MATH 3339
Statistics for the Sciences
Chapter 11: Comparing More Than Two Means

Wendy Wang
wwang60@central.uh.edu

Lecture 18 - 3339

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Outline

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

You try

In each of the situations determine the type of hypothesis test to be used.

?

1. Quart cartons of milk should contain at least 32 ounces. A sample of 22 cartons contained the following amounts in counces. Does sufficient evidence exist to conclude the mean amount of milk in cartons is less than 32 ounces? The data is: (31.5, 32.2, 31.9, 31.8, 31.7, 32.1, 31.5, 31.6, 32.4, 31.6, 31.8, 32.2, 32.1, 31.8, 31.6, 32.0, 61.6, 31.7, 32.0, 31.9, 31.8, 31.6)
   a) One Sample T Test for Means
   b) One Sample Z Test for Proportions
   c) Two Sample T Test for Means
   d) One Sample Z Test for Means

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

You try

In each of the situations determine the type of hypothesis test to be used.

2. In an experiment on relaxation techniques, subject's brain signals were measured before and after the relaxation exercises with the following results:

Person	1	2	3	4	5
Before	32	38	65	50	30
After	25	35	56	52	24

Is there sufficient evidence to suggest that the relaxation slowed the brain waves? Assume the population is normally distributed.

a) One Sample T Test for Means
b) One Sample Z Test for Proportions
c) Two Sample T Test for Means
d) Matched Pairs T Test

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

You try

3. Suppose you wish to perform a hypothesis test for a population mean. Suppose that the population standard deviation is unknown, the population is skewed to the right, and the sample is large. Would you perform a z-test or t-test?
   a) The t-test is appropriate.
   b) Either test is appropriate.
   c) The z-test is appropriate.
   d) Neither test is appropriate.

4. Which of the following statements are true?
   a) A significance test (p-value = 0.0001) rejected the null hypothesis that the population mean is 35.
   b) A report on a study says that the results are statistically significant and the P-value is 0.85.
      prob(z)
   c) The z-test statistic had a value of 0.023, and the null hypothesis was rejected at the 5% level because 0.023 < 0.05.
   d) All of these are true.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Weight Loss

- From: `http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_HypothesisTesting-ANOVA/BS704_HypothesisTesting-Anova_print.html`
- Is there a difference in the mean weight loss among different programs?
- A clinical trial is run to compare weight loss programs and participants are randomly assigned to one of the comparison programs and are counseled on the details of the assigned program.
- Participants follow the assigned program for 8 weeks.
- Three popular weight loss programs are considered.
  - Low calorie diet.
  - Low fat diet
  - Low carbohydrate diet
  - Control group

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Results

- Response variable = weight loss = weight at the end of 8 weeks - weight at beginning of the study
- The observed weight losses of twenty people in this study are as follows:

Low Calorie	Low Fat	Low Carbohydrate	Control
8	2	3	2
9	4	5	2
6	3	4	-1
7	5	2	0
3	1	3	3

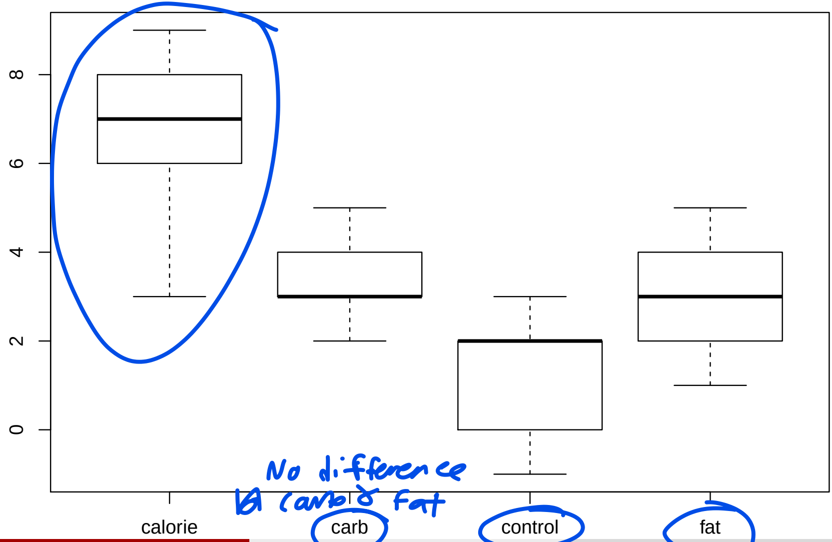
- Is there a statistically significant difference in the mean weight loss among the four diets?

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

H_a: at least 1 is different from others

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Box Plots of Weight Loss



Hypotheses

- We want to know if there is a "statistically significant difference" in the mean weight loss among the four diets.
- Null hypothesis: mean weight loss is the same among the four diets

$$H_0 : \mu_{\text{calorie}} = \mu_{\text{carb}} = \mu_{\text{control}} = \mu_{\text{fat}}$$

- Alternative hypothesis is that at least one of the mean weight loss among the four diets is different.
- Rejecting H_0 is evidence that the mean of at least one group is different from the other means.

