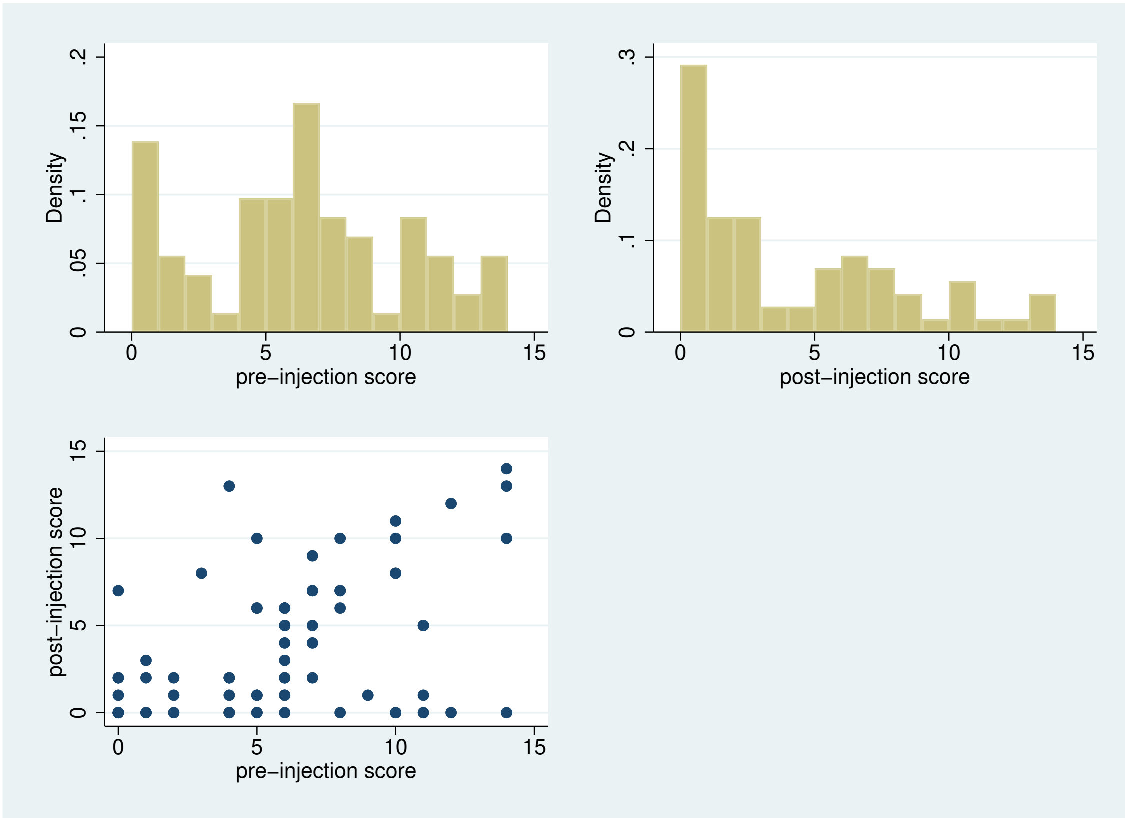
Generalized Linear Models

Session 1 - Practical Solution and Commentary

1. The code to produce the plots below is given in the Do file.

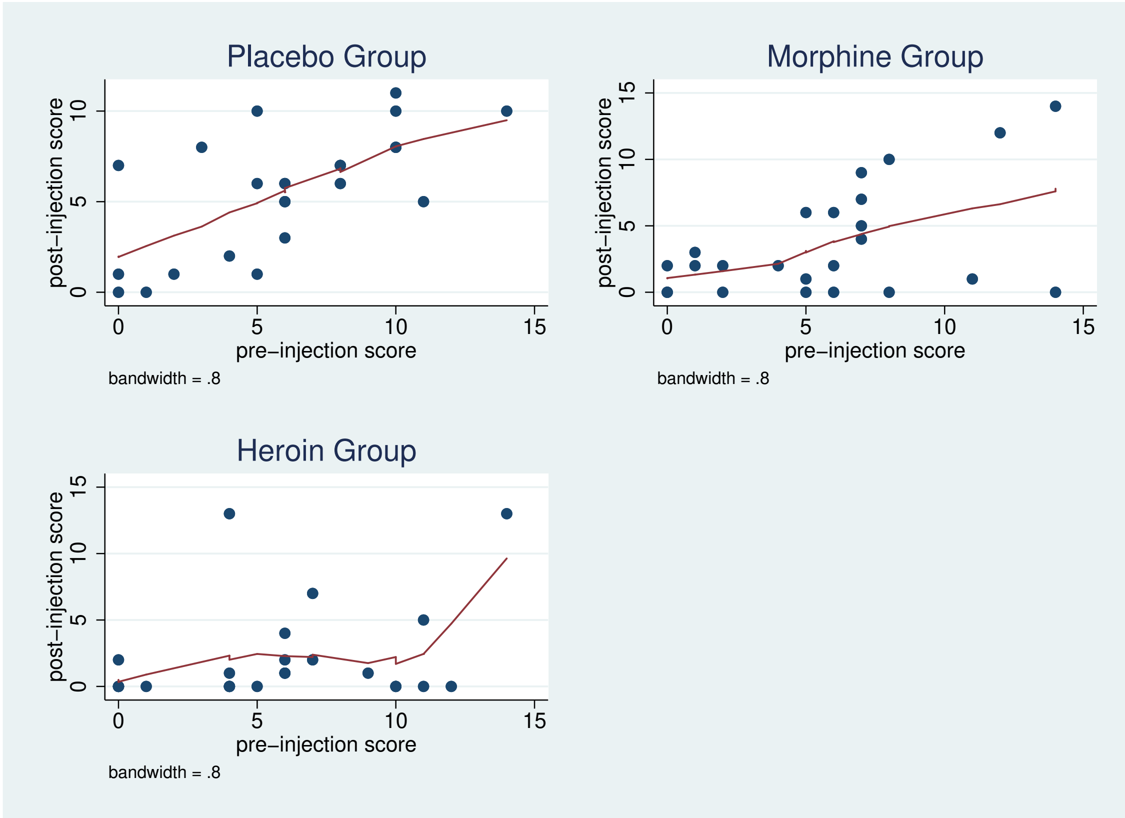



**Discussion: What can be concluded about the appropriateness of linear regression models for post-injection scores on the basis of these preliminary explorations?**

The histograms of pre- and post-injection scores suggest that these variables are not normally distributed. Note, however, that this does not necessarily invalidate inferences that we want to draw from the regression models that we are going to fit, since in a regression model it is only residuals that need to be normally distributed for valid inferences.

The scatter plot shows that pre- and post-injection scores are positively associated.

2. The code to produce the plots overleaf is given in the Do file.




