Overview of Topic X

This topic will cover

- Random vs. Fixed Effects (§25)
- Using E(MS) to Obtain Appropriate Tests in a Random or Mixed Effects Model (§25)

Chapter 25: One-way Random Effects Design

Fixed Effects vs Random Effects

- Up to this point we have been considering "fixed effects models", in which the levels of each factor were fixed in advance of the experiment and we were interested in differences in response among those specific levels.
- Now we will consider "random effects models", in which the factor levels are meant to be representative of a general population of possible levels. We are interested in whether that factor has a significant effect in explaining the response, but only in a general way. For example, we're not interested in a detailed comparison of level 2 vs. level 3, say.

- When we have both fixed and random effects, we call it a "mixed effects model". The main SAS procedure we will use is called "`proc mixed`" which allows for fixed and random effects, but we can also use `glm` with a `random` statement. We'll start first with a single random effect.

- In some situations it is clear from the experiment whether an effect is fixed or random. However there are also situations in which calling an effect fixed or random depends on your point of view, and on your interpretation and understanding. So sometimes it is a personal choice. This should become more clear with some examples.

Data for one-way design

- Y, the response variable
- Factor with levels $i = 1$ to r
- $Y_{i,j}$ is the jth observation in cell i, $j = 1$ to n_i
- A balanced design has $n = n_i$

KNNL Example

- KNNL page 1036 (`nknw964.sas`)
- Y is the rating of a job applicant
- Factor A represents five different personnel interviewers (officers), $r = 5$ levels
- $n = 4$ *different* applicants were randomly chosen and interviewed by each interviewer (i.e. 20 applicants) (applicant is *not* a factor since no applicant was interviewed more than once)
- The interviewers were <u>selected at random</u> from the pool of interviewers and the applicants were randomly assigned to interviewers.

- Here we are not so interested in the differences between the five interviewers that happened to be picked (i.e. does Joe give higher ratings than Fred, is there a difference between Ethel and Bob). Rather we are interested in quantifying and accounting for the effect of "interviewer" in general. There are other interviewers in the "population" (at the company) and we want to make inference about them too.

- Another way to say this is that with fixed effects we were primarily interested in the *means* of the factor levels (and the differences between them). With random effects, we are primarily interested in their *variances*.

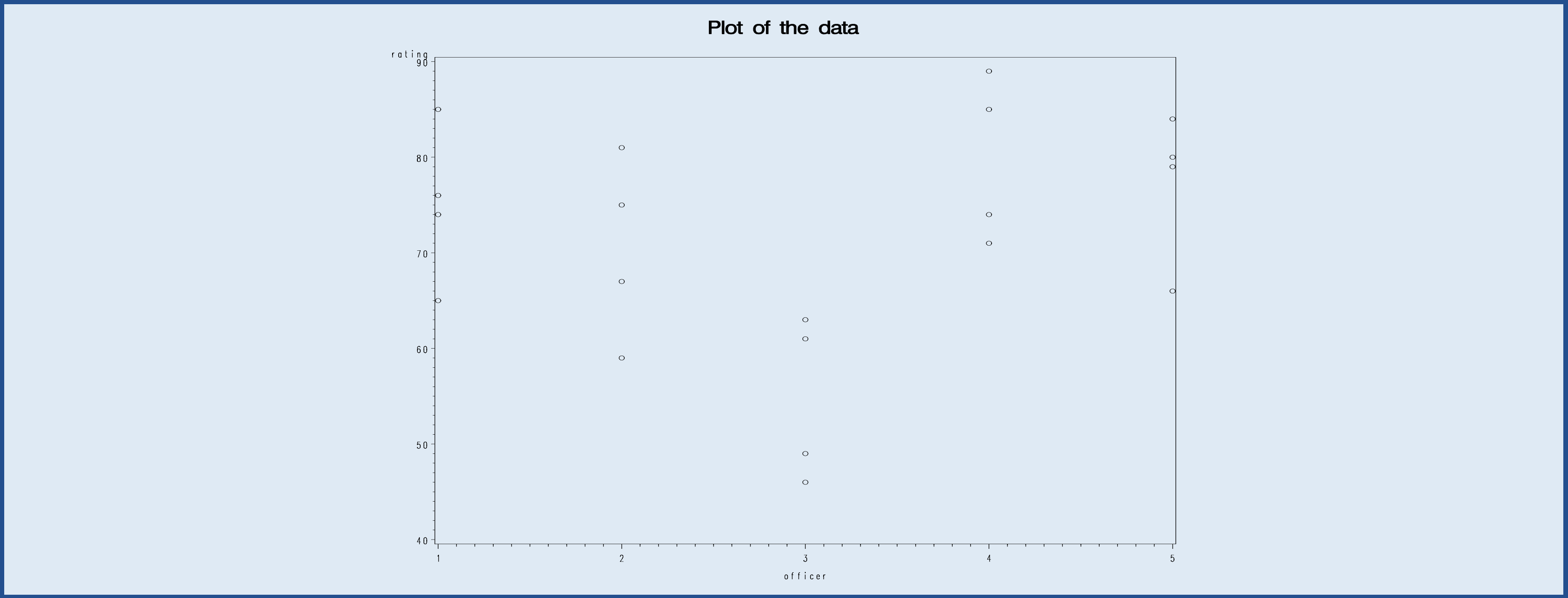
Read and check the data

```
data interview;
   infile 'h:\System\Desktop\CH24TA01.DAT';
   input rating officer;
proc print data=interview;
```

```
Obs    rating    officer
  1      76         1
  2      65         1
  3      85         1
  4      74         1
  5      59         2
  6      75         2
  7      81         2
  8      67         2
  9      49         3
 10      63         3
 11      61         3
 12      46         3
 13      74         4
 14      71         4
 15      85         4
 16      89         4
 17      66         5
 18      84         5
 19      80         5
 20      79         5
```

