Question 1. A study evaluated the effect of 8-hour nightly continuous positive airway pressure (CPAP) treatment on fasting insulin levels in patients with prediabetes and suffering from obstructive sleep apnoea. In the study 39 patients were randomly assigned to either the 8-hour nightly CPAP treatment group or an oral placebo group. The fasting insulin level was measured for each patient at the beginning of the study and after two weeks treatment.

(a) At the beginning of the study each patient's percentage of body fat was measured. The 26 patients in the CPAP treatment group had an average body fat percentage of 36.8% and a standard deviation of 7.8% while the 13 patients in the oral placebo group had an average body fat percentage of 32.7% and a standard deviation of 4.3%. Is there any evidence of a difference in the population mean body fat percentage between the two groups? State the null and alternative hypotheses, and use an appropriate test statistic to determine the P-value. What do you conclude?

(b) After the two week trial, the 26 subjects in the CPAP group experienced an average decrease in fasting insulin level of 5.7 (pmol/L) with a standard deviation of 26.7 (pmol/L). Does this give any evidence of a decrease in mean fasting insulin level for patients receiving the CPAP treatment? State the null and alternative hypotheses, and use an appropriate test statistic to determine the P-value. What do you conclude?

Question 2. A study was made to identify characteristics of children presenting at two hospitals (Casey and Clayton) suffering accidental poisoning. Patients aged under 13 and admitted to the emergency department of one of the hospitals during the period 1 July 2009 and 30 June 2012 were included in the study. Of the 1248 patients admitted to Clayton hospital the source of the poison in 496 cases was household cleaning products while of the 1057 patients admitted to Casey hospital, household cleaning products was the source in 406 cases.

(a) Overall, what proportion of accidental poisoning cases did household cleaning products cause?

(b) What is the difference between the two hospitals in the proportions of poisonings caused by household cleaning products?

(c) Give a 95% confidence interval for the true difference in the proportions of children accidentally poisoned with household cleaning products between the two hospitals. What does this interval say about accidental poisoning rates of children due to household cleaning products in the catchment areas of the two hospitals?

(d) The following table gives the summary of the major emergency department discharge categories for children presenting at the two hospitals in the study with accidental poisoning.

	Discharged to usual residence directly	Admitted to a ward of the same hospital	Transferred to a different hospital
Clayton	317	596	144
Casey	361	754	133

Based on this table, is there evidence of an association between hospital and discharge category of children?

Question 3. Disappointed by an earlier experiment, a poultry researcher has developed a new diet for improving weight gain in week-old chicks. She conducts a study to investigate the diet using 25 week-old chicks, recording the weight gain (g) over two weeks. She found the following results:

Diet	n	Mean (g)	Standard deviation (g)
Standard	10	70.3	2.56
New	15	74.2	2.81

    (a) Let μ_1 and μ_2 be the (population) mean weight gains of week-old chicks fed the standard and new diets, respectively. In terms of these parameters, what are the null and alternative hypotheses that the researcher would like to test?

    (b) What is the pooled estimate of standard deviation?

    (c) Carry out a test of the hypotheses in (a). Show your working and state your conclusion.

Question 4. A group of 10 students took turns in measuring their reaction time by having one student, the assistant, drop a ruler while another student, the subject, tries to catch it as quickly as possible. Each student had two attempts, recorded in the table below as Reaction 1 (cm) and Reaction 2 (cm). It is commonly thought that there is a learning effect with reaction time, whereby the second attempt will be faster than the first attempt.

Student	1	2	3	4	5	6	7	8	9	10
Reaction 1	20	37	35	21	19	24	20	14	19	23
Reaction 2	41	13	17	15	18	12	20	9	13	15

    (a) Let μ be the mean reduction in reaction distance from Reaction 1 to Reaction 2. In terms of this parameter, what null and alternative hypotheses would we test to determine if there is a learning effect?

    (b) The Reaction 1 measurements had mean 23.4 cm with standard deviation 7.71 cm while the Reaction 2 measurements had mean 17.3 cm with standard deviation 8.91 cm. The mean of the differences was 6.1 cm with standard deviation 12.36 cm. Based on these statistics, carry out a test to determine whether there is a learning effect. Show your working and state your conclusion.

    (c) Calculate a 95% confidence interval for the mean reduction in reaction distance from Reaction 1 to Reaction 2.

    (d) Suppose we would like to estimate the mean reduction in reaction distance at 95% confidence with a margin of error of 3 cm. For ethical reasons we would like to keep the number of subjects required to a minimum. Using the above data as a pilot study, how many subjects would we need to obtain a margin of error of 3 cm?

Question 5. Based on data from the Danish National Birth Cohort, researchers investigated whether maternal fish intake during pregnancy might influence the risk of childhood asthma. The table below gives a summary of the cases of doctor-diagnosed asthma against whether mothers consumed fish regularly or not.

Fish intake	Yes	No
Asthma	360	299
No Asthma	2088	788

Use an appropriate statistical method to decide whether there is evidence of an association between maternal fish intake during pregnancy and childhood asthma. State your null and alternative hypotheses in words or symbols. Show your working and state your conclusion.

From STAT1201, semester two 2014

Question 6. A marine researcher suspects that the length of Bluefin tuna in a certain part of the Atlantic Ocean is declining. In 2013 a random sample of 51 Bluefin tuna had an average length of 180 cm with a standard deviation of 20 cm. During the same period in 2014 a random sample of 67 Bluefin tuna had an average length of 175 cm with a standard deviation of 19 cm.

