Lecture 32

Analysis of Covariance II

STAT 512

Spring 2011

Background Reading

KNNL: Chapter 22

Topic Overview

- ANCOVA with multiple factors
- ANCOVA with Blocking
- Use of Differences

ANCOVA Model

$$Y_{ij} = \mu + \alpha_i + \beta\left(X_{ij} - \bar{X}_{\bullet\bullet}\right) + \varepsilon_{ij}$$

- $\varepsilon_{ij} \overset{iid}{\sim} N\left(0, \sigma^2\right)$ and $\sum \alpha_i = 0$ (or $\alpha_a = 0$)
- Centering covariate $\left(X_{ij} - \bar{X}_{\bullet\bullet}\right)$ means that μ will represent an overall mean
- Can extend this model to multiple factors or multiple covariates (or both)

Diagnostics

- Examine the data and residuals (check the three standard assumptions)
- Check the same-slope assumption (plots, interaction term)
- Look for outliers that are influential

Diagnostics / Remedial Measures

- Examine variances (standard deviations). Look at MSE for models run separately on each treatment group (use a BY statement in PROC REG or GLM)

- Transform if needed, use Box-Cox to assist in finding an appropriate transformation

Two-Way ANCOVA Model

- Multiple Factors:

$$Y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma\left(X_{ijk} - \bar{X}_{\bullet\bullet\bullet}\right) + \varepsilon_{ij}$$

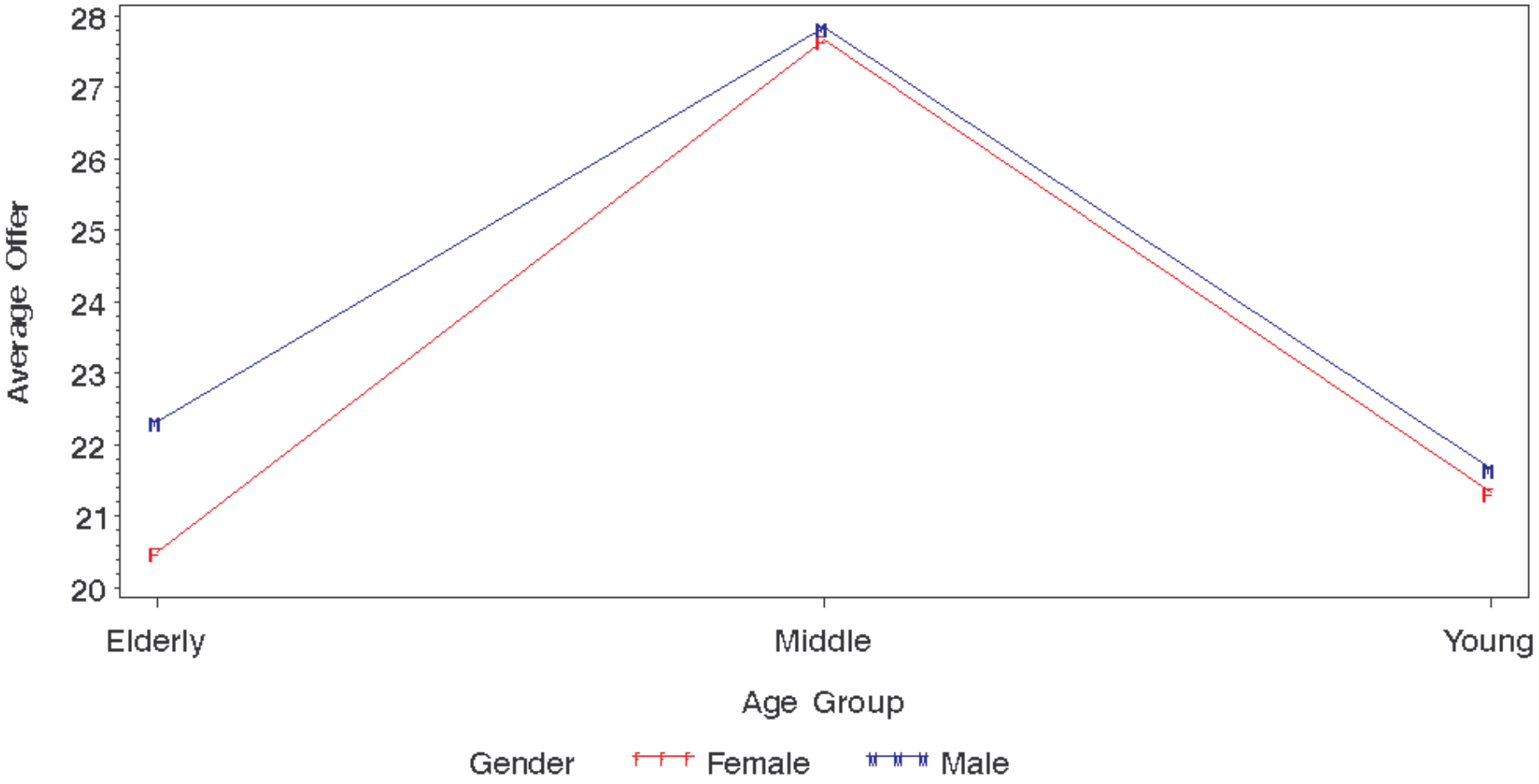
Two-Way ANCOVA Model (2)

