

## Final examination, 23. April 2007

All aids are allowed, except personal computer and personal assistance. The exam consists of 3 questions, which have equal weight (*10 points each*) and should all be answered; further detail about the points is given for specific parts of each question. The duration of the exam is 3 hours. Generally, all statistical models used should be specified, and to such detail that it is clear which terms are present and in which form.

### Questions 1+3.

Not shown.

### Question 2.

An experiment on maize germination recorded the “amylase specific activity” of sprouted maize under 32 different treatment conditions. These treatment conditions were made up by all combinations of analysis temperature (`temp_anl`; values: 10, 13, 15, 20, 25, 30, 35 and 40 °C), growth temperature (`temp_grw`; values: 13 and 25 °C) and variety of maize (B73 and O43). There were three replicates done for each set of treatment conditions, for a total of 96 observations.

- A) (*4 points*) Describe the type of experiment carried out, and specify the statistical model used in the Minitab/Stata listings for Question 2.A, either in terms of a model equation or a model formula. Use the listings to draw overall conclusions about the effects of the different factors in the experiment. State specifically which of the three factors have a statistically significant impact on the amylase activity.
- B) (*3 points*) Use the Minitab plots and tables provided for Question 2.B to explore all the significant effects, and describe in simple terms their character. Specify precisely what statistical evidence you rely on when referring to some levels as being “different”. (*Note*: No Stata output is provided for this part.)
- C) (*3 points*) For the purpose of predicting the amylase activity as a polynomial function of the analysis temperature, a number of additional analyses were carried out (Minitab/Stata listings for Question 2.C). Based on these analyses, which model do you consider as the “best” model for prediction? (explain your choice). Use your selected model to predict the amylase activity at an analysis temperature of 28 °C; explain carefully the conditions your prediction corresponds to. Give also a statistical assessment of whether this “best” model gives an acceptable fit to the data compared to the model with `temp_anl` as a factor (categorical predictor).

Minitab listing for Question 2.A:

```
MTB > GLM 'amylase' = variety| 'temp_anl'| 'temp_grw';
SUBC> Brief 2 ;
SUBC> RType 2 .
General Linear Model: amylase versus variety, temp_anl, temp_grw
```

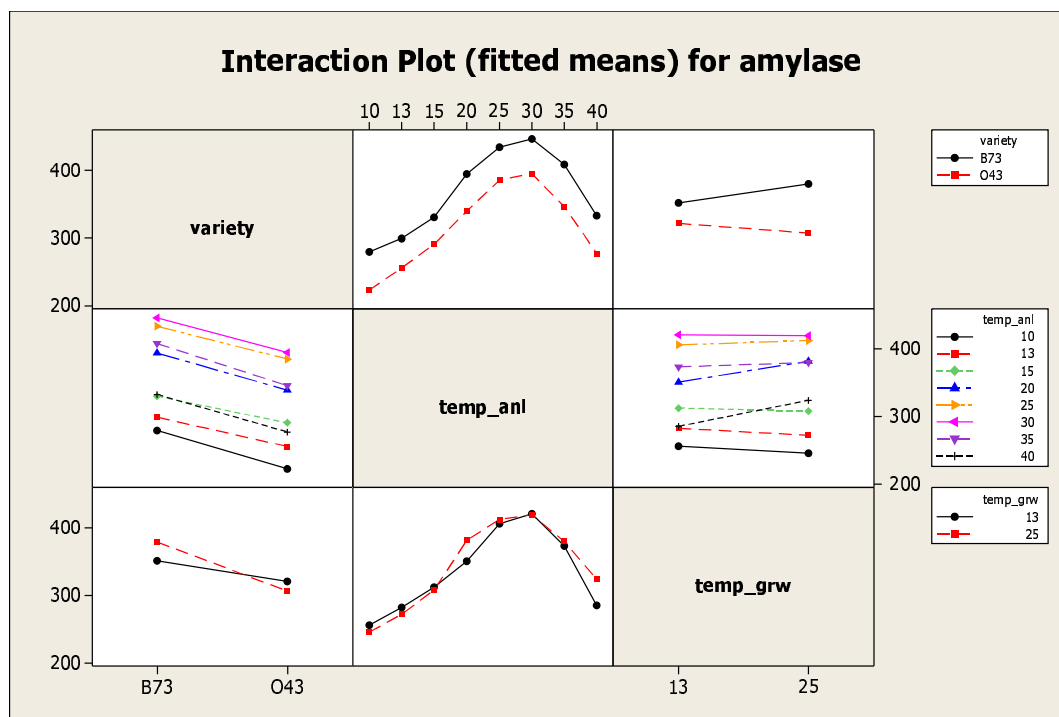
Factor	Type	Levels	Values
variety	fixed	2	B73, O43
temp_anl	fixed	8	10, 13, 15, 20, 25, 30, 35, 40
temp_grw	fixed	2	13, 25