   (a) Compute the pooled variance.
   (b) Does the data support the scientist's suspicion that the length of Bluefin tuna is declining. You may assume that the data are normally distributed with equal variances.
   (c) Determine a 95% confidence interval for the expected difference in Bluefin tuna lengths.

Question 7. Researchers asked women who were pregnant with planned pregnancies how long it took them to get pregnant (Blair and Wilcox, 1985). Length of time to pregnancy was measured according to the number of cycles between stopping birth control and getting pregnant. Women were also categorized according to whether or not they smoked, with smoking defined as having at least one cigarette per day for at least the first cycle during which they were trying to get pregnant. The observed counts are as follows:

	Pregnancy occurred after	
	First cycle	Two or more cycles
Smoker	29	71
Non-smoker	198	288

Is there evidence that there is a higher proportion of non-smokers than smokers who are able to get pregnant during the first cycle? State the null and alternative hypotheses, and use an appropriate test statistic to determine the P-value. What do you conclude?

Question 8. Plastic microparticles are contaminating the world's shorelines, and much of this pollution appears to come from fibres from washing polyester clothes, predominantly polyester fleece fabrics. Samples have been taken from beaches along the coastline in North Queensland and Victoria. 5 samples were taken from each of 18 different shorelines (9 in each state), for a total of 90 samples. The mean number of polyester microparticles found per 250mL of beach sediment was 19.6 with a standard deviation of 8.2.

(a) Construct and interpret a 99% confidence interval for the mean number of polyester microparticles per 250mL of beach sediment.
(b) The researchers were also interested in the dispersion of these polyester microparticles, comparing the contamination of beaches in the warm region of North Queensland with the colder Victorian region (where fleece clothing is worn more frequently).

Region	n	Mean	Standard deviation
North Queensland	45	16.67	8.54
Victoria	45	21.22	9.82

Carry out a test of whether there is a difference in the mean number of polyester microparticles between Victoria and North Queensland.

Question 9. How often do you use cash to pay for purchases? In a survey of 1000 Australian adults in 2014, 43% of respondents reported not having used cash to pay for any purchases in the previous 7 days.

(a) Construct and interpret a 90% confidence interval for the proportion of adult Australians who don't pay cash within a 7 day period.
(b) The researchers were interested in whether an individual's propensity to pay with cash was associated with age. The number of individuals who did not pay with cash is given by Age Group in the contingency table below.

Age Group	Cash transaction in past 7 days?		Total
	No	Yes	
Under 35	232	202	434
Over 35	198	368	566
Total	430	570	1000

Is there any evidence of a difference in the proportion of people who made no cash transaction between the two age groups?

Question 10. Cocaine addiction is very hard to break, and relapse is common among addicts trying hard to break their addiction. A study has been carried out to investigate the effectiveness of two drugs, Desipramine and Lithium, in the treatment of cocaine addiction. The subjects in the six week study were cocaine addicts seeking treatment. The 72 subjects were randomly assigned to one of three groups (Desipramine, Lithium or a Placebo) and the study was double-blind. The results are presented in the table below.

	Relapse	Did not relapse
Placebo	20	4
Lithium	18	6
Desipramine	10	14

Is there any evidence of an association between the likelihood of relapse and the choice of treatment drug? State the null and alternative hypotheses, and use an appropriate test statistic to determine the P-value. What do you conclude?

From STAT1201, semester one 2015

Question 11. Two lakes are being analysed with respect to their polychlorinated biphenyl (PCB) concentrations in fish. The PCB concentrations (in parts per billion, ppb) from 10 fish caught in Lake *A* had a sample mean and sample standard deviation of 11.170 ppb and 0.862 ppb, respectively. The sample mean and sample standard deviation of PCB concentrations from 8 fish taken from Lake *B* were 11.988 ppb and 0.738 ppb, respectively.

(a) A marine scientist wants to know if there is a difference in the true mean PCB concentrations between the two lakes. Help her answer this question by carrying out an appropriate hypothesis test. Write your conclusion in a manner that can be understood by the marine scientist.

(b) Construct and interpret a 95% confidence interval for the true difference in mean PCB concentrations between the two lakes

Question 12. In a study a researcher collected data from seven fields, where two methods of sowing (row planted and broadcast planted) wheat were compared. The response recorded was the difference between the row-planted profit and the broadcast-crop profit in each field. The average difference was -1.014286 and standard deviation of the difference was 1.606723.

(a) Clearly state the null hypothesis for this experiment, using words and symbols. Use an appropriate test statistic to determine the P-value. What do you conclude?

(b) Construct a 95% confidence interval for the difference in mean profit for the two sowing methods?

Question 13. *Gemfibrozil* is a cholesterol-lowering drug that can purportedly reduce the risk of heart attacks. To investigate this claim, two groups of middle-aged men were randomly assigned to either the drug or a placebo group, and followed over a 5-year period. The incidences of heart attacks during this period are tabulated below:

	Heart attack	No heart attack
Gemfibrozil	56	1995
Placebo	84	1946

Does Gemfibrozil lower the risk of heart disease? Answer this question by carrying out an appropriate hypothesis test. Write you conclusions in a manner that can be understood by a clinical researcher.

Construct and interpret a 90% confidence interval for the difference in heart attack rates between the drug and placebo group.

From STAT1201, semester two 2015

Question 14. Suppose we select 50 statistics students who are sitting in the front half of a lecture room and find that 10 were shorter than 150 cm. On the basis of this, construct an approximate 95% confidence interval for the proportion of students who sit at the front of a lecture who are shorter than 150 cm.