Given the relatively small numbers of observations in each trial arm an assumption of linearity seems reasonable here.

3. Your completed table should look similar to that below.

**Table 1**

| # | Linear predictor |
|---|---|
| 1 | $\eta_i = \beta_0$ |
| 2 | $\eta_i = \beta_0 + \beta_1 u_i + \beta_2 v_i$ |
| 3 | $\eta_i = \beta_0 + \beta_3 x_i$ |
| 4 | $\eta_i = \beta_0 + \beta_1 u_i + \beta_2 v_i + \beta_3 x_i$ |
| 5 | $\eta_i = \beta_0 + \beta_1 u_i + \beta_2 v_i + \beta_3 x_i + \beta_{13} u_i x_i + \beta_{23} v_i x_i$ |

Here $x_i$ and $Y_i$ are respectively the pre- and post-injection scores for individual $i$. The additional predictor variables are as follows:

$$u_i = \begin{cases} 0 & \text{Placebo or Heroin} \\ 1 & \text{Morphine} \end{cases} \quad \text{and} \quad v_i = \begin{cases} 0 & \text{Placebo or Morphine} \\ 1 & \text{Heroin.} \end{cases}$$

In each model the link function is the identity, $\eta_i = \mu_i$, and it is assumed that $Y_i \sim N(\mu_i, \sigma^2)$ and are independent.

