INTRODUCTION TO BIOSTATISTICS (STAT115)

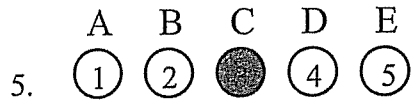
Mid Semester Test - 23 September 2010

TIME ALLOWED: ONE HOUR TWENTY MINUTES (7.00 pm –8.20 pm)

THERE ARE 50 MULTIPLE CHOICE QUESTIONS.

READ THE FOLLOWING NOTES CAREFULLY

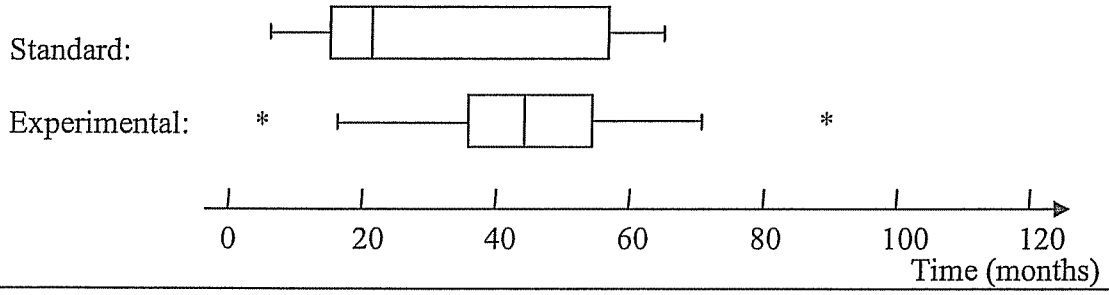
1. Answer on the coloured EXAMINATION ANSWER SHEET.
2. Write your name on the examination sheet in **CAPITAL LETTERS.**
3. Write your STUDENT ID in the space provided.
4. Carefully shade the numbers which match your **STUDENT ID.**
5. Beside Paper write your **TUTORIAL DAY AND TIME.**
6. Use a soft pencil, preferably 2B.
7. Carefully shade in the circle which matches your answer choice, as for the hypothetical question 5 below:



8. You will score 1 mark for a correct answer, and 0 for an incorrect answer or no answer.

Calculators may be used. Tables for the binomial, standard normal, t and χ^2 distributions and a summary of formulae are attached.

Information for question 1
The box-and-whisker plots for the survival times (in months) of patients with a terminal illness, one group with 68 patients who are given a standard treatment, and another group of 29 patients who are given an experimental treatment, are as follows



1. Which of the following statements is false?
   (A) The interquartile ranges for the values in the two groups are different.
   (B) The median survival time is lower in the experimental group.

Information for question 2
St John's wort has been used in folk medicine for a long time to treat depression. A study was carried out to assess the effectiveness of this medication. A group of 105 patients was randomly allocated to one of two groups. The first group was given St John's wort and the second a placebo. Symptoms were measured after 8 weeks.

2. This study is:
   (A) a case control study
   (B) a randomized controlled trial
   (C) a cohort study

Information for questions 3 - 5
A study investigated the association between physical activity and risk of death from heart disease. A group of 7142 men were followed for 20 years. Leisure time physical activity was measured at baseline and graded from 1 to 3 with 1 = mainly sendentary, 2 = social sport activity, and 3 = participation in competitive sport. The following results were obtained:

Sport Category	Person Years follow-up	Number of deaths
1	35,082	214
2	83,191	391
3	22,000	77

3. A measure of incidence of death over 20 years in those who participated in competitive sport is 3.5 per 1000 person years. This measure is the:
   (A) cumulative incidence
   (B) incidence rate

4. The relative risk comparing risk of death in those participating in competitive sport to those mainly sedentary is closest to:
   (A) 0.36 (B) 2.78 (C) 0.57 (D) 1.74 (E) 0.77

5. 253 men were excluded from the study at the start because they had evidence of heart disease. The reason for this was to:
   (A) ensure the study was randomized
   (B) reduce sample size
   (C) remove men no longer at risk of developing heart disease since their exercise behaviour may have changed.