Plot the data

```
title1 'Plot of the data';
symbol1 v=circle i=none c=black;
proc gplot data=interview;
   plot rating*officer;
run;
```



Random effects model (cell means)

This model is also called

- ANOVA Model II
- A variance components model

$$Y_{i,j} = \mu_i + \epsilon_{i,j}$$

- The μ_i are iid $N(\mu, \sigma_A^2)$. NOTE!!!!! THIS IS DIFFERENT!!!!
- The $\epsilon_{i,j}$ are iid $N(0, \sigma^2)$
- μ_i and $\epsilon_{i,j}$ are independent
- $Y \sim N(\mu, \sigma_A^2 + \sigma^2)$

Now the μ_i are random variables with a common mean. The question of "are they all the same" can now be addressed by considering whether the *variance* of their distribution, σ_A^2, is zero. Of course, the estimated means will likely be different from each other; the question is whether the difference can be explained by error (σ^2) alone.

The text uses the symbol σ_μ^2 instead of σ_A^2; they are the same thing. I prefer the latter notation because it generalizes more easily to more than one factor, and also to the factor effects model.

Two Sources of Variation

Observations with the same i (e.g. the same interviewer) are dependent, and their covariance is σ_A^2. The components of variance are σ_A^2 and σ^2. We want to get an idea of the relative magnitudes of these variance components.

Random factor effects model

Same basic idea as before... $\mu_i = \mu + \alpha_i$. The model is $Y_{i,j} = \mu + \alpha_i + \epsilon_{i,j}$.

$$\begin{aligned} \alpha &\sim N(0, \sigma_A^2) \\ \epsilon_{i,j} &\sim N(0, \sigma^2) \\ Y_{i,j} &\sim N(\mu, \sigma_A^2 + \sigma^2) \end{aligned}$$

The book uses σ_α^2 instead of σ_A^2 here. Despite the different notations, σ_α^2 and σ_μ^2 are really the same thing, because μ_i and α_i differ only by an additive constant (μ), so they have the same variance. That is why in these notes I'm using the same symbol σ_A^2 to refer to both. (With two factors we will have to distinguish between these.)

Parameters

There are two important parameters in these models: σ_A^2 and σ^2. (also μ in the F.E.M.)

The cell means $\mu_{i,j}$ are random variables, not parameters.

We are sometimes interested in estimating $\frac{\sigma_A^2}{\sigma_A^2+\sigma^2} = \frac{\sigma_A^2}{\sigma_Y^2}$.

In some applications it is called the *intraclass correlation coefficient*. It is the correlation between two observations with the same i.

ANOVA Table

- The terms and layout of the ANOVA table are the same as what we used for the fixed effects model
- The expected mean squares (EMS) are different because of the additional random effects, so we will estimate parameters in a new way.
- Hypotheses being tested are also different.

EMS and parameter estimates

$E(MSE) = \sigma^2$ as usual. We use MSE to estimate σ^2.

$E(MSA) = \sigma^2 + n\sigma_A^2$. Note that this is different from before. From this you can see that we should use $\frac{(MSA - MSE)}{n}$ to estimate σ_A^2.

Hypotheses

$$\begin{aligned} \mathrm{H}_0 &: \quad \sigma_A^2 = 0 \\ \mathrm{H}_1 &: \quad \sigma_A^2 \neq 0 \end{aligned}$$

The test statistic is $F = MSA/MSE$ with $r - 1$ and $r(n-1)$ degrees of freedom (since this ratio is 1 when the null hypothesis is true); reject when F is large, and report the p-value. Note that in the one factor analysis, the test is the same it was before. This WILL NOT be the case as we add more factors.

SAS Coding and Output

run `proc glm` with a `random` statement

```
proc glm data=interview;
   class officer;
   model rating=officer;
   random officer;
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	1579.700000	394.925000	5.39	0.0068
Error	15	1099.250000	73.283333		
Corrected Total	19	2678.950000			

`Random` statement output

```
Source                  Type III Expected Mean Square
officer                 Var(Error) + 4 Var(officer)
```

This is SAS's way of saying $E(MSA) = \sigma^2 + 4\sigma_A^2$ (note $n = 4$ replicates).

proc varcomp

This procedure gets the “variance components”.

