Question 1

A study aimed to determine if grapefruit juice has beneficial effects on the pharmacokinetics of oral digoxin, a drug often prescribed for heart ailments. Seven healthy non-smoking volunteers participated in the study. Subjects took digoxin with water for 2 weeks, no digoxin for 2 weeks, and then digoxin with grapefruit juice for 2 weeks. The peak plasma digoxin concentrations (Cmax; ng/mL) when subjects took digoxin under the two conditions are given in the following table:

Subject	1	2	3	4	5	6	7
Water	2.34	2.46	1.87	3.09	5.59	4.05	6.21
Grapefruit Juice	3.03	3.46	1.97	3.81	3.07	2.62	3.44
Decrease	-0.69	-1.00	-0.10	-0.72	2.52	1.43	2.77

While small, note that the sample size was chosen carefully by the authors. In their paper they state that "assuming an α level of 0.05, a sample size of seven subjects has a power of 85% to detect a 25% change in digoxin Cmax".

Lower values of Cmax are better since they imply that digoxin is available in the body for longer. Is there any evidence that grapefruit juice increases the effectiveness of oral digoxin, by decreasing Cmax?

(a) Identify one issue in the design of this experiment that undermines being able to use the data to answer this question. How could the design be improved? [2 marks]

Initial digoxin exposure might carry over, even after 2 weeks without.

Randomly allocate order of water/grapefruit treatments to subjects.

(b) Suppose μ is the true mean decrease in digoxin Cmax with grapefruit compared to water. Define the null and alternative hypotheses for this study in symbols. [1 mark]

H_0: $\mu = 0$

H_1: $\mu > 0$

(c) We have two sets of Cmax measurements, one for water and one for grapefruit. Briefly explain why we work with the differences rather than carrying out a two-sample t-test to compare the treatments. [1 mark]

Observations are not independent since they came from same subjects

(d) The seven differences in Cmax have mean 0.601 ng/mL with standard deviation 1.609 ng/mL. Use these values to test the hypotheses in (b). What do you conclude? [2 marks]

$$t_6 = \frac{0.601 - 0}{1.609/\sqrt{7}} = 0.988$$

P-value is $P(T_6 \geq 0.988) > 0.1$,

no evidence to suggest a decrease in Cmax with grapefruit juice.

(e) How many of the seven subjects had a lower Cmax value with grapefruit juice? Use this to find the *P*-value for a sign test of whether grapefruit juice tends to lower Cmax. What do you conclude? [2 marks]

$x = 3$.

P-value is $P(X \geq 3) = .273 + .273 + .164 + .055 + .008$

$= 0.773$,

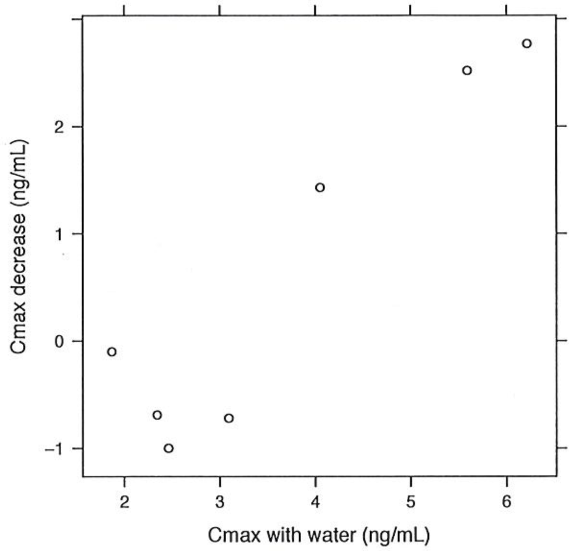
again no evidence that Cmax tends to be lower with grapefruit juice.

(f) For a sign test from seven subjects, what is the minimum number of reductions in Cmax needed to give evidence of an effect at the 5% level? [1 mark]

$P(X \geq 7) = .008$ but $P(X \geq 6) = .055 + .008 = .063$,

so need all 7 to be reductions for evidence at the 5% level.

(g) The authors speculated that the decrease in Cmax with the grapefruit juice may be related to an individual's baseline response with water. They produced the following scatterplot of the relationship observed in the data:



They calculated the Pearson correlation coefficient for this relationship to be 0.9321. Does this give any evidence that there is an underlying association between the decrease in Cmax and the Cmax with water? [3 marks]

Test $H_0: \rho = 0$ vs $H_1: \rho \neq 0$, using

$$t_5 = \frac{.9321 - 0}{\sqrt{\frac{1 - .9321^2}{5}}} = 5.75,$$

One-sided P-value is $P(T_5 \geq 5.75)$, btw .005 and .001, so two-sided P-value is btw .01 and .002, strong evidence of an association between decrease and baseline with water.

