

R06 - ANOVA and F-tests

STAT 5870 (Engineering)
Iowa State University

November 18, 2024

One-way ANOVA model/assumptions

The one-way ANOVA (ANalysis Of VAriance) model is

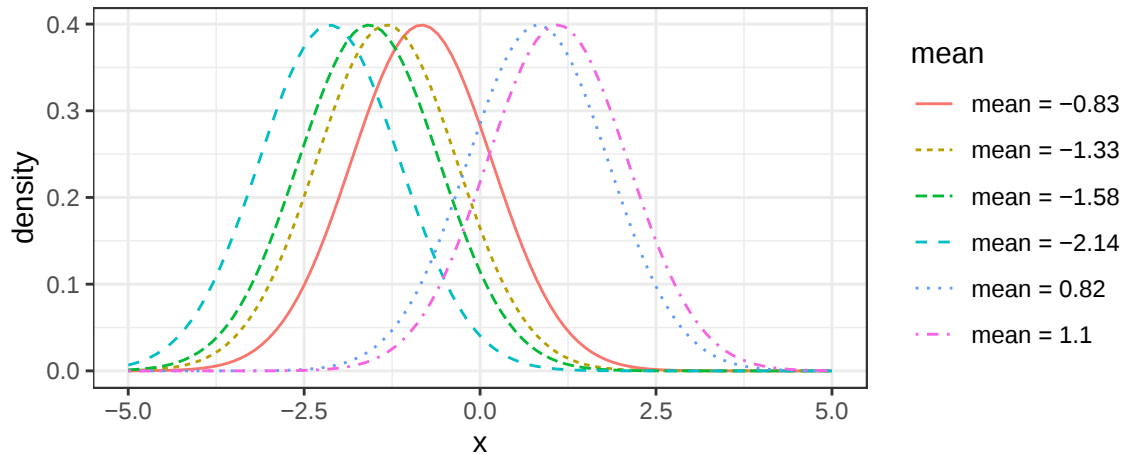
$$Y_{ig} \stackrel{ind}{\sim} N(\mu_g, \sigma^2) \quad \text{or} \quad Y_{ig} = \mu_g + \epsilon_{ig}, \epsilon_{ig} \stackrel{iid}{\sim} N(0, \sigma^2)$$

for $g = 1, \dots, G$ and $i = 1, \dots, n_g$.

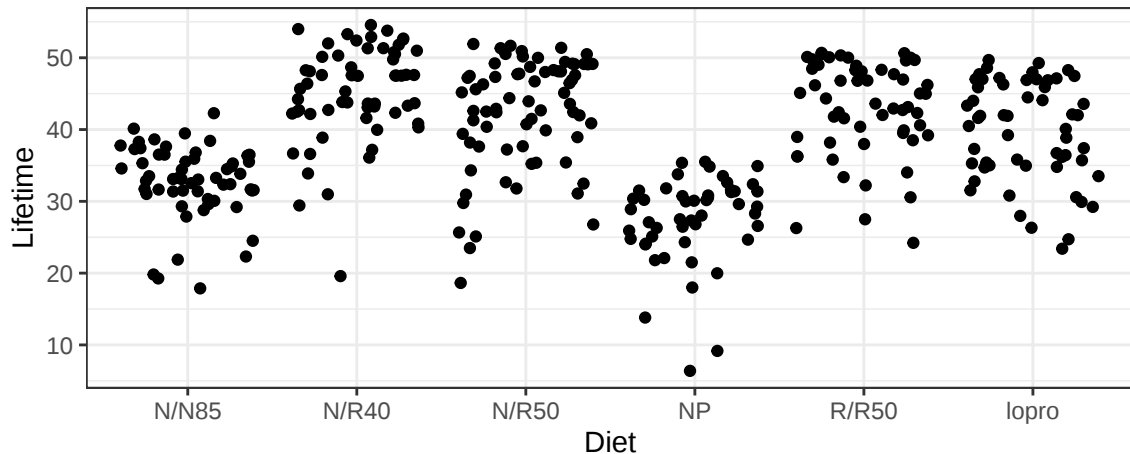
Assumptions:

- Errors
 - are normally distributed.
 - have a common variance.
 - are independent.
- Each group has its own mean.