- Basic idea remains the same. For each treatment combination we have a linear regression in which the slopes are the same, but the intercepts may differ.
- We make comparisons using least-square means, with the covariates set to their mean values (so that any differences will not be due to the level of the covariates)

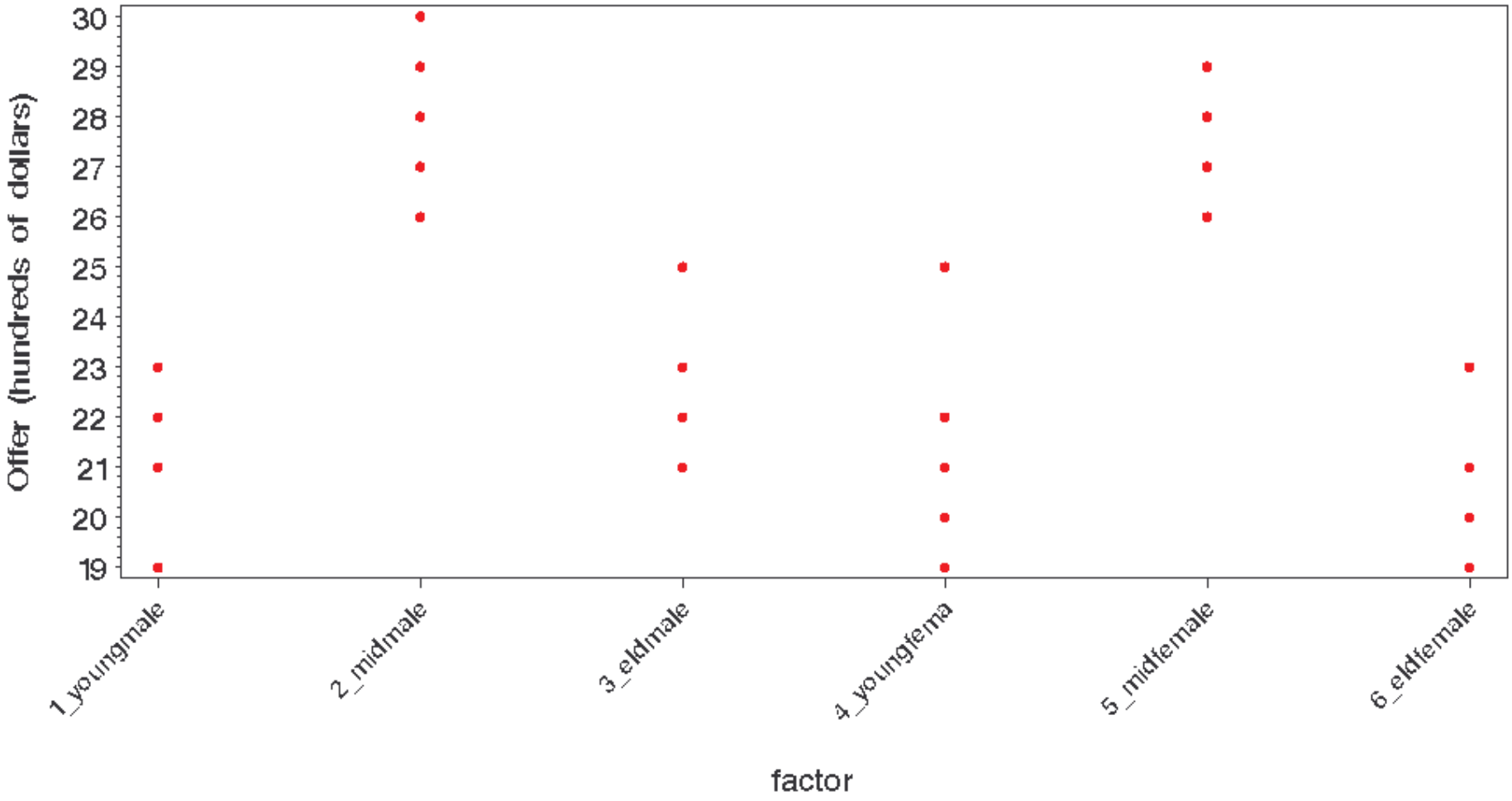
Two-way ANCOVA example

- Cash Offers Example (cashoffers_ancova.sas)
- Y is offer made by a dealer on a used car
- Factor 1 is the age of person selling the car (young, middle, elderly)
- Factor 2 is gender of the person selling the car (male, female)
- Covariate is overall sales volume for the dealer

