Topic V:
Hypothesis Testing (appendix)

Wei You

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Fall, 2023

The Normal Model

Assume $\boldsymbol{X} = (X_1, \ldots, X_n)$ is a random sample from $\mathcal{N}(\mu, \sigma)$.

Common Statistics

Based on the sample mean and sample variance $\bar{X}$ and S^2 define

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}, \quad T = \frac{\bar{X} - \mu}{S/\sqrt{n}}, \quad V = \frac{n-1}{\sigma^2} S^2.$$

We know Z has the standard normal distribution, T has the student-t distribution with DoF $n-1$, and V has the χ^2 distribution with DoF $n-1$. Moreover, Z and V are independent.

They are natural test statistics when the parameters are replaced by the values in the null hypothesis.

Quantiles

Let $p \in (0, 1)$ and $k \in \mathbb{N}$

Quantiles

- $z(p)$ denotes the quantile of order p for the standard normal distribution.
- $t_k(p)$ denotes the quantile of order p for the student-t distribution with DoF k.
- $\chi^2_k(p)$ denotes the quantile of order p for the chi-square distribution with DoF k.

Test of the Mean with Known Variance

One-sample z-test

- When μ is unknown and σ is known. For a conjectured μ_0, the statistic $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ has normal distribution $\mathcal{N}(\frac{\mu-\mu_0}{\sigma/\sqrt{n}}, 1)$. The significance level $\alpha \in (0,1)$ is given.
- $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$.
    - Reject H_0 if Z is too big or too small.
    - Define $R = \{\mathbf{x} : Z \leq z(\alpha/2) \text{ or } Z \geq z(1-\alpha/2)\}$. Under H_0, $\beta(\mu_0) = \mathbb{P}(\mathbf{X} \in R | \mu = \mu_0) = \alpha$. So the significance level (maximum type I error) is α.
- $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$.
    - Reject H_0 if Z is too big.
    - Define $R = \{\mathbf{x} : Z \geq z(1-\alpha)\}$. The significance level is $\max_{\mu \leq \mu_0} \mathbb{P}(Z \geq z(1-\alpha)) = \alpha$.

Cont.

- p-value: the two-sided test has p-value $2(1-\Phi(|Z|))$. It is the area under the standard normal PDF outside Z and $-Z$. The left(right)-tailed test has p-value $1-\Phi(Z)$ $(\Phi(Z))$.
- The power function of the two-sided test is $\beta(\mu)=\mathbb{P}\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\in(-\infty,z(\alpha/2)+\frac{\mu_0-\mu}{\sigma/\sqrt{n}})\cup(z(1-\alpha/2)+\frac{\mu_0-\mu}{\sigma/\sqrt{n}},\infty)\right)$.
  - $\beta(\cdot)$ is decreasing on $(-\infty,\mu_0)$ and increasing on (μ_0,∞). $\beta(\mu_0)=\alpha$.
  - $\beta(-\infty)=\beta(\infty)=1$.
  - Increasing n or decreasing σ makes the test uniformly more powerful.
- For the left-tailed test, $\beta(\mu)=1-\Phi\left(z(1-\alpha)+\frac{\mu_0-\mu}{\sigma/\sqrt{n}}\right)$.
  - $\beta(\cdot)$ is increasing; $\beta(\mu_0)=\alpha$; $\beta(-\infty)=0$ and $\beta(\infty)=1$.
- Derive the power function for the right-tailed test after class.

Test of the Mean with Unknown Variance

One-sample t-test

- If σ is unknown, we cannot use Z as a test statistic.
- A natural thought is to replace σ by S, the sample SD. Let $T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$. If $\mu = \mu_0$ (under H_0), then $T \sim t_{n-1}$. Similarly, consider tests of significance level α.
  - $H_0 : \mu = \mu_0$ versus $H_0 : \mu \neq \mu_0$. We reject it when T is too small or too large. Let $R = \{\mathbf{x} : T < t_{n-1}(\alpha/2) \text{ or } T > t_{n-1}(1-\alpha/2)\}$. The significance level is $\mathbb{P}(\mathbf{X} \in R | \mu = \mu_0) = \alpha$.
  - $H_0 : \mu \leq \mu_0$ versus $H_0 : \mu > \mu_0$. Let $R = \{\mathbf{x} : T > t_{n-1}(1-\alpha)\}$. The significance level is $\max_{\mu \leq \mu_0} \mathbb{P}(\mathbf{X} \in R) = \mathbb{P}(\mathbf{X} \in R | \mu = \mu_0) = \alpha$. Similar for $H_0 : \mu \geq \mu_0$.
  - The p-value of the tests are $2(1 - F_{t,n-1}(|T|))$ (two-sided) and $1 - F_{t,n-1}(T)$ (left-tailed).
  - The power function cannot be computed explicitly.

