Concept

Sampling error : the difference between sample and population

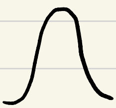
Sampling bias → Selection bias
- non-response bias
- undercoverage bias
- volunteer bias
- Convenience sampling bias
- survivorship bias

Normal distribution

$X \sim N(\mu, \sigma^2)$

" X follows a normal distribution with mean μ & standard deviation σ"

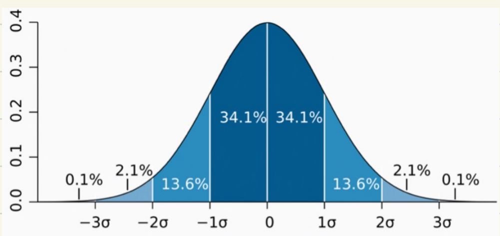
" is normally distributed variance σ^2 "



"bell - shaped"

mean, median, and <u>mode</u> are all equal

众数：正态分布中最高的点
(即最频繁出现的值)



68 - 95 - 99.7 Rule

Sampling distribution

Take samples of size n from a normal distribution repeatedly, record the mean, the distribution of sample means

Standard Error of the Mean (SEM)

a measure of how well your sample mean estimates the true population mean

$SEM = \frac{\sigma}{\sqrt{n}}$ Standard deviation of the population / sample size 同时也是 Standard deviation of $\bar{x}$ sampling distribution

The central limit theorem

In general:

If we take n independent random variables from any distribution, and take their (normalised) sum, then that sum will tend towards a normal distribution with increasing n.

将 n 的 sum 作为一次独立的实验结果 (e.g. Rolling dice)

假设 $n = 2$

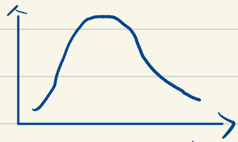
实验 1：2个骰子点数和为 6

实验 2：2个骰子点数和为 11

实验 3：2个骰子点数和为 3

⋮

点数和绘制成的 histogram：



normal distribution

当 $n = 1$ 时 ⟶ unique distribution

当 $n > 2$ 时 ⟶ 更加 tend to be normal distribution

For sample means:

Even if a population is not normally distributed, the sampling distribution (for large enough samples) will tend to be normal.

sample size ⟶ bigger

distribution of average ⟶ more normal and narrower

Procedure of hypothesis testing

(1) Formulate the null hypothesis, alternative hypothesis

(2) Design the experiment and collect data

(3) Summarise and describe the data

(4) Think about what you would expect if H_0 were true

(5) Could your data be explained by the null hypothesis

(6) Determine the probability of your data given H_0
(7) Interpret your p value and make a decision
(8) Be aware that hypothesis test are not perfect

Hypothesis Testing

{ Null hypothesis : nothing is happening
{ Alternative hypothesis : something is happening

Normal distribution :

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Types of tests

(1) Two-tailed : can be less or more (default)
(2) One-tailed : need to be justified
One bigger than the other (right tailed) ... smaller ... (left ...)

P-value :
probability of observing a value as or more extreme as the one you observed if the null hypothesis is true.

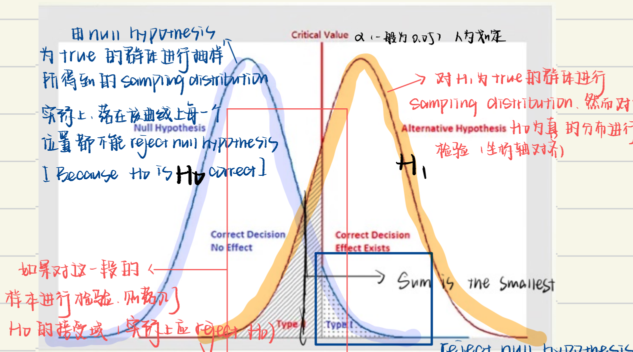
$p < 0.05$: reject the null hypothesis
$p > 0.05$: cannot reject the null hypothesis

	Something is happening	Nothing is happening
We found something	True positive (correct) probability $= 1-\beta$	False positive (Type 1 error) probability $= \alpha$
We found nothing	False negative (Type 2 error) probability $= \beta$	True negative (correct) probability $= 1-\alpha$

Example: Two Tailed

- Given: critical z values are ±1.96, α = 0.05

标准化的 normal distribution



两条曲线为H_0和H_1的总体分布 / 均值抽样分布 sampling distribution

If choosing α, we are choosing our level of protection against Type I error.

