

STA303/1002 - Methods of Data Analysis II

(Week 03 lecture note)

Wei (Becky) Lin

Jan. 17/19, 2017



Review: Topics learned last week

- t-test
 - one sample t-test
 - two sample t-test (could use lm(), anova(), t.test() in R)
 - paired t-test
- One-way ANOVA: Comparing More Than Two Means (a factor with more than two levels)
 - Assumption: $Y_{ij} \sim N(\mu_i, \sigma^2)$, i=group indicator=1, ..., r, j=observations for i-th group (1, ..., n_i), n=total sample size= $n_1 + \dots + n_r$
 - Want to test whether we have equal population means or not

$$H_0 : \mu_1 = \dots = \mu_r, \quad H_a : \exists i \neq j \text{ s.t. } \mu_i \neq \mu_j$$

- Test statistic

$$F^* = \frac{MST}{MSE} = \frac{SS_T/(r-1)}{SS_E/(n-r)} \sim F_{r-1, n-r} \text{ under } H_0$$

- Decision: rejecting H_0 at significance level α if F^* is too big, that is, we reject H_0 if

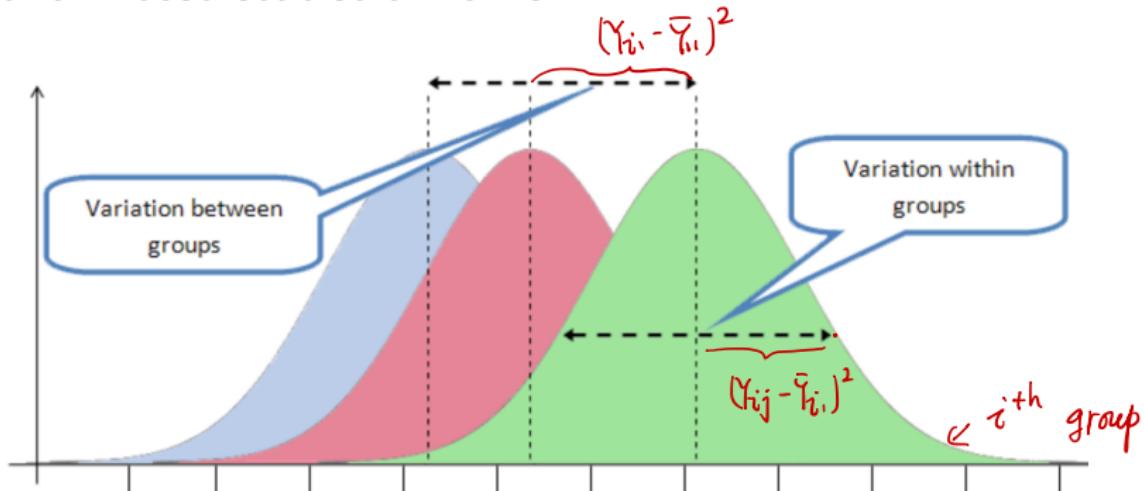
$$F^* > F_{1-\alpha, r-1, n-r} \quad \leftarrow \text{one-sided test}$$

Why the F test statistic works?

$$F^* = \frac{MST}{MSE} = \frac{SS_T/(r-1)}{SS_E/(n-r)} \sim \underbrace{F_{r-1, n-r} \text{ under } H_0}_{\text{under } H_0}$$

- $SS = \sum_{i=1}^r \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2$ defines the total sum of squares.
- $SSE = SS_W = \sum_{i=1}^r \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2 = \sum_{i=1}^r (n_i - 1) S_i^2$, error sum of squares (or the sum of squares within groups). This captures the variation within a group (help to estimate the σ^2 in the model assumed).
 - $E(SSE/(n-r)) = E(MSE) = \sigma^2$
- $SST = SS_B = \sum_{i=1}^r \sum_{j=1}^{n_i} (\bar{Y}_{i.} - \bar{Y}_{..})^2 = \sum_{i=1}^r n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2$, treatment sum of squares (or the sum of squares between groups). This measures the variation among groups (group means around the grand mean).
 - $E(SST/(r-1)) = \sigma^2 + \underbrace{\frac{1}{r-1} \sum_{i=1}^r n_i (\mu_i - \bar{\mu})^2}_{> 0 \text{ if } \mu_i \neq \mu_j \forall i \neq j; \text{ or } = 0} = \text{error + treatment,}$ under H_0 (equal means), $E(SST/(r-1)) = \sigma^2$, otherwise
 $E(SST/(r-1)) > \sigma^2$

Why the F test statistic works?



$$F^* = \frac{\underline{MST}}{\underline{MSE}} = \frac{\text{treatment+error}}{\text{error}} = \frac{SS_T/(r-1)}{SS_E/(n-r)} \sim F_{r-1, n-r} \text{ under } H_0$$

- If H_0 is true, $F^* \approx 1$, we don't have evidence to reject H_0 .
- If H_0 is false, $MST > MSE$, F^* increases. So large values of observed F^* are evidence against H_0 , and we test H_0 using a one-tailed test.

One-way ANOVA Table

Source	df	SS	Mean Square	F
Treatments	r-1	SS _T	MST=SS _T /(r-1)	F=MST/MSE
Residual	n-r	SS _E	MSE	
Total	n-1	SS		

Model Equations for one-way ANOVA

- Fixed effect model: $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$, $\rightarrow Y_{ij} \sim N(\underbrace{\mu + \alpha_i}_{= \mu_i}, \epsilon_{ij})$
 $(i = 1, 2, \dots, r, j = 1, 2, \dots, n_i)$
- $\epsilon_{ij} \sim N(0, \sigma^2)$, equal variance across all groups i.
- α_i is the main effect for the i-th treatment.

$$\hookrightarrow \alpha_i = \mu_i - \mu$$

Bartlett Test of Homogeneity of Variances

- Recall we learned how to test equal variances for two groups last week, R function is `var.test()`. Now we consider the same testing but the number of groups is more than two.
- The H_0 and H_a in the Bartlett test are defined as:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_r^2$$

H_a : At least one σ_i^2 is not equal to the others.