Tests for Standard Deviation of Normal Models

Equal-tailed chi-squared test

- From previous examples, the key to construct a test statistic is to make sure that
    - Its distribution is known under H_0.
    - The value of the statistic is closely related to whether H_0 is true.
- Consider normal samples with unknown σ and want to test $\sigma = \sigma_0$. Let $V = \frac{n-1}{\sigma_0^2} S^2$. If $\sigma = \sigma_0$, then $V \sim \chi^2_{n-1}$.
- Consider $H_0 : \sigma = \sigma_0$ versus $H_1 : \sigma \neq \sigma_0$.
    - We reject H_0 if V is too small or too large.
    - For $\alpha \in (0, 1)$, let $R = \left\{\mathbf{x} : V < \chi^2_{n-1}(\alpha/2) \text{ or } V > \chi^2_{n-1}(1 - \alpha/2)\right\}$. Its significance level is $\mathbb{P}(V \in R | \sigma = \sigma_0) = \alpha$.
- The one-sided tests follow similarly. Try to derive the p-value of given $\mathbf{x}$.

The Test for Non-normal Population

Normal approximation of non-normal population

- Consider X_i drawn from Bernoulli(p). Want to test p.
- Use the statistic $Y = \sum_{i=1}^{n} X_i$. Under $H_0 : p = p_0$, $Y \sim Binom(n, p)$. For given α, let $R = \{\mathbf{x} : Y < b_{n,p_0}(\alpha/2)$ or $Y > b_{n,p_0}(1-\alpha/2)\}$. Its significance level is α.
- The one-sided test and their power functions follow analogously.
- When n is large, $Z = \frac{Y - np_0}{\sqrt{np_0(1-p_0)}}$ is approximately standard normal under H_0. We can use $R = \{Z < z(\alpha/2) \cup Z > z(1-\alpha/2)\}$.
- The same approximation can be used for any distribution if you want to test their means. Simply use Z, T, V and the quantiles.

Two-sample Normal Model: Test Means with Known Variances

Suppose $\boldsymbol{X} = (X_1, \ldots, X_{n_1})$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_{n_2})$ are random samples from $\mathcal{N}(\mu_x, \sigma_x)$ and $\mathcal{N}(\mu_y, \sigma_y)$.

- We want to test want to test $H_0 : \mu_x = \mu_y$ (or $\mu_x \leq \mu_y$).
- Conditions: independent populations; normal distribution, or n_1 and n_2 large enough; known σ_x and σ_y.
- Use

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\sigma_x^2/n_1 + \sigma_y^2/n_2}}$$

  which has standard normal distribution when $\mu_x = \mu_y$.
- If $|Z| > z(1 - \alpha/2)$, then p-value is less than α. Equivalently, H_0 is rejected at significance level α.
- We seldom mention power in practice, which is usually hard to analyze.

Two-sample t-test

Example: We observe $X_1, \dots, X_{n_1}$, and $Y_1, \dots, Y_{n_2}$, want to test $H_0 : \mu_x - \mu_y = \delta$ (or similar hypotheses).

- Conditions: independent populations; normal distribution, or n_1 and n_2 large enough; unknown variances.
- Equal variance: use pooled estimator: $S_{xy}^2 = \frac{\sum_{i=1}^{n_1}(X_i - \bar{X})^2 + \sum_{i=1}^{n_2}(Y_i - \bar{Y})^2}{n_1 + n_2 - 2}$ (from HW4, it is UMVUE!); it this a unbiased estimator for the variance of two populations, whether or not they have the same mean.
- Unequal variance: use S_x^2 and S_y^2.
- Use

$$T = \frac{\bar{X} - \bar{Y} - \delta}{S_{xy}\sqrt{1/n_1 + 1/n_2}} \quad \text{for equal variance,}$$

$$V = \frac{\bar{X} - \bar{Y} - \delta}{\sqrt{S_x^2/n_1 + S_y^2/n_2}} \quad \text{for unequal variance.}$$

Two-sample t-test

For the equal-V case,

- Under H_0, T follows a t-distritbution $n_1 + n_2 - 2$.
- If $T < t_{n_1+n_2-2}(\alpha/2)$ or $T > t_{n_1+n_2-2}(1-\alpha/2)$, then p-value is less than α.

For the unequal case,

- Under H_0, V approxiamtely follows a t-distribution with DoF

$$\frac{(S_x^2/n_1 + S_y^2/n_2)^2}{(S_x^2/n_1)^2/(n_1-1) + (S_y^2/n_2)^2/(n_2-1)}.$$

This approximation is better done when both n_1 and n_2 are larger than 5. The test is called the Welch's t-test.

