Testing of hypothesis

- Assumed that the underlying data satistics a distribution
  - Assume its normal
  - Central limit theorem applies

⊗ - Distribution Free tests
  - does not mean "assumption" free
  - Work for a lot of parameters

• Sign test ! - for median

- Inference for these methods
  - likelihood ratio tests
  - "power" - how good a test is
  - "convergence rate" - Assumptions

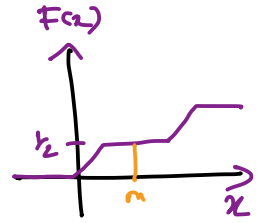
{ outside the scope of this course

Testing of Hypothesis

Sign test for the median:

- For random variable X, median m is given by

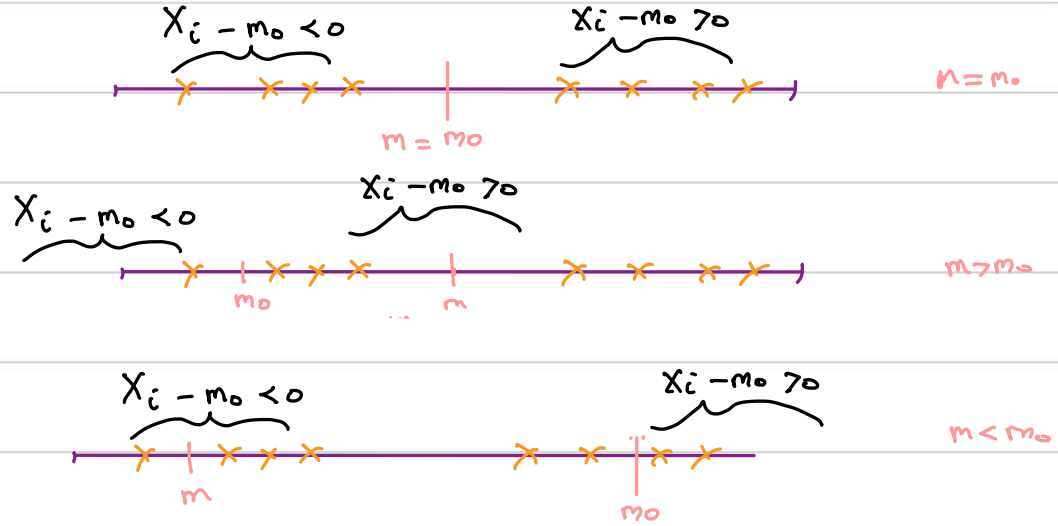
$$P(X \leq m) \quad = \frac{1}{2} = \quad P(X \geq m)$$



Note there could be many such m that satisfy the above.

- Suppose $X_1, X_2, \ldots, X_n$ are an i.i.d. sample from X
- Test:

$$H_0 := m = m_0 \qquad \text{versus} \qquad H_A = \begin{cases} m < m_0 \\ m > m_0 \\ m \neq m_0 \end{cases}$$



Design the test :

① Calculate $X_i - m_0$ for $1 \leq i \leq n$

② $S_+ = \#\{i : X_i - m_0 > 0\}$

$S_- = \#\{i : X_i - m_0 < 0\}$

Testing of Hypothesis

Sign test for the median:

- Calculate: $X_1 - m_0, X_2 - m_0, \ldots, X_n - m_0$.
- Determine:

$$S_- = \#\{i : \text{sign}(X_i - m_0) < 0\}$$

and

$$S^+ = \#\{i : \text{sign}(X_i - m_0) > 0\}$$

- Under $H_0 := m = m_0$

$$S_- \overset{d}{=} \text{Binomial}(n, \frac{1}{2}) \qquad \text{and } S^+ \overset{d}{=} \text{Binomial}(n, \frac{1}{2})$$

$H_0: m = m_0$

$S_+ \sim$ Binomial$(n, \frac{1}{2})$

$S_- \sim$ Binomial$(n, \frac{1}{2})$

- Get sample $\{x_1, \ldots, x_n\}$
- Set α
  - Calculate

$s_+ = \#\{i : x_i - m_0 > 0\}$

$s_- = \#\{i : x_i - m_0 < 0\}$

- H_A :