- Test statistics:

$$T = \frac{(n - r) \ln S_p^2 - \sum_{i=1}^r (n_i - 1) \ln S_i^2}{1 + \frac{1}{3(3-1)} (\sum_{i=1}^r 1/(n_i - 1) - 1/(n - r))} \sim \chi^2_{r-1} \text{ under } H_0$$

where S_p^2 is the pooled variance $S_p^2 = \sum_i^r (n_i - 1) S_i^2 / (n - r)$.

- Decision: we reject H_0 if

$$T_{obs} > \chi^2_{1-\alpha, r-1}$$

- R function: `bartlett.test()`

Two types of factor

- **Fixed effect:** data has been gathered from **all the levels of the factor that are of interest.**
 - Example: The purpose of an experiment is to compare the effects of three specific dosages of a drug on the response. **Dosage** is the factor; the three specific dosages in the experiment are the levels; there is no intent to say anything about other dosages.
- **Random effect:** the factor has many possible levels, interest is in all possible levels, but only a random sample of levels is included in the data
 - Example: The purpose of another experiment is to study the effect of machine operator on the quality final product. The researcher selects a random sample of operators from the large number of operators at the various facilities that manufacture the widgets. The factor is **operator**.

An excellent discussion paper about ANOVA is by Andrew Gelman, **Analysis of Variance - Why it is more important than ever**, 2005, Vol. 33, No. 1, 1-53.

<http://www.stat.columbia.edu/~gelman/research/published/AOS259.pdf>

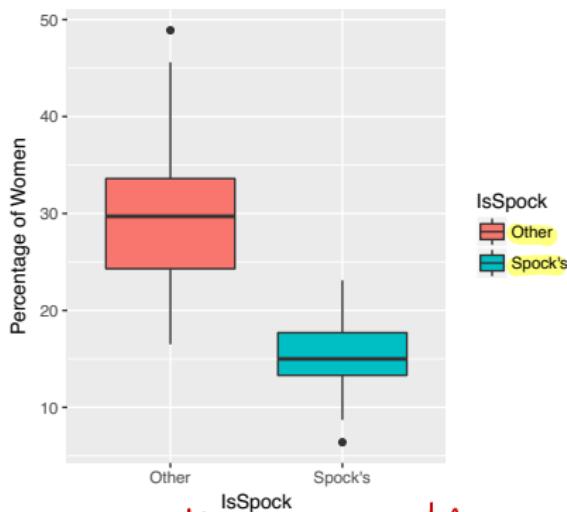
One-way ANOVA Example Spock Conspiracy Trial

Spock Conspiracy Trial

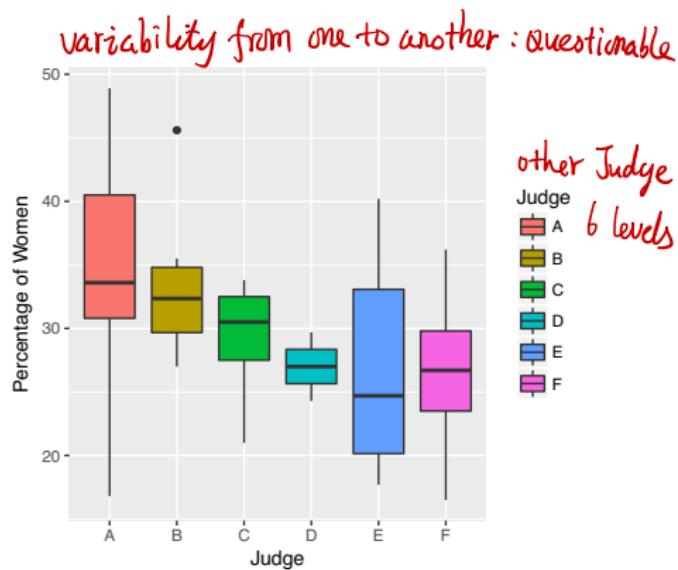
- Benjamin Spock
 - Best-selling American paediatrician, author.
 - Generally well-liked by women due to his books.
- In 1968, Dr. Spock was tried in Boston on charges of conspiring to violate the Selective Service Act by encouraging young men to resist being drafted into military service for Vietnam.
 - The defence in the case challenged the method of jury selection claiming that women were underrepresented.
 - Defence argued bias in jury selection
 - Jury selection process
 - 300 names are selected at random from eligible citizens.
 - Judge's office selects 35-200 "at random", called venire.
 - An actual jury is selected from the venire in a nonrandom process
 - Spock's judge had a history of very few women on his venires
 - There was only 1 woman on Spock's venire, and she was released by the prosecution.

Spock Conspiracy Trial

- Questions of interest
 - Q1: Is there statistical evidence that women are under-represented on Spock's judge's venires?
 - Q2: Is there evidence of a difference between any of the other six judges in the district?
- Visualization of both cases



Visually, significant diff.



Spock Conspiracy Trial

Two sample t-test : Q1

- To answer the first question, we'll use the two-sample t-test.
- Assumptions: $Y_{ij} \sim N(\mu_i, \sigma_i^2)$
- Checking equal variances for two groups: `var.test()`, `p.val = 0.2482`.
- Normality checking: looks fine for both groups. `qqnorm()`; `qqline()`.
- Hypothesis testing using `t.test(onlyspock,notspock,var.equal=T)`

$$H_0 : \mu_{Spock} - \mu_{other} = 0, \quad H_a : \mu_{Spock} - \mu_{other} \neq 0$$

```
##  
## Two Sample t-test  
##  
## data: onlyspock$Percent and notspock$Percent  
## t = -5.6697, df = 44, p-value = 1.03e-06 → very strong evidence (p<0.001)  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -20.155294 -9.584045 → 0 & CI  
## sample estimates:  
## mean of x mean of y  
## 14.62222 29.49189
```