(h) Briefly explain how your conclusion to (g) relates to your conclusion to (d). [1 mark]

Grapefruit juice has little effect for people who already have low Cmax values, giving no evidence overall. However, it seems grapefruit juice may be beneficial for people with high Cmax using water.

Question 2

In a study of factors thought to be associated with birth weight, a random sample of 100 births was selected from all the birth records at a hospital in 2001. Three variables were extracted from each record: the length of gestation (weeks), the smoking status of the mother (smoker/nonsmoker) and the birth weight (grams).

(a) The mean length of gestation for the 100 births was 38.36 weeks with standard deviation 3.36 weeks. Construct a 95% confidence interval for the mean length of gestation for all births at this hospital in 2001. [2 marks]

$t^*_{99} \simeq 1.96$, so 95% CI is

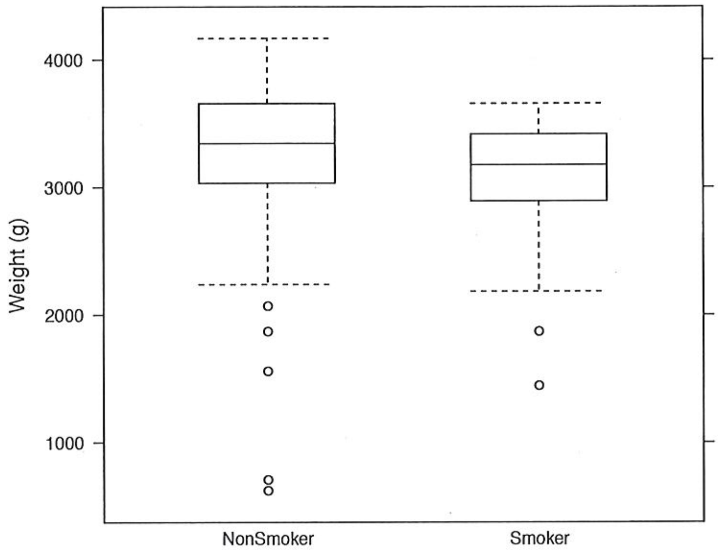
$$38.36 \pm 1.96 \times \frac{3.36}{\sqrt{100}} = 38.36 \pm 0.66,$$

so 95% confident mean length is between 37.70 and 39.02 weeks.

(b) There are claims that the mean length of gestation has been decreasing in recent decades from the traditionally held value of 40 weeks. Based on your interval in (a), or otherwise, do the data from this hospital support or contradict such claims? [1 mark]

40 is not in the above interval so this does suggest mean has decreased.

The main question of interest is whether babies born to smoking mothers tend to have lower birth weights. The following figure shows a side-by-side comparison of the two distributions from this data set.



A two-sample t-test in R gave the following output:

```
      Welch Two Sample t-test

data:  Weight by SmokingStatus
t = 1.3865, df = 18.42, p-value = 0.1817
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -131.3122   646.6612
sample estimates:
mean in group NonSmoker    mean in group Smoker
               3258.941                3001.267
```

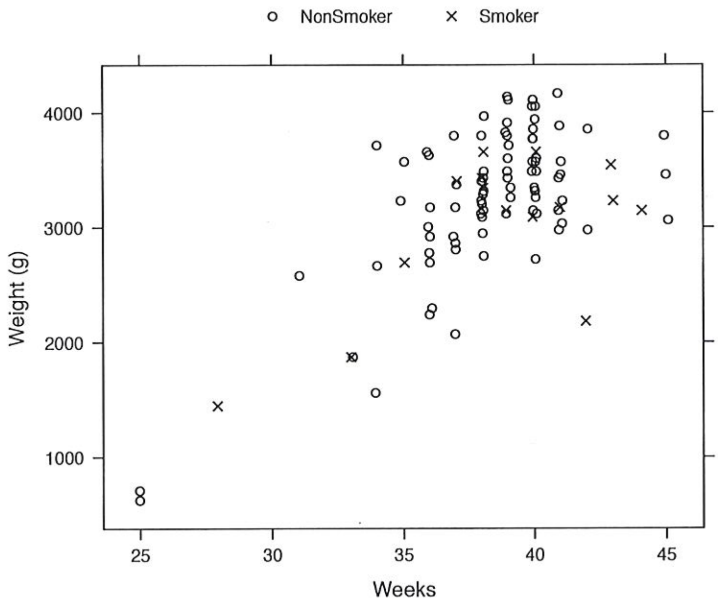
(c) On average, how much lower are birth weights for smokers compared to non-smokers? [1 mark]

$3258.941 - 3001.267 = 257.7$ g

(d) Summarise the conclusions from the t-test. [2 marks]

There is no evidence to suggest a difference in mean birth weight between smokers and non-smokers.

