WAGENINGEN UNIVERSITY **Biometris, ASN 1205**

Examination Statistics (MAT-24306)

Wednesday, May 2, 2012; 9:00 – 12:00 h

											0

program registration number

Name and first name

You can use the book by Ott & Longnecker, the Lecture Notes, a hand written summary (one A4, two-sided), and a calculator, but keep e.g. printed PowerPoint sheets and your cellphone in your bag. There are five questions with credits 26, 18, 12, 20 and 14, and with 10 bonus credits this makes a total of 100. Below, credits (per part of an exercise) are in parentheses.
You are free to answer the questions in English or in Dutch (or both).

Put your answers in the empty spaces indicated in this exam. If you use extra paper, put your name on the extra pages, and refer to these extra pages in this exam.

Question 1 (26) Taste of cheese

Several chemical procedures take place during the maturing of Cheddar cheese that affect the taste of the cheese. A researcher wants to investigate whether taste depends on concentrations of chemical elements Acetic, H_2S and Lactic in the final product. From each of 20 cheese samples a score (y) for the taste is determined (as an average of a number of subjective scores of "tasters"). Also, Acetic (x_1), H_2S (x_2) and Lactic (x_3) concentrations are measured. The data are shown on the right.

y	x_1	x_2	x_3
5.6	4.5	3.1	0.8
7.2	5.1	5.0	1.5
24.8	5.3	5.4	1.5
47.2	5.7	7.4	1.8
24.8	4.6	3.8	0.9
52.9	5.6	7.6	1.0
20.4	5.8	8.7	1.2
35.1	6.0	7.9	1.7
23.8	4.8	3.8	1.2
3.7	5.2	4.1	1.5
19.7	5.7	6.1	1.6
20.5	6.4	7.9	1.9
20.1	4.4	2.9	1.0
9.6	5.2	4.9	1.3
22.5	6.1	6.7	1.5
25.8	6.3	9.5	1.7
7.7	4.7	3.9	1.1
14.2	5.4	4.7	1.4
27.9	5.2	6.1	1.6
40.3	5.4	9.0	1.9

The researcher starts with the model:

$$\underline{y}_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \underline{\varepsilon}_i,$$

assuming independence, normality and equal variance for the random error terms $\underline{\varepsilon}_i$, where i = 1…20 for the cheese samples.

a (2) Give the practical interpretation (meaning) of coefficient β_3.

Practical interpretation :

The increase in expected score for taste when concentration Lactic is increased with one unit, while concentrations Acetic and H_2S remain the same.
[One point when the last part about concentrations Acetic and H_2S is missing]

Some SPSS output is shown on page 5.

b (2) Calculate the proportion of variation explained by the model, adjusted for degrees of freedom (show your calculations).

Proportion of variance explained (with calculations):

Adjusted $R^2 = 1 - MSE / MS(Total) = 1 - MS(Residual) / MS(Total) \quad 1 - 117.193 / (3410.9 / 19) = 0.34.$

(can be found directly in output: 0.347 because of rounding).

c1 (3) Test whether the model has any predictive power at all. Give:
(1) the outcome of the test statistic,
(2) the distribution of the test statistic under the null hypothesis of no predictive power (including any degrees of freedom),
(3) the appropriate P-value and your conclusion (in words).

Outcome test statistic: $F = 4.368$

Distribution test statistic under the null hypothesis(including any degrees of freedom):

F-distribution with degrees of freedom $df1 = 3$, $df2 = 16$

P-value: 0.02

Conclusion (in words):

We have proven (since P-value = 0.02 < 0.05) that at least one of the three prediction variables Acetic (x_1), H_2S (x_2) *or Lactic* (x_3) *has predictive power, i.e. at least one of the three coefficients* β_1, β_2 *or* β_3 *differs from 0.*

c2 (3) Test (α = 0.05) null hypothesis H_0: Lactic concentration (x_3) has no predictive value. Give:
(1) the outcome of the test statistic,
(2) the distribution of the test statistic under H_0 (including any degrees of freedom),
(3) the P-value and your conclusion (in words).

Outcome of the test statistic: t = - 0.246 (test statistic of t-test for H_0: β_3 = 0).

Distribution of the test statistic (including degrees of freedom):

t-distribution with degrees of freedom df = 16.

P-value: 0.809

Conclusion (in words):

We cannot show that Lactic concentration (x_3) has any predictive value, because the (two-sided) P-value of 0.809 exceeds α = 0.05.

d1 (2) The researcher determines the upper confidence bound of a **one-sided** 0.95 confidence interval for coefficient β_3.Without any calculations, explain whether this upper bound will be lower, higher or the same as the bound 20.333 taken from the table of coefficients in the SPSS output.

Explanation (*no calculations*)*:*

Upper bound 20.333 is the upper bound of a two-sided 0.95 confidence interval.
So it is also the upper bound of a one-sided 0.975 confidence interval.
We want the upper bound of a one-sided 0.95 interval.
The latter upper bound will be smaller than 20.333, because the true value will be part of the interval with smaller probability 0.95, the interval will be smaller, so the upper bound will be smaller.

d2 (2) Suppose that the null hypothesis is not rejected in part c2. Explain, without any calculations whether value 0 is in the **two-sided** 0.95 interval for β_3 or not?

Explanation:

The two-sided 0.95 confidence for β_3 consists of all possible values for β_3 that are not rejected by the (two-sided) t-test (α = 0.05). Since value 0 is not rejected, it will be in the two-sided 0.95 confidence interval.