Hypotheses

- We want to know if there is a "statistically significant difference" in the mean weight loss among the four diets.

- Null hypothesis: mean weight loss is the same among the four diets

$$H_0 : \mu_{\text{calorie}} = \mu_{\text{carb}} = \mu_{\text{control}} = \mu_{\text{fat}}$$

- Alternative hypothesis is that at least one of the mean weight loss among the four diets is different.

- Rejecting H_0 is evidence that the mean of at least one group is different from the other means.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Hypotheses

- We want to know if there is a "statistically significant difference" in the mean weight loss among the four diets.
- Null hypothesis: mean weight loss is the same among the four diets

$$H_0 : \mu_{\text{calorie}} = \mu_{\text{carb}} = \mu_{\text{control}} = \mu_{\text{fat}}$$

- Alternative hypothesis is that at least one of the mean weight loss among the four diets is different.
- Rejecting H_0 is evidence that the mean of at least one group is different from the other means.

Hypotheses

- We want to know if there is a "statistically significant difference" in the mean weight loss among the four diets.
- Null hypothesis: mean weight loss is the same among the four diets

$$H_0 : \mu_{\text{calorie}} = \mu_{\text{carb}} = \mu_{\text{control}} = \mu_{\text{fat}}$$

- Alternative hypothesis is that at least one of the mean weight loss among the four diets is different.
- Rejecting H_0 is evidence that the mean of at least one group is different from the other means.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Assumptions

The assumptions of analysis of variance are the same as those of the two sample t-test, but they must hold for all k groups.

- The measurements in every group is a SRS.
- We have a Normal distribution for each of the k populations.
- The variance is the same in all k populations.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Analysis: ANOVA

- ANalysis Of VAriance
- We can estimate how much variation among group means *ought* to be present from sampling error alone if the null hypothesis is true.
- ANOVA lets us determine whether there is more variance among the sample means than we would expect by chance alone.

The Formulas

- Let $\bar{X}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$ denote the average of the observations in the ith group.
- Let $N = \sum_{i=1}^{M} n_i$, be the total number of observations in all the M groups.
- Let $\bar{X}_{\cdot\cdot} = \frac{1}{N}\sum_{i=1}^{M} n_i\bar{X}_{i\cdot}$ be the average of all the observations (the grand average)

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

The Formulas

- The **treatment sum of squares** (between groups) is

$$SS(betw) = \sum_{i=1}^{M} n_i(\bar{X}_{i.} - \bar{X}_{..})^2.$$

- The **error sum of squares** (residual) is

$$SSE = SS(resid) = \sum_{i=i}^{M} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i.})^2 = \sum_{i=1}^{M} (n-1)S_i^2.$$

- The **total sum of squares** is

$$SS(tot) = \sum_{i=i}^{M} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{..})^2 = SS(betw) + SS(resid).$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Diets Example

$N=20$ $\bar{x}=3.55$ $m=4$ overall mean

Don't worry for final treatment of squares

Low Calorie	Low Fat	Low Carbohydrate	Control
8	2	3	2
9	4	5	2
6	3	4	-1
7	5	2	0
3	1	3	3

$n_1=5$ $\bar{x}_1=6.6$ $n_2=5$ $\bar{x}_2=3$ $\bar{x}_3=3.4$ $\bar{x}_4=1.2$

$$SS(trt)=\sum_{i=1}^{m} n_i(\bar{x}_i-\bar{x})^2$$

$$=5(6.6-3.55)^2+5(3-3.55)^2+5(3.4-3.55)^2+5(1.2-3.55)^2=\boxed{75.75}$$

$$SSE \sum_{i=1}^{m}(n_i-1)s_i^2=\boxed{47.2}$$

The F Test

- The **mean square for treatments** is $MSTr = \frac{SSTr}{M-1}$.
- The **mean square for error** is $MSE = \frac{SSE}{N-M}$.
- The test statistic is $F = \frac{MSTr}{MSE}$.
- This test statistic has an F distribution with parameters "numerator degrees of freedom" = M - 1 and "denominator degrees of freedom" = N - M. Where N is the total number of observations and M is the number of groups.