Analysis of Variance for amylase, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
variety	1	63773	63773	63773	99.33	0.000
temp_anl	7	327762	327762	46823	72.93	0.000
temp_grw	1	1154	1154	1154	1.80	0.185
variety*temp_anl	7	1177	1177	168	0.26	0.966
variety*temp_grw	1	10637	10637	10637	16.57	0.000
temp_anl*temp_grw	7	7151	7151	1022	1.59	0.154
variety*temp_anl*temp_grw	7	6257	6257	894	1.39	0.224
Error	64	41088	41088	642		
Total	95	458998				

S = 25.3378    R-Sq = 91.05%    R-Sq(adj) = 86.71%

Minitab output for Question 2.B:



MEANS of amylase

Rows: variety Columns: temp\_anl

	10	13	15	20	25	30	35	40	All
B73	279.0	299.0	330.0	394.0	433.5	446.0	408.0	332.5	365.2
043	222.5	255.5	290.5	339.0	385.0	394.5	345.5	277.0	313.7
All	250.8	277.3	310.3	366.5	409.3	420.3	376.7	304.8	339.5

Rows: temp\_grw Columns: temp\_anl

	10	13	15	20	25	30	35	40	All
13	256.0	282.5	312.5	351.0	406.0	421.0	373.5	285.5	336.0
25	245.5	272.0	308.0	382.0	412.5	419.5	380.0	324.0	342.9
All	250.8	277.3	310.3	366.5	409.3	420.3	376.7	304.8	339.5

Rows: variety Columns: temp\_grw

	13	25	All
B73	351.3	379.2	365.2
043	320.8	306.6	313.7
All	336.0	342.9	339.5

Minitab listing for Question 2.C:

```
MTB > GLM 'amylase' = variety | 'temp_grw' 'temp_anl';
SUBC> Covariates 'temp_anl';
SUBC> Brief 2 .
General Linear Model: amylase versus variety, temp_grw
```

Factor	Type	Levels	Values
variety	fixed	2	B73, 043
temp_grw	fixed	2	13, 25

Analysis of Variance for amylase, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
variety	1	63773	63773	63773	19.78	0.000
temp_grw	1	1154	1154	1154	0.36	0.551
variety*temp_grw	1	10637	10637	10637	3.30	0.073
temp_anl	1	90095	90095	90095	27.95	0.000
Error	91	293339	293339	3224		
Total	95	458998				

S = 56.7759 R-Sq = 36.09% R-Sq(adj) = 33.28%

Term	Coef	SE Coef	T	P
Constant	268.62	14.60	18.40	0.000
temp_anl	3.0149	0.5703	5.29	0.000

```
MTB > GLM 'amylase' = variety| 'temp_grw' 'temp_anl' 'temp_anl'* &
CONT> 'temp_anl';
SUBC> Covariates 'temp_anl';
SUBC> Brief 2 .
General Linear Model: amylase versus variety, temp_grw
```

Factor	Type	Levels	Values
variety	fixed	2	B73, 043
temp_grw	fixed	2	13, 25

Analysis of Variance for amylase, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
variety	1	63773	63773	63773	84.60	0.000
temp_grw	1	1154	1154	1154	1.53	0.219
variety*temp_grw	1	10637	10637	10637	14.11	0.000
temp_anl	1	90095	267350	267350	354.65	0.000
temp_anl*temp_anl	1	225494	225494	225494	299.13	0.000
Error	90	67846	67846	754		
Total	95	458998				

S = 27.4562 R-Sq = 85.22% R-Sq(adj) = 84.40%

Term	Coef	SE Coef	T	P
Constant	-30.26	18.67	-1.62	0.109
temp_anl	32.197	1.710	18.83	0.000
temp_anl*temp_anl	-0.59025	0.03413	-17.30	0.000

```
MTB > GLM 'amylase' = variety| 'temp_grw' 'temp_anl' 'temp_anl'* &
CONT> 'temp_anl' 'temp_anl'* 'temp_anl'* 'temp_anl';
SUBC> Covariates 'temp_anl';
SUBC> Brief 2 .
General Linear Model: amylase versus variety, temp_grw
```