```
proc varcomp data=interview;
   class officer;
   model rating=officer;
```

```
      MIVQUE(0) Estimates
Variance Component          rating
Var(officer)              80.41042
Var(Error)                73.28333
```

(Other methods are available for estimation; `mivque` is the default.)

SAS is now saying

$$\begin{aligned} \text{Var}(Error) &= \hat{\sigma}^2 = 73.28333 \text{ (notice this is just } MSE) \\ \text{Var}(officer) &= \hat{\sigma}^2_\mu = 80.41042 = \frac{(394.925 - 73.283)}{4} \\ &= \frac{(MSA - MSE)}{n}. \end{aligned}$$

As an alternative to using `proc glm` with a `random` statement, and `proc varcomp`, you could instead use `proc mixed`, which has some options specifically for mixed models.

proc mixed

```
proc mixed data=interview cl;
   class officer;
   model rating=;
   random officer/vcorr;
```

- The `cl` option after `data=interview` asks for the confidence limits.
- The `class` statement lists all the categorical variables just as in `glm`.

- The `model rating=;` line looks strange. In `proc mixed`, the `model` statement lists *only the fixed effects*. Then the random effects are listed separately in the `random` statement. In our example, there were no fixed effects, so we had no predictors on the model line. We had one random effect, so it went on the random line.

- This is different from `glm`, where all the factors (fixed and random) are listed on the model line, and then the random ones are repeated in the `random` statement.

- Just in case you're not confused enough, `proc varcomp` assumes all factors are random effects unless they are specified as fixed...

`Proc mixed` gives a huge amount of output. Here are some pieces of it.

