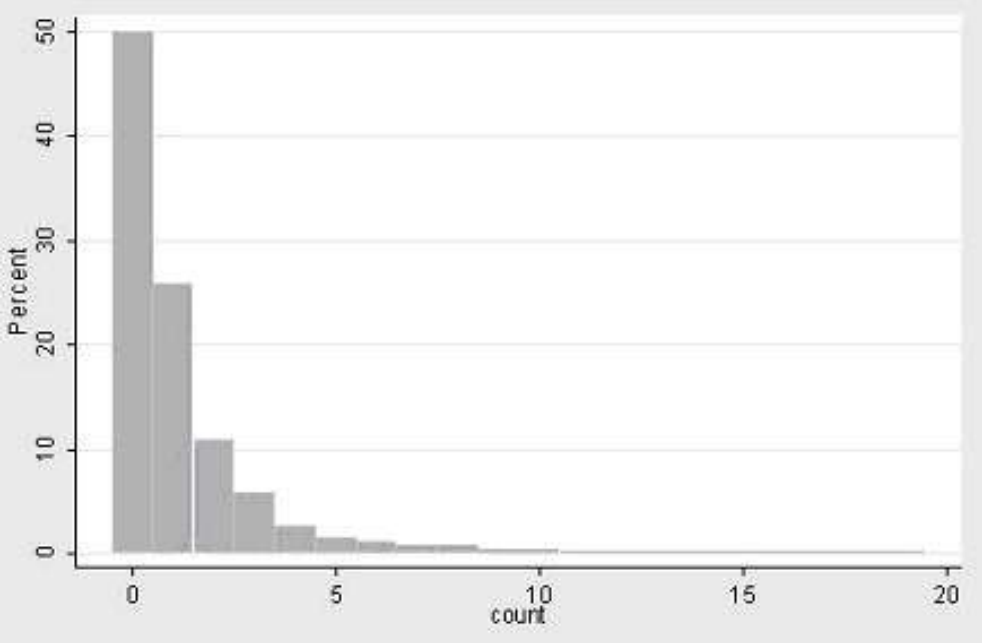
Statistics Multiple Choice Test

For each question choose ONE statement that you consider to be the most accurate.

1. In a study of heart disease, diastolic and systolic blood pressures were measured. A person was defined as 'hypertensive' if their diastolic pressure was above 90 mmHg and their systolic pressure was above 140 mmHg. The variable 'hypertensive' is:
   a) binary
   b) paired continuous
   c) biased
   d) continuous

2. In the study described in question 1, a significance test was conducted to assess whether the proportion of 'hypertensive' men was different to that in women, and produced p<0.004. This P-value means that if the proportions of men and women in the population that are hypertensive are the same, we would expect results like these, or more extreme:
   a) less than 4 times in 100
   b) less than 4 times in 10,000
   c) less than 4 times in 1,000
   d) less than 1 time in 4,000

3. This histogram shows the count of a parasite found in blood specimens from a series of patients.



The distribution of the counts is skewed to the right, which implies that:

a) you must never quote the mean – only the median is valid
b) the median is the best measure of centre
c) anything you calculate from these data is strongly influenced by the large values, so it is best to drop them before making any calculations
d) the standard error is the best measure of how widely the observations vary

4. You have conducted a Z-test on the difference in the proportions of men and women that smoke, using data from a sample of men and a sample of women. The result of the test gives $p<0.001$.

a) the null hypothesis is that the proportions in the two samples are the same
b) we can interpret $p<0.001$ to mean that the null hypothesis is definitely false
c) we can interpret $p<0.001$ to mean that the alternative hypothesis is definitely true
d) the null hypothesis is that the proportions of men and women that smoke in the populations from which the samples were taken are the same.

5. The systolic blood pressure of a random sample of 80 men was measured and the mean found to be 122 mmHg. The standard error of this estimate was 4.25 mmHg.
a) The true population mean is 122 mmHg.
b) A 95% confidence interval for the population mean is from 113.67 mmHg to 130.33 mmHg.
c) We cannot calculate a confidence interval for the mean from the information given.
d) A 95% confidence interval for the population mean is from 120.04 mmHg to 123.96 mmHg.

