STA 2002

许炜垣

Lecture 1-2

1. F- distribution

Definition 4.6.3 The F *distribution* with m and n degrees of freedom (or $F(m,n)$) is the distribution of the random variable

$$F = \frac{(X_1^2 + X_2^2 + \cdots + X_m^2)/m}{(Y_1^2 + Y_2^2 + \cdots + Y_n^2)/n}, \quad \frac{\chi^2(m)/m}{\chi^2(n)/n}$$

where $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ are i.i.d., each with the standard normal distribution. (Equivalently, $Z = (X/m)/(Y/n)$, where $X \sim \chi^2(m)$ and $Y \sim \chi^2(n)$.)

$Z \sim F(m,n) \Rightarrow \frac{1}{Z} \sim F(n,m)$

$F_\alpha(m,n) = \frac{1}{F_{1-\alpha}(n,m)}$

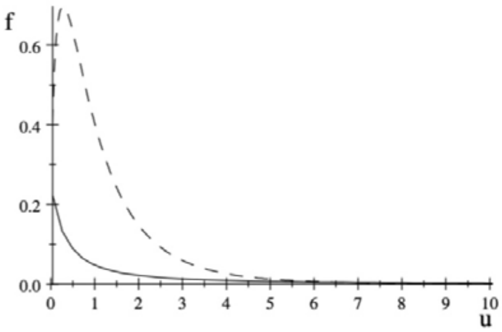
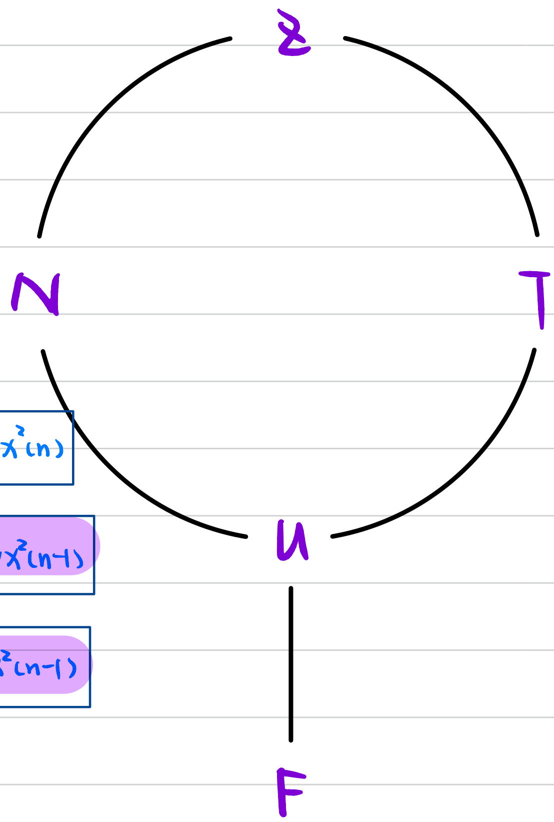


Figure 4.6.4: Plot of the $F(2,1)$ (solid line) and the $F(3,10)$ (dashed line) density functions.

2. Revision of distributions related to normal distribution



$\boxed{X_i \sim N} \Rightarrow \boxed{\frac{X_i - u}{\sigma} \sim Z} \Rightarrow \boxed{\frac{\bar{X} - u}{\frac{\sigma}{\sqrt{n}}} \sim Z}$

$\boxed{T = \frac{Z}{\sqrt{U/r}}} \Rightarrow \boxed{\begin{cases} Z = \frac{\bar{X}-u}{\frac{\sigma}{\sqrt{n}}} \\ \qquad \Rightarrow \frac{\bar{X}-u}{\frac{S}{\sqrt{n}}} \sim t(n-1) \\ U = \frac{(n-1)S^2}{\sigma^2} \end{cases}}$

$\boxed{X_i \sim N} \Rightarrow \boxed{\left(\frac{X_i - u}{\sigma}\right)^2 \sim \chi^2(1)} \Rightarrow \boxed{\sum_{i=1}^{n}\left(\frac{X_i - u}{\sigma}\right)^2 \sim \chi^2(n)}$

$\Downarrow$

$\boxed{\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^2 \sim \chi^2(n-1)}$

$\Downarrow$

$\boxed{\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)}$

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

Lecture 3

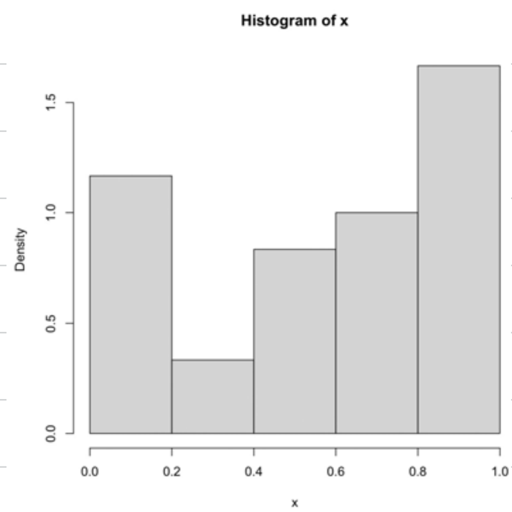
1. Data visualization

① Histogram

a. Relative frequency histogram

频率　　　　　　频数

The definition of relative frequency: Relative frequency $= \dfrac{\text{Frequency}}{\text{times of observations}} = \dfrac{f}{n}$



$$\frac{f}{n} = h(x)\cdot(c_i - c_{i-1})$$

$$\Rightarrow \underline{h(x) = \frac{f_i}{n(c_i - c_{i-1})}}$$

b. Bin

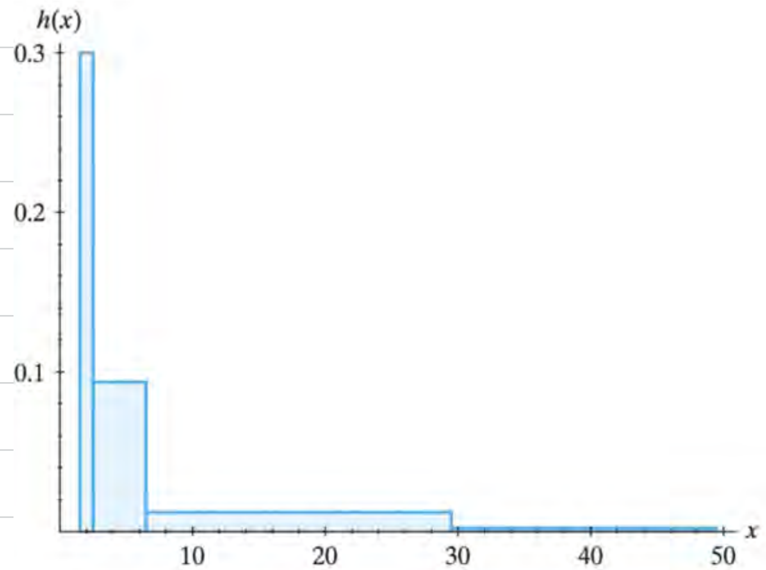
△ Bin size

i) Different bin size reveal different level of details:

- Too small: Too many details
- Too large: Too little details

ii) Uneven bins

We can use uneven bins when the data is skewed and some regions contain too few points



STEM		LEAF
1	628	:5
0	629	:
3	630	:358
3	631	:033
2	632	:77
9	633	:001446669
5	634	:01335
10	635	:0000113668
7	636	:0013689
2	637	:88
6	638	:334668
5	639	:22223
0	640	:
1	641	:2
3	642	:147
0	643	:
2	644	:02

② Stem-and-Leaf plots

63.78	63.45	63.58	63.08	63.40	64.42	63.27	63.10
63.34	63.50	63.83	63.63	63.27	63.30	63.83	63.50
63.36	63.86	63.34	63.92	63.88	63.36	63.36	63.51
63.51	63.84	64.27	63.50	63.56	63.39	63.78	63.92
63.92	63.56	63.43	64.21	64.24	64.12	63.92	63.53
63.50	63.30	63.86	63.93	63.43	64.40	63.61	63.03
63.68	63.13	63.41	63.60	63.13	63.69	63.05	62.85
63.31	63.66	63.60					

⇒

③ Box plot

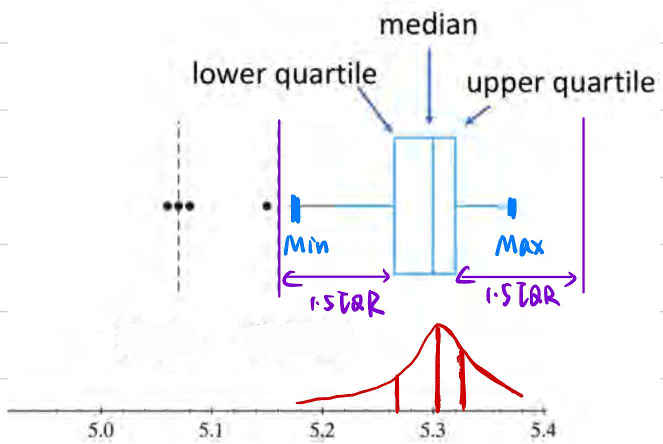
a. Concepts

Range = Difference of the extremes

Interquartile range = Difference of the first and third quartiles → it is a value

Midrange = Average of extremes

b. Explanations



upper quartile = third quartile

lower quartile = first quartile

Inner fences

outer fences

Suspected outliers

Outliers

x: value

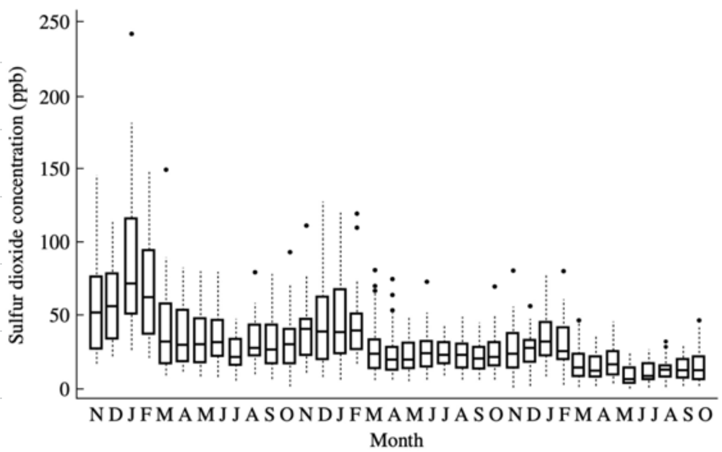
y: probability

$\xrightarrow{\text{Condense some information}}$

x: value with some information of probability (cdf)

Essence: 2 variables → 1 variable

⇓



2. Overview of statistical inference

① Overview

Data from sampling distribution (Sample) $\xrightarrow{\text{Estimation}}$ Parameters of underlying distribution (Population)

Hypothesis testing

(We know the distribution, but not the parameters)

Testing whether the parameter is correct

Estimation
- Point estimation
  - Maximum likelihood estimate (MLE) — more accurate
  - Method of moment (MM) — easier to solve
- Interval estimation

② Concepts

a. Random sample : △ members in a random sample are independently chosen; (default i.i.d)

△ members in a random sample all have nonzero probability

b. Simple random sample : △ it is a random sample;

△ members have the same probability to be chosen

3. Maximum likelihood

① Definition of likelihood

a. Original definition

$Lik(\theta) = \underline{P(\theta \mid Data)}$

A function of θ

b. Practical definition

$L(\theta) = P(Data \mid \theta)$

$$P(\theta \mid Data) = \frac{P(\theta \cap Data)}{P(Data)} = \frac{\boxed{P(Data \mid \theta)}\,\boxed{P(\theta)}}{\boxed{\sum_\theta P(Data \mid \theta) P(\theta)}}$$

Without given data, it stays the same

Stays the same

Essence: Probability ⇒ Probability function (Computable)

c. Two situations of RVs

$$\begin{cases} \text{Discrete type} & L(Data \mid \theta) = P(Data \mid \theta) \\ \text{Continuous type} & L(Data \mid \theta) = f(Data \mid \theta) \end{cases}$$

② Types of likelihood

If the X_is are i.i.d. (random sample)

$$\begin{cases} \text{Likelihood of } \theta: & L(\theta) = f(x_1 \cap x_2 \cap \cdots \cap x_n \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta) \quad (\text{"}\mid\text{"} / \text{"};\text{"}) \\ \text{Log likelihood of } \theta: & l(\theta) = \ln\left(\prod_{i=1}^{n} f(x_i \mid \theta)\right) = \sum_{i=1}^{n} \ln[f(x_i \mid \theta)] \end{cases}$$

(l(θ))

⇒ with the same results

③ Maximum likelihood estimate (mle)

a. Computing the likelihood ⟶ $L(\theta) \mid l(\theta) = \cdots$

b. Maximizing the likelihood ⟶ $\hat{\theta} = \cdots$ a function of X_is (不一定通过求一阶导)

c. Examining the maximum

$\hat{\theta}$ is an estimator

Examples

Ex: Normal

If $X_1, X_2, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$,

$$f(x_1, x_2, \ldots, x_n \mid \underline{\mu, \sigma}) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{x_i - \mu}{\sigma}\right]^2\right)$$

two paras

$$l(\mu, \sigma) = -n \log \sigma - \frac{n}{2} \log 2\pi - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2$$

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu) \qquad \frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \sigma^{-3} \sum_{i=1}^{n} (X_i - \mu)^2$$

two variables

⇒ find its partial derivative

$$\hat{\mu} = \overline{X} \qquad \hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2}$$

assign "$\hat{u} = \overline{X}$"

there are some errors

Uniform distribution

$$f(x) = \begin{cases} \frac{1}{R-L} & L \le x \le R \\ 0 & \text{otherwise.} \end{cases}$$

two paras

otherwise, $L(\theta) = 0$, obviously it is not a maximum

$$Lik(R, L) = \left(\frac{1}{R-L}\right)^n, \text{subjected to } R \ge \max(x), L \le \min(x)$$

Maximizing Lik(R,L) is equivalent to:

$$\text{minimizing } (R-L), \text{subjected to } R \ge \max(x), L \le \min(x)$$

Solution is $R = \max(x), L = \min(x)$

Lecture 4

1. Method of moments

① Sample moment

Var $\neq \hat{u}_2$!

$\hat{u}_k = \sum_{i=1}^{n} x_i^k \left(\frac{1}{n}\right)$

$\hat{u}_k = E(X_i^k) = \sum_{i=1}^{n} x_i^k f(x_i) = \sum_{i=1}^{n} x_i^k \left(\frac{1}{n}\right)$

② Process

a. Find u_is (parameters)

b. Find $\hat{u}_i$s (data)

c. $u_i = \hat{u}_i$ # of unknown paras = # of u_i's equation

Examples

Poisson distribution

$$m_Z(s) = \exp(\lambda(e^s - 1))$$

$$E(X) = m'_X(0) = \lambda e^0 \exp\left(\lambda(e^0 - 1)\right) = \lambda$$

$$E(X^2) = m''_X(0) = \lambda e^0 \exp\left(\lambda(e^0 - 1)\right) + \left(\lambda e^0\right)^2 \exp\left(\lambda(e^0 - 1)\right) = \lambda + \lambda^2$$

$$\hat{\mu}_1 = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

the method of moments estimate of λ: $\hat{\lambda} = \overline{X}$.

Let $X \sim \text{Exponential}(\lambda)$

$$m_X(s) = \lambda(\lambda - s)^{-1}$$

$$E(X) = m'_X(0) = \lambda(\lambda - 0)^{-2} = \lambda/\lambda^2 = 1/\lambda$$

$$\text{Var}(X) = E(X^2) - (E(X))^2 = (2/\lambda^2) - (1/\lambda)^2 = 1/\lambda^2$$

$$\hat{\lambda} = 1/\bar{X}$$

Use the lower-order moment in order to make the estimation more accurate

估计n个参数就同时列n个关于moment的方程

2. Difference between MLE & MM

① MLE is more accurate

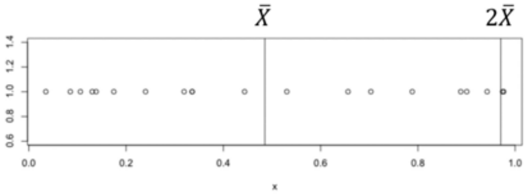
Uniform distribution:

First moment: $\int_L^R \frac{X}{R-L} dX = \frac{1}{R-L}\int_L^R XdX = \frac{1}{R-L}\int_L^R d\left(\frac{X^2}{2}\right) = \frac{R+L}{2}$

For a simple case, suppose L=0 is given, and $x_i \geq 0, \forall i$

Method of moment result: $\frac{R}{2} = \bar{X}$

But this can sometimes happen:



② MM is easier to solve

Gamma distribution:

$$f(x) = \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(a)} e^{-\lambda x}$$

when $x > 0$, with $f(x) = 0$ for $x \leq 0$.

MM:

Its MGF:

$$M(t) = \int_0^\infty e^{tx} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\, dx$$
$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{x(t-\lambda)}\, dx$$

The latter integral converges for $t < \lambda$ and can be evaluated by relating it to the gamma density having parameters α and $\lambda - t$. We thus obtain

$$M(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)}\left(\frac{\Gamma(\alpha)}{(\lambda-t)^\alpha}\right) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha$$

Take derivative against t, evaluate at t=0, to find the first two moments:

$$\mu_1 = \frac{\alpha}{\lambda}$$
$$\mu_2 = \frac{\alpha(\alpha+1)}{\lambda^2}$$

Solve for parameters when moments are estimated from sample:

$$\mu_2 = \mu_1^2 + \frac{\mu_1}{\lambda}$$
$$\lambda = \frac{\mu_1}{\mu_2 - \mu_1^2}$$
$$\alpha = \lambda\mu_1 = \frac{\mu_1^2}{\mu_2 - \mu_1^2}$$

MLE:

$$f(x|\alpha,\lambda) = \frac{1}{\Gamma(\alpha)}\lambda^\alpha x^{\alpha-1} e^{-\lambda x}, \qquad 0 \leq x < \infty$$

$$l(\alpha,\lambda) = \sum_{i=1}^n [\alpha \log\lambda + (\alpha-1)\log X_i - \lambda X_i - \log\Gamma(\alpha)]$$
$$= n\alpha\log\lambda + (\alpha-1)\sum_{i=1}^n \log X_i - \lambda\sum_{i=1}^n X_i - n\log\Gamma(\alpha)$$

Setting derivative to zero, can't find a closed form solution. Need to resort to numeric approach to solve for the parameters.

$$\frac{\partial l}{\partial\alpha} = n\log\lambda + \sum_{i=1}^n \log X_i - n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$$
$$\frac{\partial l}{\partial\lambda} = \frac{n\alpha}{\lambda} - \sum_{i=1}^n X_i$$

No need to go into details

3. Sampling distribution

① Definition

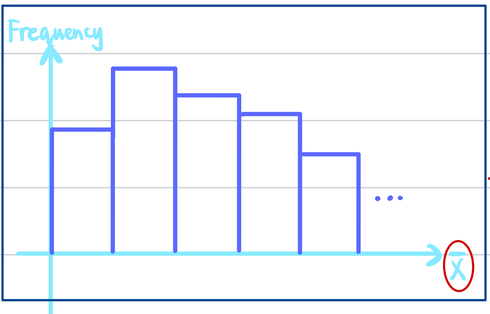
(usually, it's $\bar{X}$)

Sampling distribution is a probability distribution of a statistic obtained from a large number of samples.

