For questions 1 – 20, circle one answer only. If making an error, write your final answer clearly in the margin. Ambiguous responses will be considered incorrect.

1. The mean age of five people in a room is 30 years. One of the people whose age is 50, leaves the room. The mean age of the remaining people is 25 years.

   T F

2. For a Normal distribution, the upper and lower quartiles are respectively one standard deviation above and below the mean.

   T F

3. A correlation coefficient based on a scatter plot measures the proportion of data lying on the regression line.

   T F

4. A phone–in poll at a radio station concluded that 70% of Canadians approve of Stephen Harper, based on the responses of 8,000 callers. The conclusion is valid since the sample size is large.

   T F

5. The weights of a type of tomato are believed to have standard deviation of 28g. If two such tomatoes are weighed together, their combined weight would have standard deviation 56g.

   T F

6. A random sample of 100 students are asked if they are vegetarians. Five percent respond yes, and a 99% confidence interval for the true proportion of students who are vegetarians is found to be $(0.03, 0.07)$. This implies 99% of all random samples of 100 students will have a sample proportion that falls between 0.03 and 0.07.

   T F

7. Confidence intervals are constructed for both parameters and statistics.

T F

8. You carry out a hypothesis test that compares the means of two populations. Suppose a p-value of 0.12 is obtained. This implies that there is a 12% chance that the two population means are equal.

T F

9. In linear regression, at a fixed value of the explanatory variable, a 95% prediction interval for an observation of the response variable is always wider than a 95% confidence interval for the mean response.

T F

10. One-way ANOVA provides evidence against the null hypothesis that the population means are all equal when the between-group variation is large compared to the within-group variation.

T F

11. If the observations in a data set are all equal, then

(a) the variance of the data set equals 0.

(b) the mean of the data set equals 0.

(c) the IQR of the data set equals 0.

(d) both (a) and (c).

(e) both (b) and (c).

12. The slope of a regression line and the correlation are similar in the sense that

    (a) they both have the same sign.

    (b) they do not depend on the units of measurement of the data.

    (c) they both fall between –1 and 1 inclusive.

    (d) neither of them can be affected by outliers.

    (e) both can be used for prediction.

13. Which of the following is an incorrect statement about the correlation between two quantitative variables X and Y?

    (a) A correlation of –0.8 indicates a stronger linear association between X and Y than a correlation of 0.5.

    (b) A correlation of 0 implies X and Y are not related at all.

    (c) A correlation of –1 indicates that $Y = -X$.

    (d) Both (b) and (c).

    (e) Both (a) and (b).

14. A certified fitness coach wanted to test the effectiveness of a new fitness program in reducing weight among obese patients. Fifty female patients and fifty male patients participated in the experiment. Within each gender group, the patients were randomly assigned to one of the two fitness programs – the new and the existing fitness programs. Upon completion of the program, reduction in weight was measured for each patient. Which of the following statements is incorrect about this experiment?

    (a) There are four treatments in the study.

    (b) Gender is a blocking variable.

    (c) The patients were not guaranteed to lose weight due to the experiment.

    (d) Reduction in weight is the response variable.

    (e) Type of fitness program is the factor.

15. There are two events A and B. It is given that $P(A) = 0.7$ and $P(B) = 0.6$. Consider the following statements:

(i) A is the complement of B.

(ii) The two events are not disjoint.

(iii) A is a more probable event than B.

(iv) The probability that the two events both occur must be equal to 0.42.

Which of the above statements is correct about the two events?

(a) (iii) only.

(b) (iv) only.

(c) (ii) and (iii) only.

(d) (ii) and (iv) only.

(e) (i), (iii) and (iv) only.

