2014-05-05

Friday, May 09, 2014
9:14

中 F

Workshop on ANOVA & Nonparametrics

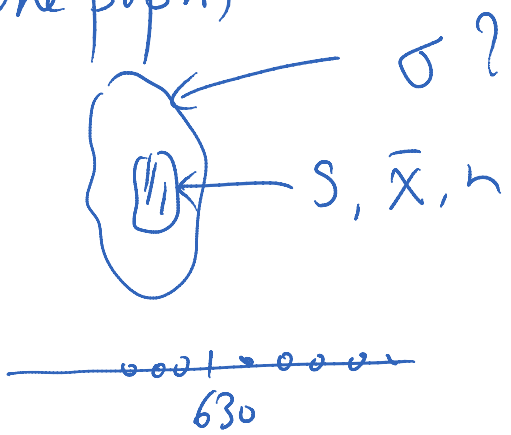
- Megastat

① ANOVA

1. Hypothesis testing (One pop'n)

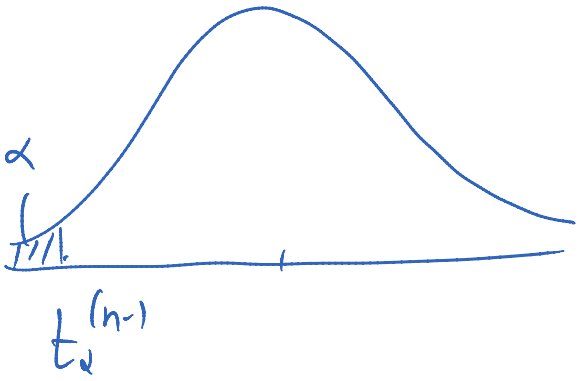
$H_0: \mu = \mu_0$ (630)

$H_a: \mu < \mu_0$



Test stat (σ unknown)

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}, \quad df = n-1$$



Ex.

→ Reject H_0

DeGroote
618
621
625
632
636
640
645
646

Hypothesis Test: Mean vs. Hypothesized Value					
	630.000	hypothesized value			
	632.875	mean DeGroote			
	10.723	std. dev.			
	3.791	std. error			
	8	n			
	7	df			
	0.76	t			
	.7635	p-value (one-tailed, lower)			

Don't reject H_0

2. HT (two popn's)

D n_1 ⬤ ⬤ R n_2

μ_1 μ_2

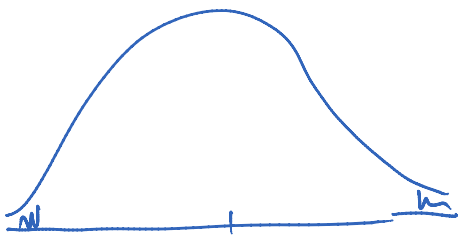
σ_1^2 σ_2^2 ← (equal!)

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}, \quad df = n_1 + n_2 - 2$$

$\mu_1 - \mu_2 = 0$



$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

Ex.

DeGroote	Rotman
618	615
621	642
625	630
632	632
636	610
640	622
645	636
646	644
	601

Hypothesis Test: Independent Groups (t-test, pooled variance)						
	DeGroote	Rotman				
	632.88	625.78	mean			
	10.72	14.79	std. dev.			
	8	9	n			
		15	df			
		7.097	difference (DeGroote - Rotman)			
		170.295	pooled variance			
		13.050	pooled std. dev.			
		6.341	standard error of difference			
		0	hypothesized difference			
		1.12	t			
		2806	p-value (two-tailed)			