Experiments / Observations / Members ⟶ a sample ⟶ samples

a sample: $\bar{X}$, S^2

samples: $\bar{X}_1, \bar{X}_2, \bar{X}_3 \dots$; $S_1^2, S_2^2, S_3^2 \dots$



单位从Xi变化至$\bar{X}_i$：探索$\bar{X}_i$的分布

根据CLT，当sample size足够大时，$\bar{X} \sim N(\mu, \frac{\sigma^2}{\sqrt{n}})$

⇒当 sample size / # of samples 都足够大时，sampling distribution 将会是很漂亮的 Normal D

② Statistics of this distribution

- Mean : $E(\bar{X}) = \mu$
- Variance : $Var(\bar{X}) = \frac{\sigma^2}{n}$
- Standard error : $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

Measure how far the sample mean to the true population mean

③ Unbiased estimates

a. Objects

Estimators (= Statistics = the function of Xi's)

b. Definition

- Unbiased : E(estimator) = Real parameter
- Biased : E(estimator) ≠ Real parameter

e.g. $E(\hat{\theta}) = \theta$, $E(\bar{X}) = \mu$, $E(S^2) = \sigma^2$ ✓

$E(S^2) = E(\sigma^2) \nRightarrow E(S) = E(\sigma)$

△ $bias(\hat{\theta}) = E(\hat{\theta}) - \theta$

△ Mean square error: $MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = (bias(\hat{\theta}))^2 + var(\hat{\theta})$

④ Review of CLT

In a sample from arbitrary distribution
- Xis are i.i.d.
- Sample size is large

⇒

$\bar{X}$s satisfy
$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

Lecture 5-6

1. Confidence interval (CI)

① Definition

A random interval calculated from the sample, that contains θ with some specified probability

If the coverage probability is $(1-\alpha)$, the interval is called a $100(1-\alpha)\%$ confidence interval

② Factors affecting the length of CI

a. Sample size ↑, the length of CI ↓; (Standard deviation ↓)

b. Standard deviation ↑, the length of CI ↑;

c. Confidence desired ↑ (α ↓), the length of CI ↑

③ error = |estimator − parameter| e.g. $|\bar{x}-u|$

ε

2. CI for a single mean

① Derivation

a. σ is known ⟶ Z

$$P\left(-z_{\frac{\alpha}{2}} \le \frac{\bar{X}-u}{\frac{\sigma}{\sqrt{n}}} \le z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

$$\Rightarrow P\left(\bar{X}-z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}} \le u \le \bar{X}+z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

b. σ is unknown ⟶ T

Since $\frac{\bar{X}-u}{\frac{S}{\sqrt{n}}} \sim t(n-1)$

then $P\left(-t_{\frac{\alpha}{2}}(n-1) \le \frac{\bar{X}-u}{\frac{S}{\sqrt{n}}} \le t_{\frac{\alpha}{2}}(n-1)\right) = 1-\alpha$

$$\Rightarrow P\left(\bar{X}-t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}} \le u \le \bar{X}+t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}\right) = 1-\alpha$$

② Four cases

will not change CI, but the name of CI will change: CI ⇒ App CI

Probability: 大写 $\bar{X}$
CI: 小写 $\bar{x}$

a. Case 1: X_i is normally distributed, σ is known

b. Case 2: X_i is arbitrarily distributed, σ is known

$(100-\alpha)\%$ CI: $\left[\bar{x}-z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}},\ \bar{x}+z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\right]$

c. Case 3: X_i is normally distributed, σ is unknown

d. Case 4: X_i is normally distributed, σ is unknown

$(100-\alpha)\%$ CI: $\left[\bar{x}-t_{\frac{\alpha}{2}}(n-1)\cdot\frac{S}{\sqrt{n}},\ \bar{x}+t_{\frac{\alpha}{2}}(n-1)\cdot\frac{S}{\sqrt{n}}\right]$

③ One-sided confidence interval

a. Case 1 & Case 2: $\left[\bar{x}-z_{\alpha}\cdot\frac{\sigma}{\sqrt{n}}, +\infty\right)$ / $\left(-\infty, \bar{x}+z_{\alpha}\cdot\frac{\sigma}{\sqrt{n}}\right]$

b. Case 3 & Case 4: $\left[\bar{x}-t_{\alpha}(n-1)\cdot\frac{S}{\sqrt{n}}, +\infty\right)$ / $\left(-\infty, \bar{x}+t_{\alpha}(n-1)\cdot\frac{S}{\sqrt{n}}\right]$

3. CI for the difference between two means

① Two cases (X_is and Y_is are ind)

Case 1 : X, Y are normally distributed, σ_X, σ_Y are known

Case 2 : X, Y are normally distributed, σ_X, σ_Y are unknown
- m & n are $\underline{\text{large}}$ (30) — Case 2.1
- m & n are small
  - $\sigma_X = \sigma_Y = \sigma$ — Case 2.2
  - $\sigma_X \neq \sigma_Y$ — Case 2.3

a. Case 1 : X, Y are normally distributed, σ_X, σ_Y are known

$$P\left(-z_{\frac{\alpha}{2}} \le \frac{\bar{x}-\bar{y}-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n}}} \le z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

$$\Rightarrow P\left(\bar{x}-\bar{y}-z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n}} \le (\mu_X-\mu_Y) \le \bar{x}-\bar{y}+z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n}}\right) = 1-\alpha$$

$$\Rightarrow (100-\alpha)\%\ CI : \left[\bar{x}-\bar{y}-z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n}},\ \bar{x}-\bar{y}+z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n}}\right]$$

b. Case 2.1 : X, Y are normally distributed, σ_X, σ_Y are unknown + m & n are large

Exp1

When m & n are large, S_X is closed to σ_X, S_Y is closed to σ_Y ($E(S^2)=\sigma^2$, $Var(S^2)=\frac{2\sigma^4}{n-1}$) → using $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

then we can do a simple substitution ☆ A trick, when (r is large, we can't find t-value, we use z-value to make a replacement

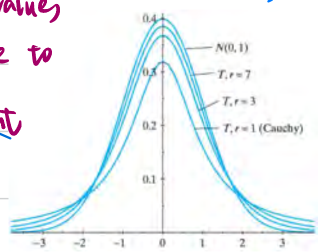
Exp2

When m & n are large, $t \to z$

then we can do a simple substitution

$$\Rightarrow (100-\alpha)\%\ CI : \left[\bar{x}-\bar{y}-z_{\frac{\alpha}{2}}\sqrt{\frac{S_X^2}{m}+\frac{S_Y^2}{n}},\ \bar{x}-\bar{y}+z_{\frac{\alpha}{2}}\sqrt{\frac{S_X^2}{m}+\frac{S_Y^2}{n}}\right]$$

两种explanation ⇒ 一种相同结果



c. Case 2.2 : X, Y are normally distributed, σ_X, σ_Y are unknown + m & n are small + $\sigma_X = \sigma_Y = \sigma$

$$Z = \frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma^2}{m}+\frac{\sigma^2}{n}}},\quad U = \frac{(m-1)S_X^2}{\sigma^2}+\frac{(n-1)S_Y^2}{\sigma^2} \sim \chi^2(m+n-2)$$

since (X_is, Y_is, X_is and Y_is) are i.i.d.

$$T = \frac{Z}{\sqrt{U/r}} = \frac{\dfrac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma^2}{m}+\frac{\sigma^2}{n}}}}{\sqrt{\left[\frac{(m-1)S_X^2}{\sigma^2}+\frac{(n-1)S_Y^2}{\sigma^2}\right]/(m+n-2)}} = \frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{\sum_{i=1}^{m}(X_i-\bar{X})^2+\sum_{i=1}^{n}(Y_i-\bar{Y})^2}{m+n-2}}\sqrt{\frac{1}{m}+\frac{1}{n}}} = \frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{S_p\cdot\sqrt{\frac{1}{m}+\frac{1}{n}}} \sim t(m+n-2)$$

Pooled estimator of σ^2 : $S_p = \sqrt{\frac{\sum_{i=1}^{m}(X_i-\bar{X})^2+\sum_{i=1}^{n}(Y_i-\bar{Y})^2}{m+n-2}}$

S_p^2 is an unbiased estimator : $E(S_p^2) = \frac{1}{m+n-2}\left[E\left(\sum_{i=1}^{m}(X_i-\bar{X})^2\right)+E\left(\sum_{i=1}^{n}(Y_i-\bar{Y})^2\right)\right]$

$= \frac{1}{m+n-2}\left[(m-1)\sigma^2+(n-1)\sigma^2\right] = \sigma^2$

$$\Rightarrow P\left(-t_{\frac{\alpha}{2}}(m+n-2) \le \frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{S_p\cdot\sqrt{\frac{1}{m}+\frac{1}{n}}} \le t_{\frac{\alpha}{2}}(m+n-2)\right) = 1-\alpha$$

$$\Rightarrow (100-\alpha)\%\ CI : \left[\bar{x}-\bar{y}-t_{\frac{\alpha}{2}}(m+n-2)\cdot S_p\sqrt{\frac{1}{m}+\frac{1}{n}},\ \bar{x}-\bar{y}+t_{\frac{\alpha}{2}}(m+n-2)\cdot S_p\sqrt{\frac{1}{m}+\frac{1}{n}}\right]$$

d. Case 2.3: X, Y are normally distributed, σ_X, σ_Y are unknown + m & n are small + $\sigma_X \neq \sigma_Y$

$\Rightarrow (100-\alpha)\%$ CI : $\left[\bar{x}-\bar{y}-t_{\frac{\alpha}{2}}(r)\cdot\sqrt{\frac{S_X^2}{m}+\frac{S_Y^2}{n}},\ \bar{x}-\bar{y}+t_{\frac{\alpha}{2}}(r)\cdot\sqrt{\frac{S_X^2}{m}+\frac{S_Y^2}{n}}\right]$

$$r = \left\lfloor \frac{\left(\frac{S_X^2}{m}+\frac{S_Y^2}{n}\right)^2}{\frac{1}{m-1}\left(\frac{S_X^2}{m}\right)^2+\frac{1}{n-1}\left(\frac{S_Y^2}{n}\right)^2} \right\rfloor$$

② Paired case

X, Y are normally distributed, σ_X, σ_Y are unknown, (X_i, Y_i) is a pair

(X_i, Y_i)s are ind, X_is and Y_is are dep

Let $D_i = X_i - Y_i$ 把双变量看成单变量

$$P\left(-t_{\frac{\alpha}{2}}(n-1) \le \frac{\bar{D}-(\mu_X-\mu_Y)}{S_D/\sqrt{n}} \le t_{\frac{\alpha}{2}}(n-1)\right) = 1-\alpha$$

$$\Rightarrow P\left(\bar{D}-t_{\frac{\alpha}{2}}(n-1)\frac{S_D}{\sqrt{n}} \le \mu_X-\mu_Y \le \bar{D}+t_{\frac{\alpha}{2}}(n-1)\frac{S_D}{\sqrt{n}}\right) = 1-\alpha$$

$$S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i-\bar{D})^2$$

$\Rightarrow (100-\alpha)\%$ CI : $\left[\bar{D}-t_{\frac{\alpha}{2}}(n-1)\frac{S_D}{\sqrt{n}},\ \bar{D}+t_{\frac{\alpha}{2}}(n-1)\frac{S_D}{\sqrt{n}}\right]$

4. CI for the difference between two proportions

① Single type

X_i's $\sim$ Bernoulli (p)

then by CLT, $\bar{X} \sim N(p, \frac{p(1-p)}{n})$

Since by point estimation, $\hat{p} = \bar{X}$

then $P(-z_{\frac{\alpha}{2}} \le \dfrac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \le z_{\frac{\alpha}{2}}) = 1-\alpha$

$\Rightarrow P(\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}) = 1-\alpha$ (we need to use $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ to approximate $\sqrt{\frac{p(1-p)}{n}}$)

$\Rightarrow (100-\alpha)\%$ CI : $[\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$

② Difference type

$$P(-z_{\frac{\alpha}{2}} \le \frac{(p_1-p_2)-(\hat{p}_1-\hat{p}_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \le z_{\frac{\alpha}{2}}) = 1-\alpha$$

$$\Rightarrow P(\hat{p}_1-\hat{p}_2 - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \le p_1 - p_2 \le \hat{p}_1-\hat{p}_2 + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}) = 1-\alpha$$

$$\Rightarrow (100-\alpha)\% \text{ CI} : [\hat{p}_1-\hat{p}_2 - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}, \hat{p}_1-\hat{p}_2 + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}]$$

Summary:

① 找点估计.

② 构造关于点估计 & 原参数的 Test-statistic 及 Test-statistic 的 Dis

③ 列概率不等式，求 CI

注意单边的界限是否为 ∞!

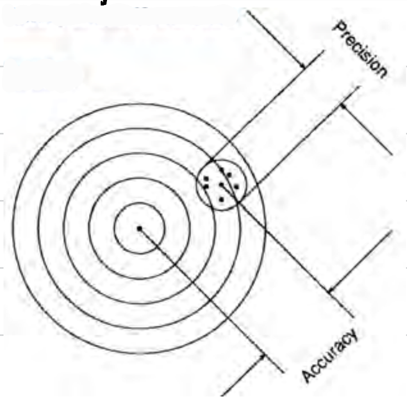
3. Sample size

① Sample size ↑, Cost ↑

② Precision & Accuracy

Precision 精 —— Precision is measured by the half-width of CI

Accuracy 准



③ Computation

a. Ascertain CI

μ
- if σ^2 is known: $(100-\alpha)\%$ CI: $[\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}]$
- if σ^2 is unknown: $(100-\alpha)\%$ CI: $[\bar{X} - z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}]$

把含n的表达式都替换掉

- $t_{\frac{\alpha}{2}}(n-1) \Rightarrow z_{\frac{\alpha}{2}}$: Since we don't know n, we use $z_{\frac{\alpha}{2}}$ instead
- S^2 : S^2 ~~should~~ will be estimated

p $(100-\alpha)\%$ CI: $[\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$ $\hat{p}$ depends on n

- Solution 1 : P is estimated as p^* (better)
- Solution 2 : Since $p(1-p) \le \frac{1}{4}$, then we take $\hat{p}(1-\hat{p}) = \frac{1}{4}$

b. Let half-length (called maximum error of the point estimate) of CI = 0.5, then find the corresponding $\lceil n \rceil$

Lecture 7-10

1. Basic concepts of Hypothesis testing

① Types of Hypotheses

a. Null hypothesis : An initial assumption
Alternative hypothesis

One-sided hypothesis test $H_0: \mu = 5$, $H_1: \mu > 5$
Two-sided hypothesis test $H_0: \mu = 5$, $H_1: \mu \neq 5$

b. Simple hypothesis
Composite hypothesis

② Test statistic T and Critical region C

Test statistic T : The statistic used to specify the Critical region
Critical region C : The region that we reject H_0

It's a statistic about our target statistic used for testing the hypothesis

By convention, we say that we fail to reject H_0, we don't use "accept"
we reject H_0

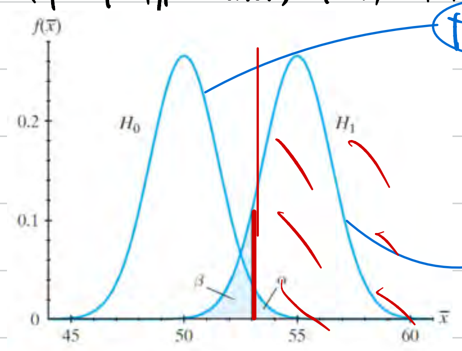
③ Types of errors

	H_0 true	H_1 true
Reject H_0	**Type I error**	Correct ☺
Fail to Reject H_0	Correct ☺	**Type II Error**

Next we define two probabilities:

$\alpha = P(\text{Type 1 error}) = P(T \in C \mid H_0)$

$\beta = P(\text{Type 2 error}) = P(T \notin C \mid H_1)$



$f(\overline{X} \mid H_0)$

$f(\overline{X} \mid H_1)$

If we change the critical region C, α and β change in different ways

Figure 8.1-1 pdf of $\overline{X}$ under H_0 and H_1

④ Significance level

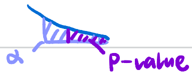
Most of the time, we fixed α, then determine the corresponding critical region.

Then at this time, the artificial upper bound of α is called Significance level.

⑤ p-value

p-value is the probability of observing 再取一个 a more extreme test statistic (in the direction that favors H_1) given that H_0 is true

We reject H_0 if p-value $\le \alpha$, the significance level.