Interaction Plot with No Covariate



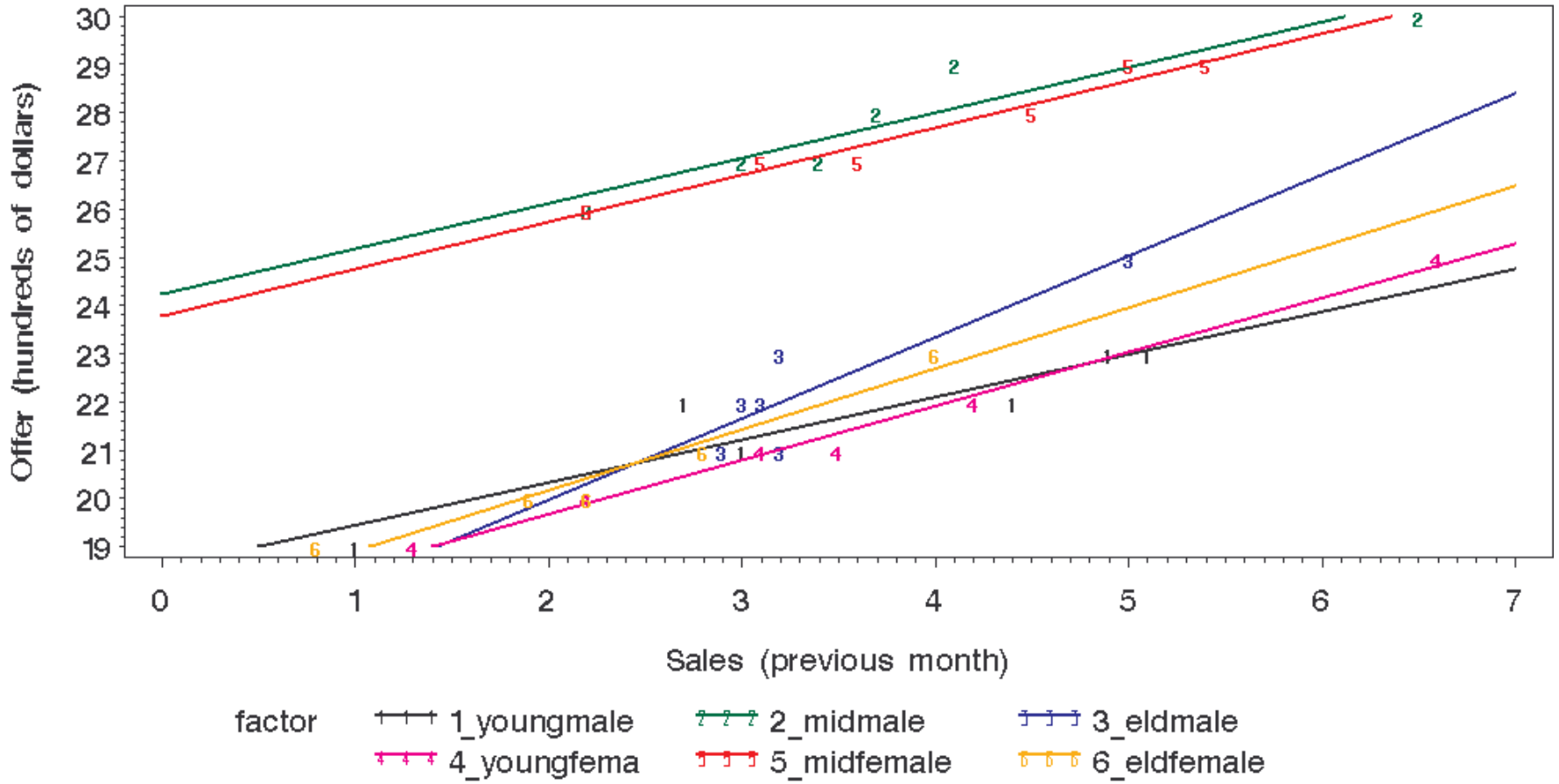
Plot of Offers against Factor Combinations w/o Covariate



Plots w/o Covariate

- Plots (and previous analysis) with simple two-way ANOVA showed differences in that middle-aged appeared to do better than the other two groups; no interaction or gender differences.

Plot of Offers vs Covariate



Covariate

- Clearly is a relationship to the covariate; higher sales means higher offers

- Plot suggests a slight interaction; maybe something different going on in the elderly-male group.

- Let's look at the ANCOVA

SAS Code for ANCOVA

```
proc glm data=cash;
  class age gender;
  model offer=sales age|gender;
  output out=diag p=pred r=resid;
  lsmeans age gender
          /tdiff pdiff cl adjust=tukey;
run;
```

- Note: May include interaction with SALES to check equality of slopes assumption.