Spock Conspiracy Trial

Two sample t-test : Q1

R skill for running two sample t test

- ① • Split data into two samples, run `t.test()` on them.
- ② • Create a binary indicator variable to differentiate Spock's and other, then run `t.test()`

```
library(Sleuth2)
jury = case0502 # 46 observations on two variables (Percent, Judge)

# Split data into two subgroups
notspock <- subset(jury, Judge != "Spock's")
onlyspock <- subset(jury, Judge == "Spock's")
① t.test(onlyspock$Percent,notspock$Percent,var.equal = T)

# Or create a binary indicator variable
jury$IsSpock=as.factor(ifelse(jury$Judge=="Spock's","Spock's","Other"))
② t.test(Percent~IsSpock,var.equal=T,data=jury)
```

Spock Conspiracy Trial

ANOVA approach : Q1

$$x = \text{ifelse}(\text{conditions}, A, B) \Rightarrow x = \begin{cases} A, & \text{cond} = T \\ B, & \text{cond} = F \end{cases}$$

```
library(Sleuth2)
jury = case0502 # 46 observations on two variables (Percent, Judge:7 levels)
jury$IsSpock=as.factor(ifelse(jury$Judge=="Spock's","Spock's","Other"))

with(jury, tapply(Percent, IsSpock, mean))
```

Υ factor FUN: mean, sum, median, sd, var, ---

```
##      Other   Spock's
## 29.49189 14.62222
```

two-sample t-test

statistic

```
summary(aov(Percent~IsSpock, data=jury))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)						
## IsSpock	1	1601	1600.6	32.15	1.03e-06 ***						
## Residuals	44	2191	49.8	"							
## ---				$(-5.6697)^2 = (t^*)^2$							
## Signif. codes:	0	'***'	0.001	'**'	0.01	'*'	0.05	'. '	0.1	' '	1

$$F^* = (t^*)^2$$

Compare with
output in slide 11

Spock Conspiracy Trial

ANOVA approach : Q2

- Assumptions: $Y_{ij} \sim_{\perp} N(\mu_i, \sigma_i^2)$, i is the other judge indicator (A-F)
- Null and alternative Hypotheses

$$H_0 : \mu_A = \mu_B = \dots = \mu_F = \mu, \quad H_a : \exists \text{ at least one } i, \text{ s.t. } \mu_i \neq \mu$$

```
library(Sleuth2)
jury = case0502 # 46 observations on two variables (Percent, Judge:7 levels)
with(jury, tapply(Percent, Judge, mean) )
```

```
## Spock's      A      B      C      D      E      F
## 14.62222 34.12000 33.61667 29.10000 27.00000 26.96667 26.80000
```

```
summary(aov(Percent~Judge, data=subset(jury, !(Judge=="Spock's"))))
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Judge       5  326.5   65.29   1.218  0.324
## Residuals  31 1661.3   53.59
```

```
# check normality : looks fine
# plot(lm(Percent~Judge, data=subset(jury, !(Judge=="Spock's"))), which=2)
```

Spock Conspiracy Trial

ANOVA approach : Q2

Checking whether variances in each of the groups (samples) are the same or not using **bartlett.test()**

```
library(Sleuth2)
jury = case0502 # 46 observations on two variables (Percent, Judge:7 levels)
bartlett.test(Percent~Judge, data=subset(jury, !(Judge=="Spock's")))
##                                     For other Judges
##   Bartlett test of homogeneity of variances
##   data: Percent by Judge
##   Bartlett's K-squared = 6.3125, df = 5, p-value = 0.277
```

Multiple Comparison (All Pairwise Comparison)

Why use ANOVA instead of multiple t-tests?

- If you are comparing means between more than two groups, why not just do **multiple two sample t-tests**?
 - One problem with this approach is the **increasing number of tests** as the number of groups increases
- The **advantage of using ANOVA over multiple t-tests is that the ANOVA F-test will identify if any of the group means are significantly different from at least one of the other group means with a single test, regardless of the number of groups being compared at specified significance level α .**

Multiple comparison

- Are all the means equal? ANOVA (analysis of variance) answers this question.
- A significant ANOVA F-test is evidence that not all group means are equal but it does not identify where the differences exist.
- Methods used to find group differences after the ANOVA null hypothesis has been rejected are called post-hoc tests.
 - Post-hoc is Latin for “after-this”
- Post-hoc multiple comparisons should only be done when the ANOVA F-test is significant.
- Multiple comparison, also called
 - Simultaneous testing
 - Multiple testing
 - Type I error inflation.

The problem with multiple t-test

- The probability of marking Type I error increases as the number of tests increase.
- $\alpha = P(\text{type I error}) = P(\text{Reject } H_0 | H_0 \text{ is true})$. $\alpha = 0.05 = 1/20$, this means that, by chance the null hypothesis will be incorrectly rejected once in every 20 tests.
- As the number of tests increases, the probability of finding a “significant” result by chance increases. That is, the overall type I error increases.

Why Multiple Testing matters?

- In general, if we perform k hypothesis tests at α level individually, what is the probability of at least one false positive?

$$P(\text{making at least 1 type I error in } k \text{ tests}) = 1 - (1 - \alpha)^k$$

Proof:

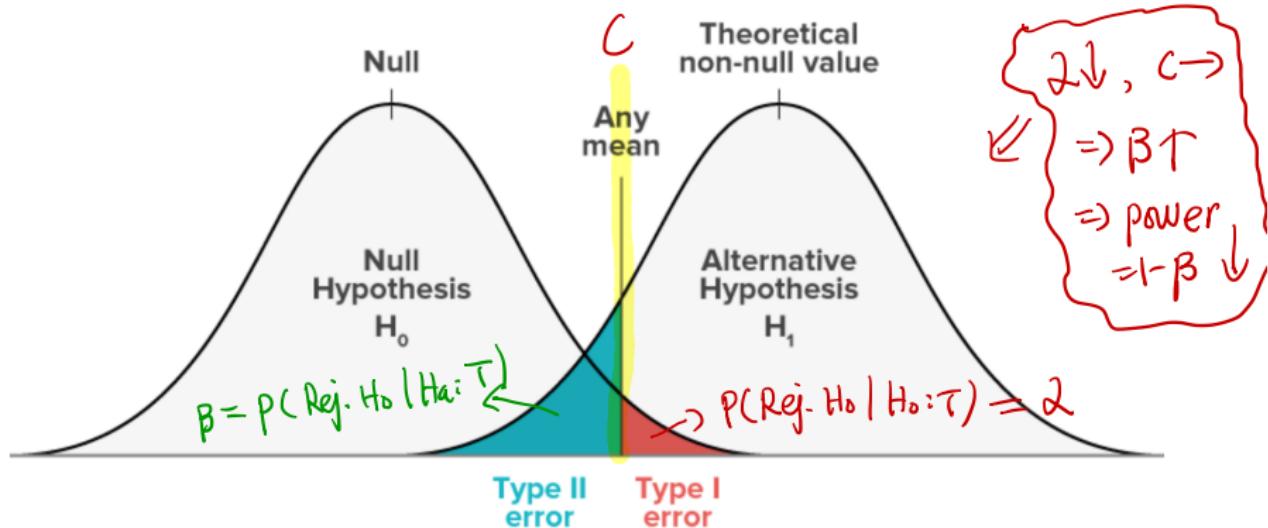
- P(making type I error in one test) = α
- P(not making type I error in one test) = $1 - \alpha$
- P(not making type I error in k independent tests) = $(1 - \alpha)^k$
- P(making at least 1 type I error in k independent tests) = $1 - (1 - \alpha)^k$

- Define the familywise error, α_{FW} as the probability of making a single type I error somewhere in the family of k multiple tests.

$$\alpha_{FW} \leq 1 - (1 - \alpha)^k$$

- e.g. $\alpha = 5\%$, $k = 4$, then $\alpha_{FW} \leq 1 - (1 - 0.05)^4 = 0.1855$. The familywise type I error rate for a family of 4 tests is 18.55%.
- Type I error is important to control and we failed to do it for this example.

Familywise error rate



The **familywise error**, α_{FW} for a family of k multiple tests with each test at α significant level is at most $1 - (1 - \alpha)^k$

- We note that we can lower the familywise error α_{FW} simply by lower the individual test error rate α .
 - If we do this for the worst case scenario (mutually independent tests) then we're covered case with dependent tests.
 - But it comes with a loss of power (increased chance of making a Type II error)
 - see comment in upper right corner.

Bonferroni Procedure

- Since $\alpha_{FW} \leq k\alpha$, we can guarantee $\alpha_{FW} \leq \alpha$ simply by using α/k in place of α for individual tests.
- This correction to individual tests is called **Bonferroni correction**.
- It is simple to understand and apply.
- It is very conservative (Type II error inflation): not very powerful, even though Type I error is controlled.
- It does not scale well with the number of tests (k).

Bonferroni Procedure

