Stat 135 Lab 10

GSI: Yutong Wang
Apr 10, 2020

Berkeley
UNIVERSITY OF CALIFORNIA

Questions?

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

Berkeley
UNIVERSITY OF CALIFORNIA

Review: Mann-Whitney test

Data: Let $X_1, ..., X_n$ be the smaller sample from F (control group), and $Y_1, ..., Y_m$ 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.

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

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

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

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

Berkeley
UNIVERSITY OF CALIFORNIA

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: $W_+ \sim N(\frac{n(n+1)}{4}, \frac{n(n+1)(2n+1)}{24})$.

Berkeley
UNIVERSITY OF CALIFORNIA

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.

Reference: stackexchange by Glen_b -Reinstate

Berkeley
UNIVERSITY OF CALIFORNIA

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.
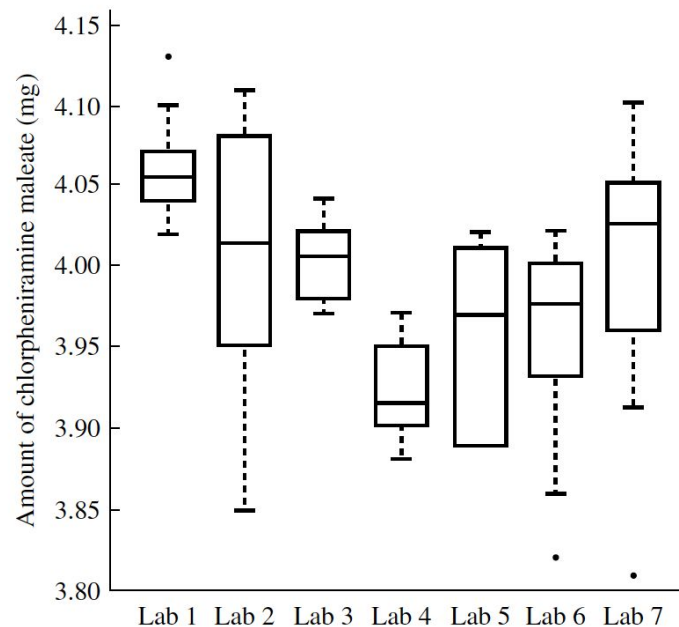
Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7
4.13	3.86	4.00	3.88	4.02	4.02	4.00
4.07	3.85	4.02	3.88	3.95	3.86	4.02
4.04	4.08	4.01	3.91	4.02	3.96	4.03
4.07	4.11	4.01	3.95	3.89	3.97	4.04
4.05	4.08	4.04	3.92	3.91	4.00	4.10
4.04	4.01	3.99	3.97	4.01	3.82	3.81
4.02	4.02	4.03	3.92	3.89	3.98	3.91
4.06	4.04	3.97	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 jth observation of the ith treatment, μ is the overall mean level, α_i is the different effect of the ith treatment, and ϵ_{ij} is the random error in the jth observation under the ith 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,

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

Hypothesis: $H_0 : \alpha_1 = \alpha_2 = ... = \alpha_I = 0$, 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 ith group.

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

(2) SSB and SSW are independent random variables.

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

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

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

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

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

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

Berkeley
UNIVERSITY OF CALIFORNIA

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

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 1
Answer Key

Pill	n	Mean Curing Index	Standard Error of Mean	Sample SD
Placebo	10	100	10	31.62
Crysvita	10	160	13	41.11

Solutions:

Here we have $I = 2$ and $J = 10$. Since we only have these summary statistics, we need to calculate SSb and SSw without the actual data. The last column can be calculated by multiplying the SE of mean by $\sqrt{10}$. Therefore,

$$SS_W = \sum (J-1)s_i^2 = 9 * (31.62^2 + 41.11^2) = 24208.71$$

$$\bar{Y}_{..} = \frac{\bar{Y}_{1.} + \bar{Y}_{2.}}{2} = \frac{100 + 160}{2} = 130$$

$$SS_B = J * \sum (\bar{Y}_{i.} - \bar{Y}_{..})^2 = 10 * (30^2 + 30^2) = 18000$$

$$F = \frac{SS_B/(I-1)}{SS_W/(I*(J-1))} = \frac{24208.71}{18000/18} = 24.209$$

According to the F-table, $F_{0.05,1,27} = 4.41 < 24.209$, so we reject the null that the new medicine has the same effect as a placebo (no effect).

