Analysis of Covariance ANCOVA

Assoc Prof Dr Norsa'adah Bachok
MBBS(Aust), MComMed(Mal), PhD(Mal)
Unit of Biostatistics and Research Methodology
School of Medical Sciences
Universiti Sains Malaysia

Objectives

- Understand types of data needed in ANCOVA
- Be able to conduct the ANCOVA analysis
- Able to interpret and present the ANCOVA results

Analysis of Covariance

- An extension of ANOVA in which main effects and interactions are assessed on DV scores after the DV has been adjusted for by the DV's relationship with one or more Covariates (CVs).
- Basic requirements
  - 1 DV (I, R) – continuous
  - 1 IV (N, O) – categorical
  - 1 CV (I, R) – continuous

Basic requirements of data

- Minimum number of CVs that are uncorrelated with each other.
- CVs should also be uncorrelated with the IVs (e.g. the CV should be collected before treatment is given) in order to avoid diminishing the relationship between the IV(s) and DV.
- A covariate is a variable that is related to the DV, which you can't manipulate, but you want to account for it's relationship with the DV.

ANCOVA

- Reduce systematic bias & within groups error.
- The influence of covariate on the dependent variable is statistically controlled.
- Control variable that not the main interest but have effect on the dependent variable.
- Confirmatory study? Exploratory?

Purposes

- Reduce bias (adjust estimates of population means on one or more variables)
- Increase power (reduce experimental error)
  - The magnitude of the reduction in the error term is related to the size of the correlation between the covariate and the dependent variable in the design
  - The larger the correlation, the greater is the reduction in the error term

$$F_{\text{ANCOVA}} = \frac{\text{Between Group variation indep var} + \text{Between Group variation COV}}{\text{Within Group variation individu diff} + \text{Within Group variation COV}}$$

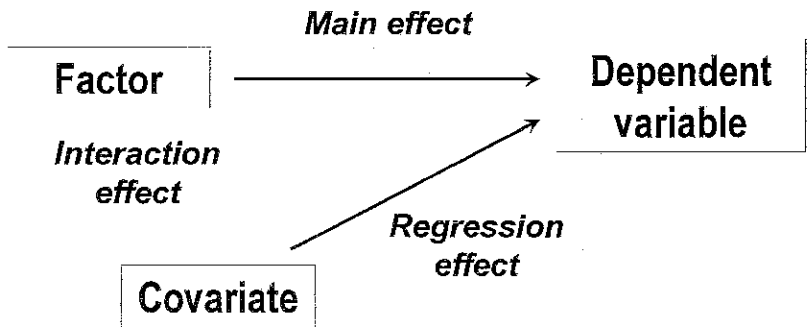
ANCOVA

- Evaluates whether the population means on the dependent variable, adjusted for differences on the covariate, differ across levels of a factor.
- Suitable for a study to examine the effect of treatment but do no have equal baseline parameters.
- If the factor has more than 2 levels, need to do post hoc test.

ANCOVA

- Variance: in Dependent Variable
  - The linear combination that is examined in ANCOVA is the Y score that is adjusted for the effects of the covariates
- Covariance: between Dependent Variable & Covariate(s)
  - Examine the proportion of shared variance between the adjusted Y score and the Independent Variable.
- Ratio: Between Groups/ Within Groups
  - The ratio of between-groups variance over within-groups variance that determine the p value.
- *F-test*
  - The significance test in ANCOVA
  - If significant, 2 or more means statistically differ after controlling for the effect of 1 or more covariates

ANCOVA

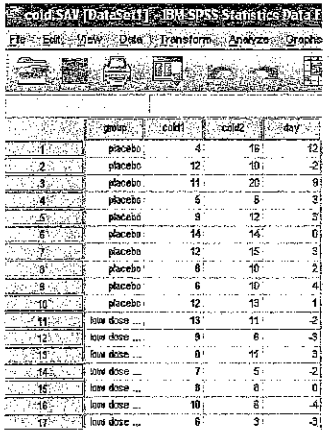


1. Dependent variable – continuous numerical
2. Study Factor – categorical with two or more levels
3. Covariate – continuous numerical

Steps ANCOVA

Step 1: Data exploration: descriptive summary by mean/%
Step 2: Bivariate exploration by one-way ANOVA / t test
Step 3: Determine main effect in the model
(no variable selection)
Step 4: Checking interactions
Step 5: Checking model assumptions
Normality of residuals
Equal Variance
Linearity of relationship
Overall Model Fitness
Step 6: Post Hoc test
Step 7: Interpretation & presentation of results

Cold file



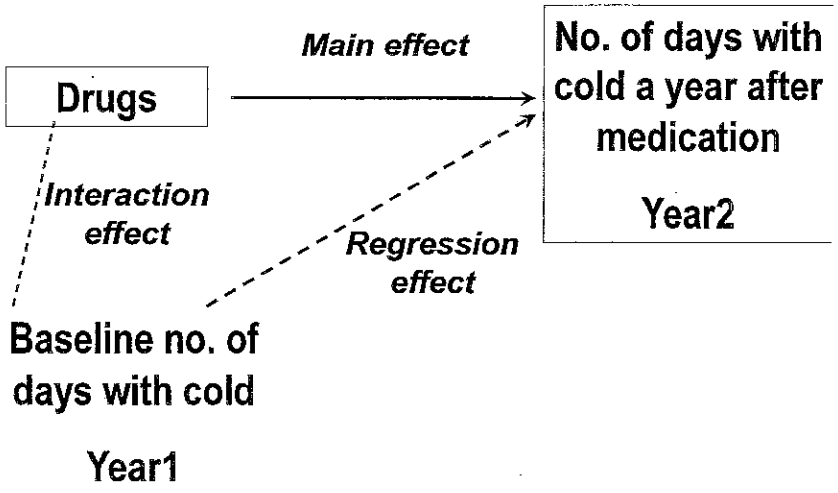
	group	cold1	cold2	day
1	placebo	4	16	12
2	placebo	12	10	-2
3	placebo	11	20	9
4	placebo	5	8	3
5	placebo	9	12	3
6	placebo	14	14	0
7	placebo	12	15	3
8	placebo	8	10	2
9	placebo	6	10	4
10	placebo	12	13	1
11	low dose ...	13	11	-2
12	low dose ...	9	6	-3
13	low dose ...	0	15	3
14	low dose ...	7	5	-2
15	low dose ...	8	8	0
16	low dose ...	10	6	-4
17	low dose ...	6	3	-3

