1.- Pair the term in the first column with the most related term in the second column and explain the relationship of each pair of terms

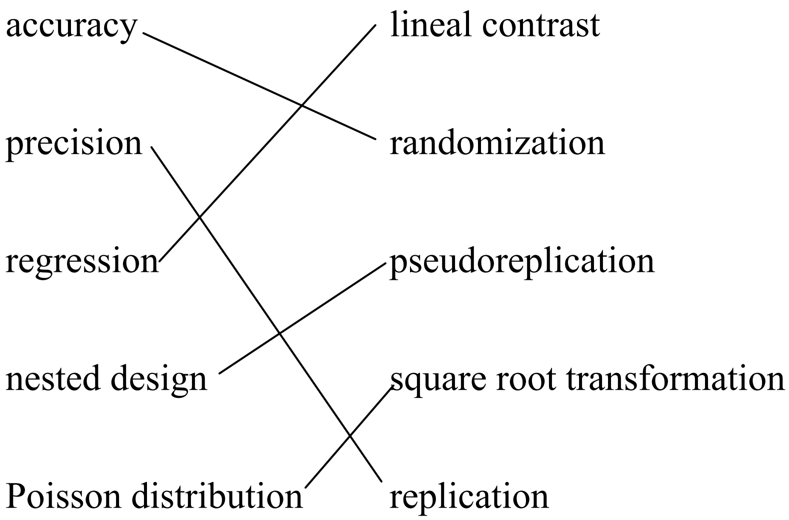
accuracy	lineal contrast
precision	randomization
regression	pseudoreplication
nested design	square root transformation
Poisson distribution	replication

2.- A population geneticist wants to test the significance of the differences in the length of the anthers among **3 races** of species "X" adapted to open grasslands (OG), forests (FO), and river sides (RS). He randomly selected 4 populations of each environment and then 6 individuals from each population, and finally randomly measured 2 anthers per individual. The objective of these additional measures is to get an idea of the variation among populations and among individual plants.

Source	DF	Sum of Squares	Mean Square	F	p<
Races	__	845.20	______		______
Populations	__	620.90	______		______
Individuals	__	1801.20	______		______
Error	__	1842.60	______		______
Corr. Total	__				

- Calculate the degrees of freedom, MS and F
- **Are there significant differences among races in the length of the anthers?**
- **Is there a significant variation among population?**
- **Is there a significant variation among individuals?**
- **Calculate the proportion of variation contributed by each level**

1.- Pair the term in the first column with the most related term in the second column and explain the relationship of each pair of terms



2.- Total 3*4*6*2=144

Source	DF	Sum of Squares	Mean Square	F	*P*
Races	2	845.20	422.6	422.6/70.1= 6.02	*P*<0.01
Populations	9	620.90	70.1	70.1/30.0= 2.34	*P*=0.03
Individuals	60	1801.20	30.0	30.0/25.6= 1.17	NS
Error	72	1842.60	25.6		
Corrected Total	143	143.00			

EXPECTED MS	
RACE	Var(Error) + 2 Var(Ind) + 12 Var(Pop) + 48 Var(Race)
POPULATION	Var(Error) + 2 Var(Ind) + 12 Var(Pop)
INDIVIDUAL	Var(Error) + 2 Var(Ind)
ERROR	Var(Error)

Variance Components		
Var(Error) =	25.6	66.7%
Var(Ind)= (30.0-25.6)/2=	2.2	5.7%
Var(Pop)= (70.1-30.0)/12=	3.3	8.6%
Var(Race)= (422.6-70.1)/48=	**7.3**	**19.0%**
Total	**38.4**	**100%**