$m \neq m_0$	$m > m_0$	$m < m_0$
$S = \min(S_+, S_-)$	$\mathbb{P}(S_- \leq s_-)$	$\mathbb{P}(S_+ \leq s_+)$
$2\mathbb{P}(S \leq \min(s_+, s_-))$	if $< \alpha$	if $< \alpha$
if $< \alpha$	reject H_0	reject H_0
reject H_0		

Testing of Hypothesis

Sign test for the median:

```
>  x=c(9.4,13.4,15.6,16.2,16.4,16.8,18.1,18.7,18.9,19.1,19.3,20.1,
> S = x-22
> Sminus = sum(S<0)
> Splus = sum(S>0)
> binom.test(sum(S > 0), length(S), p=0.5, alternative="less")

        Exact binomial test

data:  sum(S > 0) and length(S)
number of successes = 5, number of trials = 20,
p-value = 0.02069
alternative hypothesis: true probability of success is less than 0
95 percent confidence interval:
 0.0000000 0.4555824
sample estimates:
probability of success
                  0.25
```

$H_0: m = 22$

15

5

$H_A: m < 22$

n

$=$

reject H_0 !

Disadvantage

- only uses the sign
- data may have more evidence

$H_0: m = 1$

$H_A: m = 10^6$

$m = -5$

$m = 10^6$

Testing of Hypothesis

Wilcoxon-Sign test for the median:

- Calculate: $X_1 - m_0, X_2 - m_0, \ldots, X_n - m_0$.
- Calculate: $\mid X_1 - m_0 \mid, \mid X_2 - m_0 \mid, \ldots, \mid X_n - m_0 \mid$.
- Calculate the rank R_i of $\mid X_i - m_0 \mid$ in

  $\mid X_1 - m_0 \mid, \mid X_2 - m_0 \mid, \ldots, \mid X_n - m_0 \mid$.
- Calculate the rank $Z_i = 1(X_i - m_0 > 0)$.
- Compute $W = \sum_{i=1}^{n} Z_i R_i$.

Under $H_0 : m = m_0$

Ex. Distribution of W ?

U

$U = \sum_{i=1}^{n} U_i$

$U_i \equiv \begin{cases} 0 & \frac{1}{2} \\ 1 & \frac{1}{2} \end{cases}$

Range $(W) = \{0, 1, \ldots, \frac{n(n+1)}{2}\}$



$W = 0$

W - Tables



$W = \frac{n(n+1)}{2}$

Ex: $E[W] = ?$ $\quad Var[W] = ?$

Data: 5, 3.9, 5.2, 5.5, 2.8, 6.1, 6.4
2.6, 1.7, 4.3

$H_0: m = 3.7$ $H_A: m \neq 3.7$

i	x_i	$x_i - 3.7$	$\lvert x_i - 3.7 \rvert$	R_i	Z_i
1	5	1.3	1.3	5	1
2	3.9	0.2	0.2	1	1
3	5.2	1.5	1.5	6	1
4	5.5	1.8	1.8	7	1
5	2.8	−0.9	0.9	3	0
6	6.1	2.4	2.4	9	1
7	6.4	2.7	2.7	10	1
8	2.6	−1.1	1.1	4	0
9	1.7	−2.0	2.0	8	0
10	4.3	0.6	0.6	2	1

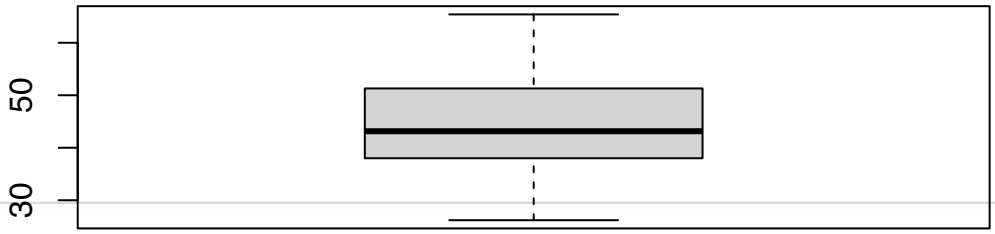
$$W = 5(1) + 1(1) + 6(1) + 7(1) + 9(1) + 10(1) + 2(1) = 40$$