16. In testing a two–sided significance test for a mean, the test statistic was –2.12 which is expected to be a value from the standard Normal distribution under the null hypothesis. The p–value of the statistic is

(a) 0.017

(b) 0.034

(c) 0.048

(d) 0.983

(e) 0.050

17. A sample of size 25 is used to conduct a test of the null hypothesis $H_0 : \mu = 30$ against the alternative $H_a : \mu > 30$. Using a t test, the degrees of freedom would be

    (a) 30

    (b) 31

    (c) 23

    (d) 24

    (e) 25

18. A regression line is fitted to a scatter plot consisting of 42 points. The fit is quite good, but it is desirable to test whether the underlying slope parameter is zero. The F test statistic for this test was found to be 5.55. Which of the following is the p–value for this statistic?

    (a) 0.0010

    (b) 0.0234

    (c) 0.0468

    (d) 0.976

    (e) 0.999

19. If examining the plot of the residuals against the explanatory variable x for a linear model, when the model fits well one would expect

    (a) the residuals to lie on a line of positive slope.

    (b) the residuals to lie on a line of negative slope.

    (c) the residuals to scatter about a line of positive slope.

    (d) there to be no variation in the residuals.

    (e) there to be no obvious pattern in the residuals.

20. In performing the usual hypothesis test in the analysis of variance using the mean–square ratio, the alternative hypothesis is that

(a) at least two of the underlying group means are different.

(b) all the underlying group means are different.

(c) all the within–group variances are equal.

(d) at least two within–group variances are different.

(e) all the within–group variances are different.

21. In a large city, 37% of all restaurants accept both master and visa credit cards, and 20% accept only master cards but not visa cards. A tourist visiting the city picks at random a restaurant at which to have lunch. Define the following events:

M = {the randomly chosen restaurant accepts master credit cards},
V = {the randomly chosen restaurant accepts visa credit cards}.

(a) Choose the appropriate responses to the following statements.

M and V are disjoint events.

(i) True

(ii) False

(iii) Insufficient information to tell.

M and V are independent events.

(i) True

(ii) False

(iii) Insufficient information to tell.

M is the complement of V.

(i) True

(ii) False

(iii) Insufficient information to tell.

(b) What does the probability 20% represent?

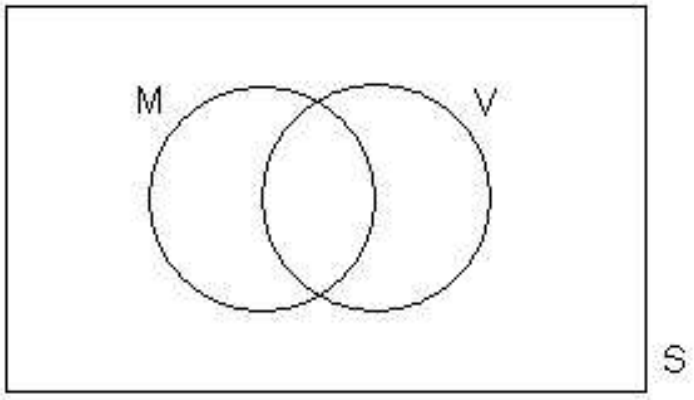
(i) $P(M \text{ and } V^c)$

(ii) $P(M^c \text{ and } V)$

(iii) $P(M \text{ or } V^c)$

(iv) $P(M^c \text{ or } V)$

Shade the corresponding region in the following Venn Diagram.



(c) The probability that a randomly chosen restaurant accepts master credit cards is (hint: use the Venn diagram)

(i) 17%

(ii) 40%

(iii) 57%

(iv) 76%

(d) Now suppose 2 restaurants are randomly chosen. What is the probability that at least one of the two restaurants accept both master and visa credit cards?

22. A type of thread is being studied for its tensile strength. Fifty pieces were tested under similar conditions, the mean tensile strength being 78.30kg and the standard deviation being 5.60kg.

    (a) Give an approximate 95% confidence interval for the mean tensile strength of the thread.

    (b) Assuming a Normal distribution for the strength of the thread, estimate the tensile strength that would be exceeded by 95% of such threads.