```
.      Covariance Parameter Estimates
Cov Parm   Estimate   Alpha     Lower     Upper
officer     80.4104    0.05   24.4572   1498.97
Residual    73.2833    0.05   39.9896    175.54
```

The estimated intraclass correlation coefficient is $\frac{\hat{\sigma}_A^2}{\hat{\sigma}_A^2+\hat{\sigma}^2} = \frac{\hat{\sigma}_A^2}{\hat{\sigma}_Y^2} = \frac{80.4104}{80.4104+73.2833} = 0.5232.$

About half the variance in rating is explained by interviewer.

Output from **`vcorr`** option

This gives the intraclass correlation coefficient.

Row	Col1	Col2	Col3	Col4
1	1.0000	0.5232	0.5232	0.5232
2	0.5232	1.0000	0.5232	0.5232
3	0.5232	0.5232	1.0000	0.5232
4	0.5232	0.5232	0.5232	1.0000

Confidence Intervals

- For μ the estimate is $\bar{Y}_{..}$, and the variance of this estimate under the random effects model becomes $\sigma^2\{\bar{Y}_{..}\} = \frac{(n\sigma_A^2+\sigma^2)}{rn}$ which may be estimated by $s^2\{\bar{Y}_{..}\} = \frac{(MSA)}{rn}$. See page 1038 for derivation if you like. To get a CI we use a t critical value with $r-1$ degrees of freedom.
- Notice that the variance here involves a combination of the two errors and we end up using MSA instead of MSE in the estimate (we used MSE in the fixed effects case).
- We may also get point estimates and CI's for σ^2, σ_A^2, and the intraclass correlation $\sigma_A^2/(\sigma_A^2+\sigma^2)$. See pages 1040-1047 for details. All of these are available in `proc mixed`.

Applications

- In the KNNL example we would like $\sigma_\mu^2/(\sigma_\mu^2 + \sigma^2)$ to be small, indicating that the variance due to interviewer is small relative to the variance due to applicants.
- In many other examples we would like this quantity to be large. One example would be measurement error - if we measure r items n times each, σ^2 would represent the error inherent to the instrument of measurement.

Two-way Random Effects Model

Data for two-way design

- Y, the response variable
- Factor A with levels $i = 1$ to a
- Factor B with levels $j = 1$ to b
- $Y_{i,j,k}$ is the kth observation in cell (i, j) $k = 1$ to $n_{i,j}$
- For balanced designs, $n = n_{i,j}$

KNNL Example

- KNNL Problem 25.15, page 1080 (`nknw976.sas`)
- Y is fuel efficiency in miles per gallon
- Factor A represents four different drivers, $a = 4$ levels
- Factor B represents five different cars of the same model, $b = 5$
- Each driver drove each car twice over the same 40-mile test course ($n = 2$)

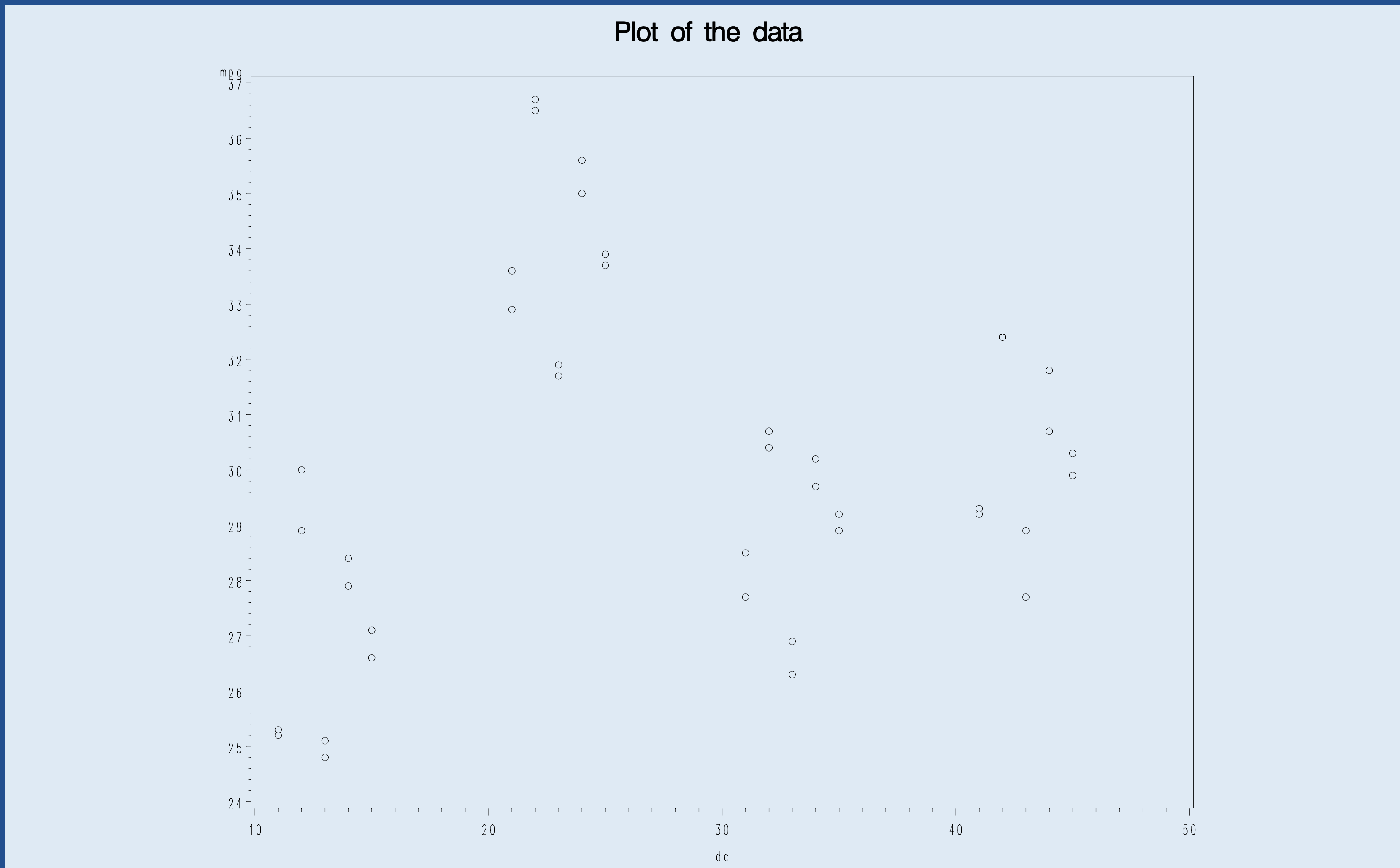
Read and check the data

```
data efficiency;
   infile 'h:\System\Desktop\CH24PR15.DAT';
   input mpg driver car;
proc print data=efficiency;
```