Factor	Type	Levels	Values
variety	fixed	2	B73, 043
temp_grw	fixed	2	13, 25

Analysis of Variance for amylase, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
variety	1	63773	63773	63773	98.17	0.000
temp_grw	1	1154	1154	1154	1.78	0.186
variety*temp_grw	1	10637	10637	10637	16.37	0.000
temp_anl	1	90095	419	419	0.64	0.424
temp_anl*temp_anl	1	225494	2421	2421	3.73	0.057
temp_anl*temp_anl*temp_anl	1	10031	10031	10031	15.44	0.000
Error	89	57814	57814	650		

Total 95 458998

S = 25.4873 R-Sq = 87.40% R-Sq(adj) = 86.56%

Term	Coef	SE Coef	T	P
Constant	148.44	48.67	3.05	0.003
temp_anl	5.584	6.956	0.80	0.424
temp_anl*temp_anl	0.5765	0.2986	1.93	0.057
temp_anl*temp_anl*temp_anl	-0.015510	0.003947	-3.93	0.000

```
MTB > GLM 'amylase' = variety | 'temp_grw' 'temp_anl' 'temp_anl'* &
CONT> 'temp_anl' 'temp_anl'* 'temp_anl'* 'temp_anl' 'temp_anl'* &
CONT> 'temp_anl'* 'temp_anl'* 'temp_anl';
SUBC> Covariates 'temp_anl';
SUBC> Brief 2 .
General Linear Model: amylase versus variety, temp_grw
```

Factor	Type	Levels	Values
variety	fixed	2	B73, 043
temp_grw	fixed	2	13, 25

Analysis of Variance for amylase, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
variety	1	63773	63773	63773	99.51	0.000
temp_grw	1	1154	1154	1154	1.80	0.183
variety*temp_grw	1	10637	10637	10637	16.60	0.000
temp_anl	1	90095	950	950	1.48	0.227
temp_anl*temp_anl	1	225494	2062	2062	3.22	0.076
temp_anl*temp_anl*temp_anl	1	10031	2063	2063	3.22	0.076
temp_anl*temp_anl*temp_anl*temp_anl	1	1418	1418	1418	2.21	0.140
Error	88	56397	56397	641		
Total	95	458998				

S = 25.3154 R-Sq = 87.71% R-Sq(adj) = 86.74%

Term	Coef	SE Coef	T	P
Constant	331.5	132.2	2.51	0.014
temp_anl	-31.47	25.85	-1.22	0.227
temp_anl*temp_anl	3.157	1.760	1.79	0.076
temp_anl*temp_anl*temp_anl	-0.08939	0.04982	-1.79	0.076
temp_anl*temp_anl*temp_anl*temp_anl	0.000743	0.000499	1.49	0.140

(Stata listing on the next page)

Stata listing for Question 2.A:

```
. anova amylase Variety temp_anl temp_grw Variety*temp_anl Variety*temp_grw temp_a
> nl*temp_grw Variety*temp_anl*temp_grw
```

Source	Partial SS	df	MS	F	Prob > F
Model	417909.454	31	13480.9501	21.00	0.0000
Variety	63772.5084	1	63772.5084	99.33	0.0000
temp_anl	327761.725	7	46823.1036	72.93	0.0000
temp_grw	1154.39974	1	1154.39974	1.80	0.1847
Variety*temp_anl	1177.1157	7	168.159386	0.26	0.9663
Variety*temp_grw	10636.5654	1	10636.5654	16.57	0.0001
temp_anl*temp_grw	7150.60158	7	1021.51451	1.59	0.1542
Variety*temp_anl*temp_grw	6256.53857	7	893.791225	1.39	0.2241
Residual	41088.1733	64	642.002707		
Total	458997.628	95	4831.55398		

Number of obs = 96 R-squared = 0.9105  
 Root MSE = 25.3378 Adj R-squared = 0.8671

Stata listing for Question 2.C:

```
. anova amylase Variety temp_grw Variety*temp_grw temp_anl, cont(temp_anl)
```

Source	Partial SS	df	MS	F	Prob > F
Model	165658.368	4	41414.5919	12.85	0.0000
Variety	63772.5084	1	63772.5084	19.78	0.0000
temp_grw	1154.39974	1	1154.39974	0.36	0.5510
Variety*temp_grw	10636.5654	1	10636.5654	3.30	0.0726
temp_anl	90094.8941	1	90094.8941	27.95	0.0000
Residual	293339.26	91	3223.50835		
Total	458997.628	95	4831.55398		