– Use W-tables –

Testing of Hypothesis

Wilcoxon-Sign test for the median:

```
> x= c(35.5, 44.5, 39.8, 33.3, 51.4, 51.3, 30.5, 48.9, 42.1, 40.3,
+      46.8, 38.0, 40.1, 36.8, 39.3, 65.4, 42.6, 42.8, 59.8, 52.4, 26.2,
+      60.9, 45.6, 27.1, 47.3, 36.6, 55.6, 45.1, 52.2, 43.5)
> median(x)
[1] 43.15
> boxplot(x)
```



$H_0: m = 45$ $H_A: m \neq 45$

```
> wilcox.test(x, mu=45, alternative="two.sided")

        Wilcoxon signed rank exact test

data:  x
V = 200, p-value = 0.5158
alternative hypothesis: true location is not equal to 45
```

– don't reject at 5% l.o.s.

Two Sample-test (Welch approximation)

Let $n, m \geq 1$, $X_1, X_2, \ldots, X_n$ be i.i.d. Normal(μ_X,σ_1^2) and $Y_1, Y_2, \ldots, Y_m$ be i.i.d. Normal(μ_Y,σ_2^2).

Unequal Variance: $\sigma_1 \neq \sigma_2$ Test Statistic:

value unknown

$$T := \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_Y)}{S_{pooled}}$$

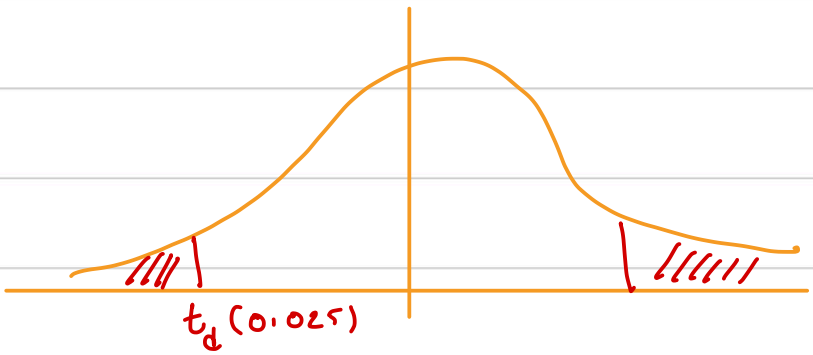
where

$$S^2_{pooled} = \frac{S_X^2}{n} + \frac{S_Y^2}{m}$$

Proof omitted

$$T \sim t_d \text{ with } d = \left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right)^2 \cdot \left(\frac{1}{n-1}\left(\frac{S_X^2}{n}\right)^2 + \frac{1}{m-1}\left(\frac{S_Y^2}{m}\right)^2\right)^{-1}.$$

$H_0: \mu_1 = \mu_2$ $\quad H_A = \begin{cases} \mu_1 > \mu_2 \\ \mu_1 \neq \mu_2 \\ \mu_1 < \mu_2 \end{cases}$



$H_A: \mu_1 > \mu_2$: $P(T > t_d(0.025)$ ....

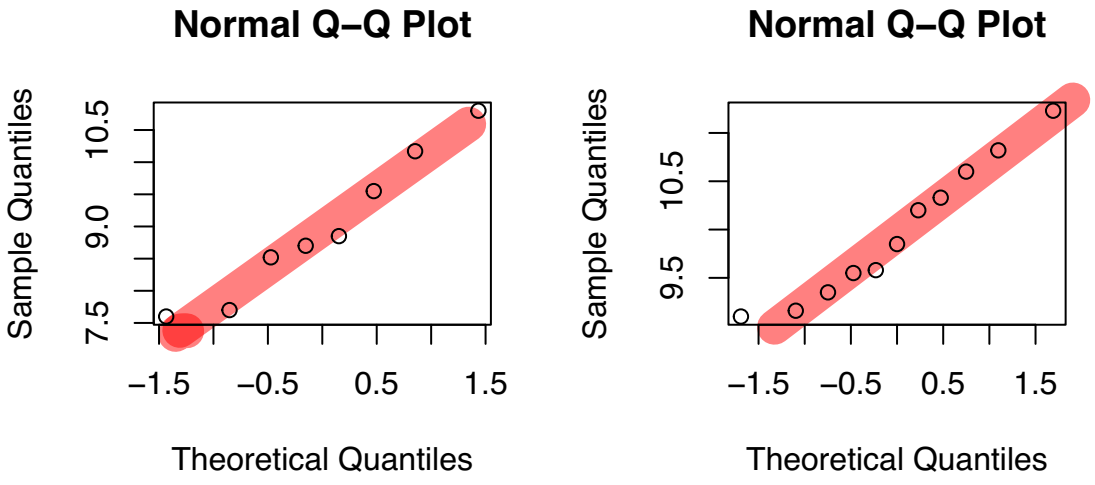
Testing for equality of mean in two populations

