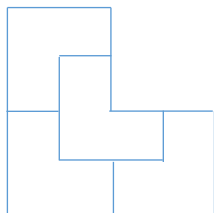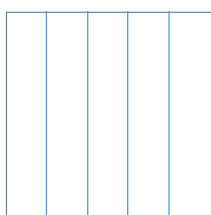## Chapter 1: Exercises – answers
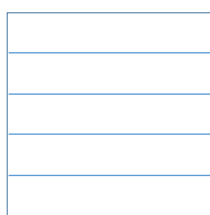
1) Divide the following shape into *four* identical shapes.

Divide the following shape into *five* identical shapes.

or

This exercise shows that it is sometimes difficult to come up with a solution to an easy task if we are caught up in certain patterns of thinking. If you followed the instructions, it very likely took you some time to solve the first, moderately difficult puzzle, which has only one solution. Paradoxically, after solving this puzzle it is much more difficult to come up with the solution to the second puzzle which has a trivial solution and which, if presented on its own, would take only seconds to solve.

2) The mean of the numbers is 2124.18.

3) A model is a simplification of complex reality that enables us to conceptualize of and analyse reality more easily. Despite being a simplification, the model must capture the main features of the reality that it represents – we often talk about the best-fitting model that needs to be chosen. In corpus statistics, we use mathematical models to analyze complex linguistic reality.

4) Select the best-fitting geometrical model for the area of Great Britain.

   a) triangle △

Use the model to calculate the area of Great Britain.

The area of a triangle is calculated as:

$$A = \frac{base \times hight}{2} = \frac{520 \times 900}{2} = 234{,}000 \text{ km}^2$$

This is very close to the actual area of Great Britain which is 229,848 km^2. We can therefore say that the model we used in this exercise (triangle) is a good model.

5) The answers to the quiz are: 1b); 2a); 3b); 4b); 5a); 6c); 7c); 8b)

6) –

7) and 9) The following table shows the male (M) and female (F) speakers in the population. The speakers in shaded cells are older speakers.

text_1	F	text_39	F	text_77	M	text_115	F	text_153	M	text_191	M
text_2	F	text_40	F	text_78	M	text_116	F	text_154	M	text_192	M
text_3	F	text_41	F	text_79	M	text_117	F	text_155	M	text_193	M
text_4	F	text_42	F	text_80	M	text_118	F	text_156	M	text_194	M
text_5	F	text_43	F	text_81	M	text_119	F	text_157	M	text_195	M
text_6	F	text_44	F	text_82	M	text_120	F	text_158	M	text_196	M
text_7	F	text_45	F	text_83	M	text_121	F	text_159	M	text_197	M
text_8	F	text_46	F	text_84	M	text_122	F	text_160	M	text_198	M
text_9	F	text_47	F	text_85	M	text_123	F	text_161	M	text_199	M
text_10	F	text_48	F	text_86	M	text_124	F	text_162	M	text_200	M
text_11	F	text_49	F	text_87	M	text_125	F	text_163	M	text_201	F
text_12	F	text_50	F	text_88	M	text_126	F	text_164	M	text_202	F
text_13	F	text_51	M	text_89	M	text_127	F	text_165	M	text_203	F
text_14	F	text_52	M	text_90	M	text_128	F	text_166	M	text_204	F
text_15	F	text_53	M	text_91	M	text_129	F	text_167	M	text_205	F
text_16	F	text_54	M	text_92	M	text_130	F	text_168	M	text_206	F
text_17	F	text_55	M	text_93	M	text_131	F	text_169	M	text_207	F
text_18	F	text_56	M	text_94	M	text_132	F	text_170	M	text_208	F
text_19	F	text_57	M	text_95	M	text_133	F	text_171	M	text_209	F
text_20	F	text_58	M	text_96	M	text_134	F	text_172	M	text_210	F
text_21	F	text_59	M	text_97	M	text_135	F	text_173	M	text_211	F
text_22	F	text_60	M	text_98	M	text_136	F	text_174	M	text_212	F
text_23	F	text_61	M	text_99	M	text_137	F	text_175	M	text_213	F
text_24	F	text_62	M	text_100	M	text_138	F	text_176	M	text_214	F
text_25	F	text_63	M	text_101	F	text_139	F	text_177	M	text_215	F
text_26	F	text_64	M	text_102	F	text_140	F	text_178	M	text_216	F
text_27	F	text_65	M	text_103	F	text_141	F	text_179	M	text_217	F
text_28	F	text_66	M	text_104	F	text_142	F	text_180	M	text_218	F
text_29	F	text_67	M	text_105	F	text_143	F	text_181	M	text_219	F
text_30	F	text_68	M	text_106	F	text_144	F	text_182	M	text_220	F
text_31	F	text_69	M	text_107	F	text_145	F	text_183	M	text_221	F
text_32	F	text_70	M	text_108	F	text_146	F	text_184	M	text_222	F
text_33	F	text_71	M	text_109	F	text_147	F	text_185	M	text_223	F
text_34	F	text_72	M	text_110	F	text_148	F	text_186	M	text_224	F
text_35	F	text_73	M	text_111	F	text_149	F	text_187	M	text_225	F
text_36	F	text_74	M	text_112	F	text_150	F	text_188	M	text_226	F
text_37	F	text_75	M	text_113	F	text_151	M	text_189	M	text_227	F
text_38	F	text_76	M	text_114	F	text_152	M	text_190	M	text_228	F

Lancaster University

text_229	F	text_282	M	text_335	F	text_388	M	text_441	F	text_494	M
text_230	F	text_283	M	text_336	F	text_389	M	text_442	F	text_495	M
text_231	F	text_284	M	text_337	F	text_390	M	text_443	F	text_496	M
text_232	F	text_285	M	text_338	F	text_391	M	text_444	F	text_497	M
text_233	F	text_286	M	text_339	F	text_392	M	text_445	F	text_498	M
text_234	F	text_287	M	text_340	F	text_393	M	text_446	F	text_499	M
text_235	F	text_288	M	text_341	F	text_394	M	text_447	F	text_500	M
text_236	F	text_289	M	text_342	F	text_395	M	text_448	F		
text_237	F	text_290	M	text_343	F	text_396	M	text_449	F		
text_238	F	text_291	M	text_344	F	text_397	M	text_450	F		
text_239	F	text_292	M	text_345	F	text_398	M	text_451	M		
text_240	F	text_293	M	text_346	F	text_399	M	text_452	M		
text_241	F	text_294	M	text_347	F	text_400	M	text_453	M		
text_242	F	text_295	M	text_348	F	text_401	F	text_454	M		
text_243	F	text_296	M	text_349	F	text_402	F	text_455	M		
text_244	F	text_297	M	text_350	F	text_403	F	text_456	M		
text_245	F	text_298	M	text_351	M	text_404	F	text_457	M		
text_246	F	text_299	M	text_352	M	text_405	F	text_458	M		
text_247	F	text_300	M	text_353	M	text_406	F	text_459	M		
text_248	F	text_301	F	text_354	M	text_407	F	text_460	M		
text_249	F	text_302	F	text_355	M	text_408	F	text_461	M		
text_250	F	text_303	F	text_356	M	text_409	F	text_462	M		
text_251	M	text_304	F	text_357	M	text_410	F	text_463	M		
text_252	M	text_305	F	text_358	M	text_411	F	text_464	M		
text_253	M	text_306	F	text_359	M	text_412	F	text_465	M		
text_254	M	text_307	F	text_360	M	text_413	F	text_466	M		
text_255	M	text_308	F	text_361	M	text_414	F	text_467	M		
text_256	M	text_309	F	text_362	M	text_415	F	text_468	M		
text_257	M	text_310	F	text_363	M	text_416	F	text_469	M		
text_258	M	text_311	F	text_364	M	text_417	F	text_470	M		
text_259	M	text_312	F	text_365	M	text_418	F	text_471	M		
text_260	M	text_313	F	text_366	M	text_419	F	text_472	M		
text_261	M	text_314	F	text_367	M	text_420	F	text_473	M		
text_262	M	text_315	F	text_368	M	text_421	F	text_474	M		
text_263	M	text_316	F	text_369	M	text_422	F	text_475	M		
text_264	M	text_317	F	text_370	M	text_423	F	text_476	M		
text_265	M	text_318	F	text_371	M	text_424	F	text_477	M		
text_266	M	text_319	F	text_372	M	text_425	F	text_478	M		
text_267	M	text_320	F	text_373	M	text_426	F	text_479	M		
text_268	M	text_321	F	text_374	M	text_427	F	text_480	M		
text_269	M	text_322	F	text_375	M	text_428	F	text_481	M		
text_270	M	text_323	F	text_376	M	text_429	F	text_482	M		
text_271	M	text_324	F	text_377	M	text_430	F	text_483	M		
text_272	M	text_325	F	text_378	M	text_431	F	text_484	M		
text_273	M	text_326	F	text_379	M	text_432	F	text_485	M		
text_274	M	text_327	F	text_380	M	text_433	F	text_486	M		
text_275	M	text_328	F	text_381	M	text_434	F	text_487	M		
text_276	M	text_329	F	text_382	M	text_435	F	text_488	M		
text_277	M	text_330	F	text_383	M	text_436	F	text_489	M		
text_278	M	text_331	F	text_384	M	text_437	F	text_490	M		
text_279	M	text_332	F	text_385	M	text_438	F	text_491	M		
text_280	M	text_333	F	text_386	M	text_439	F	text_492	M		
text_281	M	text_334	F	text_387	M	text_440	F	text_493	M		

Lancaster University

11) Main types of bias: Text sample bias, Topic bias, Non-coverage bias, Traditional text type bias, Legal considerations bias, Practicality bias, Self-selection bias

12) Appropriate research design

   a) Linguistic feature design

   b) Individual-texts/speakers design

   c) Whole corpus design or Individual-texts/speakers design

   d) Individual-texts/speakers design

13) The errors are highlighted in the table below.

Word or expression	Frequency	Frequency per million
the	5,896	5,142.17
of	30,666	26,745.23
and	27,909	24,340.72
to	26,188	2,283.98
of the	6,887	60,06.47
and the	19,530	17,033.01
Words total	2,293,194	

Explanation: 1) and 2) The frequency of the definite article in written corpora is usually between 5 and 6% of all running words. The correct numbers should therefore be 58,960 and 51,421.7. 3) The frequency of the preposition *to* per million is incorrect; it should be 22,839.8. 4) and 5) Because combinations of words are much less frequent than their individual occurrence the combination *and the* is more likely to have the frequencies in the thousands rather than tens of thousands; the correct numbers are: 1,953 and 1,703.3 6) BE06 is approximately a one-million-word corpus. The total word count displayed in the table includes part-of-speech tags; the correct wordcount should be 1,146,597.

14) Which visualization type (graph) would be appropriate in the following situations?

   a) Histogram

   b) Box plots

   c) Scatter plot

   d) Error bars
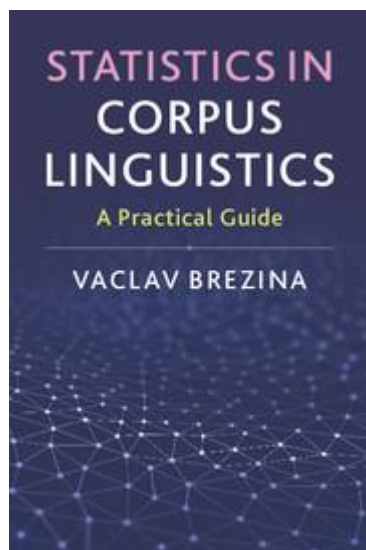
Lancaster University

Materials from Brezina, V. (2018). *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge: Cambridge University Press.

Brezina, V. (2018). *Statistics in Corpus Linguistics: A Practical Guide.* Cambridge: Cambridge University Press.

Do you use language corpora in your research or study, but find that you struggle with statistics? This practical introduction will equip you to understand the key principles of statistical thinking and apply these concepts to your own research, without the need for prior statistical knowledge. The book gives step-by-step guidance through the process of statistical analysis and provides multiple examples of how statistical techniques can be used to analyse and visualise linguistic data. It also includes a useful selection of discussion questions and exercises which you can use to check your understanding.

The book comes with a Companion website, which provides additional materials (answers to exercises, datasets, advanced materials, teaching slides etc.) and Lancaster Stats Tools online, a free click-and-analyse statistical tool for easy calculation of the statistical measures discussed in the book.