- RQ: Are people taking vitamin C have less number of days with cold symptoms?
- Group (factor)
  - 1 = placebo
  - 2 = low dose Vitamin C
  - 3 = high dose Vitamin C
- Cold1: Predays (covariate)- number of days with cold symptoms in 1st yr prior treatment
- Cold2:Postdays (dependent)- number of days with cold symptoms in 2nd yr with treatment

Research question

- Mean differences
  - Does the number of days of cold symptoms differ among groups, assuming no a priori differences among groups?
- Relationship between variables
  - Is there a relationship between the amount of Vitamin C taken and the number of days of cold symptoms, holding constant the number of days with cold symptoms in the year prior to treatment?

Example



Step 1: Data exploration: descriptive summary by mean/%

Step 2: Bivariate exploration: One-way ANOVA

Descriptives

cold2

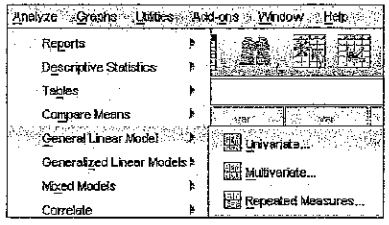
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
placebo	10	12.80	3.584	1.133	10.24	15.36	8	20
low dose vitamin C	10	7.70	3.860	1.221	4.94	10.46	3	15
high dose vitamin C	10	10.00	4.422	1.398	6.84	13.16	3	16
Total	30	10.17	4.379	.800	8.53	11.80	3	20

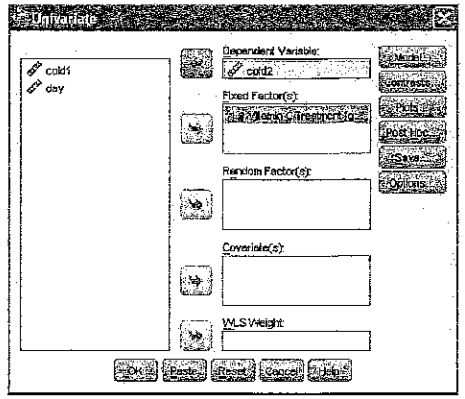
ANOVA

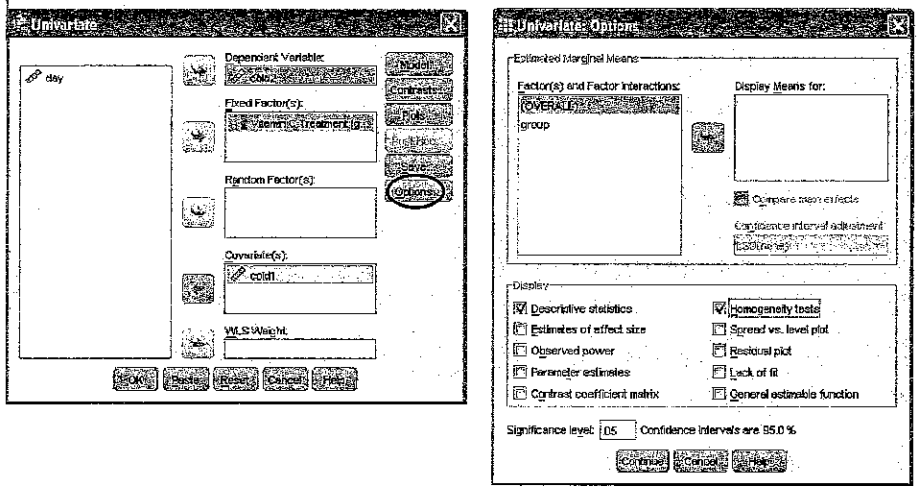
cold2

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	130.467	2	65.233	4.137	.027
Within Groups	425.700	27	15.767		
Total	556.167	29			

Step 3: Determine main effect in the model (no variable selection)







Main effect

Tests of Between-Subjects Effects

Dependent Variable:no. of days with cold symptoms 2nd yr

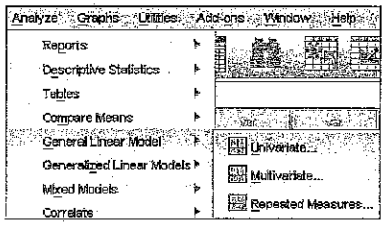
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	156.385[a]	3	52.128	3.390	.033
Intercept	149.030	1	149.030	9.692	.004
cold1	25.918	1	25.918	1.686	.206
group	140.913	2	70.457	4.582	.020
Error	399.782	26	15.376		
Total	3657.000	30			
Corrected Total	556.167	29			

