Description

Below we demonstrate gsem's group() option, which allows us to fit models in which coefficients, intercepts, and other types of parameters differ across groups of the data. We will fit a Weibull model to the following survival data

```
. use https://www.stata-press.com/data/r19/gsem_cancer
(Patient survival in drug trial)

. describe

Contains data from https://www.stata-press.com/data/r19/gsem_cancer.dta
 Observations:            48                   Patient survival in drug trial
    Variables:             4                   16 Jan 2025 15:41
                                               (_dta has notes)
─────────────────────────────────────────────────────────────────────────────
Variable      Storage   Display    Value
    name         type    format    label      Variable label
─────────────────────────────────────────────────────────────────────────────
studytime      byte     %8.0g                 Months to death or end of exp.
died           byte     %8.0g                 1 if patient died
drug           byte     %8.0g                 Drug type (1=placebo)
age            byte     %8.0g                 Patient's age at start of exp.
─────────────────────────────────────────────────────────────────────────────

Sorted by:

. notes

_dta:
  1.  Artificial data on cancer patient survival.
  2.  Drug 1 is a placebo.  Drugs 2 and 3 are alternative treatments.
```

See [SEM] **Example 47g** and [SEM] **Example 48g** for how to fit survival models using gsem. In this example, we focus on fitting multiple-group models using gsem's group() and ginvariant() options. See [SEM] **Intro 6** for background on these options.
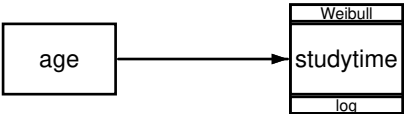
Remarks and examples

Remarks are presented under the following headings:

> *Fitting the multiple-group model*
> *Fitting the model with the Builder*

Fitting the multiple-group model

We want to fit the model

If we do not allow for group differences, we can fit this model by typing

```
. gsem (studytime <- age, family(weibull, failure(died)))
```

However, this study used one placebo and two drugs. We want to allow the intercept and auxiliary parameter to vary across the levels of drug but constrain the coefficient on age to be equal across the levels of drug. In other words, we want to fit a stratified Weibull model. We add the group(drug) and ginvariant(coef) options to our gsem command to specify that drug is the group identifier and that coefficients should not vary across groups.

```
. gsem (studytime <- age, family(weibull, failure(died))),
> group(drug) ginvariant(coef)
  (iteration log omitted)
```

Generalized structural equation model	Number of obs = 48
Grouping variable: drug	Number of groups = 3
Log likelihood = -109.28976	

```
 ( 1)  [studytime]1bn.drug#c.age - [studytime]3.drug#c.age = 0
 ( 2)  [studytime]2.drug#c.age - [studytime]3.drug#c.age = 0
```

Group: 1	Number of obs = 20
Response: studytime	No. of failures = 19
Family: Weibull	Time at risk = 180.00
Form: Proportional hazards	
Link: Log	

	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]
studytime						
age	.1212332	.0367538	3.30	0.001	.049197	.1932694
_cons	-10.36921	2.341022	-4.43	0.000	-14.95753	-5.780896
/studytime						
ln_p	.4541282	.1715663			.1178645	.7903919

Group: 2	Number of obs = 14
Response: studytime	No. of failures = 6
Family: Weibull	Time at risk = 209.00
Form: Proportional hazards	
Link: Log	

	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]
studytime						
age	.1212332	.0367538	3.30	0.001	.049197	.1932694
_cons	-14.93039	3.445179	-4.33	0.000	-21.68282	-8.177965
/studytime						
ln_p	.9413477	.2943728			.3643876	1.518308

```
Group:     3                                    Number of obs   =      14
Response:  studytime                            No. of failures =       6
Family:    Weibull                              Time at risk    = 355.00
Form:      Proportional hazards
Link:      Log
```

	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
studytime						
age	.1212332	.0367538	3.30	0.001	.049197	.1932694
_cons	-14.08495	3.242463	-4.34	0.000	-20.44006	-7.72984
/studytime						
ln_p	.6735495	.369625			-.0509022	1.398001

Notes:

1. In [SEM] **Intro 6**, we wrote that gsem classifies each parameter into one of seven classes, which are the following:

Class description	Class name
1. intercepts and cutpoints	cons
2. fixed coefficients	coef
3. latent variable coefficients	loading
4. covariances of errors	errvar
5. scaling parameters	scale
6. means of exogenous variables	means
7. covariances of exogenous latent variables	covex
8. all the above	all
9. none of the above	none

By default, classes 1, 2, and 3 are constrained to be equal, and the others are allowed to vary.

2. In the output, we have a separate table of parameter estimates for each level of drug. The coefficient on age is 0.12 in all groups, but, as expected, the other parameters differ across groups.

