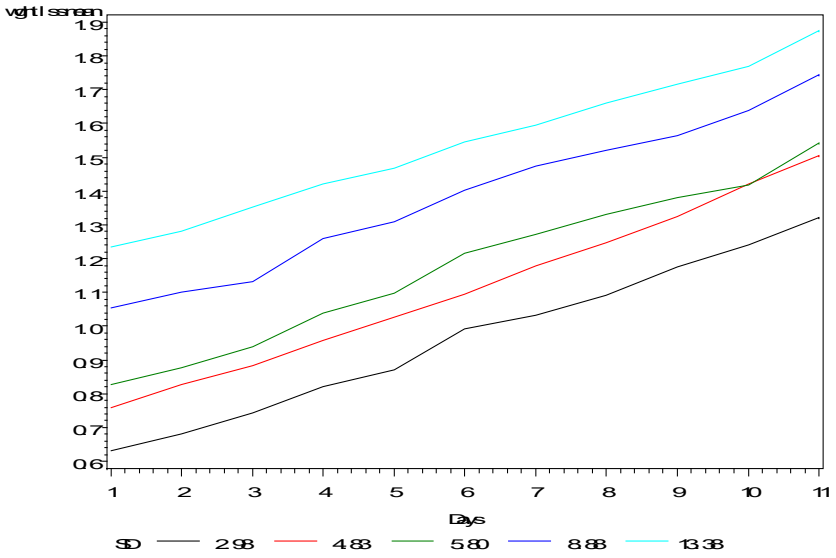
1)

```
proc sort;
by days sd;
run;

proc means mean;
var response;
by days sd;
output out=summary mean=wghtlssmean;
run;

symbol1 i=j;
proc gplot data=summary;
plot wghtlssmean*days=sd;
run;
```



appears an increase in the mean water loss as the level of saturation deficit increases.

This is a repeated measures design with Saturation Deficit (SD) the main factor and Days of Exposure the factor upon which the repeated observations are taken. There are 5 replications of the complete experiment.

a. A model for this experiment is given here:

$y_{ijk} = \mu + \alpha_i + \pi_{j(i)} + \beta_k + \alpha\beta_{ij} + \epsilon_{ijk}$, where

y_{ijk} is the water loss of the j^{th} tick during Day k under the i^{th} level of Saturation Deficit.

α_i is the fixed effect of the i^{th} level of SD

$\pi_{j(i)}$ is the random effect of tick j at SD level i, iid $N(0,\sigma_\tau^2)$ r.v.'s

β_k is the fixed effect of Day k

$\alpha\beta_{ik}$ is the fixed interaction effect of the i^{th} SD level with Day k

ϵ_{ijk} is the random effect of all other factors on water loss, iid $N(0,\sigma_\epsilon^2)$ r.v.'s

1)continued …

```
proc glm data=cayenne;
class days sd tick;
model response = sd tick(sd) days days *sd;
random tick(sd) / test;
run;
```

```
The GLM Procedure
Tests of Hypotheses for Mixed Model Analysis of Variance

Dependent Variable: Response

   Source                     DF   Type III SS   Mean Square  F Value  Pr > F

*  SD                          4     11.327052      2.831763    23.27  <.0001

   Error: MS(Tick(SD))        20      2.434098      0.121705
* This test assumes one or more other fixed effects are zero.

   Source                     DF   Type III SS   Mean Square  F Value  Pr > F

   Tick(SD)                   20      2.434098      0.121705    58.52  <.0001
*  Days                       10     13.251695      1.325169   637.19  <.0001
   Days*SD                    40      0.080764      0.002019     0.97  0.5261

   Error: MS(Error)          200      0.415942      0.002080
* This test assumes one or more other fixed effects are zero.
```

First, we test the main effect for SD which is significant with p-value < 0.0001. Thus, we can conclude that the mean water loss is different across the five levels of SD. From the profile plot, we observe that as SD increases, the mean water loss appears to increase.