Example A class has two graders A and B to see if they are consistent in evaluation. We give them a random sample of exams from the class and ask them to grade them. Suppose their scores are converted to out of 15 and are:

Grader A: 8.52 8.70 9.55 10.80 8.85 10.17 7.60 7.70

Grader B: 10.20 9.55 9.35 9.85 9.16 10.60 11.23 9.10 9.58 10.82 10.33

```
> A = c(8.52,  8.70,  9.55, 10.80,  8.85, 10.17,  7.60,  7.70)
> B= c(10.20,  9.55,  9.35,  9.85,  9.16, 10.60, 11.23,  9.10,
+      9.58, 10.82, 10.33)
> par(mfrow=c(1,2))
> qqnorm(A)
> qqnorm(B)
```



Data is normal

Testing for equality of mean in two populations

Example A class has two graders A and B to see if they are consistent in evaluation. We give them a random sample of exams from the class and ask them to grade them. Suppose their scores are converted to out of 15 and re:

```
> A = c(8.52,  8.70,  9.55, 10.80,  8.85, 10.17,  7.60,  7.70)
> B= c(10.20,  9.55,  9.35,  9.85,  9.16, 10.60, 11.23,  9.10, 9.5
> t.test(A,B)

        Welch Two Sample t-test

data:  A and B
t = -2.1944, df = 10.97, p-value = 0.05064
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.988973371  0.003291553
sample estimates:
mean of x mean of y
 8.986250  9.979091
```

$H_0: \mu_1 = \mu_2$

$H_A: \mu_1 \neq \mu_2$

← Reject the null

Testing of Hypothesis

Consider a clinical trial designed to investigate the effectiveness of a new drug for a respiratory disese. A total of 10 participants are randomized to receive either the new drug or a placebo. Participants were asked to record the number of episodes of shortness of breath over a 1 week period following receipt of the assigned treatment. The data:

Placebo: 7 5 6 4 12

New Drug: 3 6 4 2 1

H_0: The two populations are equal

versus H_1: The two populations are not equal.

Mann – Whitney test

Placebo	Drug	order in Sample		Rank Placebo	Rank Drug
7	3		1		1
5	6		2		2
6	4		3		2
4	2	4	4	4.5	4.5
12	1	5		6	
		6	6	7.5	7.5
		7		9	
		12		10	

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 5.5 + \frac{5(6)}{2} - 37 = 3$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = 5.5 + \frac{5(6)}{2} - 18 = 22$$

$$U = \min(U_1, U_2)$$

Test Statistic:

n_i = number of samples in group $i \in \{1, 2\}$

R_i = sum of the ranks for group $i \in \{1, 2\}$.

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1, \qquad U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

and

$$U = \min(U_1, U_2).$$

$U_1 + U_2 = n_1 n_2$

If Surely (Medicine is good)

$n_1 = 5 \qquad n_2 = 5$

$R_1 = 6+7+8+9+10 = 40$

$R_2 = 1+2+3+4+5 = 15$

$U_1 = 5.5 + 15 - 40 = 0$

$U_2 = 5.5 + 15 - 15 = 25$

$U = 0$!

If Surely (Medicine is not needed.)

$n_1 = 5 \qquad n_2 = 5$

$R_1 = 2+4+6+8+10 = 30$

$R_2 = 1+3+5+7+9 = 25$

$U_1 = 5.5 + 15 - 30 = 10$

$U_2 = 5.5 + 15 - 25 = 15$

$U = 10$

Smaller values of U support Effectiveness of medicine

Testing of Hypothesis

Test Statistic:

n_i = number of samples in group $i \in \{1, 2\}$

R_i = sum of the ranks for group $i \in \{1, 2\}$.