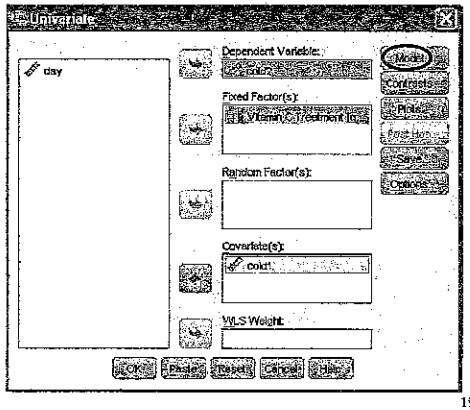
a. R Squared = .281 (Adjusted R Squared = .198)

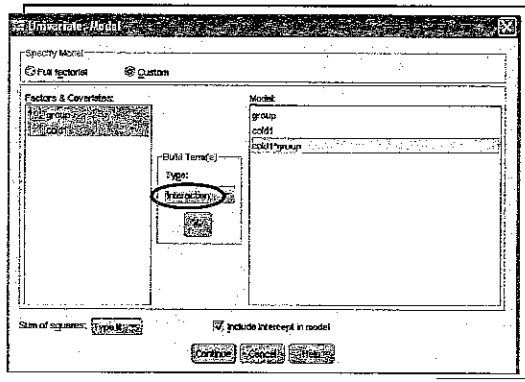
There is a significant effect of vitamin C on cold symptoms after adjusting the number of days with cold symptoms prior starting the medication

Step 4: Checking interaction effect

2-way interaction effect of baseline days with cold symptoms and group







Tests of Between-Subjects Effects

Dependent Variable:no. of days with cold symptoms 2nd yr

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	162.524[a]	5	32.505	1.982	.118
Intercept	151.281	1	151.281	9.223	.006
group	23.388	2	11.694	.713	.500
cold1	27.101	1	27.101	1.652	.211
group * cold1	6.139	2	3.070	.187	.831
Error	393.642	24	16.402		
Total	3657.000	30			
Corrected Total	556.167	29			

a. R Squared = .292 (Adjusted R Squared = .145)

Interaction effect of group*cold p value 0.831

Step 5: Checking model assumptions

1. Normality of distribution of residuals
   - For dependent variables on the covariate and in the same group
   - If the scores for the covariate alone are normally distributed the ANCOVA is robust to this assumption
2. Homogeneity of variance
   - Levene's test
   - Scatter plot residual vs predicted
3. Linearity: Covariate is linearly related to the dependent variable within all levels of the factor
   - Preferably r>0.3
   - Scatter plot residual vs covariate

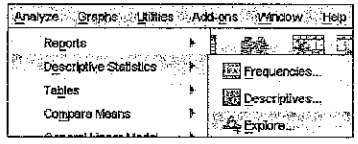
5. Homogeneity of regression slopes: the relationship of dependent variable to the covariate in each group should be the same
   - Regression slopes (β_1) for $Y=\beta_0+\beta_1X$ should be the same for all treatment levels
6. Overall Model Fitness: scatter plot of residual vs predicted

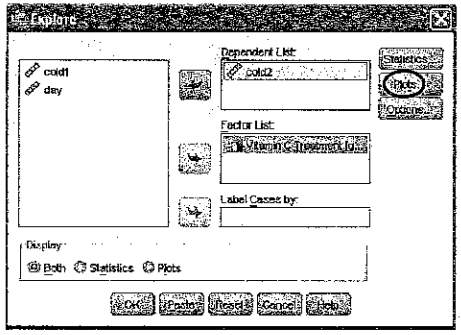
Assumptions

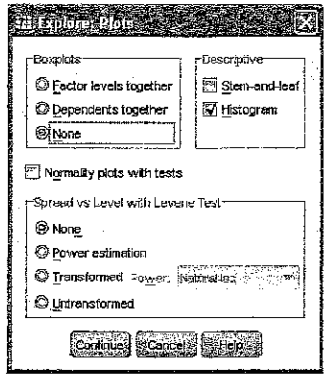
- Random samples
- Observations are independent
- Reliability of covariate