```
Obs     mpg    driver    car
  1    25.3       1       1
  2    25.2       1       1
  3    28.9       1       2
  4    30.0       1       2
  5    24.8       1       3
  6    25.1       1       3
  7    28.4       1       4
  8    27.9       1       4
...
```

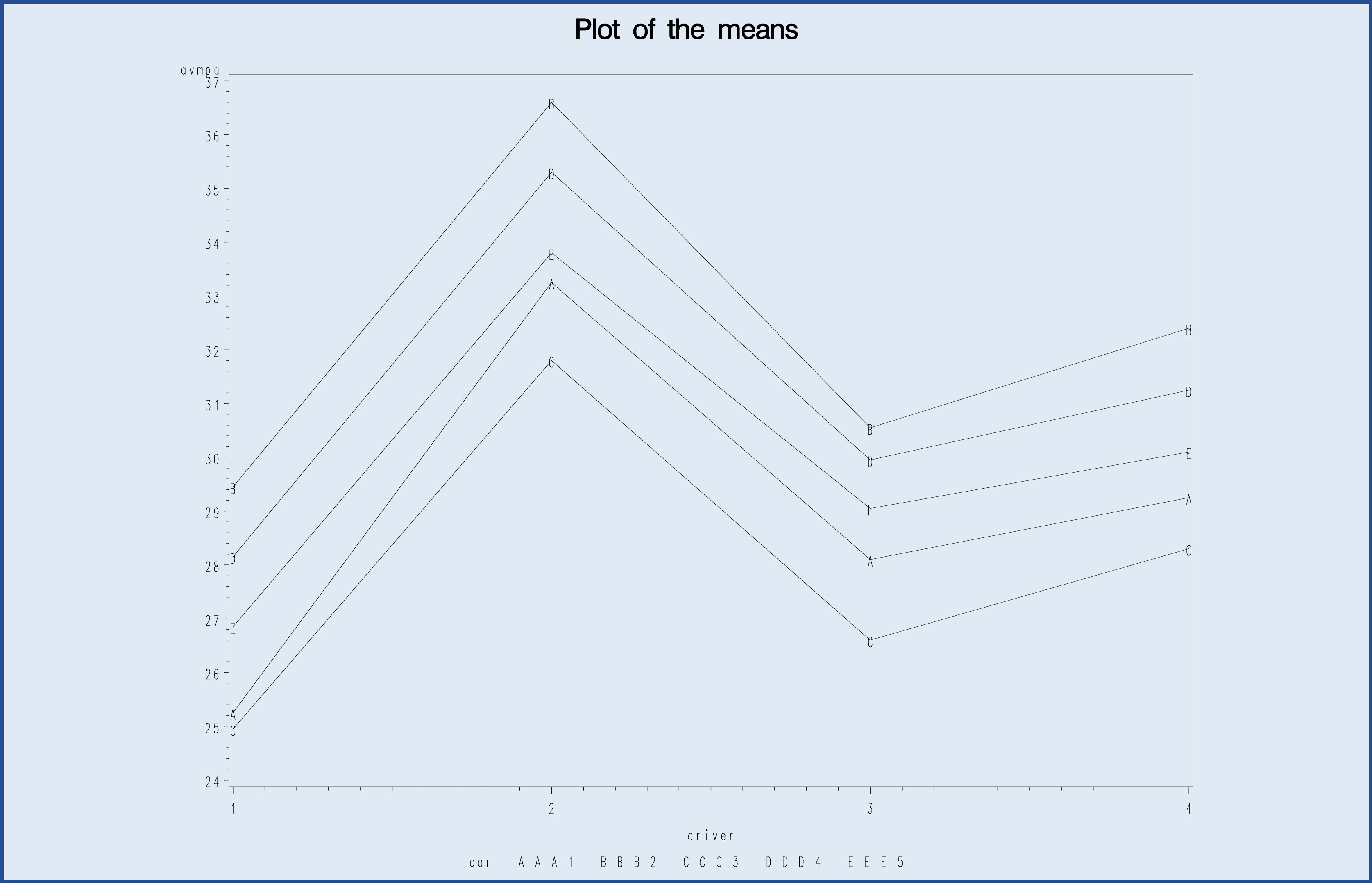
Prepare the data for a plot, and plot the data

```
data efficiency;
   set efficiency;
   dc = driver*10 + car;
title1 'Plot of the data';
symbol1 v=circle i=none c=black;
proc gplot data=efficiency;
   plot mpg*dc;
```


Plot of the data
mpg
37
36
35
34
33
32
31
30
29
28
27
26
25
24
10
20
30
40
50
dc

Find and plot the means