⇒ 在α内 ⇒ reject H_0

⑥ Process

a. Determine Test parameter. H_0 & H_1 - Test stat & its distribution

b. Use either one of the three approaches to draw conclusion

2. Hypothesis testing for normal mean

① Two cases

Case 1: σ^2 is known
Case 2: σ^2 is unknown

a. Case 1: σ^2 is known

Null Hypothesis	Test Statistic	Distribution under H_0
$H_0: \mu = \mu_0$	$Z_0 = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$	$Z_0 \sim N(0,1)$

以T为中心　　以μ_0为中心

Alternative Hypothesis	Rejection/Critical Region (H_0 is rejected)	p-value	Confidence interval Approach
$H_1: \mu \neq \mu_0$	$\lvert Z_0 \rvert > Z_{\alpha/2}$	$2P(Z > \lvert Z_0 \rvert)$	$[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$
$H_1: \mu > \mu_0$	$Z_0 > Z_\alpha$	$P(Z > Z_0)$	$[\bar{X} - z_\alpha \frac{\sigma}{\sqrt{n}}, +\infty)$
$H_1: \mu < \mu_0$	$Z_0 < -Z_\alpha$	$P(Z < Z_0)$	$(-\infty, \bar{X} + z_\alpha \frac{\sigma}{\sqrt{n}}]$

Threshold: Critical values

H_0 is rejected when p-value<α

H_0 is rejected when the CI does NOT include μ_0

25

b. Case 2: σ^2 is unknown

Null Hypothesis	Test Statistic	Distribution under H_0
$H_0: \mu = \mu_0$	$t_0 = \dfrac{\bar{X} - \mu_0}{S / \sqrt{n}}$	$t_0 \sim t(n-1)$

Alternative Hypothesis	Rejection/Critical Region (H_0 is rejected)	p-value	Confidence interval Approach
$H_1: \mu \neq \mu_0$	$\lvert t_0 \rvert > t_{\alpha/2}(n-1)$	$2P(T > \lvert t_0 \rvert)$	$[\bar{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}]$
$H_1: \mu > \mu_0$	$t_0 > t_\alpha(n-1)$	$P(T > t_0)$	$[\bar{X} - t_\alpha(n-1)\frac{S}{\sqrt{n}}, +\infty)$
$H_1: \mu < \mu_0$	$t_0 < -t_\alpha(n-1)$	$P(T < t_0)$	$(-\infty, \bar{X} + t_\alpha(n-1)\frac{S}{\sqrt{n}}]$

Threshold: Critical values

H_0 is rejected when p-value<α

H_0 is rejected when the CI does NOT include μ_0

3. Hypothesis testing for normal variance

Null Hypothesis	Test Statistic	Distribution under H0
$H_0: \sigma^2 = \sigma_0^2$	$\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2}$	$\chi_0^2 \sim \chi^2(n-1)$

Alternative Hypothesis	Rejection/Critical Region (H_0 is rejected)	p-value	Confidence interval Approach
$H_1: \sigma^2 \neq \sigma_0^2$	$\chi_0^2 > \chi_{\alpha/2}^2(n-1)$ or $\chi_0^2 < \chi_{1-\alpha/2}^2(n-1)$	Approximation $2\min\{P(\chi^2 > \chi_0^2), P(\chi^2 < \chi_0^2)\}$	$[\frac{(n-1)s^2}{\chi_{\alpha/2}^2(n-1)}, \frac{(n-1)s^2}{\chi_{1-\alpha/2}^2(n-1)}]$
$H_1: \sigma^2 > \sigma_0^2$	$\chi_0^2 > \chi_\alpha^2(n-1)$ Threshold: Critical values	$P(\chi^2 > \chi_0^2)$	$[\frac{(n-1)s^2}{\chi_\alpha^2(n-1)}, +\infty)$
$H_1: \sigma^2 < \sigma_0^2$	$\chi_0^2 < \chi_{1-\alpha}^2(n-1)$	$P(\chi^2 < \chi_0^2)$	$(\cancel{-\infty}\ 0, \frac{(n-1)s^2}{\chi_{1-\alpha}^2(n-1)}]$

H_0 is rejected when p-value<α

H_0 is rejected when the CI does NOT include σ_0^2

33

4. Hypothesis testing for equality of two normal means

① Two cases

Case 1: X_is and Y_is are ind
- σ_X^2, σ_Y^2 are known — Case 1.1
- σ_X^2, σ_Y^2 are unknown, $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ — Case 1.2
- σ_X^2, σ_Y^2 are unknown, $\sigma_X^2 \neq \sigma_Y^2$ — Case 1.3

Case 2: (X_i, Y_i)s are ind

a. Case 1.1: X_is and Y_is are ind, σ_X^2, σ_Y^2 are known

Null Hypothesis	Test Statistic	Distribution under H0
$H_0: \mu_X - \mu_Y = \Delta_0$	$Z_0 = \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$	$Z_0 \sim N(0,1)$

Alternative Hypothesis	Rejection/Critical Region (H_0 is rejected)	p-value	Confidence interval Approach
$H_1: \mu_X - \mu_Y \neq \Delta_0$	$\lvert Z_0 \rvert > Z_{\alpha/2}$	$2P(Z > \lvert Z_0 \rvert)$	$[\bar{X} - \bar{Y} - z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}, \bar{X} - \bar{Y} + z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}]$
$H_1: \mu_X - \mu_Y > \Delta_0$	$Z_0 > Z_\alpha$ Threshold: Critical values	$P(Z > Z_0)$	$[\bar{X} - \bar{Y} - z_\alpha\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}, +\infty)$
$H_1: \mu_X - \mu_Y < \Delta_0$	$Z_0 < -Z_\alpha$	$P(Z < Z_0)$	$(-\infty, \bar{X} - \bar{Y} + z_\alpha\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}]$

H_0 is rejected when p-value<α

H_0 is rejected when the CI does NOT include Δ_0

39

b. Case 1.2: X_is and Y_is are ind, σ_X^2, σ_Y^2 are unknown, $\sigma_X^2 = \sigma_Y^2 = \sigma^2$

Null Hypothesis $H_0: \mu_X - \mu_Y = \Delta_0$

Test Statistic $t_0 = \dfrac{\bar{X} - \bar{Y} - \Delta_0}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}}$ $\quad S_p^2 = \dfrac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$

Distribution under H0 $t_0 \sim t(n+m-2)$

If we are not sure, we should use F-test to determine:
- Pooled t-test
- Welch's t-test

Alternative Hypothesis	Rejection/Critical Region (H_0 is rejected)	p-value	Confidence interval Approach
$H_1: \mu_X - \mu_Y \neq \Delta_0$	$\lvert t_0 \rvert > t_{\alpha/2}(n+m-2)$	$2P(T > \lvert t_0 \rvert)$	$[\bar{X} - \bar{Y} - t_{\alpha/2}(n+m-2)S_p\sqrt{\frac{1}{n}+\frac{1}{m}}, \bar{X} - \bar{Y} + t_{\alpha/2}(n+m-2)S_p\sqrt{\frac{1}{n}+\frac{1}{m}}]$
$H_1: \mu_X - \mu_Y > \Delta_0$	$t_0 > t_\alpha(n+m-2)$	$P(T > t_0)$	$[\bar{X} - \bar{Y} - t_\alpha(n+m-2)S_p\sqrt{\frac{1}{n}+\frac{1}{m}}, +\infty)$
$H_1: \mu_X - \mu_Y < \Delta_0$	$t_0 < -t_\alpha(n+m-2)$	$P(T < t_0)$	$(-\infty, \bar{X} - \bar{Y} + t_\alpha(n+m-2)S_p\sqrt{\frac{1}{n}+\frac{1}{m}}]$

Threshold: Critical values

H_0 is rejected when p-value<α

H_0 is rejected when the CI does NOT include Δ_0

c. Case 1.3: X_is and Y_is are ind, σ_X^2, σ_Y^2 are unknown, $\sigma_X^2 \neq \sigma_Y^2$

Null Hypothesis $H_0: \mu_X - \mu_Y = \Delta_0$

Test Statistic $t_0 = \dfrac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}}$ $\quad r = \left\lfloor \dfrac{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right)^2}{\frac{1}{n-1}\left(\frac{S_X^2}{n}\right)^2 + \frac{1}{m-1}\left(\frac{S_Y^2}{m}\right)^2} \right\rfloor$

Distribution under H0 $t_0 \sim t(r)$

Alternative Hypothesis	Rejection/Critical Region (H_0 is rejected)	p-value	Confidence interval Approach
$H_1: \mu_X - \mu_Y \neq \Delta_0$	$\lvert t_0 \rvert > t_{\alpha/2}(r)$	$2P(T > \lvert t_0 \rvert)$	$[\bar{X} - \bar{Y} - t_{\alpha/2}(r)\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}, \bar{X} - \bar{Y} + t_{\alpha/2}(r)\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}]$
$H_1: \mu_X - \mu_Y > \Delta_0$	$t_0 > t_\alpha(r)$	$P(T > t_0)$	$[\bar{X} - \bar{Y} - t_\alpha(r)\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}, +\infty)$
$H_1: \mu_X - \mu_Y < \Delta_0$	$t_0 < -t_\alpha(r)$	$P(T < t_0)$	$(-\infty, \bar{X} - \bar{Y} + t_\alpha(r)\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}]$

Threshold: Critical values

H_0 is rejected when p-value<α

H_0 is rejected when the CI does NOT include Δ_0

47

d. Case 2: (X_i, Y_i)s are ind

Null Hypothesis $H_0: \mu_X - \mu_Y = \Delta_0$

Test Statistic $t_0 = \dfrac{\bar{D} - \Delta_0}{S_D/\sqrt{n}}$

Distribution under H0 $t_0 \sim t(n-1)$

Alternative Hypothesis	Rejection/Critical Region (H_0 is rejected)	p-value	Confidence interval Approach
$H_1: \mu_X - \mu_Y \neq \Delta_0$	$\lvert t_0 \rvert > t_{\alpha/2}(n-1)$	$2P(T > \lvert t_0 \rvert)$	$[\bar{D} - t_{\alpha/2}(n-1)\frac{S_D}{\sqrt{n}}, \bar{D} + t_{\alpha/2}(n-1)\frac{S_D}{\sqrt{n}}]$
$H_1: \mu_X - \mu_Y > \Delta_0$	$t_0 > t_\alpha(n-1)$	$P(T > t_0)$	$[\bar{D} - t_\alpha(n-1)\frac{S_D}{\sqrt{n}}, +\infty)$
$H_1: \mu_X - \mu_Y < \Delta_0$	$t_0 < -t_\alpha(n-1)$	$P(T < t_0)$	$(-\infty, \bar{D} + t_\alpha(n-1)\frac{S_D}{\sqrt{n}}]$

Threshold: Critical values

H_0 is rejected when p-value<α

H_0 is rejected when the CI does NOT include Δ_0

5. Hypothesis testing for equality of two normal variances

$$\begin{cases} U = \frac{(m-1)S_X^2}{\sigma_X^2} \sim \chi^2(m-1) \\ V = \frac{(n-1)S_Y^2}{\sigma_Y^2} \sim \chi^2(n-1) \end{cases}$$

$$F = \frac{U/u}{V/v} = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F(m-1, n-1)$$

系数要非常小心呀工肥囡!

Null Hypothesis	Test Statistic	Distribution under H0
$H_0: \sigma_X^2 = \sigma_Y^2$	$F_0 = \frac{S_X^2}{S_Y^2}$	$F_0 \sim F(n-1, m-1)$

不对称的双边HT都是取2×min

Alternative Hypothesis	Rejection/Critical Region (H_0 is rejected)	p-value	Confidence interval Approach (CI for $\frac{\sigma_X^2}{\sigma_Y^2}$)
$H_1: \sigma_X^2 \neq \sigma_Y^2$	$F_0 > F_{\alpha/2}(n-1, m-1)$ or $F_0 < F_{1-\alpha/2}(n-1, m-1)$	Approximation $2\min\{P(F > F_0), P(F < F_0)\}$	$[\frac{S_X^2}{S_Y^2} F_{1-\alpha/2}(m-1, n-1), \frac{S_X^2}{S_Y^2} F_{\alpha/2}(m-1, n-1)]$
$H_1: \sigma_X^2 > \sigma_Y^2$	$F_0 > F_\alpha(n-1, m-1)$	$P(F > F_0)$	$[\frac{S_X^2}{S_Y^2} F_{1-\alpha}(m-1, n-1), +\infty)$
$H_1: \sigma_X^2 < \sigma_Y^2$	$F_0 < F_{1-\alpha}(n-1, m-1)$	$P(F < F_0)$	$(\cancel{-\infty}\ 0, \frac{S_X^2}{S_Y^2} F_\alpha(m-1, n-1)]$

Threshold: Critical values

H_0 is rejected when p-value<α

H_0 is rejected when the CI does NOT include **1**

6. Hypothesis testing for proportions

Null Hypothesis

$$H_0: p = p_0$$

Test Statistic

$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Asymptotic Distribution under H0

$$Z_0 \sim N(0,1)$$

Alternative Hypothesis	Rejection/Critical Region (H_0 is rejected)		p-value	Confidence interval Approach
$H_1: p \neq p_0$	$\lvert Z_0 \rvert > Z_{\alpha/2}$	Threshold: Critical values	$2P(Z > \lvert Z_0 \rvert)$	$[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$
$H_1: p > p_0$	$Z_0 > Z_\alpha$		$P(Z > Z_0)$	$[\hat{p} - z_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \cancel{\infty}\ 1)$
$H_1: p < p_0$	$Z_0 < -Z_\alpha$		$P(Z < Z_0)$	$(\cancel{-\infty}\ 0, \hat{p} + z_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$

H_0 is rejected when p-value<α

H_0 is rejected when the CI does NOT include p_0

7. Hypothesis testing for equality of two proportions

Null Hypothesis

$$H_0: p_1 = p_2$$

Test Statistic

$$Z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad \hat{p} = \frac{\sum_{i=1}^{n_1} X_i + \sum_{i=1}^{n_2} Y_i}{n_1 + n_2}$$

Distribution under H_0

$$Z_0 \sim N(0,1)$$

Alternative Hypothesis	Rejection/Critical Region (H_0 is rejected)	p-value	Confidence interval Approach
$H_1: p_1 \neq p_2$	$\lvert Z_0 \rvert > Z_{\alpha/2}$	$2P(Z > \lvert Z_0 \rvert)$	$[\hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}, \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}]$
$H_1: p_1 > p_2$	$Z_0 > Z_\alpha$	$P(Z > Z_0)$	$[\hat{p}_1 - \hat{p}_2 - z_\alpha\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}, \cancel{\infty}\ 1)$
$H_1: p_1 < p_2$	$Z_0 < -Z_\alpha$	$P(Z < Z_0)$	$(\cancel{-\infty}\ -1, \hat{p}_1 - \hat{p}_2 + z_\alpha\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}]$

H_0 is rejected when p-value<α

H_0 is rejected when the CI does NOT include 0

区分 此时的
- CR / p-value → 需要 H_0 → pooled p
- CI → 无需 H_0 → $\hat{p}_1$、$\hat{p}_2$

非常好区别~

(pooled t-test 则无妨，一致)

Lecture 11

1. Statistical Power

Power 即 H_1 的正确可能性

① Definition

$\text{Power} = 1-\beta = P(T \in C \mid H_1)$

$\Rightarrow K(u_1) = 1-\beta(u_1) = P(T \in C \mid u=u_1)$ — u_1 is from H_1

e.g. H_1: "$u>u_0$" = "$u=u_1, u_1>u_0$" (因为最后u只能是某一个值)

$K(u_0) = 1-\beta(u_0) = P(T\in C \mid u=u_0) = P(T \in C \mid H_0) = \alpha$

2. Power for normal mean

① Power

Null Hypothesis	Test Statistic	Distribution under H_0
$H_0: \mu=\mu_0$	$Z_0 = \dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$	$Z_0 \sim N(0,1)$

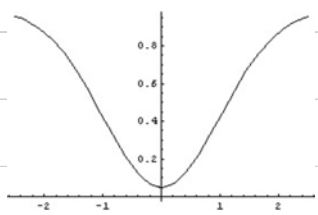
Alternative Hypothesis	Critical Region (H_0 is rejected)	Power under $\mu=\mu_1$	Sample size needed for $K(\mu_1)=1-\beta; K(\mu_0)=\alpha$
$H_1: \mu \neq \mu_0$	$\lvert Z_0 \rvert > Z_{\alpha/2}$	$\Phi\left(-Z_{\alpha/2}+\frac{\mu_1-\mu_0}{\sigma/\sqrt{n}}\right)+\Phi\left(-Z_{\alpha/2}-\frac{\mu_1-\mu_0}{\sigma/\sqrt{n}}\right)$	$n=\dfrac{(Z_\beta+Z_{\alpha/2})^2\sigma^2}{(\mu_1-\mu_0)^2}$
$H_1: \mu > \mu_0$	$Z_0 > Z_\alpha$	$\Phi\left(-Z_\alpha+\frac{\lvert\mu_0-\mu_1\rvert}{\sigma/\sqrt{n}}\right)$	$n=\dfrac{(Z_\beta+Z_\alpha)^2\sigma^2}{(\mu_1-\mu_0)^2}$
$H_1: \mu < \mu_0$	$Z_0 < -Z_\alpha$		

The factors that affect power:

- α : $\alpha\uparrow, K\uparrow$
- $\lvert u_1-u_0\rvert$: $\lvert u_1-u_0\rvert\uparrow, K\uparrow$
- σ : $\sigma\uparrow, K\downarrow$
- n : $n\uparrow, K\uparrow$

带绝对值号，所以对称~

a. Two-sided



$K(u_1) = P(Z_0 \in C \mid u=u_1) = P(|Z_0| > Z_{\frac{\alpha}{2}} \mid u=u_1) = P(Z_0 > Z_{\frac{\alpha}{2}} \mid u=u_1) + P(Z_0 < -Z_{\frac{\alpha}{2}} \mid u=u_1)$

Under $u=u_1$, $\dfrac{\bar{X}-u_1}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$, then $\left(\dfrac{\bar{X}-u_1}{\frac{\sigma}{\sqrt{n}}} + \dfrac{u_1-u_0}{\frac{\sigma}{\sqrt{n}}}\right) \sim N\left(\dfrac{u_1-u_0}{\frac{\sigma}{\sqrt{n}}}, 1\right) \Rightarrow Z_0 \sim N\left(\dfrac{u_1-u_0}{\frac{\sigma}{\sqrt{n}}}, 1\right)$

$\Rightarrow P\left(Z_0 - \dfrac{u_1-u_0}{\frac{\sigma}{\sqrt{n}}} > Z_{\frac{\alpha}{2}} - \dfrac{u_1-u_0}{\frac{\sigma}{\sqrt{n}}}\right) + P\left(Z_0 - \dfrac{u_1-u_0}{\frac{\sigma}{\sqrt{n}}} < -Z_{\frac{\alpha}{2}} - \dfrac{u_1-u_0}{\frac{\sigma}{\sqrt{n}}}\right)$

$= 1-\Phi\left(Z_{\frac{\alpha}{2}} - \dfrac{u_1-u_0}{\frac{\sigma}{\sqrt{n}}}\right) + \Phi\left(-Z_{\frac{\alpha}{2}} - \dfrac{u_1-u_0}{\frac{\sigma}{\sqrt{n}}}\right)$

$= \Phi\left(-Z_{\frac{\alpha}{2}} + \dfrac{u_1-u_0}{\frac{\sigma}{\sqrt{n}}}\right) + \Phi\left(-Z_{\frac{\alpha}{2}} - \dfrac{u_1-u_0}{\frac{\sigma}{\sqrt{n}}}\right)$