$\alpha = 0.05 = Pr$ { data provide significant evidence for H_A },
if H_0 is true ⟶ Type I error

↳ 研究者在进行假设试验时愿意接受的风险水平
（H_0为真情况下犯下 Type I error 的最大概率）

To compare 2 quantities:

(1) 2 independent samples

e.g. Happiness ratings of UoE students based in Edinburgh and in Haining

(2) 2 paired samples ⟶ take the difference, treated as 1 sample.

e.g. Satisfaction ratings of catering on ZJU International campus in 2018/19 & 2019/20 (same people)

(3) 1 sample, compare with a number

e.g. Average activity of PFC neurons with a baseline (4 Hz)

Actually compare the means depend on ⟶ { Effect size (r^2, Cohen's d)
Sample size (s.e.m. $= \sigma/\sqrt{n}$)

Cumulative Distribution Function (CDF)

effect size
⇓
两组(或多组)之间实际存在的有意义的差异

e.g. $cdf_{normal}(1.96) = 0.975$

R: pnorm(1.96)

Inverse, cdf value ⟶ qnorm(0.975)

可能由于 sample size 太小

We can have a fairly large effect size but no significant difference between the means or a very small effect size but a significant difference between the means

可能由于 sample size 过大

生成 simulation

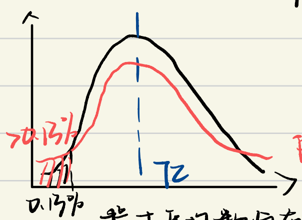
$\begin{cases} \text{rnorm(N, mean, sd)} & \text{已知分布和参数} \\ \text{sample(empirical\_distribution, N)} & \text{只有实际观测数据} \end{cases}$

Student's t distribution

已知新冠者自愈的平均时间是72小时、标准差为8小时 定值

现对100人进行临床试验　　总体标准差较难得到

发现痊愈时间为69.6小时



N(72, 0.8)

$S.E.M = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{100}} = 0.8$　　→ t分布

自由度99

样本平均数分布 H_0

t分布容错率更高　　自由度↑
接近正态分布↑

已知新冠者自愈的平均时间是72小时　　利用样本标准差替代整体标准差

现对100人进行临床试验

发现痊愈时间为69.6小时、样本标准差为8小时 变量

$t = \frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{n}}$　　t值代表偏离中心点多少个标准差

样本标准差　　样本量 > 30 默认可以用 normal distribution

双侧：≠　　单侧：>、<

p-value：默认 H_0 成立时事件发生的概率 → 犯弃真错误的实际概率

p-value < α：犯弃真错误的实际概率低于容忍程度

[实验者可以容忍弃真风险即可拒绝 H_0]

One sample：与确定的数值对比

e.g. Is the factory filling each bottle with a volume different from 500mL?

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

two sample：两组 sample 数据对比

e.g. Is the factory in Dublin filling each bottle with a volume different from the factory in Glasgow?

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}}} \qquad d.o.f = (n_1 - 1) + (n_2 - 1)$$

Equal variance — Student's t test

Unequal variance — Welch's t test

- Paired t-test: Same subjects measured at different times or under different conditions
- unpaired t-test: Different subjects ...

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

sd of the differences

Use the `t.test()` function in R to perform a *t*-test.

```
t.test(x, y = NULL,
alternative = c("two.sided", "less", "greater"),
mu = 0, paired = FALSE, var.equal = FALSE,
conf.level = 0.95, ...)
```

t.test(sample1, sample2, var.equal = FALSE)

1 sample `mu` = reference value	**2 sample** Must provide two vectors of values as `x` and `y` For Welch's t-test `var.equal = FALSE` ()
1-tailed `alternative = "greater"` or `"less"`	**2-tailed** `alternative = "two.sided"` (default)
Paired `paired = TRUE`	**Unpaired** `paired = FALSE` (default)

If not normally distributed:

- one sample: wilcox.test(x, mu = ...)
- two sample:
  - wilcox.test(x, y, paired = TRUE) — Wilcoxon signed-rank test
  - wilcox.test(x, y, paired = FALSE) — Mann-Whitney U test

Replicate: Discover the same result by designing and running the experiment from scratch: collect & analyse data

Reproduce: Draw the same conclusions from an existing dataset: analyse existing date