```
library(Sleuth2)
jury = case0502 # 46 observations on two variables (Percent, Judge:7 levels)
with(jury, pairwise.t.test(Percent, Judge, p.adj="bonf"))
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data: Percent and Judge
## Spock's ✓ significant results
##   A      B      C      D      E
## A 0.00022 -     -     -     -
## B 0.00013 1.00000 -     -     -
## C 0.00150 1.00000 1.00000 -     -
## D 0.57777 1.00000 1.00000 1.00000 -
## E 0.03408 1.00000 1.00000 1.00000 1.00000 -
## F 0.01254 1.00000 1.00000 1.00000 1.00000 1.00000
##
## P value adjustment method: bonferroni
```

$n = \# \text{ of total test}$
 $\{ p_i = P\text{.value from}$
 $\quad \downarrow \quad t\text{-test } i$

$\hookrightarrow \text{adjusted P-value}$

$= n p_i$ at 1
 $> 1 : \text{truncated}$

Confidence interval by Bonferroni Procedure

total # of $(1-\alpha)\%$ CI = k

- If we want a family of k $(1 - \alpha)\%$ CI (for means) with familywise coverage, using Bonferroni, it is

$$\hat{\theta} \pm t_{df, 1 - \frac{1}{2}(\alpha/k)} se(\hat{\theta})$$

```
library(Sleuth2)
jury = case0502 # 46 observations on two variables (Percent, Judge:7 levels)
fit = lm(Percent~Judge, data=jury)
confint(fit, level=1-0.05/nlevels(jury$Judge) )
```

```
##           0.357 % 99.643 %
## (Intercept) 8.078085 21.16636
## JudgeA      8.547341 30.44821
## JudgeB      8.647254 29.34163
## JudgeC      5.222969 23.73259
## JudgeD     -2.969585 27.72514
## JudgeE      1.997255 22.69163
## JudgeF      2.922970 21.43259
```

0 & CI

Tukey's HSD Procedure

- HSD stands for “Honestly Significant Difference”.
- Compare the difference between each pair of sample means to a critical difference

$$q = \frac{\bar{Y}_i - \bar{Y}_j}{\sqrt{\frac{1}{2} MSE(1/S_i^2 + 1/S_j^2)}} \sim q_{k,n-k} \text{ under } H_0$$

Given α , could find the critical value.

- $q_{k,n-k}$, a studentized range q distribution with a range of k and n-k degrees of freedom.
- k is the range (i.e., the number of groups). \overline{n}
of groups
- a difference between the means of Group i and j will be significant if

$$|\bar{Y}_i - \bar{Y}_j| > HSD = q_{k,\alpha} \sqrt{\frac{1}{2} MSE(1/S_i^2 + 1/S_j^2)}$$

\downarrow
critical value given α

Example: Tukey's HSD

```
library(Sleuth2)
jury = case0502 # 46 observations on two variables (Percent, Judge:7 levels)
fit = aov(Percent~Judge, data=jury)
TukeyHSD(fit, "Judge")
```

$$\text{conf. level} = 0.95 / 0.90 / \dots$$

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
```

```
## Fit: aov(formula = Percent ~ Judge, data = jury)
```

```
##
```

