

Use of SAS
August, 2012

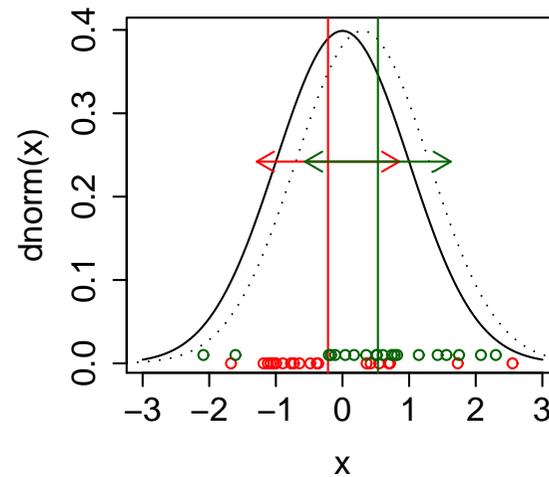
3. *t*-test and One-way ANOVA

Comparing two samples

Two groups,

x_{11}, \dots, x_{1n_1}

x_{21}, \dots, x_{2n_2}



$N(\mu_1, \sigma_1^2)$

$N(\mu_2, \sigma_2^2)$

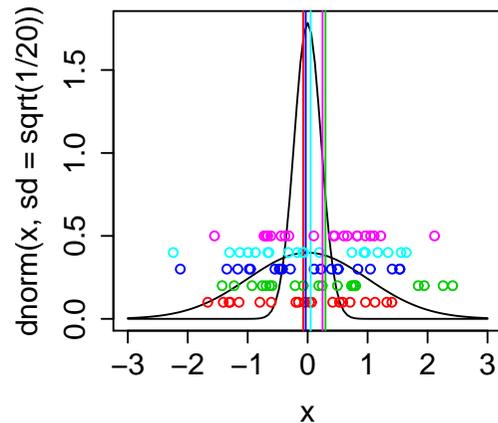
(\bar{x}_1, s_1^2)

(\bar{x}_2, s_2^2)

Significant difference between \bar{x}_1 and \bar{x}_2 ?

Null hypothesis $H_0 : \mu_1 = \mu_2$

Two-sample t -test



$$\text{SEM} = s/\sqrt{n}$$

Standard error of mean

$$\text{SEDM} = \sqrt{\text{SEM}_1^2 + \text{SEM}_2^2}$$

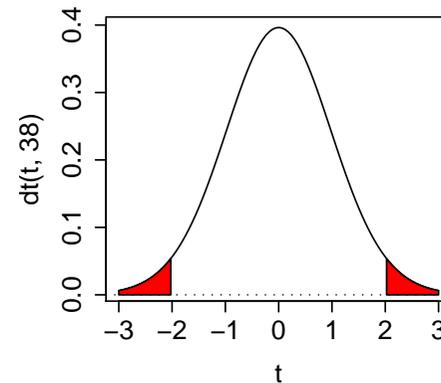
Standard error of difference of means

$$t = \frac{\bar{x}_2 - \bar{x}_1}{\text{SEDM}}$$

test-statistic t measures disagreement between data and H_0

The p -value

$$t = \frac{\bar{x}_2 - \bar{x}_1}{\text{SEDM}}$$



t : measures disagreement between data and H_0

If H_0 is true: distribution of t is symmetric around 0

p : the prob. of having observed a more extreme t -value

if $p < 5\%$: H_0 is rejected

Two tests: same or different variances?

Assume $\sigma_1^2 = \sigma_2^2$ before testing $\mu_1 = \mu_2$?

Same variance:

- Natural under null hypothesis (same distributions)
- Nice theory.

Separate variances:

- Looks specifically for difference in means
- Approximative theory.

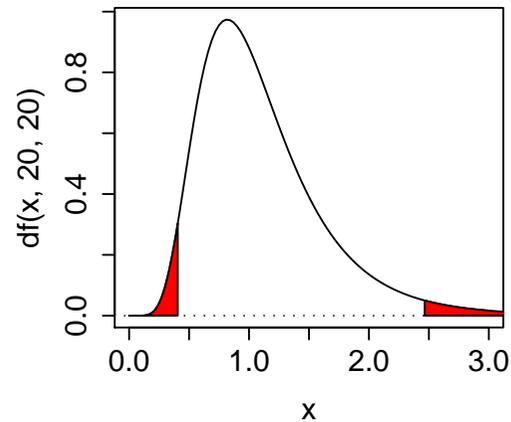
Test for same variance

Test statistic

$$F = s_1^2/s_2^2$$

F-distribution with (f_1, f_2) degrees of freedom

$$f_1 = n_1 - 1 \quad f_2 = n_2 - 1$$



t-test in SAS

Example from Anders Juul

IDEA: Compare men and women with respect to $\log(\text{SIGF1})$ for 20–30 year olds.