4. 1. Overall mean model

Linear predictor $\eta_i = \beta_0$

```
. reg mentact

      Source |       SS       df       MS              Number of obs =      72
-------------+------------------------------           F(0, 71)      =    0.00
       Model |           0     0           .           Prob > F      =       .
    Residual |    1117.875    71  15.7447183           R-squared     =  0.0000
-------------+------------------------------           Adj R-squared =  0.0000
       Total |    1117.875    71  15.7447183           Root MSE      =   3.968

------------------------------------------------------------------------------
     mentact |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   3.791667   .4676287     8.11   0.000     2.859241    4.724092

------------------------------------------------------------------------------
```

2. Drugs model

Linear predictor $\eta_i = \beta_0 + \beta_1 u_i + \beta_2 v_i$

```
. reg mentact i.treat

      Source |       SS       df       MS              Number of obs =      72
-------------+------------------------------           F(2, 69)      =    4.83
       Model |      137.25     2      68.625           Prob > F      =  0.0109
    Residual |     980.625    69  14.2119565           R-squared     =  0.1228
-------------+------------------------------           Adj R-squared =  0.0974
       Total |    1117.875    71  15.7447183           Root MSE      =  3.7699

------------------------------------------------------------------------------
     mentact |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |
Morphine ..  |     -1.875   1.088269    -1.72   0.089    -4.046038    .2960375
Heroin Gr..  |     -3.375   1.088269    -3.10   0.003    -5.546038   -1.203962
             |
       _cons |   5.541667   .7695225     7.20   0.000     4.006511    7.076822
------------------------------------------------------------------------------
```

3. Pre-inj model

Linear predictor $\eta_i = \beta_0 + \beta_3 x_i$

```
. reg mentact prement

      Source |       SS       df       MS              Number of obs =      72
-------------+------------------------------           F(1, 70)      =   18.49
       Model |  233.54726      1   233.54726           Prob > F      =  0.0001
    Residual |  884.32774     70  12.6332534           R-squared     =  0.2089
-------------+------------------------------           Adj R-squared =  0.1976
       Total |   1117.875     71  15.7447183           Root MSE      =  3.5543

------------------------------------------------------------------------------
     mentact |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     prement |   .4587228   .1066892     4.30   0.000     .2459379    .6715078
       _cons |    1.09667   .7538827     1.45   0.150    -.4069016    2.600242
------------------------------------------------------------------------------
```

4. Drugs + Pre-inj model

Linear predictor $\eta_i = \beta_0 + \beta_1 u_i + \beta_2 v_i + \beta_3 x_i$

```
. reg mentact i.treat prement

      Source |       SS       df       MS              Number of obs =      72
-------------+------------------------------           F(3, 68)      =   11.03
       Model | 365.819593      3  121.939864           Prob > F      =  0.0000
    Residual | 752.055407     68  11.0596383           R-squared     =  0.3272
-------------+------------------------------           Adj R-squared =  0.2976
       Total |   1117.875     71  15.7447183           Root MSE      =  3.3256

------------------------------------------------------------------------------
     mentact |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |
Morphine ..  |   -1.76151   .9603436    -1.83   0.071    -3.677845    .1548261
Heroin Gr..  |  -3.318255   .9601002    -3.46   0.001    -5.234105   -1.402405
             |
     prement |   .4539615   .0998574     4.55   0.000     .2546991    .6532238
       _cons |   2.817898   .9054238     3.11   0.003     1.011153    4.624643
------------------------------------------------------------------------------
```

5. Drugs + Pre-inj + Drugs-by-Pre-inj interaction model

Linear predictor $\eta_i = \beta_0 + \beta_1 u_i + \beta_2 v_i + \beta_3 x_i + \beta_{13} u_i x_i + \beta_{23} v_i x_i$