Problem 1
Answer Key - contd

Recall that the t statistic for 2 independent samples with size n is

$$t_{2n-2} = \frac{\bar{X} - \bar{Y} - 0}{s_p \sqrt{\frac{2}{n}}}$$

where

$$s_p = \sqrt{\frac{(n-1)(s_x^2 + s_y^2)}{2n - 2}} = \sqrt{\frac{s_x^2 + s_y^2}{2}}$$

Let $a = \frac{\bar{X} - \bar{Y}}{2}$, so $\bar{X} - a = -(\bar{Y} - a) = b$

$$\therefore (\bar{X} - \bar{Y})^2 = [(\bar{X} - a) - (\bar{Y} - a)]^2$$
$$= b^2 + b^2 - 2b(-b) = 4b^2$$
$$= 2 * [(\bar{X} - a)^2 + (\bar{Y} - a)^2]$$
$$[(\bar{X} - a)^2 + (\bar{Y} - a)^2] = \frac{(\bar{X} - \bar{Y})^2}{2}$$

Now let's rewrite our F statistic where $I = 2, J = n$

$$F_{1,2(N-1)} = \frac{J[(\bar{X} - a)^2 + (\bar{Y} - a)^2]/(I - 1)}{(J - 1)(s_x^2 + s_y^2)/(I * (J - 1))}$$
$$= \frac{\frac{n}{2}(\bar{X} - \bar{Y})^2}{s_p^2} = (t_{2n-2})^2$$

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

Berkeley
UNIVERSITY OF CALIFORNIA

Problem 2: Rice 12.2.22
Hints

Are there systematic differences between the labs?

$\Longleftrightarrow$ Is there any α_i (different effect of the ith 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.

© AG

We fail to reject the null hypothesis.

Berkeley
UNIVERSITY OF CALIFORNIA

Problem 2: Rice 12.2.22
Answer Key

Compute it by hand:

```r
I = ncol(tablet2)
J = nrow(tablet2)
Y_i_bar <- colMeans(tablet2) #This is the mean within each lab
SS_B <- J*sum((Y_i_bar - mean(Y_i_bar))^2) #Has df = I-1
SS_T <- sum((tablet2 - mean(Y_i_bar))^2)
SS_W <- (J-1)*sum(sapply(tablet2, var)) #Has df = (J-1)I
F_stat <- (SS_B/(I-1))/(SS_W/(I*(J-1)))
p_val <- 1 - pf(F_stat, df1 = 6, df2 = 63)
p_val
```