1. Open dataset: SET statement;
2. Compute $\text{LSIGF1} = \text{LOG}(\text{SIGF1})$
3. Start PROC TTEST
4. Use WHERE to select subgroup
5. Specify dependent variable
6. — and classification
7. Do not forget RUN

Code for the *t*-test

```
libname juul 'p:\sas\data\juul';  
data juul;  
    set juul.juul2;  
    lsigf1=log(sigf1);  
run;  
proc ttest data=juul;  
    where age > 20 and age < 30;  
    var lsigf1;  
    class sexnr;  
run;
```

Variable: lsigf1

sexnr	N	Mean	Std Dev	Std Err	Minimum	Maximum
1	23	5.5992	0.1981	0.0413	5.2470	6.0355
2	18	5.6306	0.2551	0.0601	5.1059	6.0064
Diff (1-2)		-0.0314	0.2247	0.0707		

sexnr	Method	Mean	95% CL Mean	Std Dev
1		5.5992	5.5136 5.6849	0.1981
2		5.6306	5.5038 5.7575	0.2551
Diff (1-2)	Pooled	-0.0314	-0.1745 0.1116	0.2247
Diff (1-2)	Satterthwaite	-0.0314	-0.1801 0.1173	

sexnr	Method	95% CL Std Dev
1		0.1532 0.2804
2		0.1914 0.3825
Diff (1-2)	Pooled	0.1841 0.2886
Diff (1-2)	Satterthwaite	

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	39	-0.44	0.6594
Satterthwaite	Unequal	31.422	-0.43	0.6697

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	17	22	1.66	0.2629

Exercise: t -test

Consider again the Juul data with variables

- Age (years)
- Height (cm)
- Menarche (No/Yes: 1/2)
- Sex (M/F: 1/2)
- Serum IGF1, growth hormone ($\mu\text{g/ml}$)
- Tanner stage (1–5)
- Testis volume (ml)
- Weight (kg)

Here the main aim is to compare the IGF1-level in boys and girls *above the age of 5 years*.

1. For each Tanner-stage, test if the IGF1-level is the same in boys and girls. The distribution of IGF1 seems to be skewed, but $\log(\text{IGF1})$ can be assumed to follow a normal distribution.

One-way ANOVA

Comparing more than two groups

$$\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k \quad s_1, s_2, \dots, s_k$$

Joint test for any differences between the groups.

Why not just pairwise t-tests?

MASS SIGNIFICANCE
LOSS OF OVERVIEW

The fewer tests, the better.

Notation and Models

x_{ij} observation no. j in group no. i
(i.e., x_{35} the 5th observation in group 3)

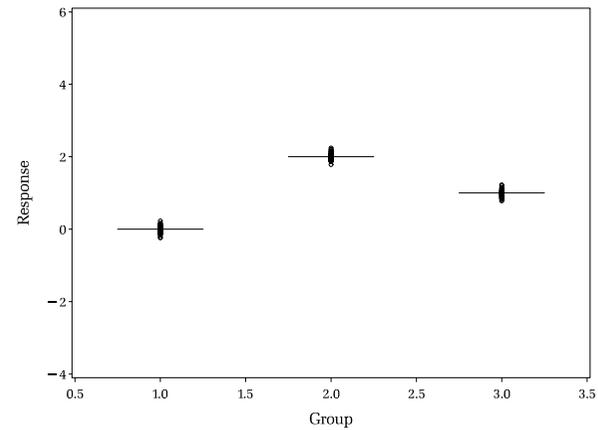
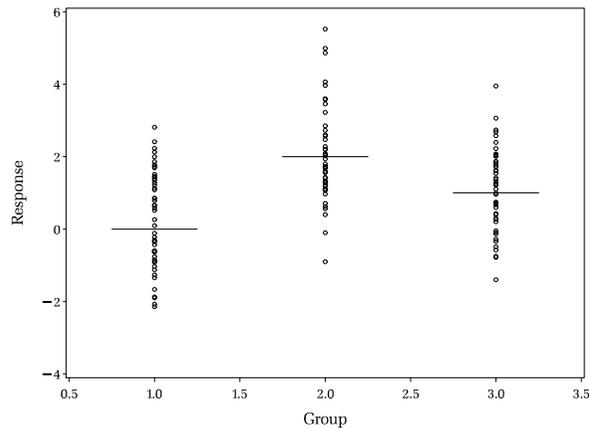
The model

$$X_{ij} = \mu_i + \epsilon_{ij} \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

The hypothesis of no differences between groups

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

Variation within and between groups



Main idea:

If the variation between group means is large compared to the variation within groups, it is a sign that the hypothesis is wrong.