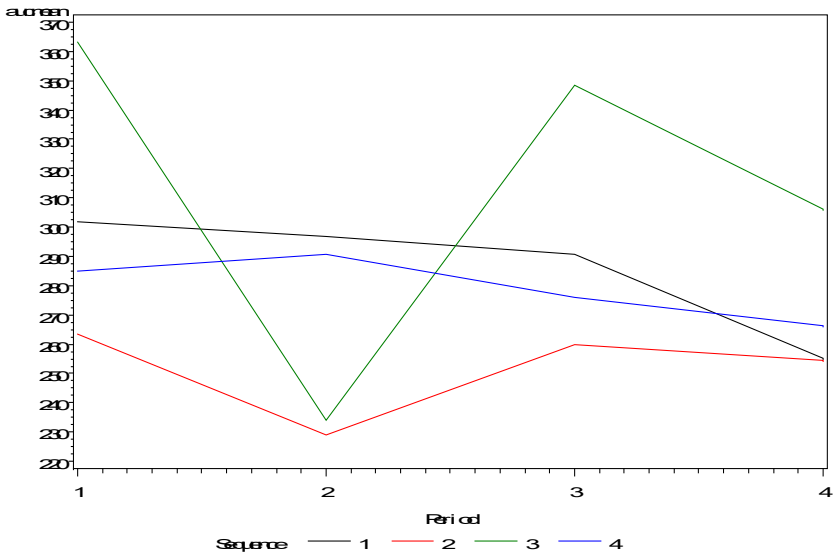
Yes, it is plausible for the increase in mean whole-body weight loss for the cayenne tick over the study to be consistent for each of the levels of SD … there is not a significant interaction between SD and Days (p-value = 0.5261). Furthermore, the profile plot displays nearly parallel lines for the five levels of SD.

2)

```
proc sort;
by period sequence;
run;

proc means mean;
var auc;
by period sequence;
output out=summary mean=aucmean;
run;

symbol1 i=j;
proc gplot data=summary;
plot aucmean*period=sequence;
run;
```




```
Period=1 Analysis Variable : AUC

Mean
301.0707692

Period=2 Analysis Variable : AUC

Mean
262.4526923

Period=3 Analysis Variable : AUC

Mean
291.7342308

Period=4 Analysis Variable : AUC

Mean
269.7811538
```

Some evidence for a period effect as the means across the periods vary …

2)continued …

Treatment=A

The MEANS Procedure

Analysis Variable : AUC

Mean
296.9607692

Treatment=B

Analysis Variable : AUC

Mean
284.3461538

Treatment=C

Analysis Variable : AUC

Mean
284.2238462

Treatment=D

Analysis Variable : AUC

Mean
259.5080769

some evidence of different AUC means as the means vary …

2) continued …

The model is as follows:

$y_{ijk} = \mu + \delta_i + \beta_{j(i)} + \gamma_k + \tau_{d(i,k)} + \varepsilon_{ijk}$; $i = 1,2,3,4$; $j(1) = j(3) = 1,2,3,4,5,6$;
$j(2) = j(4) = 1,2,3,4,5,6,7$; $k = 1,2,3,4$ where

δ_i is the fixed sequence effect

$\beta_{j(i)}$ is the random patient within treatment effect

γ_k is the fixed effect of the k^{th} time period

$\tau_{d(i,k)}$ is the fixed effect of the i^{th} treatment

ε_{ijk} is the random effect due to all other factors

```
PROC GLM DATA = angina;
    CLASS sequence treatment period subject;
    MODEL AUC = sequence subject(sequence) treatment period ;
    TEST H=sequence  E=subject(sequence);
RUN;
```

Dependent Variable: AUC

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	31	1259105.235	40616.298	6.39	<.0001
Error	72	457663.569	6356.438		
Corrected Total	103	1716768.804			

R-Square	Coeff Var	Root MSE	AUC Mean
0.733416	28.34650	79.72728	281.2597

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Sequence	3	49286.962	16428.987	2.58	0.0598
Subject(Sequence)	22	1162481.582	52840.072	8.31	<.0001
Treatment	3	19187.195	6395.732	1.01	0.3951
Period	3	28149.496	9383.165	1.48	0.2282

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Sequence	3	49286.962	16428.987	2.58	0.0598
Subject(Sequence)	22	1162481.582	52840.072	8.31	<.0001
Treatment	3	21657.666	7219.222	1.14	0.3405
Period	3	28149.496	9383.165	1.48	0.2282

Tests of Hypotheses Using the Type III MS
for Subject(Sequence) as an Error Term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Sequence	3	49286.96207	16428.98736	0.31	0.8173

The p-value for treatment effect is 0.340 > 0.05 which implies there is no significant difference in the AUC means across treatment. As the hypothesis test for equality of treatments was not rejected, it is concluded that no pair of treatments significantly differ from each other.