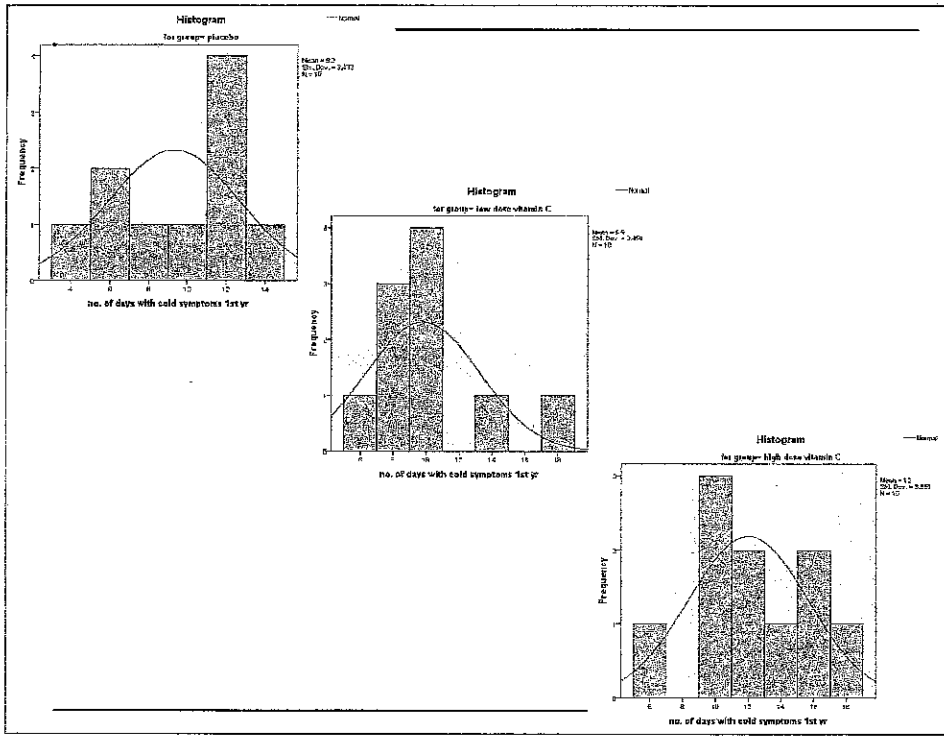
} By study design and method

1. Assumption: Normality of distribution

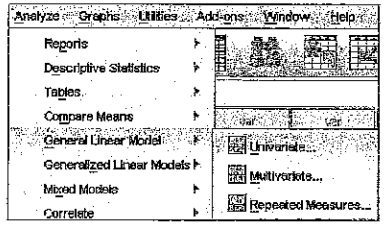


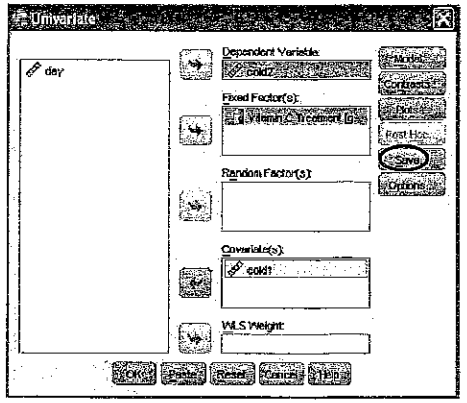


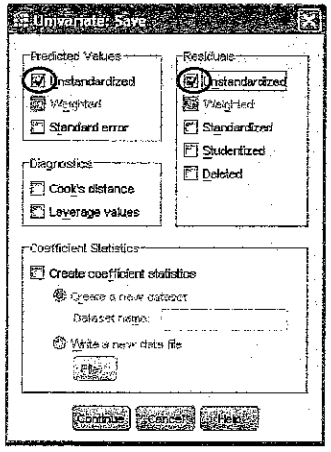




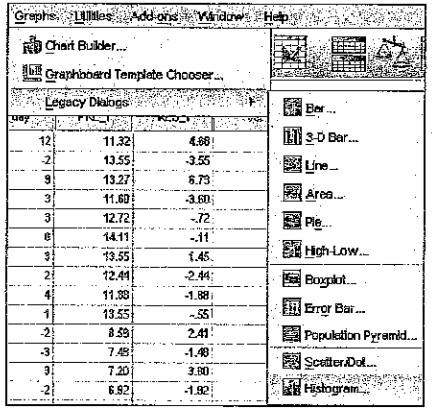
Residual, predicted values



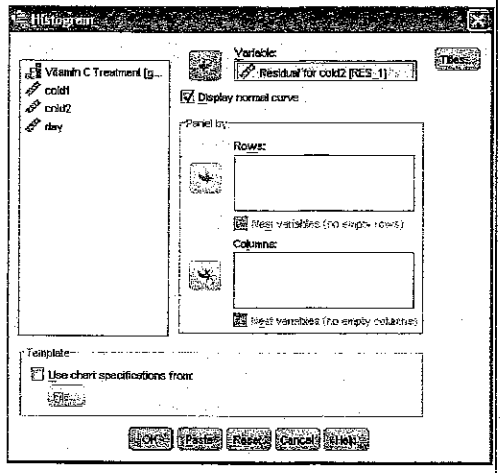




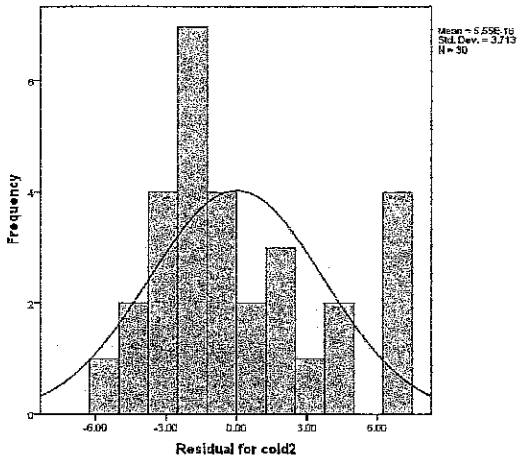
PRE_1	RES_1
11.32	4.68
13.55	-3.55
13.27	6.73
11.60	-3.60
12.72	-.72
14.11	-.11
13.55	1.45
12.44	-2.44
11.88	-1.88
13.55	-.55
8.59	2.41
7.48	-1.48
7.20	3.80
6.92	-1.92
7.20	.80
7.76	-1.76
9.64	-3.64
7.76	7.24
9.99	-.99
7.48	-4.48
8.33	3.67
11.11	-2.11
9.72	-4.72
9.16	-6.16
11.67	.33
9.44	-3.44
10.56	1.44
10.84	-1.84
9.44	6.56
9.72	8.28



