

Stat 135 Lab 10

GSI: Yutong Wang
Apr 10, 2020

Questions?



Source: Phillip Kraskoff / shutterstock

Sandy Beach

<https://www.thecrazytourist.com/15-best-beaches-oahu/>

To-Do Today

1. Mann-Whitney/Wilcoxon rank sum test
2. Wilcoxon signed rank test
3. One-Way ANOVA (Analysis of Variance)
 - a. Analysis of variance
 - b. F distribution
4. Bonferroni Correction

Review: Mann-Whitney test

Data: Let X_1, \dots, X_n be the smaller sample from F (control group), and Y_1, \dots, Y_m be the larger sample from G (treatment group).

Hypothesis: $H_0 : F = G, H_1 : F \neq G$.

Test: There are two possible test statistics:

1. T_Y is the Mann-Whitney test statistic. It is the rank sum of larger sample in the combined sample.
2. $U_Y = \sum_{i=1}^n \sum_{j=1}^m I(X_i < Y_j)$ is the Wilcoxon test statistic.

$$T_Y \sim N\left(\frac{m(m+n+1)}{2}, \frac{mn(m+n+1)}{12}\right). \text{ Reject } H_0 \text{ if } \left| \frac{T_Y - \frac{m(m+n+1)}{2}}{\sqrt{\frac{mn(m+n+1)}{12}}} \right| > z_{\alpha/2}$$

$$U_Y = T_Y - \frac{m(m+1)}{2}$$

$$U_Y \sim N\left(\frac{mn}{2}, \frac{mn(m+n+1)}{12}\right). \text{ Reject } H_0 \text{ if } \left| \frac{U_Y - \frac{mn}{2}}{\sqrt{\frac{mn(m+n+1)}{12}}} \right| > z_{\alpha/2}$$

Review: Comparison between t -test and Mann-Whitney test

Note that t -test is uniformly most powerful when the assumptions of t -test are met, and Mann-Whitney test is 95% as powerful as t -test.

Hypothesis: $H_0 : \mu_X - \mu_Y = 0, H_1 : \mu_X - \mu_Y \neq 0$.

When the assumptions of t -test are not met, the Mann-Whitney test is preferred. To check if the normal assumption of 2-sample t -test is met, you should use boxplot or QQ-plot in R.

Transforming data that violates assumptions



Review: Wilcoxon signed rank test

For paired data, e.g.,

before	after	diff	abs(diff)	rank	signed rank
25	27	2	2	2	2
29	25	-4	4	3	-3
60	59	-1	1	1	-1
27	37	10	10	4	4

Hypothesis: $H_0 : \mu_{\text{diff}} = 0, H_1 : \mu_{\text{diff}} \neq 0$.

Test statistic: W_+ = sum of positive signed ranks, i.e., rank times sign(difference).

Thm regarding computing p-value: $W_+ \sim N\left(\frac{n(n+1)}{4}, \frac{n(n+1)(2n+1)}{24}\right)$.

Review: Wilcoxon signed rank test

You should use the signed rank test when the data are paired.

```
wilcox.test(x, y, paired = TRUE)
```

You should use the rank-sum test when the data are not paired.

```
wilcox.test(x, y, paired = FALSE)
```

How to decide whether the data is paired or not?