```
proc means data=efficiency;
   output out=effout mean=avmpg;
   var mpg;
   by driver car;
title1 'Plot of the means';
symbol1 v='A' i=join c=black;
symbol2 v='B' i=join c=black;
symbol3 v='C' i=join c=black;
symbol4 v='D' i=join c=black;
symbol5 v='E' i=join c=black;
proc gplot data=effout;
   plot avmpg*driver=car;
```


Plot of the means
avmpg
37
36
35
34
33
32
31
30
29
28
27
26
25
24
1
2
3
4
driver
car A A A 1 B B B 2 C C C 3 D D D 4 E E E 5

Random Effects Model

Random cell means model

$$Y_{i,j,k} = \mu_{i,j} + \epsilon_{i,j,k}$$

- $\mu_{i,j} \sim N(\mu, \sigma_\mu^2)$. NOTE!!!!! THIS IS DIFFERENT!!!
- $\epsilon_{i,j,k} \sim^{iid} N(0, \sigma^2)$ as usual
- $\mu_{i,j}$, $\epsilon_{i,j,k}$ are independent
- The above imply that $Y_{i,j,k} \sim N(\mu, \sigma_\mu^2 + \sigma^2)$

Dependence among the $Y_{i,j,k}$ can be most easily described by specifying the covariance matrix of the vector $(Y_{i,j,k})$

Random factor effects model

$Y_{i,j,k} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{i,j} + \epsilon_{i,j,k}$, where

$$\begin{aligned} \alpha_i &\sim N(0, \sigma_A^2) \\ \beta_j &\sim N(0, \sigma_B^2) \\ (\alpha\beta)_{i,j} &\sim N(0, \sigma_{AB}^2) \\ \sigma_Y^2 &= \sigma_A^2 + \sigma_B^2 + \sigma_{AB}^2 + \sigma^2 \end{aligned}$$

Now the component σ_μ^2 from the cell means model can be divided up into three components - A, B, and AB. That is, $\sigma_\mu^2 = \sigma_A^2 + \sigma_B^2 + \sigma_{AB}^2$

Parameters

- There are five parameters in this model: μ, σ_A^2, σ_B^2, σ_{AB}^2, σ^2
- The cell means are random variables, not parameters!!!

ANOVA Table

- The terms and layout of the ANOVA table are the same as what we used for the fixed effects model
- However, the expected mean squares (EMS) are different.

EMS and parameter estimates

$$\begin{aligned}
\mathsf{E}(MSA) &= \sigma^2 + bn\sigma_A^2 + n\sigma_{AB}^2 \\
\mathsf{E}(MSB) &= \sigma^2 + an\sigma_B^2 + n\sigma_{AB}^2 \\
\mathsf{E}(MSAB) &= \sigma^2 + n\sigma_{AB}^2 \\
\mathsf{E}(MSE) &= \sigma^2
\end{aligned}$$

Estimates of the variance components can be obtained from these equations or other methods.

- Note the patterns in the EMS: (these hold for balanced data).

- They all contain σ^2. For MSA, it also contains all the σ^2's that have an A in the subscript (σ^2_A and σ^2_{AB}); similarly for the other MS terms.

- The coefficient of each term (except the first) is the product of n and all letters *not* represented in the subscript. It is also the total number of observations at each fixed level of the level corresponding to the subscript (e.g. there are nb observations for each level of A)

Hypotheses

$$\mathsf{H}_{0A} : \sigma_A^2 = 0; \quad \mathsf{H}_{1A} : \sigma_A^2 \neq 0$$
$$\mathsf{H}_{0B} : \sigma_B^2 = 0; \quad \mathsf{H}_{1B} : \sigma_B^2 \neq 0$$
$$\mathsf{H}_{0AB} : \sigma_{AB}^2 = 0; \quad \mathsf{H}_{1AB} : \sigma_{AB}^2 \neq 0$$

Hypothesis H_{0A}

- $H_{0A} : \sigma_A^2 = 0$; $H_{1A} : \sigma_A^2 \neq 0$
- $E(MSA) = \sigma^2 + bn\sigma_A^2 + n\sigma_{AB}^2$
- $E(MSAB) = \sigma^2 + n\sigma_{AB}^2$
- $E(MSE) = \sigma^2$
- Need to look for the ratio that will be 1 when H_0 is true and bigger than 1 when it is false. So this hypothesis will be tested by $F = \frac{MSA}{MSAB}$ (not the usual fixed effects test statistic). The degrees of freedom for the test will be the degrees of freedom associated to those mean squares: $a - 1, (a - 1)(b - 1)$.
- Notice you can no longer assume that the denominator is MSE!!!!! (Note that the test using MSE is done by SAS, but it is not particularly meaningful (it sort of tests both main and interaction at once).)

Hypothesis H_{0B}

- $H_{0B}: \sigma_B^2 = 0$; $H_{1B}: \sigma_B^2 \neq 0$
- $E(MSB) = \sigma^2 + an\sigma_B^2 + n\sigma_{AB}^2$
- $E(MSAB) = \sigma^2 + n\sigma_{AB}^2$
- $E(MSE) = \sigma^2$
- So H_{0B} is tested by $F = \frac{MSB}{MSAB}$ with degrees of freedom $b-1, (a-1)(b-1)$.

Hypothesis H_{0AB}

- $\text{H}_{0AB} : \sigma^2_{AB} = 0$; $\text{H}_{1AB} : \sigma^2_{AB} \neq 0$
- $\text{E}(MSAB) = \sigma^2 + n\sigma^2_{AB}$
- $\text{E}(MSE) = \sigma^2$
- So H_{0AB} is tested by $F = \frac{MSAB}{MSE}$ with degrees of freedom $(a-1)(b-1), ab(n-1)$.

Run **proc glm**

```
proc glm data=efficiency;
   class driver car;
   model mpg=driver car driver*car;
   random driver car driver*car/test;
```

Regular ANOVA Tables

- Model and error output

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	19	377.4447500	19.8655132	113.03	<.0001
Error	20	3.5150000	0.1757500		
Corrected Total	39	380.9597500			

- Factor effects output

Source	DF	Type I SS	Mean Square	F Value	Pr > F
driver	3	280.2847500	93.4282500	531.60	<.0001
car	4	94.7135000	23.6783750	134.73	<.0001
driver*car	12	2.4465000	0.2038750	1.16	0.3715

- Type III SS Table is identical to Type I SS Table.

Only the interaction test is valid here: the test for interaction is $MSAB/MSE$, but the tests for main effects should be $MSA/MSAB$ and $MSB/MSAB$ which are done with the `test` statement, not $/MSE$ as is done here. (However, if you do this the main effects are significant as shown below.)

Lesson: just because SAS spits out a p-value, doesn't mean it is for a meaningful test!

Random statement output