The researcher considers dropping concentrations Acetic (x_1) and Lactic (x_3) from the model. Choose the right output from page 5 to answer the next questions.

e (4) To feel really confident about this she performs an F-test ($\alpha = 0.05$). Give:

(1) the null and alternative hypotheses H_0 and H_a in the notation of the model,
(2) the outcome of the test statistic,
(3) the distribution of the test statistic under H_0,
(4) the P-value and your conclusion (in words).

Null hypothesis and alternative hypothesis:

H_0: $\beta_1 = \beta_3 = 0$ versus H_a: either $\beta_1 \neq 0$ or $\beta_3 \neq 0$ or both $\beta_1 \neq 0$ and $\beta_3 \neq 0$.

Outcome test statistic: $F = 1.223$.

Distribution test statistic under the null hypothesis(including any degrees of freedom):

F-distribution with degrees of freedom df1 = 2, df2 = 16.

P-value: 0.32

Conclusion (in words):

We cannot show that variables x_1 and x_3 have any predictive value. So, it makes sense to drop these two variables from the model.

Output first part of Question 1, starting with three explanatory variables

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1535.844	3	511.948	4.368	.020[a]
	Residual	1875.094	16	117.193		
	Total	3410.938	19			

a. Predictors: (Constant), x3, x2, x1

b. Dependent Variable: y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	47.837	33.522		1.427	.173	-23.225	118.900
	x1	-11.819	8.935	-.518	-1.323	.205	-30.761	7.123
	x2	7.100	2.356	1.092	3.014	.008	2.107	12.094
	x3	-2.669	10.850	-.066	-.246	.809	-25.671	20.333

a. Dependent Variable: y

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.605[a]	.366	.331	10.9590	.366	10.401	1	18	.005
2	.671[b]	.450	.347	10.8256	.084	1.223	2	16	.320

a. Predictors: (Constant), x2

b. Predictors: (Constant), x2, x3, x1

ANOVA[c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1249.137	1	1249.137	10.401	.005[a]
	Residual	2161.801	18	120.100		
	Total	3410.938	19			
2	Regression	1535.844	3	511.948	4.368	.020[b]
	Residual	1875.094	16	117.193		
	Total	3410.938	19			

a. Predictors: (Constant), x2

b. Predictors: (Constant), x2, x3, x1

c. Dependent Variable: y

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.372[a]	.138	.037	13.1504	.138	1.362	2	17	.283
2	.671[b]	.450	.347	10.8256	.312	9.085	1	16	.008

a. Predictors: (Constant), x3, x1

b. Predictors: (Constant), x3, x1, x2

ANOVA[c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	471.087	2	235.543	1.362	.283[a]
	Residual	2939.851	17	172.932		
	Total	3410.938	19			
2	Regression	1535.844	3	511.948	4.368	.020[b]
	Residual	1875.094	16	117.193		
	Total	3410.938	19			

a. Predictors: (Constant), x3, x1

b. Predictors: (Constant), x3, x1, x2

c. Dependent Variable: y

From now on we proceed with the more simple model $\underline{y}_i = \beta_0 + \beta_2 x_{2i} + \underline{\varepsilon}_i$.

There is SPSS output for this simplified model below.

Output second part of Question 1, simplified model with x_2 (H_2S) only

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1249.137	1	1249.137	10.401	.005[a]
	Residual	2161.801	18	120.100		
	Total	3410.938	19			

a. Predictors: (Constant), x2

b. Dependent Variable: y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.627	7.634		-.082	.935
	x2	3.935	1.220	.605	3.225	.005

a. Dependent Variable: y

Below are some copied lines from output with the simplified model (‘res’ and ‘sep’ are short for ‘residual’ and ‘standard error of prediction’):

y	x2	res	sep
5.6	3.1	-5.97265	4.22945
7.2	5.0	-11.84981	2.69797
24.8	5.4	4.17606	2.53287
47.2	7.4	18.70537	3.04048
24.8	3.8	10.47261	3.56775
…	…	…	…
25.8	9.5	-10.95885	5.00355
7.7	3.9	-7.02093	3.48007
14.2	4.7	-3.66920	2.87044
27.9	6.1	4.52131	2.45980
40.3	9.0	5.50882	4.48158

f1 (2) Calculate an estimate for the expected taste score of a cheese sample with an H_2S concentration of 5.0, and obtain the corresponding standard error from the output.

Estimate:

*estimated value = -0.627 + 3.935 * 5 = 19.048.*

Standard error:

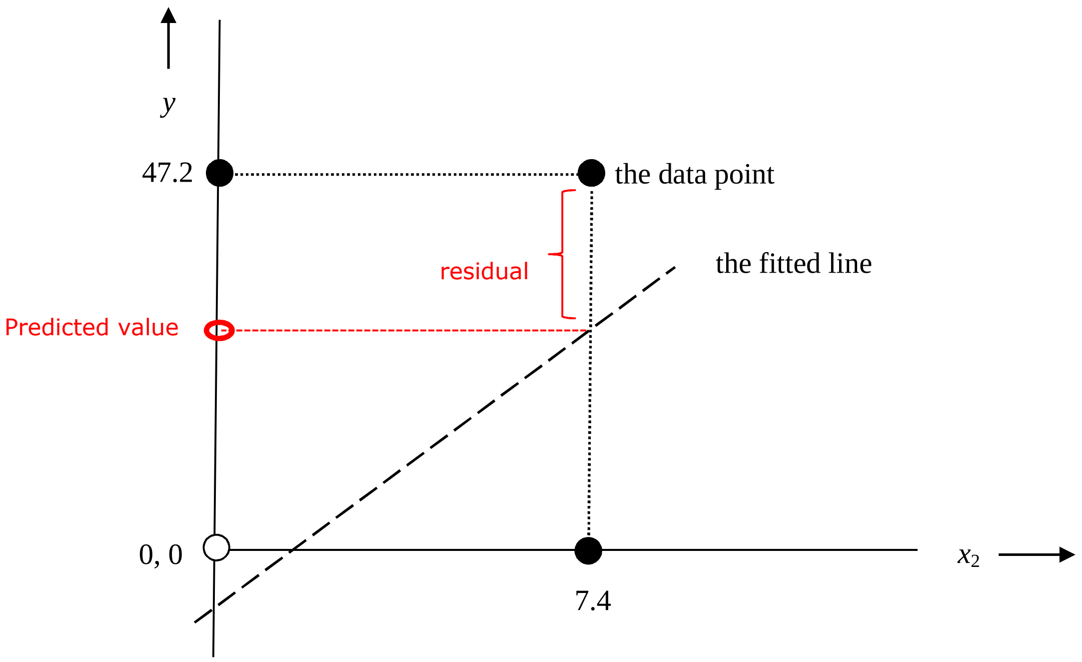
read from copied lines above (sep for x2 equal 5) that se = 2.69797.

f2 (3) Calculate a two-sided 0.95-confidence interval for the expected taste score of any cheese sample with an H_2S concentration of 5.0 (if you could not find answers to f1, use 20.0 as an estimate and 3.0 as the standard error).

Confidence interval:

*Confidence bounds are 19.048 ±2.101 * 2.69797, so the 0.95 confidence interval is (13.38, 24.72) (2.101 is from table 2, row with df = 18, column with right-tail probability 0.05 / 2 = 0.025).*

g (3) Below is a sketch of the fitted line and the data point x_2 = 7.4 and y = 47.2. Indicate, in this sketch, the fitted (predicted) value and the residual corresponding to this data point.
When you are in doubt about your drawing talent, add a brief comment under the picture.



Comment:

Residual is distance (parallel to the vertical y-axis) from the data point to the fitted line.
Predicted (fitted) value is the y-value of the point on the line that corresponds to the data point, so the residual is the difference between the value of response y and the predicted value on the line.

Question 2 (18) **Noise and headache**

The effectiveness of different treatments to reduce sensibility for noise will be studied for people who regularly suffer from headache. A psychologist selects 20 persons with migraine and 20 persons with stress headache. In each group of 20 persons, randomly, six persons are assigned to treatment 1 (relaxation training), six persons to treatment 2 (meditation), and eight persons to treatment 3 (control, no treatment). After a fixed period of time, each person listens individually to a sound that is gradually increased in volume. The volume at which a person indicates that the sound is experienced as decidedly unpleasant, is the response *y*. The model for analysis, with main effects and interactions, reads as follows:

$$\underline{y}_{ijk} = \mu + \tau_i + \beta_j + \tau\beta_{ij} + \underline{\varepsilon}_{ijk},$$

with indices *i* for migraine or stress headache, *j* for treatments 1, 2, 3, and *k* for persons within combination of headache group and treatment ($k = 1 \ldots 6$ or 8). Error terms $\underline{\varepsilon}_{ijk}$ are assumed to be independently normally distributed around 0 with equal variance σ^2.

a1(1) Briefly explain why it is important that the treatments are randomly assigned (within each group of 20 persons).

Explanation:

Treatments are randomly assigned within groups to avoid confounding of treatment effects with patient characteristics , e.g. age, that may also have an effect on the response.

a2(2) (1) Explain why the study is observational with respect to the factor headache.
(2) What kind of risk do you see with respect to validity of conclusions for any differences between migraine or stress headache?

Explanation:

You cannot randomly assign migraine or stress headache to patients. The selected patients already suffer from one of these two types of headache. Which makes this part of the study observational.

Risk:

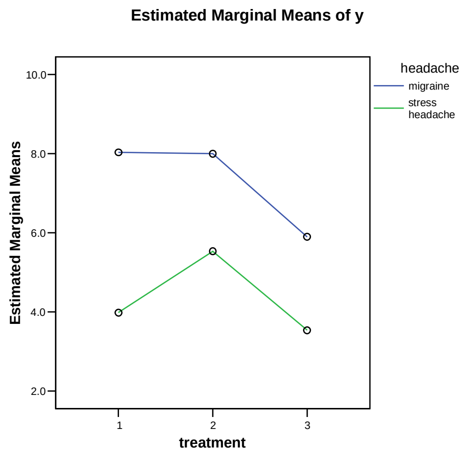
So the two types of headache may be confounded with other factors, such as age. That means that any difference that we find between groups could be, partially or completely, an e.g. age effect as well.

b1(2) Explain what the presence of (sizeable) interaction ($\tau\beta_{ij}$) would imply for the differences between the three treatments, in terms of this study.

Explanation:

It would mean that the difference in expected response between (some of) the treatments is not the same for patients with the two types of headache.

b2 (2) On the right a profile plot is shown of the six mean responses for the three treatments within the migraine and stress headache groups. How should this plot ideally look like, when there is no interaction (explain your answer)?



Explanation:

Ideally, the two broken lines that connect the means for the treatments for each of the two headache groups should be parallel.

c1 (2) Would it matter for the test results of F-tests for main effects and interactions whether we used type I or type II sums of squares for this data? Explain your answer.