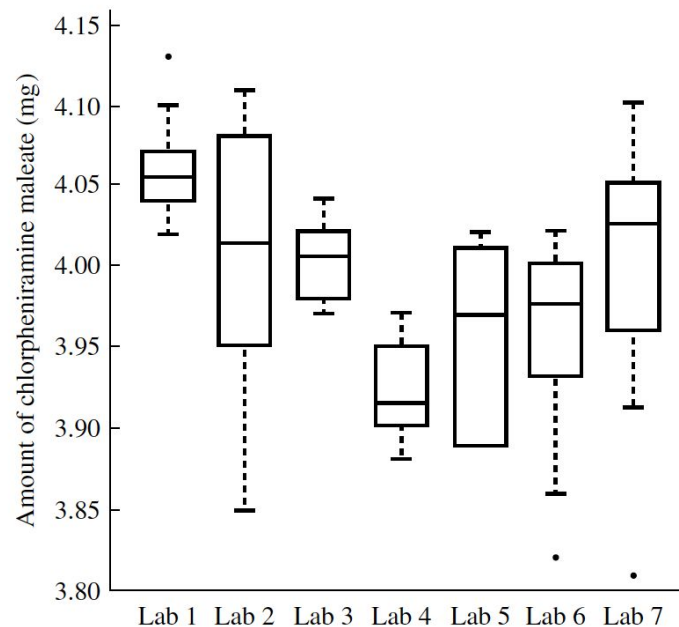
Pairing comes from consideration of what was sampled, instead of whether two groups have the same sample size or not.

The pairs of values are at least somewhat positively dependent, while unpaired values are not dependent. Often the dependence-pairing occurs because they're observations on the same unit (repeated measures), e.g. before-versus-after measure.

Review: One-Way ANOVA

(1) Problem Setup

There are I groups, each containing J samples. The I groups will be referred to generically as treatments, or levels. E.g., $I = 7$, $J = 10$ in Rice 12.2.



Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7
4.13	3.86	4.00	3.88	4.02	4.02	4.00
4.07	3.85	4.02	3.88	3.95	3.86	4.02
4.04	4.08	4.01	3.91	4.02	3.96	4.03
4.07	4.11	4.01	3.95	3.89	3.97	4.04
4.05	4.08	4.04	3.92	3.91	4.00	4.10
4.04	4.01	3.99	3.97	4.01	3.82	3.81
4.02	4.02	4.03	3.92	3.89	3.98	3.91
4.06	4.04	3.97	3.90	3.89	3.99	3.96
4.10	3.97	3.98	3.97	3.99	4.02	4.05
4.04	3.95	3.98	3.90	4.00	3.93	4.06

FIGURE 12.1 Boxplots of determinations of amounts of chlorpheniramine maleate in tablets by seven laboratories.

Review: One-Way ANOVA

(2) Models and hypothesis

Notation: Y_{ij} = the j th observation of the i th treatment, μ is the overall mean level, α_i is the different effect of the i th treatment, and ϵ_{ij} is the random error in the j th observation under the i th treatment.

Model:

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$
$$\sum_{i=1}^I \alpha_i = 0$$

The errors $\epsilon_{ij} \sim N(0, \sigma^2)$ are assumed to be independent from each other. Thus,

$$Y_{ij} \stackrel{\text{i.i.d.}}{\sim} N(\mu + \alpha_i, \sigma^2)$$

Hypothesis: $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_I = 0$, i.e., $\mathbb{E}[Y_{ij}] = \mu$.
 H_1 : at least one α_i is not equal to zero.

Review: One-Way ANOVA

(3) Analysis of variance

Notation:

$$\text{SST} = \sum_{i=1}^I \sum_{j=1}^J (Y_{ij} - \bar{Y})^2, \text{ total sum of squares}$$

$$\text{SSB} = \sum_{i=1}^I \sum_{j=1}^J (\bar{Y}_{i.} - \bar{Y})^2 = J \sum_{i=1}^I (\bar{Y}_{i.} - \bar{Y})^2, \text{ sum of squares between groups}$$

$$\text{SSW} = \sum_{i=1}^I \sum_{j=1}^J (Y_{ij} - \bar{Y}_{i.})^2 = (J - 1) \sum_{i=1}^I s_i^2, \text{ sum of squares within groups}$$

where s_i^2 is the sample variance in the i th group.

Facts: (1) $\text{SST} = \text{SSW} + \text{SSB}$

(2) SSB and SSW are independent random variables.

