

# 10 Tests for Comparing $k$ Populations

MTH 3240 Environmental Statistics

Spring 2020

## Objectives

Objectives:

- Interpret sums of squares, degrees of freedom, and mean squares.
- Carry out a one-factor ANOVA  $F$  test for differences among  $k$  population means.

## Introduction

- We've seen how to test for a difference between **two** population means  $\mu_1$  and  $\mu_2$  (e.g. using the *two-sample t test*).
- We're sometimes interested in testing for differences among **several** population means  $\mu_1, \mu_2, \dots, \mu_k$ .

- (cont'd)

**Example:** Are there differences among **phosphate** concentrations in streams for **three** regions that differ by **land-use type** (*highly urbanized, moderately urbanized, rural*)?

The hypotheses would be

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

$H_a$ : The  $\mu_i$ 's aren't all equal.

where  $\mu_1, \mu_2,$  and  $\mu_3$  are the **population mean phosphate** concentrations for the **three land-use types**.

Notes

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

- The procedures we'll look at can also be used to compare responses to **several treatments** in an **experiment**.

**Example:** Snakes are randomized to **three** treatment groups fed **different levels of selenium (Se)** (1, 10, and 20  $\mu\text{g/g}$ ) to determine the effect (if any) on **Se in their offspring**. The hypotheses would be

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

$H_a$ : The  $\mu_i$ 's aren't all equal.

where  $\mu_1$ ,  $\mu_2$ , and  $\mu_3$  are the **population mean Se** concentrations in offspring of snakes fed the **three Se diets**.

- We'll look at two procedures for testing for **differences** among  $k$  **population means**:

1. The **one-factor ANOVA  $F$  test**.
2. The **Kruskal-Wallis test**.

The **ANOVA  $F$  test** requires a **normality** assumption (or large sample sizes).

The **Kruskal-Wallis test** is a **nonparametric** test (i.e. **doesn't** rely on a normality assumption).

- We'll refer to the  $k$  samples as **groups**.

The **categorical explanatory variable** that distinguishes the **groups** is called the **factor**. Its **levels** are the categories that distinguish the **groups**.

**Example:** In the study of phosphate for three land-use types, the **factor** is **land-use type**, with three **levels**: *highly urbanized, moderately urbanized, and rural*.

**Example:** In the study of Se in offspring of snakes fed three levels of Se, the **factor** is **dietary Se**, with three **levels**: 1, 10, and 20  $\mu\text{g/g}$ .

## Introduction to ANOVA

- **One-factor analysis of variance (ANOVA)** is a procedure for deciding if there are differences among  $k$  **population means**.

The **populations** correspond to **levels** of a **factor**.

Notes

---

---

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

---

---

- The **null hypothesis** is that there are **no differences** among the **population means**  $\mu_1, \mu_2, \dots, \mu_k$ .

**Null Hypothesis:**

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k.$$

- The **alternative hypothesis** is the following.

**Alternative Hypothesis:**

$$H_a : \text{The } \mu_i \text{'s aren't all equal.}$$

Note that  $H_a$  **isn't**

$$H_a : \mu_1 \neq \mu_2 \neq \dots \neq \mu_k.$$

**Example**

A quality assurance study was carried out to compare **lead measurements** made in water sent to  $k = 5$  **laboratories**.

Differences among the five labs' results may signify improperly calibrated equipment or poorly trained technicians.

A vat of wastewater was split into **50** specimens randomized to the labs ( $n = 10$  each) for analysis.

The **lead measurements** ( $\mu\text{g/L}$ ) and their summary statistics are on the next slide.

