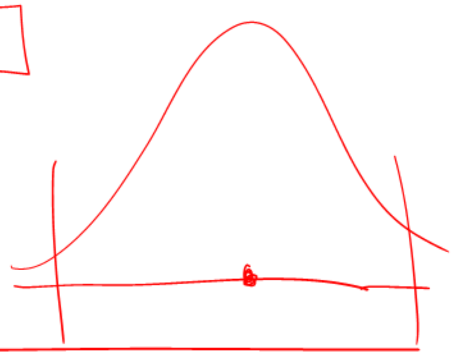
Topic 5: trimmed mean : $\alpha = 0.1$

$\underbrace{0, 1,}$ $\underbrace{10, 12, 11, 13, \ldots, 40,}$ $\underbrace{75, 90}$.

$\bar{X}$

pop mean : μ ← most common estimator : $\bar{X}$ = sample mean.

CI for μ : $\bar{X} \pm \boxed{}$



⇓

ST2132

$\boxed{\text{parameter}}$ ← estimate ← $\boxed{\text{estimator}}$

pop. mean : μ ← ← $\bar{X}$: unbiased estimator of μ.

$E(\bar{X}) = \mu$

pop. location ← trimmed mean → robust

$\bar{X}$ → not robust.

Contingency table : conditional prob :

		Disease yes	No	Total
X	yes	a	b	a + b
	No	c	d	c + d

$$\Pr(\text{Diseased} \mid X = \text{yes}) = \frac{a}{a+b} := P_1$$

$$\Pr(D \mid X = \text{No}) = \frac{c}{c+d} := P_2$$

P_1 vs P_2 :
- $P_1 - P_2$ ①
- P_1 / P_2 ②

③ : $OR = \frac{ad}{bc}$ ←

For a general parameter: $\underline{\theta}$ ⇐ a point estimate for θ : $\hat{\theta}$:

$$\hat{\theta} \pm \underline{\text{margin of error}}$$

$$= \boxed{\text{multiplier}} * \text{s.error of } \hat{\theta} .$$

---

parameter = pop mean μ ⇐ point estimate : $\bar{X}$

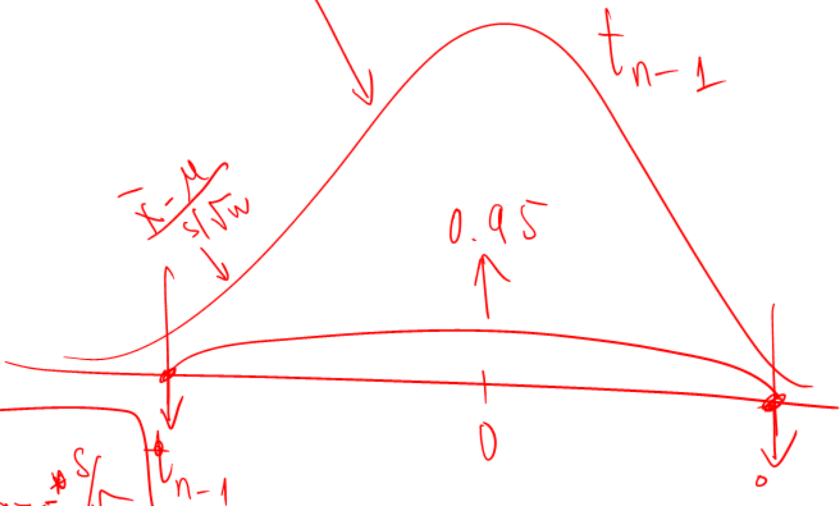
$$\bar{X} \pm \underline{\text{margin of error}}$$

CLT

$$\Rightarrow \bar{X} \sim \underline{N(\mu ; \sigma^2/n)}$$

$$\Rightarrow \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \Rightarrow \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

⇒



95% CI for μ

⇒ $$\boxed{\bar{X} \pm t_{n-1, 0.975} * s/\sqrt{n}}$$

---

parameter: pop. odds ratio : θ ⇐ point estimate : $\hat{\theta}$ ← can get from data

CI : $\hat{\theta} \pm \underline{\text{margin of error}}$.