```
Source      Type III Expected Mean Square
driver      Var(Error) + 2 Var(driver*car) + 10 Var(driver)
car         Var(Error) + 2 Var(driver*car) + 8 Var(car)
driver*car  Var(Error) + 2 Var(driver*car)
```

Random/test output

```
.                     The GLM Procedure
Tests of Hypotheses for Random Model Analysis of Variance
Dependent Variable: mpg
Source   DF  Type III SS  Mean Square  F Value  Pr > F
driver    3   280.284750    93.428250   458.26  <.0001
car       4    94.713500    23.678375   116.14  <.0001
Error    12     2.446500     0.203875
Error: MS(driver*car)
```

This last line says the denominator of the F-tests is the $MSAB$.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
driver*car	12	2.446500	0.203875	1.16	0.3715
Error: MS(Error)	20	3.515000	0.175750		

For the interaction term, this is the same test as was done above.

proc varcomp

```
proc varcomp data=efficiency;
   class driver car;
   model mpg=driver car driver*car;
```

```
      MIVQUE(0) Estimates
Variance Component            mpg
Var(driver)               9.32244
Var(car)                  2.93431
Var(driver*car)           0.01406
Var(Error)                0.17575
```

Mixed Models

Two-way mixed model

Two way mixed model has

- One fixed main effect
- One random main effect
- The interaction is considered a random effect

Tests (Restricted Mixed Models)

- Fixed main effect is tested by interaction in the denominator
- Random main effect is tested by error
- Interaction is tested by error
- Notice that these are *backwards* from what you might intuitively extrapolate from the two-way random effects and two-way fixed effects model

See Table 25.5 (page 1052) and below for the EMS that justify these statements. Also see Table 25.6 for the tests (page 1053).

Notation for two-way mixed model

Y, the response variable

A, the fixed effect (a levels)

B, the random effect (b levels)

We'll stick to balanced designs ($n_{i,j} = n$)

Factor effects parameterization

$$Y_{i,j,k} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{i,j} + \epsilon_{i,j,k}$$

Where

- μ is the overall mean,
- α_i are fixed (but unknown) fixed main effects with $\sum_i \alpha_i = 0$,
- β_j are $N(0, \sigma_B^2)$ independent random main effects,
- $(\alpha\beta)_{i,j}$ are random interaction effects.

- Randomness is "catching" so the interaction between a fixed and a random effect is considered random and has a distribution.
- **Restricted Mixed Model**
  - **The interactions are also subject to constraints kind of like fixed effects;**
  - $(\alpha\beta)_{i,j} \sim N\left(0, \frac{a-1}{a}\sigma^2_{AB}\right)$ **subject to the constraint** $\sum_i (\alpha\beta)_{i,j} = 0$ **for each** j**;**
  - **Because of the constraints,** $(\alpha\beta)_{i,j}$ **having the same** j **(but different** i**) are negatively correlated, with covariance** **Cov**$((\alpha\beta)_{i,j}, (\alpha\beta)_{i',j}) = -\frac{\sigma^2_{AB}}{a}$**;**
- **Unrestricted Mixed Model (SAS)**
  - **No constraints on the interaction effects;**
  - $(\alpha\beta)_{i,j} \overset{iid}{\sim} N\left(0, \sigma^2_{AB}\right)$.
- **The two models have different Expected Mean Squares (EMS), so suggest different test statistics for hypothesis tests!**

Expected Mean Squares (Restricted Mixed Model)

$$\begin{aligned}
\mathsf{E}(MSA) &= \sigma^2 + \frac{nb}{a-1}\sum_i \alpha_i^2 + n\sigma_{\alpha\beta}^2 \\
\mathsf{E}(MSB) &= \sigma^2 + na\sigma_\beta^2 \\
\mathsf{E}(MSAB) &= \sigma^2 + n\sigma_{\alpha\beta}^2 \\
\mathsf{E}(MSE) &= \sigma^2
\end{aligned}$$

Different denominators will be needed to test for the various effects.

$$\begin{aligned}
\mathsf{H}_{0A}: &\quad \text{all } \alpha_i = 0 \text{ is tested by } F = \tfrac{MSA}{MSAB} \\
\mathsf{H}_{0B}: &\quad \sigma_B^2 = 0 \text{ is tested by } F = \tfrac{MSB}{MSE} \\
\mathsf{H}_{0AB}: &\quad \sigma_{AB}^2 = 0 \text{ is tested by } F = \tfrac{MSAB}{MSE}.
\end{aligned}$$

So, though it seems counterintuitive at first, the fixed effect is tested by the interaction, and the random effect is tested by the error.

Expected Mean Squares (Unrestricted Mixed Model)

$$
\begin{aligned}
\mathsf{E}(MSA) &= \sigma^2 + \frac{nb}{a-1}\sum_i \alpha_i^2 + n\sigma_{\alpha\beta}^2 \\
\mathsf{E}(MSB) &= \sigma^2 + na\sigma_\beta^2 + n\sigma_{\alpha\beta}^2 \\
\mathsf{E}(MSAB) &= \sigma^2 + n\sigma_{\alpha\beta}^2 \\
\mathsf{E}(MSE) &= \sigma^2
\end{aligned}
$$

Different denominators will be needed to test for the various effects.

$$
\begin{aligned}
\text{H}_{0A}: &\quad \text{all } \alpha_i = 0 \text{ is tested by } F = \tfrac{MSA}{MSAB} \\
\text{H}_{0B}: &\quad \sigma_B^2 = 0 \text{ is tested by } F = \tfrac{MSB}{MSAB} \\
\text{H}_{0AB}: &\quad \sigma_{AB}^2 = 0 \text{ is tested by } F = \tfrac{MSAB}{MSE}.
\end{aligned}
$$

Both fixed and random effects are tested by the interaction.

SAS (`proc glm`) assumes unrestricted mixed models and writes EMS out for you but it uses the notation $Q(A)$ to denote the fixed quantity $\frac{nb}{a-1}\sum_i \alpha_i^2$. It uses the names Var$(Error) = \sigma^2$, Var$(B) = \sigma_B^2$, and Var$(A \times B) = \sigma_{AB}^2$. (It doesn't actually use the names A and B; it uses the variable names.)

Example: KNNL Problem 25.16

(`nknw1005.sas`)

Y - service time for disk drives

A - make of drive (fixed, with $a = 3$ levels)

B - technician performing service (random, with $b = 3$ levels)

The three technicians for whom we have data are selected at random from a large number of technicians who work at the company.