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1, \qquad U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

and

$$U = \min(U_1, U_2).$$

- U - tables - available

Testing of Hypothesis

Consider a clinical trial designed to investigate the effectiveness of a new drug for a respiratory disese. A total of 10 participants are randomized to receive either the new drug or a placebo. Participants were asked to record the number of episodes of shortness of breath over a 1 week period following receipt of the assigned treatment. The data:

Placebo: 7 5 6 4 12

New Drug: 3 6 4 2 1

```
> Placebo=c(7, 5, 6, 4, 12)
> NewDrug=c(3, 6, 4, 2, 1)
> wilcox.test(NewDrug, Placebo)

        Wilcoxon rank sum test with continuity correction

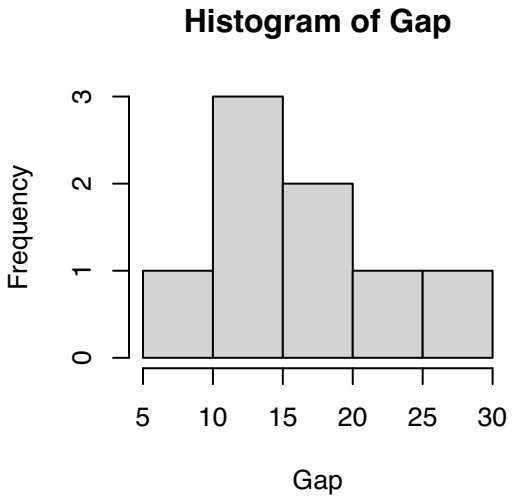
data:  NewDrug and Placebo
W = 3, p-value = 0.05855
alternative hypothesis: true location shift is not equal to 0
```

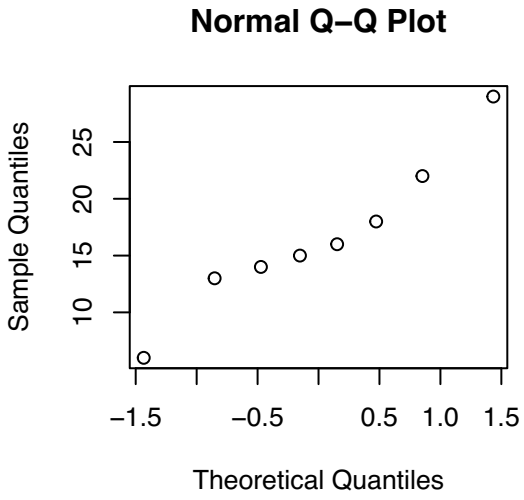
Testing of Hypothesis

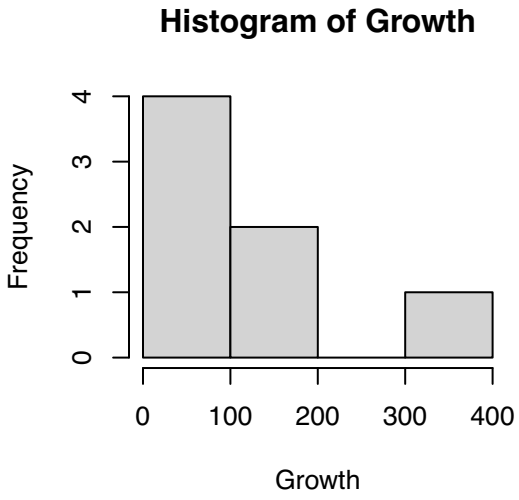
Example: Does soil respiration affect plant growth?

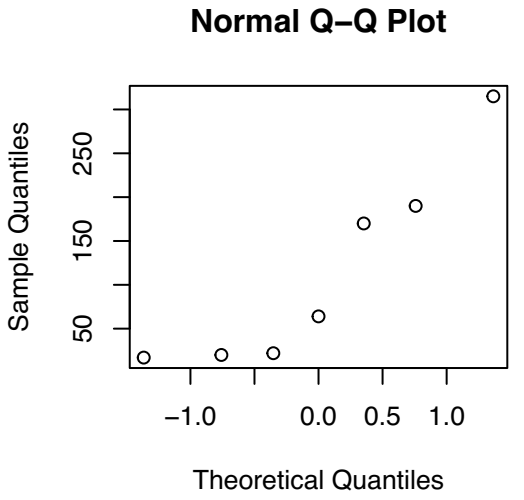
Soil cores taken from 2 locations in a forest: under an opening in the forest canopy ("gap") and at a nearby area under heavy tree growth ("growth"). Measured: amount of carbon dioxide given off by each soil core (mol CO_2 /g soil/hr). Data: *Gap* 22 29 13 16 15 18 14 6 and *Growth* 17 20 170 315 22 190 64

```
> Gap=c(22, 29, 13, 16, 15, 18, 14, 6)
> Growth=c(17, 20, 170,315, 22, 190, 64)
> par(mfrow=(c(2,2)))
> hist(Gap); qqnorm(Gap);hist(Growth);qqnorm(Growth)
```









Man Whitney Test

Example: Does soil respiration affect plant growth?