ANOVA assumptions graphically



Consider the mice data set



One-way ANOVA F-test

Are any of the means different?

Hypotheses in English:

H_0 : all the means are the same

H_1 : at least one of the means is different

Statistical hypotheses:

$$\begin{array}{ll} H_0 : & \mu_g = \mu \text{ for all } g & Y_{ig} \stackrel{iid}{\sim} N(\mu, \sigma^2) \\ H_1 : & \mu_g \neq \mu_{g'} \text{ for some } g \text{ and } g' & Y_{ig} \stackrel{ind}{\sim} N(\mu_g, \sigma^2) \end{array}$$

An ANOVA table organizes the relevant quantities for this test and computes the pvalue.

ANOVA table

A start of an ANOVA table:

Source of variation	Sum of squares	d.f.	Mean square
Factor A (Between groups)	$SSA = \sum_{g=1}^G n_g (\bar{Y}_g - \bar{Y})^2$	$G - 1$	$\frac{SSA}{G-1}$
Error (Within groups)	$SSE = \sum_{g=1}^G \sum_{i=1}^{n_g} (Y_{ig} - \bar{Y}_g)^2$	$n - G$	$\frac{SSE}{n-G} (= \hat{\sigma}^2)$
Total	$SST = \sum_{g=1}^G \sum_{i=1}^{n_g} (Y_{ig} - \bar{Y})^2$	$n - 1$	

where

- G is the number of groups,
- n_g is the number of observations in group g ,
- $n = \sum_{g=1}^G n_g$ (total observations),
- $\bar{Y}_g = \frac{1}{n_g} \sum_{i=1}^{n_g} Y_{ig}$ (average in group g),
- and $\bar{Y} = \frac{1}{n} \sum_{g=1}^G \sum_{i=1}^{n_g} Y_{ig}$ (overall average).

ANOVA table

An easier to remember ANOVA table:

Source of variation	Sum of squares	df	Mean square	F-statistic	p-value
Factor A (between groups)	SSA	$G - 1$	$MSA = SSA/G - 1$	MSA/MSE	(see below)
Error (within groups)	SSE	$n - G$	$MSE = SSE/n - G$		
Total	$SST = SSA + SSE$	$n - 1$			

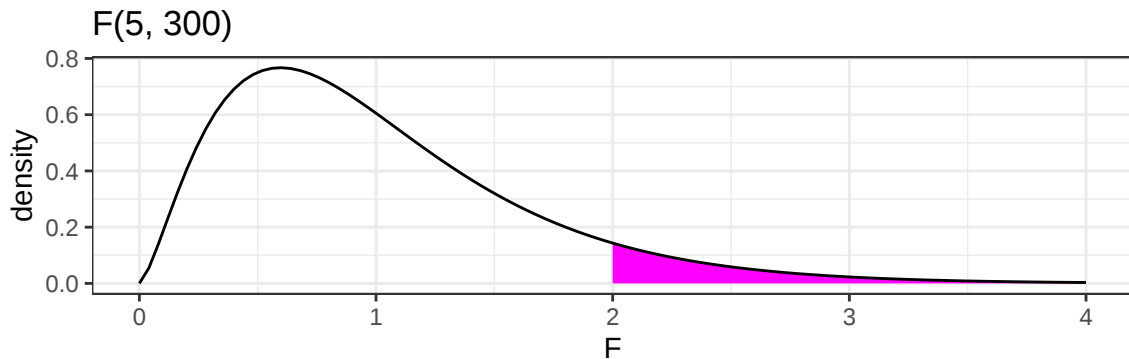
Under H_0 ($\mu_g = \mu$),

- the quantity MSA/MSE has an F-distribution with $G - 1$ numerator and $n - G$ denominator degrees of freedom,
- larger values of MSA/MSE indicate evidence against H_0 , and
- the p-value is determined by $P(F_{G-1, n-G} > MSA/MSE)$.