(3) Under the null hypothesis, $\frac{\text{SST}}{\sigma^2} \sim \chi_{IJ-1}^2$, $\frac{\text{SSW}}{\sigma^2} \sim \chi_{I(J-1)}^2$.

(Recall that $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$)

(4) Because of chi-squared subtraction property, we know $\frac{\text{SSB}}{\sigma^2} \sim \chi_{I-1}^2$.

Review: One-Way ANOVA

(4) F distribution

Let U, V be independent χ^2 random variables, with degrees of freedom a, b , respectively. The distribution

$$F_{a,b} = \frac{U/a}{V/b}$$

is called the **F distribution**.

Assuming the null hypothesis being true, the **F test statistic** is

$$F_{I-1, I(J-1)} = \frac{\frac{SSB/\sigma^2}{I-1}}{\frac{SSW/\sigma^2}{I(J-1)}}$$

The **p-value** can be calculated in R using $1 - \text{pf}(\text{F t.s.}, I - 1, I(J - 1))$. We will reject the null hypothesis if it is larger than 0.05.

Review: One-Way ANOVA

(4) Caveats about F distribution

Assumptions: The normal theory F test has several assumptions:

- (1) our data Y_{ij} is normal.
- (2) the variance of each treatment group is the same σ^2 .
- (3) all the observations are independent.

If you do not satisfy all the assumptions above, do a non-parametric test (Kruskal-Wallis test) instead!

Problem 1

Crysvita is a new medicine for the treatment of X-linked hypophosphatemia (a genetic disease), which was just approved by FDA in April 2018. Below is a simplified version of some test data for this pill where a high curing index corresponds to a positive effect.

Construct a F-test assuming that each observation follows a normal distribution with the same variances. Show that your F statistics is equal to the square of t-statistic you would get from a 2-sample t-test.

Pill	n	Mean Curing Index	Standard Error of Means
Placebo	10	100	10
Crysvita	10	160	13

Problem 1

Hints

There are $I = 2$ groups (placebo and crystiva), each containing $J = 10$ samples.

To construct such **F test**, we need to

- (1) Calculate SSB and SSW.
- (2) Calculate the F test statistic.
- (3) Compute the p-value.
- (4) Decide whether to reject the null.

To construct the **2-sample t- test**, we

- (1) Decide whether the data is paired or not.
- (2) If it is unpaired, do the samples have the same SD?
- (3) Calculate the corresponding t test statistic.
- (4) Compute the p-value.

Problem 2: Rice 12.2.22

Referring to Section 12.2, the file *tablets1.txt*, *tablets2.txt* gives the measurements on chlorpheniramine maleate tablets from another manufacturer.

Are there systematic differences between the labs?

If so, which pairs differ significantly?

How do these data compare to those given for the other manufacturer in Section 12.2?

See data on bCourses.

	Lab1	Lab2	Lab3	Lab4	Lab5	Lab6	Lab7
1	4.15	3.93	4.10	4.16	4.14	4.09	4.10
2	4.08	3.92	4.10	4.14	4.14	4.08	4.05
3	4.09	4.08	4.05	4.16	4.14	4.08	4.05
4	4.08	4.09	4.07	4.02	4.17	4.12	4.06
5	4.01	4.06	4.06	4.15	4.13	4.17	4.06
6	4.01	4.06	4.03	4.15	4.24	4.15	4.12
7	4.00	4.02	4.04	4.18	4.18	4.12	4.07
8	4.09	4.00	4.03	4.12	4.14	4.10	4.18
9	4.08	4.01	4.03	4.09	4.25	4.12	4.15
10	4.00	4.01	4.06	4.04	4.17	4.12	4.18

Problem 2: Rice 12.2.22

Hints

Are there systematic differences between the labs?

↔ Is there any α_i (different effect of the i th treatment) that is nonzero?

The above is the hypothesis of F test.