There is considerable variability in birth weights for both groups in the above figure. However, some of this variability may be explained by the different gestation lengths for each birth. The following scatterplot shows the relationship between birth weight and gestation length, with smokers and non-smokers shown by the plotting symbols:



The R output below gives the summary of the multiple regression model for birth weight based on both gestation length and smoking status:

```
lm(formula = Weight ~ Weeks + SmokingStatus, data = births)

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)           -1724.42     558.84  -3.086  0.00265 **
Weeks                   130.05      14.52   8.957 2.39e-14 ***
SmokingStatusSmoker    -294.40     135.78  -2.168  0.03260 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 484.6 on 97 degrees of freedom
Multiple R-squared:  0.4636,  Adjusted R-squared:  0.4525
F-statistic: 41.92 on 2 and 97 DF,  p-value: 7.594e-14
```

(e) Based on the model output, what is the estimated birth weight for a birth at 35 weeks gestation to a non-smoking mother? [1 mark]

$-1724.42 + 130.05 \times 35 - 294.40 \times 0$

$= 2827$ g

(f) Briefly interpret the value '130.05' in the output. [1 mark]

For each additional week of gestation, mean birth weight increases by 130.05 g (regardless of smoking status)

(g) Why do the residuals have 97 degrees of freedom? [1 mark]

We need to estimate 3 parameters from the data (b_0, b_1, b_2) so DF = 100 − 3 = 97.

(h) Based on the multiple regression model, is there any evidence of a difference in mean birth weight between smoking and non-smoking mothers? Justify your conclusion with reference to the R output above. [2 marks]

P-value for smoking status effect is 0.0326, so moderate evidence to suggest a difference in mean birth weight between smokers and non-smokers (after taking into account gestation).

(i) Briefly explain why the conclusion from the multiple regression model might be different to the conclusion from the two-sample t-test in (d). [2 marks]

There was a lot of variability within groups in (d), making it hard to detect a difference.

The multiple regression model explains a lot of this variability in terms of gestation length, allowing the smoking effect to be seen.

Question 3

Energy drinks have become widely popular among adolescents and are also consumed by athletes, paricularly those who have just begun their sporting career. A recent paper presented a study on the consumption of energy drinks by teenagers engaged in sports, including quantity consumed and factors that might be associated with consumption.

A total of 707 students, selected randomly from sports classes at various schools, completed a questionnaire on energy drink consumption. The following table shows the cross-tabulation of regular energy drink consumption by gender:

	Energy Drinks		
Gender	Yes	No	
Female	192	90	282
Male	296	129	425
	488	219	707

(a) Was this an observational or experimental study? Briefly justify your answer. [1 mark]

Observational.

No treatment/intervention was applied.

(b) Overall, what proportion of the students consumed energy drinks? [1 mark]

$\frac{488}{707} = 0.690$

(c) What is the estimated difference in the proportions of females and males who consume energy drinks? [1 mark]

Male: $\hat{p}_M = \frac{296}{425} = 0.6965$

Female: $\hat{p}_F = \frac{192}{282} = 0.6809$

so $\hat{p}_M - \hat{p}_F = 0.0156$

(d) Assuming this sample is representative of all teenagers engaged in sports, give a 95% confidence interval for the true difference in the proportions of females and males who consume energy drinks. What does the interval say about the difference in energy drink consumption between genders? [3 marks]

$$se(\hat{p}_M - \hat{p}_F) = \sqrt{\frac{.6965(1-.6965)}{425} + \frac{.6809(1-.6809)}{282}}$$

$$= 0.0356,$$

so 95% CI is

$$0.0156 \pm 1.96 \times 0.0356$$

$$= 0.0156 \pm 0.0698,$$

so we are 95% confident difference is between -0.054 and +0.085.

Since 0 is in this range, no evidence of a difference between males and females in energy drink consumption.

Another factor recorded on the questionnaire was the frequency of practising sports. The following table gives the summary of results for the 681 students who indicated they practised at least once a week:

	Energy Drinks		
Practising Sports	Yes	No	
Daily	328	146	474
2-3 times per week	28	13	41
Once per week	114	52	166
	470	211	681

(e) Based on this table, is there evidence of an association between energy drink consumption and frequency of practising sports? [5 marks]

Assuming independence, expected values are

	Yes	No
Daily	327.1	146.9
2-3	28.3	12.7
Once	114.6	51.4

So χ^2 statistic is

$$\chi^2_2 = \frac{(328-327.1)^2}{327.1} + \frac{(146-146.9)^2}{146.9} + \dots + \frac{(52-51.4)^2}{51.4} = 0.028$$

P-value is $P(\chi^2_2 \geq 0.028) > .975$

no evidence to suggest energy drink consumption is associated with frequency of practising sports.