**Measured Lead Concentrations**

Lab 1	Lab 2	Lab 3	Lab 4	Lab 5
3.4	4.5	5.3	3.2	3.3
3.0	3.7	4.7	3.4	2.4
3.4	3.8	3.6	3.1	2.7
5.0	3.9	5.0	3.0	3.2
5.1	4.3	3.6	3.9	3.3
5.5	3.9	4.5	2.0	2.9
5.4	4.1	4.6	1.9	4.4
4.2	4.0	5.3	2.7	3.4
3.8	3.0	3.9	3.8	4.8
4.2	4.5	4.1	4.2	3.0

$$\begin{aligned} \bar{Y}_1 &= 4.30 & \bar{Y}_2 &= 3.97 & \bar{Y}_3 &= 4.46 & \bar{Y}_4 &= 3.12 & \bar{Y}_5 &= 3.34 \\ S_1 &= 0.904 & S_2 &= 0.440 & S_3 &= 0.642 & S_4 &= 0.764 & S_5 &= 0.737 \end{aligned}$$

Notes

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

Notes

---

---

---

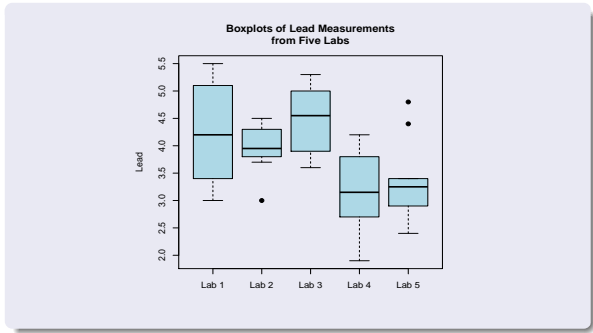
---

---

---

---

---



Notes

---

---

---

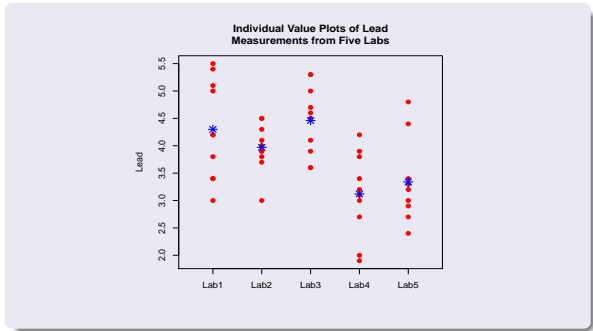
---

---

---

---

---



Notes

---

---

---

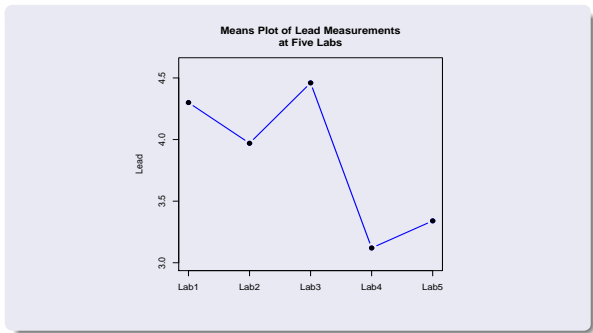
---

---

---

---

---



Notes

---

---

---

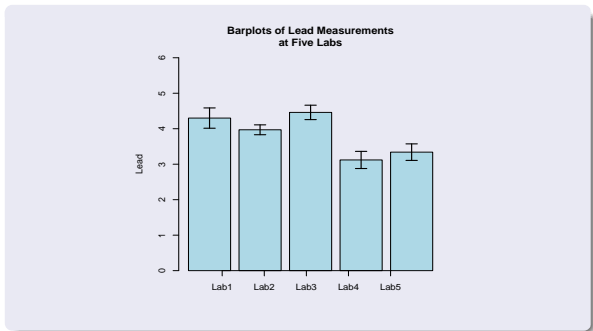
---

---

---

---

---



- We'll suppose we have **random samples** (groups), each of size  $n$ , from  $k$  populations.
- In practice, the sample (group) sizes **don't** all have to have the same, but the notation gets messier.
- Also, in practice, the samples (groups) could be **treatment groups** in an **experiment**.