Answer (*with explanation*)*:*

With 6, 6 and 8 people per treatment within each headache group we have an orthogonal design. So, it does not matter whether we use type I or type II sums of squares: they will be the same.

For the next questions, you can use the SPSS output below.

Output first part of Question 2

Between-Subjects Factors

		Value Label	N
headache	1	migraine	20
	2	stress headache	20
treatment	1		12
	2		12
	3		16

Descriptive Statistics

Dependent Variable: y

headache	treatment	Mean	Std. Deviation	N
migraine	1	8.033	.9501	6
	2	8.000	1.7321	6
	3	5.900	1.5014	8
	Total	7.170	1.7290	20
stress headache	1	3.983	1.0553	6
	2	5.533	1.0614	6
	3	3.538	1.8981	8
	Total	4.270	1.6349	20
Total	1	6.008	2.3216	12
	2	6.767	1.8802	12
	3	4.719	2.0547	16
	Total	5.720	2.2170	40

Tests of Between-Subjects Effects

Dependent Variable: y

Source	Type II Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	119.970[a]	5	23.994	11.376	.000
Intercept	1308.736	1	1308.736	620.481	.000
headache	84.100	1	84.100	39.872	.000
treatment	30.184	2	15.092	7.155	.003
headache * treatment	5.686	2	2.843	1.348	.273
Error	71.714	34	2.109		
Total	1500.420	40			
Corrected Total	191.684	39			

a. R Squared = .626 (Adjusted R Squared = .571)

Parameter Estimates

Dependent Variable: y

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	3.538	.513	6.889	.000	2.494	4.581
[headache=1]	2.363	.726	3.253	.003	.887	3.838
[headache=2]	0[a]	.	.	.	.	.
[treatment=1]	.446	.784	.568	.573	-1.148	2.040
[treatment=2]	1.996	.784	2.545	.016	.402	3.590
[treatment=3]	0[a]	.	.	.	.	.
[headache=1] * [treatment=1]	1.688	1.109	1.521	.137	-.567	3.942
[headache=1] * [treatment=2]	.104	1.109	.094	.926	-2.150	2.358
[headache=1] * [treatment=3]	0[a]	.	.	.	.	.
[headache=2] * [treatment=1]	0[a]	.	.	.	.	.
[headache=2] * [treatment=2]	0[a]	.	.	.	.	.
[headache=2] * [treatment=3]	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

c2(3) Test ($\alpha = 0.05$) whether there is interaction between headache and treatment.

(1) Give the outcome of the test statistic,

(2) give the distribution of the test statistic when there is no interaction (with degrees of freedom),

(3) is interaction between headache and treatment statistically significant or not?

Outcome test statistic: F-test, test statistic F = 1.348

Distribution of test statistic (including any degrees of freedom):

F-distribution, df1 =2 , df2 = 34.

Statistically significant or not:

P-value = 0.273 > 0.05, so the interaction is not statistically significant.

The psychologist decides to drop the interaction terms from the model.

So, from now on, the additive model, without interaction, is assumed.

There is output from an analysis with this simplified model below.

Output second part of Question 2

Tests of Between-Subjects Effects

Dependent Variable: y

Source	Type II Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	114.284[a]	3	38.095	17.718	.000
Intercept	1308.736	1	1308.736	608.713	.000
headache	84.100	1	84.100	39.116	.000
treatment	30.184	2	15.092	7.019	.003
Error	77.400	36	2.150		
Total	1500.420	40			
Corrected Total	191.684	39			

a. R Squared = .596 (Adjusted R Squared = .563)

Parameter Estimates

Dependent Variable: y

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	3.269	.434	7.536	.000	2.389	4.148
[headache=1]	2.900	.464	6.254	.000	1.960	3.840
[headache=2]	0[a]	.	.	.	.	.
[treatment=1]	1.290	.560	2.303	.027	.154	2.425
[treatment=2]	2.048	.560	3.657	.001	.912	3.184
[treatment=3]	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

d1 (3) Test ($\alpha = 0.05$) whether there are systematic differences between the three treatments. Give:

(1) the outcome of the test statistic,

(2) the distribution of the test statistic when there is no difference between the treatments in their effect upon the expected response (including any degrees of freedom),

(3) the P-value and your conclusion (in words).

Outcome of test statistic: F-test, F = 7.019.

Distribution of the test statistic(including any degrees of freedom):

F-distribution, df1 = 2, df2 = 36.

P-value:0.003

Conclusion (in words):

P-value = 0.003 < 0.05, so we have shown that there are differences among the three treatments.

d2 (2) Next, the psychologist decides to perform two pairwise comparisons between treatment 'relaxation training' and the control, and between 'meditation' and the control (treatments 1 and 2 versus treatment 3), with Fisher's (protected) LSD (α = 0.05).

(1) Give the two estimated pairwise differences and their standard errors,

(2) are these two pairwise differences significant, according to Fisher's LSD method?

Answer:

From the output: estimated difference 1 versus 3 is 1.290, and estimated difference 2 versus 3 is 2.048.
From the output: respective standard errors are 0.560 and 0.560.
From the output: respective P-values are 0.027 and 0.001.
So both differences are significant with Fisher's LSD (since this is the same as applying the t-test twice)..

d3 (1) If the psychologist wanted to have an experiment wise error rate of at most 0.05 in part d2, what method would you recommend for the pairwise comparisons? Explain your answer.

Answer (*with explanation*)*:*

These are comparisons of treatments with the control treatment, so Dunnett's method would be appropriate.

