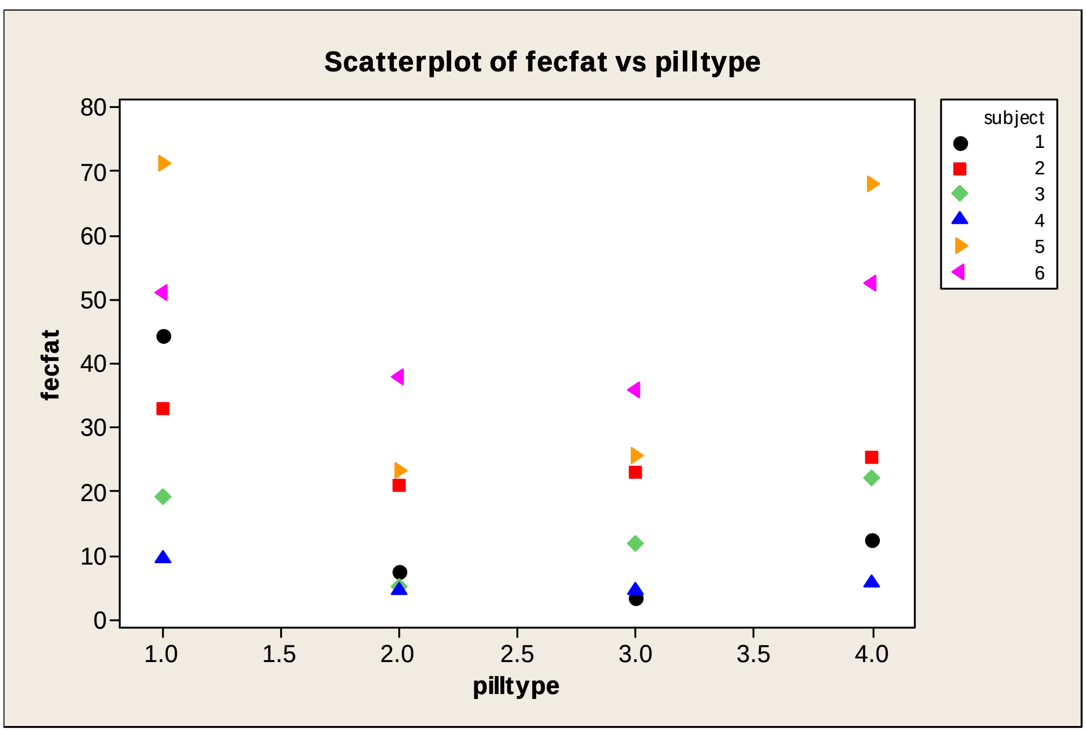
Fecal Fat Example on p.254

The model (from page 256 equation (8.2)) is $\mathbf{FECFAT_{st} = \mu + SUBJECT_s + PILLTYPE_t + \varepsilon_{st}}$, for s = 1 … 6 subjects, t = 1 … 4 pill-types, $\varepsilon_{st} \sim N(0,\sigma_\varepsilon^2)$ and $\mathbf{SUBJECT_s} \sim N(0,\sigma_{subj}^2)$ - thus subjects are "random".

General Linear Model: fecfat versus subject, pilltype

```
Factor    Type    Levels  Values
subject   random       6  1, 2, 3, 4, 5, 6
pilltype  fixed        4  1, 2, 3, 4

Analysis of Variance for fecfat, using Adjusted SS for Tests
Source    DF  Seq SS  Adj SS  Adj MS      F      P
subject    5  5588.4  5588.4  1117.7  10.45  0.000
pilltype   3  2008.6  2008.6   669.5   6.26  0.006
Error     15  1605.0  1605.0   107.0
Total     23  9202.0

S = 10.3440   R-Sq = 82.56%   R-Sq(adj) = 73.26%

Expected Mean Squares, using Adjusted SS
        Expected Mean Square
   Source    for Each Term
1  subject   (3) + 4.0000 (1)
2  pilltype  (3) + Q[2]
3  Error     (3)

Variance Components, using Adjusted SS
         Estimated
Source       Value
subject      252.7
Error        107.0
```