- Notation:**
  - $k$  = The number of populations being compared.
  - $n$  = The common sample size for the  $k$  groups (samples).
  - $Y_{ij}$  = The  $j$ th observation in the  $i$ th group.  
The first subscript,  $i$ , indicates the **group**, taking values  $1, 2, \dots, k$ . The second,  $j$ , distinguishes **individuals** within a group, taking values  $1, 2, \dots, n$ .

- (cont'd):
  - $\bar{Y}_i$  = The sample mean for the  $i$ th group (the ***i*th group mean** or ***i*th factor level mean**).
  - $S_i$  = The sample standard deviation for the  $i$ th group.
  - $N$  = The **overall sample size** for the  $k$  groups combined.  
Note:  $N = kn$ .
  - $\bar{Y}$  = The **overall sample mean** (of all  $N$  observations in the  $k$  groups combined).

Group 1	Group 2	...	Group $k$
$Y_{11}$	$Y_{21}$	...	$Y_{k1}$
$Y_{12}$	$Y_{22}$	...	$Y_{k2}$
$\vdots$	$\vdots$	...	$\vdots$
$Y_{1n}$	$Y_{2n}$	...	$Y_{kn}$
$\bar{Y}_1$	$\bar{Y}_2$	...	$\bar{Y}_k$
$S_1$	$S_2$	...	$S_k$

Notes

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

- When the group sizes are equal, the **overall mean**  $\bar{Y}$  is the **average** of the  $k$  **group means**:

$$\bar{Y} = \frac{1}{k} \sum_{i=1}^k \bar{Y}_i .$$

**Example**

For the lead measurements made at five labs,

$$n = 10 \quad \text{and} \quad N = 50,$$

and the **overall mean** is obtained either by averaging the **50** lead measurements or, as below, by averaging the **five group means**:

$$\begin{aligned} \bar{Y} &= \frac{1}{k} \sum_{i=1}^k \bar{Y}_i \\ &= \frac{1}{5} (4.30 + 3.97 + 4.46 + 3.12 + 3.34) \\ &= \mathbf{3.838} . \end{aligned}$$

- Key Idea:** In ANOVA, we'll decide if differences among the group means are **statistically significant** by comparing the variation **between groups** to the variation **within groups**.
- Between-groups variation** refers to variation among *group means*.
- Within-groups variation** refers to variation of *individual observations* within the groups.

- On the next slide, the **between-groups** variation is the same for the two plots, but the **within-groups** variation is smaller in the first plot.  
  
Differences among the three groups turn out to be statistically significant in the first plot, but not the second.

Notes

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

Notes

---

---

---

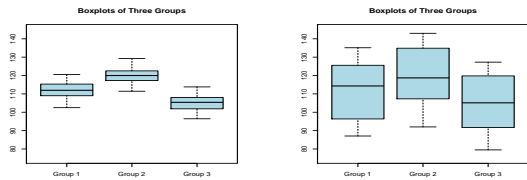
---

---

---

---

---



**Example**

For the lead measurement data from the five labs, we want to decide if there are **any differences** among the **five labs**. So we're testing

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

$$H_a : \text{The } \mu_i \text{'s aren't all equal}$$

where  $\mu_1, \mu_2, \dots, \mu_5$  are the true (unknown) **population mean** measurement results for the five labs.

The null hypothesis says there are **no differences** among the five labs. The alternative says **at least two** of the labs **differ** from each other.

Statistical software produces the following output called an **ANOVA table**.

Source	DF	SS	MS	F	P-value
Lab	4	13.813	3.453	6.77	0.000
Error	45	22.945	0.510		
Total	49	36.758			

We'll see how each value was computed and how to interpret them later.

For now, just note that because the p-value is less than  $\alpha = 0.05$ , we **reject the null hypothesis**.

There are **statistically significant differences** among the five labs' mean lead measurement results.

