One-way ANOVA:
Comparing several means

PSLS chapter 24

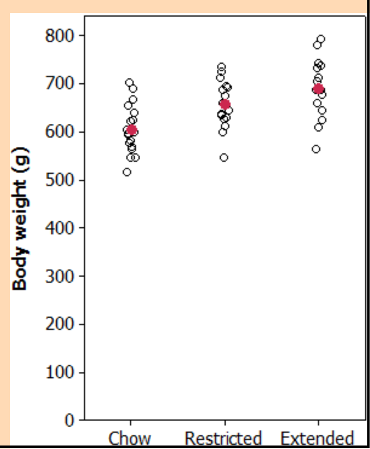
Male lab rats were randomly assigned to 3 groups differing in access to food for 40 days. The rats were weighted (in grams) at the end.

- Group 1: access to chow only.
- Group 2: chow plus access to cafeteria food restricted to 1 hour every day.
- Group 3: chow plus extended access to cafeteria food (all day long every day).

Cafeteria food: bacon, sausage, cheesecake, pound cake, frosting and chocolate

Chow	Restricted	Extended
516	546	564
547	599	611
546	612	625
564	627	644
577	629	660
570	638	679
582	645	687
594	635	688
597	660	706
599	676	714
606	695	738
606	687	733
624	693	744
623	713	780
641	726	794
655	736	
667		
690		
703		

	N	Mean	StDev
Chow	19	605.63	49.64
Restricted	16	657.31	50.68
Extended	15	691.13	63.41



Handling multiple comparisons statistically

- The **first** step in examining multiple populations statistically is to test for an **overall** statistical significance as evidence of any difference among the parameters we want to compare.

  ➔ **ANOVA *F* test**

- **If** that overall test showed statistical significance, then a detailed follow-up analysis can examine all pair-wise parameter comparisons to define which parameters differ from which and by how much.

  ➔ *more complex methods (see Chapter 26)*

Reminder: Hypotheses tests

A null hypothesis H_0: a hypothesized model of the population distribution(s) specifying the parameter value(s)

A test statistic: a measure of how much the sample statistic(s) deviate from the parameter(s) under H_0, relative to how much variability is to be expected for the statistic(s)

A test *P*-value: the probability of obtaining a test statistic at least as extreme as that computed, if H_0 was true

related measures

The ANOVA F test

The **ANOVA *F* test** for comparing k means tests:

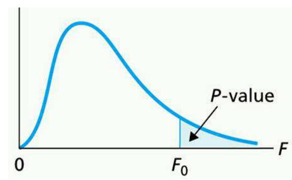
H_0: All means μ_i are equal ($\mu_1 = \mu_2 \ldots = \mu_k$)

H_a: At least one mean μ_i differs (H$_0$ is not true)

The ***F* statistic** compares the variation of the sample means with the variation of individuals in the samples.

$$\boxed{F = \frac{\text{MSG}}{\text{MSE}}}$$

The test P-value is the probability, if H_0 was true, of obtaining a test statistic at least as extreme as that computed. When H_0 is true, F has the ***F* distribution** with degrees of freedom $k - 1$ and $N_{\text{total}} - k$.

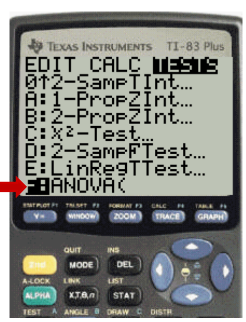


TI calculator : one-way ANOVA for means

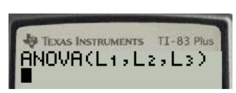
If you have the raw data, enter each random sample into its own list. Then

STAT then

ANOVA



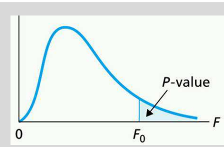
Identify the k lists containing the k random samples, with a comma between lists



See videos for TI-83 and CrunchIt! on Sapling

If you are given F and can find the degrees of freedom,

2nd DISTR: **Fcdf**(F statistic, 1E99, df$_{num}$, df$_{den}$) = P-value.



FYI

$$F = \frac{\text{MSG}}{\text{MSE}} = \frac{\text{variation among sample means}}{\text{variation among individuals within samples}}$$

Source of variation	Sum of squares SS	df	Mean square MS	*F*	*P*-value
Between "groups"	$\sum n_i(\bar{x}_i - \bar{x})^2$	$k - 1$	MSG = SSG/df_G	$F = \frac{\text{MSG}}{\text{MSE}}$	Tail area above *F*
Within groups, "error"	$\sum (n_i - 1)s_i^2$	$N_{total} - k$	MSE = SSE/df_E		
Total	SST = SSG + SSE $\sum (x_{ij} - \bar{x})^2$	$N_{total} - 1$			

R^2 = SSG/SST **Root MSE = s_p (pooled standard deviation)**

MSG: mean square for groups. It measures the variability of the sample means.

MSE: mean square for error, or **pooled sample variance s_p^2**. It measures the variability within the samples.

ANOVA assumptions

1. The *k* samples must be **independent random samples.** The individuals in each sample are completely unrelated.

2. Each population represented by the *k* samples must be **normally distributed**. However, the test is robust to deviations from normality (skew, mild outliers) for large enough samples thanks to the central limit theorem.

3. The ANOVA F-test requires that **all *k* populations have the same standard deviation σ.**

There are tests to check for equality of variance. However, they are tricky and may be sensitive to deviations from the normality assumption or require equal sample sizes.

A conservative approach: The ANOVA *F* test is approximately correct when the largest sample standard deviation is no more than ~ 2 times the smallest sample standard deviation.