Problem 2: Rice 12.2.22
Answer Key

Compute it using R function:

```
tablet1_tidy <- gather(tablet1, key = lab, value = measurements)
tablet2_tidy <- gather(tablet2, key = lab, value = measurements)

fit1 <- lm(measurements ~ lab, data = tablet1_tidy)
anova(fit1)
fit2 <- lm(measurements ~ lab, data = tablet2_tidy)
anova(fit2)
```
```
Analysis of Variance Table

Response: measurements
          Df  Sum Sq   Mean Sq F value   Pr(>F)
lab        6 0.15343 0.0255724  11.902 6.975e-09 ***
Residuals 63 0.13536 0.0021486
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
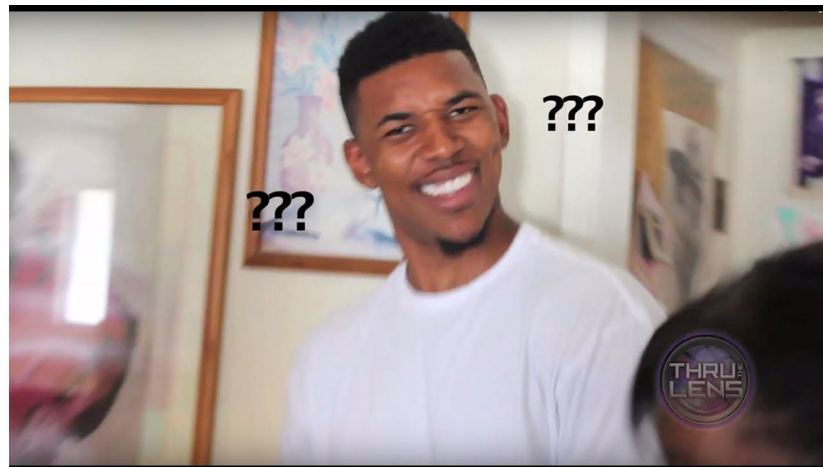
Problem 2: Rice 12.2.22

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??
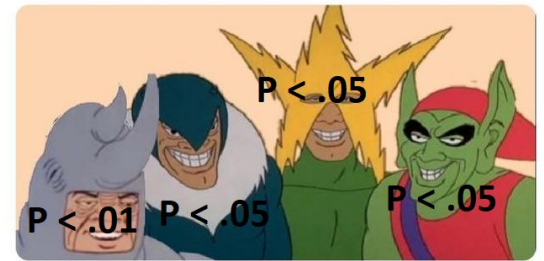
Review: Bonferroni Correction

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

P < .05
P < .01   P < .05   P < .05

Me and the significant boys after Bonferroni correction

P < .01

Review: Bonferroni Correction

A $100(1 - \frac{\alpha}{k})\%$ CI for the ith 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 2: Rice 12.2.22

Answer Key

```{r}
get_bound <- function(vec, lower){ #By default returns the
  center = mean(vec)
  n = length(vec)
  halfwidth = qt(1 - 0.025/7, df = n - 1)*sd(vec)/sqrt(n)
  if(lower == T) {
    return(center - halfwidth)
  }
  else{
    return(center + halfwidth)
  }
}
upperbounds <- sapply(tablet2, get_bound, lower = F)
lowerbounds <- sapply(tablet2, get_bound, lower = T)
```
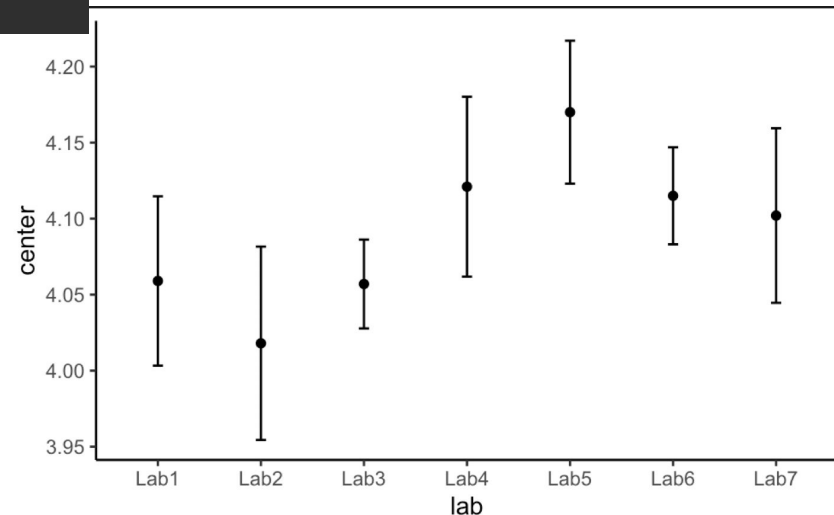
Problem 3

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.

Problem 3

We make a one way ANOVA test with the following assumptions:

1. Data in each treatment group is normally distributed.

2. Data have equal variance across treatment groups.

```r
I = 3
J = 5
SS_W <- (J - 1)*(20^2 + 15^2 + 22^2)
Y_bar_i <- c(135, 160, 155)
Ybar <- mean(Y_bar_i) #Total Mean
SS_B <- J*(sum((Y_bar_i - Ybar)**2))
F_stat <- (SS_B/(I - 1))/(SS_W/(I*(J-1)))
cutoff <- qf(0.95, df1 = I-1, df2 = I*(J-1))
F_stat < cutoff
```