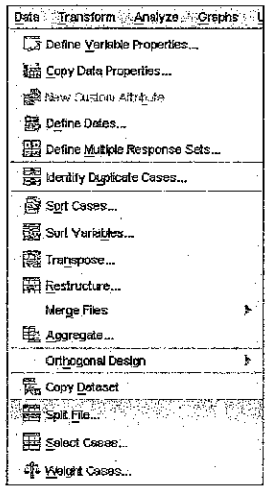
day	PRE_1	RES_1	var
12	11.32	4.68	
2	13.55	-3.55	
9	13.27	6.73	
3	11.60	-3.60	
3	12.72	-.72	
8	14.11	-.11	
3	13.55	1.45	
2	12.44	-2.44	
4	11.88	-1.88	
1	13.55	-.55	
-2	8.59	2.41	
-3	7.48	-1.48	
9	7.20	3.80	
-2	6.92	-1.92	



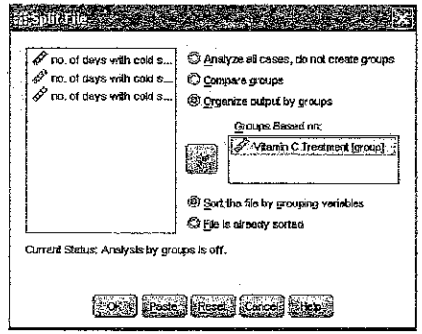
Distribution of residuals



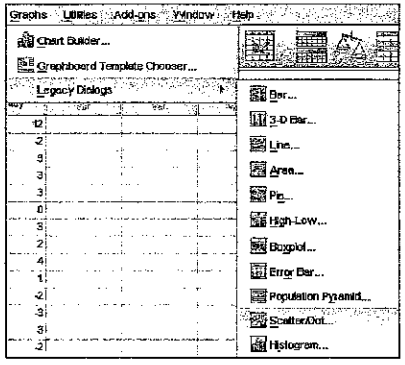
2. Linearity
4. Homogeneity of regression slopes

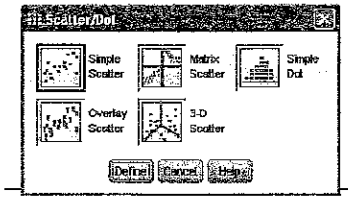


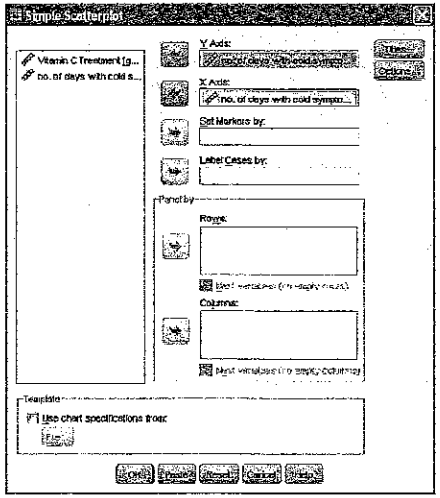
Assumption need to be checked for each group of independent variable



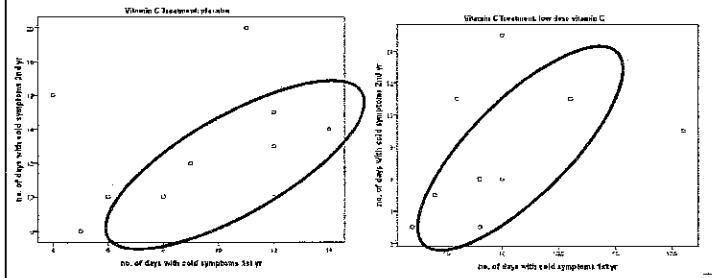
Scatter plot



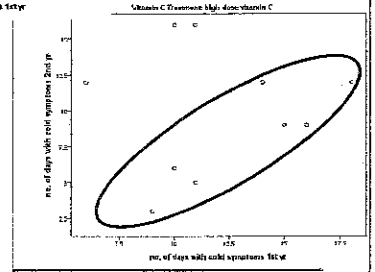




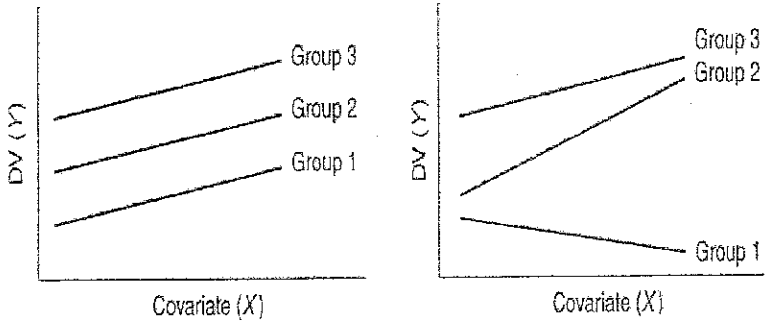
Assumption: Linearity
Homogeneity of regression slopes



- Slopes between covariate and DV are similar across groups
- Indicates no interaction between Dependent Variable and covariate
- If slopes differ, covariate behaves differently depending on which group (i.e., heterogeneity of regression)
- When slopes are similar (what we desire), Y is adjusted similarly across groups



Homogeneity of Regression



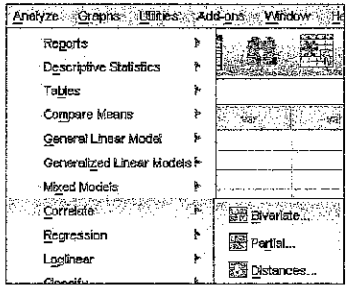
(a) Homogeneity of regression (slopes)