Information for questions 6 - 10
The prevalence of prostate cancer in men aged 70 and over is 1 in 50. A blood test measures the level of an antigen. The test is positive (indicating prostate cancer) in 90% of diseased cases but also incorrectly positive in 20% of disease-free cases. Men aged 70 and over present randomly for screening by the test.

Use this space for a tree diagram to help answer #6-10.

6. The sensitivity of the test is
(A) 0.10 (B) 0.80 (C) 0.20 (D) 0.02 (E) 0.90

7. The specificity of the test is
(A) 0.20 (B) 0.90 (C) 0.98 (D) 0.80 (E) 0.10

8. The probability that the test is positive is
(A) 0.018 (B) 0.214 (C) 0.784 (D) 0.802 (E) 0.020

9. If the test is positive, the probability that the patient has prostate cancer is closest to
(A) 0.018 (B) 0.023 (C) 0.084 (D) 0.916 (E) 0.388

10. The positive predictive value of the test is closest to
(A) 0.018 (B) 0.023 (C) 0.084 (D) 0.916 (E) 0.388

11. A random variable, Z, has the standard normal distribution.
The value of $\mathbf{Pr(0 < Z < 1.5)}$ is:
(A) 0.9332 (B) 0.8664 (C) 0.0668 (D) 0.1336 (E) 0.4332

Information for questions 12 - 18
Glucose levels in healthy adults follow a random variable, X, which is normally distributed with mean $\mu_X = 5.31$ mmol/litre and standard deviation $\sigma_X = 0.56$ mmol/litre.

12. The probability that a healthy adult has a glucose level above 5.5 mmol/litre is closest to:
(A) 0.133 (B) 0.367 (C) 0.266 (D) 0.663 (E) 0.734

13. The distribution of means of samples of size 16 drawn from the distribution X has standard deviation closest to:
(A) 0.56 (B) 0.04 (C) 0.14 (D) 0.08 (E) 0.31

14. The probability that the mean glucose level for a sample of size 16 is more than 5.5 mmol/litre is closest to:
(A) 0.0869 (B) 0.4131 (C) 0.1738 (D) 0.3669 (E) 0.9131

Further information for questions 15 - 18
A random sample of 16 adult patients on a particular treatment produces glucose levels with mean $\bar{x} = 5.72$ mmol/litre and standard deviation $s = 0.64$ mmol/litre.

15. The estimated standard error of the sample mean is closest to:
(A) 0.14 (B) 0.64 (C) 0.04 (D) 0.56 (E) 0.16

16. The number of degrees of freedom associated with this sample is:
(A) 15 (B) 16 (C) 30 (D) 14 (E) 32

17. A 95% confidence interval for the true mean glucose level of patients on this treatment is closest to:
(A) 5.72 ± 0.21 (B) 5.72 ± 0.16 (C) 5.72 ± 0.34
(D) 5.72 ± 0.40 (E) 5.72 ± 0.44

18. There is evidence that adult patients on this treatment have a glucose level which is higher than that for healthy adults:
(A) True (B) False

Information for questions 19 – 25
Suppose that the proportion of children who develop chronic bronchitis in the first year of life is 5%. A random sample of 20 children (multiple births excluded) is selected.

19. The probability that only one of the 20 children will develop chronic bronchitis is closest
(A) 0.0500 (B) 0.7359 (C) 0.3774 (D) 0.6226 (E) 0.9500

20. The probability that two or more of the 20 children will develop chronic bronchitis is closest to:
(A) 0.2641 (B) 0.0754 (C) 0.1000 (D) 0.7359 (E) 0.1887

21. Clinic records show that children in the first year of life develop chronic bronchitis in two out of 20 cases when both parents are chronic bronchitis. This provides evidence at a 5% level of significance that there is higher incidence of chronic bronchitis when both parents are chronic bronchitis.
(A) True (B) False

Further information for questions 22 - 25
A random sample of 100 children (multiple births excluded) is selected. X is the binomial random variable for the number of these children who develop chronic bronchitis.