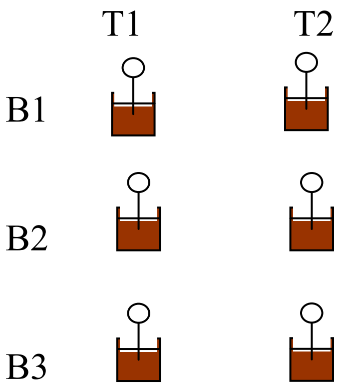
TWO NESTED FACTORS

```
data ex99_3;
input species $ flask sample mes protein @@;
cards;
A 1 1 1 3.7  A 1 1 2 3.9  A 1 2 1 3.9  A 1 2 2 3.7
A 2 1 1 3.3  A 2 1 2 3.1  A 2 2 1 3.5  A 2 2 2 3.7
A 3 1 1 3.4  A 3 1 2 3.2  A 3 2 1 4.0  A 3 2 2 4.3
A 4 1 1 3.6  A 4 1 2 3.5  A 4 2 1 3.8  A 4 2 2 3.5
B 1 1 1 4.1  B 1 1 2 4.4  B 1 2 1 4.0  B 1 2 2 4.2
B 2 1 1 3.6  B 2 1 2 3.7  B 2 2 1 3.9  B 2 2 2 3.4
B 3 1 1 3.8  B 3 1 2 3.9  B 3 2 1 3.7  B 3 2 2 3.1
B 4 1 1 3.5  B 4 1 2 3.2  B 4 2 1 3.9  B 4 2 2 3.8
C 1 1 1 4.1  C 1 1 2 4.5  C 1 2 1 4.5  C 1 2 2 4.9
C 2 1 1 3.9  C 2 1 2 3.6  C 2 2 1 4.6  C 2 2 2 4.8
C 3 1 1 4.2  C 3 1 2 4.1  C 3 2 1 4.8  C 3 2 2 4.5
C 4 1 1 4.7  C 4 1 2 4.3  C 4 2 1 4.0  C 4 2 2 4.2;

proc glm;
 class species flask sample;
 model protein= species flask(species) sample(flask species);
 random flask(species) sample(flask species);
 test h=species e=flask(species);
 test h=flask(species) e=sample(flask species);

proc varcomp Method= Type1;
  class species flask sample;
  model protein= species flask(species) sample(flask species);
run; quit;
```

Variance Source	DF	Sum of Squares	F Value	Pr > F	Error Term	Mean Sq	Variance Component	% of Total
Total	47	10.067				0.21	0.27	100.0
species	2	4.777	14.81	0.0014	flask	2.38	0.13	49.9
flask	9	1.450	0.68	0.7135	sample	0.16	-0.01	0.0
sample	12	2.840	5.68	0.0002	Error	0.23	0.09	35.0
Error	24	1.000				0.04	0.04	14.9

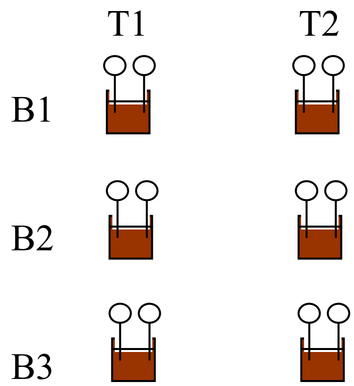
You get the same result using `sample(flask*species)`

RCBD 1 obs./ cell



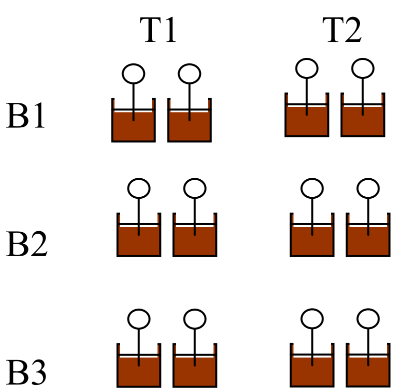
class B T;
model y= B T;
Tukey non-additivity test required

RCBD 1 obs./cell with subsamples



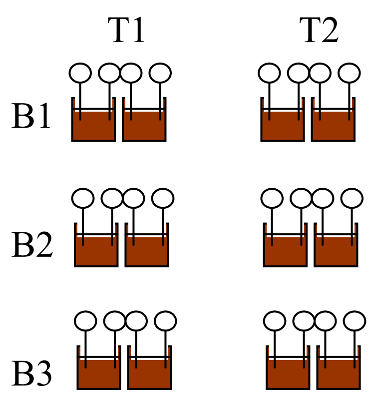
class B T **pot**;
model y= B T **pot(B*T)**;
random **pot(B*T)**;
test h= T e= **pot(B*T);**
Tukey non-additivity test required

RCBD >1 obs./ cell



class B T;
model y= B T B*T;
Tukey non-additivity test
not required

RCBD >1 obs./cell with subsamples



class B T pot;
model y=B T B*T **pot(B*T)**;
random **pot(B*T)**;
test h= T e= **pot(B*T);**
test h= B e= **pot(B*T)**;
test h= T*B e= **pot(B*T)**;
or
means T/Tukey e= **pot(B*T)**;
Tukey non-additivity test not required

RCBD calculation of Number of Blocks (STD pg 241)

Table 9.2 (STD P. 207). % oil in Flax seed inoculated with rust at different stages.