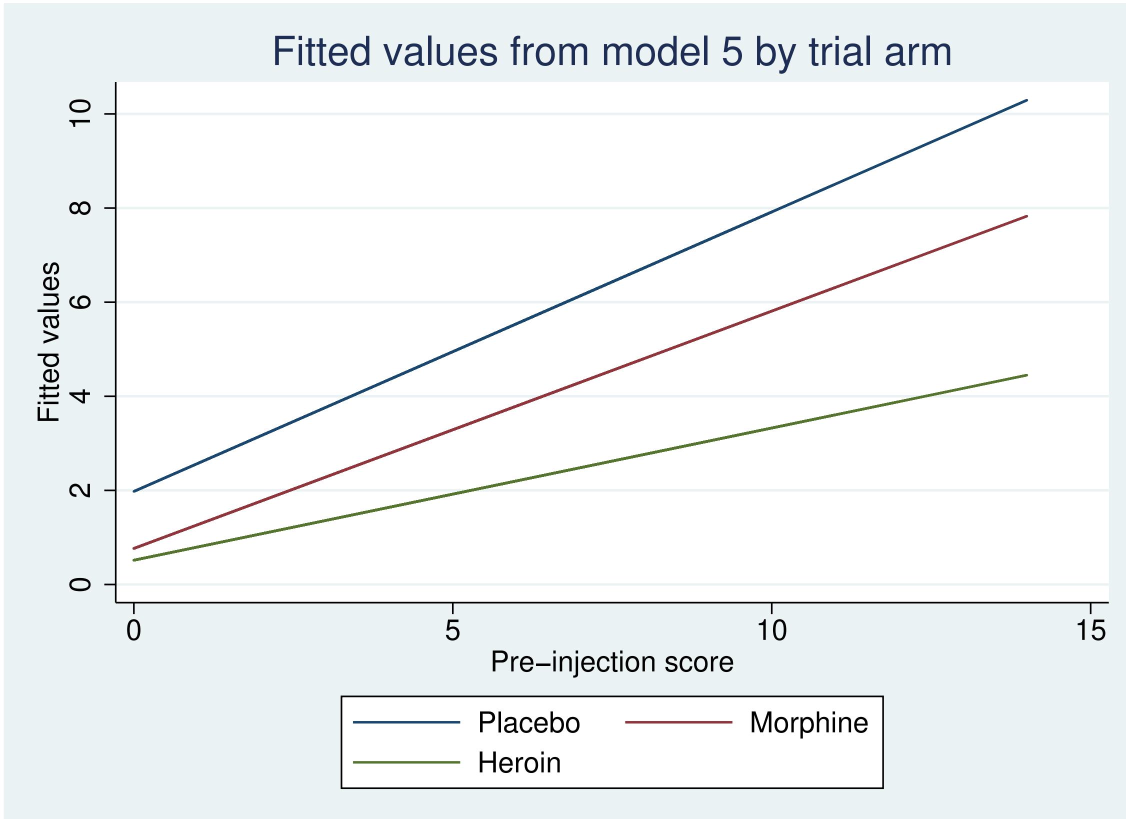
```
. reg mentact i.treat##c.prement

      Source |       SS       df       MS              Number of obs =      72
-------------+------------------------------           F(5, 66)      =    6.93
       Model | 384.747393      5  76.9494786           Prob > F      =  0.0000
    Residual | 733.127607     66   11.107994           R-squared     =  0.3442
-------------+------------------------------           Adj R-squared =  0.2945
       Total |   1117.875     71  15.7447183           Root MSE      =  3.3329

------------------------------------------------------------------------------
     mentact |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |
Morphine ..  |  -1.211742   1.750342    -0.69   0.491    -4.706413    2.282929
Heroin Gr..  |  -1.461968   1.771855    -0.83   0.412    -4.999591    2.075655
             |
     prement |   .5939394   .1834682     3.24   0.002     .2276333    .9602455
             |
      treat#|
   c.prement |
Morphine ..  |  -.0895258   .2483459    -0.36   0.720    -.5853644    .4063129
Heroin Gr..  |  -.3129855   .2503829    -1.25   0.216    -.8128911    .1869202
             |
       _cons |    1.97803   1.294069     1.53   0.131    -.6056616    4.561722
------------------------------------------------------------------------------
```

The table below shows the number of parameters, the residual sums of squares, and the degrees of freedoms for each of the five models.

**Table 2**

| Model ($\mu_i$) | no. pars | RSS | rdf |
|---|---|---|---|
| 1. $\beta_0$ | 1 | 1117.9 | 71 |
| 2. $\beta_0 + \beta_1 u_i + \beta_2 v_i$ | 3 | 980.6 | 69 |
| 3. $\beta_0 + \beta_3 x_i$ | 2 | 884.3 | 70 |
| 4. $\beta_0 + \beta_1 u_i + \beta_2 v_i + \beta_3 x_i$ | 4 | 752.1 | 68 |
| 5. $\beta_0 + \beta_1 u_i + \beta_2 v_i + \beta_3 x_i + \beta_{13} u_i x_i + \beta_{23} v_i x_i$ | 6 | 733.1 | 66 |

5. The code to produce the following plot is given in the Do file.




The fitted values from model 5 lie along three straight lines, one for each trial arm, with the slopes and intercepts being derived from the Stata output on the previous page as follows.

