Continuous Distributions pmf vs (pdf & cdf)

① probability distribution

$f(x)$ = function that describes possible values of x and their probabilities

② simple with discrete: pmf $= f(x) = P(X=x)$

eg. $x \sim U_{Disc}(a,b)$ $n=b-a+1$, pmf $= f(x) = \frac{1}{n}$

$x \sim U_{Disc}(2,5)$ $n = 5-2+1 = 4$, pmf $= f(x) = \frac{1}{4}$

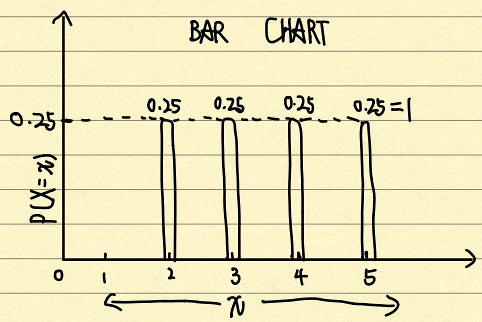
③ What about continuous ?

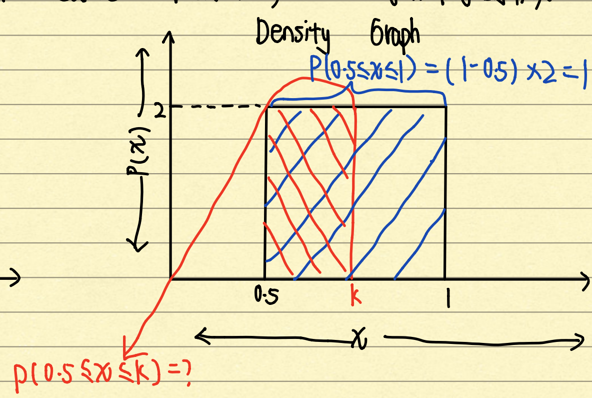
pdf is equivalent to pmf as both describe possible values of x and their possibilities

④ but this description is not as simple

pdf $= f(x)$ = height of $f(x)$ on density graph

cdf $= F(x)$ = area under the curve $= P(X \leq x)$ cdf是pdf的积分





⑤ $X \sim U_{CONT}(0.5, 1)$

pdf $= f(x) = \frac{1}{b-a} = \frac{1}{1-0.5} = 2$

Can't be a probability as $2 > 1$

⑥ Find probabilities by finding $\overset{\text{integration}}{\boxed{\text{area under the curve}}}$ of the pdf → cdf

if pdf is $f(x) = \frac{1}{b-a}$

then cdf is $F(x) = P(a \leq x \leq k) = \int_k^a \left(\frac{1}{b-a}\right) dx$

$= \ln\Big|_a^k \frac{1}{b-a}(x) = \frac{k-a}{b-a}$ =>

eg. if $k = 0.75$, $P(0.5 \leq x \leq 0.75) = \frac{0.75-0.5}{1-0.5} = 0.5 = 50\%$

⑦ So why does $P(x \leq k) = P(x<k)$?

$P(x \leq k) = P(x<k) + P(x=k)$ $\Big\}$ → $P(x=k) = \int_k^k \left(\frac{1}{b-a}\right) dx = \ln\Big|_k^k \frac{1}{b-a}(x) = 0$

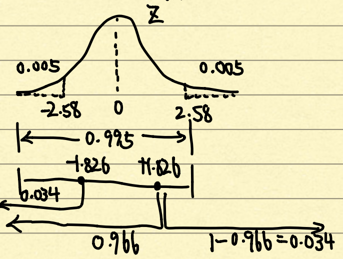
⑧ In a continuous distribution, there are an infinite number of values in even a tiny integral. So, the probability of a random variable will be <u>exactly equal</u> to any specific value is 0

Hypothesis test (means)

Sample: $\bar{x} = 24$, sd $= 3$, $n = 30$

Null Hypothesis: $H_0: \mu = 24+1 = 25$ $\quad H_A: \mu \neq 25$ $\quad \alpha = 0.01$

