Name: _____

This examination has 11 pages. Please check it as your first step.

Answers are to be written on the question paper, in the spaces provided.

Unless specifically asked to in a particular question, you do not need to complete any detailed calculations. Instead, answer as though you were setting up the computing task for your research assistant to complete.

If there is a possibility of ambiguity, be clear about 1- and 2-sided hypotheses, levels of confidence, degrees of freedom, whether 2 sample or 1 sample procedures, paired or unpaired, which table or formula from your textbook is relevant, etc.

Attempt all questions. The allocation of points will be as follows:

Q1 __/10
Q2 __/25
Q3 __/20 (best 4 of the 5 portions a-e)
Q4 __/50
Q5 __/25 (best 5 of the 6 portions a-f)
Q6 __/60 (best 6 of the 7 portions a-g)

total ___/180

Since the points add up to a total of 180, and since the examination is about 3 hours or 180 minutes, you can plan accordingly.

Q1 A new meaning for the abbreviation 'SE' ??? 10 POINTS

The following is part of a table in a recent paper from the Annals of Internal Medicine on a randomized placebo-controlled trial of low-dose aspirin in patients with chronic stable angina (paper courtesy of Leslie Brailsford)

"Baseline Characteristics of Participants with Chronic Stable Angina in the U.S. Physicians' Health Study

Characteristic	Aspirin Group (n=119)	Placebo Group (n=102)
Mean age, years	63.6 ± 9.3	62 4 ± 8.6
Mean systolic blood pressure, mm Hg	132.5 ± 13.0	132.5 ± 14.4
Mean diastolic blood pressure, mm Hg	80.3 ± 7 8	80.2 ± 7.9
Mean cholesterol level, mmol/L	5.9 ± 1.1	5.8 ± 1.3

Plus-minus values are mean ± SE"

If you were checking this paper for typographical and other errors before it was published, what SE (Statistical Error !!!) would you have noticed?

Q2 Judging the error in a measurement or average of measurements 15 POINTS

[If it helps, assume Gaussian variation for both parts of this question]

Say whether each assertion is true or false, and give a reason.

a "If all you have is one measurement, you cannot estimate the likely size of the random error in it--you would have to take another measurement, and see how much it changes" 5 POINTS

b "If all you have is one hundred measurements, you cannot estimate the likely size of the random error in their average--you would have to take another mean, and see how much it changes" 10 POINTS

Q3 Premature Death in Jazz Musicians 20 POINTS

"Comparison with national values showed that 70 (82%) of the musicians exceeded their life expectancy" (paragraph 7)

Part 1: Leave aside the possibility of a biased comparison. To keep it simple, consider a random sample (n=86) of persons (all of one specific race and sex) with a certain characteristic identified at birth in 1880, whose longevity is compared with the life expectancy (average lifetime) of their sex- and race-specific 1880 birth cohort. Also, assume that every one (including each of the 86) of the birth cohort is now dead and that you have the complete lifetable.

a What is the approximate shape of the frequency distribution of lifetimes in this 1880 birth cohort? 5 POINTS

b Choose an index (i.e. a measure) which summarizes the central tendency of lifetimes and formulate the null hypothesis in terms of this index. 5 POINTS

c Approximately (or precisely if you can say so) what percentage of the 1880 cohort lived past the index of central tendency for the cohort? Explain your answer. 5 POINTS

d For the index you chose, suggest a data-analysis which tests the hypothesis that persons with this certain characteristic of interest have shorter lifetimes than those in the 1880 birth cohort. 5 POINTS

e Suggest a 2nd index and give a corresponding test. 5 POINTS

Q3 Premature Death in Jazz Musicians (continued)

Part 2: OPTIONAL (for students who wish to comment on validity of the comparison made by author) In one or two sentences, explain one bias in the author's comparison; and say whether it gives the jazz musicians an inbuilt longevity advantage or disadvantage?

[Hint: there are two biases: one explains why, on average, bishops live longer than priests (or full professors longer than assistant professors??); the other is because of the synthetic/artificial nature of lifetables]

Q4 Paracetamol and Fever 50 POINTS

a	Entry was limited to children with temperatures between 38°C and 41°C. Given the mean of 38.9 °C and the SD of 0.9, what can you say about the shape of the frequency distribution over the 38°-41° interval? 5 POINTS

b	"We estimated a sample size requirement of 210 subjects completing the trial" (Sample size –– paragraph 5 of Methods) Give the formula by which the authors estimated this (identify what numbers go with what parameters, but leave the details to your assistant [who has not taken a statistics course]) 5 POINTS

c	"Student's t- test and Mann-Whitney (alias Wilcoxon) test..." (Statistical testing –– paragraph 5 of Methods) Why did the authors use the Mann-Whitney (alias Wilcoxon) test? In light of the n's and the shape of the distribution of duration of fever, was their concern about the use of the t test justified? 5 POINTS

Q4 Paracetamol and Fever (continued)

d	"The mean duration of fever..." [paragraph 4 of Results] Explain in a sentence, in non-technical words, the phrase "the differences were statistically non-significant" 5 POINTS

e	"The 95% CI for the differences between the paracetamol and placebo groups for duration of fever was -10.0 to +7.1 h" Explain in non-technical words what this statement says. 5 POINTS

f	How does this CI add to what is shown in Figure 1? 5 POINTS

g	How was the CI calculated? 5 POINTS

Q4 Paracetamol and Fever (continued)

h Before the study, the authors anticipated a SD of 2 days (48 hours) for the duration of fever. The SD of the duration of fever observed in the n=225 is not reported explicitly.