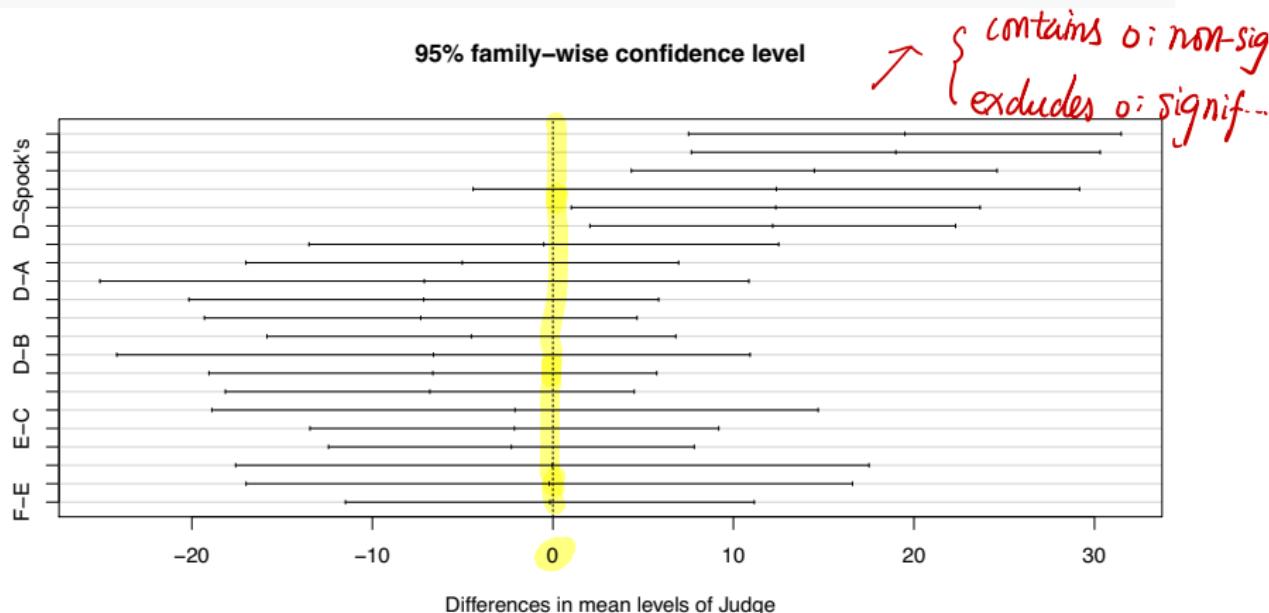
	diff	lwr	upr	p adj
## A-Spock's	19.49777727	7.514685	31.480870	0.0001992
## B-Spock's	18.99444405	7.671486	30.317402	0.0001224
## C-Spock's	14.47777732	4.350216	24.605339	0.0012936
## D-Spock's	12.37777758	-4.416883	29.172438	0.2744263
## E-Spock's	12.34444443	1.021486	23.667402	0.0248789
## F-Spock's	12.17777766	2.050216	22.305339	0.0098340
## B-A	-0.50333322	-13.512422	12.505755	0.9999997
## C-A	-5.01999995	-17.003092	6.963092	0.8470098
## D-A	-7.11999969	-25.094638	10.854639	0.8777485
## E-A	-7.15333284	-20.162421	5.855755	0.6146239
## F-A	-7.31999961	-19.303092	4.663093	0.4936380
## C-B	-4.51666673	-15.839625	6.806291	0.8742030
## D-B	-6.61666648	-24.158118	10.924785	0.9003280

↳ adjusted
p-value
(reference paper:
check portal
Wright (1992))

Example: Tukey's HSD

A visual representation of the tests

```
library(Sleuth2)
jury = case0502 # 46 observations on two variables (Percent, Judge:7 levels)
fit = aov(Percent~Judge,data=jury)
plot(TukeyHSD(fit, "Judge"))
```



Dunnett's Test

- When testing all pairwise comparisons with Bonferroni or Tukey, the group are treated symmetrically.
- Sometimes, we are interested in comparing one control group (Y_c) to all other $k-1$ groups.
- In this situation, we can do better than Bonferroni and Tukey HSD using Dunnett's Test
- Dunnett's method:
 - Compare the pairwise differences, $|\bar{Y}_i - \bar{Y}_c|$, to a critical difference

$$T = t_{\alpha, k-1, n-k} \sqrt{\frac{1}{2} MSE(1/S_i^2 + 1/S_j^2)}$$

- t_α comes from Dunnett's t table
- Only use when one of the groups is a control group
- Only interested in comparing the "other" groups to the control group