$\begin{cases} H_0: u=0 \\ H_1: u\neq 0 \end{cases}$

b. One-sided

△ For $\begin{cases} H_0: u=u_0, \\ H_1: u>u_0 \quad (u_1>u_0) \end{cases}$

$K(u_1) = P(Z_0 > Z_\alpha \mid u=u_1) = P\left(Z_0 - \dfrac{u_1-u_0}{\frac{\sigma}{\sqrt{n}}} > Z_\alpha - \dfrac{u_1-u_0}{\frac{\sigma}{\sqrt{n}}}\right)$

$= 1-\Phi\left(Z_\alpha - \dfrac{u_1-u_0}{\frac{\sigma}{\sqrt{n}}}\right)$

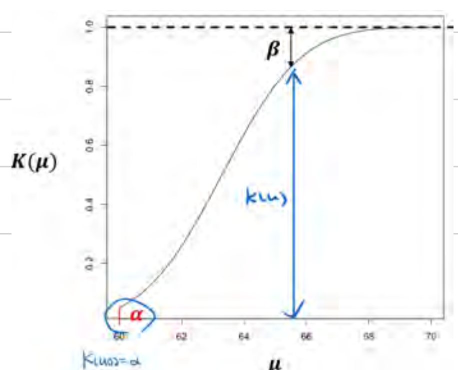
$= \Phi\left(-Z_\alpha + \dfrac{u_1-u_0}{\frac{\sigma}{\sqrt{n}}}\right)$

△ For $\begin{cases} H_0: u=u_0, \\ H_1: u<u_0 \quad (u_1<u_0) \end{cases}$

$K(u_1) = P(Z_0 < -Z_\alpha \mid u=u_1) = P\left(Z_0 - \dfrac{u_1-u_0}{\frac{\sigma}{\sqrt{n}}} < -Z_\alpha - \dfrac{u_1-u_0}{\frac{\sigma}{\sqrt{n}}}\right)$

$= \Phi\left(-Z_\alpha - \dfrac{u_1-u_0}{\frac{\sigma}{\sqrt{n}}}\right)$

$$\boxed{\Rightarrow \text{For one sided, } K(u_1) = \Phi\left(-Z_\alpha + \dfrac{|u_1-u_0|}{\frac{\sigma}{\sqrt{n}}}\right)}$$



② Sample size

a. Two-sided

can be ignored (n is large)

$$K(\mu_1)=\Phi\left(-Z_{\frac{\alpha}{2}}+\frac{\mu_1-\mu_0}{\frac{\sigma}{\sqrt{n}}}\right)+\underline{\Phi\left(-Z_{\frac{\alpha}{2}}-\frac{\mu_1-\mu_0}{\frac{\sigma}{\sqrt{n}}}\right)}=1-\beta$$

$$\Rightarrow -Z_{\frac{\alpha}{2}}+\frac{\mu_1-\mu_0}{\frac{\sigma}{\sqrt{n}}}=Z_\beta$$

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

b. One-sided

$$K(\mu_1)=\Phi\left(-Z_\alpha+\frac{|\mu_1-\mu_0|}{\frac{\sigma}{\sqrt{n}}}\right)=1-\beta$$

$$\Rightarrow -Z_\alpha+\frac{|\mu_1-\mu_0|}{\frac{\sigma}{\sqrt{n}}}=Z_\beta$$

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

Factors Affecting the Sample Size

(1) The sample size increases as σ^2 increases.

(2) The sample size increases as the significance level is made smaller (α decreases).

(3) The sample size increases as the required power increases ($1-\beta$ increases).

(4) The sample size decreases as the absolute value of the distance between the null and alternative means ($|\mu_0-\mu_1|$) increases.

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

two-sided

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

one-sided

3. Power for equality of two normal means

$n=m,\ \sigma_X=\sigma_Y=\sigma,$	Null Hypothesis	Test Statistic	Distribution under H_0
	$H_0:\mu_X=\mu_Y$	$Z_0=\dfrac{\bar{X}-\bar{Y}}{\sigma\sqrt{2/n}}$	$Z_0\sim N(0,1)$

Alternative Hypothesis	Critical Region (H_0 is rejected)	Power under $\mu_X-\mu_Y=\Delta$	Sample size needed for $K(\Delta)=1-\beta$; $K(0)=\alpha$
$H_1:\mu_X\neq\mu_Y$	$\lvert Z_0\rvert>Z_{\alpha/2}$	$\Phi\left(-Z_{\alpha/2}+\frac{\Delta}{\sigma}\sqrt{\frac{n}{2}}\right)+\Phi\left(-Z_{\alpha/2}-\frac{\Delta}{\sigma}\sqrt{\frac{n}{2}}\right)$	$n=2\dfrac{(Z_\beta+Z_{\alpha/2})^2\sigma^2}{\Delta^2}$
$H_1:\mu_X>\mu_Y$	$Z_0>Z_\alpha$	$\Phi\left(-Z_\alpha+\frac{\lvert\Delta\rvert}{\sigma}\sqrt{\frac{n}{2}}\right)$	$n=2\dfrac{(Z_\beta+Z_\alpha)^2\sigma^2}{\Delta^2}$
$H_1:\mu_X<\mu_Y$	$Z_0<-Z_\alpha$		

when $\begin{cases} m=n=n \\ \sigma_X=\sigma_Y=\sigma \end{cases}$

Not necessary, we can use $\sqrt{\sigma_X^2+\sigma_Y^2}$ to replace σ

计算sample size时有相同的sample size当然会更好，但方差其实不作统一要求，这里这样书写仅是为了简化表达

a. Two-sided

$$Z_0=\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}\sim N\left(\frac{\Delta}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}},1\right)$$

$$K(\Delta)=P(|Z_0|>Z_{\frac{\alpha}{2}}\mid\mu_X-\mu_Y=\Delta)=P(Z_0>Z_{\frac{\alpha}{2}}\mid\mu_X-\mu_Y=\Delta)+P(Z_0<-Z_{\frac{\alpha}{2}}\mid\mu_X-\mu_Y=\Delta)$$

$$=P\left(Z_0-\frac{\Delta}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}>Z_{\frac{\alpha}{2}}-\frac{\Delta}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}\right)+P\left(Z_0-\frac{\Delta}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}<-Z_{\frac{\alpha}{2}}-\frac{\Delta}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}\right)$$

$$=\Phi\left(-Z_{\frac{\alpha}{2}}+\frac{\Delta}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}\right)+\Phi\left(-Z_{\frac{\alpha}{2}}-\frac{\Delta}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}\right)$$

4. Power for proportion

Null Hypothesis	Test Statistic	Distribution under H_0
$H_0: p = p_0$	$Z_0 = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$	$Z_0 \sim N(0,1)$

Alternative Hypothesis	Critical Region (H_0 is rejected)	Power under $p = p_1$	Sample size needed for $K(\Delta) = 1-\beta; K(0) = \alpha$
$H_1: p \neq p_0$	$\lvert Z_0 \rvert > Z_{\alpha/2}$	$\Phi\left(\frac{-Z_{\alpha/2}\sqrt{p_0(1-p_0)} + \sqrt{n}(p_1 - p_0)}{\sqrt{p_1(1-p_1)}}\right) + \Phi\left(\frac{-Z_{\alpha/2}\sqrt{p_0(1-p_0)} - \sqrt{n}(p_1 - p_0)}{\sqrt{p_1(1-p_1)}}\right)$	$n = \frac{\left(Z_\beta\sqrt{p_1(1-p_1)} + Z_{\alpha/2}\sqrt{p_0(1-p_0)}\right)^2}{(p_1-p_0)^2}$
$H_1: p > p_0$	$Z_0 > Z_\alpha$	$\Phi\left(\frac{-Z_\alpha\sqrt{p_0(1-p_0)} + \sqrt{n}(\lvert p_1 - p_0\rvert}{\sqrt{p_1(1-p_1)}}\right)$	$n = \frac{\left(Z_\beta\sqrt{p_1(1-p_1)} + Z_\alpha\sqrt{p_0(1-p_0)}\right)^2}{(p_1-p_0)^2}$
$H_1: p < p_0$	$Z_0 < -Z_\alpha$		

a. Two-sided

Since $\dfrac{\hat{p}-p_1}{\sqrt{\frac{p_1(1-p_1)}{n}}} \sim N(0,1)$, then $\dfrac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \left(\dfrac{\hat{p}-p_1}{\sqrt{\frac{p_1(1-p_1)}{n}}} \times \dfrac{\sqrt{p_1(1-p_1)}}{\sqrt{p_0(1-p_0)}} + \dfrac{p_1-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right) \sim N\left(\dfrac{p_1-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}, \dfrac{p_1(1-p_1)}{p_0(1-p_0)}\right)$

$$K(p_1) = P(|Z_0| > Z_{\frac{\alpha}{2}} \mid p=p_1) = P(Z_0 > Z_{\frac{\alpha}{2}} \mid p=p_1) + P(Z_0 < -Z_{\frac{\alpha}{2}} \mid p=p_1)$$

先把分母构造好，

$$= P\left(\frac{Z_0 - \frac{p_1-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}}{\sqrt{\frac{p_1(1-p_1)}{p_0(1-p_0)}}} > \frac{Z_{\frac{\alpha}{2}} - \frac{p_1-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}}{\sqrt{\frac{p_1(1-p_1)}{p_0(1-p_0)}}}\right) + P\left(\frac{Z_0 - \frac{p_1-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}}{\sqrt{\frac{p_1(1-p_1)}{p_0(1-p_0)}}} < \frac{-Z_{\frac{\alpha}{2}} - \frac{p_1-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}}{\sqrt{\frac{p_1(1-p_1)}{p_0(1-p_0)}}}\right)$$

后构造分子

$$= \Phi\left(\frac{-Z_{\frac{\alpha}{2}} + \frac{p_1-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}}{\sqrt{\frac{p_1(1-p_1)}{p_0(1-p_0)}}}\right) + \Phi\left(\frac{-Z_{\frac{\alpha}{2}} - \frac{p_1-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}}{\sqrt{\frac{p_1(1-p_1)}{p_0(1-p_0)}}}\right)$$

$$= \Phi\left(\frac{-Z_{\frac{\alpha}{2}}\sqrt{p_0(1-p_0)} + \sqrt{n}(p_1-p_0)}{\sqrt{p_1(1-p_1)}}\right) + \Phi\left(\frac{-Z_{\frac{\alpha}{2}}\sqrt{p_0(1-p_0)} - \sqrt{n}(p_1-p_0)}{\sqrt{p_1(1-p_1)}}\right)$$

Lecture 12

1. Order statistic

① Definition

Let $X_1, \cdots, X_n$ are i.i.d and $X_1, \cdots, X_n \sim f$, then $X_{(k)}$ = k-th smallest of $X_1, \cdots, X_n$

② cdf & pdf

a. Overview

$P(X_{(k)} \le x)$

$$F_{X_{(k)}}(x) = \sum_{l=k}^{n} \binom{n}{l} F(x)^l (1-F(x))^{n-l},$$

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!1!(n-k)!}[F(x)]^{k-1} f(x)[1-F(x)]^{n-k}.$$

≥k个

x

$X_{(k)}$

小于x的至少有k个.

b. Derivation

△ $X_{(n)}$

$F_{X_{(n)}}(x) = P(X_{(n)} \le x) = P(\max\{X_1, \cdots, X_n\} \le x) = P(X_1 \le x, \cdots, X_n \le x) = F(x)^n$

$f_{X_{(n)}}(x) = nF(x)^{n-1} \cdot f(x)$

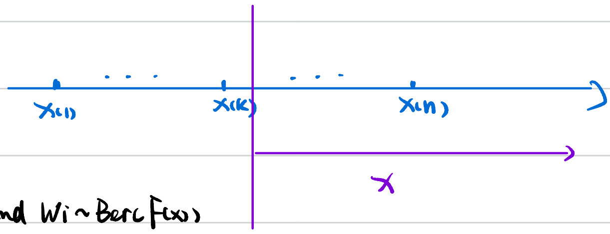
△ $X_{(1)}$

$F_{X_{(1)}}(x) = 1 - P(X_{(1)} > x) = 1 - (1-F(x))^n$

$f_{X_{(1)}}(x) = n(1-F(x))^{n-1} \cdot f(x)$

△ $X_{(k)}$

$F_{X_{(k)}}(x) = P(X_{(k)} \le x)$



Let W denote the # of X_i where $X_i \le x$

then $W = \sum_{i=1}^{n} W_i$, $W \sim Bin(n, F(x))$, where $W_i = I\{X_i \le x\}$ and $W_i \sim Ber(F(x))$

then $F_{X_{(k)}}(x) = P(X_{(k)} \le x) = P(W \ge k) = \sum_{l=k}^{n} P(W=l) = \sum_{l=k}^{n} \binom{n}{l} (F(x))^l (1-F(x))^{n-l} = \sum_{l=k}^{n-1} \binom{n}{l} (F(x))^l (1-F(x))^{n-l} + F(x)^n$

$f_{X_{(k)}}(x) = \sum_{l=k}^{n-1} \binom{n}{l} l (F(x))^{l-1} f(x) (1-F(x))^{n-l} - \sum_{l=k}^{n-1} \binom{n}{l} (F(x))^l (n-l)(1-F(x))^{n-l-1} f(x) + n(F(x))^{n-1} \cdot f(x)$

$= \sum_{l=k}^{n} \binom{n}{l} l f(x) (F(x))^{l-1} (1-F(x))^{n-l} - \sum_{l=k}^{n-1} \binom{n}{l} (F(x))^l (n-l)(1-F(x))^{n-l-1} f(x)$

$= \sum_{l=k}^{n} \binom{n}{l} l f(x) (F(x))^{l-1} (1-F(x))^{n-l} - \underline{\sum_{l'=k+1}^{n} \binom{n}{l'-1} (n-l'+1) f(x) F(x)^{l'-1} (1-F(x))^{n-l'}}$

Let $l' = l+1$

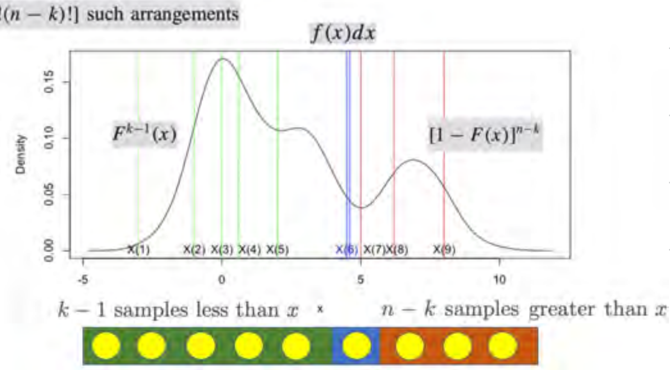
$= \sum_{l=k}^{n} \frac{n!}{(l-1)!(n-l)!} f(x) (F(x))^{l-1} (1-F(x))^{n-l} - \sum_{l'=k+1}^{n} \frac{n!}{(l'-1)!(n-l')!} f(x) F(x)^{l'-1} (1-F(x))^{n-l'}$

$= \frac{n!}{(k-1)!(n-k)!} f(x) (F(x))^{k-1} (1-F(x))^{n-k}$

$= \frac{n!}{(k-1)!1!(n-k)!} (F(x))^{k-1} f(x) (1-F(x))^{n-k}$

pdf of Trinomial Distribution

$n!/[(k-1)!1!(n-k)!]$ such arrangements



③ Example: Uniform distribution

a. pdf

As for Uniform distribution $[0,1]$ $\begin{cases} F(x)=x \\ f(x)=1 \end{cases}$

then $f_{X_{(k)}}(x) = \frac{n!}{(k-1)!1!(n-k)!} x^{k-1}(1-x)^{n-k}$

As for Beta distribution $Beta \sim (\alpha, \beta)$, $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} = \frac{(\alpha+\beta-1)!}{(\alpha-1)!(\beta-1)!} x^{\alpha-1}(1-x)^{\beta-1}$

then $X_{(k)} \sim Beta(k, n-k+1)$

Summary: if $X \sim Uni[0,1]$, then $X_{(k)} \sim Beta(k, n-k+1)$

b. mean & variance

△ mean

i) by Definition

$$E(X_{(k)}) = \int_0^1 x \frac{n!}{(k-1)!1!(n-k)!} x^{k-1}(1-x)^{n-k} dx = \int_0^1 \frac{n!}{(k-1)!1!(n-k)!} x^{k}(1-x)^{n-k} dx = \int_0^1 \frac{n!}{(k-1)!(n-k)!} x^{k+1-1}(1-x)^{n-k+1-1} dx = \frac{n!}{(k-1)!(n-k)!} \cdot \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)}$$
$$= \frac{n!}{(k-1)!(n-k)!} \times \frac{k!(n-k)!}{(n+1)!} = \frac{k}{n+1}$$

ii) by Beta distribution

Since $X_{(k)} \sim Beta(k, n-k+1)$

then $E(X_{(k)}) = \frac{\alpha}{\alpha+\beta} = \frac{k}{n+1}$

$\boxed{\text{联想 Dis 和 pdf}}$

△ variance

i) by Definition

$$E(X_{(k)}^2) = \int_0^1 x^2 \frac{n!}{(k-1)!1!(n-k)!} x^{k-1}(1-x)^{n-k} dx = \int_0^1 \frac{n!}{(k-1)!1!(n-k)!} x^{k+1}(1-x)^{n-k} dx = \frac{n!}{(k-1)!(n-k)!} \int_0^1 x^{k+2-1}(1-x)^{n-k+1-1} dx = \frac{n!}{(k-1)!(n-k)!} \cdot \frac{\Gamma(k+2)\Gamma(n-k+1)}{\Gamma(n+3)}$$
$$= \frac{n!}{(k-1)!(n-k)!} \times \frac{(k+1)!(n-k)!}{(n+2)!} = \frac{k(k+1)}{(n+1)(n+2)}$$

$$\Rightarrow Var(X_{(k)}) = E(X_{(k)}^2) - \left(E(X_{(k)})\right)^2 = \frac{k(k+1)}{(n+1)(n+2)} - \frac{k^2}{(n+1)^2} = \frac{k(k+1)(n+1) - k^2(n+2)}{(n+1)^2(n+2)} = \frac{k(n+1-k)}{(n+1)^2(n+2)}$$

ii) by Beta distribution

Since $X_{(k)} \sim Beta(k, n-k+1)$

then $Var(X_{(k)}) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{k(n+1-k)}{(n+1)^2(n+2)}$

④ Expectation of $F(X_{(k)})$

Let $Y = F(X)$

then $F_Y(y) = P(Y \le y) = P(F(X) \le y) = P(X \le F^{-1}(y)) = F_X(F_X^{-1}(y)) = y$

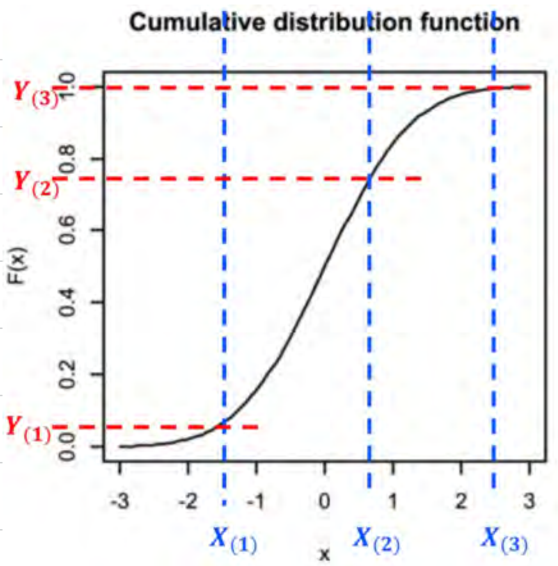
then $Y \sim Uni[0,1]$

Let $Y_i = F(X_i)$

then $F(X_i) \sim Uni[0,1]$ $\qquad Y_i \sim Uni[0,1]$

then $F(X_{(k)}) \sim Beta(k, n-k+1)$ $\qquad Y_{(i)} \sim Beta(k, n-k+1)$

then $E(Y_{(k)}) = E(F(X_{(k)})) = \frac{k}{n+1}$



任意分布的cdf ~ Uni[0,1] !