22. The mean of X is:
(A) 2 (B) 10 (C) 15 (D) 5 (E) 12

23. The standard deviation of X is closest to:
(A) 4.75 (B) 2.24 (C) 5.00 (D) 22.59 (E) 2.18

24. Using the normal approximation to the binomial distribution, the probability that at least 10 out of 100 children in the first year of life develop chronic bronchitis is closest to:
(A) 0.4803 (B) 0.0197 (C) 0.0294 (D) 0.4941 (E) 0.0059

25. A study from a second clinic of 100 children whose parents are both chronic bronchitis has 10 children in their first year of life who are chronic bronchitis. This study shows evidence of higher incidence of chronic bronchitis when both parents are chronic bronchitis.
(A) True (B) False

Information for questions 26 - 36

The consumption of fat increases the risk of heart disease through increasing the level of LDL cholesterol. A study was carried out to compare the effects of butter and margarine. The **data driven** research hypothesis was that butter raised LDL cholesterol levels more than margarine. 200 people were randomly allocated to one of two controlled diets. The two diets were identical except that the first diet included butter, while in the second, butter was replaced by margarine. After 45 days the following results for LDL cholesterol (mg/dl) were obtained.

	Sample Size (n_i)	Sample Mean ($\bar{x}_i$)	Sample variance (s_i^2)
Butter	100	174	1024
Margarine	100	168	676

26. This type of study is one of the following. Which one?
(A) a cohort study
(B) a case-control study
(C) a randomized controlled trial

27. The purpose of randomization is
(A) to obtain two groups which are likely to be comparable
(B) to obtain a representative sample from a population

28. If μ_B and μ_M are the respective population mean cholesterol levels for the people on the butter and margarine diets, appropriate null and alternative hypotheses for this **data driven** study are

(A) $H_0 : \mu_B - \mu_M = 0$, $H_A : \mu_B - \mu_M < 0$
(B) $H_0 : \mu_B = 0$, $H_A : \mu_M > 0$
(C) $H_0 : \mu_M - \mu_B > 0$, $H_A : \mu_M - \mu_B = 0$
(D) $H_0 : \mu_B - \mu_M = 0$, $H_A : \mu_B - \mu_M > 0$

29. The pooled estimate, s_p^2, for the common **variance** of the two samples is (to two decimal places)
(A) 29.00 (B) 850.00 (C) 29.15 (D) 14.50

30. The **estimated standard error** of the difference between the mean cholesterol levels in the butter and margarine groups is closest to
(A) 17.00 (B) 0.11 (C) 119.00 (D) 10.79 (E) 4.12

31. The value of the standardized test statistic is closest to
(A) 1.46 (B) 0.35 (C) 54.55 (D) 1.96 (E) 1.89

32. The p-value associated with the test statistic in the last question (assuming the standard normal can be used) is closest to
(A) 0.9279 (B) 0.4279 (C) 0.0416 (D) 0.0721 (E) 0.0271

33. The p-value in the last question suggests rejection of the null hypothesis in favour of the alternative at the 5% level of significance.
(A) True (B) False

34. The 95% confidence interval for the difference between the means is $-2.08 < \mu_B - \mu_M < 14.08$. A difference of 12 mg/dl is considered clinically important. This study produces
(A) a p-value > 0.05 and a conclusive result
(B) a p-value < 0.05 and an inconclusive result
(C) a p-value > 0.05 and an inconclusive result
(D) a p-value < 0.05 and a conclusive result.