Test stat: $\dfrac{\bar{x} - \mu}{sd/\sqrt{n}} = \dfrac{24-25}{3\sqrt{30}} = -1.826$



Z test: (2 sided)

Z-score for $+1.826 = 0.966$

$1 - 0.966 = 0.034$

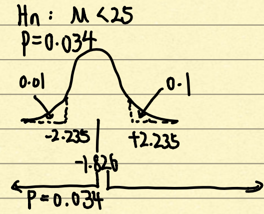
p-val $= 0.034 \times 2 = 0.068$

Z test: (1 sided)

$H_0: \mu \geq 25$

$H_A: \mu < 25$

$P = 0.034$



$\dfrac{\mu_0 - \bar{x}}{sd/\sqrt{n}}$: - 2 sided test would be the same
- interprete the test statistic differently for one sided test (by changing sign)

- t-test: test statistic doesn't change: -1.826
- need df: here $n-1 = 30-1 = 29$
- in this course, only 2 sided tests

look up $\alpha = 0.01$ and df $= 29$ in table

$\rightarrow 2.76$

$-2.76 < -1.826 < +2.76$

therefore fail to reject Ho (null Hypothesis)

Two sample test:

Sample 1: $\bar{x}_1 = 20$, $\mu_1 = 40$, $sd_1 = 5$

Sample 2: $\bar{x}_2 = 30$, $\mu_2 = 50$, $sd_2 = 8$

difference in means

$H_0: \mu_1 - \mu_2 = 3$ $\quad H_A: \mu_1 - \mu_2 \neq 3$

test-statistic: $\dfrac{(\mu_1 - \mu_2) - 3}{SE(\mu_1 - \mu_2)}$

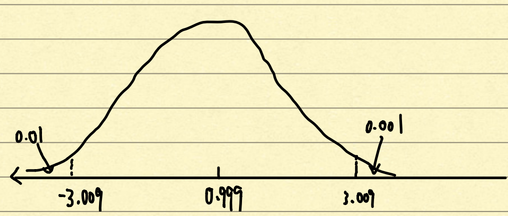
Assume equal variance

calculate pooled sd: $s = \sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-2)s_2^2}{n_1+n_2-2}}$

$$s = \sqrt{\frac{(40-1)5^2 + (50-1)8^2}{40+50-2}} = 6.84$$

$$SE(\mu_1 - \mu_2) = s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 6.84 \times \sqrt{\frac{1}{40} + \frac{1}{50}} = 1.45$$

Hence: test statistic $= \dfrac{20-30}{1.45} = -6.90$



Z test: 3.009 has 0.999 to the left

t test : $df = n_1 + n_2 - 2 = 40 + 50 - 1 - 1 = 88$

α - level $= 0.01$ 查 t-table

CV ≈ 2.635

$-7.90 < -2.635$

Therefore reject H_0 (null Hypothesis)

Assumed unequal variance

Z- test : $SE(\mu_1 - \mu_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{5^2}{40} + \frac{8^2}{50}} = 1.38$

test statistic $= \frac{20 - 30}{1.38} = -7.25$

Hypothesis test (proportions)

sample size : 50   23 observation classed as positive

$\alpha = 0.01$ $H_0 = \pi$ (population proportion) $= 0.5$ $H_n = \pi \neq 0.5$

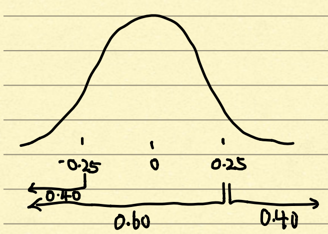
$\bar{\pi} = \frac{23}{50} = 0.46$

$n\pi_0 = 50\,(0.5) = 25$ $n(1 - \pi_0) = 50\,(1 - \frac{1}{2}) = 25$

Test Statistic $= \frac{\bar{\pi} - \pi_0}{SE(\bar{\pi})}$ $SE(\bar{\pi}) = \sqrt{\frac{\bar{\pi}\,(1 - \bar{\pi})}{n}} = \sqrt{\frac{0.46\,(0.54)}{50}} = 0.07$