Sampling distn of $\hat{\theta}$ is <u>NOT</u> normal, but :

sampling distn of $\underline{\log \hat{\theta}}$ can be <u>approximated by normal</u>.

$\boxed{\text{Flow}}$ → ⇒ <u>form CI</u> for $\underline{\log \theta}$ ⇒ take expo...

⇒ get CI for θ .

$\theta \leftarrow \hat{\theta}$

$\log\theta \leftarrow \log\hat{\theta}$: CI for $\log\theta$: $\log\hat{\theta} \pm \boxed{\text{margin of error.}}$

$= \boxed{\text{multiplier}} * \text{S.E of } (\log\hat{\theta})$

from $Z \sim N(0,1)$

$$\boxed{\log\hat{\theta} \pm \underbrace{Z_{\alpha/2}}_{\approx q_{1-\frac{\alpha}{2}}} * \text{ASE}(\log\hat{\theta}) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}}$$

a

$\underline{\text{CI} = (0.8 ; 5)}$ for θ $\ni 1 \Rightarrow$ OR could equal to 1

pop oddsratio

$t_{n-1} \to Z$

df very large then $t_{df} \approx Z$

CI for μ :

$$\boxed{\bar{X} \pm Z_{\frac{\alpha}{2}} * s/\sqrt{n}}$$

$t_{n-1;}$

$\boxed{\text{t\_test}}$ → 2132 / 1131 / 3131

Males → $\mu \neq 1.7\text{m}$ → pop. mean

Sample : 100 = n ⇒

point estimate : sample mean : $\bar{X}$

$\bar{X} \xrightarrow{\text{close}} 1.7 \overset{\text{then}}{\Rightarrow} \boxed{\mu \approx 1.7}$ is supported

$\bar{X} \xrightarrow{\text{far}} 1.7$ then $\boxed{\mu = 1.7}$ is against

to perform a test : checking if data provide evidence against some statement.

null hypothesis H_0 : $\mu = 1.7$

alternative hypo. $H_1 : H_A$ $\mu \neq 1.7$ / $\mu < 1.7$ / $\mu > 1.7$ — one of these 3 only.

⇒ find evidence that data support H_0: $\bar{X}$ vs 1.7

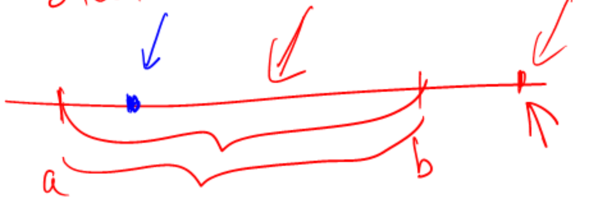
if $\bar{X}$ is close to 1.7 ⇒ data is for H_0.

if $\bar{X}$ is far from 1.7 ⇒ data is against H_0.

$$\boxed{\frac{\bar{X} - 1.7}{\text{s.e of } \bar{X}}}$$:= called test statistic

⇒ find distn of test statistic under H_0.

assume H_0 is true ⇒ then find the distn of the test statistic. $\mu = 1.7$ H_0.



X → Y : are → associated? → independent? ⇒ a test.

H_0 : X & Y are indpd H_1 : X & Y are associated.

⇒ assume H_0 is true ⇒ what will happen to the table?
⇒ calculate the values for the cells.
expected counts.

⇒ Compare the observed counts & expected counts.

$O_i \to E_i \quad i = \overline{1,4}$

$\boxed{E_i =}$

$$\frac{|O_i - E_i|^2}{\ldots}$$ → test statistic.