Sums of squares

Variation (**W**ithin) groups: $SSD_W = \sum_i \sum_j (x_{ij} - \bar{x}_i)^2$

\bar{x}_i mean for group i

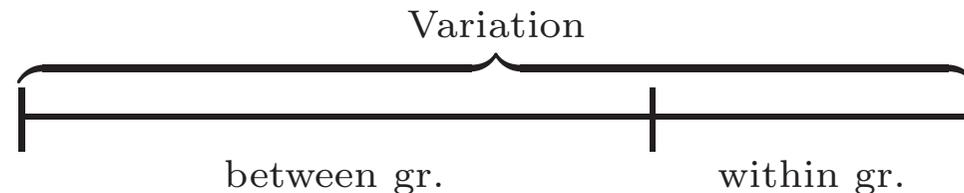
Variation (**B**etween) groups: $SSD_B = \sum_i \sum_j (\bar{x}_i - \bar{x}_.)^2$

$\bar{x}_.$ total (grand) mean

Can be mathematically proven that

$$SSD_B + SSD_W = SSD_{\text{total}} = \sum_i \sum_j (x_{ij} - \bar{x}_.)^2$$

The model (grouping) *explains* part of the variation

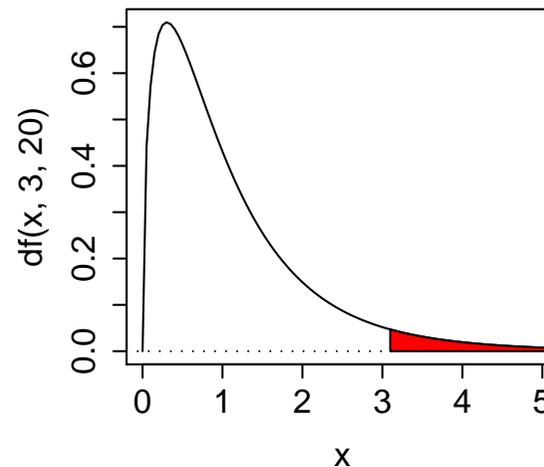


F-test for identical group means

We reject the hypothesis if the variation between groups is large compared to the variation within groups. Consider

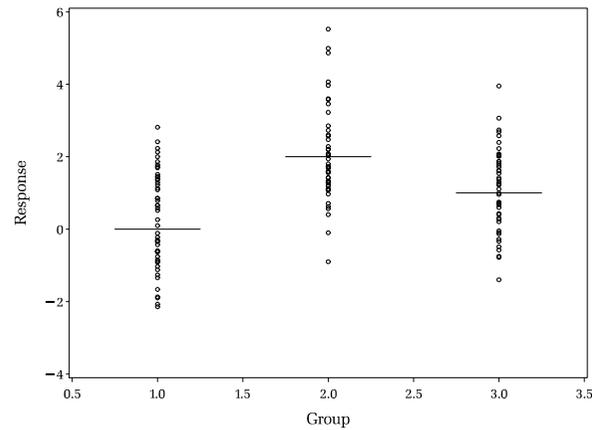
$$F = [\text{SSD}_B / (k - 1)] / [\text{SSD}_W / (N - k)]$$

If group differences are coincidental then F follows an F -distribution:



If F too large: Reject the hypothesis that the groups are identical.

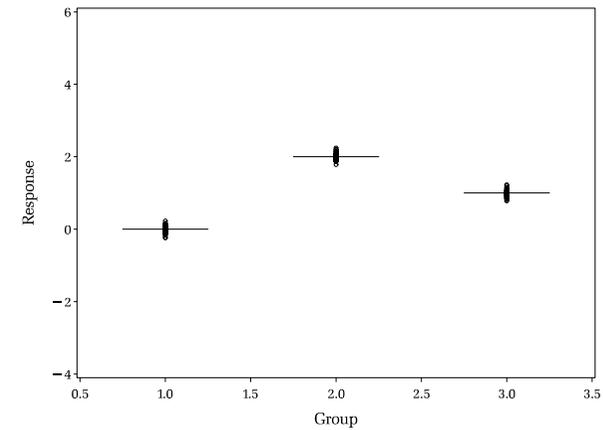
Testing for identical group means



high variation within grp.