F-distribution

F -distribution has two parameters:

- numerator degrees of freedom (ndf)
- denominator degrees of freedom (ddf)



One-way ANOVA F-test (by hand)

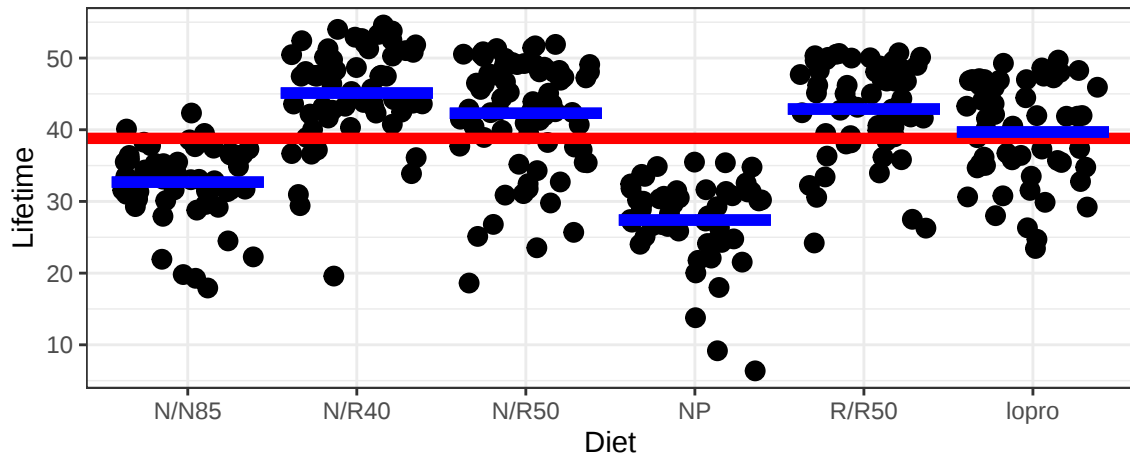
```
# A tibble: 7 x 4
  Diet      n mean  sd
<chr> <int> <dbl> <dbl>
1 N/N85    57 32.7  5.13
2 N/R40    60 45.1  6.70
3 N/R50    71 42.3  7.77
4 NP       49 27.4  6.13
5 R/R50    56 42.9  6.68
6 lopro    56 39.7  6.99
7 Total   349 38.8  8.97
```

So

$$\begin{aligned}
 SSA &= 57 \times (32.7 - 38.8)^2 + 60 \times (45.1 - 38.8)^2 + 71 \times (42.3 - 38.8)^2 + 49 \times (27.4 - 38.8)^2 \\
 &\quad + 56 \times (42.9 - 38.8)^2 + 56 \times (39.7 - 38.8)^2 = 12734 \\
 SST &= (349 - 1) \times 8.97^2 = 28000 \\
 SSE &= SST - SSA = 28000 - 12734 = 15266 \\
 G - 1 &= 5 \\
 n - G &= 349 - 6 = 343 \\
 n - 1 &= 348 \\
 MSA &= SSA / G - 1 = 12734 / 5 = 2547 \\
 MSE &= SSE / n - G = 15266 / 343 = 44.5 = \hat{\sigma}^2 \\
 F &= MSA / MSE = 2547 / 44.5 = 57.2 \\
 p &= P(F_{5,343} > 57.2) < 0.0001
 \end{aligned}$$

F statistic is off by 0.1 relative to the table later, because of rounding of 8.97. The real SST is 28031 which would be the F statistic of 57.1.

