

GILLES PARADIS
NAME

86/92

Instructions: Write answers in spaces provided. If you wish, you may add in any calculations/reasons for your choice of answer. When describing a statistical test, mention whether it is one or two sided and the degrees of freedom (if appropriate).

5/5 1. What is the mode 5, median 6, mean 7, and (approximate) standard deviation 3 of the set of 9 observations 4, 5, 5, 5, 6, 6, 9, 10, 13?

5/5 2. Rank the following 3 distributions in terms of relative spread, using the ranks 1 = "most concentrated" to 3 = "most spread out".

X = height Y = height Z = weight

$$\mu_X = 60'' \quad \mu_Y = 150 \text{ cm} \quad \mu_Z = 60 \text{ kg}$$

$$\sigma_X = 2'' \quad \sigma_Y = 3 \text{ cm} \quad \sigma_Z = 5 \text{ kg}$$

Rank: 2 1 3

3. Rank each of the following 3 observations in terms of the percentiles they fall into in their respective Gaussian distributions. Use the ranks 1 = "lowest centile value of the 3" to 3 = "highest of the 3".

$$x = 73'' \quad y = 160 \text{ cm} \quad z = 1.90 \text{ m}$$

$$\mu_x = 68'' \quad \mu_y = 150 \text{ cm} \quad \mu_z = 1.80 \text{ m}$$

$$\sigma_x = 2'' \quad \sigma_y = 6 \text{ cm} \quad \sigma_z = 0.05 \text{ m}$$

Rank: 3 1 2

4. Rank the following 3 sample means from 1 = "least reliable" to 3 = "most reliable".

$$\bar{x} = 100 \text{ cm}$$

$$\bar{y} = 50 \text{ kg}$$

$$\bar{z} = 20 \text{ years}$$

$$n_x = 100$$

$$n_y = 25$$

$$n_z = 100$$

$$s_x = 20 \text{ cm}$$

$$s_y = 10 \text{ kg}$$

$$s_z = 5 \text{ years}$$

Rank: 1 3 2

5. For each of the following types of data, say which of the 5 statistical or probability distributions

Binomial Normal (Gaussian) Poisson Uniform Other

best approximates it.

Data

Distribution

a) Birthweights. Normal ✓

b) Numbers of Sudden Infant Death Syndrome (SIDS) cases from year to year. Poisson ✓

c) Numbers of children born in different months of year. Uniform ✓

d) Number of decayed/missing/filled teeth (out of 32) per person in a population of adults. Binomial ✓ other

e) The percentile scores of a random sample of candidates taking a standardized test. Normal ✓ uniform

f) The \bar{y} 's obtained from different samples of (n=)20 lengths of stay in a hospital. Normal ✓

g) The numbers of drivers wearing seatbelts in different random samples of (n=)20 drivers. Normal (?) Binomial → Normal if $\pi \rightarrow 0$

1% of the population
in each percentile.
It is the values of each
percentile in terms of raw score
that will vary

6. Which of the following statements are TRUE, which are FALSE and which WOULD BE TRUE IF n LARGE ENOUGH?

- a) 95% of the observations that form a Gaussian distribution are within 2SD of the mean. True
- b) 95% of the observations that form a Binomial Distribution are within 2SD of the mean (consult Binomial Tables if you like). True if n is large enough
- c) The \bar{y} 's from different samples form a Gaussian distribution. True if you sample from the same population
- d) The SD of the numbers of decayed/missing/filled (DMF) teeth in a sample of adults was 20 DMF teeth (see 5d). FALSE
- e) The SD in a larger sample is, as a rule, larger than the SD in a smaller sample. False
- f) To halve the standard error of the mean, one must double n . False

7. A snow-removal contractor offers you a choice of

- a) a contract, costing \$100, to clear your driveway/sidewalks after each major snowstorm of the winter. OR
- b) the same service, but on a snowstorm by snowstorm basis, which will cost you \$20 per snowstorm.

You look up the meteorological records for the past century and find that the average number, λ , of snowstorms per winter is 7.

- If you base your probabilities on the Poisson distribution, what are the chances that, in this coming winter, there will be
- a) exactly 5 snowstorms 2.01%
- b) more than 5 70%
- c) less than 5 17%
- What is the mean 7, the variance 7 and the modal 7 number of snowstorms per winter if the distribution does in fact follow a Poisson distribution? this will be 2 mode 6 and 7

8. A recent study examined the change in knowledge of risk factors for breast cancer and breast self examination techniques following a 2-hour workshop for $n = 16$ female staff. Each participant's knowledge was tested before and 6 months after the workshop. On both the 'pre' and 'post' test, 'knowledge' was measured using the same 30 item quiz, which produced a score of 0 if all questions were answered incorrectly to 30 if all answered correctly. The following data resulted

	'Pre' Scores	'Post' Scores
mean	14	18
SD	7	6

• What is an appropriate statistical test of whether knowledge is increased following the workshop? Paired sample "t" test, one tailed, $df = 15$

• Do you have enough information to perform the test? No, you cannot agree

• If not, what more do you need? The SDs you need the individual scores to calculate

• If the appropriate test indicated a statistically significant increase in scores, can you conclude that the workshop did in fact improve knowledge? we can conclude

that there is an increase in the mean score after the workshop and after passing the post test. Whether this indicates a increase in knowledge is another story

• Suggest one improvement in the design of this evaluation study. Increase in the mean score may only reflect that I remember the questions in the workshop. I may not want to choose a different test trying to measure the same construct.

also might have a control group

Gilles PARADIS

Student's Name

post test

9. For each of the following confidence intervals, give the corresponding p values (eg. $p < 0.05$, $p < 0.01$, $p < 0.05$, $p > 0.05$ etc)

- 95% C.I. for μ : 47 ± 3 $44-50$ Test of $\mu = 45$: $p > 0.05$
- 99% C.I. for $\mu_Y - \mu_X$: 10 ± 6 $4-16$ Test of $\mu_Y = \mu_X$: $p < 0.01$
- 95% C.I. for $\mu_Y - \mu_X$: 5 ± 4 Test of $\mu_Y = \mu_X$: $0.01 < p < 0.05$
- 99% C.I. for $\mu_Y - \mu_X$: 5 ± 6

10.