How could one reconstruct this SD from the results given [assume that the SD is the same in the two treatment groups]? 5 POINTS

i "Children..were more likely to be rated.as having at least a 1-category improvement in activity...." [2nd last paragraph of Results]

What tests could be used to compare the two groups? Do they all give the same answer? 5 POINTS

j *"On the basis of ...completing the trial"* [sample size considerations, first sentence of paragraph 5 of Methods]

"There were no significant differences between groups in mean duration of subsequent fever" [Abstract]

If these two statements were the ONLY information you were given about the trial, what could you conclude? 5 POINTS

Q5 Melatonin and Delayed Sleep 25 POINTS

a	What sample size formula or table would you have shown the authors if they had consulted you concerning sample sizes before doing their study? 5 POINTS

b	What is it about the study design that makes the required sample size so much smaller than that in Kramer's study? 5 POINTS

c	What do <u>you</u> consider would be a clinically significant advance in sleep onset time? 5 POINTS

d	"In all 8 subjects sleep onset time was earlier during melatonin treatment than during placebo" [Abstract] List 3 possible tests of these data, putting them in order of increasing statistical power [do not carry out the tests, but give references] 5 POINTS

Q5 Melatonin and Delayed Sleep (continued)

e	Set up the calculation from which the p<0.01 for the 3.49 versus 2.12 [Table II, sleep onset time, melatonin versus placebo] was derived 5 POINTS

f	Are there sufficient data in the table to allow you to verify the calculation? 5 POINTS

Q6 Triazolam and Memory Impairment 60 POINTS

a "5 of the 6 ... , with a total of 12 episodes.." [Abstract]

Which is more statistically correct? Why? 10 POINTS

(i) to base inferences from the observed rate of 5/6 on the Binomial distribution with n=6?

(ii) to base inferences from the observed rate of 40% [12 episodes in 30 subject-drug nights] on the Binomial with n=30?

b "In the temazepam group there were no such episodes of memory impairment"

From these data, find a 95% confidence interval for the <u>proportion of persons</u> who will have episodes of memory impairment with this drug. 10 POINTS

c "this rate for triazolam was statistically significant from those for either the temazepam or placebo group (each $p<0.05$)" [3rd sentence, 2nd paragraph of Results –– daytime memory assessment]

What test would you perform here [limit yourself to triazolam versus placebo; be clear about denominators, degrees of freedom, sample sizes etc.]? 10 POINTS

Q7 Triazolam and Memory Impairment (continued)

d Convert the 12 observations in each row of the figure (top right, p829) into a separate 2x2 table for each subject.

What statistical test would you perform on the data for each separate 2x2 table? Are such "n=1" tests legitimate?. 10 POINTS

e "a small sample size predisposes to type II error" [3rd paragraph of Discussion]

Use ordinary words to explain what type II error is, and why it is more likely with small n's. 10 POINTS

f Can a very small sample size predispose to type I error? [You may wish to use subjects 2 and 3 to make your point] 10 POINTS

g What about Type I error with more moderate size samples? 10 POINTS

Q5 Limb-reduction defects and chorion villus sampling

SIR,--Dr Firth's report (March 30) on transverse limb-reduction defects after chorion villus sampling (CVS) at 56-66 days' gestation prompted us to examine our registry data. The possibility of an association between CVS and birth defects is a matter of such concern that a report of our preliminary findings is warranted. From 6604 cases of malformation reported in 1988-90 we selected the 118 with a transverse limb-reduction defect, this being the one in all of Firth's cases. The controls were the other 6486 malformed cases. 4 cases of limb-reduction and 15 controls were delivered after CVS (odds ratio 15.14, 95% confidence interval 4.18-49.65; p[Fisher's exact]=0.000305). (Letter to The Lancet Vol 337: May 4,1991)

a Write down the null hypothesis (in symbols, rather than in words)

b Write down an alternative hypothesis (again, in symbols). Say why you chose the one you did.

c Write down a table showing the reported data.

d List the other tables that Fisher's approach considers are even greater evidence for a positive association than the observed one.

Q5 Limb-reduction defects and chorion villus sampling (continued)

e Write down the expression for, but do not calculate, the (conditional) probability of the observed table.

f What further calculation is needed to obtain the author's p-value (the author's alternative hypothesis is one sided)

g If the number of cases of limb-reduction delivered after CVS had been 6 rather than 4, would Fisher's Exact test have been indicated? Why/why not?