23. Eight marksmen, labeled A, B,. . . ,H, shot at targets with two types of rifle. Their scores were as in the table below:

	Marksman							
	A	B	C	D	E	F	G	H
Rifle Type 1	93	99	90	87	85	94	88	91
Rifle Type 2	89	93	86	92	78	90	91	87

Test the hypothesis that the rifles are equal quality. State carefully the null and alternative hypotheses, any distributional assumptions you make and your final conclusions.

24. A study was conducted to determine whether an expectant mother's cigarette smoking has any effect on the bone mineral content of her otherwise healthy child. A sample of 30 newborns whose mothers smoked during pregnancy has a mean bone mineral content of 0.092 g/cm and a standard deviation of 0.026 g/cm; a sample of 72 infants whose mothers did not smoke has a mean of 0.105 g/cm and a standard deviation of 0.025 g/cm.

(a) Do the data suggest that the population mean bone mineral content of newborns differ between mothers who smoked and those who did not smoke during pregnancy? Use a significance level of $\alpha = 0.05$. Define clearly the parameter(s) and variable(s) that relate to your test, and state any assumption(s) that you use.

(b) Based on the results obtained from part (a), you can confidently say that (circle all that apply):

(i) smoking causes a decrease in the bone mineral content in the newborns.

(ii) smoking is associated with the bone mineral content in the newborns.

(iii) smoking has no effect on the bone mineral content in the newborns.

(iv) smoking is independent of the bone mineral content in the newborns.

(c) Would you expect a 95% confidence interval for the true difference in the population means to contain the value 0? (Circle one)

Yes No

Briefly justify your answer.

25. How well does the size of a house determine the annual tax house owners are paying? Nineteen houses are randomly selected from a city. The house size (measured in square feet of living space) and the amount of annual tax (in dollars) are recorded for each of the 19 houses. Here are the summary statistics for the two variables:

    house size : mean = 1456 sqft, standard deviation = 370 sqft
    annual tax : mean = \$1707, standard deviation = \$323

    The linear regression line that predicts the amount of annual tax from the house size has a slope of \$0.81 per square foot.

    (a) Below is a partial ANOVA table for the regression line. Complete the table.

Source	Sum of Squares	df	Mean Square	F
Model	1610825			
Error	262428			
Total				

    (b) Is there a significant linear relationship between the size of a house and the amount of annual tax charged? Carry out an appropriate test using the 1% significance level.

(c) Find the value of r^2, where r is the correlation between the size of a house and the amount of annual tax charged. Interpret this value in the context of this question.

(d) The 95% confidence interval for the population slope is found to be (0.64, 0.98) dollars per square foot. Using this information, find the standard error of the slope of the regression line.

(e) Give a point estimate for the mean annual tax paid by owners owning a 1500-square feet house.

26. A study investigated whether month of birth impacts on the time a baby learns to crawl. Parents with children born in January, May or October were asked the age, in weeks, at which their child could crawl one metre within a minute. The data are summarised below:

		Crawling age		
		Mean	st. dev.	size
Birth month	January	29.84	7.08	34
	May	28.58	8.07	29
	October	33.83	6.93	40

The data from each birth month are assumed to follow a Normal distribution. The analysis is via ANOVA, with an incomplete ANOVA table given below:

Source	Sum of squares	df	MS	F
Between groups	505.26			
Error			53.45	
Total				

(a) Circle which, if any, of the following statements you consider to be correct:

(i) It is inappropriate to use ANOVA here since there is evidence that the Normal distributions underlying each sample have different variances.

(ii) It is inappropriate to use ANOVA here since the sample sizes are unequal.

(iii) It would be inappropriate to calculate correlations for these data.

(iv) This is a randomised block design experiment.

.

(b) State clearly the null hypothesis for the AVOVA test.

(c) Compute the test statistic for the test in (b).

(d) Test the hypothesis at the 5% significance level, providing a range for the p-value of the test.

27. In a certain city, 25% of residents are European. Suppose 120 people are called for jury duty, and only 24 of them are European. Does this indicate that Europeans are under-represented in the jury selection system? Carry out an appropriate hypothesis test at the 10% significance level. Remember to define the parameter(s) that relates to your test.