Question 3 (12) **Allergy and food**

One of the possible causes for a particular allergy may be the presence of a substance A in food. In a study, a standard meal is compared with an A-poor meal. From 100 persons with the allergy, 60 persons are randomly assigned to the standard meal, and 40 persons to the A-poor meal. For each person it is observed whether there is an allergic reaction (yes or no) after eating the assigned meal. Persons are assumed to react independently from each other. The observations are shown in the table on the right.

meal * reaction Crosstabulation

Count

		reaction		
		yes	no	Total
meal	standard	48	12	60
	A poor	29	11	40
Total		77	23	100

Let π_1 and π_2 be the probabilities for an allergic reaction with the standard and A-poor meal, respectively.

a (2) Give an estimate for π_1, and an estimate for the corresponding standard error.

Estimate: 48 / 60 = 0.8.

Standard error: $\sqrt{0.8 * (1 - 0.8) / 60} = 0.0516$.

The researcher decides to use Pearson's chi-square statistic to test the null hypothesis H_0: $\pi_1 = \pi_2$.

b1 (2) To obtain the outcome of the test statistic, a separate term is calculated for each of the four cells (meal standard or poor / reaction yes or no) in the table, and these four terms are added up. Calculate the term corresponding to the standard meal and an allergic reaction.

Calculation of the term:

Term is $(\text{Observed} - \text{Expected})^2 / \text{Expected}$ *, with Observed = 48, and Expected = 77 * 60 / 100 = 46.2. So, required term is* $(48 - 46.2)^2 / 46.2 = 0.0701$.

b2 (1) For the test the researcher wants to use an approximation with the chi-square distribution. What is the number of degrees of freedom?

*Degrees of freedom: (number of rows -1) * (number of columns – 1) =(2 - 1) * (2 - 1) = 1.*

The researcher discovers in the SPSS output below that there is apparently no need for an approximation (with the chi-square distribution).

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	.762[b]	?	.383	.469	.263	
Continuity Correction[a]	.398	?	.528			
Likelihood Ratio	.753	?	.385	.469	.263	
Fisher's Exact Test				.469	.263	
Linear-by-Linear Association	.755[c]	?	.385	.469	.263	.130
N of Valid Cases	100					

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 9.20.

c. The standardized statistic is .869.

b3(3) (1) Explain why you need a one-sided test.
(2) Give the exact P-value the researcher should use for the comparison of the two meals.
(3) Give your conclusion ($\alpha = 0.05$) in words.

Explanation:

You are only interested whether the A-poor meal has a smaller probability for an allergic reaction, e.g. $H_a: \pi_1 > \pi_2$ or $H_a: \pi_1 - \pi_2 > 0$

Exact P-value: 0.263 (note that 48 / 60 > 29 / 40).

Conclusion (in words):

We cannot show (since P-value 0.263 > 0.05) that substance A is a possible cause for the allergy.

The length of the 0.95 confidence interval for probability π_2 is 0.28 (so half of the length is 0.14, not shown in the output, you can take this for granted). The researcher is somewhat disappointed by this: she hoped for a half length of 0.1.

c (3) Suppose that the researcher asked you to calculate, prior to the experiment, how many people are needed for the A-poor meal for a 0.95 interval for π_2 with an expected half length of 0.10. Also suppose that she has no idea beforehand what π_2 could be.
What would be your answer (and explain your calculations)?

Answer (with explanation):

No idea about π_2*, so assume worst case* $\pi_2 = 0.5$ *for calculation of interval length.*
Required number (Ott & Longnecker, page 505) $n = 1.96^2 * 0.5 * 0.5 / 0.1^2 = 96.04 \approx 96$.
Normal approximation OK.

Question 4 (20) **Active time of tablets**

Three types of tablets (T1, T2 or T3) are designed to release their active agent slowly in the body. The types of tablets are compared in a study, where 24 test persons are randomly assigned to T1, T2 and T3, with eight persons per type of tablet. Test persons are individually instructed on how to take the tablet, avoiding any contact between them. Per person the active time *y* is measured. This is the time period (in minutes) during which the concentration of the active agent in the blood stream is above a specific threshold level. The longer the active time, the better. The researchers expect the body mass index (BMI, variable *x*) to be an important indicator for the speed of physical and chemical processes in the body. The model used for analysis reads as follows:

$$\underline{y}_{ij} = \beta_0 + \tau_i + \beta_1 x_{ij} + \underline{\varepsilon}_{ij} ,$$

with effect τ_i for the *i*-th type of tablet (i = 1, 2, 3 for T1, T2 or T3), index j for the test persons (j = 1 … 8 per type of tablet) and coefficient β_1 for BMI. Error terms $\underline{\varepsilon}_{ij}$ are assumed to be independently normally distributed around zero with constant variance σ^2.