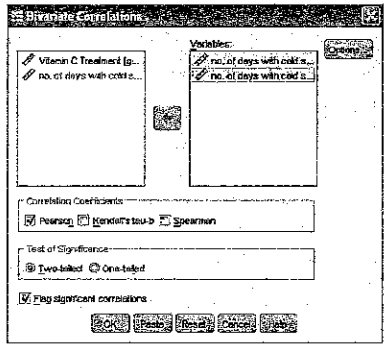
(b) Heterogeneity of regression (slopes)

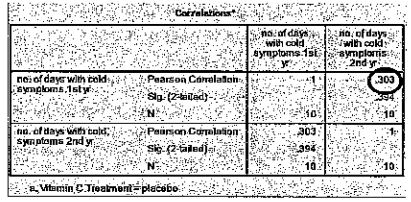
Covariate

- Want high r with DV
- Low with other covariates
- Is statistically controlled in an adjusted DV
- If covariate correlates with IV → heterogeneity of regression (violation of an assumption)

Correlation





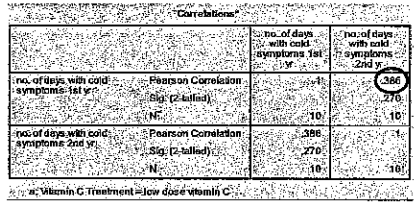


Correlations[a]

		no. of days with cold symptoms 1st yr	no. of days with cold symptoms 2nd yr
no. of days with cold symptoms 1st yr	Pearson Correlation	1	.303
	Sig. (2-tailed)	.	.394
	N	10	10
no. of days with cold symptoms 2nd yr	Pearson Correlation	.303	1
	Sig. (2-tailed)	.394	.
	N	10	10

a. Vitamin C Treatment = placebo

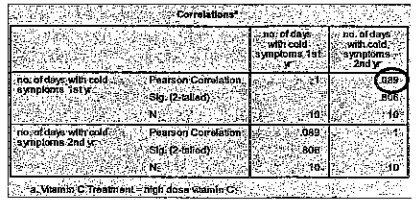
Comment on *r*?



Correlations[a]

		no. of days with cold symptoms 1st yr	no. of days with cold symptoms 2nd yr
no. of days with cold symptoms 1st yr	Pearson Correlation	1	.386
	Sig. (2-tailed)	.	.270
	N	10	10
no. of days with cold symptoms 2nd yr	Pearson Correlation	.386	1
	Sig. (2-tailed)	.270	.
	N	10	10

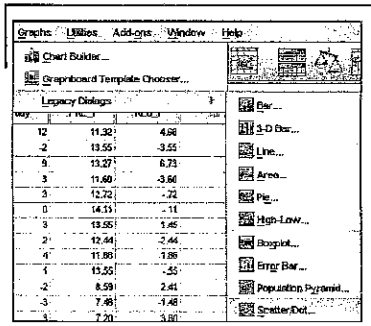
a. Vitamin C Treatment = low dose vitamin C

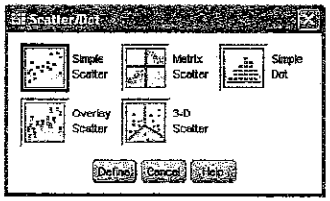


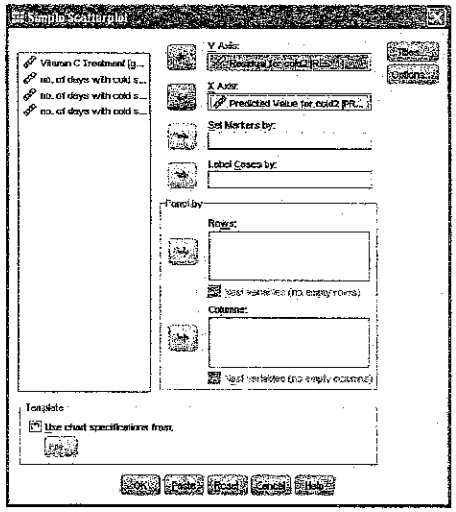
Correlations[a]

		no. of days with cold symptoms 1st yr	no. of days with cold symptoms 2nd yr
no. of days with cold symptoms 1st yr	Pearson Correlation	1	.089
	Sig. (2-tailed)	.	.806
	N	10	10
no. of days with cold symptoms 2nd yr	Pearson Correlation	.089	1
	Sig. (2-tailed)	.806	.
	N	10	10