Output

Source	DF	SS	MS	F Value	Pr > F
sales	1	63.37	63.37	221.58	<.0001
age	2	232.49	116.24	406.45	<.0001
gender	1	1.55	1.55	5.40	0.0273
age*gender	2	0.19	0.10	0.34	0.7142
Error	29	8.3	0.286		
Total	35	398.9			

- Gender effect shows up once covariate included in model (size of effect is very small, but it is significant)

LSMEANS / Multiple Comparisons

#	age	LSMEAN	95% Conf	Limits
1	Elderly	22.03	21.70	22.35
2	Middle	27.24	26.91	27.56
3	Young	21.40	21.09	21.72

i/j	1	2	3
1		<.0001	0.0241
2	<.0001		<.0001
3	0.0241	<.0001	

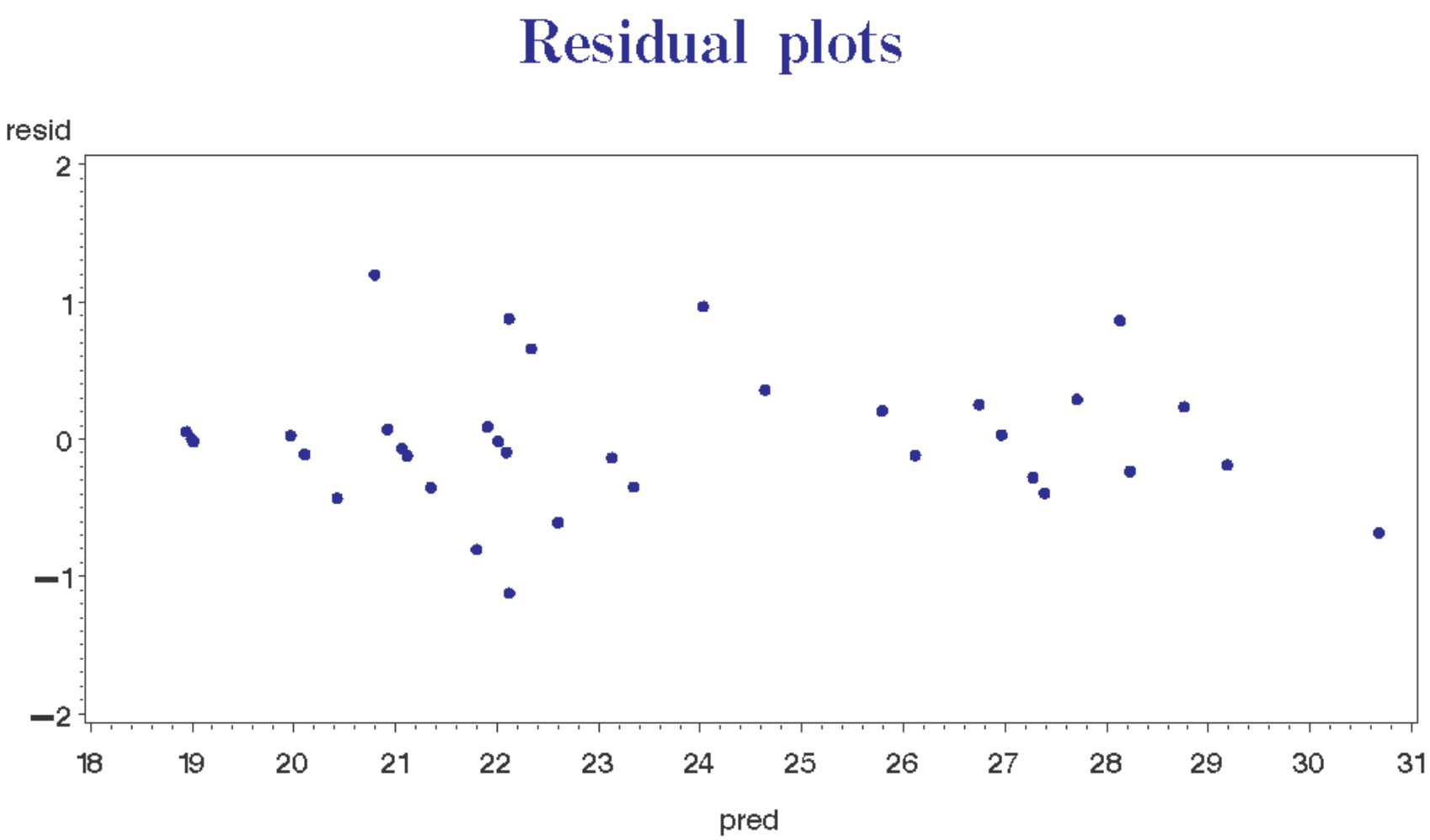
- Note: Different results here too!!! Effect we saw for age is still there, and additionally elderly get significantly better offer than young.