**Sums of Squares**

- We'll measure **between-groups variation** and **within-groups variation** using **sums of squares**.
- Recall that one measure of variation in a data set  $X_1, X_2, \dots, X_n$  is the **sample variance** (squared standard deviation),

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

Its numerator is an example of a **sum of squares**.

Notes

Notes

Notes

Notes

- We measure **between-groups variation** by the **treatment sum of squares**, denoted **SSTr**.

**Treatment Sum of Squares:**

$$SSTr = n \sum_{i=1}^k (\bar{Y}_i - \bar{Y})^2.$$

**SSTr** will be **large** if there are **large differences** among the group means  $\bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_k$  (which would suggest there are differences among the population means  $\mu_1, \mu_2, \dots, \mu_k$ ).

- We measure **within-groups variation** by the **error sum of squares**, denoted **SSE**.

**Error Sum of Squares:**

$$SSE = \sum_{i=1}^k \sum_{j=1}^n (Y_{ij} - \bar{Y}_i)^2.$$

The **error sum of squares** is sometimes also called the **residual sum of squares**. It reflects **random variation** in the data.

**SSE** will be **large** if there's **substantial** variation among individual observations ( $Y_{ij}$ 's) **within** groups, and small otherwise.

## ANOVA Partition of the Variation

- We also measure the **total variation** in the data using the so-called **total sum of squares**, denoted **SSTo**.

**Total Sum of Squares:**

$$SSTo = \sum_{i=1}^k \sum_{j=1}^n (Y_{ij} - \bar{Y})^2.$$

**SSTo** reflects **total variation** due both to differences among the group means **and** random variation among individual observations within groups.

- It can be shown that

**ANOVA Partition:**

$$SSTo = SSTr + SSE.$$

This splits the total variation in the data as:

$$\text{Total Variation} = \text{Between-Groups Variation} + \text{Within-Groups Variation}$$

Notes

---

---

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

---

---



Example

For the data on lead measurements at five labs, statistical software gives the following values for the sums of squares.

$$SSTo = 36.758$$

$$SSTr = 13.813$$

$$SSE = 22.945.$$

We see that indeed  $SSTo = SSTr + SSE$  since

$$36.758 = 13.813 + 22.945.$$

This indicates that more than a third of the total variation in the lead measurements (**13.813** out of **36.758**) is the result of differences among the five labs' group means.

## Degrees of Freedom

- Associated with each sum of squares is a quantity called its **degrees of freedom** (abbreviated **df**).
- **Df** will be used later to determine the **p-value** of the **ANOVA F test**.
- The **df** of a sum of squares is the **number of deviations**, among those used to compute the sum of squares, that are **"free to vary."**

- To illustrate, recall that the **sample variance** (squared standard deviation) of a data set  $X_1, X_2, \dots, X_n$  is

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1},$$

and the numerator is a **sum of squares**.

The (non-squared) deviations  $X_i - \bar{X}$  used to compute the sum of squares always **add up to zero**.

Thus any  $n - 1$  of the deviations determine the remaining one, so only  $n - 1$  are **"free to vary"**

The **df** associated with the sum of squares is therefore  **$n - 1$** .

Notes

---

---

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

---

---

- Here are the **df** associated with each of the sums of squares  $SSTo$ ,  $SSTr$ , and  $SSE$ .

**Degrees of Freedom:** For one-factor ANOVA, the degrees of freedom are:

$$df \text{ for } SSTo = N - 1$$

$$df \text{ for } SSTr = k - 1$$

$$df \text{ for } SSE = k(n - 1) = N - k$$

- The **df**, like the associated sums of squares, are additive in the following sense.