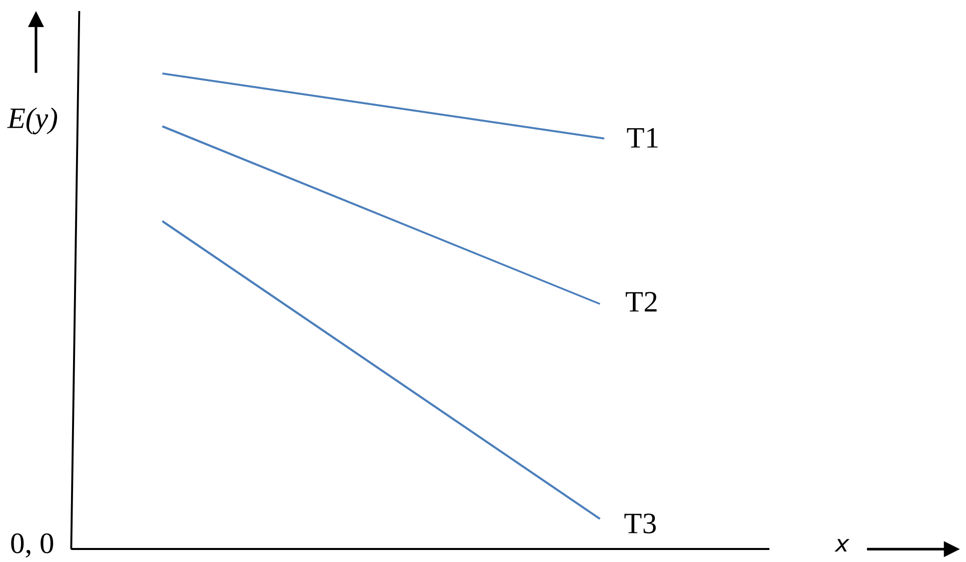
Two test persons failed to show up, and final sample sizes for T1, T2, T3 were 7, 8, 7 (j = 1…7 or 8).

a (2) Explain why the researchers decided to include covariate *x* (BMI) in the model.

Explanation:

They hope that inclusion of covariate x will give a more accurate comparison between tablets T1, T2 and T3, since the analysis will "correct" for differences in BMI among test persons.

b (3) Below is a sketch of the **systematic part** of the model, with expected values for *y* along the vertical and variable *x* along the horizontal axis, and separate lines for the three types of tablets. Explain what is decidedly wrong in this sketch.



Explanation:

according to the model all three lines in the sketch should have the same slope (β_1), which is evidently not true: the three lines that are shown have different slopes.

For the next questions you can use the output from SPSS from page 19.

Take care, output is provided for different models!

c (3) Use an F-test ($\alpha = 0.05$) to investigate whether there are systematic differences between T1, T2 and T3 with respect to their effect upon expected active time of a tablet. Give:
(1) the outcome of the test statistic,
(2) the distribution of the test statistic (including any degrees of freedom), when T1, T2 and T3 have the same effect on active time,
(3) the P-value, and your conclusion (in words).

Outcome of test statistic: F = 5.874.

Distribution of test statistic (including any degrees of freedom):

F-distribution, df1 = 2, df2 = 18.

P-value: 0.011.

Conclusion (in words):

We have proven that there are differences among tablets T1, T2 and T3 with respect to their effect upon active time.

d1(2) Give an estimate of β_1 and an estimate of the corresponding standard error.

Estimate: -3.859.

Standard error: 2.318.

d2(1) Give the name of the statistical criterion that was used to obtain the estimate in part d1.

Name of statistical criterion: the least squares criterion (least squares estimation also OK).

Tablets T1 and T2 are new, but tablet T3 is more familiar and added as a control. Tablet T1 showed promising results in in-vitro experiments. At the least, the researchers hope to show in this study with test persons that T1 is superior to T3.

e1 (2) Formulate a suitable null hypothesis H_0 and alternative hypothesis H_a in terms of the parameters of the model for comparing T1 and T3.

Null hypothesis: H_0: $\tau_1 - \tau_3 = 0$ *or* H_0: $\tau_1 = \tau_3$.

Alternative hypothesis: H_a: $\tau_1 - \tau_3 > 0$ *or* H_a: $\tau_1 > \tau_3$.

e2 (3) Test ($\alpha = 0.05$) whether tablet T1 performs **better** than T3. Give:
(1) the outcome of the test statistic,
(2) the distribution (including any degrees of freedom) of the test statistic when T1 does not perform better than T3.
(3) the P-value, and your conclusion (in words).

Outcome of test statistic: test statistic of t-test is t = 2.297.

Distribution of test statistic: t-distribution with df =18.

P-value: 0.034 / 2 = 0.017.

Your conclusion (in words):

We have proven (because P-value 0.017 < 0.05) that tablet T1 performs better (has a higher expected active time) than tablet T3.

e3 (1) Give a 0.95 confidence interval for the expected difference in active time between tablets of type T1 and T3.

Confidence interval: from the output (3.139, 70.569).

f (1) Is there evidence, on the basis of a test ($\alpha = 0.05$) , that covariate x (BMI) is indeed an important covariate to include in the model?

Answer: no, the (two-sided) P-value for H_0: $\beta_1 = 0$ (t-test) is 0.113 > 0.05.

Extra: even when we would use a one-sided P-value, the answer would still be no.

In the model, the effect of BMI on expected active time is supposed to be the same for the three types of tablet T1, T2 and T3. The researchers wonder whether this is an appropriate assumption.

g (3) Test ($\alpha = 0.05$) whether the (linear) effect of BMI is the same for T1, T2 and T3. Give:
(1) the outcome of the test statistic,
(2) the distribution of the test statistic (including any degrees of freedom), when the BMI effect is the same for T1, T2 and T3,
(3) the P-value, and your conclusion (in words).

Outcome test statistic: test for interaction between tablets and BMI, outcome F = 0.254.

Distribution of test statistic: F-distribution, df1 = 2, df2 = 16.

P-value: 0.779.