Two-sample Paired t-test

Example: Similar to the previous example, observe (X_1, Y_1), ..., (X_n, Y_n), want to test $H_0 : \mu_x = \mu_y$ (or $\mu_x \leq \mu_y$).

- Conditions: correlated populations; normal distribution, or n large enough.
- This time, we use
$$T = \frac{\bar{X} - \bar{Y}}{S_{x-y}/\sqrt{n}}.$$
Under H_0, it is t-distribution with DoF $n - 1$. How is it different from the unpaired version?

Two-sample F-test

Example: Observe $X_1, \ldots, X_{n_1}$, and $Y_1, \ldots, Y_{n_2}$, want to test $H_0 : \sigma_x^2/\sigma_y^2 = \rho$ $(H_0 : \sigma_x^2/\sigma_y^2 > \rho)$.

- Normal population or large samples
- Under H_0, $F = \frac{S_x^2}{\rho S_y^2}$ follows a F distribution with DoF $n_1 - 1$ and $n_2 - 1$.
- For one-sided test, if $F < F_{n_1-1,n_2-1}(\alpha)$, reject H_0.
- For two-sided test, if $F < F_{n_1-1,n_2-1}(\alpha/2)$ or $F > F_{n_1-1,n_2-1}(1-\alpha/2)$, reject H_0.

χ^2-test for One-sample Bernoulli Model

Example: Suppose $X_1, \ldots, X_n$ is a random sample from Bernoulli distribution with p. We want to test H_0: $p = p_0$.

- We have mentioned this test. Define the statistic $Z = \frac{\sum_{i=1}^n X_i - np_0}{\sqrt{np_0(1-p_0)}}$. It is approximately normal for large n. So we reject it if $|Z| \geq z(1-\alpha/2)$.
- Equivalently, $R = \{V = Z^2 > \chi_1^2(1-\alpha)\}$
- Simple algebra implies

$$V = \frac{(\sum_{i=1}^n X_i - np_0)^2}{np_0} + \frac{(\sum_{i=1}^n (1-X_i) - n(1-p_0))^2}{n(1-p_0)}$$
$$= \frac{(O_0 - E_0)^2}{E_0} + \frac{(O_1 - E_1)^2}{E_1}.$$

χ^2-test for Multi-sample Bernoulli Model

Now for multi-sample. Let $X_i = (X_{i,1}, \ldots, X_{i,n_i})$ be a random sample from Bernoulli population i for $i \in \{1, \ldots, m\}$.

- The completely specified case
  - The null hypothesis H_0: the probabilities are $(p_1, \ldots, p_m)$.
  - Define

$$V = \sum_{i=1}^{m} \left(\frac{(\sum_{j=1}^{n_i} X_{i,j} - n_i p_i)^2}{n_i p_i} + \frac{(\sum_{j=1}^{n_i} (1 - X_{i,j}) - n_i(1 - p_i))^2}{n_i(1 - p_i)} \right)$$
$$= \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i} \approx \chi_m^2.$$

  - The test statistic measures the discrepancy between the expected and observed frequencies, over all outcomes and all samples.
  - The DoF here is m.
  - Reject if $V > \chi_m^2(1 - \alpha)$.

χ^2-test for Multi-sample Bernoulli Model

Let $X_i = (X_{i,1}, \ldots, X_{i,n_i})$ be a random sample from Bernoulli population i for $i \in \{1, \ldots, m\}$.

- The equal probability case
  - The null H_0: $p_1 = p_2 = \cdots = p_m$.
  - Under H_0, the m samples can be combined to form one large sample of Bernoulli trials. Thus, a natural approach is to estimate p and then define the test statistic that measures the discrepancy between the expected and observed frequencies.
  - We can consider the test statistic

$$V = \sum_{i=1}^{m} \left(\frac{\left(\sum_{j=1}^{n_i} X_{i,j} - n_i p\right)^2}{n_i p} + \frac{\left(\sum_{j=1}^{n_i} (1 - X_{i,j}) - n_i(1-p)\right)^2}{n_i(1-p)} \right)$$

    where $p = \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n_i} X_{i,j}$ is the total frequency.
  - It can be shown that $V \sim \chi^2_{m-1}$ approximately.
  - We lose one DoF because we have to estimate p.