PLASMA CAFFEINE CONCENTRATIONS IN OUTPATIENTS

SIR.—Concern has been expressed over the possible role of excessive caffeine ingestion in the development of various diseases, including cancers of the urinary tract^{1,2} and coronary heart disease arising from hyperlipidaemia.³ Excessive caffeine intake can also result in "caffeinism", a syndrome manifesting as an anxiety state with symptoms of irritability, headache, agitation, and nervousness.^{4,6} Caffeine withdrawal is associated with headaches, irritability, and drowsiness.^{5,8} Both conditions may be misdiagnosed.

Caffeine is consumed largely in coffee, tea, and cola drinks but also in non-prescription analgesics, chocolate, and "tonic" preparations, including 'Lucozade'.

Most data on the harmful effects of caffeine have involved surveys of the consumption of caffeine-containing beverages in which the number of cups drunk per day is the measure of caffeine intake. Gilbert^{9,10} has questioned this approach since there are large differences in the caffeine content of beverages according to the method of preparation. Ground coffee beans produce a drink generally containing more caffeine than the drink prepared from instant coffee and tea. Published estimates of the caffeine content of cups of tea and coffee reveal marked differences from study to study.⁹ Absorption rates of caffeine from different beverages also vary and peak plasma caffeine levels achieved after ingestion of a given amount of caffeine depend on the type of drink.¹¹ Some, though by no means all, of the clinical effects of caffeine are related to its concentration in the blood. It is therefore surprising that plasma levels of caffeine have not been more extensively studied as an index of caffeine status.

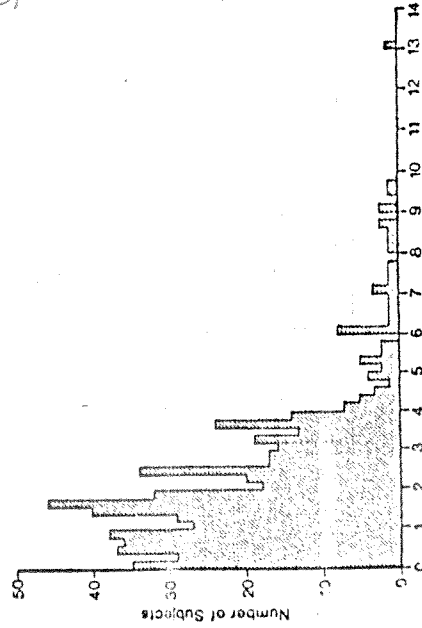
We have measured random plasma caffeine concentrations in 600 adults (300 men and 300 women) who presented to this department for biochemical investigations as outpatients. The patients were not fasting and no restriction had been placed on their caffeine intake. Caffeine was measured by radioimmunoassay¹² with ¹⁸-H₃-Hj-caffeine tracer and an antiserum kindly donated by Dr C. E. Cook. The inter-batch coefficient of variation was 5.5%. Results are shown in the figure.

Values for plasma caffeine ranged from less than 0.2 mg/l to 13.1 mg/l. The mean concentration was 2.12 mg/l and there was no significant difference between men and women. In 21 men and 14 women (i.e., 5.8% of the sample), the plasma caffeine concentration was less than 0.2 mg/l, and in 95% of the population it was less than 5.6 mg/l. The mean peak plasma caffeine level one hour after consumption of two cups of strong coffee is about 5.3 mg/l.¹¹ In the sample reported here, 18 men and 16 women, (5.7% of the sample), had values in excess of this, suggesting a high daily intake of caffeine.

We believe that these findings will be of interest to workers involved in epidemiological and clinical studies of caffeine.

Department of Clinical Biochemistry,
St Luke's Hospital,
Guildford, Surrey GU1 1NT

JANET M. SMITH
SHEENA PEARSON
VINCENT MARKS



Plasma caffeine concentrations.

Plasma Concentrations in Outpatients

1. What kind of sample is this? What kind of sample would you have selected? (Answer in terms of person, place and time.)

10 They probably took a systematic sample of subjects presenting at the clinic. I would use non-patients, in non-hospital or clinic context (ideally ordinary individuals in a variety of context even though this is not really feasible) and at various times of the day (random selection of context)

2. What do the authors mean when they say the "interbatch coefficient of variation was 5.5%"? It means that the variation in the plasma caffeine level

4 distribution was only 5.5% between the various batch of tests they did. It indicates the reliability of the measurement, to the means of these batches?

3. Is the mean [of 2.12 mg/l] a good descriptor of the findings? Why/Why not? Give a sentence summary of your own (using guessed-at summary statistics if not available in text). No because the distribution is skewed to the right. One might want to say most of the people caffeine val were from 0 to 1 with the median of 1 1/2 mg/l for example. - 3 give the 90iles.

4. What statistical test would you use to compare concentrations in men and women? Independent sample t test / two tailed

5. Do you agree with the statement that 5.7% of the sample probably have a high daily intake of caffeine? Not necessarily, it depends when these people had their last cup of coffee. It may just be that they had their last cup of coffee just before the test. Also different people may have different

GILLES PARADIS

Student's Name