Treatment	Block 1	Block 2	Block 3	Block 4	Mean	$\sum_i (\bar{Y}_{i.} - \bar{Y}_{..})^2$
Seedling	34.4	35.9	36.0	34.1	35.1	0.174
Early bloom	33.3	31.9	34.9	37.1	34.3	1.480
Late Bloom	34.4	34.0	34.5	33.1	**34.0**	2.300
Full Bloom	36.8	36.6	37.0	36.4	36.7	1.400
Ripening	36.3	34.9	35.9	37.2	36.0	0.234
Control	36.4	37.3	37.7	36.7	**37.0**	2.200
***R^2*= 0.64**				**Mean**	35.517	**Sum=7.788**

Source of Variation	Df	MS	F
Blocks	r-1=3	1.06	
Treatments	t-1=5	6.35	4.8**
Error σ^2	(r-1)(t-1)=15	**1.326**	

To calculate the **exact power**: $\phi = q*\sqrt{n} / \sigma$ with $q^2 = \Sigma\tau_i^2/k$

$\phi = \sqrt{(7.788/6)} * \sqrt{4} / \sqrt{1.326}$ $\phi = 1.98$

Looking in the **Power chart** $df_1=5$ $df_2=15$ Power ≈ **0.90**

$$\phi = \sqrt{\frac{r * (\sum \tau_i^2 / k)}{MS_{error}}}$$

r= number of blocks
***k*= number of treatments**

Analyst n=18 blocks=3 Power for treatment = 0.68 (block 0.13)
n=24 blocks=4 Power for treatment = **0.90** (block 0.18)

Approximate formula for ϕ with only the extreme means (37.0 and 34.0, d=3.0)

$$\phi = \sqrt{\frac{r * d^2}{2k * MS_{error}}}$$

$\sum \tau^2 = (d/2)^2 + (d/2)^2 = d^2/2$

Then ϕ = SQRT {[r * (3^2)]/[2*6*1.326]}

for r = 4 ϕ = 1.5 $df_1=5$ $df_2= (6-1)*(4-1)=15$ Power ≈ 0.70

for **r= 6**, ϕ = 1.84 $df_1=5$ $df_2= (6-1)*(6-1)=25$ Power ≈ 0.90

Average of variances and MSE in an RCBD

Source of Variation	Df	MS	F
Blocks **Original**	r-1=3	1.06	
Treatments	t-1=5	6.35	4.8**
Error σ^2	(r-1)(t-1)=15	**1.326**	

Table 9.2 (STD P. 207). % oil in Flax seed inoculated with rust at different stages.
Add 5 to block I and subtract 3 from block IV

Treatment	Blk I+5	Blk II	Blk III	Blk IV-3	Mean	Var
Seedling	39.4	35.9	36	31.1	35.6	11.65
Early bloom	38.3	31.9	34.9	34.1	34.8	7.05
Late Bloom	39.4	34	34.5	30.1	34.5	14.54
Full Bloom	41.8	36.6	37	33.4	37.2	12.00
Ripening	41.3	34.9	35.9	34.2	36.575	10.41
Control	41.4	37.3	37.7	33.7	37.525	9.91
R^2= 0.64						**10.93**

Source of Variation	Df	SS	MS	F
Blocks I+5 IV-3	r-1=3	**176.78**	58.9	
Treatments	t-1=5	31.76	6.35	4.8**
Error σ^2	(r-1)(t-1)=15	**19.89**	**1.326**	

In an RCBD the MSE is not the average of the s^2 within each treatment!