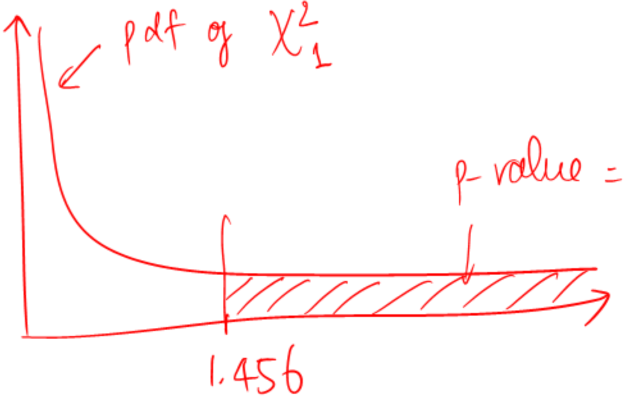
if this distance / test statistic is small ⇒ $O_i \approx E_i$
⇒ data is supporting H_0.

if the test statistic is large ⇒ $O_i \neq E_i$
⇒ data provides evidence against H_0.

⇒ p-value : helps to quantify how data against H_0.

if p-value is large ⇒ data do not provide enough evidence against H_0

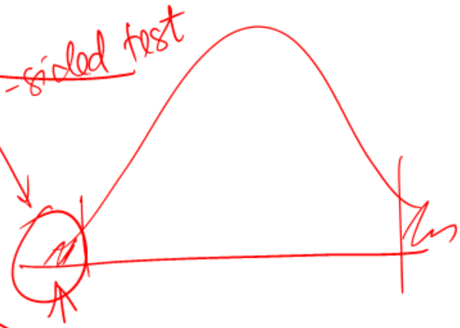
p-value is small ⇒ data provide strong/enough evidence against H_0.



p-value = right area of test statistic. ⇒ (for χ^2 test)

$H_0: \mu = 1.7$ vs $H_1: \mu \neq 1.7$ 2-sided test

p-value = 0.0005



$H_0: \mu = 1.7$ vs $H_1: \mu < 1.7$. 1 sided test.

p-value

17/9/2021

χ^2 test : to test the indpd between 2 variables:

H_0 : 2 variables are indpd ⇐ true ⇒ expected count in each cell ⇒ E_i

H_1 : ______ associated.

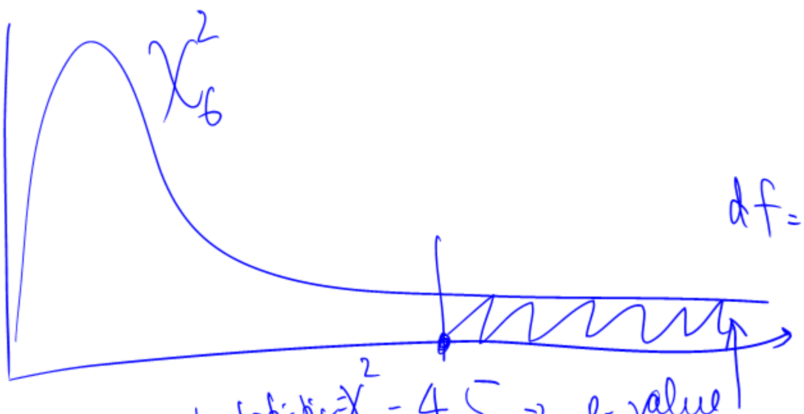
⇒ Test statistic : $\sum_i \frac{(|O_i - E_i| - 0.5)^2}{E_i}$ ← H_0 χ^2_1 if table = 2×2 $(2-1)\times(2-1) = 1$