Multiple Comparisons (2)

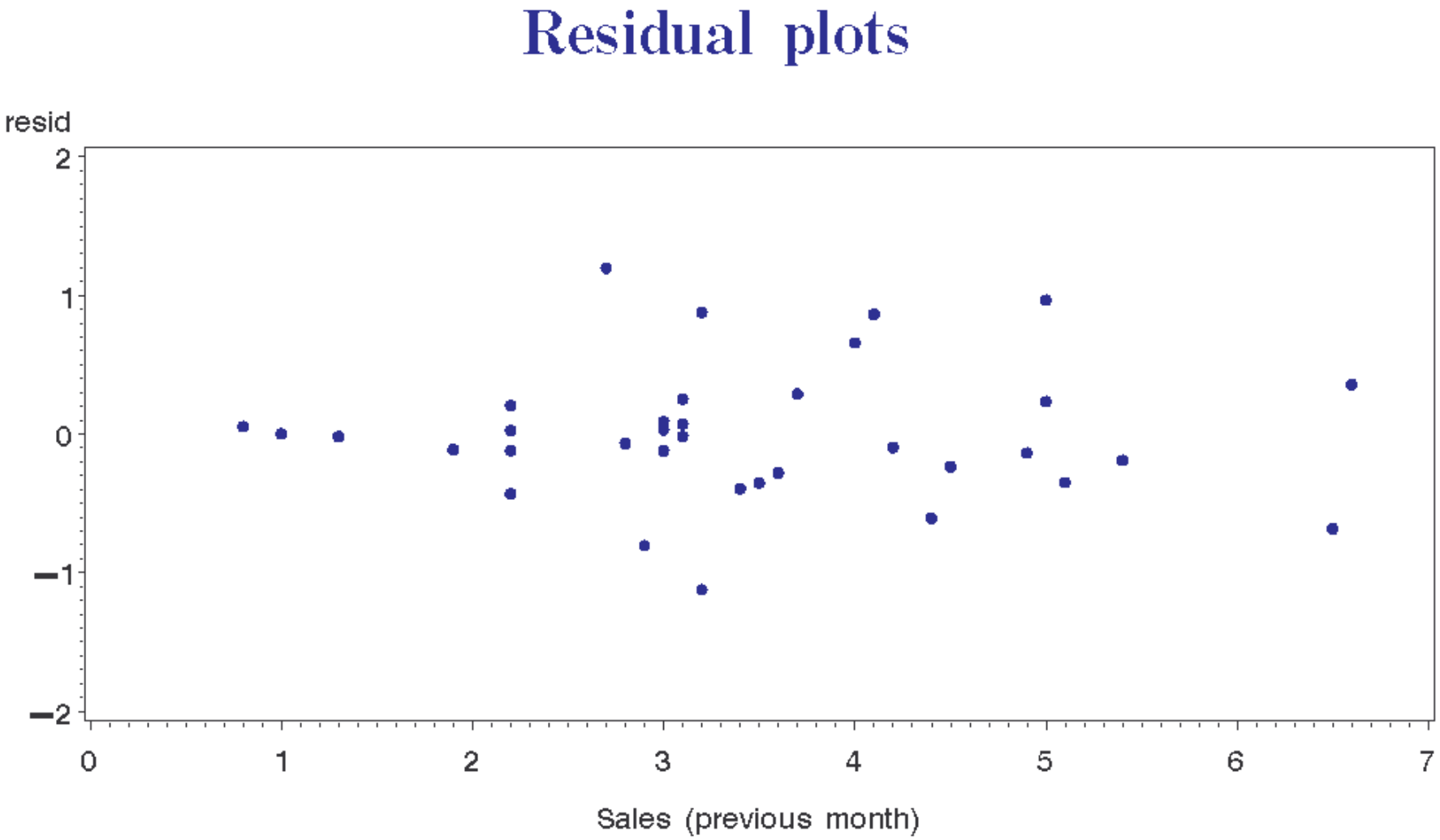
gender	offer LSMEAN	Pr > |t|
Female	23.3464846	0.0273
Male	23.7646265	

- Gender effect is significant (Male > Female) but the size of the effect is quite small (only half the size of the difference between elderly/young)

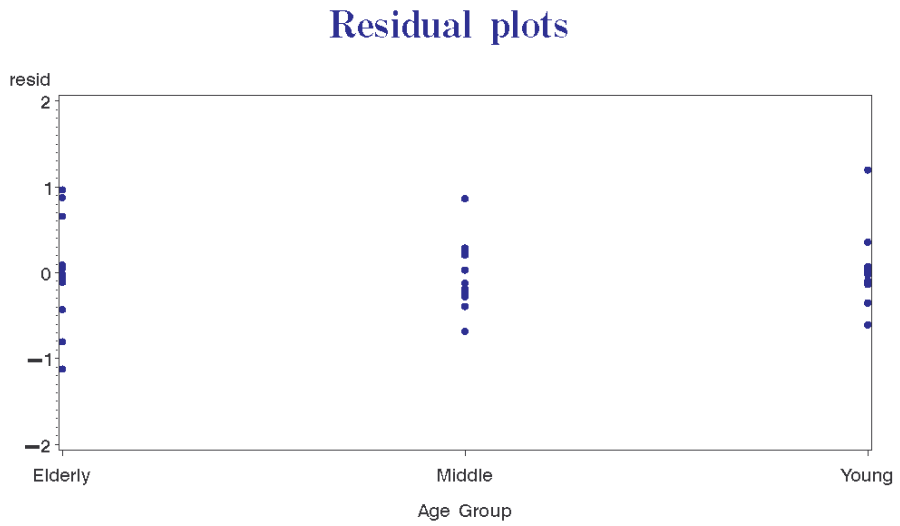
Residual Analysis (1)