However, if you merge back the MS Block into the error and transform this in a CRD then it is the same

(**19.89**+**176.78**= 196.67 Total error without the blocks 196.67/(3+15)=**10.93**

Weighted Average
TRT1= 3 (based on 10 measures)
TRT2= 4 (based on 5 measures)
TRT3= 5 (based on 2 measures)
Average of the three averages= (3+4+5)/3= **4**
Weighted Average of the three averages= (3*10+4*5+5*2)/17= (30+20+10)/17= **3.53**

Average of Square roots		SQRT
	9.0	3.0
	16.0	4.0
Average	**12.5**	3.5 and de-transformed=**12.25**

The de-transformed mean from a square root transformation is < than the mean of the original data.

A graduate student wanted to carry out an experiment to see the effects of **four different growing light conditions** (100, 600, 1100, 1600 µmol of quanta·m^{-2}·s^{-1}) on photosynthesis of soybean leaves. He had 16 plants and **four growth chambers** each of which had **four isolated compartments where light intensity can be regulated independently.** Growth chambers were from different years and have different levels of precision in temperature control. In his **1st experimental design**, he decided to assign one light level per each growth chamber: light levels of 100, 600, 1100, 1600 were assigned randomly to growth chambers 'A', 'B', 'C', 'D' respectively. Then he planned to assign randomly 1 plant to each of the 4 compartments in each chamber and to take photosynthesis measurements of 2 randomly selected leaves from each plant (8 measurements per treatment).

His professor suggested an alternative design (**2nd experimental design**): randomize the 4 light intensities within each chamber to the four isolated compartments and place one plant in each compartment. After one week he measured the photosynthetic rate of 2 leaves per plant and obtained the following information.

Growth chamber	**Leaf**	**Light level**			
		100	**600**	**1100**	**1600**
A	1	4.5	6.8	5.9	6.0
	2	3.8	5.8	8.6	7.5
B	1	3.7	7.0	6.2	7.9
	2	4.1	7.9	7.8	9.0
C	1	2.5	4.9	8.7	9.1
	2	5.1	5.9	10.2	7.1
D	1	2.5	4.8	4.9	5.4
	2	3.3	6.7	6.5	6.5

Which is the best design?

Indicate the number of replications in each design

Is the first design a CRD or an RCBD?

The mean and variance of linear functions

1) Add a constant C to each observation

$\sum (Y+C) = nC + \sum Y$

Example: Add 10 to each observation

(3, 5) Mean 4, $s^2 = 2$

(13, 15) Mean 14, $s^2 = 2$

Mean of the new variable = Previous mean + 10

σ^2 new variable = Previous σ^2

2) Multiply each observation by a constant

$\sum (Y*C) = C * \sum Y$

Example: Multiply each observation * 10

(3, 5) Mean 4, $s^2 = 2$

(30, 50) Mean 40, $s^2 = 200$

Mean of the new variable = Previous mean * 10

σ^2 of the new variable = Previous σ^2 * 100

3) Add two independent random variables

$Z = X_1 + X_2$

$X_1 \sim N(\mu_1, \sigma^2_1)$ and $X_2 \sim N(\mu_2, \sigma^2_2) \Rightarrow X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma^2_1 + \sigma^2_2)$

Mean of Z = Mean of X_1 + mean of X_2

Variance of Z = $\sigma^2_1 + \sigma^2_2$

4) Subtract two independent random variables

$Z = X_1 - X_2$

$X_1 \sim N(\mu_1, \sigma^2_1)$ and $X_2 \sim N(\mu_2, \sigma^2_2) \Rightarrow X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma^2_1 + \sigma^2_2)$

Mean of Z = Mean of X_1 - mean of X_2

Variance of Z = $\sigma^2_1 + \sigma^2_2$

Assume we have 32 animals, 4 different ranches. We randomly assign 8 animals to the four treatments in each ranch:

RCBD with more than one replication per cell

```
data lambs;
input sex_est $ block gain @@;
cards;
f0 1 46    f0 2 51    f0 3 61   f0 4 50
m0 1 49    m0 2 53    m0 3 66   m0 4 56
f3 1 56    f3 2 52    f3 3 68   f3 4 56
m3 1 53    m3 2 64    m3 3 73   m3 4 58

f0 1 48    f0 2 53    f0 3 62   f0 4 52
m0 1 51    m0 2 55    m0 3 68   m0 4 58
f3 1 58    f3 2 54    f3 3 70   f3 4 58
m3 1 55    m3 2 66    m3 3 75   m3 4 60

proc glm data=lambs order=data;
*Without order=data, SAS reads alphabetically f0 f3 m0 m3;
  class block sex_est;
  model gain= block sex_est block*sex_est;
  output out=check  p= pred  r= resi;

  contrast 'sex'         sex_est 1 -1  1 -1;
  contrast 'estrogen'    sex_est 1  1 -1 -1;
  contrast 'interaction' sex_est 1 -1 -1  1;
  means sex_est/tukey;

proc univariate data= check normal;
  var resi;

proc glm data=lambs;
  class sex_est;
  model gain= sex_est;
  means sex_est/ HOVTEST = LEVENE;

run; quit;
```

Two subsamples per animal, 16 animals

```
data lambs;
input sex_est $ block animal gain @@;
cards;
f0 1 1 46    f0 2 1 51    f0 3 1 61   f0 4 1 50
m0 1 2 49    m0 2 2 53    m0 3 2 66   m0 4 2 56
f3 1 3 56    f3 2 3 52    f3 3 3 68   f3 4 3 56
m3 1 4 53    m3 2 4 64    m3 3 4 73   m3 4 4 58

f0 1 1 48    f0 2 1 53    f0 3 1 62   f0 4 1 52
m0 1 2 51    m0 2 2 55    m0 3 2 68   m0 4 2 58
f3 1 3 58    f3 2 3 54    f3 3 3 70   f3 4 3 58
m3 1 4 55    m3 2 4 66    m3 3 4 75   m3 4 4 60

proc glm data=lambs order=data;
*Without order=data, SAS reads alphabetically f0 f3 m0 m3;
*With order=data, SAS reads f0 m0 f3 m3: Contrast are different!;
  class block sex_est animal;
  model gain= block sex_est animal(block*sex_est);
  random animal(block*sex_est);
  test h=sex_est e=animal(block*sex_est);
*The same results would have been obtained with sex_est(block);

  contrast 'sex'         sex_est 1 -1  1 -1 / e=animal(block*sex_est);
  contrast 'estrogen'    sex_est 1  1 -1 -1 / e=animal(block*sex_est);
  contrast 'interaction' sex_est 1 -1 -1  1 / e=animal(block*sex_est);
  means sex_est/tukey e=animal(block*sex_est);

*Next is Tukey wih incorrect error for comparison;
  means sex_est/tukey;

proc varcomp Method= Type1;
  class block sex_est animal;
  model gain= block sex_est animal(block*sex_est);

run;
quit;
```

```
data lambs;
   input sex_est $ @;
   do block = 1 to 4;
     input gain @;
     output;
     end;
cards;
f0  47  52  62  51
m0  50  54  67  57
f3  57  53  69  57
m3  54  65  74  59
proc glm;
  proc glm;
  class sex_est;
  model gain= sex_est; run;
```

WITHOUT BLOCK

Dependent Variable: gain

Source	DF	SS	MS	F Value	Pr > F
Model	3	208	69.3	1.29	0.32 NS
Error	12	646	53.8		
Corrected Total	15	854			

R-Square	Coeff Var	Root MSE	gain Mean
0.2436	**12.65**	**7.337**	**58**
208/854	Root MSE/58	=”average” s	

R-Square **Proportion of variation explained by the model**
= SST / Total SS

MSE= is the average of the s^2 from the four treatment (in a CRD but not in an RCBD).

Root MSE =is the square root of the MSE of the ANOVA table. (≈ average s compared to an average s^2

Coeff Var **=** Root MSE/ General mean (variation per unit of mean)

Source	DF	SS	MS	F Value	Pr > F
sex_est	3	208	69.3	1.29	0.32

WITH BLOCK

```
  class block sex_est;
  model gain=block sex_est;
```

Dependent Variable: gain

Source	DF	SS	MS	F Value	Pr > F
Model	6	784	30.7	16.80	0.0002
Error	9	70	7.8		
Corrected Total	15	854			

R-Square	Coeff Var	Root MSE	gain Mean
0.918	**4.808**	**2.788**	**58**
784/854	2.788/58*100		

Source	DF	SS	MS	F Value	Pr > F
block	3	576	192.0	24.69	0.0001
sex_est	3	208	69.3	8.91	0.0046