$\boxed{\text{Order Stat 的 cdf 的期望} = C}$

⇓

Q-Q plot

2. Q-Q plot

① Theoretical quantile vs Sample quantile

a. Theoretical quantile : π_p

$F(\pi_p) = p \Rightarrow \pi_p = F^{-1}(p)$

For standard normal distribution, we have $\pi_p = z_{1-p}$

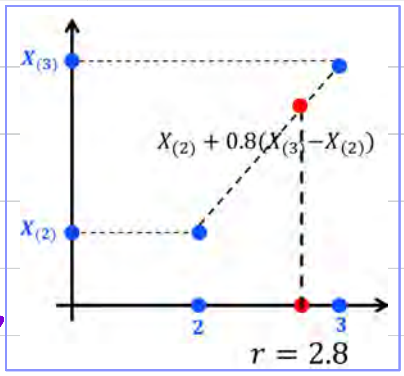
b. Sample quantile: $\hat{\pi}_p = x_{(k)}$

△ Derivation

Since $E(F(X_{(k)})) = \frac{k}{n+1}$, (同时取cdf再构造等式)

$F(\pi_p) = p$

then $X_{(k)}$ is an estimator of π_p for $p = \frac{k}{n+1}$



△ Computation

Let $r = k = (n+1)p$

$$\begin{cases} \text{if } r \text{ is an integer}, & \hat{\pi}_p = x_{(r)} \\ \text{if } r \text{ is not an integer}, & \hat{\pi}_p = x_{(\lfloor r \rfloor)} + (r - \lfloor r \rfloor)(x_{(\lfloor r \rfloor + 1)} - x_{(\lfloor r \rfloor)}) \end{cases}$$

($\hat{\pi}_p = x_{(r^*)}$, where $r^* = \lfloor r \rfloor + (r - \lfloor r \rfloor)$, then we use a linear approximation)

② Quantile-Quantile plot

a. x-axis = Theoretical quantile

$\pi_{\frac{i}{n+1}}$ (set according to our hypothesis)

b. y-axis = Sample quantile

$x_{(i)}$

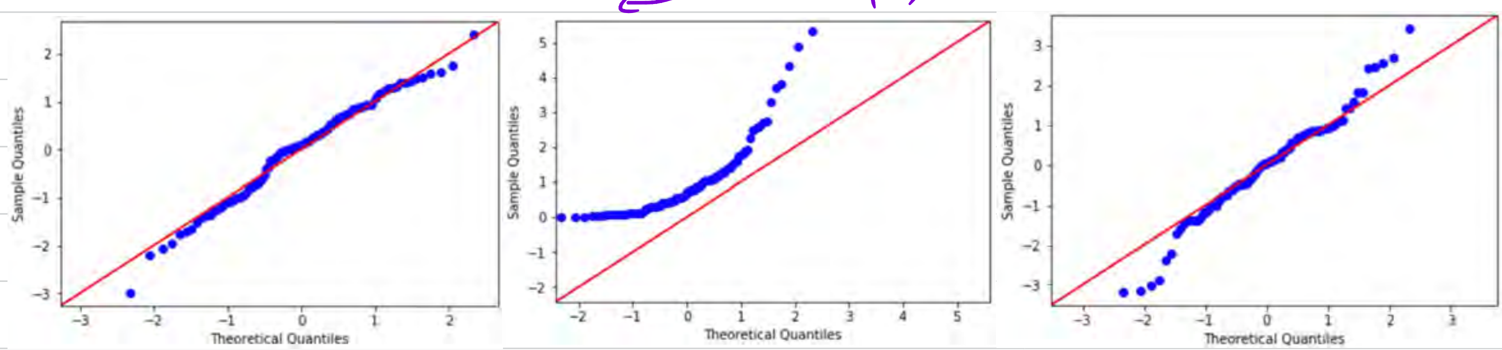
$\rightarrow$ plot $(\pi_{\frac{i}{n+1}}, x_{(i)}) \rightarrow$

Good fit: The pairs are close to $y=x$

Poor fit: The pairs are not close to $y=x$

注意是以k/i为单位，而不是p（一个点一个点…）

examples



✓ ✗ ✗

Draw a vertical line, then we can observe the discrepancy between the x-values from Hypothesis-Distribution and Real Distribution for the same p.

Lecture 15–16

1. Distribution-free CI for Percentiles

① "Distribution-free"

- No assuming normal distribution — if normal, then Mean = Median
- No requirement on the sample size

② Fundamentally

$x_{(k)}$ is a point estimator for π_p where $p = \frac{k}{n+1}$.

then if we want to construct CI for π_p, we just need to construct CI for $x_{(k)}$ (no need to be symmetric)

③ CI for $\pi_{\frac{1}{2}}$

when n is small (≤ 20) — Binomial

a. CI: $(x_{(i)}, x_{(j)})$ $\hat{=}$ confidence coefficient

b. Confidence level of $(x_{(i)}, x_{(j)})$:

Let $W = \sum \mathbb{1}\{X_i \le \pi_{\frac{1}{2}}\}$, then $W \sim Bin(n, \frac{1}{2})$

$$P(x_{(i)} < \pi_{\frac{1}{2}} < x_{(j)}) = P(i \le W \le j-1)$$
$$= \sum_{l=i}^{j-1} \binom{n}{l} (1-\tfrac{1}{2})^l (\tfrac{1}{2})^{n-l}$$
$$= \sum_{l=i}^{j-1} \binom{n}{l} (\tfrac{1}{2})^n$$

$i-1$, $x_{(i)}$, $j-i-1$, $x_{(j)}$, $n-j$

$\pi_{\frac{1}{2}}$

A useful table to find CI for $\pi_{\frac{1}{2}}$ according to n and significance level. (由P找区间)

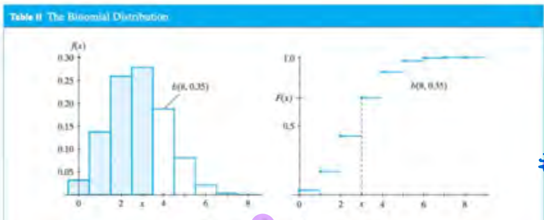
Table 7.5-1 Information for confidence intervals for m

n	$(i, n+1-i)$	$P(Y_i < m < Y_{n+1-i})$	n	$(i, n+1-i)$	$P(Y_i < m < Y_{n+1-i})$
5	(1,5)	0.9376	13	(3,11)	0.9776
6	(1,6)	0.9688	14	(4,11)	0.9426
7	(1,7)	0.9844	15	(4,12)	0.9648
8	(2,7)	0.9296	16	(5,12)	0.9232
9	(2,8)	0.9610	17	(5,13)	0.9510
10	(2,9)	0.9786	18	(5,14)	0.9692
11	(3,9)	0.9346	19	(6,14)	0.9364
12	(3,10)	0.9614	20	(6,15)	0.9586

i j $x_{(i)}$ $x_{(j)}$

close to 95%

Table II The Binomial Distribution



$$F(x) = P(X \le x) = \sum_{k=0}^{x} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

n	x	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
		p									
2	0	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
	1	0.9975	0.9900	0.9775	0.9600	0.9375	0.9100	0.8775	0.8400	0.7975	0.7500
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
	1	0.9928	0.9720	0.9392	0.8960	0.8438	0.7840	0.7182	0.6480	0.5748	0.5000
	2	0.9999	0.9990	0.9966	0.9920	0.9844	0.9730	0.9571	0.9360	0.9089	0.8750
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
	1	0.9860	0.9477	0.8905	0.8192	0.7383	0.6517	0.5630	0.4752	0.3910	0.3125
	2	0.9995	0.9963	0.9880	0.9728	0.9492	0.9163	0.8735	0.8208	0.7585	0.6875
	3	1.0000	0.9999	0.9995	0.9984	0.9961	0.9919	0.9850	0.9744	0.9590	0.9375
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

$W \sim Bin(n, p)$

$\Rightarrow n - W \sim Bin(n, 1-p)$

使得其 ≤ 0.5

when n is large (> 20) $\xrightarrow{CLT}$ Normal

a. CI: $(x_{(i)}, x_{(j)})$

b. Confidence level of $(x_{(i)}, x_{(j)})$:

$$P(x_{(i)} < \pi_{\frac{1}{2}} < x_{(j)}) = P(i \le W \le j-1) \quad \text{Discrete}$$
$$= P(i - \tfrac{1}{2} \le W \le j - \tfrac{1}{2}) \quad \text{Continuous}$$

往外延伸

Half-unit correction

$$= P\left(\frac{i-\frac{1}{2}-\frac{n}{2}}{\sqrt{\frac{n}{4}}} \le \frac{W-\frac{n}{2}}{\sqrt{\frac{n}{4}}} \le \frac{j-\frac{1}{2}-\frac{n}{2}}{\sqrt{\frac{n}{4}}}\right)$$

$W \sim Bin(n, \frac{1}{2}) \xrightarrow{CLT} W \sim N(\frac{n}{2}, \frac{n}{4})$

$$\approx \Phi\left(\frac{j-\frac{1}{2}-\frac{n}{2}}{\sqrt{\frac{n}{4}}}\right) - \Phi\left(\frac{i-\frac{1}{2}-\frac{n}{2}}{\sqrt{\frac{n}{4}}}\right)$$

非常小心CLT的使用！(n>20)

④ CI for π_p

when n is small ≤ 20	when n is large > 20
a. CI: $[X_{(i)}, X_{(j)})$	a. CI: $[X_{(i)}, X_{(j)})$
b. Confidence level of $(X_{(i)}, X_{(j)})$:	b. Confidence level of $(X_{(i)}, X_{(j)})$
$P(X_{(i)} < \pi_p < X_{(j)}) = P(i \le W \le j-1)$	$P(X_{(i)} < \pi_p < X_{(j)}) = P(i \le W \le j-1)$
$= \sum_{l=i}^{j-1} \binom{n}{l} p^l (1-p)^{n-l}$	$= P(i-\frac{1}{2} \le W \le j-\frac{1}{2})$
~~查表~~ 算出来	$W \sim Bin(n,p) \xrightarrow{CLT} W \sim N(np, np(1-p))$
	$= P(\frac{i-\frac{1}{2}-np}{\sqrt{np(1-p)}} \le \frac{W-np}{\sqrt{np(1-p)}} \le \frac{j-\frac{1}{2}-np}{\sqrt{np(1-p)}})$
	$= \Phi(\frac{j-\frac{1}{2}-np}{\sqrt{np(1-p)}}) - \Phi(\frac{i-\frac{1}{2}-np}{\sqrt{np(1-p)}})$

2. Distribution-free Hypothesis test for Percentiles

① Sign Test

Let m be π_p.

$$\begin{cases} H_0: m = m_0 \\ H_1: \begin{cases} \text{Two-sided}: m \ne m_0 \\ \text{One-sided}: m > m_0 \mid m < m_0 \end{cases} \end{cases}$$

$W = \sum I\{X_i \le m_0\}$, $W \sim Bin(n, p)$

Next, we set two-sided H_1 as the example, to introduce two sub-methods in Sign Test

a. Critical Region

△ Find CR according to the def of α → 查表 大表 (当n>20时，无法查表，此时可用CLT转化)

△ Make judgement according to CR

b. p-value

△ Compute p-value

△ Compare p-value & α

c. CI

△ Find CI according to α

△ Judge.

(CI不属于ST or WT)

Take the difference between real data & assumed median into consideration

② Wilcoxon Test

Let m be $\pi_{\frac{1}{2}}$ (in Wilcoxon Test, our test parameter can be just the median)

$H_0: m = m_0$

H_1: Two-sided: $m \neq m_0$

One-sided: $m > m_0$ | $m < m_0$

$W = \sum$ signed rank, $\dfrac{W-0}{\sqrt{\frac{n(n+1)(2n+1)}{6}}} \sim N(0,1)$

CLT so powerful!

Known → Normal

Unknown → Normal

Step 1: Find ranks

Step 2: Find signed ranks

Step 3: Compute W

$\mu = 0$ (Since we assumed mean = median)

$\sigma^2 = \sum_{k=1}^{n} k^2$

☆ Adjustment:

If $X = m_0$, rank $= 0$ (n 要跟着减少)

If $|X_i - m_0| = |X_j - m_0|$, $R_i = R_j = \frac{R_i + R_j}{2}$

a. Critical Region

Obviously

b. p-value

△ Compute p-value: $2P(W \geq w_0)$ We need to do half-unit correction, e.g.: $P(W \geq 29) = P(W \geq 28)$

$P(W \leq 29) = P(W \leq 30)$

Since W always differ by 2, then the half unit is 1

△ Compare p-value & α

(Whole-unit) continuity correction:

$$P(W \geq 25; H_0) = 1 - P(W \leq 23; H_0) \approx 1 - \Phi\left(\frac{24}{\sqrt{385}}\right) = 0.11,$$

Lecture 17–18

1. Chi-square Goodness-of-fit test

单次实验

① Multinomial – Chisqu Theorem

If $(Y_1, \cdots Y_K) \sim \text{Multi}(n, p_1, \cdots, p_K)$, then $\sum_{i=1}^{K} \frac{(Y_i - np_i)^2}{np_i} \sim \chi^2(\underline{K-1})$ as $n \to \infty$

p_K 可通过$(1-p_1-p_2 \cdots)$求得, 故自由度 = # 可自由变量 = $K-1$

$\Rightarrow \chi^2 = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(K-1)$, where $\begin{cases} O_i: \text{Observed frequency} \\ E_i: \text{Expected frequency} \end{cases}$

② Process

若是自己分, 则令E_i的等一些

分组 $\Rightarrow$ Data: $Y_1, \cdots, Y_n \sim \text{Multi}(n, p_1, \cdots, p_K)$

Hypothesis: $\begin{cases} H_0: p_i = p_{i0}, \text{ for } i = 1, 2, 3, \cdots, n \\ H_1: H_0 \text{ is not true} \end{cases}$ (Good fit)

Chi-squ test stat: $\chi^2 = \sum_{i=1}^{K} \frac{(Y_i - np_{i0})^2}{np_{i0}}$, $\underline{\chi^2 \sim \chi^2(K-1)}$

To be further discussed as follows

$\Rightarrow \begin{cases} \text{Critical region}: & \chi^2 > \chi^2_\alpha(K-1) \\ \text{p-value}: & P(\chi^2 > \chi^2_0) \end{cases}$

a. When parameter is specified: Expected frequency is determined

b. When parameter is not specified: Expected frequency is determined by parameter

$\underline{\text{Two ways to estimate parameter}}$ — $\begin{cases} \text{Minimum chi-squ estimate} \\ \text{Maximum likelihood estimate} \checkmark \end{cases}$

△ Minimum chi-squ estimate

e.g.