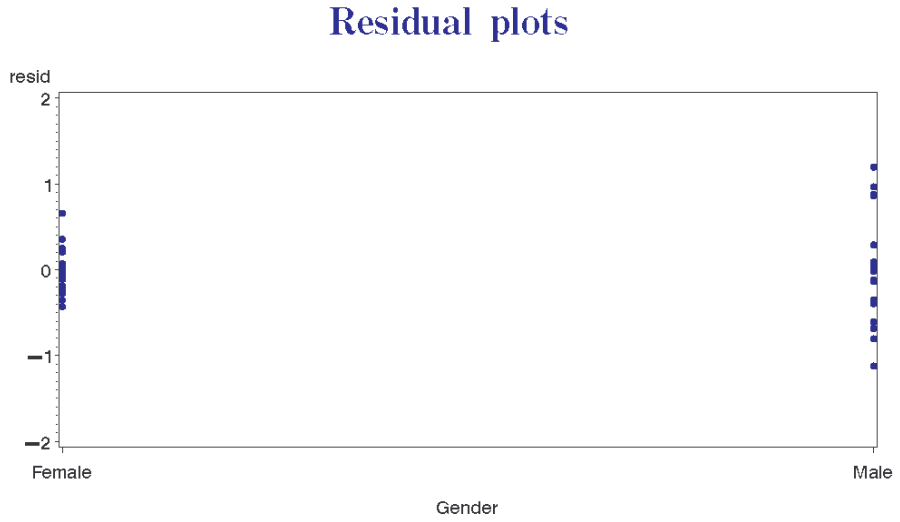


Residual Analysis (2)