$F(M-1, N-M) = F(4-1, 20-4)$

The ANOVA Table

Source of Variation	degrees of freedom	Sum of Squares	Mean Square	F
Treatments	M - 1	SSTr	MSTr	$\frac{MSTr}{MSE}$
Error	N - M	SSE	MSE	
Total	N - 1	SST		

$(\cancel{M}-1)+(N-\cancel{M})=N-1$

$SST = SST_R + SSE$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

ANOVA for Diets

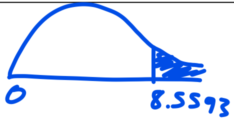
$N-M=20-4=16$
$(N-1)-(M-1)=$

Did in above

Source of Variation	degrees of freedom	Sum of Squares	Mean Square	F
Diets	$4-1=3$	75.75	$\frac{75.75}{3}$	$F=\frac{(75.75/3)}{(47.2/16)}=8.5593$
Error	same $<$ $20-4=16$, $19-3=16$	47.2	$\frac{47.2}{16}$	
Total	$N-1=19$			

$\boxed{\text{p-value = 1 - pf(f,M - 1, N - M)}}$

$=1-pf(8.5593, 3, 16)$
$=0.00128$



UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

```
> diet.lm=lm(Loss~Diet,data=diet)
> anova(diet.lm)
Analysis of Variance Table

Response: Loss
Df Sum Sq Mean Sq F value   Pr(>F)
Diet       3  75.75   25.25  8.5593 0.001278 **
Residuals 16  47.20    2.95
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Tukey's Method (The T Method)

SkIP

The T-method is used to determine which pair (or pairs) of means differs significanctly.

1. Select α, determine $Q_{\alpha,k,N-k}$. in R it is qtukey$(1-\alpha, k, N-k)$ where, k = number of groups and N = total sample size.

2. Calculate $w = Q_{\alpha,k,N-k}\sqrt{\text{MSE}/j}$. Where j = the number of elements in each group.

3. List the sample means in increasing order and underline those pairs that differ by less than w.

4. Any pair of sample means not underscored by the same line corresponds to a pair of population or treatment means that are judged significantly different.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Tukey's Method for Diet Example

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Multiple Comparisons

- If our p-value is small for the ANOVA F test, this implies that at least one of the means is different from the other.
- Which one(s) are different?
- We could do a t-test for each pair of means.
- Problem: when we do multiple t-tests our P(Type 1 error) becomes greater than α.
- Solution: There are methods of adjustments to reduce the significance level of the pairwise test enough so that the probability of one or more type I errors in the whole set of comparisons is less than α.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Multiple Comparisons

- If our p-value is small for the ANOVA F test, this implies that at least one of the means is different from the other.
- Which one(s) are different?
- We could do a t-test for each pair of means.
- Problem: when we do multiple t-tests our P(Type 1 error) becomes greater than α.
- Solution: There are methods of adjustments to reduce the significance level of the pairwise test enough so that the probability of one or more type I errors in the whole set of comparisons is less than α.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Multiple Comparisons

- If our p-value is small for the ANOVA F test, this implies that at least one of the means is different from the other.
- Which one(s) are different?
- We could do a t-test for each pair of means.
- Problem: when we do multiple t-tests our P(Type 1 error) becomes greater than α.
- Solution: There are methods of adjustments to reduce the significance level of the pairwise test enough so that the probability of one or more type I errors in the whole set of comparisons is less than α.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Multiple Comparisons

- If our p-value is small for the ANOVA F test, this implies that at least one of the means is different from the other.
- Which one(s) are different?
- We could do a t-test for each pair of means.
- Problem: when we do multiple t-tests our P(Type 1 error) becomes greater than α.
- Solution: There are methods of adjustments to reduce the significance level of the pairwise test enough so that the probability of one or more type I errors in the whole set of comparisons is less than α.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

The Bonferroni Method

- The **Bonferroni Method** of adjustments reduces the significance level for the pairwise test to α/k, where k is the number of comparisons.
- R Code:

```
> attach(diet)
> pairwise.t.test(Loss,Diet,"bonferroni")

Pairwise comparisons using t tests with pooled SD

data:  Loss and Diet

        calorie  carb     control
carb    0.05695  -        -
control 0.00083  0.35914  -
fat     0.02632  1.00000  0.70193

P value adjustment method: bonferroni
```

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

The Bonferroni Method

- The **Bonferroni Method** of adjustments reduces the significance level for the pairwise test to α/k, where k is the number of comparisons.
- R Code:

```
> attach(diet)
> pairwise.t.test(Loss,Diet,"bonferroni")

Pairwise comparisons using t tests with pooled SD

data:  Loss and Diet

        calorie carb    control
carb    0.05695 -       -
control 0.00083 0.35914 -
fat     0.02632 1.00000 0.70193

P value adjustment method: bonferroni
```

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example MPG

Is there a difference in the average miles per gallon for different makes of automobiles? The following table shows the mean mpg of three different makes of automobiles. The data is on the data sets list called mpg.https://www.math.uh.edu/~wwang/MATH3339_summer2020//mpg.txt

Make	n	$\bar{X}$	S
Honda	5	29.9	1.468
Toyota	6	33.04	2.1173
Nissan	4	29.3	1.3115

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS