

Final examination, 22 April 2024 (for students without VHM 8120)

All aids are allowed, except a computer and personal assistance. Restricted use of some computer-like devices (including tablets and smartphones) is permitted under the rules described at the VHM 802 course homepage. 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. Your answers should generally (unless specified otherwise) be based on the information provided. Nevertheless, if at some point you think it is necessary to carry out additional analysis in statistical software, explain carefully the purpose of your proposed analysis and how you would implement it in the statistical software.

Question 1.

An instrument called a “Visiplume” measures ultraviolet light. By comparing absorption in clear air and absorption in polluted air, the concentration of SO₂ in the polluted air can be estimated. A standard method for measuring SO₂ is in use by the Environmental Protection Agency, and a study was conducted to compare the two methods across a range of air samples. The recorded response was the ratio of the Visiplume reading to the standard reading. The four experimental conditions were: measurements of SO₂ in an inflated bag ($n = 9$), measurements of a smoke generator with SO₂ ($n = 11$), and measurements at two coal-fired plants ($n = 5$ and $n = 6$). The ratios recorded for each of the four experimental conditions are listed in the table below.

Experimental condition	n	ratio of SO ₂ measurement by Visiplume versus standard method											
1: Bag	9	1.055	1.272	0.824	1.019	1.069	0.983	1.025	1.076	1.100			
2: Smoke	11	1.131	1.236	1.161	1.219	1.169	1.238	1.197	1.252	1.435	0.827	3.188	
3: Plant no. 1	5	0.798	0.971	0.923	1.079	1.065							
4: Plant no. 2	6	0.950	0.978	0.762	0.733	0.823	1.011						

Several statistical analyses for these data are shown in the attached Stata listings. For simplicity of presentation, the included graphical displays are from Minitab, but similar graphs could be obtained from Stata. Note that the order of the data in the worksheet corresponds to the order (left to right within rows, rows entered top-down) in the table above.

- A) (*3 points*) Explain the statistical model used in the listing for part A); you may either give a model formula with the relevant model assumptions, or give a verbal description of the model and its assumptions. Discuss, based on the information provided and any additional calculations you find useful, whether the assumptions behind the model seem to be met to a satisfactory degree, and identify as closely as possible the source of any problem(s). Defer a discussion of how you might want to deal with any problem(s) with the model assumptions to part B).
- B) (*3 points*) Review and explain the analyses presented in the listings for part B). Interpret the results, and use them to select a model for the data which satisfactorily meets the model’s

assumptions. If you have a choice between several satisfactory models, discuss briefly the pros and cons of each of them.

- C) (4 points) Irrespective of your conclusions in A) and B), choose for this part one model to focus on for the statistical inference, preferably your “best” model from parts A) and B). Carry out a statistical analysis with focus on the following specific study objectives:

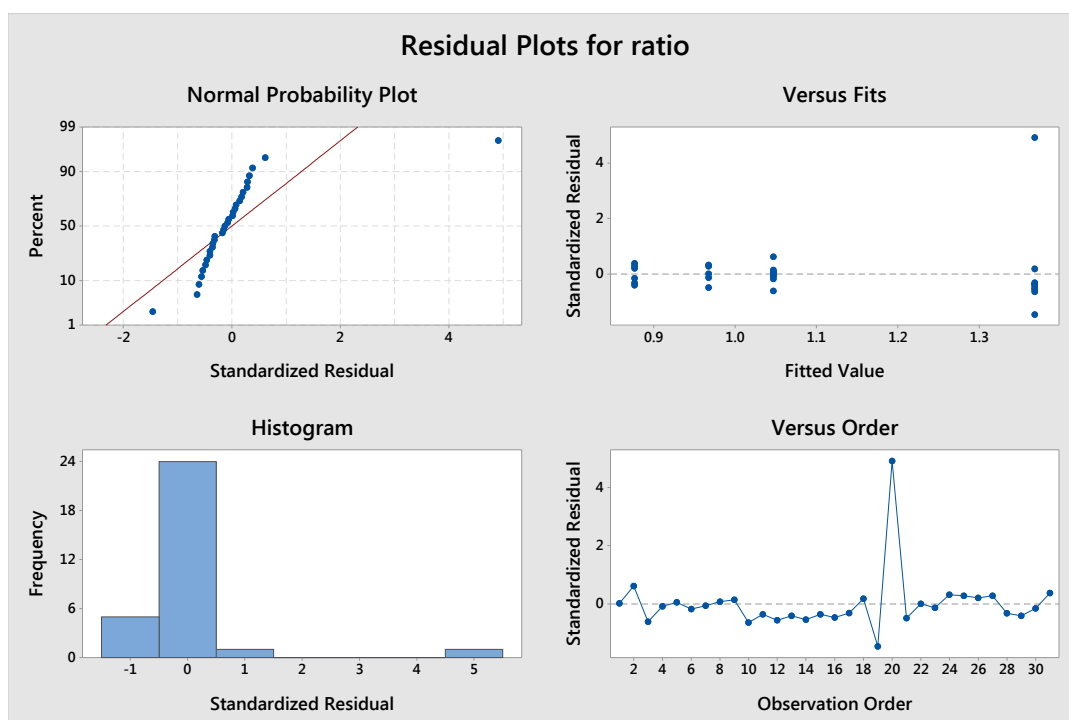
“We are interested in whether the Visiplume instrument, relative to the standard method, performs the same across all experimental conditions, between the coal-fired plants, and between the generated smoke and the real coal-fired smoke. Additionally, do the Visiplume recordings seem to differ from those of the standard reading for any of the four experimental conditions?”

Stata listing (extract) for Question 1, part A):

```
. oneway ratio exp_cond, tab
```

exp_cond	Summary of ratio		
	Mean	Std. Dev.	Freq.
1	1.047	.1172156	9
2	1.3684545	.62027482	11
3	.96720001	.11473973	5
4	.87616667	.11862448	6
Total	1.115129	.41952605	31

Source	Analysis of Variance				F	Prob > F
	SS	df	MS			
Between groups	1.1997191	3	.399906366	2.65	0.0693	
Within groups	4.08034417	27	.151123858			
Total	5.28006327	30	.176002109			



Stata listing (extract) for Question 1, part B):

```
. xi: boxcox ratio i.exp_cond
i.exp_cond      _Iexp_cond_1-4      (naturally coded; _Iexp_cond_1 omitted)
...
Number of obs   =          31
LR chi2(3)      =          17.60
Prob > chi2     =          0.001
Log likelihood = 10.800206
```

```
-----+-----
      ratio |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
    /theta | -1.571682   .4543732    -3.46  0.001   -2.462237   -.6811265
-----+-----
```

```
-----+-----
      Test           Restricted      LR statistic      P-value
      H0:      log likelihood      chi2      Prob > chi2
-----+-----
theta = -1      9.8950442           1.81           0.178
theta = 0      2.6253124           16.35          0.000
theta = 1     -12.556104           46.71          0.000
-----+-----
```

```
. gen tratio=ratio^(-1.5)
. oneway tratio exp_cond, tab
```

```
-----+-----
      |           Summary of tratio
exp_cond |           Mean   Std. Dev.     Freq.
-----+-----
      1 |     .9540181   .17052966           9
      2 |     .74392347  .26592019          11
      3 |     1.0755431  .20724611           5
      4 |     1.2556413  .25940834           6
-----+-----
      Total |     .95744788  .29171443          31
-----+-----
```

```
-----+-----
                        Analysis of Variance
Source                SS           df      MS           F      Prob > F
-----+-----
Between groups      1.10487355           3   .368291184     6.87     0.0014
Within groups      1.44804564           27  .05363132
-----+-----
Total                2.55291919           30  .085097306
-----+-----
```

```
. gen ratiox1=ratio if ratio<3
(1 missing value generated)
```

```
. xi: boxcox ratiox1 i.exp_cond
i.exp_cond      _Iexp_cond_1-4      (naturally coded; _Iexp_cond_1 omitted)
...
Number of obs   =          30
LR chi2(3)      =          19.54
Prob > chi2     =          0.000
Log likelihood = 20.837215
```

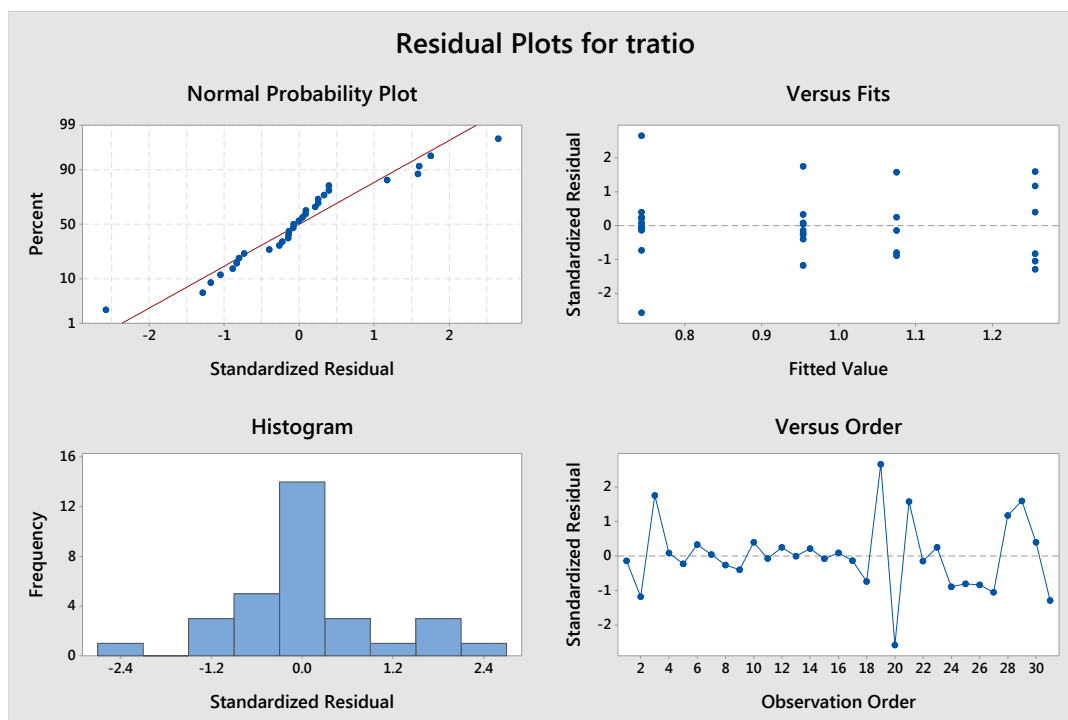
```
-----+-----
      ratio |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
    /theta |  1.193858   .9156761     1.30  0.192   -.6008343   2.98855
-----+-----
```

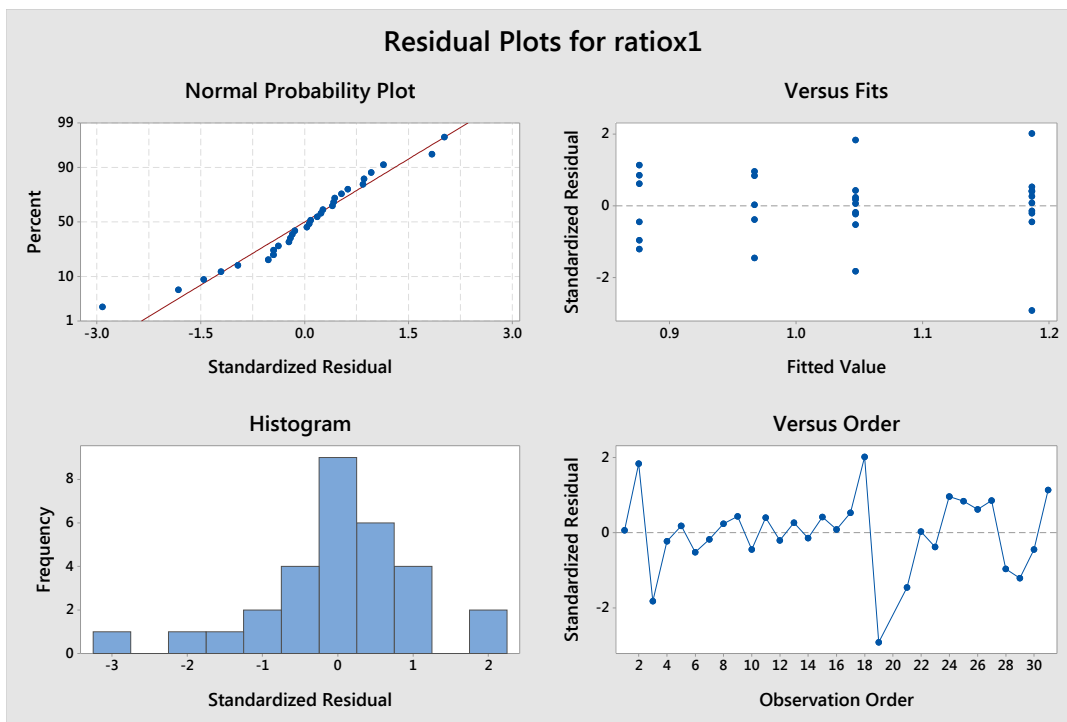
Test	Restricted	LR statistic	P-value
H0:	log likelihood	chi2	Prob > chi2
theta = -1	18.169343	5.34	0.021
theta = 0	20.01679	1.64	0.200
theta = 1	20.814913	0.04	0.833

. oneway ratiox1 exp_cond, tab

exp_cond	Summary of ratio		
	Mean	Std. Dev.	Freq.
1	1.047	.1172156	9
2	1.1865	.15113954	10
3	.96720001	.11473973	5
4	.87616667	.11862448	6
Total	1.0460333	.17019693	30

Source	Analysis of Variance				
	SS	df	MS	Prob > F	
Between groups	.401518822	3	.133839607	7.94	0.0006
Within groups	.438524081	26	.016866311		
Total	.840042903	29	.028966997		





Question 2.

Polychlorinated biphenyls (PCBs) are a collection of compounds that are no longer produced in the United States but are still found in the environment. Evidence suggests they can cause harmful health effects when consumed. Because PCBs can accumulate in fish, efforts have been made to identify areas where fish contain excessive amounts, so that recommendations concerning consumption limits can be made. There are over 200 types of PCBs. We consider data from a study across various lakes in the United States where the amounts of PCBs in fish were measured. The data include 69 sites (for which further demographic information is available, including their location) and measures of 7 specific PCB compounds as well as the PCB total (only available with a comprehensive and expensive panel of analyses), all measured in parts per billion (ppb).

Not all types of PCBs are equally harmful. A scale has been developed to convert the raw amount of each type of PCB to a toxic equivalent (TEQ). There are four toxic equivalent variables in the data: one for the total TEQ from all PCBs found in the samples, similar ones for dioxin and furans, and finally a total TEQ computed as the sum of these three variables. Some descriptive information about the variables is shown in the Stata listings below.

```
. list in 1/5
```

pcb28	pcb52	pcb118	pcb126	pcb138	pcb153	pcb180	pcb	teqpcb	teqdioxin	teqfuran	teq
.421	.532	.720	0	1.46	1.59	.738	19.20	.189	.609	.14	.938
.025	.030	.236	0	.64	.92	.664	6.01	.068	.900	0	.969
.076	.134	1.54	0	3.29	3.9	1.15	24.97	.330	.620	.03	.980
.152	.466	1.94	.0055	3.94	5.44	1.33	37.44	.924	.005	.07	.999
.116	.243	1.47	.0059	3.18	3.65	2.14	30.18	.927	0	.09	1.017

```
. tabstat pcb28-teq, stat(mean sd min p25 p50 p75 max skewness) c(stat)
```

Variable	Mean	SD	Min	p25	p50	p75	Max	Skewness
pcb28	.5264638	.9484019	.006	.126	.262	.432	6.93	4.900894
pcb52	.9580435	1.59828	.02	.228	.477	.892	9.06	3.578909
pcb118	3.256319	3.019118	.236	1.49	2.42	3.89	18.9	2.724947
pcb126	.0097159	.0082499	0	.0054	.0074	.0147	.0317	.7694042
pcb138	6.826812	5.862651	.64	3.18	4.92	8.65	32.3	2.054121
pcb153	8.274783	8.679789	.86	3.3	5.44	9.87	43.5	2.190855
pcb180	4.15842	4.986439	.395	1.24	2.69	4.49	31.5	3.216147
pcb	68.46737	59.3906	6.0996	30.183	47.9596	91.6305	318.7461	1.989725
teqpcb	1.718211	1.342831	.06837	.9235	1.2633	2.39235	5.41694	1.028463
teqdioxin	.7127181	.5773875	0	.222	.64353	1.005	3.24836	1.517973
teqfuran	.238813	.2046309	0	.113	.1897	.2985	1.126	2.308835
teq	2.669744	1.634529	.9384	1.35751	2.06698	3.73025	6.49311	.8727413

- A) (5 points) Explain the analysis whose results are shown in the Stata listing below and are also illustrated in the two figures obtained in Stata using the commands indicated. Interpret the results, and describe for what objective(s) the analysis might be informative or useful. Indicate briefly what additional analyses it might be of interest to carry out from the results obtained, for each objective you included above. Finally, discuss whether the analysis in your view has been carried out in the right way or you would want to suggest any changes to improve its validity or usefulness.
- b) (5 points) Describe some (at least two) multivariable or multivariate analyses that it would be relevant to perform for these data, excluding the same (type of) analysis as in A). Make sure to describe clearly what the specific purpose of each analysis would be and how the results, including any graphical representations, from the analysis would contribute to the understanding of the data or constitute a relevant use of the data. For each analysis, indicate (briefly) relevant assumptions, settings and analytical steps.

Stata listing and graphs for Question 2:

```
. pca pcb28-pcb
```

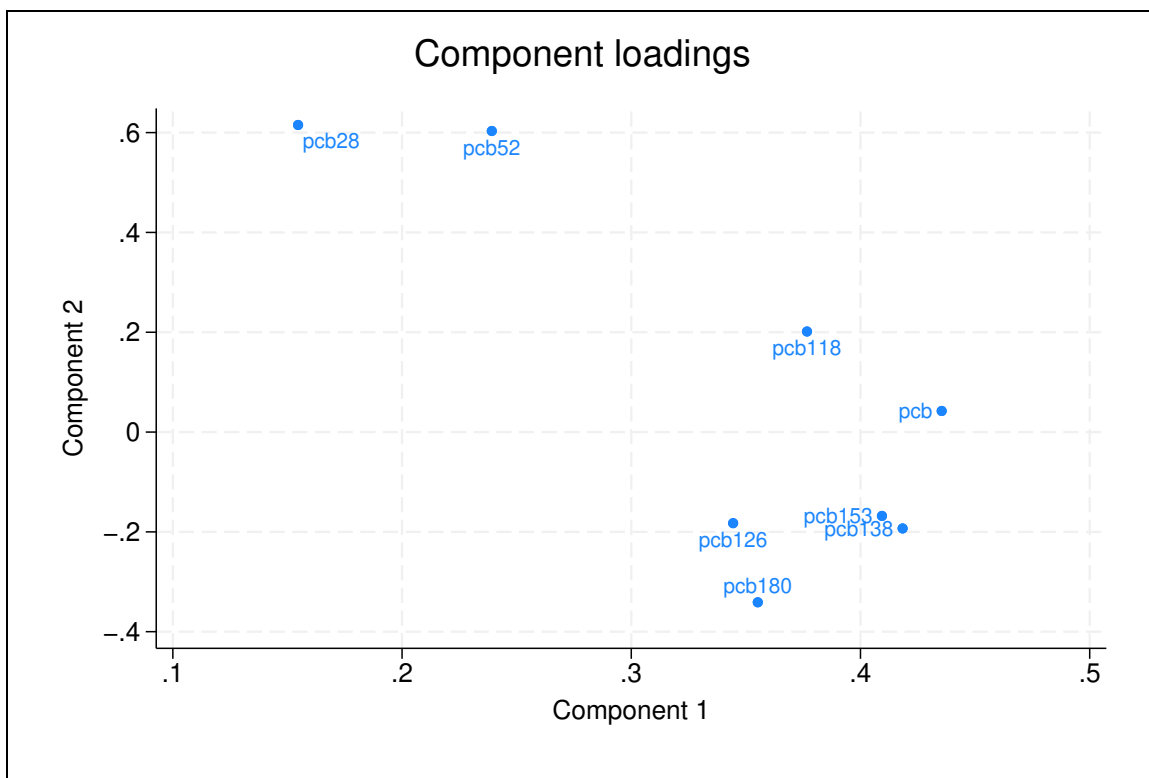
```
Principal components/correlation          Number of obs   =          69
                                           Number of comp. =           8
                                           Trace           =           8
Rotation: (unrotated = principal)        Rho             =        1.0000
```

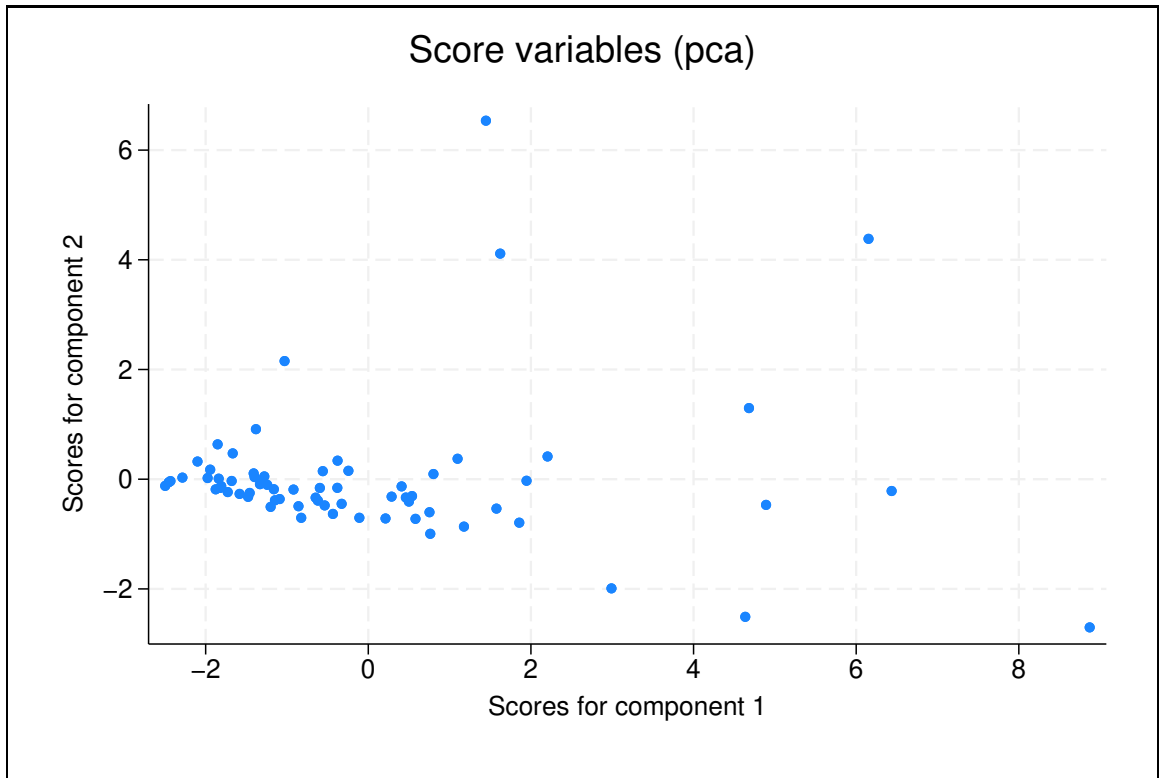
Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	5.15691	3.49499	0.6446	0.6446
Comp2	1.66191	1.11889	0.2077	0.8524
Comp3	.54302	.140193	0.0679	0.9202
Comp4	.402827	.270608	0.0504	0.9706
Comp5	.132218	.0727049	0.0165	0.9871
Comp6	.0595135	.0223618	0.0074	0.9945
Comp7	.0371517	.0307008	0.0046	0.9992
Comp8	.00645083	.	0.0008	1.0000

Principal components (eigenvectors)

Variable	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6	Comp7	Comp8	Unexpl
pcb28	0.1546	0.6153	0.4928	0.5067	-0.3013	0.0424	0.0669	0.0295	0
pcb52	0.2392	0.6032	-0.1358	-0.3083	0.6202	0.0647	-0.1223	0.2483	0
pcb118	0.3767	0.2015	-0.5078	-0.2005	-0.5030	-0.2990	0.3934	0.1492	0
pcb126	0.3444	-0.1826	-0.4589	0.7126	0.2931	0.1990	0.0632	-0.0071	0
pcb138	0.4184	-0.1931	0.0825	0.0294	-0.1763	-0.3074	-0.7522	0.2978	0
pcb153	0.4094	-0.1681	0.1543	-0.2788	-0.2170	0.8015	0.0424	0.1066	0
pcb180	0.3552	-0.3411	0.4822	-0.0282	0.3166	-0.3304	0.4992	0.2562	0
pcb	0.4354	0.0422	0.0841	-0.1441	0.0848	-0.1380	-0.0713	-0.8657	0

. loadingplot
 . scoreplot





Question 3.

A study was carried out at a University Medical Centre (Instititue of Oncology, Ljublana, Yugoslavia) with the purpose of gaining insights into prognostic factors for a successful outcome in the treatment of breast cancer in women. The data comprise 286 women who were diagnosed and treated for breast cancer, and the outcome of interest is recurrence of the tumour after treatment. A total of nine variables representing demographics of the women and information about the tumour are included and briefly summarized in the table below.

Variable	Description	Values
recur	recurrence of tumour	0 (no) / 1 (yes)
malig	degree of malignancy	categories: 1, 2, 3 ~ increasing malignancy
age	age in years	intervals: 10–19, 20–29, ..., 80–89, 90–99
menop	time of menopause	1 (≥ 40), 2 (< 40), 3 (prior to menopause)
tsize	tumour size (<i>mm</i>)	intervals: 0–4, 5–9, ..., 50–54, 55–59
invnodes	number of lymph nodes involved	intervals: 0–2, 3–5, ..., 33–35, 36–39
nodecaps	presence of nodal caps	no/yes
breast	affected breast	categories: left, right
tloc	tumour location	categories: left-up, left-low, right-up, right-low, central
irrad	radiation therapy	no/yes

To further describe the data, a list of the first 10 observations in the dataset and a set of cross-tabulations are provided, from Stata 18 statistical software.

```
. list recur-irrad in 1/10, sep(10)
```

```

+-----+
| recur  malig   age   menop  tsize  invnodes  nodecaps  breast  tloc  irrad |
+-----+
|      0      3  30-39  premeno  30-34      0-2        no    left  left_low  no |
|      0      2  40-49  premeno  20-24      0-2        no    right right_up  no |
|      0      2  40-49  premeno  20-24      0-2        no    left  left_low  no |
|      0      2  60-69    ge40   15-19      0-2        no    right  left_up  no |
|      0      2  40-49  premeno   0-4      0-2        no    right right_low  no |
|      0      2  60-69    ge40   15-19      0-2        no    left  left_low  no |
|      0      2  50-59  premeno  25-29      0-2        no    left  left_low  no |
|      0      1  60-69    ge40   20-24      0-2        no    left  left_low  no |
|      0      2  40-49  premeno  50-54      0-2        no    left  left_low  no |
|      0      2  40-49  premeno  20-24      0-2        no    right  left_up  no |
+-----+

```

```
. table recur malig
```

```

-----
|          |          malig
|          |          1    2    3    Total
-----+-----
recur     |
  0       |  59   102   40    201
  1       |  12    28   45     85
Total    |  71   130   85    286
-----

```

. table recur age, nototal

	age					
	20-29	30-39	40-49	50-59	60-69	70-79
recur						
0	1	21	63	71	40	5
1	0	15	27	25	17	1

. table recur menop, nototal

	menop		
	ge40	lt40	premeno
recur			
0	94	5	102
1	35	2	48

. table recur tsize, nototal

	tsize											
	0-4	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	5-9	50-54	
recur												
0	7	27	23	34	36	35	12	16	2	4	5	
1	1	1	7	16	18	25	7	6	1	0	3	

. table recur invnodes, nototal

	invnodes						
	0-2	12-14	15-17	24-26	3-5	6-8	9-11
recur							
0	167	1	3	0	19	7	4
1	46	2	3	1	17	10	6

. table recur nodecaps, nototal

	nodecaps		
	?	no	yes
recur			
0	5	171	25
1	3	51	31

(continues on the next page)

```
. table recur breast, nottotal
```

	breast	
	left	right
recur		
0	103	98
1	49	36

```
. table recur tloc, nottotal
```

		tloc				
	?	central	left_low	left_up	right_low	right_up
recur						
0	0	17	75	71	18	20
1	1	4	35	26	6	13

```
. table recur irrad, nottotal
```

	irrad	
	no	yes
recur		
0	164	37
1	54	31

- A) (3 points) Explain the statistical models used in the two Stata listings presented for Question 3.A; for each, you may either give a model formula or a verbal description of the model, including the outcome/parameter being modelled and the predictor variables involved (and their assumed relationship with the outcome). Regardless of any concerns you might have about the validity of the models and results (which you will get the opportunity to discuss in B) below), interpret the estimates obtained in both models for the predictor `malig`. Your interpretation should include both a quantitative statement of the size of the effect and a statement about its statistical significance.
- B) (4 points) Briefly review the assumptions behind the two statistical models. Next, discuss for each of the two models, any issues you might have about the validity and/or adequacy of the results. For each issue you identify, include also a (brief) suggestion for how to deal with it.
- C) (3 points) Use the information provided in all the Stata listings to outline a strategy for developing a good predictive model for recurrence of breast cancer from these data. Explain and motivate the different steps you would go through in developing the model, while explaining how you use the information from the analyses already carried out in the process.

Stata listing for Question 3.A:

```
. logit recur i.malig i.age i.menop i.tsize i.invnode i.nodecaps i.breast i.tloc i.irrad
```

```
note: 1.age != 0 predicts failure perfectly;
      1.age omitted and 1 obs not used.
```

```
note: 10.tsize != 0 predicts failure perfectly;
      10.tsize omitted and 4 obs not used.
```

```
note: 4.invnodes != 0 predicts success perfectly;
      4.invnodes omitted and 1 obs not used.
```

```
note: 1.tloc != 0 predicts success perfectly;
      1.tloc omitted and 1 obs not used.
```

```
note: 6.age omitted because of collinearity.
note: 6.tloc omitted because of collinearity.
Iteration 0: Log likelihood = -169.83384
Iteration 1: Log likelihood = -140.3533
Iteration 2: Log likelihood = -138.19205
Iteration 3: Log likelihood = -138.07837
Iteration 4: Log likelihood = -138.07793
Iteration 5: Log likelihood = -138.07793
```

Logistic regression

```
Number of obs = 279
LR chi2(30) = 63.51
Prob > chi2 = 0.0003
Pseudo R2 = 0.1870
```

Log likelihood = -138.07793

recur	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
-----+-----						
malig						
2	-.3764823	.4538131	-0.83	0.407	-1.26594	.5129751
3	.9709063	.4578884	2.12	0.034	.0734614	1.868351
age						
20-29	0	(empty)				
30-39	.3975972	1.389702	0.29	0.775	-2.326169	3.121363
40-49	-.019551	1.347575	-0.01	0.988	-2.66075	2.621648
50-59	-.1261622	1.292141	-0.10	0.922	-2.658711	2.406387
60-69	.6241452	1.312171	0.48	0.634	-1.947663	3.195953
70-79	0	(omitted)				
menop						
lt40	.756859	.9736399	0.78	0.437	-1.15144	2.665158
premeno	.6334752	.5016704	1.26	0.207	-.3497806	1.616731
tsize						
10-14	-2.195231	1.666422	-1.32	0.188	-5.461359	1.070897
15-19	-.0420423	1.309794	-0.03	0.974	-2.609192	2.525107
20-24	.3642982	1.252763	0.29	0.771	-2.091072	2.819669
25-29	.2614743	1.266136	0.21	0.836	-2.220107	2.743056
30-34	.2407442	1.258719	0.19	0.848	-2.226299	2.707788
35-39	-.0742186	1.372649	-0.05	0.957	-2.764562	2.616125
40-44	-.2712883	1.355841	-0.20	0.841	-2.928687	2.38611
45-49	-.4483498	1.821691	-0.25	0.806	-4.018799	3.1221

5-9		0	(empty)				
50-54		.8835831	1.461422	0.60	0.545	-1.980751	3.747918
invnodes							
12-14		.9152215	1.447415	0.63	0.527	-1.921661	3.752104
15-17		.5552194	1.012261	0.55	0.583	-1.428775	2.539214
24-26		0	(empty)				
3-5		.7434623	.5176915	1.44	0.151	-.2711943	1.758119
6-8		1.065333	.7006924	1.52	0.128	-.3079992	2.438665
9-11		1.185844	.8641259	1.37	0.170	-.5078121	2.879499
nodecaps							
no		.2307829	.9819815	0.24	0.814	-1.693865	2.155431
yes		.6352681	.9980074	0.64	0.524	-1.32079	2.591327
breast							
right		-.3995402	.3394561	-1.18	0.239	-1.064862	.2657816
tloc							
?		0	(empty)				
central		-.9576701	.8111159	-1.18	0.238	-2.547428	.6320878
left_low		-.5550162	.5029879	-1.10	0.270	-1.540854	.430822
left_up		-.7128619	.4883703	-1.46	0.144	-1.67005	.2443263
right_low		-1.034276	.7658016	-1.35	0.177	-2.53522	.4666673
right_up		0	(omitted)				
irrad							
yes		.3763273	.3792367	0.99	0.321	-.366963	1.119618
_cons		-1.504909	1.850158	-0.81	0.416	-5.131152	2.121334

. logit recur malig age menop tsize invnode nodecaps breast tloc irrad

Iteration 0: Log likelihood = -174.02401
Iteration 1: Log likelihood = -152.05447
Iteration 2: Log likelihood = -151.32883
Iteration 3: Log likelihood = -151.32617
Iteration 4: Log likelihood = -151.32617

Logistic regression

Number of obs = 286
LR chi2(9) = 45.40
Prob > chi2 = 0.0000
Pseudo R2 = 0.1304

Log likelihood = -151.32617

recur		Coefficient	Std. err.	z	P> z	[95% conf. interval]
malig		.765511	.2206747	3.47	0.001	.3329965 1.198025
age		.0414884	.2037459	0.20	0.839	-.3578463 .4408232
menop		.209243	.2106346	0.99	0.321	-.2035932 .6220792
tsize		.0673424	.069434	0.97	0.332	-.0687457 .2034305
invnodes		.19264	.0815101	2.36	0.018	.0328831 .3523969
nodecaps		.2325435	.3475202	0.67	0.503	-.4485837 .9136706
breast		-.3291949	.2944294	-1.12	0.264	-.9062659 .247876
tloc		.0559859	.1332823	0.42	0.674	-.2052426 .3172144

```

      irradi |   .3401894   .3309041    1.03   0.304   -.3083708   .9887497
      _cons |  -4.564675   1.514391   -3.01   0.003   -7.532826  -1.596523
-----

```

Stata listing for Question 3.C:

```

. stepwise, pr(0.05): logit recur (i.malign) (i.age) (i.menop) (i.tsize) (i.invnode)
(i.nodcaps) (i.breast) (i.tloc) (i.irradi)

```

```

note: 1b.malign omitted because of estimability.
note: 1b.age omitted because of estimability.
note: 6.age omitted because of estimability.
note: 1b.menop omitted because of estimability.
note: 1b.tsize omitted because of estimability.
note: 10.tsize omitted because of estimability.
note: 1b.invnodes omitted because of estimability.
note: 4.invnodes omitted because of estimability.
note: 1b.nodcaps omitted because of estimability.
note: 1b.breast omitted because of estimability.
note: 1b.tloc omitted because of estimability.
note: 6.tloc omitted because of estimability.
note: 1b.irradi omitted because of estimability.
note: 7 obs omitted because of estimability.

```

Wald test, begin with full model:

```

p = 0.6421 >= 0.0500, removing 2.tsize 3.tsize 4.tsize 5.tsize 6.tsize 7.tsize 8.tsize
9.tsize 11.tsize
p = 0.6601 >= 0.0500, removing 2.nodcaps 3.nodcaps
p = 0.6202 >= 0.0500, removing 2.age 3.age 4.age 5.age
p = 0.5232 >= 0.0500, removing 2.tloc 3.tloc 4.tloc 5.tloc
p = 0.4102 >= 0.0500, removing 2.breast
p = 0.3100 >= 0.0500, removing 2.irradi
p = 0.1766 >= 0.0500, removing 2.menop 3.menop

```

Logistic regression

```

Number of obs =   279
LR chi2(7)     =   41.04
Prob > chi2    =   0.0000
Pseudo R2     =   0.1208

```

Log likelihood = -149.31629

```

-----
      recur | Coefficient  Std. err.      z    P>|z|    [95% conf. interval]
-----+-----
      malign |
      2 |   .0281144   .3987065    0.07   0.944   -.7533359   .8095647
      3 |    1.2504    .409881    3.05   0.002   .4470482   2.053752
      |
      invnodes |
      12-14 |    1.09299   1.253176    0.87   0.383   -1.36319   3.54917
      15-17 |    .5902411   .8689519    0.68   0.497   -1.112873   2.293356
      3-5 |    1.071937   .3956434    2.71   0.007   .2964899   1.847384
      6-8 |    1.286473   .5531134    2.33   0.020   .2023909   2.370555
      9-11 |    1.457992   .6989154    2.09   0.037   .0881433   2.827841
      |
      _cons |  -1.650243   .3226307   -5.11   0.000   -2.282588  -1.017898
-----

```