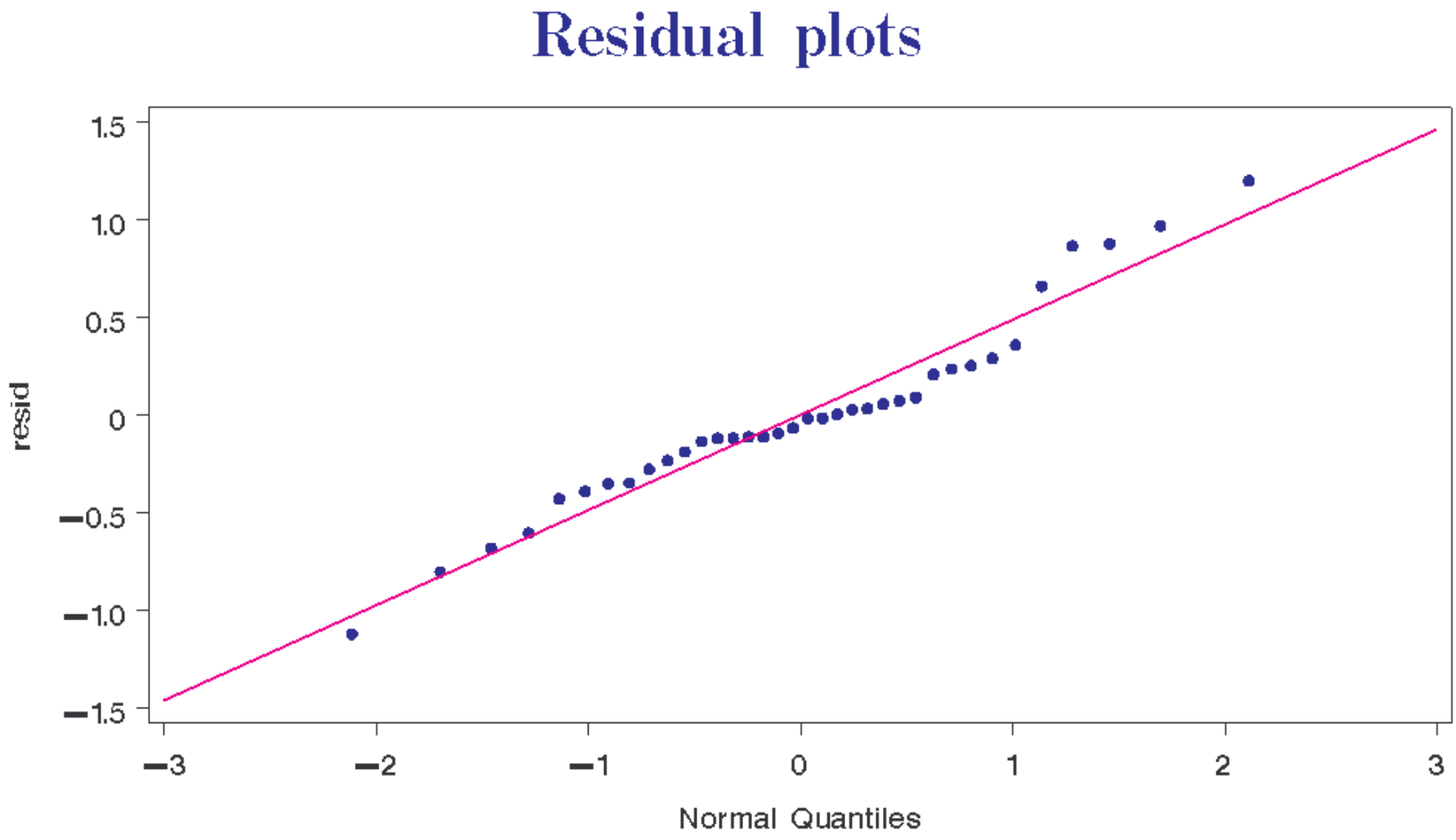


Residual Analysis (3)





Residual Analysis (4)



Testing Equality of Slopes

- For one-factor, one covariate: Test interaction of factor and covariate

- For two-factor, one covariate: Test interaction of covariate and two-way factor interaction

sales	1	54.83	54.83	213.06	<.0001
age	2	30.49	15.25	59.24	<.0001
gender	1	0.03	0.03	0.10	0.7554
age*gender	2	0.68	0.34	1.32	0.2847
sales*age*gender	5	2.12	0.43	1.65	0.1863

Example (Auditor Training)

- See auditor_ancova.sas for data and coding
- Firm testing the effectiveness of three training methods (home-study, local training, or national training).
- 30 data points; Response variable is a proficiency score (higher = better)

Example (Auditor Training)

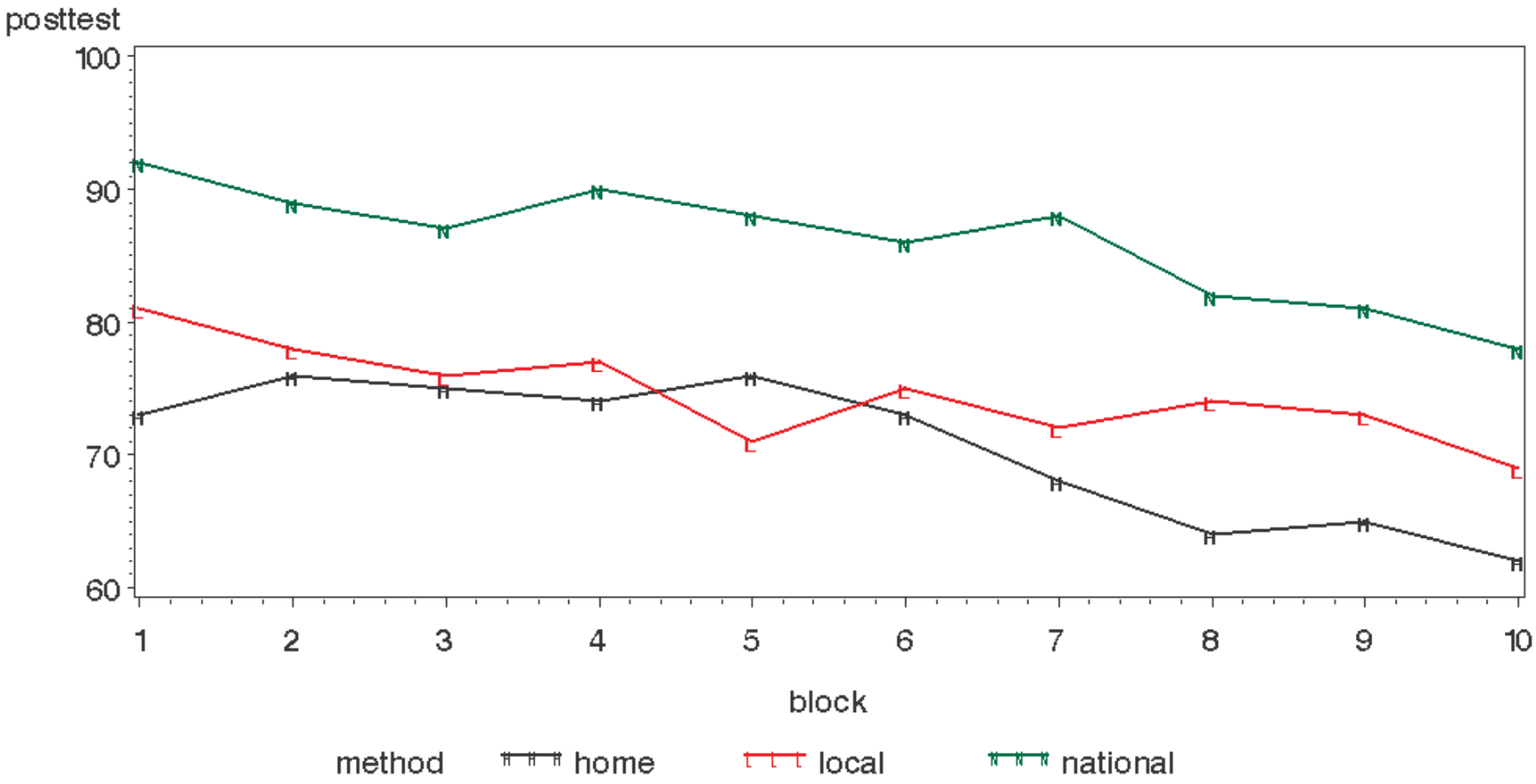
- Block on length of time since graduation (Block #1 is the most recent graduates)
- Additionally, a pretest score is available. This will be used as a covariate.
- Previous results (without the covariate) indicated national training was to be preferred. Local (group) training was 2nd best, not much better than home.