2)continued …

```
PROC SORT DATA = angina; BY subject period;
DATA angina2; SET angina;
    KEEP subject sequence treatment period AUC CO;
    CO = LAG(treatment); IF period = 1 tHEN CO = '0';
RUN;

PROC GLM DATA = angina2;
    CLASS sequence treatment period subject co;
    MODEL AUC = sequence subject(sequence) treatment period co;
    TEST H=sequence  E=subject(sequence);
RUN;
```

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Sequence	3	21181.592	7060.531	1.10	0.3546
Subject(Sequence)	22	1162481.582	52840.072	8.24	<.0001
Treatment	3	10289.688	3429.896	0.54	0.6598
Period	2	13366.728	6683.364	1.04	0.3580
CO	3	15337.201	5112.400	0.80	0.4995

Tests of Hypotheses Using the Type III MS
for Subject(Sequence) as an Error Term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Sequence	3	21181.59180	7060.53060	0.13	0.9390

The p-value for the Carryover effect is 0.4995 > 0.05 which implies there is no significant evidence of a carryover effect. Therefore, it can be assumed the drugs of a previous treatment have washed out of the system prior to the next treatment being administered. Treatment effect remains non-significant …

2)continued …

```
proc sort data=angina;
by period;
run;

proc glm data=angina;
class treatment;
model auc = treatment;
by period;
run;
```

Period=1

The GLM Procedure

Dependent Variable: AUC

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	34805.4226	11601.8075	0.71	0.5565
Error	22	359636.8016	16347.1273		
Corrected Total	25	394442.2242			

R-Square	Coeff Var	Root MSE	AUC Mean
0.088240	42.46705	127.8559	301.0708

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Treatment	3	34805.42260	11601.80753	0.71	0.5565

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Treatment	3	34805.42260	11601.80753	0.71	0.5565

A similar result is obtained using just the first period. There is not an effect due to the treatment (p-value = 0.557). The crossover design is more suitable because the variability in response from individual patients is reduced by having each patient respond to each of the four treatments.

3)

16.1

$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{1i} x_{2i} + \beta_6 x_{1i} x_{3i} + \beta_7 x_{1i} x_{4i} + \epsilon_i$ for $i = 1, \ldots, 24$

x_1 = covariate

$x_2 = \begin{cases} 1 & \text{if Treatment 2 is applied} \\ 0 & \text{if Otherwise} \end{cases}$

$x_3 = \begin{cases} 1 & \text{if Treatment 3 is applied} \\ 0 & \text{if Otherwise} \end{cases}$

$x_4 = \begin{cases} 1 & \text{if Treatment 4 is applied} \\ 0 & \text{if Otherwise} \end{cases}$

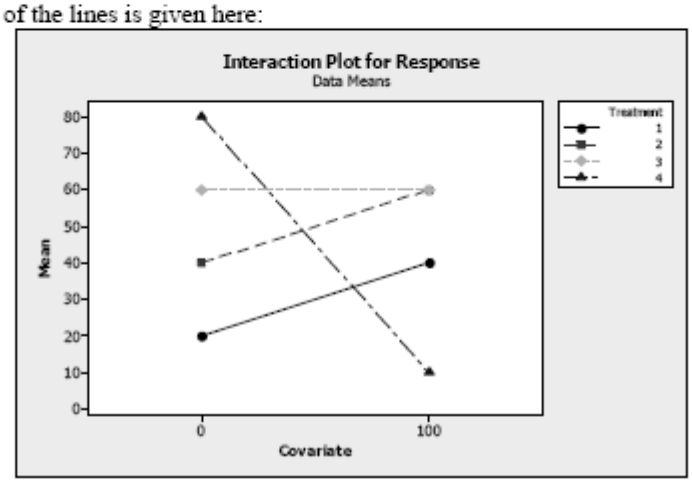
The coefficients are identified through the mean response under each treatment

Treatment	Mean Response
1	$\beta_0 + \beta_1 x_1$
2	$(\beta_0 + \beta_2) + (\beta_1 + \beta_5) x_1$
3	$(\beta_0 + \beta_3) + (\beta_1 + \beta_6) x_1$
4	$(\beta_0 + \beta_4) + (\beta_1 + \beta_7) x_1$