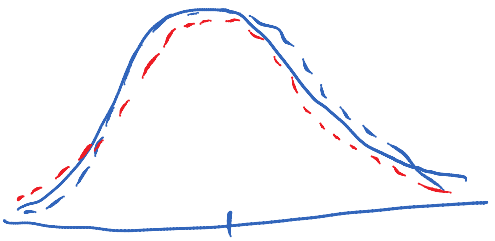
3. ANOVA

$H_0: \mu_1 = \mu_2 = \cdots = \mu_n$

H_{ai}: at least two differ

3 MBA prog's D R S

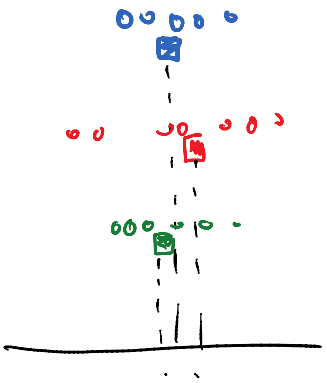
If all equal



D

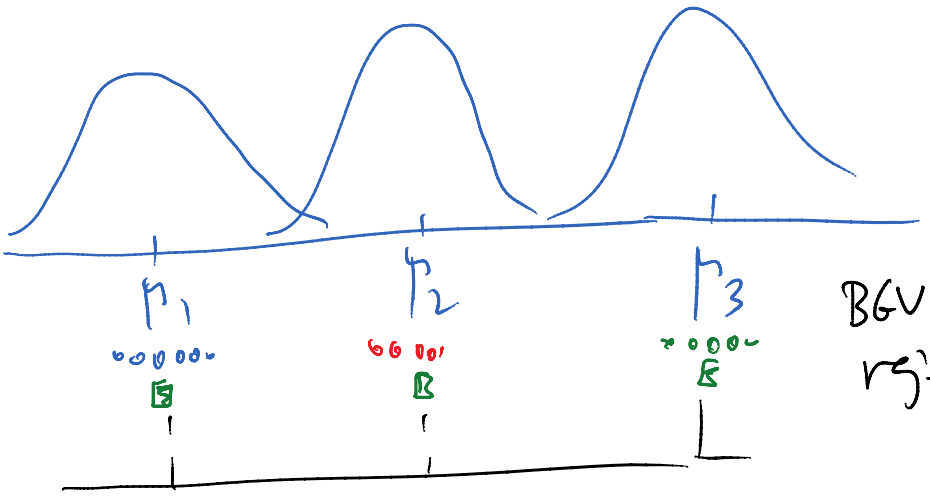
R

S



BGV < WGV

If all diff



BGV > WGV
reject

Compare Between group variability
with within

Ex

DeGroote	Rotman	Schulich
618	615	600
621	642	599
625	630	630
632	632	602
636	610	630

640	622	605
645	636	
646	644	
	610	

One factor ANOVA					
	Mean	*n*	*Std. Dev*		
	632.9	8	10.72	DeGroote	
	626.8	9	13.11	Rotman	
	611.0	6	14.86	Schulich	
	624.8	23	15.05	Total	
ANOVA table					
Source	*SS*	*df*	*MS*	*F*	*p-value*
Treatment	1,699.48	2	849.741	5.17	.0154
Error	3,284.43	20	164.222		
Total	4,983.91	22			
Post hoc analysis					
p-values for pairwise t-tests					
		Schulich	Rotman	DeGroote	
		611.0	626.8	632.9	
Schulich	611.0				
Rotman	626.8	.0300			
DeGroote	632.9	.0049	.3392		

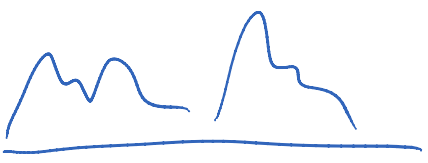
② Non-parametric methods

Assumptions in ANOVA

- normality
- variances equal

小心

What if



Nonparametrics

- One pop
  - Sign test
- Two pop
  - Mann/Whitney
- Three or more
  - Kruskall/Wallis

1. Sign test

$H_0: M_d = 630$

$H_a: M_d < 630$

π: proportion of students who have <630

$H_0: \pi = 0.5$

$H_a: \pi < 0.5$

X: total # who get <630

$X \sim Bin(n = 8, \pi = 0.5)$

618 621 625 | 632 636 640 645 646 $\quad n=8$

3 $\quad$ $M_d = 630$ $\quad$ 5

we expected $\mu_x = \pi(n) = .5(8) = 4$

$$p\text{-value} = Pr(X \le 3) = \sum_{x=0}^{3} \frac{8!}{x!(8-x)!} (0.5)^x (0.5)^{8-x} = .363$$

Ex.