F is small

H_0 is *not* rejected



small variation within grp.

F is large

H_0 is rejected

One-way ANOVA in SAS

IDEA: Compare boys in different Tanner stage with respect to their log SIGF1

1. This time generate a new data set, juulboys
2. Select: $\text{SEXNR} = 1, \text{AGE} < 20$
3. Use GLM
4. MODEL statement: What is described by what?
5. Remember to say that **tanner** is a grouping (CLASS)

Code for ANOVA

```
libname juul 'p:\sas\data\juul';  
data juulboys;  
    set juul.juul2;  
    lsigf1 = log(sigf1);  
    if sexnr = 1 and 0 < age < 20;  
run;  
proc glm data=juulboys;  
    class tanner;  
    model lsigf1 = tanner / solution;  
run;
```

(PROC ANOVA can also be used)

Output

The GLM Procedure

Class Level Information

Class	Levels	Values
tanner	5	1 2 3 4 5

Number of Observations Read 546

Number of Observations Used 400

The GLM Procedure

Dependent Variable: lsigf1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	91.3807356	22.8451839	117.57	<.0001
Error	395	76.7537608	0.1943133		
Corrected Total	399	168.1344964			

R-Square	Coeff Var	Root MSE	lsigf1 Mean
----------	-----------	----------	-------------

0.543498 7.836512 0.440810 5.625077

Source	DF	Type I SS	Mean Square	F Value	Pr > F
tanner	4	91.38073563	22.84518391	117.57	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
tanner	4	91.38073563	22.84518391	117.57	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	6.123871632 B	0.04110573	148.98	<.0001
tanner 1	-0.984254861 B	0.05197816	-18.94	<.0001
tanner 2	-0.342559094 B	0.08248145	-4.15	<.0001
tanner 3	-0.045918085 B	0.10068806	-0.46	0.6486
tanner 4	0.110391165 B	0.08810206	1.25	0.2109
tanner 5	0.000000000 B	.	.	.

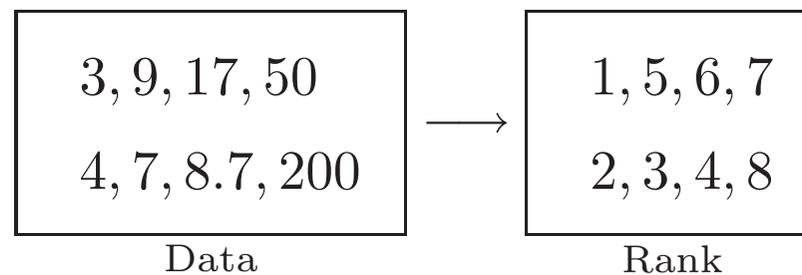
NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

Nonparametric tests

Mann-Whitney test (alias Wilcoxon)

Kruskal-Wallis test

Idea: “t-test” or “ANOVA” on *ranks*



Distribution is (in principle) known under null hypothesis. Does not depend on data following a Normal distribution.

Other “scores” than ranks can also be used

Nonparametric tests in SAS

- ```
proc npar1way wilcoxon data=juul.juul2;
 where sexnr = 1 and 0 < age < 20;
 var sigf1;
 class tanner;
run;
```
- (Mann-Whitney is obtained if there are only two groups to compare)
- It is the `wilcoxon` option that select rank scores and thus the Kruskal-Wallis/Mann-Whitney test, see manual for alternatives.

# Output

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable sigf1  
Classified by Variable tanner

| tanner | N   | Sum of<br>Scores | Expected<br>Under H0 | Std Dev<br>Under H0 | Mean<br>Score |
|--------|-----|------------------|----------------------|---------------------|---------------|
| 1      | 192 | 20758.00         | 38496.00             | 1155.20917          | 108.114583    |
| 2      | 38  | 8222.00          | 7619.00              | 677.99178           | 216.368421    |
| 3      | 23  | 6569.50          | 4611.50              | 538.28582           | 285.630435    |
| 4      | 32  | 10387.00         | 6416.00              | 627.30283           | 324.593750    |
| 5      | 115 | 34263.50         | 23057.50             | 1046.52522          | 297.943478    |

Average scores were used for ties.

Kruskal-Wallis Test

|                 |          |
|-----------------|----------|
| Chi-Square      | 254.3465 |
| DF              | 4        |
| Pr > Chi-Square | <.0001   |