16.2

a. At least one of β_5, β_6, β_7 is not zero

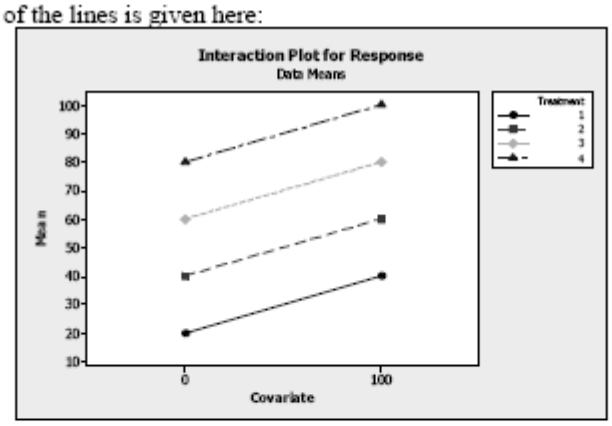
A plot of the lines is given here:



3)continued …

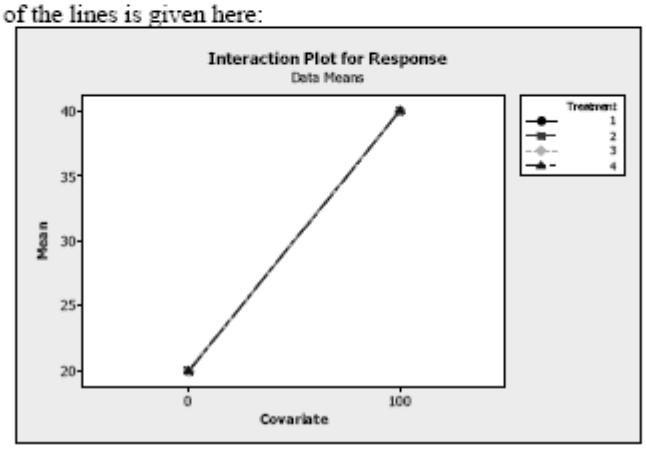
b. $\beta_5 = \beta_6 = \beta_7 = 0$ but at least one of β_2, β_3, β_4 is not zero

A plot of the lines is given here:



c. $\beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$

A plot of the lines is given here:



4)

16.5

a. $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{1i} x_{2i} + \beta_5 x_{1i} x_{3i} + \epsilon_i$ for $i = 1,\ldots,30$

x_1 = Number of Cigarettes

$$x_2 = \begin{cases} 1 & \text{if Treatment I is applied} \\ 0 & \text{if Otherwise} \end{cases}$$

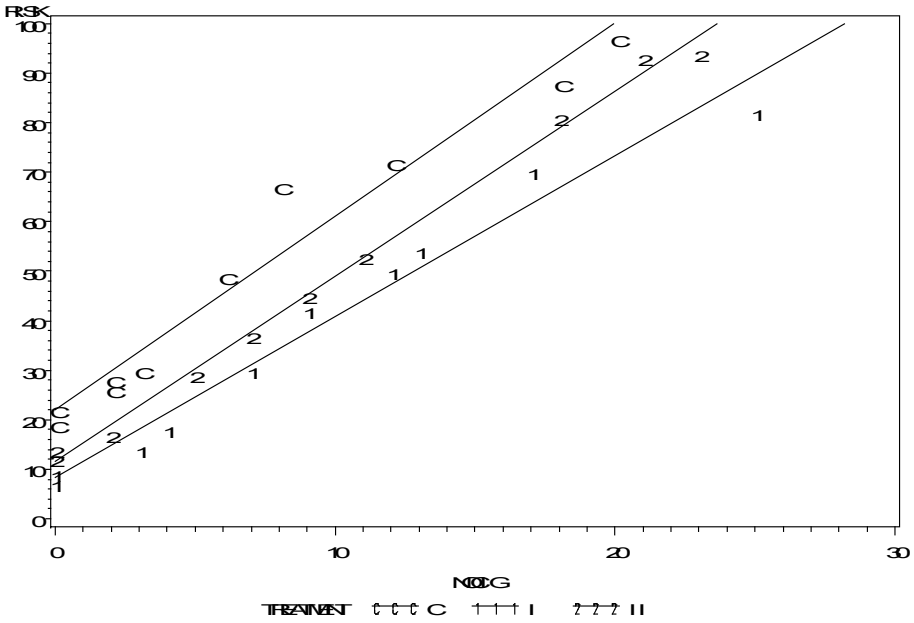
$$x_3 = \begin{cases} 1 & \text{if Treatment II is applied} \\ 0 & \text{if Otherwise} \end{cases}$$