6. The difference between two percentages is tested by a Z-test, which shows that $p=0.089$
a) there is weak evidence for a true population difference in the percentages
b) the null hypothesis is true
c) the null hypothesis is false
d) it would be advisable to confirm this result using a chi-squared test

7. The mean and standard deviation of the birth weights of a representative sample of 153 newborns are 3250 grams and 428 grams respectively. On the basis of these figures, a 95% confidence interval for the population mean birth weight runs from 3181 to 3319 grams.
a) 95% of the individual newborn birth weights are between 3181 and 3319g
b) the mean birth weight in this sample of 153 newborns is probably between 3181 and 3319g
c) the mean of the population from which the 153 newborns came is definitely between 3181 and 3319 grams
d) we can be 95% confident that the range from 3181 to 3319 grams includes the mean of the population from which the 153 newborns were taken

8. The difference between two proportions is tested statistically by both a Z-test and by a chi-squared test. Z is 1.31 and chi-squared is 1.72.
   a) the P-value is greater than 0.05 ($p>0.05$)
   b) the P-value is less than 0.001 ($p<0.001$)
   c) neither (a) nor (b) are correct and the P-values from the two tests will be different
   d) only the Z-test is valid because it makes a direct comparison of two proportions

9. A cross-sectional survey including 1,461 men found that of 959 men who work near home, 8.1% were HIV infected, and among 502 men who work away from home, 12.4% were HIV infected.

The difference in HIV prevalence was 4.3% with a 95% confidence interval (CI): 1.0% to 7.4%.

   a) the 95% CI shows that we have underestimated the true difference in HIV prevalence by as much as 7.4%.
   b) the 95% CI is not valid because the two samples of men are different sizes.
   c) there is good evidence against the null hypothesis of no difference in HIV prevalence in the two groups, because the 95% CI does not include 0%.
   d) the difference in HIV prevalence is not valid – an odds ratio should be used instead.

10. The table below shows the prevalence of a disease by age group in a sample of adults over the age of 30.

Age (years)	30–39	40–49	50–59	60+
Sample size	63	59	50	69
Disease prevalence	35%	39%	48%	52%

The overall chi-squared is 4.90 with $p=0.18$; Chi-squared test for trend is 4.77 with $p=0.03$.

   a) there is no evidence for any change in disease prevalence with age because the overall chi-squared test is not statistically significant
   b) the sample sizes are not the same in each age group so the results cannot be trusted
   c) there is reasonable evidence that disease prevalence increases with age
   d) the chi-squared test for trend is not appropriate because only three of the age groups have the same age range of 10 years

11. Disability was assessed in a national survey using the Activities of Daily Living (ADL) checklist. The results are presented in the table below and a 2x2 chi-squared statistic is calculated.

	Disability		
Gender	***Yes***	***No***	Total
Men	1123	1470	2593
	43.30%	56.70%	100.00%
	39.10%	55.00%	46.70%
Women	1752	1204	2956
	59.30%	40.70%	100.00%
	60.90%	45.00%	53.30%
Total	2875	2674	5549
	51.80%	48.20%	100.00%
	100.00%	100.00%	100.00%

The value of the chi-squared statistic is x^2 (1) = 140.9, p<0.001.

a) There is no evidence of an association between gender and disability
b) More men than women have disability in carrying out Activities of Daily Living.
c) There is strong evidence for an association between gender and disability in carrying out Activities of Daily Living.
d) The sample size is small, the Chi-squared is not appropriate for these data.

12. The mean Body Mass Index in two large samples of adults from different age groups are 20.6 Kg/m^2 and 26.1 Kg/m^2. The 95% confidence intervals for the population means are 17.1 to 24.1 and 22.8 to 29.4 Kg/m^2 , respectively.

a) because the two confidence intervals overlap, we know that there is only weak evidence for a difference between the population means
b) these confidence intervals cannot be compared unless we know the sample sizes in each group
c) a significance test is needed before we can say whether the two group means differ
d) to compare the two groups it would be better to quote 99% confidence intervals

13. A 95% confidence interval for a population proportion has been calculated based on the proportion in a random sample of 50 people. If a random sample of 500 people had been taken instead:
    a) the 95% confidence interval would have been 10 times wider
    b) the 95% confidence interval would have been one tenth as wide
    c) the 95% confidence interval would have been about one third as wide
    d) the 95% confidence interval would be the same width.

14. The table below shows the results from a study to assess an educational intervention to promote breast feeding. Mothers in the intervention group attended extra sessions about breast feeding during antenatal care. The outcome was breastfeeding 3 months after birth.

	Breast-feeding at 3 months				
	No		Yes		Total
Control	170	59%	81	48%	251
Education	117	41%	86	52%	203
Total	287		167		454

The study report stated that the risk ratio for the effect of the intervention on breast feeding at 3 months is 52%/48%, which equals 1.08.

    a) this ratio is the correct measurement of the intervention effect
    b) this ratio is not correct because it uses column percentages
    c) a risk ratio should never be quoted before a significance test has been conducted
    d) the sample size is too small for the risk ratio to be valid

15. In the table for question 14, the chi-square value is 4.92, p=0.027
    a) the chi-squared and P-value are not appropriate here, because the table shows column percentages, not row percentages.
    b) the chi-squared and P-value are appropriate.
    c) there is no evidence for an effect of the intervention because the P-value is not small enough
    d) it is not appropriate to use a chi-squared test to compare two percentages which are used to calculate a risk ratio; a Z-test of proportions should be used instead.