Advanced Course in Marketing 08
Statistical tests and comparison

Takumi Tagashira

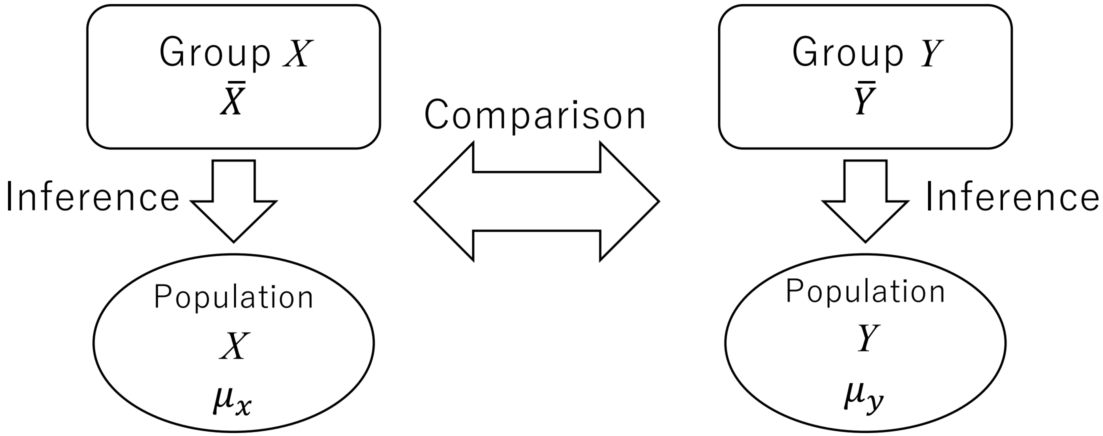
Email: takumi.tagashira@r.hit-u.ac.jp

Aims of this session

- We ensured the core ideas of the statistical tests based on population mean test in the previous session.
- Other practical statistical test examples:
    - Mean comparison test (between two groups)
    - Mean comparison test between more than two groups（ANOVA）
- We will use the different test statistics for these purposes, but the test procedures and the core concepts are coherent.

Mean comparison tests

- We want to check whether the population means differ between two group (X vs. Y).
  - E.g., Gender, region
- We consider the relationship between population and samples in tests.



Mean comparison test

- Suppose we want to check whether the population means differ between two group (X vs. Y).
  - E.g. gender, region, etc.
  - Marketing research often involves this type of tests.
- The sample mean difference is $(\bar{X} - \bar{Y})$.
- In this condition, the test hypotheses are:
- H0: Population means of two groups are identical.
- H1: Population means of two groups are different.

Mean comparison test

- Suppose we want to check whether the population means differ between two group (X vs. Y).
  - E.g. gender, region, etc.
  - Marketing research often involves this type of tests.
- The sample mean difference is $(\bar{X} - \bar{Y})$.
- In this condition, the test hypotheses are:
- H0: $\mu_X = \mu_Y$
- H1: $\mu_X \neq \mu_Y$
- We can rewrite H0 as: $\mu_X - \mu_Y = \mathbf{0}$

Intuition of test statistics

- Intuition of test statistic of population mean is as follows:
- Test statistic= $\frac{Point\ estimate - Null\ value}{Std.Dev.(or\ error)} = \frac{(\bar{X}-\bar{Y})-(\mu_x-\mu_y)}{Std.err.}$

- $(\mu_x - \mu_y)$ is unknown…
- When H0 is true; $\mu_x - \mu_y = 0$
- Thus, we employ the following (calculable) test statistics:
- $\frac{(\bar{X}-\bar{Y})}{Std.err.} \sim t(m+n-1)$
- It is known that the test statistic follows t-distribution when H0 is true.

Detailed points

- The basic mean comparison tests assume that the variances between two groups are identical ($\sigma_X = \sigma_X = \sigma$).
  - i.e., heteroscedasticity
- When two variances are not identical (i.e., heteroscedastic), we use a t-test method called Welch’s t-test.
  - If your results mention 'Welch’s test,' please don't be concerned.
  - We will skip the details.

Mean comparison tests in R

- It is not difficult to run the tests.
- We use ~ (tilde) to model the relationship between outcome and categorical variables
- Now, we test whether the amount spent differs by gender using the “idpos_cust” dataset.

```
idpos_cust <- readr::read_csv("data/idpos_customer.csv")
t.test(monetary ~ gender, data = idpos_cust) #Welch
t.test(monetary ~ gender, data = idpos_cust,var.equal=T) #Homoscedastic t-test
```

- Results are slightly different?
- Which method is more appropriate?