The coefficients are identified through the mean response under each treatment

Treatment	Mean Response
C	$\beta_0 + \beta_1 x_1$
I	$(\beta_0 + \beta_2) + (\beta_1 + \beta_4) x_1$
II	$(\beta_0 + \beta_3) + (\beta_1 + \beta_5) x_1$

```
symbol1 v='C' i=rl c=black;
symbol2 v='1' i=rl c=black;
symbol3 v='2' i=rl c=black;

proc gplot data=hyper;
plot risk*nocig=treatment;
run;
```



It appears there is a relationship between risk and number of cigarettes smoked. The slopes look the same. Parallelism appears to be a reasonable assumption to make.

4)continued …

```
proc glm data=hyper;
class treatment;
model risk = treatment nocig treatment*nocig; ** test for same slope **;
run;
```

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TREATMENT	2	469.53955	234.76978	11.57	0.0003
NOCIG	1	22164.04588	22164.04588	1092.45	<.0001
NOCIG*TREATMENT	2	127.08382	63.54191	3.13	0.0619 ←

With p-value = 0.0619, insufficient evidence to indicate that the lines are not parallel.

```
proc glm data=hyper;
class treatment;
model risk = treatment nocig;
run;
```

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TREATMENT	2	1820.11191	910.05595	38.54	<.0001 ←
NOCIG	1	22390.49589	22390.49589	948.13	<.0001

Strong evidence the adjusted mean ratings are not all the same for the three treatments.

The covariate spans the space for all treatment levels so extrapolation – not concern.

3 ANCOVA assumptions okay –
- linear b/w response & covariate,
- same slope,
- covariate free of treatment influence

Examining diagnostic plots – residuals appear to be normal and of equal variance.

5)

16.7

a. $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{1i} x_{2i} + \beta_5 x_{1i} x_{3i} + \epsilon_i$ for $i = 1,\ldots,30$

x_1 = average monthly sales for the 12 months prior to promotion

$$x_2 = \begin{cases} 1 & \text{if Promotion B is used} \\ 0 & \text{if Otherwise} \end{cases}$$

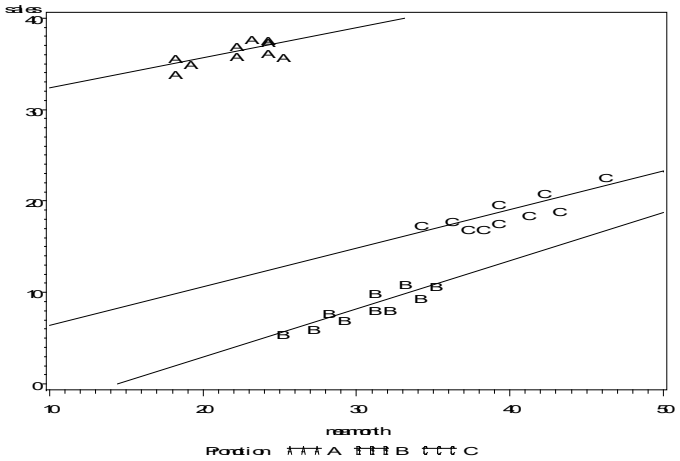
$$x_3 = \begin{cases} 1 & \text{if Promotion C is used} \\ 0 & \text{if Otherwise} \end{cases}$$

The coefficients are identified through the mean response under each treatment

Promotion	Mean Response
A	$\beta_0 + \beta_1 x_1$
B	$(\beta_0 + \beta_2) + (\beta_1 + \beta_4) x_1$
C	$(\beta_0 + \beta_3) + (\beta_1 + \beta_5) x_1$

```
symbol1 v='A' i=rl c=black;
symbol2 v='B' i=rl c=black;
symbol3 v='C' i=rl c=black;

proc gplot data=marketing;
plot sales*meanmonth=promotion;
run;
```



The slopes appear to be the same under the three promotions. The extrapolation problem appears to be an issue due to the non-overlap of the covariate under differing groups.

```
proc glm data=marketing;
class promotion;
model sales = promotion meanmonth promotion*meanmonth; ** test for same slope **;
run;
```

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Promotion	2	82.13933826	41.06966913	41.13	<.0001
meanmonth	1	46.15362167	46.15362167	46.22	<.0001
meanmonth*Promotion	2	1.50329749	0.75164875	0.75	0.4818 ←

… insufficient evidence to indicate that the lines are not parallel.