Example: Dunnett's Test

```
library(multcomp)
jury = case0502
summary(glht(aov(lm(Percent~Judge,data=jury)),linfct=mcp(Judge="Dunnett")))

##
##    Simultaneous Tests for General Linear Hypotheses
##
##    Multiple Comparisons of Means: Dunnett Contrasts
##
##
## Fit: aov(formula = lm(Percent ~ Judge, data = jury))
##
## Linear Hypotheses:
##   Control Grp Estimate Std. Error t value Pr(>|t|) 
## A - Spock's == 0  19.498    3.857   5.056 < 0.001 ***
## B - Spock's == 0  18.994    3.644   5.212 < 0.001 ***
## C - Spock's == 0  14.478    3.259   4.442 < 0.001 ***
## D - Spock's == 0  12.378    5.405   2.290  0.13025
## E - Spock's == 0  12.344    3.644   3.388  0.00885 **
## F - Spock's == 0  12.178    3.259   3.736  0.00332 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Results on Spock's conspiracy trial

- Are these adjustments necessary?
 - Use if snooping, but maybe not with pre-planned tests.
- Bonferroni, Tukey HSD and Dunnett all conclude that Spock's judge is different from A,B,C,E,F
 - But not judge D.

↑ check and compare
B: slides 23, 24
T: slides 26, 27
D: slide 29



Take a break, and see you on Thursday

Basic of Two-way ANOVA

- Response variable Y_{ijk} is continuous
- Have two categorical explanatory variables (call them **Factor A** and **Factor B**)
- Factor A has levels $i=1$ to a
- Factor B has levels $j=1$ to b
- Each combination of levels (i,j) labels the treatment combination or cell
- A third subscript k indicates observation number in cell (i,j) . $k = 1$ to n_{ij} (for now assume balanced design; equal sample sizes with $n_{ij} = r$)
- Could analyze as a one-way ANOVA by taking each (i,j) combination as a different level of a single factor.

Example: Paint Data

		application	
		D	S
primer	1	4.0,4.5,4.3	5.4,4.9,5.6
	2	5.6,4.9,5.4	5.8,6.1,6.3
	3	3.8,3.7,4.0	5.5,5.0,5.0

Paints are commonly applied to metal surfaces. An experiment was performed to investigate the effect of paint primers (3 levels: three different primers: 1,2,3), and the application method (2 levels: two application methods: spray (S), dip(D)) on the paint adhesion force (adhesion: the response variable).

- Factor A: paint primer, $i=1,2,3$
- Factor B: application method, $j=D,S$
- $r=3$: three observations for each treatment combination (i,j) .

↳ balanced design, $n_{ij} = r = 3$

Cell mean model

cell mean
↓

$$Y_{ijk} = \underline{\mu_{ij}} + \epsilon_{ijk}, \text{ where } \epsilon_{ijk} \sim \perp N(0, \sigma^2)$$

- Estimate μ_{ij} by cell mean \bar{Y}_{ij} .
- Estimate factor level means (mean for one level of given factor across all levels of other factor), as follows:

$$\hat{\mu}_{i..} = \bar{Y}_{i..}, \quad \hat{\mu}_{.j} = \bar{Y}_{.j}.$$

$\hat{\mu}_{i..}$ = marginal
mean of A
at level i

- Estimate grand mean by $\hat{\mu} = \bar{Y}_{...}$
- Disadvantage – need contrasts to separate effects

$\hat{\mu}_{.j} = \bar{Y}_{.j}$
= marginal mean
of B at level j

Factor Effect model

measures main effect of A

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk},$$

measures main effect of B

- where $\epsilon_{ijk} \sim_{\perp} N(0, \sigma^2)$
- and $\sum_i^a \alpha_i = \sum_j^b \beta_j = \sum_i (\alpha \beta)_{ij} = \sum_j (\alpha \beta)_{ij} = 0$
- Constraints required to keep model from being over-parameterized.
- Advantage: Effects can be analyzed separately. This is the model we want to use.
- Using indicator variables, this is still a multiple regression problem

{ For factor A with a levels \rightarrow create a-1 dummy variables
For factor B with b levels \rightarrow create b-1 dummy variables

Two-way ANOVA: main question of interest

- Are there **main effects** for group variables?

$$H_0 : \alpha_1 = \dots = \alpha_a = 0, \quad H_0 : \beta_1 = \dots = \beta_b = 0$$