**Table 3**

| Trial arm | Fitted intercept | Fitted slope |
|---|---|---|
| Placebo | 1.97803 | 0.5939394 |
| Morphine | $1.97803 + -1.211742 = 0.7662884$ | $0.5939394 + -0.0895258 = 0.5044136$ |
| Heroin | $1.97803 + -1.461968 = 0.5160623$ | $0.5939394 + -0.3129855 = 0.2809539$ |

6. i) To compare the fits of models 3 and 4 compute the relevant F-statistic from the entries in Table 2.

$$F = \frac{(884.3 - 752.1)/(70 - 68)}{752.1/68} = 5.98$$

**Discussion: What do you conclude from this test?**

Comparison with the $F_{2,68}$ distribution gives p=0.004. There is therefore strong evidence of a difference in the post-injection scores between the groups after adjusting for pre-injection scores.

This test can also be carried out in Stata using the following command (after re-fitting model 4.).

```
. test 2.treat 3.treat

( 1)  2.treat = 0
( 2)  3.treat = 0

      F(  2,    68) =    5.98
           Prob > F =    0.0041
```

ii) To compare the fits of models 4 and 5 again compute the relevant F-statistic from the entries in Table 2.

$$F = \frac{(752.1 - 733.1)/(68 - 66)}{733.1/66} = 0.85$$

**Discussion: What do you conclude from this test?**

Comparison with the $F_{2,66}$ distribution gives p=0.43. There is therefore no evidence of an interaction between group and pre-injection score.

This test can also be carried out in Stata using the following command (after re-fitting model 5.).

```
. test 2.treat#c.prement 3.treat#c.prement

( 1)  2.treat#c.prement = 0
( 2)  3.treat#c.prement = 0

      F(  2,    66) =    0.85
           Prob > F =    0.4312
```

**Discussion: Why can a F-test not be used to compare the fits of models 2 and 3?**

Models 2 and 3 are not nested, so cannot be compared using an analogous approach to that above.

7. The parameter estimates and standard errors are identical. However, the confidence intervals and p-values are slightly different. This is because the `glm` command approximates the exact t-distribution with the (asymptotically valid) z distribution. This is reported in the output table (heading of third column of results).

8. Key parameter estimates (and SEs) from models 2 & 4 are as follows.

**Table 4**

| | Mean | Mean difference from Placebo | SE | Adj. Mean difference from placebo | SE |
|---|---|---|---|---|---|
| Placebo | 5.542 | 0 | – | 0 | – |
| Morphine | 3.667 | -1.875 | 1.088 | -1.762 | 0.960 |
| Heroin | 2.167 | -3.375 | 1.088 | -3.318 | 0.960 |

9. **Discussion: Write a paragraph (or three!) to summarise your conclusions from these analyses.**

   The F-test (from `reg mentact i.treat`) gives $p = 0.0109$, indicating moderately strong evidence that the groups differ in respect of their post-injection scores. Looking at the p-values corresponding to the individual coefficients, we see that the estimated difference between morphine and placebo is not formally statistically significant at the 5% level, but is close ($p = 0.089$), whilst the estimated difference between heroin and placebo is formally statistically significant ($p = 0.003$).

   Adjusting for "prement" changes the estimated group differences somewhat, but not dramatically. This is as we would expect: since the trial is randomized, the purpose of baseline adjustment is to improve the efficiency of treatment effect estimates (hence the reduced standard error), but we do not expect systematic changes in between-group differences since there can be no confounding.

   The F-test for the interaction parameters in the last model gives $p = 0.43$, indicating no evidence against the null hypothesis that the effects of the drugs are the same irrespective of pre-injection mental activity. So, we cannot say that there is evidence that the effects of the drugs are different for subjects with different levels of pre-injection mental activity.

10. i) Using lincom
    After fitting model 4 the lincom command can be usefully used to obtain a 95% confidence interval for the difference in the post-injection scores on heroin and morphine after adjusting for pre-injection scores.

```
. lincom 3.treat-2.treat

 ( 1)  - 2.treat + 3.treat = 0

------------------------------------------------------------------------------
     mentact |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -1.556745   .9601002    -1.62   0.110    -3.472595    .3591049
------------------------------------------------------------------------------
```

    ii) Re-parameterise the model
    We can use the following command to change the reference category to the second level of the treatment variable, which corresponds to the morphine group.

    `regress mentact b2.treat prement`

    The coefficient for the heroin group now represents the difference between the morphine and heroin groups.