$= \frac{0.46 - 0.50}{0.07} = -0.25$

Look up + 0.25 in normal distribution => 0.6



p- value $= 2 \times 0.40 = 0.80$

fail to reject H_0

2 sample test

$\alpha = 0.01$

sample 1: $\pi_1 = 40\%$ $n = 100$ $H_0: \bar{\pi}_1 - \bar{\pi}_2 = 2$

sample 2: $\pi_2 = 50\%$ $n = 50$ $H_n: \bar{\pi}_1 - \bar{\pi}_2 \neq 2$

$n_1\bar{\pi}_1 = 100 \times 0.40 = 40$, so $n_1(1 - \bar{\pi}_1) = 60$

$n_2\bar{\pi}_2 = 50 \times 0.50 = 25$, so $n_2(1 - \bar{\pi}_2) = 25$

Test Statistic : $\frac{\bar{\pi}_1 - \bar{\pi}_2 - \overset{\alpha \times 2}{0.02}}{SE(\bar{\pi}_1 - \bar{\pi}_2)}$ $\bar{\pi}_0 = \frac{\bar{\pi}_1 n_1 + \bar{\pi}_2 n_2}{n_1 + n_2} = \frac{40 + 25}{100 + 50} = 0.43$

$SE(\bar{\pi}_1 - \bar{\pi}_2) = \sqrt{\bar{\pi}_0 (1 - \bar{\pi}_0)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

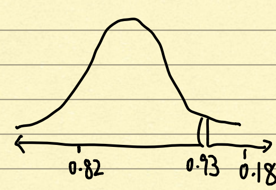
$= \sqrt{(0.43)(0.57)\left(\frac{1}{100} + \frac{1}{50}\right)} = 0.086$

$= \frac{|0.40 - 0.50| - 0.02}{0.086}$ => $-0.10 - 0.02$

=> $0.50 - 0.40 - 0.02$

$+0.10 - 0.02$ => $\frac{0.10 - 0.02}{0.086} = 0.93$ → 0.82 (normal table)

看 0.40 − 0.50 是否使用了绝对值



p-value = 2 × 0.18 = 0.36
fail to reject H_0 $\alpha = 0.01$
CI for $\pi_1 - \pi_2$ (difference in means) $\longrightarrow \bar{\pi}_1 - \bar{\pi}_2 \pm 1.96\,SE(\bar{\pi}_1 - \bar{\pi}_2)$
$= 0.40 - 0.50 \pm 1.96 \times (0.086)$

Chi Square Test

	pos	neg	Total
sample 1	40	60	100
sample 2	25	25	50
Total	65	85	150

	pos	neg
sample 1	$\frac{100 \times 65}{150} = 43.3$	$\frac{100 \times 85}{150} = 56.7$
samle 2	$\frac{50 \times 65}{150} = 21.7$	$\frac{50 \times 85}{150} = 28.3$

Expected value : $\frac{\text{Row Total} \times \text{Col Total}}{\text{Total No. of obs}}$

Test Statistic = $\underset{\text{All cells}}{\text{Sum}}$ (observed − expected)^2 / expected = −0.25 + 0.19 + 0.50 − 0.38
= 0.06

	pos	neg
sample 1	$\frac{(40-43.3)^2}{43.3}$	$\frac{(60-56.7)^2}{56.7}$
Sample 2	$\frac{(25-21.7)^2}{21.7}$	$\frac{(25-28.3)^2}{28.3}$

degree of freedom = (rows − 1) (columns − 1)
= (2−1) (2−1)
= 1×1 = 1

$\alpha = 0.001$ CV = 10.83
since 0.06 < 10.83
fail to reject H_0