5)continued …

```
proc glm data=marketing;
class promotion;
model sales = promotion meanmonth;
run;
```

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Promotion	2	2942.452285	1471.226143	1502.03	<.0001 ←
meanmonth	1	51.467134	51.467134	52.54	<.0001

Strong evidence the adjusted mean sales are not all the same for the three promotions.

3 ANCOVA assumptions okay –
linear b/w response & covariate,
same slope,
covariate free of treatment influence

Examining diagnostic plots – residuals appear to be normal and of equal variance.

However, Extrapolation Problem exists …

---

a. Yes. The covariate covers a different space for the three variables so the assumptions are not wholly satisfied. The extrapolation problem is a major problem.

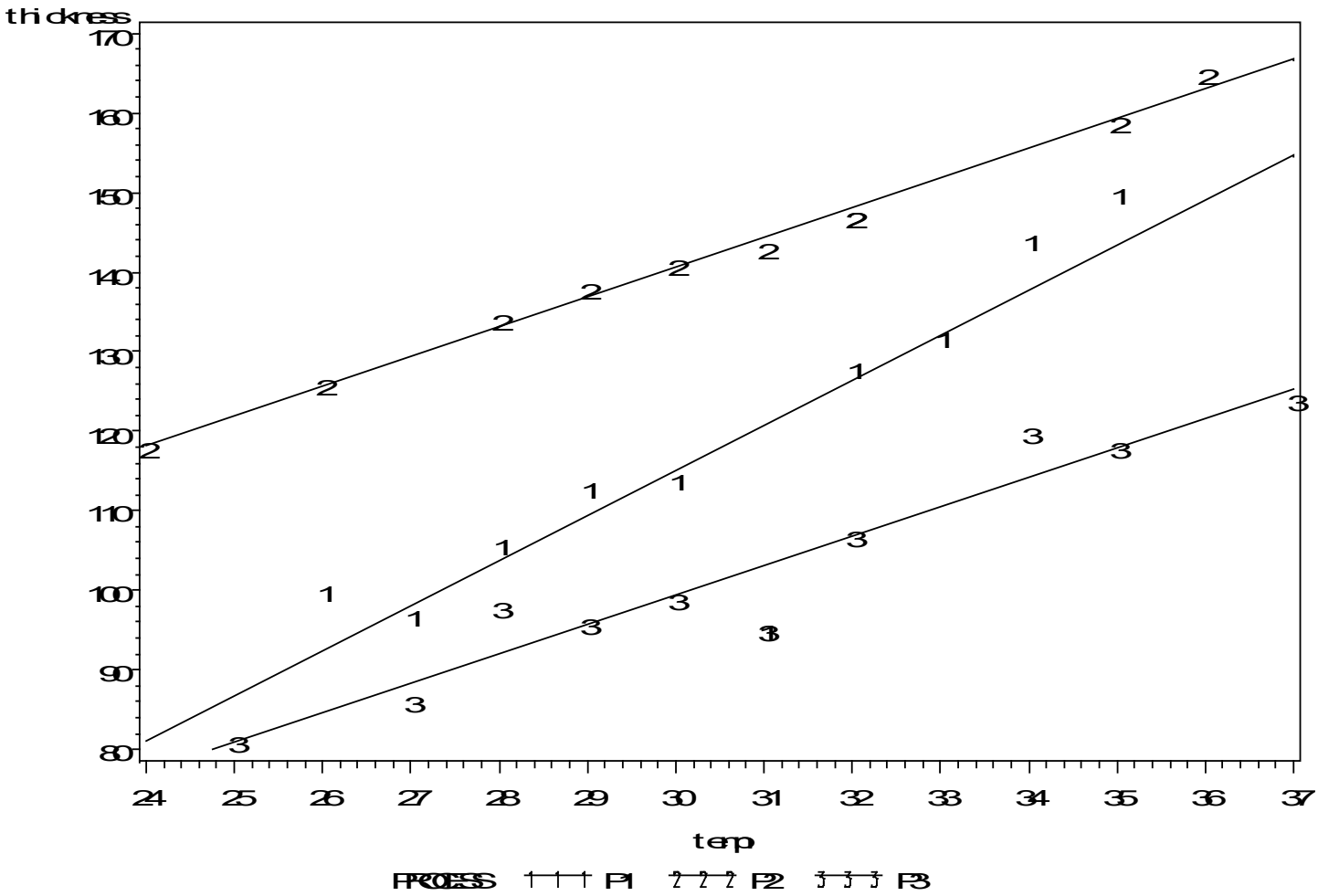
b. To avoid this problem, it would have been a good idea to block on the covariate x (average monthly sales in the previous 12 months) and fit a randomized complete block design. This would cause the treatments (promotions) to be randomly applied to the whole range of covariates.