Goodness-of-fit test (with unspecified parameter)

Example

Data

$X_1, \ldots, X_n \sim \text{Bin}(4, p)$

p unknown

Null Hypothesis

$H_0: p_i = p_{i0}, \quad i = 1, 2, \ldots, k,$

$p_{i0} = P(A_i) = \frac{4!}{(i-1)!(5-i)!} p^{i-1}(1-p)^{5-i}, i = 1, 2, \ldots, 5$

p_{0i} also unknown (depend on *p*)

	observed	expected
$A_1 = \{0\}$	7	$100(1-p)^4$
$A_2 = \{1\}$	18	$100 * 4p(1-p)^3$
$A_3 = \{2\}$	40	$100 * 6p^2(1-p)^2$
$A_4 = \{3\}$	31	$100 * 4p^3(1-p)^1$
$A_5 = \{4\}$	4	$100p^4$

Estimate *p* from data and plug-in

Goodness-of-fit test (with unspecified parameter)

Estimate *p*

	observed	expected
$A_1 = \{0\}$	7	$100(1-p)^4$
$A_2 = \{1\}$	18	$100 * 4p(1-p)^3$
$A_3 = \{2\}$	40	$100 * 6p^2(1-p)^2$
$A_4 = \{3\}$	31	$100 * 4p^3(1-p)^1$
$A_5 = \{4\}$	4	$100p^4$

(depend on *p*)

$$\chi^2(p) = \frac{(y_1 - 100(1-p)^4)^2}{100(1-p)^4} + \frac{(y_2 - 100 * 4p(1-p)^3)^2}{100 * 4p(1-p)^3} + \frac{(y_3 - 100 * 6p^2(1-p)^2)^2}{100 * 6p^2(1-p)^2} + \frac{(y_4 - 100 * 4p^3(1-p)^1)^2}{100 * 4p^3(1-p)^1} + \frac{(y_5 - 100p^4)^2}{100p^4}$$

Estimate *p* as the $\tilde{p}$ that minimizes $\chi^2(p)$

Called the minimum chi-square estimator

☆ 因为我们用$p_1, p_2, \cdots, p_n$去构造$\tilde{p}$, 所以我们要进一步减去一个自由度, 故而这里 $\chi^2(\tilde{p}) \sim \chi^2(5-1-1)$

$\hookrightarrow \begin{cases} K \text{ categories} \\ 1 \text{ dependent } p_i \\ d \text{ ~~unknown~~ para}_s \text{ estimated} \end{cases} \rightarrow \chi^2 \sim \chi^2(K-1-d)$

△ Maximum likelihood estimate

$p = \hat{p} \Rightarrow \chi^2$

Sometimes when E_i is too small, χ^2 will not be a accurate test stat, hence we will combine some small groups into a group. (Be careful for the # of the categories) (通常令 $E_i > 5$)

Testing Poisson

The mean of the assumed Poisson distribution in this example is unknown and must be estimated from the sample data. The estimate of the mean number of defects per board is the sample average, that is, $(32\cdot0 + 15\cdot1 + 9\cdot2 + 4\cdot3)/60 = 0.75$. From the Poisson distribution with parameter 0.75, we may compute p_i, the theoretical, hypothesized probability associated with the ith class interval. Since each class interval corresponds to a particular number of defects, we may find the p_i as follows:

$$p_1 = P(X = 0) = \frac{e^{-0.75}(0.75)^0}{0!} = 0.472$$

$$p_2 = P(X = 1) = \frac{e^{-0.75}(0.75)^1}{1!} = 0.354$$

$$p_3 = P(X = 2) = \frac{e^{-0.75}(0.75)^2}{2!} = 0.133$$

$$p_4 = P(X \geq 3) = 1 - (p_1 + p_2 + p_3) = 0.041$$

The expected frequencies are computed by multiplying the sample size $n = 60$ times the probabilities p_i. That is, $E_i = np_i$. The expected frequencies follow:

Number of Defects	Probability	Expected Frequency
0	0.472	28.32
1	0.354	21.24
2	0.133	7.98
3 (or more)	0.041	2.46

Since the expected frequency in the last cell is less than 3, we combine the last two cells:

Number of Defects	Observed Frequency	Expected Frequency
0	32	28.32
1	15	21.24
2 (or more)	13	10.44

2. Chi-square test for Homogeneity 多次实验是否结果类似

① Process

	A_1	A_2	A_3	$\cdots$	A_k
First experiment	$p_{1,1}$	$p_{1,2}$	$p_{1,3}$	$\cdots$	$p_{1,k}$
Second experiment	$p_{2,1}$	$p_{2,2}$	$p_{2,3}$	$\cdots$	$p_{2,k}$
……	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
h-th experiment	$p_{h,1}$	$p_{h,2}$	$p_{h,3}$	$\cdots$	$p_{h,k}$

Data: $Y_{11}, \dots, Y_{1k} \sim \text{Multi}(n, p_1, \dots, p_k)$

$\dots$

$Y_{h1}, \dots, Y_{hk} \sim \text{Multi}(n, p_1, \dots, p_k)$

Hypothesis: $\begin{cases} H_0: p_{11} = p_{21} = \dots = p_{h1} = p_1, \dots \\ H_1: H_0 \text{ is not true} \end{cases}$ 我们想要检测不同次实验下的多项分布是否一致

Test Stat: When p_{ij} are known: $\chi^2 = \sum_{i=1}^{h} \sum_{j=1}^{k} \frac{(y_{ij} - n_i p_{ij})^2}{n_i p_{ij}}$, $\chi^2 \sim \chi^2(\underbrace{h(k-1)}_{})$ 有h行，每行都有k-1个自由变量

When we need to estimate $\hat{p}_{ij} = \hat{p}_j = \underbrace{\frac{Y_{1j} + \dots + Y_{hj}}{h}}_{\text{用样本数据估计}}$: $\chi^2 = \sum_{i=1}^{h} \sum_{j=1}^{k} \frac{(y_{ij} - n_i \hat{p}_j)^2}{n_i \hat{p}_j}$, $\chi^2 \sim \chi^2(\underbrace{(h-1)(k-1)}_{})$

实际损失了(k-1)个estimated params，故减(k-1)个自由度

$\Rightarrow \begin{cases} \text{Critical region} \\ \text{p-value} \end{cases}$

To better compute Test Stat, we use the Contingency Table

	Grade					
	A	B	C	D	F	Total
Group I	8 (6)	13 (11)	16 (15)	10 (13)	3 (5)	50
Group II	4 (6)	9 (11)	14 (15)	16 (13)	7 (5)	50
Total	12	22	30	26	10	100

expected: $50 \times \hat{p}_A = 50 \times 0.12 = 6$

$50 \times \hat{p}_B = 11$

$\vdots$

$50 \times \hat{p}_F = 5$

(i) Compute expected frequency

(ii) Compute observed frequency

(iii) Compute test Stat and do HT

小技巧：当自己分组时，令 $\hat{p}_1 = \hat{p}_2 = \dots = \hat{p}_k = \frac{1}{k}$ 可以有效减小计算量

	Categories				
	A_1	A_2	A_3	A_4	Total
U	2 (5)	4 (5)	4 (5)	10 (5)	20
V	8 (5)	6 (5)	6 (5)	0 (5)	20
Total	10	10	10	10	
	$\hat{p}_{A_1} = \frac{1}{4}$	$\hat{p}_{A_2} = \frac{1}{4}$	$\hat{p}_{A_3} = \frac{1}{4}$	$\hat{p}_{A_4} = \frac{1}{4}$	

3. Chi-square test for Independence 多次类别是否独立

① Process

Data: $Y_{1,1} \cdots Y_{1,k}, \cdots, Y_{n,k} \sim \text{Multi}(n, p_{11}, p_{12}, \cdots, p_{nk})$

Hypothesis: $\begin{cases} H_0: \text{Two attributes are ind, i.e. } p_{ij} = p_{i\cdot} \times p_{\cdot j} \\ H_1: \text{Two attributes are dep.} \end{cases}$

$p_{i\cdot}$ & $p_{\cdot j}$ represent marginal distribution for A & B

$\begin{cases} p_{i\cdot} = \sum_{j=1}^{k} p_{ij} \\ p_{\cdot j} = \sum_{i=1}^{n} p_{ij} \end{cases}$

Test Stat: When $p_{i\cdot}$ & $p_{\cdot j}$ are known: $\chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{k} \frac{(Y_{ij} - np_{i\cdot}p_{\cdot j})^2}{np_{i\cdot}p_{\cdot j}} \sim \chi^2(nk-1)$

When we need to estimate $p_{i\cdot}$ and $p_{\cdot j}$ $\begin{cases} \hat{p}_{i\cdot} = \frac{\sum_{j=1}^{k} Y_{ij}}{n} \\ \hat{p}_{\cdot j} = \frac{\sum_{i=1}^{n} Y_{ij}}{n} \end{cases}$: $\chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{k} \frac{(Y_{ij} - n\hat{p}_{i\cdot}\hat{p}_{\cdot j})^2}{n\hat{p}_{i\cdot}\hat{p}_{\cdot j}} \sim \chi^2((n-1)(k-1))$

$nk-1-(n-1)-(k-1) = (n-1)(k-1)$

$\Rightarrow \begin{cases} \text{Critical region} \\ \text{p-value} \end{cases}$

Lecture 20-21

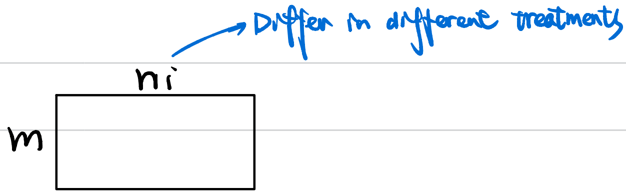
1. One-factor ANOVA

① Basics

Table 13-1 Tensile Strength of Paper (psi)

Hardwood Concentration (%)	Observations 1	2	3	4	5	6	Totals	Averages
5	7	8	15	11	9	10	60	10.00
10	12	17	13	18	19	15	94	15.67
15	14	18	19	17	16	18	102	17.00
20	19	25	22	23	18	20	127	21.17
							383	15.96

Observation



Treatment: levels of factors ("Treatment" 只针对单Factor)

② Some Notations

a. Sample mean of i^{th} treatment : $\bar{x}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$, $i=1,2,\cdots,m$

b. Grand mean : $\bar{x}_{\cdot\cdot} = \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij} = \frac{1}{n}\sum_{i=1}^{m} n_i\bar{x}_{i\cdot}$

c. Total sum-of-squares : $SS(TO) = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{\cdot\cdot})^2$

d. Between-treatment sum-of-squares: $SS(T) = \sum_{i=1}^{m} n_i(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})^2$ (Weighted sum)

e. Error sum-of-squares : $SS(E) = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i\cdot})^2$

f. Within-treatment sum-of-squares : $SS(E_i) = \sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i\cdot})^2$

$\underbrace{x_{ij}}_{\text{真实}} - \underbrace{(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})-\bar{x}_{\cdot\cdot}}_{\text{估计}}$

$SS(E) = \sum_{i=1}^{m} SS(E_i)$

$SS(TO) = SS(E) + SS(T)$ (※)

总SS = 行内总SS + 行间总SS

proof:

$$SS(TO) = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{\cdot\cdot})^2$$
$$= \sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i\cdot}+\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})^2$$
$$= \sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i\cdot})^2 + \sum_{i=1}^{m}\sum_{j=1}^{n_i}(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})^2 + 2\sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i\cdot})(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})$$
$$= \sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i\cdot})^2 + \sum_{i=1}^{m} n_i(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})^2$$
$$= SS(E) + SS(T)$$

③ Process

Assumption: in each i^{th} treatment, observations $\sim N(\mu_i, \sigma^2)$

Hypothesis: $\begin{cases} H_0: \mu_1=\mu_2=\cdots=\mu_m=\mu \\ H_1: H_0 \text{ is not true} \end{cases}$ treatment 对数据没有影响

Test Stat: $F = \frac{SS(T)/(m-1)}{SS(E)/(n-m)} = \frac{[SS(T)/\sigma^2]/(m-1)}{[SS(E)/\sigma^2]/(n-m)} \sim F(m-1, n-m)$

区分SS & S !

Mean square: $\begin{cases} MS(T) = \frac{SS(T)}{m-1} \\ MS(E) = \frac{SS(E)}{n-m} \end{cases}$

n 这里是所有的个数

$\Rightarrow \begin{cases} \text{Critical region} : F > F_\alpha(m-1, n-m) \\ \text{p-value} : P(F(m-1,n-m) > F_0) \end{cases}$

2. Two-factor ANOVA

"Factor" ≜ "Treatment"

b

Factor 2 (4 levels)

Car	Gasoline 1	2	3	4	$\bar{X}_{i\cdot}$
1	26	28	31	31	29
2	24	25	28	27	26
3	25	25	28	26	26
$\bar{X}_{\cdot j}$	25	26	29	28	27

Factor 1 (3 levels)

a

① Some Notations

a. Sample mean of the i-th row factor: $\bar{x}_{i\cdot} = \frac{1}{b}\sum_{j=1}^{b} x_{ij}$

b. Sample mean of the j-th col factor: $\bar{x}_{\cdot j} = \frac{1}{a}\sum_{i=1}^{a} x_{ij}$

c. Overall sample mean: $\bar{x}_{\cdot\cdot} = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b} x_{ij}$

d. Total sum-of-square: $SS(TO) = \sum_{i=1}^{a}\sum_{j=1}^{b} (x_{ij} - \bar{x}_{\cdot\cdot})^2$ $\quad ab-1$

e. Sum-of-squares of Factor A: $SS(A) = b\sum_{i=1}^{a} (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2$ $\quad a-1$

f. Sum-of-squares of Factor B: $SS(B) = a\sum_{j=1}^{b} (\bar{x}_{\cdot j} - \bar{x}_{\cdot\cdot})^2$ $\quad b-1$

g. Error sum-of-squares: $SS(E) = \sum_{i=1}^{a}\sum_{j=1}^{b} [\underbrace{x_{ij}}_{u_{ij}} - \underbrace{(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})}_{\hat{\alpha}_i} - \underbrace{(\bar{x}_{\cdot j} - \bar{x}_{\cdot\cdot})}_{\hat{\beta}_j} - \underbrace{\bar{x}_{\cdot\cdot}}_{\hat{u}}]^2$ $\quad (a-1)(b-1)$

$SS(TO) = SS(A) + SS(B) + SS(E)$

算TO 算A 算B 算E

② Process

Assumption: For the (i,j)-th combination, data are sampled from a distribution $x_{ij} \sim N(u_{ij}, \sigma^2)$,

where $u_{ij} = u + \underbrace{\alpha_i}_{i^{th}\text{ row effect}} + \underbrace{\beta_j}_{j^{th}\text{ col effect}}$ $\quad \left(\sum_{i=1}^{a}\alpha_i = 0,\ \sum_{j=1}^{b}\beta_j = 0\right)$ 选取u的原则

即使是 $\sum_{i=1}^{a}\alpha_i = C$, $\sum_{j=1}^{b}\beta_j = C$, 我们也可将其化为如下: $u_{ij} = (u + \bar{\alpha} + \bar{\beta}) + (\alpha_i - \bar{\alpha}) + (\beta_j - \bar{\beta})$

Hypothesis:

- Is u_{ij} affected by row? (factor A)
  - $H_{0A}: \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$
  - H_{1A}: H_{0A} is not true
- Is u_{ij} affected by col? (factor B)
  - $H_{0B}: \beta_1 = \beta_2 = \cdots = \beta_b = 0$
  - H_{1B}: H_{0B} is not true

Test Stat:

$$F_A = \frac{SS(A)/(a-1)}{SS(E)/[(a-1)(b-1)]} \Rightarrow$$

- Critical region: $F \geq F_\alpha(a-1, (a-1)(b-1))$
- p-value: $P(F_\alpha(a-1, (a-1)(b-1)) \geq F_A)$

$$F_B = \frac{SS(B)/(b-1)}{SS(E)/[(a-1)(b-1)]} \Rightarrow$$

- Critical region: $F \geq F_\alpha(b-1, (a-1)(b-1))$
- p-value: $P(F_\alpha(b-1, (a-1)(b-1)) \geq F_B)$

(建议:用间接法算SS(E))

3. Two-factor ANOVA with $c>1$ observations per cell

	Brand of Gasoline (Factor 2, b=4 levels)							
Car	1		2		3		4	
1	31.0	24.9	26.3	30.0	25.8	29.4	27.8	27.3
	26.2	28.8	25.2	31.6	24.5	24.8	28.2	30.4
2	30.6	29.5	25.5	26.8	26.6	23.7	28.1	27.1
	30.8	28.9	27.4	29.4	28.2	26.1	31.5	29.1
3	24.2	23.1	27.4	28.1	25.2	26.7	26.3	26.4
	26.8	27.4	26.4	26.9	27.7	28.1	27.9	28.8

Factor 1 (a=3 levels)

① Some Notations

a. Sample mean of the (i,j)th interaction effect : $\bar{x}_{ij\cdot} = \frac{1}{c}\sum_{k=1}^{c} x_{ijk}$

b. Sample mean of the i-th row factor : $\bar{x}_{i\cdot\cdot} = \frac{1}{bc}\sum_{j=1}^{b}\sum_{k=1}^{c} x_{ijk}$

c. Sample mean of the j-th col factor : $\bar{x}_{\cdot j\cdot} = \frac{1}{ac}\sum_{i=1}^{a}\sum_{k=1}^{c} x_{ijk}$

d. Overall sample mean : $\bar{x}_{\cdot\cdot\cdot} = \frac{1}{abc}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} x_{ijk}$

e. Total sum-of-square : $SS(TO) = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(x_{ijk}-\bar{x}_{\cdot\cdot\cdot})^2$ $\quad abc-1$

f. Sum-of-squares of Factor A : $SS(A) = bc\sum_{i=1}^{a}(\bar{x}_{i\cdot\cdot}-\bar{x}_{\cdot\cdot\cdot})^2$ $\quad a-1$

g. Sum-of-squares of Factor B : $SS(B) = ac\sum_{j=1}^{b}(\bar{x}_{\cdot j\cdot}-\bar{x}_{\cdot\cdot\cdot})^2$ $\quad b-1$

h. Sum-of-squares of interaction : $SS(AB) = c\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{ij\cdot}-\bar{x}_{i\cdot\cdot}-\bar{x}_{\cdot j\cdot}+\bar{x}_{\cdot\cdot\cdot})^2$ $\quad (a-1)(b-1)$

i. Error sum-of-squares : $SS(E) = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(x_{ijk}-\bar{x}_{ij\cdot})^2$ $\quad ab(c-1)$

$SS(TO) = SS(A) + SS(B) + SS(AB) + SS(E)$

$df_{TO} \; df_{A} \; df_{B} \; df_{AB} \; df_{E}$

② Process

Assumption: For the (i,j)-th combination, c observations are sampled from a distribution $X_{ijk} \sim N(\mu_{ij}, \sigma^2)$, $k=1,\cdots,c$

where $\mu_{ij} = \mu + \alpha_i + \beta_j + \underbrace{\gamma_{ij}}_{(i,j)^{th}\text{ interaction effect}}$ $\left(\sum_{i=1}^{a}\alpha_i = 0,\ \sum_{j=1}^{b}\beta_j = 0,\ \sum_{i=1}^{a}\gamma_{ij} = 0,\ \sum_{j=1}^{b}\gamma_{ij} = 0\right)$

Hypothesis:

- Is μ_{ij} affected by row?
  - $H_{0A}: \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$
  - H_{1A}: H_{0A} is not true
- Is μ_{ij} affected by col?
  - $H_{0B}: \beta_1 = \beta_2 = \cdots = \beta_b = 0$
  - H_{1B}: H_{0B} is not true
- Is μ_{ij} affected by interaction between row & col?
  - $H_{0AB}: \gamma_{ij} = 0$, for all i,j
  - H_{1AB}: H_{0AB} is not true

Test Stat: $F_A = \frac{SS(A)/(a-1)}{SS(E)/[ab(c-1)]} \Rightarrow$
- Critical region: $F \geq F_\alpha(a-1, ab(c-1))$
- p-value : $P(F_\alpha(a-1, ab(c-1)) \geq F_A)$

$F_B = \frac{SS(B)/(b-1)}{SS(E)/[ab(c-1)]} \Rightarrow$
- Critical region: $F \geq F_\alpha(b-1, ab(c-1))$
- p-value : $P(F_\alpha(b-1, ab(c-1)) \geq F_B)$

$F_{AB} = \frac{SS(AB)/[(a-1)(b-1)]}{SS(E)/[ab(c-1)]} \Rightarrow$
- Critical region: $F \geq F_\alpha((a-1)(b-1), ab(c-1))$
- p-value : $P(F_\alpha((a-1)(b-1), ab(c-1)) \geq F_{AB})$

