1 An authority on the physiology of sleep claims that, for every hour of sleep lost below a person's normal requirement, relative IQ decreases ten points. Suppose that he based his claim on trials with 30 subjects, who completed intelligence tests after varying amounts of sleep deprivation. The following results were obtained from an analysis of the experimental data using simple linear regression.

a (intercept) = 100.2 $s_a = 5.6$
b (slope) = 10.5 $s_b = 2.9$

Source of variation	DF	SS	MS
Total	19	1118.8	
Regression	1	526.5	
Residual		592.3	

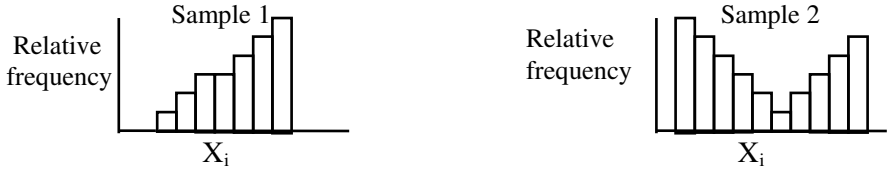
a) Complete the above regression ANOVA table.

b) Write the equation that relates IQ (Y) to number of hours sleep lost (X).

c) Present a complete statistical analysis to test the hypothesis that there is no relationship between IQ and amount of sleep lost (i.e., that the slope of the regression line equals zero).

d) What proportion of total variation in IQ is explained by the regression results above?

2 Plotted below are frequency distributions of two random samples ($n_1 = 125$, $n_2 = 158$). You are asked to test the hypothesis that the samples were drawn from identical populations.



a) What statistical test do you choose?

b) Why?

3 Research question: does the size of clutches laid by American Robins nesting at northern latitudes differ from clutch sizes laid by southern Robins? You recorded clutch size of ten randomly selected northern nests and 16 randomly selected southern nests. Independently collected data suggest that clutch sizes were normally distributed, with equal variances among clutches at northern and southern latitudes.

$H_0: \mu_1 = \mu_2$ $\overline{X}_1 = 5.0$
$H_A: \mu_1 \neq \mu_2$ $\overline{X}_2 = 2.0$
$\alpha = 0.05$ $s_{\overline{X}_1 - \overline{X}_2} = 1.2$

Complete a statistical analysis to answer the research question. For full credit, include the following:

1) Calculate the test statistic, t_{calc}. Show all formulas and values used.
2) State your statistical decision about the null hypothesis.
3) State your conclusion, in words, about the research question.

4 For a sample of 10 measurements, the mean ($\overline{X}$) is 10.0, and the variance (s^2) is 40.00. What are the 95% confidence limits for the population mean (μ)?

5 Calvin awoke late for school, but he remembered that his social science report was due that day. He decided to determine the average number of pieces of litter per block, for the 20 blocks between his home and school. Since Calvin was late, he could not count litter on each block. Instead, he decided to take a sample. He ran until he got tired, and then he counted the litter on the block where he stopped to rest. Then he resumed running until he tired, and again counted litter on the block where he stopped to rest. Using this sampling scheme, Calvin counted litter on 5 out of the 20 blocks. Calvin's companion, Hobbes, predicted that the teacher would not be impressed. Hobbes was correct: the teacher made Calvin start over because the study was flawed statistically.

a) Explain the statistical flaw in Calvin's study.

b) What detailed instructions could you give to Calvin to improve his study?

6 Research question: do body lengths of wild salmon fry vary more than body lengths of hatchery raised fry? Designate wild fry as population 1, and hatchery fry as population 2.

a) State the null and alternative hypotheses.

b) Body lengths of twenty randomly selected wild fry were measured. $s_1^2 = 11.0$.

Body lengths of 26 randomly selected hatchery fry were measured. $s_2^2 = 5.0$.

Assuming a significance level of 0.05, complete a statistical analysis to answer the research question. For full credit include the following:

1) Calculate the test statistic. Show all formulas and values used.
2) State your statistical decision about the null hypothesis.
3) State your conclusion, in words, about the research question.

7 Tannin concentrations were measured in ten acorns sampled from each of four species of oaks. Single factor analysis of variance rejected the hypothesis that mean tannin concentrations were equal in acorns from all four oak species. Determine which means differ, using the ANOVA table below and a significance level $\alpha = 0.01$. For full credit, please show your work, including formulas that you use.