Number of obs = 96 R-squared = 0.3609  
 Root MSE = 56.7759 Adj R-squared = 0.3328

```
. xi: regress amylase i.variety*i.temp_grw temp_anl
i.variety      _Ivariety_1-2      (_Ivariety_1 for variety==B73 omitted)
i.temp_grw     _Itemp_grw_13-25   (naturally coded; _Itemp_grw_13 omitted)
i.var~y*i.tem~w _IvarXtem_#_#      (coded as above)
```

Source	SS	df	MS	Number of obs =	96
Model	165658.368	4	41414.5919	F( 4, 91) =	12.85
Residual	293339.26	91	3223.50835	Prob > F =	0.0000
				R-squared =	0.3609
				Adj R-squared =	0.3328
Total	458997.628	95	4831.55398	Root MSE =	56.776

amylase	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
_Ivariety_2	-30.49583	16.3898	-1.86	0.066	-63.05217	2.060499
_Itemp_grw~25	27.9875	16.3898	1.71	0.091	-4.568834	60.54383
_IvarXtem~25	-42.10417	23.17868	-1.82	0.073	-88.14578	3.937441
temp_anl	3.014876	.5702742	5.29	0.000	1.882096	4.147656
_cons	280.4004	17.71755	15.83	0.000	245.2067	315.5942

```
. anova amylase Variety temp_grw Variety*temp_grw temp_anl temp_anl*temp_anl, cont
> (temp_anl)
```

```
Number of obs = 96 R-squared = 0.8522
Root MSE = 27.4562 Adj R-squared = 0.8440
```

Source	Partial SS	df	MS	F	Prob > F
Model	391151.902	5	78230.3805	103.78	0.0000
Variety	63772.5084	1	63772.5084	84.60	0.0000
temp_grw	1154.39974	1	1154.39974	1.53	0.2191
Variety*temp_grw	10636.5654	1	10636.5654	14.11	0.0003
temp_anl	267350.367	1	267350.367	354.65	0.0000
temp_anl*temp_anl	225493.535	1	225493.535	299.13	0.0000
Residual	67845.7253	90	753.841392		
Total	458997.628	95	4831.55398		

```
. gen temp_anl2=temp_anl^2
. xi: regress amylase i.variety*i.temp_grw temp_anl temp_anl2
i.variety      _Ivariety_1-2      (_Ivariety_1 for variety==B73 omitted)
i.temp_grw     _Itemp_grw_13-25   (naturally coded; _Itemp_grw_13 omitted)
i.var~y*i.tem~w _IvarXtem_#_#     (coded as above)
```

Source	SS	df	MS	Number of obs =	96
Model	391151.902	5	78230.3805	F( 5, 90) =	103.78
Residual	67845.7253	90	753.841392	Prob > F =	0.0000
				R-squared =	0.8522
				Adj R-squared =	0.8440
Total	458997.628	95	4831.55398	Root MSE =	27.456

amylase	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
_Ivariety_2	-30.49583	7.925914	-3.85	0.000	-46.24205	-14.74962

_Itemp_gr~25		27.9875	7.925914	3.53	0.001	12.24129	43.73371
_IvarXtem~25		-42.10417	11.20894	-3.76	0.000	-64.37267	-19.83566
temp_anl		32.19742	1.709703	18.83	0.000	28.80079	35.59404
temp_anl2		-.5902537	.0341281	-17.30	0.000	-.658055	-.5224523
_cons		-18.47803	19.28838	-0.96	0.341	-56.79777	19.84171

```
. anova amylase Variety temp_grw Variety*temp_grw temp_anl temp_anl*temp_anl temp_
> anl*temp_anl*temp_anl, cont(temp_anl)
```

Number of obs = 96 R-squared = 0.8740  
Root MSE = 25.4873 Adj R-squared = 0.8656