Graphical comparison



R code and output for one-way ANOVA

```
m <- lm(Lifetime ~ Diet, case0501)
anova(m)
```

Analysis of Variance Table

Response: Lifetime

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Diet	5	12734	2546.8	57.104	< 2.2e-16 ***
Residuals	343	15297	44.6		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

There is evidence against the null model $Y_{ig} \stackrel{ind}{\sim} N(\mu, \sigma^2)$, i.e. our data seem incompatible with this model.

General F-tests

The one-way ANOVA F-test is an example of a general hypothesis testing framework that uses F-tests. This framework can be used to test

- composite alternative hypotheses or, equivalently,
- a full vs a reduced model.

The general idea is to balance the amount of variability remaining when moving from the reduced model to the full model measured using the sums of squared errors (SSEs) relative to the amount of complexity, i.e. parameters, added to the model.

Testing full vs reduced models

If $Y_{ig} \stackrel{ind}{\sim} N(\mu_g, \sigma^2)$ for $g = 1, \dots, G$ and we want to test the hypotheses

- $H_0 : \mu_g = \mu$ for all g
- $H_1 : \mu_g \neq \mu_{g'}$ for some g and g'

think about this as two models:

- $H_0 : Y_{ig} \stackrel{ind}{\sim} N(\mu, \sigma^2)$ (reduced)
- $H_1 : Y_{ig} \stackrel{ind}{\sim} N(\mu_g, \sigma^2)$ (full)

We can use an F-test to calculate a p-value for tests of this type.

Nested models: full vs reduced

Two models are **nested** if the **reduced** model is a special case of the **full** model.

For example, consider the full model

$$Y_{ig} \stackrel{ind}{\sim} N(\mu_g, \sigma^2).$$

One special case of this model occurs when $\mu_g = \mu$ and thus

$$Y_{ig} \stackrel{ind}{\sim} N(\mu, \sigma^2).$$

is a reduced model and these two models are nested.

Calculating the sum of squared residuals (errors)

Model	Full	Reduced
Assumption	$H_1 : Y_{ig} \overset{iid}{\sim} N(\mu_g, \sigma^2)$	$H_0 : Y_{ig} \overset{iid}{\sim} N(\mu, \sigma^2)$
Mean	$\hat{\mu}_g = \bar{Y}_g = \frac{1}{n_g} \sum_{i=1}^{n_g} Y_{ig}$	$\hat{\mu} = \bar{Y} = \frac{1}{n} \sum_{g=1}^G \sum_{i=1}^{n_g} Y_{ig}$
Residual	$r_{ig} = Y_{ig} - \hat{\mu}_g = Y_{ig} - \bar{Y}_g$	$r_{ig} = Y_{ig} - \hat{\mu} = Y_{ig} - \bar{Y}$
SSE	$\sum_{g=1}^G \sum_{i=1}^{n_g} r_{ig}^2$	$\sum_{g=1}^G \sum_{i=1}^{n_g} r_{ig}^2$

General F-tests

Do the following

1. Calculate

Extra sum of squares =

Residual sum of squares (reduced) - Residual sum of squares (full)

2. Calculate

Extra degrees of freedom =

of mean parameters (full) - # of mean parameters (reduced)

3. Calculate F-statistics

$$F = \frac{\text{Extra sum of squares} / \text{Extra degrees of freedom}}{\text{Estimated residual variance in full model } (\hat{\sigma}^2)}$$

4. A pvalue is $P(F_{\text{ndf}, \text{ddf}} > F)$

- numerator degrees of freedom (ndf) = Extra degrees of freedom
- denominator degrees of freedom (ddf): df associated with $\hat{\sigma}^2$

Mice lifetimes

Consider the hypothesis that mice on all diets have a common mean lifetime except NP.