Lecture 22–24

1. Basics of Simple Linear Regression

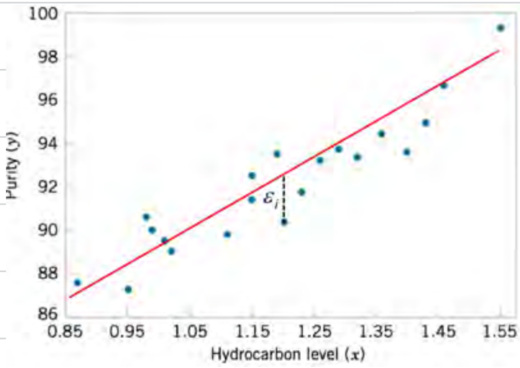
① Definition

Simple linear regression model:

$\boxed{Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i}$, $i = 1, \cdots, n$

- X_i: Regressor or Predictor
- Y_i: Response
- β_0: Intercept
- β_1: Slope
- ε_i: Random error

β_0, β_1: Regression coefficient



② Distribution of $Y|x$

x_i is given $\Rightarrow (\beta_0 + \beta_1 X_i) = c$

$\varepsilon_i \sim N(0, \sigma^2)$

$\Rightarrow Y_i | x_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2) \Rightarrow \underline{Y|x \sim N(\beta_0 + \beta_1 x, \sigma^2)}$

2. Point estimation of Simple Linear Regression (β_0, β_1)

① Least square estimates

Idea: Minimizing $\sum_{i=1}^{n} \varepsilon_i^2$

$L = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$

$$\Rightarrow \begin{cases} \left.\dfrac{\partial L}{\partial \beta_0}\right|_{\hat\beta_0, \hat\beta_1} = -2\sum_{i=1}^{n}(y_i - \hat\beta_0 - \hat\beta_1 x_i) = 0 \\ \left.\dfrac{\partial L}{\partial \beta_1}\right|_{\hat\beta_0, \hat\beta_1} = -2\sum_{i=1}^{n}(y_i - \hat\beta_0 - \hat\beta_1 x_i)x_i = 0 \end{cases}$$

$$\Rightarrow \begin{cases} \hat\beta_0 = \bar y - \hat\beta_1 \bar x \\ \hat\beta_1 = \dfrac{S_{xy}}{S_{xx}} = \dfrac{\sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}} \end{cases}$$

乘积和 - 和的乘积

$S_{xx} = n\,\mathrm{Var}(x)$ (Just a Notation)

$\Rightarrow \hat y_i = \hat\beta_0 + \hat\beta_1 x_i$, $e_i = y_i - \hat y_i$ is called "Residual"

计算步骤:

计算 S_{xy} & S_{xx}

计算 $\hat\beta_1$

计算 $\hat\beta_0$

计算 $\hat y_i = \hat\beta_0 + \hat\beta_1 x_i$

② MLE

Since $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$, $i = 1, \cdots, n$,

Then $L(\beta_0, \beta_1) = \left(\dfrac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\sum(y_i - \beta_0 - \beta_1 x_i)^2/(2\sigma^2)}$

$l(\beta_0, \beta_1) = -n\ln(\sqrt{2\pi\sigma^2}) - \sum(y_i - \beta_0 - \beta_1 x_i)^2/(2\sigma^2)$

$$\Rightarrow \begin{cases} \left.\dfrac{\partial l(\beta_0, \beta_1)}{\partial \beta_0}\right|_{\hat\beta_0, \hat\beta_1} = \sum(y_i - \hat\beta_0 - \hat\beta_1 x_i)/\sigma^2 = 0 \\ \left.\dfrac{\partial l(\beta_0, \beta_1)}{\partial \beta_1}\right|_{\hat\beta_0, \hat\beta_1} = \sum x_i(y_i - \hat\beta_0 - \hat\beta_1 x_i)/\sigma^2 = 0 \\ \left.\dfrac{\partial l(\beta_0, \beta_1, \sigma^2)}{\partial \sigma^2}\right|_{\hat\beta_0, \hat\beta_1, \hat\sigma^2} = -\dfrac{n}{2}\dfrac{1}{\hat\sigma^2} + \sum(y_i - \hat\beta_0 - \hat\beta_1 x_i)^2/(2\hat\sigma^4) = 0 \end{cases}$$

$$\Rightarrow \begin{cases} \hat\beta_0 = \bar y - \hat\beta_1 \bar x \\ \hat\beta_1 = \dfrac{S_{xy}}{S_{xx}} = \dfrac{\sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}} \end{cases}$$

(The same as LSE)

$\Rightarrow \hat\sigma^2_{MLE} = \dfrac{1}{n}\sum_{i=1}^{n}(y_i - \hat\beta_0 - \hat\beta_1 x_i)^2$ (Biased)

应该是n-2

③ Unbiased estimate of σ^2

a. Notations

$$\begin{cases} S_{xx}=\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}x_i^2-\frac{(\sum_{i=1}^{n}x_i)^2}{n} \quad (n-1)S_x^2 \\ S_{xy}=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{n}x_iy_i-\frac{\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n} \\ S_{yy}=SS(T)=\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}y_i^2-\frac{(\sum_{i=1}^{n}y_i)^2}{n} \\ SS(E)=\sum_{i=1}^{n}e_i^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \end{cases}$$

(这里是 $\hat{y}$ 而不是 $\bar{y}$)

b. Derivation

Since $SS(E)=\sum_{i=1}^{n}e_i^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)^2$,

and $\frac{\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)^2}{\sigma^2}\sim\chi^2(n-2)$,

then $E(SS(E))=(n-2)\sigma^2$

then $\hat{\sigma}^2=\frac{SS(E)}{n-2}$

SS(E)的计算:

1) 根据定义式计算

2) 计算式 (preferred)

$$\begin{aligned} SS(E)&=\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)^2=\sum_{i=1}^{n}(y_i-\bar{y}+\bar{y}-\hat{\beta}_0-\hat{\beta}_1x_i)^2 \\ &=\sum_{i=1}^{n}(y_i-\bar{y}+\bar{y}-\sum_{i=1}^{n}y_i(\frac{1}{n}-\frac{\bar{x}}{S_{xx}}(x_i-\bar{x}))-\hat{\beta}_1x_i)^2 \\ &=\sum_{i=1}^{n}(y_i-\bar{y}+\bar{y}-\bar{y}+\sum_{i=1}^{n}y_i\frac{\bar{x}}{S_{xx}}(x_i-\bar{x})-\hat{\beta}_1x_i)^2 \\ &=\sum_{i=1}^{n}[(y_i-\bar{y})+\hat{\beta}_1\bar{x}-\hat{\beta}_1x_i]^2 \\ &=\sum_{i=1}^{n}[(y_i-\bar{y})-\hat{\beta}_1(x_i-\bar{x})]^2 \\ &=\sum_{i=1}^{n}(y_i-\bar{y})^2+\sum_{i=1}^{n}\hat{\beta}_1^2(x_i-\bar{x})^2-2\sum_{i=1}^{n}(y_i-\bar{y})\hat{\beta}_1(x_i-\bar{x}) \\ &=\sum_{i=1}^{n}(y_i-\bar{y})^2+\hat{\beta}_1\sum_{i=1}^{n}[\hat{\beta}_1(x_i-\bar{x})^2-2(y_i-\bar{y})(x_i-\bar{x})] \\ &=\sum_{i=1}^{n}(y_i-\bar{y})^2+\hat{\beta}_1\sum_{i=1}^{n}[\hat{\beta}_1(x_i-\bar{x})^2-(y_i-\bar{y})(x_i-\bar{x})-(y_i-\bar{y})(x_i-\bar{x})] \\ &=\sum_{i=1}^{n}(y_i-\bar{y})^2-\hat{\beta}_1\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) \\ &=SST-\hat{\beta}_1S_{xy} \end{aligned}$$

(好难呀……)

$\left(=S_{yy}-\frac{(S_{xy})^2}{S_{xx}}\right)$

3. CI for Simple Linear Regression (β_0,β_1)

① For β_0 & β_1

Since $\hat{\beta}_1=\frac{1}{S_{xx}}\sum_{i=1}^{n}y_i(x_i-\bar{x})$, then $\hat{\beta}_0=\sum_{i=1}^{n}y_i(\frac{1}{n}-\frac{\bar{x}}{S_{xx}}(x_i-\bar{x}))$

And, $Y_i\sim N(\beta_0+\beta_1X_i,\sigma^2)$

$$\Rightarrow \begin{cases} E(\hat{\beta}_1)=\beta_1 \\ Var(\hat{\beta}_1)=\frac{\sigma^2}{S_{xx}} \end{cases} \quad \& \quad \begin{cases} E(\hat{\beta}_0)=\beta_0 \\ Var(\hat{\beta}_0)=\sigma^2[\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}] \end{cases} \quad \& \quad Cov(\hat{\beta}_0,\hat{\beta}_1)=-\sigma^2\frac{\bar{x}}{S_{xx}}$$

(not independent)

当 $\bar{x}$ 是对称式时, ind.

$$\Rightarrow \begin{cases} \hat{\beta}_1\sim N(\beta_1,\frac{\sigma^2}{S_{xx}}) \\ \hat{\beta}_0\sim N(\beta_0,\sigma^2(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}})) \end{cases}$$

(点估计)

Next, we deal with σ^2

Since $\frac{\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)^2}{\sigma^2}\sim\chi^2(n-2)$, then:

$$\begin{cases} \frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}/\sqrt{S_{xx}}}\sim t(n-2) \\ \frac{\hat{\beta}_0-\beta_0}{\hat{\sigma}\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}}}\sim t(n-2) \end{cases}, \text{ where } \hat{\sigma}^2=\frac{SS(E)}{n-2}$$

$$\Rightarrow \begin{cases} \text{CI for } \beta_1: (\hat{\beta}_1-t_{\alpha/2}(n-2)\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}},\ \hat{\beta}_1+t_{\frac{\alpha}{2}}(n-2)\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}) \\ \text{CI for } \beta_0: (\hat{\beta}_0-t_{\frac{\alpha}{2}}(n-2)\sqrt{\hat{\sigma}^2(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}})},\ \hat{\beta}_0+t_{\frac{\alpha}{2}}(n-2)\sqrt{\hat{\sigma}^2(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}})}) \end{cases}$$

S_{xx} 越小, x越聚集, CI越长, CL越大

x_0 代表某个已知的 x 值~

② For $(\beta_0+\beta_1 x_0)$ → $\widehat{\mu_{Y|x_0}}$ (mean response)

Since $\dfrac{(\hat\beta_0+\hat\beta_1 x_0)-(\beta_0+\beta_1 x_0)}{\hat\sigma\sqrt{\left(\frac{1}{n}+\frac{(x_0-\bar x)^2}{S_{xx}}\right)}} \sim t(n-2)$,

then CI for $(\beta_0+\beta_1 x_0)$: $\left(\widehat{\mu_{Y|x_0}} - t_{\frac{\alpha}{2}}(n-2)\sqrt{\hat\sigma^2\left(\frac{1}{n}+\frac{(x_0-\bar x)^2}{S_{xx}}\right)},\ \widehat{\mu_{Y|x_0}} + t_{\frac{\alpha}{2}}(n-2)\sqrt{\hat\sigma^2\left(\frac{1}{n}+\frac{(x_0-\bar x)^2}{S_{xx}}\right)}\right)$

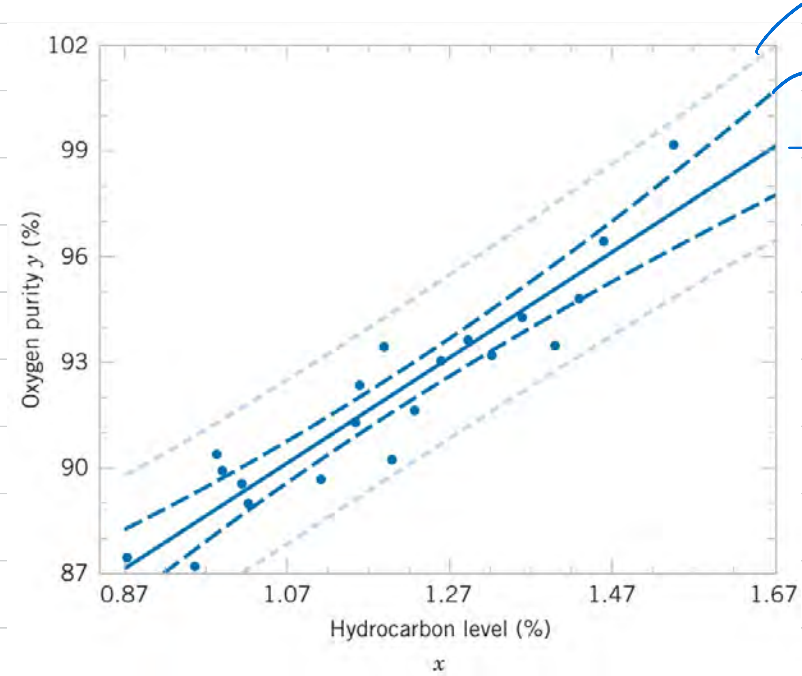
③ For y_0 (prediction)

Since $y_0=\beta_0+\beta_1 x_0+\varepsilon_i$, and $\varepsilon_i \sim N(0,\sigma^2)$,

then $\dfrac{\hat y_0 - y_0}{\hat\sigma\sqrt{\left(1+\frac{1}{n}+\frac{(x_0-\bar x)^2}{S_{xx}}\right)}} \sim t(n-2)$,

n要达到已有的值！

then CI for y_0: $\left(\hat y_0 - t_{\frac{\alpha}{2}}(n-2)\sqrt{\hat\sigma^2\left(1+\frac{1}{n}+\frac{(x_0-\bar x)^2}{S_{xx}}\right)},\ \hat y_0 + t_{\frac{\alpha}{2}}(n-2)\sqrt{\hat\sigma^2\left(1+\frac{1}{n}+\frac{(x_0-\bar x)^2}{S_{xx}}\right)}\right)$



CI for y_0

CI for $y_0-\varepsilon_i$

$\hat y_0$ | $\widehat{\mu_{Y|x_0}}$

我们可以把原来的model改写成 $Y_i=(\beta_0+\beta_1\bar x)+\beta_1(x_i-\bar x)+\varepsilon_i$, $\Rightarrow Y_i=\beta_0'+\beta_1 x_i+\varepsilon_i$

这样做的好处在于使regression沿y轴对称，即 $\bar x=0$

因此

$\hat\beta_1 \sim N\left(\beta_1,\frac{\sigma^2}{S_{xx}}\right)$, $\hat\beta_0 \sim N\left(\beta_0,\frac{\sigma^2}{n}\right)$

(不变)　　(表达式得到简化)

$\hat\sigma^2=\dfrac{SS(E)}{n-2}$ (不变)

$y_0-\varepsilon_i$ & y_0 不变

重点: CI for β_0 & β_1

Summary

Regression model $\boxed{Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i}$ $\varepsilon_i \sim N(0,\sigma^2)$

Parameter	Estimate	Confidence interval
β_0	$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right)\right)$	$[\hat{\beta}_0 - t_{\alpha/2}(n-2)\sqrt{\hat{\sigma}^2\left(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right)}, \hat{\beta}_0 + t_{\alpha/2}(n-2)\sqrt{\hat{\sigma}^2\left(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right)}]$
β_1	$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$	$[\hat{\beta}_1 - t_{\alpha/2}(n-2)\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}, \hat{\beta}_1 + t_{\alpha/2}(n-2)\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}]$
σ^2	$\hat{\sigma}^2 := \frac{1}{n-2}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$	

$$\frac{\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1(x_i - \bar{x})\right)^2}{\sigma^2} \sim \chi^2(n-2)$$

Summary

Regression model $\boxed{Y_i = \beta_0 + \beta_1 (X_i - \bar{X}) + \epsilon_i}$ $\varepsilon_i \sim N(0,\sigma^2)$

Parameter	Estimate	Confidence interval
β_0	$\hat{\beta}_0 = \bar{y} \sim N\left(\beta_0, \frac{\sigma^2}{n}\right)$	$[\hat{\beta}_0 - t_{\alpha/2}(n-2)\sqrt{\frac{\hat{\sigma}^2}{n}}, \hat{\beta}_0 + t_{\alpha/2}(n-2)\sqrt{\frac{\hat{\sigma}^2}{n}}]$
β_1	$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$	$[\hat{\beta}_1 - t_{\alpha/2}(n-2)\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}, \hat{\beta}_1 + t_{\alpha/2}(n-2)\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}]$
σ^2	$\hat{\sigma}^2 := \frac{1}{n-2}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1(x_i - \bar{x})\right)^2$	

$$\frac{\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1(x_i - \bar{x})\right)^2}{\sigma^2} \sim \chi^2(n-2)$$

4. HT for Simple Linear Regression (For original LR model: $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$)

① For β_0 & β_1

For β_1	For β_0
$H_0: \beta_1 = \beta_{10}$	$H_0: \beta_0 = \beta_{00}$
$H_1: \beta_1 \neq \beta_{10}$	$H_1: \beta_0 \neq \beta_{00}$
$t_0 = \dfrac{\hat{\beta}_1 - \beta_{10}}{\sqrt{\hat{\sigma}^2 / S_{xx}}} \sim t(n-2)$	$t_0 = \dfrac{\hat{\beta}_0 - \beta_{00}}{\sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]}} \sim t(n-2)$

$\Rightarrow$
- CR : $|t_0| > t_{\frac{\alpha}{2}}(n-2)$
- p-value : $2P(t > t_0)$

Specific case: we want to test whether the LR is significant. (i.e. whether the slope $\beta_1 = 0$)

These hypotheses relate to the significance of regression.
Failure to reject H_0 is equivalent to concluding that there is no linear relationship between *x* and *Y*.