Intuition of results

```
> t.test(monetary ~ gender, data = idpos_cust) #Welch

        Welch Two Sample t-test

data:  monetary by gender
t = 1.7895, df = 1334.2, p-value = 0.07377
alternative hypothesis: true difference in means between group female and group male is not equal to 0
95 percent confidence interval:
 -253.4545 5518.7863
sample estimates:
mean in group female   mean in group male
            17511.03             14878.36
```

t-value

P-value is less than 10% significance level, but not 5%. H0 cannot be rejected in 5% significance level.

Confidence interval

Sample mean

Homoscedasticity test in R

- We use “var.test()” function.
- The result shows that we can reject the null hypothesis:
- It means that the variances are not identical→ Welch’s test

```
var.test(monetary ~ gender, data = idpos_cust, ratio = 1)

##  F test to compare two variances
##
## data:  monetary by gender
## F = 1.9913, num df = 984, denom df = 501, p-value < 2.2e-16
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  1.706794 2.314410
## sample estimates:
## ratio of variances
##           1.991318
```

Tests in R

```
##t-test: mean comparison (Homogeneous variance)
t.test(varname ~ groupname, data = dataset, var.equal=TRUE)
t.test(dataset$varname ~ dataset$groupname, var.equal=TRUE)

#F-test: A comparison of variances
var.test(varname ~ groupname, data = dataset, ratio = 1)

##t-test: Welch's S.E.
t.test(varname ~ groupname, data = dataset, var.equal=FALSE)
```

Analysis of Variance

- We have seen methods to compare mean of two groups.
- Is it possible if the number of groups is more than two?
  - e.g. Comparison of sales among different regions in Japan.

- We have seen methods to compare mean of two groups.

- Is it possible if the number of groups is more than two?
  - e.g. Comparison of sales among different regions in Japan.

- Yes
- Analysis of Variance (ANOVA)

Independent variables and level

- Independent variable represents a factor that may impact on a focal outcome (i.e. dependent variable).
  - e.g. Regional areas that affect sales
- Levels (groups)
  - Conditions of the independent variable
  - e.g. region A, B, and C
- ANOVA assesses if the different levels of the independent variable impact on the dependent variable

ANOVA in R

- We use a dataset “tips” in a package called “reshape2”.

```
library(reshape2)
```

- Variables
  - total_bill (USD)
  - tip (USD)
  - sex: Gender of the person who paid the bill
  - smoker: Whether there is a smoker in the group
  - day: Day of the transaction
  - time
  - size: The number of people

ANOVA in R

- We use “aov()” to execute anova and “anova()” function to report the result.
- E.g.: Checking whether the amount of tip are different between days.

```
s <- aov(tip ~ day, data = tips)
anova(s)
```

- Results ?

Results

```
> anova(s)
Analysis of Variance Table

Response: tip
           Df Sum Sq Mean Sq F value Pr(>F)
day         3   9.53  3.1753  1.6724 0.1736
Residuals 240 455.69  1.8987
```

Sum of variance between groups:
Total variation between the group means and the overall mean

Sum of squared residuals (i.e. within group variation)

Results

```
> anova(s)
Analysis of Variance Table

Response: tip
           Df Sum Sq Mean Sq F value Pr(>F)
day         3   9.53  3.1753  1.6724 0.1736
Residuals 240 455.69  1.8987
```

Sum Sq/DF

P-value：
We can't reject
H0 this time.

F-value（test statistic）

- Then, what is the null hypothesis (H0) for ANOVA?

Test of ANOVA

- Null hypothesis for AVOVA can be specified as follows.
- H_0: All mean values are identical.
- What is the alternative hypothesis?

Test of ANOVA

- Null hypothesis for AVOVA can be specified as follows.
- H_0:All mean values are identical.
- H_1: At least one mean is different.

Example: summary table

- ANOVA results are often summarized as follows.

	Sum-Sq	D.F.	Mean-Sq	F-value	p-value
Treatment variable					
Residuals					

- When the null hypothesis is rejected, researchers often conduct “pairwise-comparison”.
  - The ANOVA results do not answer which group is unique.
  - e.g., Tukey’s multiple pairwise-comparisons.

Post-hoc analysis for ANOVA

- The rejection of the null hypothesis concludes that at least one group differs from others.
  - We want to check which one is unique.
- However, when we conduct pair wise multiple comparisons, it leads to the inflated probability of type 1 error occurring.
  - Suppose there are groups A, B, and C; the pair wise comparison captures the difference between A vs. B, B vs. C, and B vs. C.
- Tukey's pair wise mean comparison is a widely employed method that assesses the type 1 error issue.
  - In general, this method is less likely to reject the null hypotheses.