6)

```
symbol1 v='1' i=rl c=black;
symbol2 v='2' i=rl c=black;
symbol3 v='3' i=rl c=black;

proc gplot data=coating;
plot thickness*temp=process;
run;

proc glm data=coating;
class process;
model thickness = process temp process*temp;
run;
```



6)continued …

c. $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{1i} x_{2i} + \beta_5 x_{1i} x_{3i} + \epsilon_i ;\ i = 1,\ldots,30$

$x_1 =$ temperature

$$x_2 = \begin{cases} 1 & \text{if Process 2 is applied} \\ 0 & \text{if Otherwise} \end{cases}$$

$$x_3 = \begin{cases} 1 & \text{if Process 3 is applied} \\ 0 & \text{if Otherwise} \end{cases}$$

The coefficients are identified through the mean response under each treatment

Treatment	Mean Response
P1	$\beta_0 + \beta_1 x_1$
P2	$(\beta_0 + \beta_2) + (\beta_1 + \beta_4) x_1$
P3	$(\beta_0 + \beta_3) + (\beta_1 + \beta_5) x_1$

6)continued …

The GLM Procedure

Class Level Information

Class	Levels	Values
PROCESS	3	P1 P2 P3

Number of Observations Read 30
Number of Observations Used 30

The SAS System

The GLM Procedure

Dependent Variable: thickness

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	14051.66564	2810.33313	68.92	<.0001
Error	24	978.63436	40.77643		
Corrected Total	29	15030.30000			

R-Square	Coeff Var	Root MSE	thickness Mean
0.934889	5.290508	6.385643	120.7000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
PROCESS	2	7879.400000	3939.700000	96.62	<.0001
temp	1	5937.794200	5937.794200	145.62	<.0001
temp*PROCESS	2	234.471440	117.235720	2.88	0.0760

Source	DF	Type III SS	Mean Square	F Value	Pr > F
PROCESS	2	373.192483	186.596242	4.58	0.0207
temp	1	6172.080221	6172.080221	151.36	<.0001
temp*PROCESS	2	234.471440	117.235720	2.88	0.0760 ←

… insufficient evidence to indicate that the lines are not parallel.

6) continued …

The covariate spans the space for all treatment levels so extrapolation – not concern.

3 ANCOVA assumptions okay –
linear b/w response & covariate,
same slope,
covariate free of treatment influence

Examining diagnostic plots – residuals appear to be normal and of equal variance.

```
proc glm data=coating;
class process;
model thickness = process temp;
lsmeans process / cl;
run;
```

Source	DF	Type III SS	Mean Square	F Value	Pr > F
PROCESS	2	8686.821896	4343.410948	93.09	<.0001 ←
temp	1	5937.794200	5937.794200	127.26	<.0001

Strong evidence: adjusted mean thicknesses are not all the same for the three processes.

PROCESS	thickness LSMEAN
P1	118.040085
P2	142.780597
P3	101.279318

PROCESS	thickness LSMEAN	95% Confidence Limits	
P1	118.040085	113.599977	122.480193
P2	142.780597	138.336968	147.224225
P3	101.279318	96.834590	105.724046

```
proc glm data=coating;
class process;
model thickness = process;
run;
```

Source	DF	Type III SS	Mean Square	F Value	Pr > F
PROCESS	2	7879.400000	3939.700000	14.88	<.0001 ←

At least one process significantly differs in thickness from the others [without incorporating temperature in the model]. The conclusions match in this case, but this may not be true at all times because adjusting the treatment means for the covariate effect on the response variable may alter the conclusions about the differences in the treatment means.