Source	Partial SS	df	MS	F	Prob > F
Model	401183.215	6	66863.8692	102.93	0.0000
Variety	63772.5084	1	63772.5084	98.17	0.0000
temp_grw	1154.39974	1	1154.39974	1.78	0.1859
Variety*temp_grw	10636.5654	1	10636.5654	16.37	0.0001
temp_anl	418.650723	1	418.650723	0.64	0.4242
temp_anl*temp_anl	2421.49842	1	2421.49842	3.73	0.0567
temp_anl*temp_anl*temp_anl	10031.3131	1	10031.3131	15.44	0.0002
Residual	57814.4123	89	649.600138		
Total	458997.628	95	4831.55398		

```
. gen temp_anl3=temp_anl^3
. xi: regress amylase i.variety*i.temp_grw temp_anl temp_anl2 temp_anl3
i.variety      _Ivariety_1-2      (_Ivariety_1 for variety==B73 omitted)
i.temp_grw     _Itemp_grw_13-25   (naturally coded; _Itemp_grw_13 omitted)
i.var~y*i.tem~w _IvarXtem_#_#     (coded as above)
```

Source	SS	df	MS	Number of obs = 96
Model	401183.215	6	66863.8692	F( 6, 89) = 102.93
Residual	57814.4123	89	649.600138	Prob > F = 0.0000
Total	458997.628	95	4831.55398	R-squared = 0.8740

Adj R-squared = 0.8656  
Root MSE = 25.487

amylase	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
_Ivariety_2	-30.49583	7.357537	-4.14	0.000	-45.1151 -15.87657
_Itemp_gr~25	27.9875	7.357537	3.80	0.000	13.36823 42.60677
_IvarXtem~25	-42.10417	10.40513	-4.05	0.000	-62.77894 -21.4294
temp_anl	5.584125	6.955883	0.80	0.424	-8.237067 19.40532
temp_anl2	.5765036	.2985953	1.93	0.057	-.0167989 1.169806
temp_anl3	-.0155102	.003947	-3.93	0.000	-.0233527 -.0076677
_cons	160.2242	48.87315	3.28	0.001	63.11428 257.3341

```
. anova amylase Variety temp_grw Variety*temp_grw temp_anl temp_anl*temp_anl temp_
> anl*temp_anl*temp_anl temp_anl*temp_anl*temp_anl*temp_anl, cont(temp_anl)
```

```
Number of obs = 96 R-squared = 0.8771
Root MSE = 25.3154 Adj R-squared = 0.8674
```

Source	Partial SS	df	MS	F	Prob > F
Model	402601.109	7	57514.4442	89.74	0.0000
Variety	63772.5084	1	63772.5084	99.51	0.0000
temp_grw	1154.39974	1	1154.39974	1.80	0.1830
Variety*temp_grw	10636.5654	1	10636.5654	16.60	0.0001
temp_anl	949.734593	1	949.734593	1.48	0.2267
temp_anl*temp_anl	2062.05151	1	2062.05151	3.22	0.0763
temp_anl*temp_anl*temp_anl	2062.91609	1	2062.91609	3.22	0.0762
temp_anl*temp_anl*temp_anl*temp_anl	1417.89396	1	1417.89396	2.21	0.1405
Residual	56396.5183	88	640.869526		
Total	458997.628	95	4831.55398		

```
. gen temp_anl4=temp_anl^4
. xi: regress amylase i.variety*i.temp_grw temp_anl temp_anl2 temp_anl3 temp_anl4
i.variety      _Ivariety_1-2      (_Ivariety_1 for variety==B73 omitted)
i.temp_grw     _Itemp_grw_13-25   (naturally coded; _Itemp_grw_13 omitted)
i.var~y*i.tem~w _IvarXtem~#_#      (coded as above)
```

Source	SS	df	MS	Number of obs =	96
Model	402601.109	7	57514.4442	F( 7, 88) =	89.74
Residual	56396.5183	88	640.869526	Prob > F =	0.0000
Total	458997.628	95	4831.55398	R-squared =	0.8771
				Adj R-squared =	0.8674
				Root MSE =	25.315

amylase	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
_Ivariety_2	-30.49583	7.307927	-4.17	0.000	-45.0188 -15.97287
_Itemp_gr~25	27.9875	7.307927	3.83	0.000	13.46453 42.51047
_IvarXtem~25	-42.10417	10.33497	-4.07	0.000	-62.64275 -21.56559
temp_anl	-31.47358	25.85412	-1.22	0.227	-82.85321 19.90606
temp_anl2	3.156547	1.759735	1.79	0.076	-.3405569 6.653651
temp_anl3	-.0893858	.049821	-1.79	0.076	-.1883946 .009623
temp_anl4	.0007429	.0004995	1.49	0.140	-.0002497 .0017356
_cons	343.2343	132.2677	2.59	0.011	80.38003 606.0885