**Additivity of Degrees of Freedom:**

$$df \text{ for } SSTo = df \text{ for } SSTr + df \text{ for } SSE.$$

**Example**

For the study of lead measurements at five labs, we have  $k = 5$ ,  $n = 10$ , and  $N = 50$ , so

$$df \text{ for } SSTo = N - 1 = 49,$$

$$df \text{ for } SSTr = k - 1 = 4,$$

and

$$df \text{ for } SSE = N - k = 45.$$

As expected, the **df** are additive since

$$49 = 4 + 45.$$

## Mean Squares

- **SSTr** measures **between-groups** variation and **SSE** measures **within-groups** variation.
- But  $SSTr$  and  $SSE$  aren't directly comparable to each other (they're affected differently by  $n$  and  $k$ ).
- Instead, we'll compare so-called **mean squares** (next slide).

## Notes

---

---

---

---

---

---

---

---

## Notes

---

---

---

---

---

---

---

---

## Notes

---

---

---

---

---

---

---

---

## Notes

---

---

---

---

---

---

---

---

## Mean Squares

- A **mean square** is defined as a **sum of squares** divided by its **df**:

$$\text{Mean Square} = \frac{\text{Sum of Squares}}{\text{Degrees of Freedom}}$$

**Example:** The **sample variance**  $S^2$  is a familiar example of a **mean square**.

- For one-factor ANOVA, the **mean square for treatments** **MSTr** is

**Mean Square for Treatments:**

$$\text{MSTr} = \frac{\text{SSTr}}{k - 1}$$

and the **mean squared error** **MSE** is

**Mean Squared Error:**

$$\text{MSE} = \frac{\text{SSE}}{N - k}$$

- **MSTr** and **MSE** are measures of **between-groups** and **within-groups variation** that are *directly comparable* to each other.
- They'll be used to compute the **ANOVA  $F$  test statistic** (next slide).

## The ANOVA $F$ Test

- Here's the **one-factor ANOVA  $F$  test statistic**:

**One-Factor ANOVA  $F$  Test Statistic:**

$$F = \frac{\text{MSTr}}{\text{MSE}}$$

Notes

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

- We can think of  $F$  as

$$F = \frac{\text{Between-Groups Variation}}{\text{Within-Groups Variation}}$$

Notes

---

---

---

---

---

---

---

---

- The **between-groups variation** reflects variation among the group means,  $\bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_k$ , which in turn reflects differences among  $\mu_1, \mu_2, \dots, \mu_k$ .
  - If  $H_0$  was true, ...
    - ... it can be shown that the **between- and within-groups** variation would be **approximately equal**.
  - But if  $H_a$  was true, ...
    - ... the **between-groups** variation would be **larger** than **within-groups**.
- Thus ...
  1.  $F$  will be approximately **1** (most likely) if  $H_0$  is true.
  2.  $F$  will be **greater than 1** (most likely) if  $H_a$  is true.

Notes

---

---

---

---

---

---

---

---

Large values of  $F$  (larger than 1) provide evidence in favor of  $H_a$ : The  $\mu_i$ 's aren't all equal.

Notes

---

---

---

---

---

---

---

---

- Now suppose the  $k$  samples are from **normal** populations that have the **same standard deviation**  $\sigma$ .  
If either
  - 1 The populations are all normal, or
  - 2 The sample size  $n$  for each sample is large,
 the **null distribution** is as follows.

Notes

---

---

---

---

---

---

---

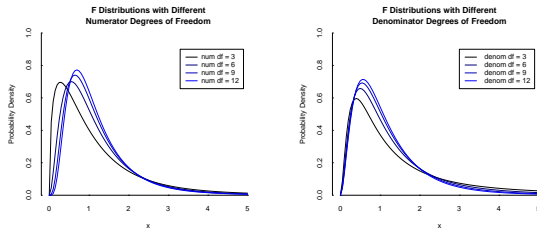
---

**Sampling Distribution of  $F$  Under  $H_0$ :** If  $F$  is the one-factor ANOVA  $F$  test statistic, then when

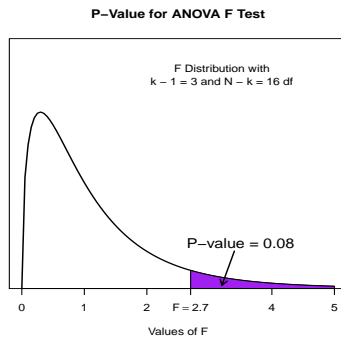
$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