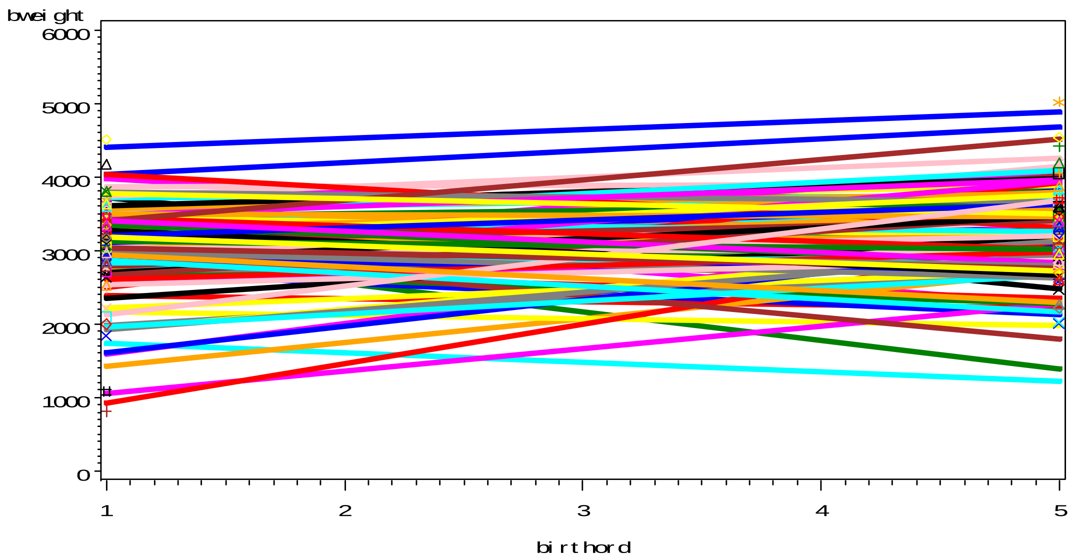


Now, "finish the story" by performing a MCP (MSP = mean separation procedure)

Fecal Fat Multiple Comparisons (using Bonferroni method)

```
Least Squares Means for fecfat

pilltype   Mean
1         38.08
2         16.53
3         17.42
4         31.07

Bonferroni Simultaneous Tests
Response Variable fecfat
All Pairwise Comparisons among Levels of pilltype

pilltype = 1  subtracted from:

          Difference       SE of             Adjusted
pilltype    of Means  Difference  T-Value     P-Value
2             -21.55       5.972   -3.608      0.0155
3             -20.67       5.972   -3.461      0.0210
4              -7.02       5.972   -1.175      1.0000

pilltype = 2  subtracted from:

          Difference       SE of             Adjusted
pilltype    of Means  Difference  T-Value     P-Value
3             0.8833       5.972   0.1479      1.0000
4            14.5333       5.972   2.4335      0.1676

pilltype = 3  subtracted from:

          Difference       SE of             Adjusted
pilltype    of Means  Difference  T-Value     P-Value
4              13.65       5.972    2.286      0.2235
```