总结: HT for β_0 & β_1

Summary: Inference on the Slope

Null Hypothesis	Test Statistic	Distribution under H_0
H_0: $\beta_1 = \beta_{1,0}$	$t_0 = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\hat{\sigma}^2/S_{xx}}}$	$t_0 \sim t(n-2)$

Alternative Hypothesis	Rejection/Critical Region (H_0 is rejected)	p-value	Confidence interval Approach
H_1: $\beta_1 \neq \beta_{1,0}$	$\lvert t_0 \rvert > t_{\alpha/2}(n-2)$	$2P(T > \lvert t_0 \rvert)$	$[\hat{\beta}_1 - t_{\alpha/2}(n-2)\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}, \hat{\beta}_1 + t_{\alpha/2}(n-2)\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}]$
H_1: $\beta_1 > \beta_{1,0}$	$t_0 > t_\alpha(n-2)$	$P(T > t_0)$	$[\hat{\beta}_1 - t_\alpha(n-2)\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}, +\infty)$
H_1: $\beta_1 < \beta_{1,0}$	$t_0 < -t_\alpha(n-2)$	$P(T < t_0)$	$(-\infty, \hat{\beta}_1 + t_\alpha(n-2)\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}]$

Threshold: Critical values

H_0 is rejected when p-value<α

H_0 is rejected when the CI does NOT include $\beta_{1,0}$

Summary: Inference on the Intercept

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

Original model

Null Hypothesis	Test Statistic	Distribution under H_0
H_0: $\beta_0 = \beta_{0,0}$	$t_0 = \frac{\hat{\beta}_0 - \beta_{0,0}}{\sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]}}$	$t_0 \sim t(n-2)$

Alternative Hypothesis	Rejection/Critical Region (H_0 is rejected)	p-value	Confidence interval Approach
H_1: $\beta_0 \neq \beta_{0,0}$	$\lvert t_0 \rvert > t_{\alpha/2}(n-2)$	$2P(T > \lvert t_0 \rvert)$	$[\hat{\beta}_0 - t_{\alpha/2}(n-2)\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}, \hat{\beta}_0 + t_{\alpha/2}(n-2)\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}]$
H_1: $\beta_0 > \beta_{0,0}$	$t_0 > t_\alpha(n-2)$	$P(T > t_0)$	$[\hat{\beta}_0 - t_\alpha(n-2)\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}, +\infty)$
H_1: $\beta_0 < \beta_{0,0}$	$t_0 < -t_\alpha(n-2)$	$P(T < t_0)$	$(-\infty, \hat{\beta}_0 + t_\alpha(n-2)\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}]$

Exercise: what is the test statistic and rejection region under the alternative model?

H_0 is rejected when p-value<α

H_0 is rejected when the CI does NOT include $\beta_{0,0}$

Summary: Inference on the Intercept

$Y_i = \beta_0 + \beta_1(X_i - \bar{X}) + \epsilon_i$

Null Hypothesis	Test Statistic	Distribution under H_0
H_0: $\beta_0 = \beta_{0,0}$	$t_0 = \frac{\hat{\beta}_0 - \beta_{0,0}}{\hat{\sigma}/\sqrt{n}}$	$t_0 \sim t(n-2)$

Alternative Hypothesis	Rejection/Critical Region (H_0 is rejected)	p-value	Confidence interval Approach
H_1: $\beta_0 \neq \beta_{0,0}$	$\lvert t_0 \rvert > t_{\alpha/2}(n-2)$	$2P(T > \lvert t_0 \rvert)$	$[\hat{\beta}_0 - t_{\alpha/2}(n-2)\sqrt{\frac{\hat{\sigma}^2}{n}}, \hat{\beta}_0 + t_{\alpha/2}(n-2)\sqrt{\frac{\hat{\sigma}^2}{n}}]$
H_1: $\beta_0 > \beta_{0,0}$	$t_0 > t_\alpha(n-2)$	$P(T > t_0)$	$[\hat{\beta}_0 - t_\alpha(n-2)\sqrt{\frac{\hat{\sigma}^2}{n}}, +\infty)$
H_1: $\beta_0 < \beta_{0,0}$	$t_0 < -t_\alpha(n-2)$	$P(T < t_0)$	$(-\infty, \hat{\beta}_0 + t_\alpha(n-2)\sqrt{\frac{\hat{\sigma}^2}{n}}]$

H_0 is rejected when p-value<α

H_0 is rejected when the CI does NOT include $\beta_{0,0}$

HT

② For correlation (此时我们开始把X也看作是RV了)

a. Recap of Bivariate Dis knowledge

△ $Cov(X,Y) = \sigma_{XY}$ (注意没有"2")

△ X & Y are ind $\underset{\times}{\overset{\checkmark}{\rightleftarrows}}$ $\rho = 0$

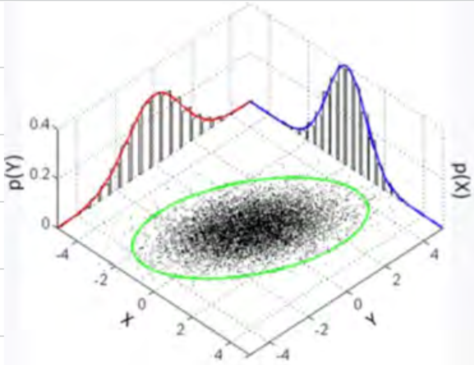
但是对于Bivariate normal，逆命题成立

△ $\rho = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y} \Rightarrow R = \hat{\rho} = \dfrac{S_{XY}}{S_X S_Y}$

$$= \frac{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2}}$$

△ pdf for Bivariate Normal Distribution:

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right)$$



△ BND's conditional distribution is still normal:

$$Y|X=x \sim N\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X),\ (1-\rho^2)\sigma_Y^2\right)$$

Simplified SLR model: $\mu_{Y|x} = \beta_0 + \beta_1(x-\mu_X)$ $\boxed{\beta_1 = \rho\frac{\sigma_Y}{\sigma_X}}$

$\Downarrow$

$\hat{\beta}_1 = R\dfrac{S_Y}{S_X}$

b. Process → 构造关于R的式子

$$\begin{cases} H_0: \rho = 0 \\ H_1: \begin{cases} \rho \neq 0 \\ \rho < 0 \\ \rho > 0 \end{cases} \end{cases}$$

Construct Z

Since with OSLR model: $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

Under H_0, we have: (Conditioned on $x_1, \cdots, x_n$)

按Cov的定义拆开~

$$\sigma_{XY} = 0 \Rightarrow Cov(X_i, Y_i) = Cov(X_i, \beta_0 + \beta_1 X_i + \varepsilon_i) = \underbrace{Cov(X_i,\beta_0)}_{0} + \underbrace{Cov(X_i,\beta_1 X_i)}_{\beta_1 Var(X_i)} + \underbrace{Cov(X_i,\varepsilon_i)}_{0}$$

$$\Rightarrow \beta_1 = 0 \Rightarrow \hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma_Y^2}{\sum(X_i-\bar{X})^2}\right) \longrightarrow \underline{\hat{\beta}_1 \sim N\left(0, \frac{\sigma_Y^2}{\sum(X_i-\bar{X})^2}\right)}$$

Construct U

we have $\dfrac{\sum_{i=1}^{n}(Y_i-\bar{Y}-\hat{\beta}_1(X_i-\bar{X}))^2}{\sigma_Y^2} = \dfrac{\sum_{i=1}^{n}(Y_i-\bar{Y}-(\frac{S_Y}{S_X}R)(X_i-\bar{X}))^2}{\sigma_Y^2} = \dfrac{(n-1)S_Y^2 + (\frac{S_Y}{S_X}R)^2(n-1)S_X^2 - 2(\frac{S_Y}{S_X}R)(n-1)S_X S_Y R}{\sigma_Y^2} = \dfrac{(n-1)S_Y^2(1-R^2)}{\sigma_Y^2} \sim \chi^2(n-2)$

Construct T (the final test stat)

$$\frac{\frac{S_Y}{S_X}R\sqrt{(n-1)S_X^2}}{\sqrt{(n-1)S_Y^2(1-R^2)/n-2}} \sim t(n-2) \Rightarrow T_0 = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}} \sim t(n-2)$$

Since $\forall X = x$, test stat $\sim t(n-2)$, then we can remove the premise "Conditioned on $x_1, \cdots, x_n$"

After successfully constructing our test Stat, we conduct CR & p-value testing

CR: $|t_0| > t_{\frac{\alpha}{2}}(n-2)$

p-value: $2P(T > t_0)$

If we seem R as our test Stat,

then CR : $|R| > \underbrace{r_{\frac{\alpha}{2}}(n-2)}_{\text{查表.}}$

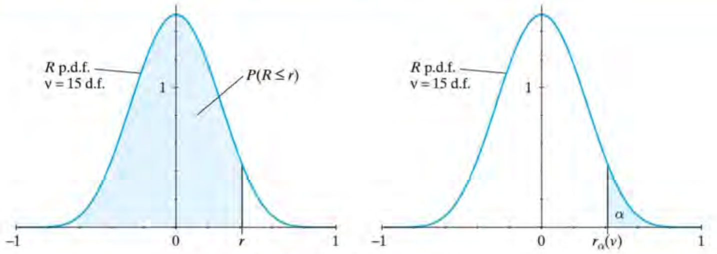
① 实际上，R的Dis是对称的

② 且 $R\uparrow$, $t_0\uparrow$

两个test Stat 正相关

$t_0 = f(R)$

Table IX Distribution Function of the Correlation Coefficient R, $\rho = 0$



$$P(R \le r) = \int_{-1}^{r} \frac{\Gamma[(n-1)/2]}{\Gamma(1/2)\Gamma[(n-2)/2]}(1-w^2)^{(n-4)/2}\,dw$$

$\nu = n-2$ degrees of freedom	$P(R \le r)$ 0.95	0.975	0.99	0.995
	$r_{0.05}(\nu)$	$r_{0.025}(\nu)$	$r_{0.01}(\nu)$	$r_{0.005}(\nu)$
1	0.9877	0.9969	0.9995	0.9999
2	0.9000	0.9500	0.9800	0.9900
3	0.8053	0.8783	0.9343	0.9587
4	0.7292	0.8113	0.8822	0.9172
5	0.6694	0.7544	0.8329	0.8745
6	0.6215	0.7067	0.7887	0.8343
7	0.5822	0.6664	0.7497	0.7977
8	0.5493	0.6319	0.7154	0.7646
9	0.5214	0.6020	0.6850	0.7348
10	0.4972	0.5759	0.6581	0.7079
11	0.4761	0.5529	0.6338	0.6835
12	0.4575	0.5323	0.6120	0.6613
13	0.4408	0.5139	0.5922	0.6411
14	0.4258	0.4973	0.5742	0.6226
15	0.4123	0.4821	0.5577	0.6054
16	0.4000	0.4683	0.5425	0.5897
17	0.3887	0.4555	0.5285	0.5750
18	0.3783	0.4437	0.5154	0.5614
19	0.3687	0.4328	0.5033	0.5487
20	0.3597	0.4226	0.4920	0.5367
25	0.3232	0.3808	0.4450	0.4869
30	0.2959	0.3494	0.4092	0.4487
35	0.2746	0.3246	0.3809	0.4182
40	0.2572	0.3044	0.3578	0.3931
45	0.2428	0.2875	0.3383	0.3721
50	0.2306	0.2732	0.3218	0.3541
60	0.2108	0.2500	0.2948	0.3248
70	0.1954	0.2318	0.2736	0.3017
80	0.1829	0.2172	0.2565	0.2829
90	0.1725	0.2049	0.2422	0.2673
100	0.1638	0.1946	0.2300	0.2540

r值

Summary: testing correlation

Null Hypothesis

$H_0: \rho = 0$

Test Statistic $t = \dfrac{R\sqrt{n-2}}{\sqrt{1-R^2}}$ **Distribution under H_0** $t \sim t(n-2)$

Alternative Hypothesis	Rejection/Critical Region (H_0 is rejected)	p-value
$H_1: \rho \neq 0$	$\lvert t\rvert > t_{\alpha/2}(n-2)$	$2P(T > \lvert t\rvert)$
$H_1: \rho > 0$	$t > t_\alpha(n-2)$	$P(T > t)$
$H_1: \rho < 0$	$t < -t_\alpha(n-2)$	$P(T < t)$

H_0 is rejected when p-value<α

Test Statistic

R

Alternative Hypothesis	Rejection/Critical Region (H_0 is rejected)
$H_1: \rho \neq 0$	$\lvert R\rvert > r_{\alpha/2}(n-2)$
$H_1: \rho > 0$	$R > r_\alpha(n-2)$
$H_1: \rho < 0$	$R < -r_\alpha(n-2)$

if we want to change H_0 from "$\rho=0$" to "$\rho=0.5$", then:

$$\frac{1}{2}\ln\left(\frac{1+R}{1-R}\right) \overset{app}{\sim} N\left(\frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right), \frac{1}{n-3}\right)$$

then we can construct our test stat:

$$\frac{\frac{1}{2}\ln\left(\frac{1+R}{1-R}\right) - \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)}{\sqrt{\frac{1}{n-3}}}$$ ← 此时 $\rho\in[-1,1]$

$\Rightarrow$ CR: $|z| > z_{\frac{\alpha}{2}}$

p-value: $2P(z > z_0)$

From the test stat above, we can further construct CI for ρ:

Since $P\left(-z_{\frac{\alpha}{2}} \le \frac{\frac{1}{2}\ln\left(\frac{1+R}{1-R}\right) - \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)}{\sqrt{\frac{1}{n-3}}} \le z_{\frac{\alpha}{2}}\right) \approx 1-\alpha$

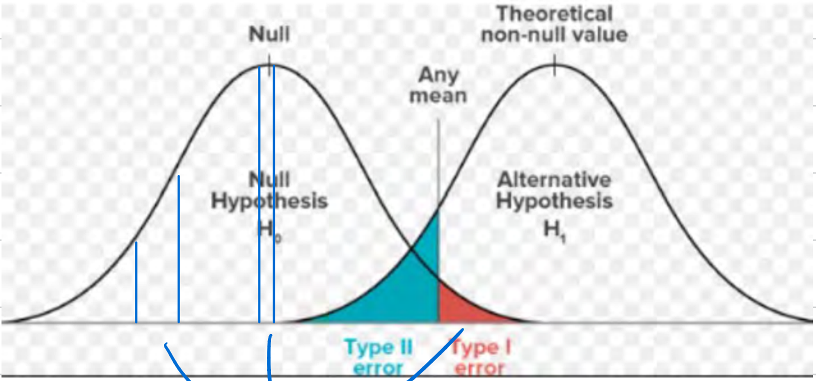
then (取指数) $CI = \left(\frac{1+R-(1-R)e^{\frac{2z_{\frac{\alpha}{2}}}{\sqrt{n-3}}}}{1+R+(1-R)e^{\frac{2z_{\frac{\alpha}{2}}}{\sqrt{n-3}}}}, \frac{1+R-(1-R)e^{\frac{-2z_{\frac{\alpha}{2}}}{\sqrt{n-3}}}}{1+R+(1-R)e^{\frac{-2z_{\frac{\alpha}{2}}}{\sqrt{n-3}}}}\right)$

Lecture 25-26

1. Likelihood Ratio Test

① Definition of Best Critical Region

Given α, we need to find C that maximizes power. → minimize β



α可以取任何区域，但仅当取右侧尾端时，我们能最大化power (最小化β)

⇒ Formal definition:

$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta = \theta_1 \end{cases}$, $\alpha = P(C|\theta_0)$, then C is a best CR of size α if $\forall D$ of size α, we have $P(C|\theta_1) \geq P(D|\theta_1)$

maximize power

② Neyman-Pearson Lemma

(Neyman–Pearson Lemma) Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution with pdf or pmf $f(x;\theta)$, where θ_0 and θ_1 are two possible values of θ. Denote the joint pdf or pmf of $X_1, X_2, \ldots, X_n$ by the likelihood function

$$L(\theta) = L(\theta; x_1, x_2, \ldots, x_n) = f(x_1;\theta) f(x_2;\theta) \cdots f(x_n;\theta).$$

If there exist a positive constant k and a subset C of the sample space such that

(a) $P[(X_1, X_2, \ldots, X_n) \in C; \theta_0] = \alpha$,

(b) $\dfrac{L(\theta_0)}{L(\theta_1)} \leq k$ for $(x_1, x_2, \ldots, x_n) \in C$, and

(c) $\dfrac{L(\theta_0)}{L(\theta_1)} \geq k$ for $(x_1, x_2, \ldots, x_n) \in C'$,

then C is a best critical region of size α for testing the simple null hypothesis H_0: $\theta = \theta_0$ against the simple alternative hypothesis $H_1: \theta = \theta_1$.

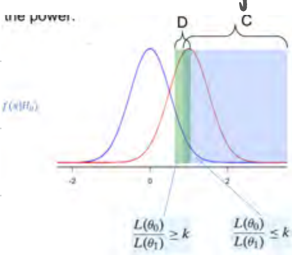
求k → 求c

如果真实数据更支持H_1，则我们当ratio小于某个k时拒绝H_0

如果真实数据更支持H_0，则我们当ratio大于某个k时拒绝H_1

Brief proof

Assume C is BCR of size α and there exists another CR of size α D, then $\alpha = \int_C L(\theta_0) = \int_D L(\theta_0)$ (assuming cont. Dis)



then we have $\int_{C \cap D'} L(\theta_0) = \int_{C' \cap D} L(\theta_0)$

$C \cap D'$, $C' \cap D$

by NP Lemma, $\begin{cases} k \int_{C\cap D'} L(\theta_1) \geq \int_{C \cap D'} L(\theta_0) \\ k \int_{C' \cap D} L(\theta_1) \leq \int_{C' \cap D} L(\theta_0) \end{cases}$ ⇒ $\int_C L(\theta_1) \geq \int_D L(\theta_1)$

maximize power

③ Uniformly most powerful test (H₁)

a. Definition

由θ决定 → "Uniformly"

$$\begin{cases} \text{Most powerful test} : \text{The test using BCR at } \theta=\theta_1 \\ \text{Uniformly most powerful test} : \text{The test using BCR at each pt in } H_1 \end{cases} \Rightarrow \begin{cases} \text{BCR is the same} \\ k \text{ in NP Lemma is different} \end{cases}$$

How to prove UMPT? ⇒ Prove $\forall \theta_1$ in H_1, $\frac{L(\theta_0)}{L(\theta_1)} \le k$

关注ratio在H₁区间内的单调性

b. 当H₁是two-sided时, UMPT通常不存在

④ Composite null hypothesis

a. Likelihood ratio

$$\begin{cases} H_0: \theta \in w \\ H_1: \theta \in w' \end{cases}, \text{ where } w \ \& \ w' \text{ are mutually exclusive and } w \cup w' = \Omega$$

parameter space 不一定是全集

then we define the likelihood ratio as follows:

$$\lambda = \frac{L(\hat{w})}{L(\hat{\Omega})} = \frac{\max_{\theta \in w} L(\theta)}{\max_{\theta \in \Omega} L(\theta)}$$

有时我们需要分类讨论

mle

we can futher explore the property of λ

i) $\lambda \in [0,1]$

ii) A small value of λ will lead to the rejection of H_0

b. Critical region for the likelihood ratio test

$$\lambda = \frac{L(\hat{w})}{L(\hat{\Omega})} \le k$$, where $k \in (0,1)$ and k is set according to α

(具体的examples请看PPT)