The schools in the study were from different cities so the authors were also interested in whether there might be differences in energy drink consumption between the cities. They used R to conduct a chi-squared analysis, giving the following results:

```
	Pearson's Chi-squared test

data:  table(data$City, data$EnergyDrink)
X-squared = 9.8619, df = 1, p-value = 0.001687
```

(f) How many cities were represented in the study? [1 mark]

Since df=1, there must have been 2 cities.

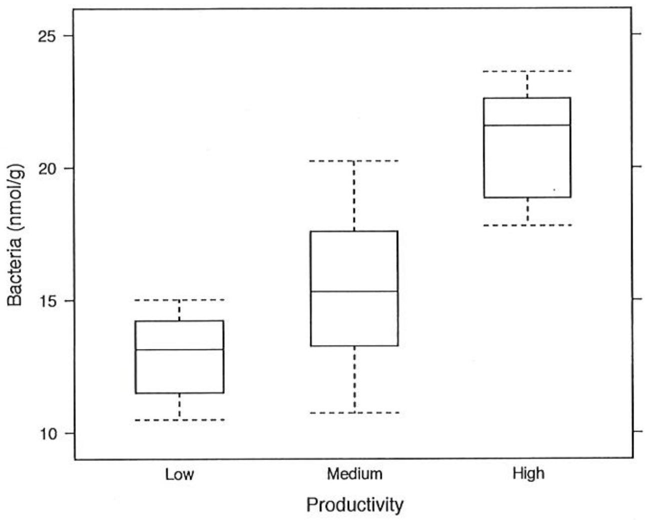
(g) Briefly summarise the conclusion from this chi-squared test. [1 mark]

Strong evidence of a difference in energy drink consumption between the two cities.

Question 4

Albic soil is a common low-yielding soil in eastern China and increasing its productivity is crucial to ongoing food security in the region. A study investigated factors that affect soil quality to gain a better understanding of those factors that are ultimately related to crop yields. They took soil samples from 36 sites and classified each site as either 'Low', 'Medium' or 'High' productivity on the basis of the mean annual crop yield over the previous five years.

One measurement made was the total concentration of bacterial communities in the soil (nmol/g). The figure below shows side-by-side box plots of the bacterial concentration for each productivity level:



The table below shows the observed sample means and standard deviations of the bacterial concentrations for each of the productivity levels.

Productivity	n	Mean	SD
Low	12	12.9	1.51
Medium	12	15.5	2.98
High	12	20.9	2.05

The researchers used one-way analysis of variance (ANOVA) to compare the mean bacterial concentrations between the productivity levels.

(a) An important assumption for one-way ANOVA is that groups and observations are independent of each other. Give one example of how that assumption might be compromised in a study like this. [1 mark]

If soil samples came from nearby areas they might not be independent.

(b) List two assumptions for one-way ANOVA that can be assessed using the above figure. Briefly comment on the validity of each assumption for these data. [2 marks]

Normal variability: each group seems roughly symmetric

Constant variability: 'Medium' has more variability than 'Low', though not quite double.

(c) The one-way ANOVA gave a total sum of squares of 569.6 and a residual sum of squares of 168.9. What are the units of these values? [1 mark]

$nmol^2/g^2$

(d) Using the values from (c), complete the ANOVA table below. [3 marks]

Source	DF	SS	MS	F
Productivity	2	400.7	200.35	39.15
Residuals	33	168.9	5.118	
Total	35	569.6		

(e) Give bounds on the P-value for the F-test. What do you conclude? [1 mark]

P-value is $P(F_{2,33} \geq 39.15) < 0.001$,

strong evidence of a difference in mean bacterial concentration between the three productivity levels.

(f) What is the R^2 value for this model of bacterial concentration? Briefly interpret the value. [2 marks]

$$R^2 = \frac{400.7}{569.6} = 0.7035,$$

so roughly 70% of bacterial concentration variability is explained by production level.

(g) Based on this model and the R^2 value, would you recommend increasing the bacteria in the soil to improve productivity? Briefly explain why or why not. [1 mark]

No, it is not necessarily a causal relationship. (Whatever improves productivity might also increase bacteria.)

(h) In addition to measuring bacterial communities, the researchers also made 25 other measurements of physical, chemical, biochemical and biological properties of each soil sample. Briefly explain an issue in comparing all these measurements between the different productivity levels. What would you recommend to overcome this issue? [2 marks]

Increases the chance of a Type I error.

Use some kind of correction (such as Bonferroni) to avoid the issue.

END OF EXAMINATION