GA Babies – p.262ff: only examining first- and last-born in families of 5



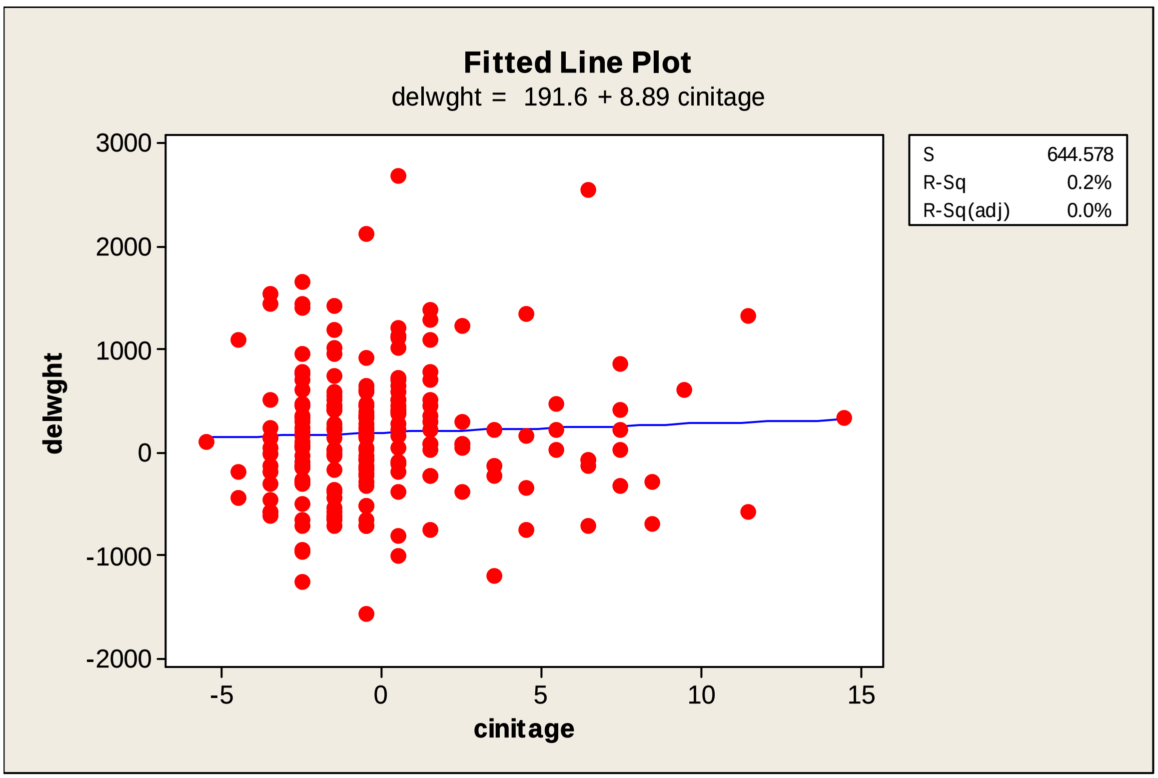
A. Paired t-test

One-Sample T: delwght

```
Test of mu = 0 vs not = 0

Variable    N   Mean  StDev  SE Mean       95% CI          T      P
delwght   200  191.6  643.6     45.5  (101.9, 281.4)  4.21  0.000
```

B. Now take account of centered initial age (CINITAGE): initial age mean = 17.545 subtracted from each value of initial age

Model Function: DELWGHT = BW5 – BW1 = $\beta_0 + \beta_1$*CINITAGE + ε



Regression Analysis: delwght versus cinitage

```
The regression equation is
delwght = 192 + 8.9 cinitage

Predictor     Coef  SE Coef     T      P
Constant    191.64    45.58  4.20  0.000
cinitage      8.89    14.16  0.63  0.531

S = 644.578   R-Sq = 0.2%   R-Sq(adj) = 0.0%
```

```
Analysis of Variance
Source            DF        SS      MS     F      P
Regression         1    163789  163789  0.39  0.531
Residual Error   198  82265157  415481
```