Equal sample sizes make the ANOVA more robust to deviations from the equal σ rule.

When equal σ cannot be assumed, data may need to be "transformed" (for ex. log of data) to meet this assumption, or other tests may be needed.

Male lab rats were randomly assigned to 3 groups differing in access to food for 40 days. The rats were weighted (in grams) at the end.

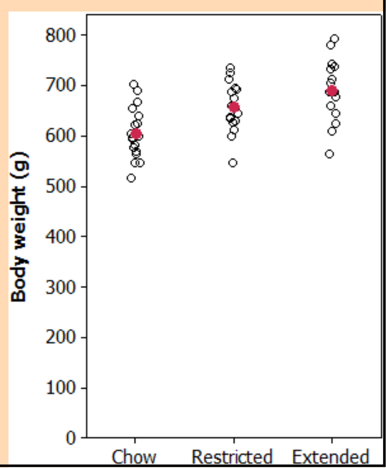
Conditions required for ANOVA:

•Independent random samples?

•Normal populations (Normal data) **<u>or</u>** large enough sample sizes?

•Equal population standard deviations (similar sample standard deviations)?

	N	Mean	StDev
Chow	19	605.63	49.64
Restricted	16	657.31	50.68
Extended	15	691.13	63.41



Meconium is a newborn's first stool, indicative of the fetus' exposure to outside contaminants during pregnancy. Do cigarette chemicals reach the fetus in the womb?
Cotinine is the metabolized form of nicotine. Its presence in the meconium was assessed for 3 independent random samples of newborns with different mother smoking status.

Are the conditions for ANOVA met?

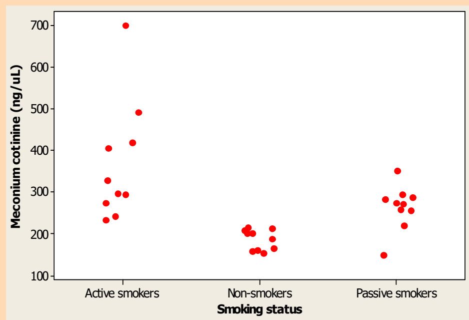
A) Yes B) No C) Not enough info

Level	N	Mean	StDev
Active smokers	10	367.20	143.67
Non-smokers	10	185.00	24.23
Passive smokers	10	263.40	52.54

•Independent random samples?

•Normal data **<u>or</u>** large enough sample sizes?

•Similar sample standard deviations?



Male lab rats were randomly assigned to 3 groups differing in access to food for 40 days. The rats were weighted (in grams) at the end.

Does the type of access to food influence the body weight of lab rats significantly?

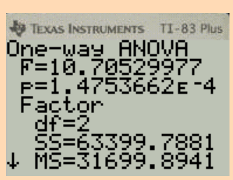
Chow	Restricted	Extended
516	546	564
547	599	611
546	612	625
564	627	644
577	629	660
570	638	679
582	645	687
594	635	688
597	660	706
599	676	714
606	695	738
606	687	733
624	693	744
623	713	780
641	726	794
655	736	
667		
690		
703		

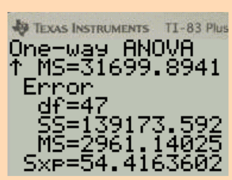
Minitab

```
One-way ANOVA: Chow, Restricted, Extended

Source  DF      SS     MS      F      P
Factor   2   63400  31700  10.71  0.000
Error   47  139174   2961
Total   49  202573

S = 54.42   R-Sq = 31.30%   R-Sq(adj) = 28.37%
```



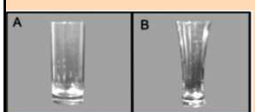


We found very strong evidence (*P* = 0.0001) that, <u>on average</u>, the body weights of male lab rats is influenced by the type of food available.

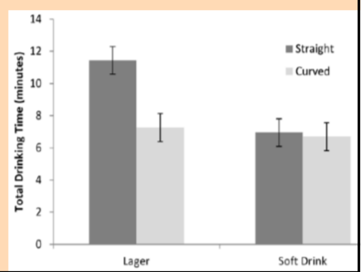
Healthy adult volunteers who drink socially but are not alcoholics were randomly assigned to a drink type: 12-oz beer or soft drink in a curved or straight glass.

Participants were told to consume the beverage at their own pace while watching a nature documentary. A wordsearch task was included to disguise the study aim.

Researchers videotaped the participants to determine the time it took each one to finish the drink. **Do the drink features significantly influence drinking time?**



Group	N (count)	Mean	Std. Error
1:	20	11.5	.85
2:	19	7.2	.89
3:	20	7	.86
4:	20	6.7	.88



A) There is very strong evidence ($P = 0.0004$) that healthy adults take longer to finish their drink when it contains beer in a straight glass.

B) There is very strong evidence ($P = 0.0004$) that, on average, healthy adults take longer to finish their drink when it contains beer in a straight glass.

C) There is very strong evidence ($P = 0.0004$) that, on average, the time it takes healthy adults to finish their drink depends on its features (glass-drink combination).

D) There is very strong evidence ($P = 0.0004$) that a healthy adult will take a different time to finish a glass of beer if the glass is curved instead of straight.

E) ANOVA is not appropriate here.

Group	N (count)	Mean	Std. Error
1:	20	11.5	.85
2:	19	7.2	.89
3:	20	7	.86
4:	20	6.7	.88

ANOVA Table...

Source of Variation	Sum of Squares	d.f.	Variance	F	p
Between Groups:	309.9797	3	103.3266	6.9144	0.0004
Within Groups:	1120.7682	75	14.9436		
Total:	1430.7479	78