Tukey's pairwise comparison in R

```
tukey_result <- TukeyHSD(s)
tukey_result
```

```
Fit: aov(formula = tip ~ day, data = tips)

$day
                 diff        lwr       upr     p adj
Sat-Fri    0.25836661 -0.6443694 1.1611026 0.8806455
Sun-Fri    0.52039474 -0.3939763 1.4347658 0.4558054
Thur-Fri   0.03671477 -0.8980753 0.9715049 0.9996235
Sun-Sat    0.26202813 -0.2976929 0.8217492 0.6203822
Thur-Sat  -0.22165184 -0.8141430 0.3708394 0.7678581
Thur-Sun  -0.48367997 -1.0937520 0.1263921 0.1724212
```

Difference in means

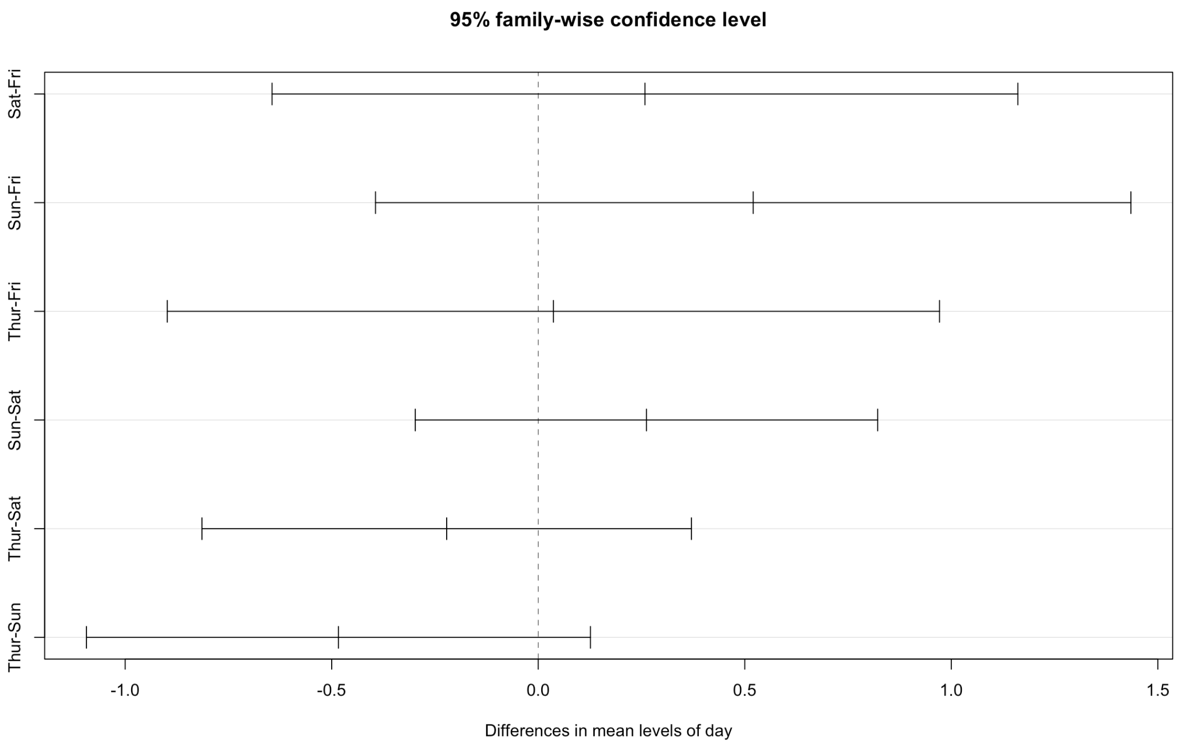
Confidence interval

P-value

None of pairs have the significant result.

Visualization of the confidence interval

```
plot(tukey_result)
```



Interpretation

- Pair wise comparison checked the mean differences for each day.
- The original null hypothesis showed at least one group differs from others.

```
library(tidyverse)
tips %>%
  group_by(day) %>%
  summarize(tipM = mean(tip))

 day    tipM
  <chr> <dbl>
1 Fri    2.73
2 Sat    2.99
3 Sun    3.26
4 Thur   2.77
```

- It seems Sunday has the highest amount of tips.
- We will generate a dummy variable and check “Sunday vs. others” statistically.

Post-hoc analysis (Sunday vs. other days)

```
#Generating day-dummy variables

tipsD <- tips %>%
  mutate(D_sun = ifelse(day == "Sun",1,0),
         D_sat = ifelse(day == "Sat",1,0),
         D_Fri = ifelse(day == "Fri",1,0),
         D_Thu = ifelse(day == "Thur",1,0))

##t-test (Sun vs. other days)
t.test(tip ~ D_sun, data = tipsD, var.equal=F)
```

```
	Welch Two Sample t-test

data:  tip by D_sun
t = -2.0754, df = 166.61, p-value = 0.03949
alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
95 percent confidence interval:
 -0.72793131 -0.01816518
sample estimates:
mean in group 0 mean in group 1
       2.882083        3.255132
```

Smaller than 0.05.
→Rejecting H0 with 5% significance level.