Let

$$Y_{ig} \stackrel{ind}{\sim} N(\mu_g, \sigma^2)$$

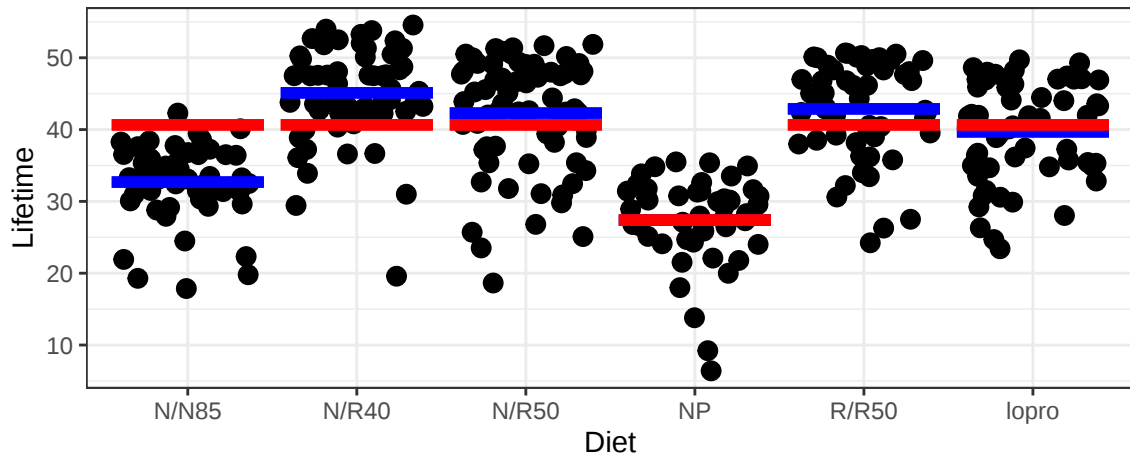
with $g = 1$ being the NP group then the hypotheses are

- $H_0 : \mu_g = \mu$ for $g \neq 1$
- $H_1 : \mu_g \neq \mu_{g'}$ for some $g, g' = 2, \dots, 6$

As models:

- $H_0 : Y_{i1} \stackrel{iid}{\sim} N(\mu_1, \sigma^2)$ and $Y_{ig} \stackrel{iid}{\sim} N(\mu, \sigma^2)$ for $g \neq 1$
- $H_1 : Y_{ig} \stackrel{ind}{\sim} N(\mu_g, \sigma^2)$

As a picture



Making R do the calculations

```
case0501$NP = factor(case0501$Diet == "NP")

modR = lm(Lifetime ~ NP, case0501) # (R)educed model
modF = lm(Lifetime ~ Diet, case0501) # (F)ull model
anova(modR,modF)
```

Analysis of Variance Table

Model 1: Lifetime ~ NP

Model 2: Lifetime ~ Diet

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	347	20630				
2	343	15297	4	5332.2	29.89	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Lack-of-fit F-test for linearity

Let Y_{ig} be the i^{th} observation from the g^{th} group where the group is defined by those observations having the same explanatory variable value (X_g).

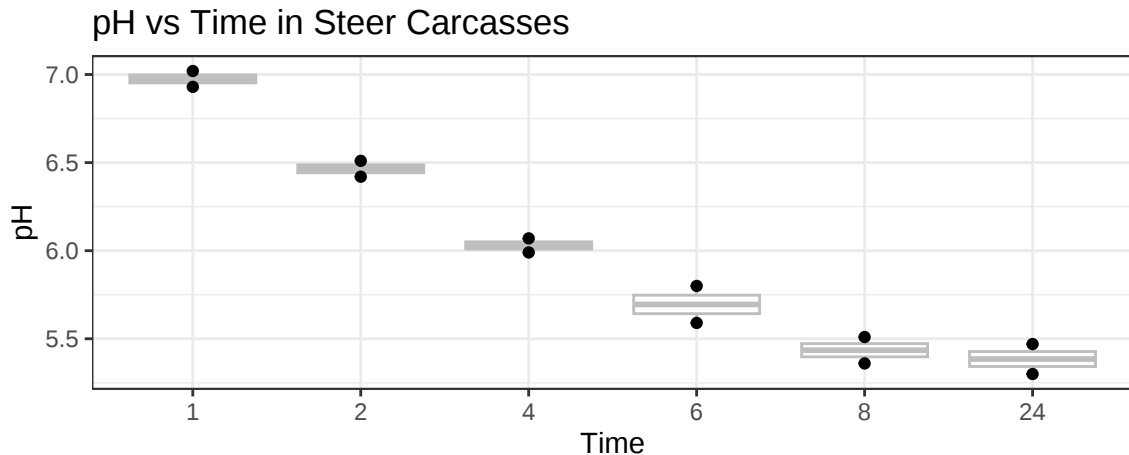
Two models:

ANOVA: $Y_{ig} \stackrel{ind}{\sim} N(\mu_g, \sigma^2)$ (full)

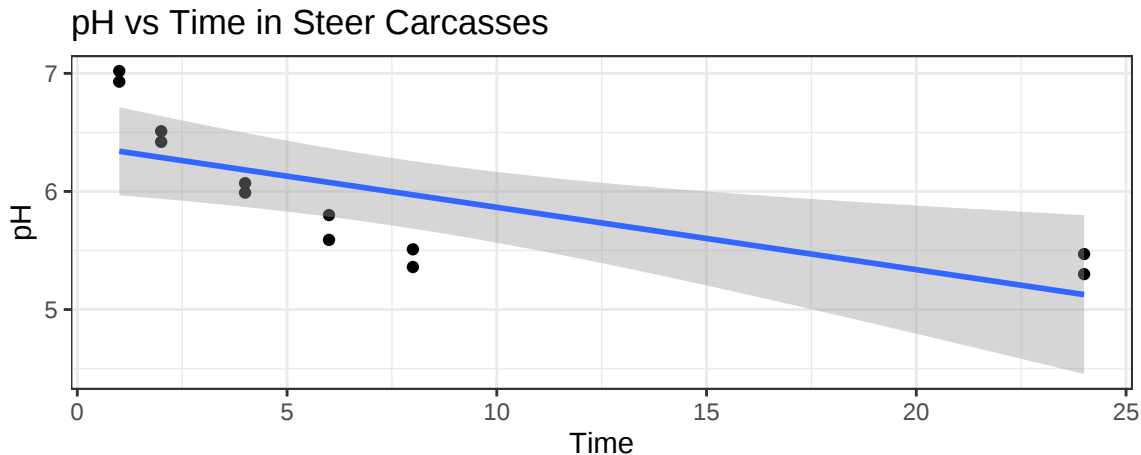
Regression: $Y_{ig} \stackrel{ind}{\sim} N(\beta_0 + \beta_1 X_g, \sigma^2)$ (reduced)

- Regression model is reduced:
 - ANOVA has G parameters for the mean
 - Regression has 2 parameters for the mean
 - Set $\mu_g = \beta_0 + \beta_1 X_g$.
- Small pvalues indicate a lack-of-fit, i.e. the regression (reduced) model is not adequate.
- Lack-of-fit F-test requires multiple observations at a few X_g values!

pH vs Time - ANOVA



pH vs Time - Regression



Lack-of-fit F-test in R

```
# Use as.factor to turn a continuous variable into a categorical variable
m_anova = lm(pH ~ as.factor(Time), Sleuth3::ex0816)
m_reg   = lm(pH ~ Time, Sleuth3::ex0816)
anova(m_reg, m_anova)
```

Analysis of Variance Table

Model 1: pH ~ Time

Model 2: pH ~ as.factor(Time)

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	10	1.97289				
2	6	0.05905	4	1.9138	48.616	0.0001048 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

There is evidence the data are incompatible with the null hypothesis that states the means of each group fall along a line.

Summary

- Use F-tests for comparison of full vs reduced model
 - One-way ANOVA F-test
 - General F-tests
 - Lack-of-fit F-tests

Think about F-tests as comparing models.