35. A second larger study produced 95% confidence interval $-0.04 < \mu_B - \mu_M < 11.81$. This study produces
(A) a p-value > 0.05 and a conclusive result
(B) a p-value > 0.05 and an inconclusive result

36. Studies are more powerful if smaller samples are taken. This statement is:
(A) True (B) False

37. Researchers were interested in investigating the relationship between coronary heart disease and smoking status. To answer their research question they enrolled a group of people with coronary heart disease and a group of people without coronary heart disease into the study and asked them about their smoking history. This study design is:
(A) cross-sectional
(B) case-control
(C) cohort

38. In comparison to a cohort study, a case-control study:
(A) takes less time and is less expensive
(B) takes more time and is less expensive
(C) takes less time and is more expensive

Information for questions 39 - 44
A cohort of 1980 men followed for 28 years were categorized by whether they had developed herpes and whether they had been circumcised by age 3. The results are in the following table:

	Herpes		
	Yes	No	Total
Circumcised by age 3	48	722	770
Not circumcised by age 3	98	1022	1120
Total	146	1744	1890

39. The odds of a man developing herpes if circumcised by age 3 is closest to:
(A) 0.062 (B) 0.490 (C) 0.096 (D) 0.066 (E) 0.329

40. The ratio of the odds of herpes for men circumcised by age 3 compared with men not circumcised by age 3 is closest to:
(A) 0.712 (B) 1.442 (C) 1.404 (D) 0.794 (E) 0.693

41. The value of the chi-squared statistic for the 2 × 2 table is 4.053. There is evidence at the 5% level of an association between the development of herpes and circumcision status.
(A) True (B) False

42. The estimated standard error of the log of the odds ratio is closest to:
(A) 0.033 (B) 0.358 (C) 0.183 (D) 0.029 (E) 0.170

43. The 95% confidence interval for the odds ratio is $0.48 < OR < 0.99$. It follows that men who have been circumcised by age 3 are more likely to develop herpes and this result is statistically significant at the $\alpha = 0.05$ level.
(A) True (B) False

44. Both the relative risk and the odds ratio can be calculated for this cohort study.
(A) True (B) False

Information for questions 45 - 50

The following table shows trunk diameter Y (in cms) at shoulder height of chestnut trees and age X (in years) for 14 trees.

Diameter (Y)	Age (X)
2.0	4
2.5	8
7.5	8
8.8	10
8.8	13
11.3	16
13.8	20
11.8	23
15.0	28
15.0	30
20.0	33
17.5	35
17.5	38
18.8	42

$\bar{x} = 22.00$

$\bar{y} = 12.16$

$\sum (x_i - \bar{x})^2 = 2028.00$

$\sum (y_i - \bar{y})^2 = 423.11$

$\sum (x_i - \bar{x})(y_i - \bar{y}) = 869.10$

45. The slope (β_1) of the least squares regression line of Y on X is closest to
(A) 0.43 (B) 2.05 (C) 4.79 (D) 2.33 (E) 0.21

46. The equation of the least squares regression line is closest to
(A) $\hat{y} = 7.54 + 0.21x$ (B) $\hat{y} = 12.16 + 0.43x$
(C) $\hat{y} = 2.70 + 0.43x$ (D) $\hat{y} = 2.70 + 2.33x$

47. The standard deviation of the sample points about the fitted regression line is $s_e = 2.05$. The estimated standard error of the slope of the regression line is closest to
(A) 0.10 (B) 0.05 (C) 4.20 (D) 2.05 (E) 0.22

48. A 95% confidence interval for the regression slope requires a t-value with degrees of freedom equal to
(A) 13 (B) 24 (C) 14 (D) 26 (E) 12

49. The value from the t-table for the 95% confidence interval in the previous question is
(A) 2.160 (B) 1.782 (C) 1.960 (D) 2.179 (E) 1.771

50. The 95% confidence interval for the regression slope is $0.33 < \beta_1 < 0.53$. There is evidence at the 5% level of significance that increased tree age results in greater trunk diameter.
(A) True (B) False