Simple ANOVA

Source	DF	SS	MS	F Value	Pr > F
Method	2	1295	647.5	32.04	<.0001
Error	27	546	20.2		
Total	29	1841			

GRP	Mean	N	method
A	86.100	10	national
B	74.600	10	local
B	70.600	10	home

ANOVA with Blocks



ANOVA with Blocks

Source	DF	SS	MS	F Value	Pr > F
block	9	433	48.2	7.72	0.0001
method	2	1295	647.5	103.75	<.0001
Error	18	113	6.24		
Total	29	1841			

method	LSMEAN	GRP
national	86.1	A
local	74.6	B
home	70.6	C

*All p-values < 0.05

ANCOVA with Blocks

Source	DF	Squares	Mean Square	F Value	Pr > F
Model	12	1728.367335	144.030611	21.80	<.0001
Error	17	112.332665	6.607804		
Corrected Total	29	1840.700000			

R-Square	Coeff Var	Root MSE	posttest Mean
0.938973	3.334066	2.570565	77.10000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
pretest	1	344.395741	344.395741	52.12	<.0001
block	9	91.787592	10.198621	1.54	0.2107
method	2	1292.184002	646.092001	97.78	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
pretest	1	0.000669	0.000669	0.00	0.9921
block	9	74.377119	8.264124	1.25	0.3298
method	2	1292.184002	646.092001	97.78	<.0001

ANCOVA with Blocks

- Type I SS : Pretest is significant alone, but block is not significant in a model with pretest (but we saw previously that it was significant when pretest was not in the model).

- Type III SS : Pretest and block are not significant when other factors in model.

- Method is significant when all other factors are in the model.

ANCOVA without Blocks

Source	DF	SS	MS	F Value	Pr > F
pretest	1	359	359	49.99	<.0001
method	2	1310	655	91.18	<.0001
Error	26	187	7.1		
Total	29	1841			

method	LSMEAN	GRP
national	86.14	A
local	74.61	B
home	70.54	C

*All p-values < 0.05

Summary of Results

- In this case it turns out that you always will identify the national training as the best.

- Notice the slight differences in each analysis – we don't actually need both concomitant variables (either use the block, or use the pretest, the information is about the same).

Block vs Pretest

Source	DF	SS	MS	F Value	Pr > F
BLOCK	9	3052	339	40.70	<.0001
Error	20	167	8.3		
Total	29	3219			

- 94% of pretest is explained by block – these variables are essentially performing identical functions in the analysis

Blocking vs. ANCOVA (1)

- Sometimes researchers have a choice between
  - CRD with covariance analysis (ANCOVA)
  - RCBD with blocks formed by means of the concomitant variable

Blocking vs. ANCOVA (2)

- If regression between response and concomitant variable is linear, about equally efficient. If not linear – RCBD more effective.
- RCBD are free of assumptions about the nature of relationship between concomitant (blocking) variable and response. ANCOVA assumes linear relationship w/equal slopes between groups.
- RCBD may require more df for blocking variable and thus leave less for the error.

Use of Differences

- For a posttest/prettest study, there are two possible options for analysis:
  - ANCOVA with posttest as response and prettest as a covariate
  - ANOVA using difference (posttest-prettest) as the response.
- If the slope parameter $\beta=1$, then these analyses are essentially equivalent.
- If slope parameter is not near 1, then ANCOVA may be more effective than the use of differences.

Use of Differences

- For the cracker example from lecture 31, $\hat{\beta} = 0.9$ and also 1 is in the 95% CI. Using the difference of current-previous period sales as the response and conducting one-way ANOVA should be sufficient.
- For the auditor example $\hat{\beta} = 0.33$ and 1 is not in the 95% CI. Better to use ANCOVA.
- See KNNL section 22.5 for more details.

Upcoming...

- Multi-Factor ANOVA (Chapter 24)