is true, the  $F$  follows a so-called  $F$  **distribution** with  $k - 1$  **numerator df** and  $N - k$  **denominator df**. We write this as

$$F \sim F(k - 1, N - k).$$



- **P-values** and **rejection regions** are obtained from the *upper tail* of the  $F(k - 1, N - k)$  **distribution**, as shown on the next slide (when the observed test statistic value is  $F = 2.7$ ).



Notes

Notes

Notes

Notes

### One-Factor ANOVA $F$ Test for $\mu_1, \mu_2, \dots, \mu_k$

**Assumptions:** Data are random samples from  $k$  populations (or responses of individuals to  $k$  treatments in a randomized experiment), the one-factor ANOVA model is appropriate, the  $\epsilon_{ij}$ 's are independent and either they follow a  $N(0, \sigma)$  distribution or the  $k$  sample sizes  $n$  are all large.\*

**Null hypothesis:**  $H_0: \mu_1 = \mu_2 = \dots = \mu_k$  or  $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$ .

**Test statistic value:**  $F = \frac{MSTr}{MSE}$ .

**Decision rule:** Reject  $H_0$  if p-value  $< \alpha$  or  $F$  is in rejection region.

Alternative hypothesis	P-value = area under $F$ -distribution with $k - 1$ and $N - k$ d.f.:	Rejection region = $F$ values such that:**
$H_a$ : Not all $\mu_i$ 's are equal or $H_a$ : Not all $\alpha_i$ 's equal 0	to the right of $F$	$F \geq F_{\alpha, k-1, N-k}$

\* The ANOVA  $F$  test can also be carried out (using statistical software) when the  $k$  sample sizes  $n_1, n_2, \dots, n_k$  aren't all the same. The sample sizes are considered to be large if they're all at least 15, unless the samples exhibit strong skewness, in which case they should all be at least 40.

\*\*  $F_{\alpha, k-1, N-k}$  is the  $100(1 - \alpha)$ th percentile of the  $F$  distribution with  $k - 1$  and  $N - k$  d.f.

### The ANOVA Table

- The results of an **analysis of variance** are summarized in a **one-factor ANOVA table** having the form shown below.

#### One-Factor ANOVA Table:

Source of Variation	DF	SS	MS	F	P-value
Factor	$k - 1$	SSTr	$MSTr = SSTr / (k - 1)$	$F = MSTr / MSE$	p
Error	$N - k$	SSE	$MSE = SSE / (N - k)$		
Total	$N - 1$	SSTo			

#### Example

For the lead measurement data from the five labs, we want to decide if there are **any differences** among the **five labs**. So we're testing

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

$$H_a: \text{The } \mu_i\text{'s aren't all equal}$$

where  $\mu_1, \mu_2, \dots, \mu_5$  are the true (unknown) **population mean** measurement results for the five labs.

The null hypothesis says there are **no differences**. The alternative says **at least two** of the labs **differ** from each other.

#### Notes

#### Notes

#### Notes

#### Notes

Here's the **ANOVA table** (again), obtained using statistical software.

Source	DF	SS	MS	F	P-value
Lab	4	13.813	3.453	6.77	0.000
Error	45	22.945	0.510		
Total	49	36.758			

Using a level of significance  $\alpha = 0.05$ , since the p-value is less than  $\alpha$ , we **reject the null hypothesis**.

There are **statistically significant differences** among the five labs' mean lead measurement results.

(Later, we'll see how to check the **normality** and **common standard deviation** assumptions to validate these hypothesis test results.)

Notes

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---

Notes

---

---

---

---

---

---

---

---