a. Vitamin C Treatment = high dose vitamin C

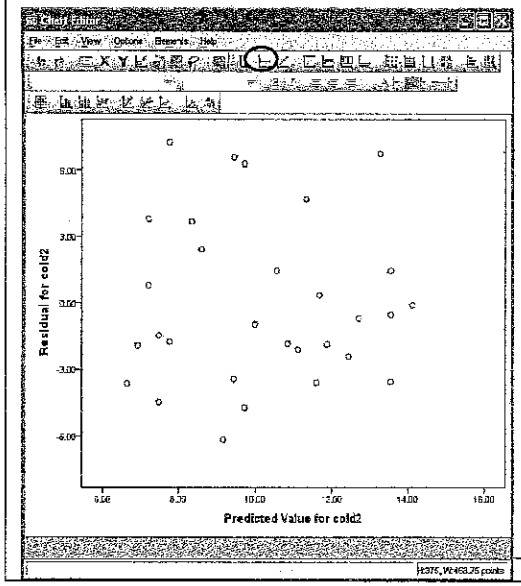


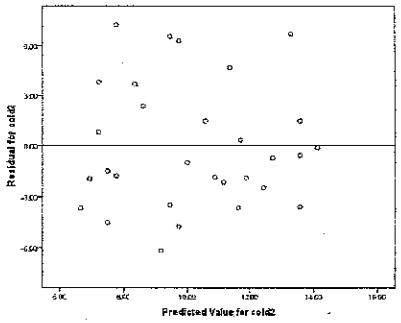




37

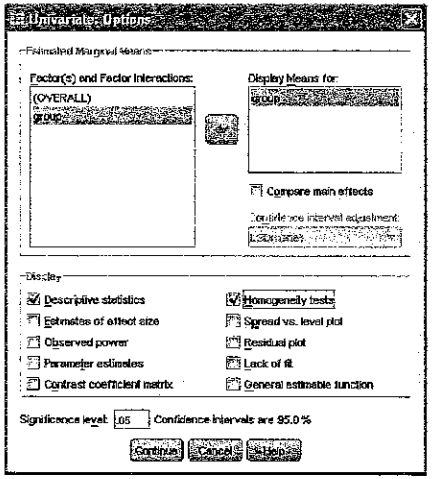
Residual vs predicted





38

2. Homogeneity of Covariance



Levene's Test of Equality of Error Variances[a]

Dependent Variable:no. of days with cold symptoms 2nd yr

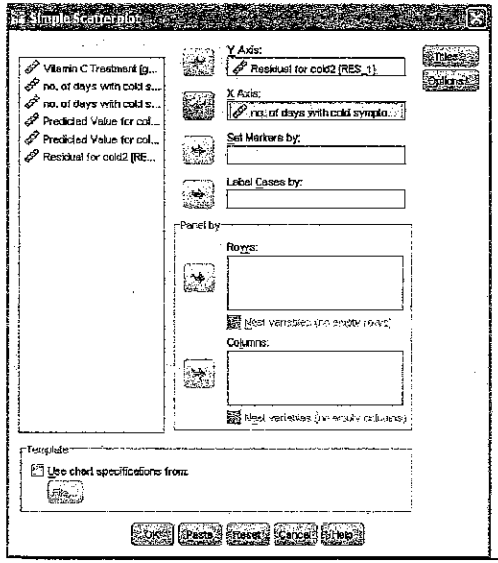
F	df1	df2	Sig.
.722	2	27	.495

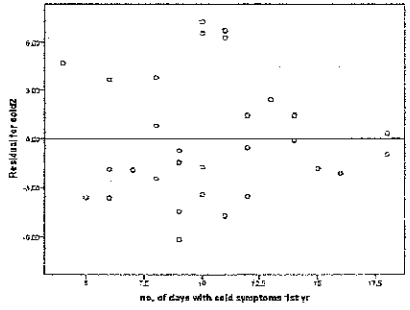
Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + cold1 + group

39

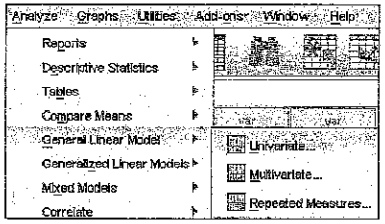
Residual vs covariate

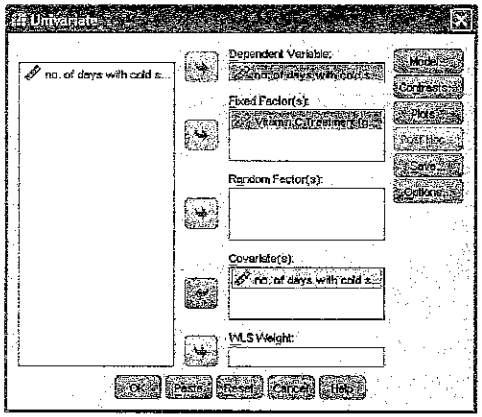




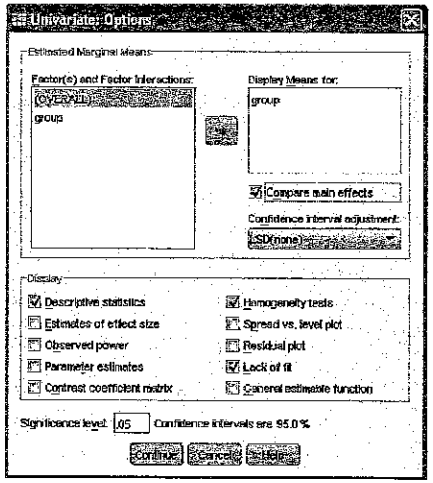
40

5. Fitness of model





Fitness of model



Lack of Fit Tests

Dependent Variable:no. of days with cold symptoms 2nd yr

Source	Sum of Squares	df	Mean Square	F	Sig.
Lack of Fit	227.115	19	11.953	.485	.901
Pure Error	172.667	7	24.667		

Step 6: Post Hoc Pairwise Comparisons

Pairwise Comparisons

Dependent Variable:no. of days with cold symptoms 2nd yr

(I) Vitamin C Treatment	(J) Vitamin C Treatment	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
placebo	low dose vitamin C	6.239*	1.757	.006	1.628	8.851
	high dose vitamin C	3.552	1.847	.065	-.244	7.349
low dose vitamin C	placebo	-6.239*	1.757	.006	-8.851	-1.628
	high dose vitamin C	-1.687	1.816	.362	-5.420	2.046
high dose vitamin C	placebo	-3.552	1.847	.065	-7.349	.244
	low dose vitamin C	1.687	1.816	.362	-2.046	5.420