Group D = 1 has a higher mean.

Summary

- ANOVA requires post-hoc analyses.
- Tukey's test is a widely employed method that modifies type 1 error inflation.
  - There are other similar methods.
- Interpret the results carefully!
  - Refer to the groups you compared precisely.
  - E.g. Sunday vs. others, or Sunday vs. Saturday etc.

Exercise: Interpretation of results

- Use tips data and check if the amount of total_bill is different across days.

- No submission required.

Supplementary slides

(Sup.) One-way ANOVA

- One-way ANOVA captures one independent variable.
  - We skip two-way ANOVA
- e.g.
- Consider a question of whether the length of individual study impacts the final exam score of Advanced course in Marketing.
- 40 random samples from the enrolled students will be divided into the following 4 groups.
  - “More than 3 times a week”, “Twice a week”, “Once a week”, “Less than once a week”
- Check if the exam scores are different due to the above-specified levels at the end of the course.
  - We often express such group division as j ($j = 1,2,3,4$) group.

(Sup.) ANOVA I

- Let J denote the number of groups, overall n samples and group j has n_j samples.
- Let y_{ij} denote the value of individual i in group j.

$$y_{ij} = \mu_j + e_{ij}, \qquad e_{ij} \sim N(0, \sigma^2)$$

- where $i = 1,\ldots, n_j$, $j = 1,\ldots, J$ and all y_{ij} are independent ($y_{ij} \sim N(\mu_j, \sigma^2)$)

$$\mu = \frac{n_1}{n}\mu_1 + \ldots + \frac{n_J}{n}\mu_J = \frac{1}{n}\sum_{j=1}^{J} n_j \mu_j$$

(Sup.) ANOVA II

- $\alpha_j = \mu_j - \mu$: Group effect

$$y_{ij} = \mu + \alpha_j + e_{ij}$$

- When $\alpha_1 = \cdots = \alpha_J = 0$ (i.e. There is no difference across groups), and

$$y_{ij} = \mu + e_{ij}.$$

Test of ANOVA

- Null hypothesis for AVOVA can be specified as follows.

- H_0: $\alpha_{\mathbf{1}} = \cdots = \alpha_J = \mathbf{0}$

- What is the alternative hypothesis?

Test of ANOVA

- Null hypothesis for AVOVA can be specified as follows.
- H_0: $\boldsymbol{\alpha_1 = \cdots = \alpha_J = 0}$
- H_1: At least one mean is different.
- Using F-test

$$F = \frac{n-J}{J-1} \times \frac{\sum_{j=1}^{J} n_j(\bar{y_j} - \bar{y})^2}{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij} - \bar{y_j})^2} \sim F(J-1,\ n-J)$$

Test of ANOVA

$$F = \frac{n-J}{J-1} \times \frac{\sum_{j=1}^{J} n_j(\bar{y_j} - \bar{y})^2}{\sum_{j=1}^{J} \sum_{i=1}^{n_j} (y_{ij} - \bar{y_j})^2} \sim F(J-1,\ n-J)$$

- It is known that F follows f-distribution with degree of freedom (J-1, n-J), when the null hypothesis is true.
- Assume a significance level α, and
  - $F > F_\alpha \rightarrow$ reject H_0
  - $F \leq F_\alpha \rightarrow$ accept H_0

(Sup.) Test of ANOVA

$$F = \frac{n-J}{J-1} \times \frac{\sum_{j=1}^{J} n_j (\bar{y_j} - \bar{y})^2}{\sum_{j=1}^{J} \sum_{i=1}^{n_j} (y_{ij} - \bar{y_j})^2} \sim F(J-1,\ n-J)$$

- It is known that F follows f-distribution with degree of freedom (J-1, n-J), when the null hypothesis is true.
- Assume a significance level α, and
  - $F > F_\alpha \rightarrow$ reject H_0
  - $F \leq F_\alpha \rightarrow$ accept H_0

ANOVA in R

- F refers to a group variable

```
s<-aov(y ~ F)
anova(s)
```

- Without an intercept term

```
s<-aov(y ~ F-1)
anova(s)
```

- Pairwise-comparisons (Tukey’s comparison)
    - Modified confidence intervals are used for multiple comparison.

```
TukeyHSD(s)
```

ANOVA

- ANOVA can compare the population means between more than two groups.
- ANOVA can be applied to models with more than one factors
  - e.g. How do the consumers' intention to buy a brand vary with different levels of price and different levels of distribution?
  - i.e. Two-way ANOVA (Main and interaction effects)
- Main effects: Effects of each independent variable
- Interaction effects: The effects of one factor on the dependent variable depend on the level (category) of the other factor.
  - Interaction term in regression model captures the same mechanism as this model