- (1) Calculate SSB, SSW.
- (2) Calculate the F test statistic.
- (3) Calculate the p-value.
- (4) Reject/ do not reject the null hypothesis.



We accept the null hypothesis.



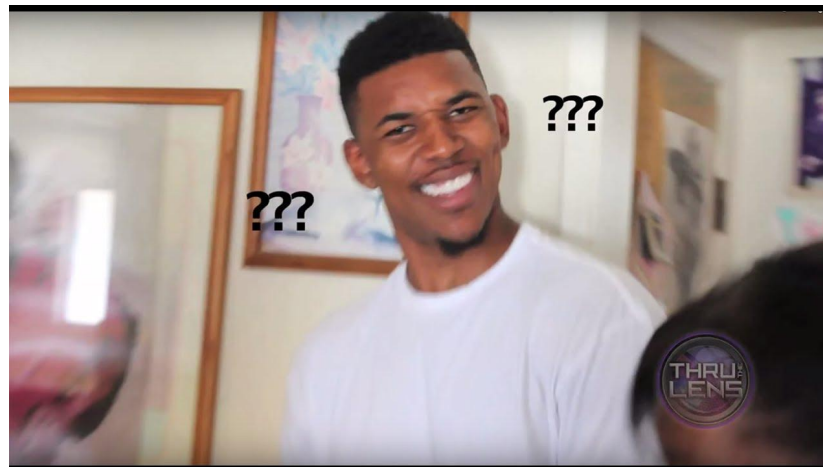
We fail to reject the null hypothesis.

Problem 2: Rice 12.2.22

Questions for you...

Now, we reject the null hypothesis, but which pairs are significantly different?

If we increase the number of groups, it seems like we will be more prone to reject the null (i.e., the different effect in least one group is non-zero). But should the p-value be related to the number of groups??

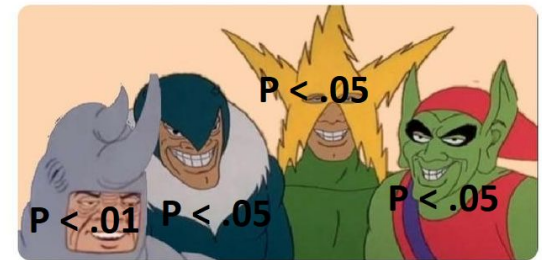


Review: Bonferroni Correction

If k null hypotheses are to be tested, a desired overall type I error rate of at most α can be guaranteed by testing each null hypothesis at level α/k .

Equivalently, if k confidence intervals are each formed to have confidence level $100(1 - \alpha/k)\%$, they all hold simultaneously with confidence level at least $100(1 - \alpha)\%$.

Me and the significant boys



Me and the significant boys after Bonferroni correction



Review: Bonferroni Correction

A $100(1 - \frac{\alpha}{k})\%$ CI for the i th group is $\bar{Y}_i \pm t_{J_i-1}(\frac{\alpha/2}{k}) \frac{S_i}{\sqrt{J_i}}$, where S_i is the within-group sample SD for group i .

Find such CI for all the groups, and see if there is any pair of groups whose CI do not overlap with each other.

Problem 2: Rice 12.2.22

Problem - contd

Referring to Section 12.2, the file *tablets1.txt*, *tablets2.txt* gives the measurements on chlorpheniramine maleate tablets from another manufacturer.

If so, which pairs differ significantly?

Hint:

Create 7 confidence intervals with Bonferroni correction for each lab, and see if there is any pair of labs whose CI do not overlap with each other.

Problem 3

The following data were obtained in a study of the effects of different diets on weight at the end of an experimental period.

Diet	n	Mean Weight (lb)	Standard Deviation
Pizza	5	135	20
Ice Cream	5	160	15
Ice Water	5	155	22

What effect did diet have on weight? Do a complete analysis and discuss the results. Use a 5% significance level. For full credit, discuss all assumptions in your test.