C. A variant of **Repeated Measures (RM) analysis** is performed in Table 8.6 on p.264. The key term is the interaction between birth order and centered initial age (called "_IbirXcini~5"). That this term is non-significant (p = 0.528) tells us that the difference in birthweights between the first- and last-born is not related to (centered) initial age, exactly the same question answered in part B above. We get essentially the same results using SAS.

```
data three;
  set one; if birthord=1 or birthord=5;
  d5=(birthord=5); d5cinitage=d5*cinitage;
proc genmod data=three;
  class momid;
  model bweight=d5 cinitage d5*cinitage/dist=normal link=identity;
  repeated subject=momid / type=exch;
run;
```

```
                              The GENMOD Procedure

                                Model Information
                         Data Set                  WORK.THREE
                         Distribution                  Normal
                         Link Function               Identity
                         Dependent Variable           bweight

                     Number of Observations Read         400
                     Number of Observations Used         400

                             Class Level Information
       Class       Levels     Values
       momid          200     39 62 79 80 84 92 108 113 125 135 199 200 221 247
                              304 336 547 853 960 1232 1448 1638 1706 1785 1856
                              2083 2166 2292 2301 2383 2519 2598 2613 2647 2735
                              2822 2899 2906 2918 2928 3044 3168 3308 3377 3431
                              3438 3469 3477 3480 3504 3526 3668 3797 3838 3936
                              ...

                              Parameter Information
                           Parameter        Effect
                           Prm1             Intercept
                           Prm2             d5
                           Prm3             cinitage
                           Prm4             d5*cinitage

                     Criteria For Assessing Goodness Of Fit
             Criterion                 DF         Value        Value/DF
             Deviance                 396  129458162.97     326914.5530
             Scaled Deviance          396      400.0000          1.0101
             Pearson Chi-Square       396  129458162.97     326914.5530
```

```
                  Scaled Pearson X2         396         400.0000          1.0101
                  Log Likelihood                       -3105.0562
          Algorithm converged.

                            Analysis Of Initial Parameter Estimates
                                    Standard     Wald 95% Confidence        Chi-
   Parameter       DF     Estimate     Error            Limits             Square     Pr > ChiSq
   Intercept        1     3016.555    40.2272    2937.711    3095.399    5623.19        <.0001
   d5               1     191.6400    56.8898     80.1380    303.1420      11.35        0.0008
   cinitage         1      25.1398    12.4992      0.6418     49.6378       4.05        0.0443
   d5*cinitage      1       8.8918    17.6765    -25.7536     43.5372       0.25        0.6149
   Scale            1     568.8984    20.1136    530.8114    609.7183
NOTE: The scale parameter was estimated by maximum likelihood.

                                     GEE Model Information
                       Correlation Structure                     Exchangeable
                       Subject Effect                      momid (200 levels)
                       Number of Clusters                                 200
                       Correlation Matrix Dimension                         2
                       Maximum Cluster Size                                 2
                       Minimum Cluster Size                                 2
          Algorithm converged.

                                     Exchangeable Working
                                          Correlation
                                  Correlation     0.3682623519

                              Analysis Of GEE Parameter Estimates
                               Empirical Standard Error Estimates

                                      Standard    95% Confidence
                Parameter   Estimate     Error        Limits            Z Pr > |Z|
                Intercept   3016.555   40.3205 2937.528 3095.582    74.81   <.0001
                d5          191.6400   45.3501 102.7555 280.5245     4.23   <.0001
                cinitage     25.1398   12.2092   1.2101  49.0695     2.06   0.0395
                d5*cinitage   8.8918   15.1661 -20.8332  38.6168     0.59   0.5577
```