Soil cores taken from 2 locations in a forest: under an opening in the forest canopy ("gap") and at a nearby area under heavy tree growth ("growth"). Measured: amount of carbon dioxide given off by each soil core (mol CO_2 /g soil/hr). Data: *Gap* 22 29 13 16 15 18 14 6

and

Growth 17 20 170 315 22 190 64

```
> Gap=c(22, 29, 13, 16, 15, 18, 14, 6)
> Growth=c(17, 20, 170,315, 22, 190, 64)
> wilcox.test(Gap, Growth)

        Wilcoxon rank sum test with continuity correction


data:  Gap and Growth
W = 6.5, p-value = 0.015
alternative hypothesis: true location shift is not equal to 0
```

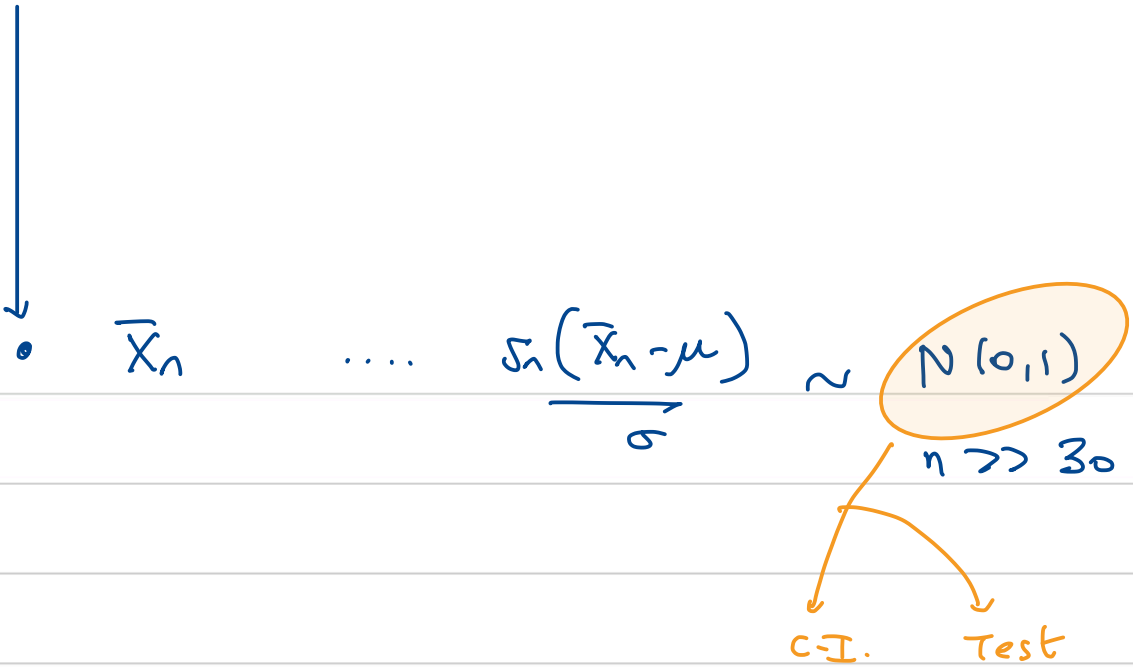
Bootstrap-Resampling Method (Effron)

Given $X_1, X_2, \ldots X_n$ i.i.d sample from a population $X \sim f(\cdot \mid \theta)$ with $\theta \in \Theta \subset \mathbb{R}^d$.

Point estimation: $\hat{\theta} = g(X_1, X_2, \ldots, X_n)$

Confidence intervals: based on information from the sampling distribution, including the standard error.

What if the underlying distribution is unknown?



