-\delta\sqrt{n}/\sigma)^2/2} \mathrm{d}t \\ &= \frac{1}{\sqrt{2\pi}} \int_{-z_{\alpha/2}-\delta\sqrt{n}/\sigma}^{z_{\alpha/2}-\delta\sqrt{n}/\sigma} e^{-t^2/2} \mathrm{d}t \\ &\approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_{\alpha/2}-\delta\sqrt{n}/\sigma} e^{-t^2/2} \mathrm{d}t \overset{!}{=} \beta, \end{aligned}$$

where we set $-z_\beta = z_{\alpha/2} - \delta\sqrt{n}/\sigma$.

Determining Sample Size using β

Normal case. Suppose the sample mean $\overline{X}$ follows a normal distribution with unknown mean μ and known variance σ^2, and we have hypothesis

$$H_0 : \mu = \mu_0, \qquad H_1 : |\mu - \mu_0| \geq \delta_0.$$

Choosing the sample size n.

$$n \approx \frac{(z_{\alpha/2} + z_\beta)^2 \sigma^2}{\delta^2},$$

where $z_{\alpha/2}$ and z_β satisfies that

$$\Phi(z_{\alpha/2}) = 1 - \alpha/2, \qquad \Phi(z_\beta) = 1 - \beta,$$

given cumulative distribution function Φ of standard normal distribution.

Determining Sample Size using β

OC curves. Still taking the F-test as an example, when $n_1 = n_2 = n$, the abscissa is given by

$$\lambda = \frac{\sigma_1}{\sigma_2}.$$

If we have the hypotheses

$$H_0 : \sigma_1 = \sigma_2, \qquad H_1 : \max\left(\frac{\sigma_1}{\sigma_2}, \frac{\sigma_2}{\sigma_1}\right) \geq \delta,$$

Shall we use $\lambda = \delta$ or $\lambda = 1/\delta$ or both? As we have calculated,

$$P[\text{accept } H_0 | H_1 \text{ true}] \leq \max\left\{F\left(\frac{1}{\delta^2} f_2\right) - F\left(\frac{1}{\delta^2} f_1\right), F\left(\delta^2 f_2\right) - F\left(\delta^2 f_1\right)\right\},$$

However, with equal sizes, $f_2 = 1/f_1$, and thus

$$\begin{aligned} F\left(\delta^2 f_2\right) - F\left(\delta^2 f_1\right) &= F\left(\frac{\delta^2}{f_1}\right) - F\left(\frac{\delta^2}{f_2}\right) = 1 - F\left(\frac{1}{\delta^2} f_1\right) - \left[1 - F\left(\frac{1}{\delta^2} f_2\right)\right] \\ &= F\left(\frac{1}{\delta^2} f_2\right) - F\left(\frac{1}{\delta^2} f_1\right). \end{aligned}$$

The two cases $\sigma_1/\sigma_2 = \delta$ or $\sigma_2/\sigma_1 = \delta$ are equivalent.

Hypothesis Tests

- Fisher's Null Hypothesis Test
- Neyman-Pearson Decision Theory
- Null Hypothesis Significance Testing

Null Hypothesis Significance Testing

- H_0 and H_1 are set up, but H_1 is always the logical negation of H_0.
- Either a hypothesis test is performed by finding a critical region, or the test statistic is evaluated and a p-value is found to reject or accept H_1.
- In either case, there is no meaningful discussion of β, since H_1 is exactly the negation of H_0.

Setting up Hypotheses

- Setting up H_0. In most of the cases, when we want to find evidence ***for*** an event, H_0 is set up to be the opposite of this event.
  Assignment 6.2.(i). Is there evidence that the success rate is greater for longer tears (μ_1)?

$$H_0 : \mu_1 \leq \mu_2$$

- Alternative hypothesis H_1. In addition to testing equality and inequality, we would like to test an additional ***amount*** of such difference.
  Assignment 6.3. Is there sufficient evidence to conclude that the two population standard deviations differ by at least 10%?

$$H_0 : \sigma_1 = \sigma_2, \qquad H_1 : \max\left(\frac{\sigma_1}{\sigma_2}, \frac{\sigma_2}{\sigma_1}\right) \geq 1.1$$

Critical Region vs. Confidence Interval

Suppose $X_1, \ldots, X_n$ is a sample of size n from a normal population X with mean μ and known variance σ^2. We have

$$H_0 : \mu \leq \mu_0, \qquad H_1 : \mu > \mu_0.$$

Then what is the corresponding confidence interval? We know that we reject H_0 at significance level α if

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha \quad \Leftrightarrow \quad \mu_0 < \overline{X} - z_\alpha \frac{\sigma}{\sqrt{n}}.$$

In this case the null value μ_0 falls ***outside*** of the confidence interval. Therefore, the corresponding confidence interval is given by

$$\mathrm{CI} = \left[\overline{X} - z_\alpha \frac{\sigma}{\sqrt{n}}, \infty\right).$$

Similarly, if the null hypothesis is $H_0 : \mu \geq \mu_0$, the corresponding one-sided confidence interval is given by

$$\mathrm{CI} = \left(-\infty, \overline{X} + z_\alpha \frac{\sigma}{\sqrt{n}}\right].$$

Thanks for your attention!

Good luck for Final exam!