Conclusion (in words):

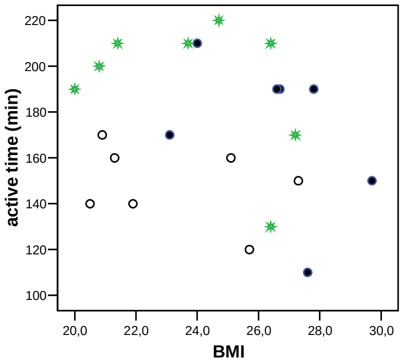
We cannot show (P-value is 0.779 > 0.05) that the three lines have a different slope, so there is no reason to doubt the initial assumption of parallel lines (a common slope).

Extra: actually, since we could not show that the common slope was different from zero, BMI does not seem to be a useful addition to the model.

SPSS output for Question 4, note that there is output for different models!

Report

Tablet		BMI	active time (min)
T1	Mean	26.500	172.86
	N	7	7
T2	Mean	23.825	192.50
	N	8	8
T3	Mean	23.243	148.57
	N	7	7
Total	Mean	24.491	172.27
	N	22	22



Tests of Between-Subjects Effects

Dependent Variable: active time (min)

Source	Type II Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	9152.857[a]	3	3050.952	4.347	.018
Intercept	14867.247	1	14867.247	21.183	.000
Tablet	8245.681	2	4122.840	5.874	.011
BMI	1945.065	1	1945.065	2.771	.113
Error	12633.506	18	701.861		
Total	674700.000	22			
Corrected Total	21786.364	21			

a. R Squared = .420 (Adjusted R Squared = .323)

Parameter Estimates

Dependent Variable: active time (min)

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	238.259	54.798	4.348	.000	123.133	353.385
[Tablet=1]	36.854	16.048	2.297	.034	3.139	70.569
[Tablet=2]	46.175	13.777	3.351	.004	17.229	75.120
[Tablet=3]	0[a]	.	.	.	.	.
BMI	-3.859	2.318	-1.665	.113	-8.729	1.011

a. This parameter is set to zero because it is redundant.

Tests of Between-Subjects Effects

Dependent Variable: active time (min)

Source	Type II Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	9541.154[a]	5	1908.231	2.493	.075
Intercept	15389.520	1	15389.520	20.108	.000
Tablet	771.309	2	385.655	.504	.613
BMI	1945.065	1	1945.065	2.541	.130
Tablet * BMI	388.296	2	194.148	.254	.779
Error	12245.210	16	765.326		
Total	674700.000	22			
Corrected Total	21786.364	21			

a. R Squared = .438 (Adjusted R Squared = .262)

Parameter Estimates

Dependent Variable: active time (min)

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	190.620	96.916	1.967	.067	-14.832	396.072
[Tablet=1]	151.495	163.826	.925	.369	-195.801	498.792
[Tablet=2]	99.394	131.808	.754	.462	-180.026	378.814
[Tablet=3]	0[a]	.	.	.	.	.
BMI	-1.809	4.145	-.436	.668	-10.597	6.979
[Tablet=1] * BMI	-4.578	6.471	-.707	.489	-18.296	9.140
[Tablet=2] * BMI	-2.284	5.574	-.410	.687	-14.101	9.534
[Tablet=3] * BMI	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

Question 5 (14) **Food and fitness**

School	Treatment	y
1	BE	5.50
1	BE	5.82
1	B	4.68
1	B	5.26
1	C	5.62
1	C	5.05
2	BE	6.07
2	BE	5.40
2	B	5.85
2	B	5.15
2	C	5.99
2	C	5.41
…	…	…
15	BE	5.57
15	BE	6.05
15	B	6.06
15	B	6.42
15	C	6.27
15	C	6.92

A researcher is interested in the possible association between food consumption and fitness of children. From each of 15 schools, six classes with pupils of about 12 years of age are involved. Within each school, in two classes a brochure about healthy food is handed out and through meetings parents are involved as well (treatment "brochure & extra" or BE), in two classes only the brochure is handed out, without meetings with the parents ("brochure" or B), and in two classes no treatment is applied ("Control" or C). BE, B and C are randomly assigned to the six classes of each school. Fitness per pupil is expressed by an average score, between 0 (poor) and 10 (excellent), from several tests taken over a period of a week during the sports lessons.

The study lasts for a year and the response *y* is the difference between the average fitness score per class at the end and at the start of the study. Part of the data for the 15 * 6 = 90 school classes is shown above on the right. The data are analysed with a mixed model:

$$\underline{y}_{ijk} = \mu + \tau_i + \underline{\beta}_j + \underline{\tau\beta}_{ij} + \underline{\epsilon}_{ijk},$$

with fixed effects for treatment (τ_i , $i = 1, 2, 3$ for BE, B and C, respectively), and independent normally distributed random effects for schools ($\underline{\beta}_j$), random interaction between treatments and schools ($\underline{\tau\beta}_{ij}$), and random residual error terms ($\underline{\epsilon}_{ijk}$) ($i$ = 1,2, 3; j = 1…15; k = 1, 2), with zero means and components of variance σ_β^2, $\sigma_{\tau\beta}^2$ and σ^2, respectively.

a1 (1) Motivate why schools are introduced as random effects and treatments as fixed effects.

Answer:

Suppose that you would repeat the experiment, then you would look at the same treatments, but since you would like your conclusions to apply to any randomly chosen school, schools would not be the same, you would like to include as many schools as possible in your data. So, treatments are fixed and schools are random.

a2 (1) Which component of variance represents the variation between classes of the same school that receive the same treatment?

Answer: the component for residual variance σ^2.

a3 (2) Suppose that some schools get many of their pupils from well-to-do areas (more prosperous areas, with more highly educated inhabitants) and others do not.
Explain how this could affect the variance component for interaction.

Answer:

You might imagine that this makes a difference for e.g. the notice parents take of the brochure and the attendance of the meetings. Hence, for some schools differences between treatments can be larger than for other schools. This means (random) interaction between schools and treatments.

We assume, like SPSS, that $\tau_3 = 0$.

b1(1) What is the practical interpretation of μ in terms of this study?

Interpretation of μ: this is the expected difference in fitness for the control treatment C.

b2 (1) Give the expression for the **expected** difference in score between classes with treatments BE and B, in terms of the parameters of the model.

Expression: $\tau_1 - \tau_2$.

For the next questions you can use the SPSS output on page 23.

c (4) Test with an F-test whether there are differences among BE, B and C with respect to their effect upon the expected score for fitness.

(1) give the null and alternative hypothesis H_0 and H_a in terms of the parameters of the model,
(2) give the expression for the test statistic F in terms of mean squares,
(3) give the distribution of test statistic F (with degrees of freedom) under H_0.
(4) when would you reject H_0, for values of F that are (too) small or (too) large?
(5) give the P-value, and your conclusion (in words).

H_0 and H_a: $H_0: \tau_1 = \tau_2 = \tau_3$, *and H_a: for at least two different treatments i and j we have* $\tau_i \neq \tau_j$.

Expression for test statistic F:

*F = MS(treatment) / MS(school * treatment) (see below 2nd table SPSS output).*

Distribution of test statistic F (with degrees of freedom):

F-distribution with degrees of freedom df1 =2, df2 = 28.

Small or large values for rejection: You reject for large values of F.

P-value: 0.002.

Conclusion:

We have shown that there are differences among the treatments in their effect upon the expected difference in fitness.

For parts d1 and d2 below, suppose that we take a random school, choose a random class within that school, hand out the brochure only (treatment B) in that class, and determine the response *y* for that class after a year.

d1 (1) Give an estimate for the variance (the population variance for B) of the response *y*.

Answer: sum of estimated components of variance = $\hat{\sigma}_{\beta}^{2} + \hat{\sigma}_{\tau\beta}^{2} + \hat{\sigma}^{2}$ = 0.335 + 0.302 + 0.190 = 0.83

(estimated values for variance components can be read from the output).

d2 (1) Give an estimate for the expected value (the population mean for B) of the response *y*.

Answer: the sample mean of B, which is 5.89 (1st table SPSS output).

From the last table with SPSS output you can read the standard error of the difference between BE and control as 0.112. This is actually wrong, because in the calculation SPSS is mistakenly using the mean square for error (residual).

e(2) What mean square should SPSS use instead to get the right result (which is 0.23)?

*Answer: MS(school * treatment).*

Extra:

$$Var(\hat{\tau}_1) = Var(\bar{y}_{1}..-\bar{y}_{3}..) = 2 * \left(\frac{\hat{\sigma}_{\tau\beta}^{2}}{15} + \frac{\hat{\sigma}^{2}}{30}\right) = 2 * \frac{\left(2\hat{\sigma}_{\tau\beta}^{2} + \hat{\sigma}^{2}\right)}{30} =$$

$2 * MS(school * treatment)$ / 30 *=2 * 0.794 / 30 = 0.0529, so estimated se =√0.0529=0.23.*

Output for Question 5

Means

Difference of average scores for fitness beginning and end of experiment.

treatment	Mean	N	Std. Deviation
BE = brochure & extra	6.3377	30	.92879
B = brochure	5.8893	30	.85628
C = control	5.4113	30	.90493
Total	5.8794	90	.96514

Variance Estimates

Component	Estimate
Var(school)	.335
Var(school * treatment)	.302
Var(Error)	.190

Dependent Variable: fitness
Method: ANOVA (Type I Sum of Squares)

Tests of Between-Subjects Effects

Dependent Variable: difference of average scores for fitness beginning and end of experiment

Source		Type II Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	3111.108	1	3111.108	1109.461	.000
	Error	39.258	14	2.804[a]		
treatment	Hypothesis	12.876	2	6.438	8.108	.002
	Error	22.232	28	.794[b]		
school	Hypothesis	39.258	14	2.804	3.532	.002
	Error	22.232	28	.794[b]		
treatment * school	Hypothesis	22.232	28	.794	4.185	.000
	Error	8.538	45	.190[c]		

a. MS(school)
b. MS(treatment * school)
c. MS(Error

Expected Mean Squares

Source	Variance Component			
	Var(school)	Var(school * treatment)	Var(Error)	Quadratic Term
Intercept	6.000	2.000	1.000	Intercept, treatment
treatment	.000	2.000	1.000	treatment
school	6.000	2.000	1.000	
school * treatment	.000	2.000	1.000	
Error	.000	.000	1.000	

Dependent Variable: fitness
Expected Mean Squares are based on Type I Sums of Squares.
For each source, the expected mean square equals the sum of the coefficients in the cells times the variance components, plus a quadratic term involving effects in the Quadratic Term cell.

Pairwise Comparisons

Dependent Variable: difference of average scores for fitness at the beginning and at the end of the experiment

(I) treatment	(J) treatment	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a] Lower Bound	Upper Bound
BE = brochure & extra	B = brochure	.448*	.112	.000	.222	.675
	C = control	.926*	.112	.000	.700	1.153
B = brochure	BE = brochure & extra	-.448*	.112	.000	-.675	-.222
	C = control	.478*	.112	.000	.251	.705
C = control	BE = brochure & extra	-.926*	.112	.000	-1.153	-.700
	B = brochure	-.478*	.112	.000	-.705	-.251

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**).