2×3 $i = 1 \to 6 = r \times c$

H_0 $\chi^2_{(r-1)\times(c-1)}$ table = r×c

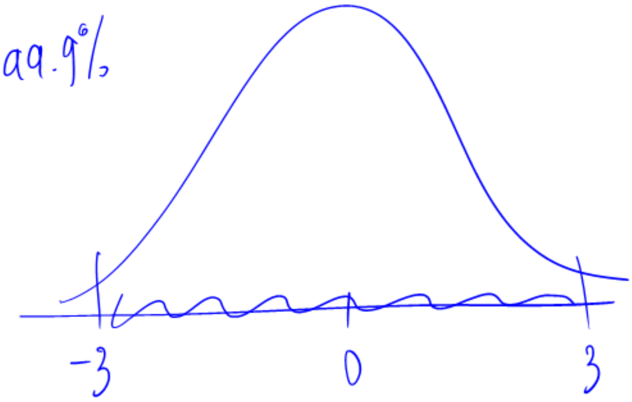
$df = (r-1) * (c-1)$

df = 6



test statistic = χ^2 = 4.5 ⇒ p-value

99.9%

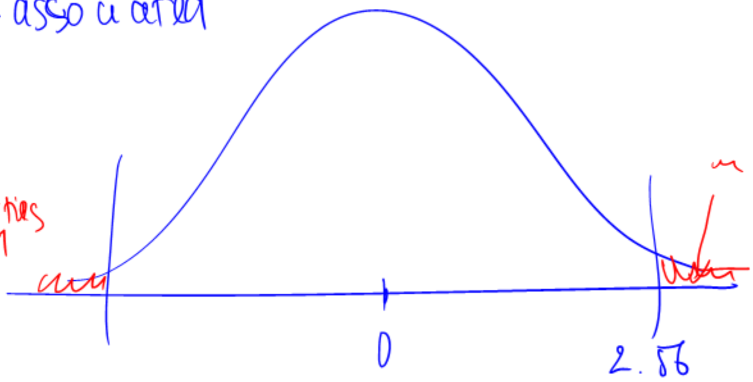


H_0 : Acohol & M are indpd

H_1 : ______ associated

$M = 2.56 \overset{H_0}{\sim} N(0,1)$

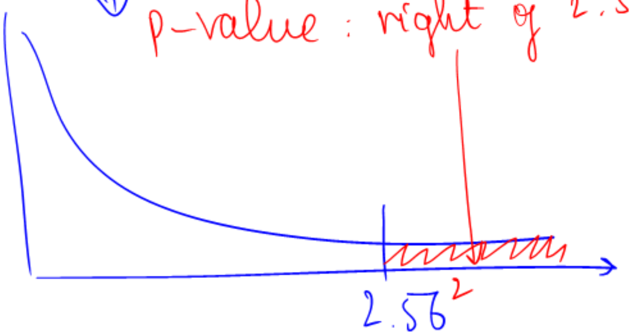
p-value = 2 tail probabilities



$M^2 = 2.56^2 \sim \chi^2_1$

$\Downarrow$ p-value : right of 2.56^2 = 0.0104

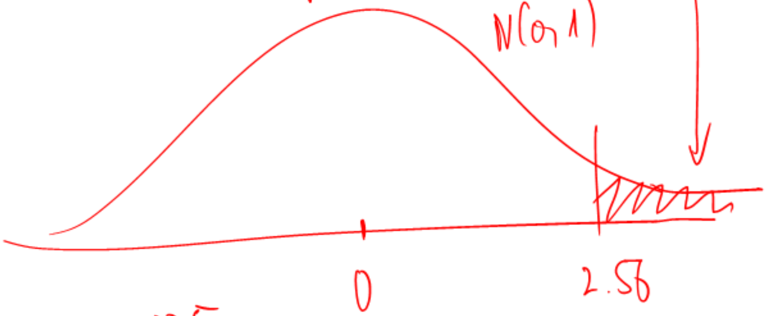
$\Downarrow$ small $\Rightarrow$ strong evidence against H_0.



1-sided test : H_0 : indpd

$\rightarrow H_1$: positive association

$M = 2.56 \Rightarrow$ p-value : right area of 2.56 under $N(0,1)$



$\Rightarrow$ p-value = 0.005

another 1 sided test :

H_0 : indpd

H_1 : negative association

$M = 2.56 \sim N(0,1)$

$\Downarrow$

p-value = left area of 2.56

= 0.995