- Are there **interaction effects**?

$$H_0 : (\alpha\beta)_{ij} = 0, 1 \leq i \leq a, 1 \leq j \leq b$$

use the factor effect

model

Constraints needed for identifiability

- $\sum_{i=1}^a \alpha_i = 0$
- $\sum_{j=1}^b \beta_j = 0$
- $\sum_{j=1}^b (\alpha\beta)_{ij} = 0, 1 \leq i \leq a$
- $\sum_{i=1}^a (\alpha\beta)_{ij} = 0, 1 \leq j \leq b$

} The one widely used in
most text book

Two-way ANOVA: main question of interest

once interaction is significant,
then keep all main effects
↙ terms in model.

- If the **interaction term** is statistically significant, we usually immediately decide that both **main effects** (Factors A and B) are also significant, no matter what the result of the test.
- It can happen that **both main effects tests are not significant but the interaction is**. This situation is called **masking**.

As usual, we divide the **total sum of squares (SS)** into the parts attributable to differences between the treatment means or main effects, written as **SSA** and **SSB**, the part attributable to **the interaction**, written as **SSAB**, and the part attributable to random variation, or **SSE**.

decomposition of sum of squares

$$SS = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij.}} (Y_{ijk} - \bar{Y}_{...})^2 = SS_{\text{Total}} : \text{need } Y_{ijk} \text{ information only.}$$

$$SSA = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij.}} (\bar{Y}_{i..} - \bar{Y}_{...})^2$$

$$SSB = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij.}} (\bar{Y}_{.j.} - \bar{Y}_{...})^2$$

$$SSAB = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij.}} (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$$

$$SSE = SS - SSA - SSB - SSAB = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij.}} (Y_{ijk} - \bar{Y}_{ij.})^2$$

F-tests for two-way ANOVA

Extra terms are ≥ 0
⇒ F-test and one-sided

Assume balanced design, for each cell (i,j) , we have $r = n_{ij}$ replicates.

SS	df	MS	E(MS)
SSA	a-1	SSA/(a-1)	$\sigma^2 + br \sum_{i=1}^a \frac{\alpha_i^2}{(a-1)}$
SSB	b-1	SSB/(b-1)	$\sigma^2 + ar \sum_{j=1}^b \frac{\beta_j^2}{(b-1)}$
SSAB	(a-1)(b-1)	SSAB/(a-1)(b-1)	$\sigma^2 + r \sum_{i=1}^a \sum_{j=1}^b \frac{(\alpha_i \beta_j)^2}{(a-1)(b-1)}$
SSE	ab(r-1)	SSE/ab(r-1)	σ^2

When the effects are **fixed**

- Ratio of MSAB/MSE tests for interaction effect (test this first since interpretation of main effects depend on significance of interaction).
- Ratio of MSA / MSE tests for factor A main effect
- Ratio of MSB / MSE tests for factor B main effect

F-tests for two-way ANOVA

- Interaction: H_0 : all $(\alpha\beta)_{ij} = 0$, H_a : not all $(\alpha\beta)_{ij} = 0$
 - Test statistic: $F_{AB}^* = MSAB/MSE \sim F_{(a-1)(b-1), ab(r-1)}$ under H_0
 - Reject H_0 if $F_{AB}^* > F_{1-\alpha, (a-1)(b-1), ab(r-1)}$
- Main effect A: H_0 : all $\alpha_1 = \dots = \alpha_a = 0$, H_a : not all α_i equal 0
 - Test statistic: $F_A^* = MSA/MSE \sim F_{(a-1), ab(r-1)}$ under H_0
 - Reject H_0 if $F_A^* > F_{1-\alpha, (a-1), ab(r-1)}$
- Main effect B: H_0 : all $\beta_1 = \dots = \beta_b = 0$, H_a : not all β_j equal 0
 - Test statistic: $F_B^* = MSB/MSE \sim F_{(b-1), ab(r-1)}$ under H_0
 - Reject H_0 if $F_B^* > F_{1-\alpha, (b-1), ab(r-1)}$

Two-way ANOVA Table



Source	SS	df	Mean Square	F
Factor A	SSA	(a-1)	$MSA = \frac{SSA}{(a-1)}$	$F = MST/MSE$
Factor B	SSB	(b-1)	$MSB = \frac{SSB}{(b-1)}$	$F = MSB/MSE$
Interaction AB	SSAB	(a-1)(b-1)	$MSAB = \frac{SSAB}{(a-1)(b-1)}$	$F = MSAB/MSE$
Error	SSE	ab(r-1)	MSE	

Revisit: Pain data example

Visualization: boxplot of both factors

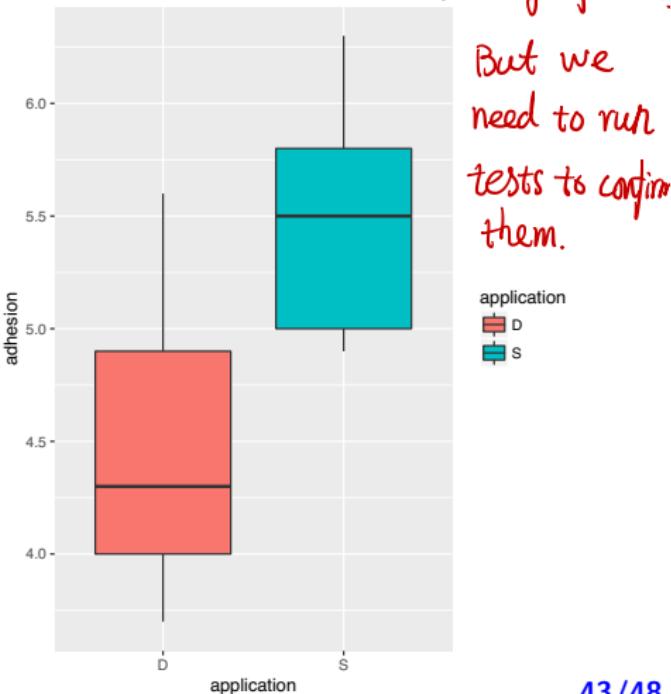
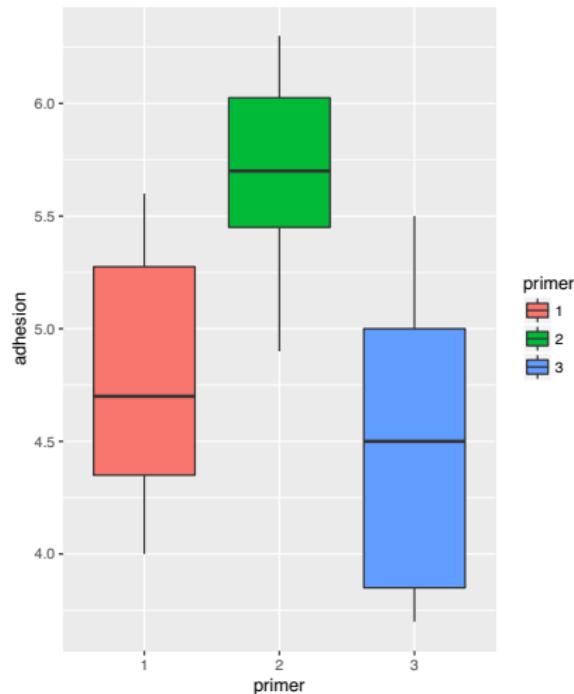
```
library(ggplot2)
library(gridExtra)