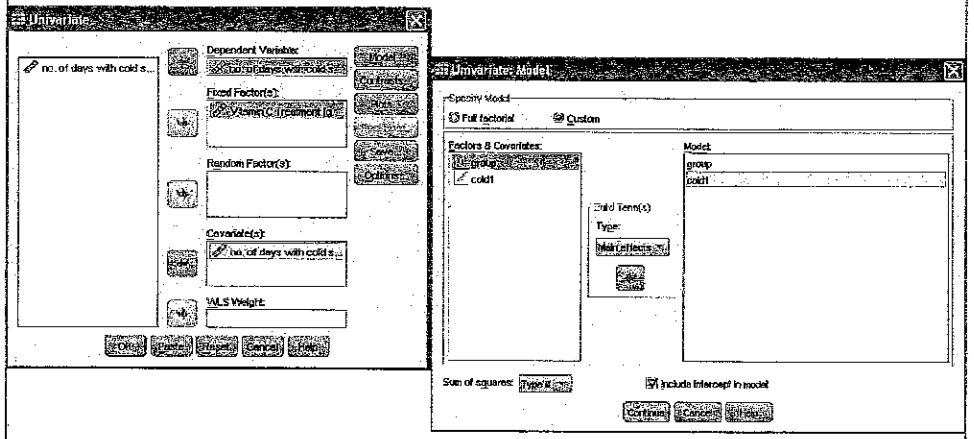
Based on estimated marginal means

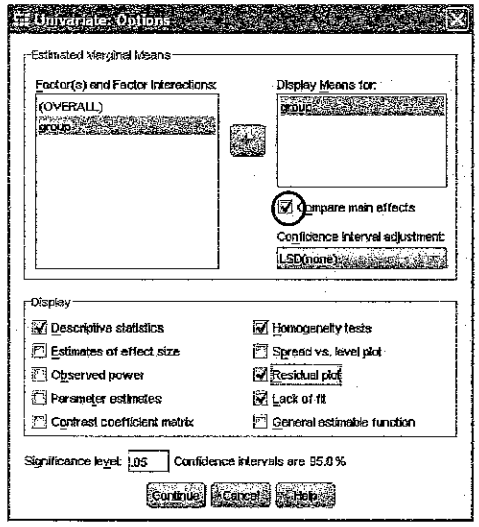
*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

There is no relationship between the amount of Vitamin C taken and the number of days of cold symptoms, holding constant the number of days with cold symptoms in the year prior to treatment.

Step 7: Interpretation & presentation of results





Descriptive Statistics

Dependent Variable:no. of days with cold symptoms 2nd yr

Vitamin C Treatment	Mean	Std. Deviation	N
placebo	12.80	3.584	10
low dose vitamin C	7.70	3.860	10
high dose vitamin C	10.00	4.422	10
Total	10.17	4.379	30

Descriptive statistics: mean no. with cold symptoms without adjustment of baselines

Tests of Between-Subjects Effects

Dependent Variable:no. of days with cold symptoms 2nd yr

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	156.385[a]	3	52.128	3.390	.033
Intercept	149.030	1	149.030	9.692	.004
group	140.913	2	70.457	4.582	.020
cold1	25.918	1	25.918	1.686	.206
Error	399.782	26	15.376		
Total	3657.000	30			
Corrected Total	556.167	29			

a. R Squared = .281 (Adjusted R Squared = .198)

The number of days of cold symptoms differ among groups when holding constant the number of days with cold symptoms in the prior to treatment

Estimates

Dependent Variable:no. of days with cold symptoms 2nd yr

Vitamin C Treatment	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
placebo	13.097[a]	1.261	10.505	15.689
low dose vitamin C	7.858[a]	1.246	5.297	10.419
high dose vitamin C	9.545[a]	1.289	6.896	12.194

a. Covariates appearing in the model are evaluated at the following values: no. of days with cold symptoms 1st yr = 10.37.

Adjusted descriptive statistics: after adjustment of baselines

Table 1. Adjusted mean of cold symptoms after adjustment with the number of days the year prior

Treatment group	Days with cold symptoms in second year		P value**
	Mean (95% CI)*	Adjusted Mean** (95% CI)	
Placebo	12.80 (10.24, 15.36)	13.10 (10.50, 15.69)	0.020
Low dose vit C	7.70 (4.94, 10.46)	7.86 (5.30, 10.42)	
High dose vit C	10.17 (8.53, 11.80)	9.54 (6.90, 12.19)	

*One –way ANOVA

**ANCOVA, results adjusted for baseline numbers with cold symptoms

Placebo vs low dose p=0.006,

Placebo vs high dose p=0.065,

Low vs high dose p=0.362 (post hoc analysis LSD)

ANCOVA

- Control for the effects due to the covariate and seeing if there are differences in groups
- ANCOVA is primarily concerned with reducing error variance/increasing power, when used appropriately for experimental setups.
- The difference between ANCOVA and MLinR is the language used to describe the results (mean differences vs. coefficients etc.), some different options for analysis because of the categorical variable (post hocs, effect size), and an additional assumption (homogeneity of regression)

Linear Models: a summary

	Y	Model effect(s)
One-way ANOVA	Y	A
Two-way ANOVA	Y	A, B
Two-way ANOVA (with inter)	Y	A, B, A*B
Simple Linear Regression	Y	X
Multiple Regression	Y	X1, X2, X3...
Multivariate Regression	Y1, Y2, Y3...	X1, X2, X3...
MANOVA	Y1, Y2, Y3...	A
ANCOVA	Y	A, B... X1, X2, X3...

Thank you

norsaadah@usm.my