χ^2-test for One-sample Multinomial Model

- One-sample multinomial test: $X_1, \ldots, X_n$ drawn from multinomial trials with outcome possible outcomes $(a_1, \ldots, a_k)$. Want to test H_0: the probabilities are $(f_1, \ldots, f_k)$ with $\sum_{j=1}^k f_j = 1$.
- Construct
$$V = \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j}.$$
where $O_j = \sum_{i=1}^n \mathbb{1}_{X_i = a_j}$ is the observations of a_j, and $E_j = nf_j$ is the expected number of observations of a_j.
- Bernoulli is the special case of $k = 2$.
- Approximately, $V \sim \chi^2_{k-1}$. (See Theorem 6.9, Jun Shao for details.)
- Can you generalize it to multi-sample case? And test equal PMF? Hint: DoFs are $m(k-1)$ and $(m-1)(k-1)$.

Goodness of Fit Tests

- The one-sample multinomial model can be used to test whether a sample is drawn from a particular distribution.
- **Example:** Observe $X_1,\ldots,X_n$, want to test $H_0 : X_i \sim \mathcal{N}(0,1)$. We can partition $\mathbb{R}$ into, say, 10 regions denoted A_j.
  - The null hypothesis is $\mathbb{P}(X \in A_j) = \mathbb{P}(Z \in A_j)$ for all j.
  - Use the multinomial χ^2 test.
- We usually want to partition the domain $\mathbb{R}$ as much as possible, but make sure each region should have more than 5 samples.
- If it is not a single distribution but a parametric family, say, $\mathcal{N}(\mu, \sigma)$, then estimate μ and σ first and then proceed as above.

Test of Independence

- Observe discrete samples $(X_1, Y_1), \ldots, (X_n, Y_n)$, want to test H_0: X is independent of Y.
- Build a frequency table of X and Y. Say $\#\{X = a_i, Y = b_j\} = M_{ij}$, and the marginal frequency $M_{\cdot j}$ and $M_{i\cdot}$.
- Construct the following test statistics:

$$V = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(M_{ij} - M_{i\cdot} M_{\cdot j}/n)^2}{M_{i\cdot} M_{\cdot j}/n}.$$

  The limiting distribution of V is $\chi^2_{(I-1)(J-1)}$. How to find the DoF? $(I-1)(J-1) = IJ - (I-1) - (J-1) - 1$.
- For continuous random variables, first group them by a partition.

A Non-Parametric Test – Kolmogorov-Smirnov Test

If we do not specify the family of distribution, we many consider the following hypotheses

$$H_0 : F = F_0, \quad \text{versus} \quad H_1 : F \neq F_0.$$

Let F_n be the empirical cdf (sufficient statistic!) and let

$$D_n(F) = \sup_x |F_n(x) - F(x)|.$$

Intuitively, if H_0 were true, one would expect $D_n(F_0)$ should be small.

Kolmogorov-Smirnov Test

The statistic $D_n(F_0)$ is called the Kolmogorov-Smirnov statistic. The rejection region is $D_n(F_0) > c$. Furthermore, for $t > 0$

$$\lim_{n\to\infty} \mathbb{P}(\sqrt{n} D_n(F) \leq t) = 1 - 2\sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2t^2}.$$

Union-Intersection Tests

If the null hypothesis can be expressed as

$$H_0 : \theta \in \bigcap_{\gamma \in \Gamma} \Theta_\gamma$$

Consider the test: $H_{0,\gamma} : \theta \in \Theta_\gamma$ v.s. $H_{1,\gamma} : \theta \in \Theta_\gamma^c$. If the rejection region for text γ is $\{\mathbf{x} : T_\gamma(\mathbf{x}) \in R_\gamma\}$, then the rejection region for the test is

$$\bigcup_{\gamma \in \Gamma} \{\mathbf{x} : T_\gamma(\mathbf{x}) \in R_\gamma\}$$

Example: (Normal LRT revisit) For a sample from $\mathcal{N}(\mu, \sigma^2)$.
Test $H_0 : \mu = \mu_0 = \{\mu : \mu \leq \mu_0\} \cap \{\mu : \mu \geq \mu_0\}$ versus $H_1 : \mu \neq \mu_0$.

$$\text{Reject } H_0 \text{ if } \frac{\bar{x} - \mu_0}{S/\sqrt{n}} \geq t^* \text{ or } \frac{\bar{x} - \mu_0}{S/\sqrt{n}} \leq -t^*$$

Intersection-Union Tests

If the null hypothesis can be expressed as

$$H_0 : \theta \in \bigcup_{\gamma \in \Gamma} \Theta_\gamma$$

Consider the test: $H_{0,\gamma} : \theta \in \Theta_\gamma$ v.s. $H_{1,\gamma} : \theta \in \Theta_\gamma^c$. If the rejection region for test γ is $\{\mathbf{x} : T_\gamma(\mathbf{x}) \in R_\gamma\}$, then the rejection region for the test is

$$\bigcap_{\gamma \in \Gamma} \{\mathbf{x} : T_\gamma(\mathbf{x}) \in R_\gamma\}$$