1. You have data drawn from a normal distribution with a known variance of 16. You want to test if the mean of your distribution is equal to 2 with significance level $\alpha = 0.05$.You collect 16 data points with sample mean 1.5.

   (a) Write down the null and the alternative hypothesis.

   (b) Which test statistic are you going to use? Draw the distribution of your test statistic under the null hypothesis, and draw your rejection region.

   (c) Find the p-value for this hypothesis and decide if you will reject H_0.

2. Data is drawn from a binomial(5, θ) distribution, where θ is unknown. Here is the table of probabilities $p(x|\theta)$ for 2 values of θ:

x	0	1	2	3	4	5
$\theta = 0.5$	0.031	0.156	0.313	0.313	0.156	0.031
$\theta = 0.6$	0.010	0.077	0.230	0.346	0.259	0.078

   You want to run a significance test on the value of θ. You have the following:
   $H_0 : \theta \leq 0.5$
   $H_1 : \theta > 0.5$
   Significance level: $\alpha = 0.1$.

   Hint: You can take the equality $\theta = 0.5$ as your null hypothesis in the following calculations.

   (a) Find the rejection region (for which outcomes of your experiment will you reject the null hypothesis)?.

   (b) Suppose you run an experiment and the data gives $x = 4$. Compute the p-value of this data.

   (c) Compute the probability of making a type II error if in reality $\theta = 0.6$?

3. Specimens of milk from a number of dairies in three different districts were analyzed, and the concentration of the radioactive isotope strontium-90 was measured in each specimen. Suppose that specimens were obtained from four dairies in the first district, from six dairies in the second district, and three dairies in the third district, and the results are as follows:

   District 1: 6.4, 5.8, 6.5, 7.7,
   District 2: 7.1, 9.9, 11.2, 10.5, 6.5, 8.8,
   District 3: 9.5, 9.0, 12.1.

   You run a one-way ANOVA and get the following outcome:

Source	SS	df	MS	F	$P(F > f)$
Groups	24.5908	?	?	?	0.0333
Error	25.24	?	?		
Total	49.8308	?			

   (a) What is the hypothesis you are testing?

   (b) Fill in the question marks.

   (c) Do you reject the null hypothesis? What is your conclusion?

4. The following regression output is for predicting annual murders per million from percentage living in poverty in a random sample of 20 metropolitan areas.

	Estimate	Std. Error	t value	$P(\|T\| > t)$
(Intercept)	-29.901	7.789	-3.839	0.001
poverty	2.559	0.390	6.562	0.000
				$R^2 = 70.52\%$

   (a) Write out the linear model.

   (b) Interpret the intercept.

   (c) Interpret the slope.

(d) Calculate the correlation coefficient.

(e) What are the hypotheses for evaluating whether poverty percentage is a significant predictor of murder rate?

(f) Is poverty percentage significant for predicting murder rate? Justify your answer.

5. We test 10 hypotheses and get the following p-values:

H_{01}	0.0011
H_{02}	0.031
H_{03}	0.017
H_{04}	0.32
H_{05}	0.11
H_{06}	0.90
H_{07}	0.07
H_{08}	0.006
H_{09}	0.004
H_{10}	0.0009

(a) Suppose that we wish to control the Type I error for each null hypothesis at level $\alpha = 0.05$. Which null hypotheses will we reject?

(b) Now suppose that we wish to control the FWER at level $\alpha = 0.05$. Which null hypotheses will we reject? Justify your answer.

(c) Now suppose that we wish to control the FDR at level $q = 0.05$. Which null hypotheses will we reject? Justify your answer.

6. Consider a model that predicts a newborn's weight using several predictors (gestation length, parity, age of mother, height of mother, weight of mother, smoking status of mother) using data from 1,236 newborns. The following table describes the outpout of a multiple linear regression model:

	Error	SE	t value	$P(\|T\| > t)$
(Intercept)	-80.41	14.35	-5.60	0.0000
gestation	0.44	0.03	15.26	0.0000
parity	-3.33	1.13	-2.95	0.0033
age	-0.01	0.09	-0.10	0.9170
height	1.15	0.21	5.63	0.0000
weight	0.05	0.03	1.99	?
smoke:yes	-8.40	0.95	-8.81	0.0000
				Adjusted R^2 : 0.2541

(a) Write the equation of the regression model that includes all of the variables.

(b) Interpret the slopes of age and smoking in this context.

(c) Is mother's weight a significant predictor for the weight of a newborn?

(d) The table below shows adjusted R-squared values for all models we evaluate in the first step of the backwards elimination process.

	Model	Adjusted R^2
1	No gestation	0.1031
2	No parity	0.2492
3	No age	0.2547
4	No height	0.2311
5	No weight	0.2536
6	No smoking status	0.2072

Which, if any, variable should be removed from the model first?