D. Finally, let's **regress BW5 on CINITAGE and BW1** (see p.265) via

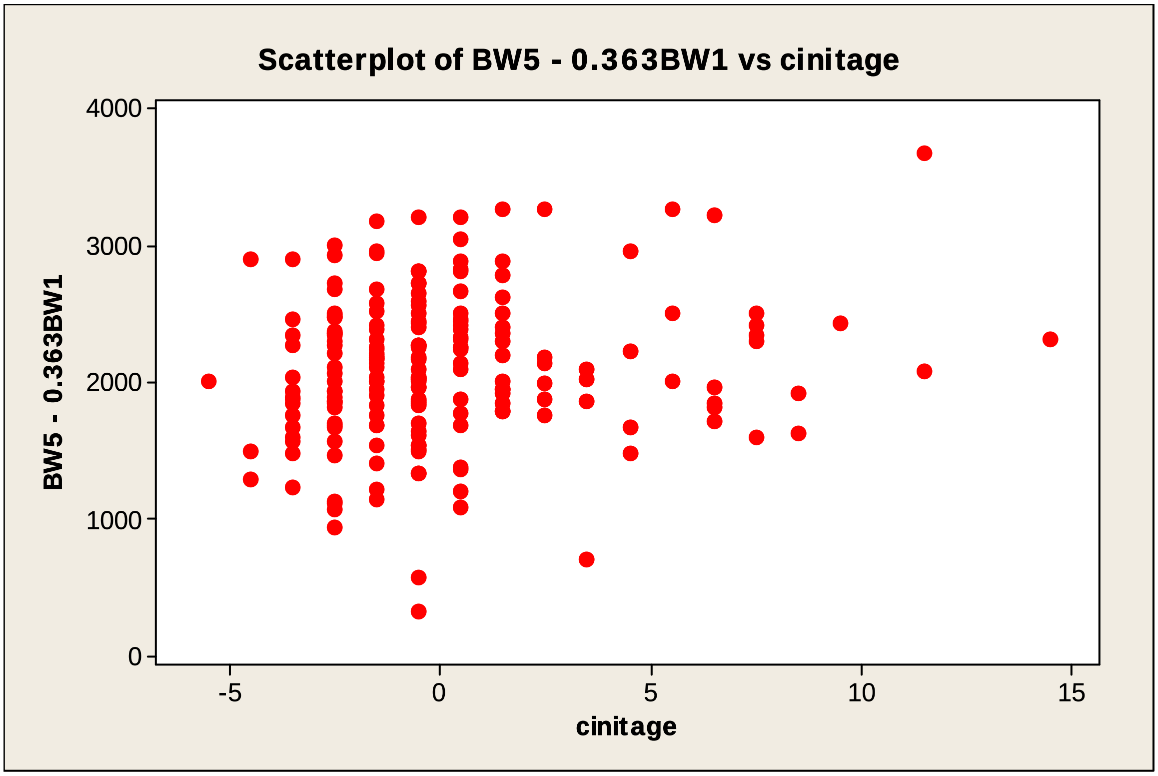
Model Function: BW5 = $\beta_0 + \beta_1$*CINITAGE + β_2*BW1 + ε

Regression Analysis: bw5 versus cinitage, bw1

```
The regression equation is
bw5 = 2114 + 24.9 cinitage + 0.363 bw1

Predictor       Coef   SE Coef      T      P
Constant      2113.6     202.7  10.43  0.000
cinitage       24.91     11.82   2.11  0.036
bw1          0.36286   0.06604   5.49  0.000

S = 532.527   R-Sq = 16.4%   R-Sq(adj) = 15.6%

Analysis of Variance
```

Source	DF	SS	MS	F	P
Regression	2	10961363	5480682	19.33	0.000
Residual Error	197	55866154	283585		
Total	199	66827517			

Note that β_2 above is estimated to be 0.363, so we let Y = BW5 – 0.363*BW1 in the following graph, and now note the significant relationship with CINITAGE (p = 0.036).



The above results notwithstanding, the authors recommend against using first-born birthweight as a covariate in observational studies such as this one (see discussion on p.265 center and bottom). In this case, we would conclude that the difference between first- and last-born birthweights does not depend upon the initial age of the mother. But, the conclusion is really secondary here – here, the authors are demonstrating that there are several methods to analyze repeated measures data. Obviously, the only one of these that is useful when we have more than two time-points is the third method (repeated measures) above; this method is called the GEE method, short for Generalized Estimating Equations. The authors claim that it is preferred as it is "robust", that is, it assumes no specific variance-covariance pattern.