```
data service;
      infile 'h:\stat512\datasets\ch19pr16.dat';
      input time tech make k;
      mt = make*10+tech;
proc print data=service;
proc glm data=service;
      class make tech;
      model time = make tech make*tech;
      random tech make*tech/test;
```

```
.                          The GLM Procedure
Dependent Variable: time
                                  Sum of
Source                DF         Squares   Mean Square   F Value   Pr > F
Model                  8     1268.177778    158.522222      3.05   0.0101
Error                 36     1872.400000     52.011111
Corrected Total       44     3140.577778

      R-Square      Coeff Var      Root MSE     time Mean
      0.403804       12.91936      7.211873      55.82222

Source       DF      Type I SS    Mean Square   F Value   Pr > F
make          2      28.311111      14.155556      0.27   0.7633
tech          2      24.577778      12.288889      0.24   0.7908
make*tech     4    1215.288889     303.822222      5.84   0.0010
```

We have $MSA = 14.16$, $MSB = 12.29$, $MSAB = 303.82$, and $MSE = 52.01$.

Assuming an **UNRESTRICTED MIXED MODEL**, SAS outputs the EMS as follows,

```
.                  The GLM Procedure
Source       Type III Expected Mean Square
make         Var(Error) + 5 Var(make*tech) + Q(make)
tech         Var(Error) + 5 Var(make*tech) + 15 Var(tech)
make*tech    Var(Error) + 5 Var(make*tech)

      Tests of Hypotheses for Mixed Model Analysis of Variance
Dependent Variable: time
Source                DF   Type III SS   Mean Square F Value   Pr > F
make                   2     28.311111     14.155556    0.05   0.9550
tech                   2     24.577778     12.288889    0.04   0.9607
Error:MS(make*tech) 4   1215.288889   303.822222

Source               DF   Type III SS   Mean Square   F Value   Pr > F
make*tech             4   1215.288889   303.822222       5.84   0.0010
Error: MS(Error)   36   1872.400000    52.011111
```

Assuming a **RESTRICTED MIXED MODEL**, we should test the effects as follows.

- To test the fixed effect make we must use the interaction:

  $F_A = MSA/MSAB = 14.16/303.82 = 0.05 \ldots$ with 2,4 df ($p = 0.955$)

- To test the random effect tech and the interaction, we use error:

  $F_B = MSB/MSE = 12.29/52.01 = 0.24 \ldots$ with $2, 36$ df ($p = 0.7908$)

- To test the interaction effect, we use error:

  $F_{AB} = MSAB/MSE = 303.82/52.01 = 5.84 \ldots$ with $4, 36$ df ($p = 0.001$)

Three-way models

- We can have zero, one, two, or three random effects (etc)
- EMS indicate how to do tests
- In some cases the situation is complicated and we need approximations, e.g. when all are random, use $MS(AB) + MS(AC) - MS(ABC)$ to test A.