DeGroote		Sign Test				
618			630	hypothesized value		
621			634	median DeGroote		
625			3	below		
632			0	equal		
636			5	above		
640						
645			8	n		
646						
				binomial		
			.3633	p-value (one-tailed, lower)		

2. Mann-Whitney Rank Sum test (Two pop)

Pop. 1 $f_1(x)$ n_1

Pop. 2 $f_2(x)$ n_2

Rank n_1+n_2 in ↗ order

T_1: Sum of ranks in n_1

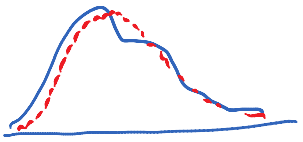
T_2: Sum of ranks in n_2

$$T = \begin{cases} T_1 \text{ if } n_1 \le n_2 \\ T_2 \text{ if } n_1 > n_2 \end{cases}$$

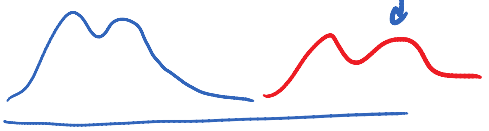
H_0: $f_1(x)$ and $f_2(x)$ are identical

H_a: " " " not "

11212211



111111 2222



Reject H_0 if $T \le T_L$ or $T \ge T_U$

in tables

Our ex.

n_2 \ n_1	8	
	T_L	T_U
9	54	90

will find $T = 83$

DeGroote	Rotman		Wilcoxon - Mann/Whitney Test			
618	615					
621	642			n	sum of ranks	
625	630			8	83	DeGroote
632	632			9	70	Rotman
636	610			17	153	total
640	622					
645	636				72.000	expected value
646	644				10.380	standard deviation

	601				1.060	z corrected for ties
					.2892	p-value (two-tailed)
			No.	Label	Data	Rank
			1	DeGroote	618	4
			2	DeGroote	621	5
			3	DeGroote	625	7
			4	DeGroote	632	9.5
			5	DeGroote	636	11.5
			6	DeGroote	640	13
			7	DeGroote	645	16
			8	DeGroote	646	17
			9	Rotman	615	3
			10	Rotman	642	14
			11	Rotman	630	8
			12	Rotman	632	9.5
			13	Rotman	610	2
			14	Rotman	622	6
			15	Rotman	636	11.5
			16	Rotman	644	15
			17	Rotman	601	1

3. Kruskall-Wallis test (nonparametric equivalent of ANOVA)

m popn's

no assumptions needed

D	R	S
⋮	⋮	⋮
n_1	n_2	n_3

H_0: $\eta_1 = \eta_2 = \eta_3$ ← medians

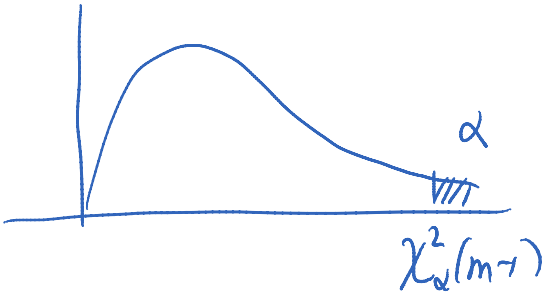
H: at least two differ

Procedure:

1. Rank data (ties assigned mid-value)
2. R_i : Sum of ranks in ith group, $i = 1, \ldots, m$
3. Test stat

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{m} \frac{R_i^2}{n_i} - 3(n+1)$$

H approx $\chi^2(m-1)$



Ex.

DeGroote	Rotman	Schulich	Kruskal-Wallis Test					
618	615	600						
621	642	599		*Median*	*n*	*Avg. Rank*		
625	630	630		634.00	8	15.63	DeGroote	
632	632	602		630.00	9	12.78	Rotman	
636	610	630		603.50	6	6.00	Schulich	
640	622	605		630.00	23		Total	
645	636							
646	644					7.124	H (corrected for ties)	
	610					2	d.f.	
						.028	p-value	
					multiple comparison values for avg. ranks			
					8.29	(.05)	10.17	(.01)

			No.	Label	Data	Rank		
			1	DeGroote	618	8		
			2	DeGroote	621	9		
			3	DeGroote	625	11		
			4	DeGroote	632	15.5		
			5	DeGroote	636	17.5		
			6	DeGroote	640	19		
			7	DeGroote	645	22		
			8	DeGroote	646	23		
			9	Rotman	615	7		
			10	Rotman	642	20		
			11	Rotman	630	13		
			12	Rotman	632	15.5		
			13	Rotman	610	5.5		
			14	Rotman	622	10		
			15	Rotman	636	17.5		
			16	Rotman	644	21		
			17	Rotman	610	5.5		
			18	Schulich	600	2		
			19	Schulich	599	1		
			20	Schulich	630	13		
			21	Schulich	602	3		
			22	Schulich	630	13		
			23	Schulich	605	4		