adhesion=c(4.0,4.5,4.3,5.6,4.9,5.4,3.8,3.7,4.0,
           5.4,4.9,5.6,5.8,6.1,6.3,5.5,5,5)
primer=factor(rep(1:3,each=3),2)
application=factor(rep(1:2,each=9),labels=c("D","S"))
paint=data.frame(adhesion,primer,application)
p1=ggplot(paint,aes(x=primer,y=adhesion,fill=primer))+geom_boxplot()
p2=ggplot(paint,aes(x=application,y=adhesion,fill=application))+geom_boxplot()
grid.arrange(p1, p2, ncol=2)
```

Revisit: Pain data example

Visualization: boxplot of both factors

- Main effect of primer seems significant.
- Main effect of application seems significant too.



visually, both main effects are significant

But we need to run tests to confirm them.

Revist: Pain data example

Visulization: interaction plot

```
library(ggplot2)
library(gridExtra)

adhesion=c(4.0,4.5,4.3,5.6,4.9,5.4,3.8,3.7,4.0,
           5.4,4.9,5.6,5.8,6.1,6.3,5.5,5,5)
primer=factor(rep(1:3,each=3),2))
application=factor(rep(1:2,each=9),labels=c("D","S"))
par(mfrow=c(1,2))

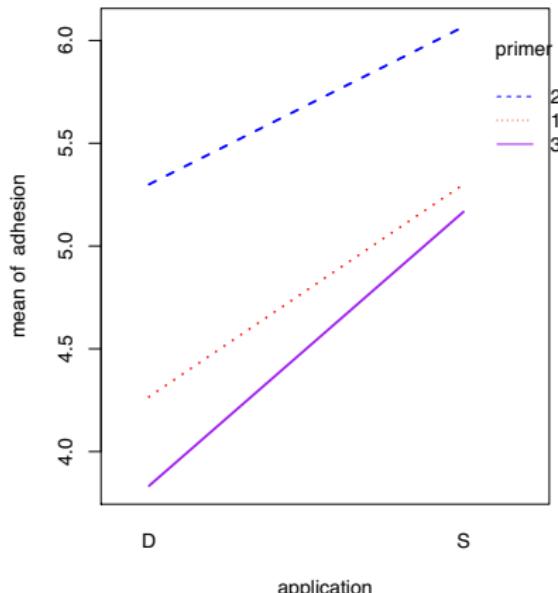
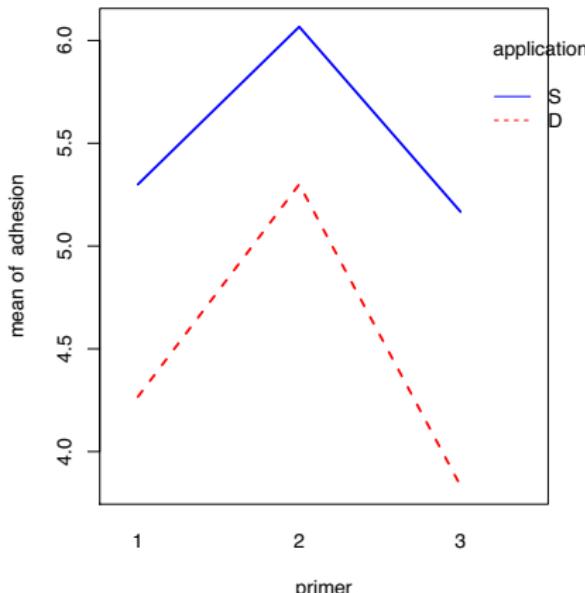
① interaction.plot(primer,application,adhesion,lwd=2,col=c("red","blue"))
② interaction.plot(application, primer,adhesion,lwd=2,col=c("red","blue","purple"))
```



Revist: Pain data example

Visualization: interaction plot

- Main effect of primer is bigger for application S than for application D.
- Main effect of application is increased from primer=3 to primer=1, and primer=2.
- Lines are nearly parallel, seems no interaction effect.



Revist: Pain data example

```
adhesion=c(4.0,4.5,4.3,5.6,4.9,5.4,3.8,3.7,4.0,  
          5.4,4.9,5.6,5.8,6.1,6.3,5.5,5,5)  
primer=factor(rep(rep(1:3,each=3),2))  
application=factor(rep(1:2,each=9),labels=c("D","S"))  
fit=lm(adhesion~primer*application)  
anova(fit)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: adhesion
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
## primer	2	4.5811	2.2906	27.8581	3.097e-05 ***
## application	1	4.9089	4.9089	59.7027	5.357e-06 ***
## primer:application	2	0.2411	0.1206	1.4662	0.2693

```
## Residuals
```

```
## ---
```

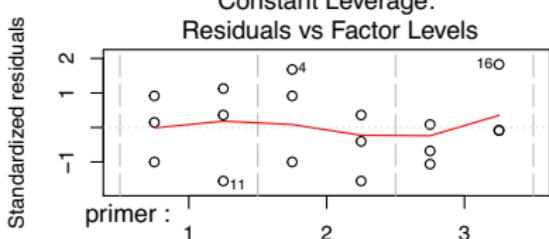
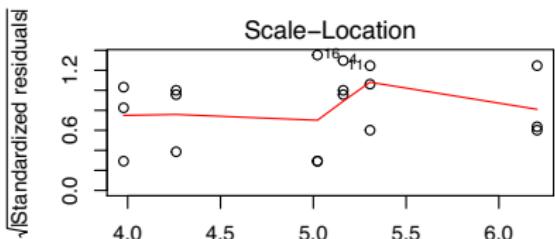
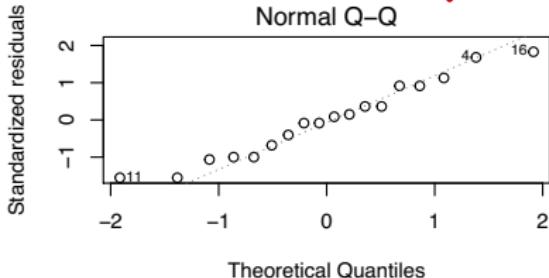
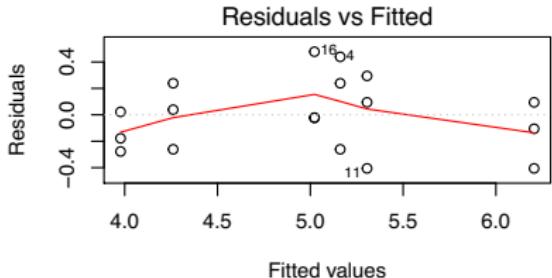
```
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3 sig.
→ non-significant

Revist: Pain data example

```
adhesion=c(4.0,4.5,4.3,5.6,4.9,5.4,3.8,3.7,4.0,  
         5.4,4.9,5.6,5.8,6.1,6.3,5.5,5,5)  
primer=factor(rep(rep(1:3,each=3),2))  
application=factor(rep(1:2,each=9),labels=c("D","S"))  
fit=lm(adhesion~primer+ application)  
par(mfrow=c(2,2)); plot(fit)
```

Assumption checking:
looks fine



After Lecture This Week

Practice problems

- Check if you receive A0 submission notice from Crowdmark or not, if not, write an email to instrutor.
- **A0 is due 10pm, Saturday, Jan. 21 on Crowdmark**
- Try the posted practice problem (solution is available too)
- Try all the R example in slides.
- Review all types of t-test and one-way ANOVA analysis.

Topics for next week:

- R Example: two-way ANOVA
- Discussion about two-way ANOVA