We can replay the results with the byparm option to request that results be sorted by parameter rather than by groups. This output makes it easy to compare estimates across the groups. Alternatively, we could have added the byparm option when we fit the model.

```
. gsem, byparm
```

Generalized structural equation model   Number of obs   =   48
Grouping variable: drug   Number of groups =   3

Group: 1   Number of obs   =   20
Response: studytime   No. of failures =   19
Family: Weibull   Time at risk   = 180.00
Form: Proportional hazards
Link: Log

Group: 2   Number of obs   =   14
Response: studytime   No. of failures =   6
Family: Weibull   Time at risk   = 209.00
Form: Proportional hazards
Link: Log

Group: 3   Number of obs   =   14
Response: studytime   No. of failures =   6
Family: Weibull   Time at risk   = 355.00
Form: Proportional hazards
Link: Log

Log likelihood = −109.28976

(1) [studytime]1bn.drug#c.age − [studytime]3.drug#c.age = 0
(2) [studytime]2.drug#c.age − [studytime]3.drug#c.age = 0

	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
studytime						
age						
1	.1212332	.0367538	3.30	0.001	.049197	.1932694
2	.1212332	.0367538	3.30	0.001	.049197	.1932694
3	.1212332	.0367538	3.30	0.001	.049197	.1932694
_cons						
1	−10.36921	2.341022	−4.43	0.000	−14.95753	−5.780896
2	−14.93039	3.445179	−4.33	0.000	−21.68282	−8.177965
3	−14.08495	3.242463	−4.34	0.000	−20.44006	−7.72984
/studytime						
ln_p						
1	.4541282	.1715663			.1178645	.7903919
2	.9413477	.2943728			.3643876	1.518308
3	.6735495	.369625			−.0509022	1.398001

The estimated intercepts, labeled _cons in the first section of the table, do not vary much across groups. We can perform a Wald test of whether these coefficients are the same using the test command. If we replay the results by typing gsem, coeflegend, we find that we can refer to the three intercepts as _b[studytime:1.drug], _b[studytime:2.drug], and _b[studytime:3.drug]. Therefore, our test command is

```
. test _b[studytime:1.drug]=_b[studytime:2.drug]=_b[studytime:3.drug]
```

(1) [studytime]1bn.drug − [studytime]2.drug = 0
(2) [studytime]1bn.drug − [studytime]3.drug = 0

        chi2(2) =   5.49
      Prob > chi2 =   0.0641

Using a 5% significance level, we fail to reject the null hypothesis that the intercepts are equal. If we wanted to refit the model based on these findings, we could include cons in the ginvariant() option as follows and constrain both the coefficients and the intercepts across groups.

```
. gsem (studytime <- age, family(weibull, failure(died))),  ///
  group(drug) ginvariant(coef cons)
```

Fitting the model with the Builder

Use the diagram in *Fitting the multiple-group model* above for reference.

1. Open the dataset.

   In the Command window, type

   ```
   . use https://www.stata-press.com/data/r19/gsem_cancer
   ```

2. Open a new Builder diagram.

   Select menu item **Statistics > SEM (structural equation modeling) > Model building and estimation**.

3. Put the Builder in gsem mode by clicking on the $\overset{G}{\text{SEM}}$ button.

4. Create the Weibull regression component for studytime.

   Select the Add regression component tool, ⌐, and then click in the diagram about one-third of the way in from the left and halfway down.

   In the resulting dialog box,

   a. select studytime in the *Dependent variable* control;

   b. check *Make response generalized*;

   c. select Weibull, Log in the *Family/Link* control;

   d. select the *Select variables* radio button (it may already be selected);

   e. use the *Independent variables* control to select the variable age;

   f. select Left in the *Independent variables' direction* control;

   g. click on **OK**.

   If you wish, move the component by clicking on any variable and dragging it.

5. Specify the censoring variable.

   a. Choose the Select tool, ⬉.

   b. Click on the box for studytime.

   c. In the Contextual Toolbar, click on the **Properties...** button.

   d. In the resulting *Variable properties* dialog box, click on the **Failure and truncation...** button in the **Variable** tab.

   e. In the resulting *Failure and truncation* dialog box, select died in the *Failure variable* control. Click on **OK**.

   f. Click on **OK** in the *Variables properties* dialog box.

6. Clean up.

   The box for _t is created closer to the independent variables than it is in the example diagram. Use the Select tool, ⬉, and click on the box for _t. Drag it to the right to allow more space for results along the paths.

7. Estimate.

   Click on the **Estimate** button, ▶, in the Standard Toolbar.

   In the resulting dialog box, do the following:

   a. Select the **Group** tab.

   b. Select the *Group analysis* radio button. Select the variable `drug` in the *Group variable* control.

   c. Select *Fixed coefficients* in the *Parameters that are equal across groups* control.

   d. Click on **OK**.

   e. In the Standard Toolbar, use the *Group* control to toggle between results for group 1 and group 2.

   You can open a completed diagram in the Builder by typing

   ```
   . webgetsem gsem_grp
   ```

Also see

[SEM] **Example 47g** — Exponential survival model

[SEM] **Example 48g** — Loglogistic survival model with censored and truncated data

[SEM] **Intro 5** — Tour of models

[SEM] **Intro 6** — Comparing groups

[SEM] **gsem** — Generalized structural equation model estimation command