Species	Mean tannin concentration $(\overline{X}_i)$
Quercus lobata	11.0
Quercus agrifolia	9.1
Quercus douglasii	7.9
Quercus chrysolepis	10.7

Source of variation	Sum of Squares (SS)	Degrees of freedom (DF)	Mean Square (MS)
Total	102.0	39	
Species	30.0	3	10.0
Error	72.0	36	2.0

8 You are studying growth rates in Douglas fir seedlings at ambient and elevated concentrations of CO_2. You want to be able to detect a difference between the means of your two samples of 1.0 mg/day with 90% power. Your significance level is 0.05. Preliminary data suggest that the variances in your samples will be equal, and approximately 1.0 $(mg/day)^2$. You have enough materials to measure the growth of 20 seedlings in each sample. Will this be enough for 90% power, or do you need to ask your boss for additional funds to collect larger samples? For full credit, please show your work.

9 Does drinking extra coffee help people stay awake (e.g., to study)? To study this question, 16 people were divided into four groups of four. Each group drank one of four volumes of coffee. The table below shows the *d*, the volume of coffee consumed (ml); $X = \log d$, the log-transformed volume (log to base *e*); *Y*, a measure of alertness in study subjects; $\overline{Y}$, the mean alertness of the four subjects that consumed volume *d*; and *s*, the standard deviation of alertness at each coffee volume.

d	$X = \log d$	Y_1	Y_2	Y_3	Y_4	$\overline{Y}$	***s***
1	0.00	21	24	26	25	24.0	2.2
10	2.30	36	38	36	35	36.3	1.3
100	4.61	43	47	45	49	46.0	2.6
1000	6.91	54	56	58	53	55.3	2.2

The following results were obtained from an analysis of the data above using simple linear regression.

a (intercept) = 24.85 $s_a = 0.89$
b (slope) = 4.49 $s_b = 0.21$

Source of variation	DF	SS	MS
Total		2205.75	
Regression		2142.45	
Residual		63.30	

a) Write a simple linear regression model that relates alertness (*Y*) to volume of coffee consumed (*d*). Use $X = \log d$, where *d* is the coffee volume as the independent variable. Use only symbols in your model; do not use numbers. Identify all terms in the model.

b) Write an equation that relates that relates alertness (*Y*) to volume of coffee consumed (*d*). Use $X = \log d$, and insert numerical estimates for all parameter values.

c) Complete the above regression ANOVA table.

d) Present a complete statistical analysis to test the hypothesis that there is no relationship between alertness and volume of coffee consumed (i.e., that the slope of the regression line equals zero).

e) What proportion of total variation in alertness is explained by the regression results above?

f) Predict alertness of individuals who drink the following volumes of coffee: (i) 100 ml; (ii) 150 ml.

10 Research question: What determines survival rates of grizzly bears in the Greater Yellowstone Ecosystem? Using data from radio-collared bears obtained over a 20-year period, a researcher fits seven models to grizzly survival data, with the following results.

Model	Model variables	K	AIC_c	ΔAIC_c	$\exp(-0.5\Delta_i)$	w_i
1	none (survival rate estimated using mean)	2	1140.1		0.000	
2	road density	3	1120.3		0.006	
3	human visitation rate	3	1121.1		0.004	
4	whitebark pine nut production	3	1113.0		0.235	
5	road density, human visitation	5	1113.5		0.183	
6	road density, pine nut production	5	1111.7		0.449	
7	roads, humans, pine nut production	9	1110.1		1.000	
					sum = 1.877	

a) Complete the table above, by filling in values for ΔAIC_c and w_i.

b) Which model performs best according to Akaike's Information Criteria? How do you know?

c) What is the probability that the model identified in (b) really is the best among the models considered?

d) What is the confidence set for the best model, among the models considered?

e) Explain why the analysis outlined above provides a more informative answer to the research question than would statistical hypothesis testing (e.g., performing a series of regressions to determine whether the effect of each variable differs significantly from zero)?

11 Four models were fit using data on numbers of Pacific herring spawning in eelgrass habitat at Cherry Point and predictor variables listed below.

Model	Variables included
1	Herring abundance @ Cherry Pt. 3, 4, and 5 years ago
2	Herring abundance @ SW Vancouver Island in preceding year
3	Contaminant concentrations in water samples near spawning sites
4	Ocean conditions: SST, PDO index, upwelling index

Below are results of model selection analysis. Akaike weights are listed under the column headed "w_i." Also listed are (hypothetical) predictions for herring abundance next year using each model

model	ΔAIC_c	w_i	Predicted number (10^6 fish)
1	1.20	0.310	10.1
2	3.10	0.120	30.0
3	9.30	0.005	21.1
4	0.00	0.565	12.5

a) Which individual model best explains herring abundance at Cherry Point? How do you know?

b) Predict the number of herring that will spawn at Cherry Point next year by combining predictions from all of the models.

c) Explain why the prediction in (b) using all models likely would be more accurate than the prediction using only the best single model?