- Treat the sample as Population
- Resample from it with replacement

Estimation, intervals

Bootstrap-Resampling Method

Suppose we want to estimate median of a population X $\hat{M}$

- Given $X_1, X_2, \ldots X_n$ i.i.d sample from a population X
- Sample with replacement B times from $X_1, X_2, \ldots X_n$
  - Compute sample median, M_i,for each trial $i = 1, \ldots B$. – each Bootstrap resample
  - Obtain an approximate distribution of the sample median.
- If we have the approximate distribution, we can find an estimate of the standard error of the sample median by finding the standard deviation of $M_1, \ldots M_B$.

- Create Histogram of $\{M_1, \ldots, M_B\}$
- Calculate its mean, variance
- intervals

Bootstrap-Resampling Method

Suppose we want to estimate parameter θ of a population X

- Given $X_1, X_2, \ldots X_n$ i.i.d sample from a population X. Suppose an estimate for θ is given by $\hat{\theta} = g(X_1, X_2, \ldots X_n)$
- Sample with replacement B times from $X_1, X_2, \ldots X_n$

$$X^*_{11}, X^*_{12}, X^*_{13}, \ldots, X^*_{1n}$$

$$\ldots$$

$$X^*_{B1}, X^*_{B2}, X^*_{B3}, \ldots, X^*_{Bn}$$

- Compute

$$\hat{\theta}^\star_i = g(X^\star_{i1}, X^\star_{i2}, \ldots X^\star_{in})$$

for $i = 1, \ldots B$.

- Using $\{\hat{\theta}^\star_1, \hat{\theta}^\star_2, \ldots, \hat{\theta}^\star_B\}$ to compute
  - Bootstrap-Bias

$$\frac{1}{B}\sum_{i=1}^{B} \hat{\theta}^\star_i - \hat{\theta} \quad \equiv \quad \text{“}\theta - \hat{\theta}\text{”}$$

  - Standard Error $(\hat{\theta})$

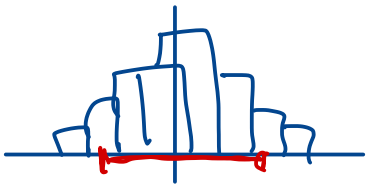
$$\sqrt{\frac{1}{B-1}\sum_{j=1}^{B}(\hat{\theta}^\star_j - \bar{\theta}^\star)^2}$$

  - Confidence Interval

    Use Standard Error $(\hat{\theta})$ around $\hat{\theta}$

    or

    If distribution of $\{\hat{\theta}^\star_1, \hat{\theta}^\star_2, \ldots, \hat{\theta}^\star_B\}$ is Normal



$$[\hat{q}^\star_{0.25}, \hat{q}^\star_{0.975}]$$

for 95% confidence interval.

Bootstrap-Resampling Method

Suppose we want to estimate parameter θ of a population X

- Given $X_1, X_2, \ldots X_n$ i.i.d sample from a population X. Suppose an estimate for θ is given by $\hat{\theta} = g(X_1, X_2, \ldots X_n)$
- Sample with replacement B times from $X_1, X_2, \ldots X_n$

$$X^*_{11}, X^*_{12}, X^*_{13}, \ldots, X^*_{1n}$$

$$\ldots$$

$$X^*_{B1}, X^*_{B2}, X^*_{B3}, \ldots, X^*_{Bn}$$

- Compute

$$\hat{\theta}^\star_i = g(X^\star_{i1}, X^\star_{i2}, \ldots X^\star_{in})$$

for $i = 1, \ldots B$.

- Using $\{\hat{\theta}^\star_1, \hat{\theta}^\star_2, \ldots, \hat{\theta}^\star_B\}$ to compute
  - Bootstrap-Bias

$$\frac{1}{B}\sum_{i=1}^{B} \hat{\theta}^\star_i - \hat{\theta}$$

  - Standard Error $(\hat{\theta})$

$$\sqrt{\frac{1}{B-1}\sum_{j=1}^{B}(\hat{\theta}^\star_j - \bar{\theta}^\star)^2}$$

  - Confidence Interval

    Use Standard Error $(\hat{\theta})$ around $\hat{\theta}$

    or

    If distribution of $\{\hat{\theta}^\star_1, \hat{\theta}^\star_2, \ldots, \hat{\theta}^\star_B\}$ is Normal

$$[\hat{q}^\star_{0.25}, \hat{q}^\star_{0.975}]$$

    for 95% confidence interval.