MATH 6630 Sample Questions
November 2024

Question 1: (20 marks in 2 parts)

Consider using the Exponential(θ) distribution to model the lifetime of statistically independent electrical components:

$$f(x;\theta) = \frac{1}{\theta}\exp\{-x/\theta\} \quad x, \theta > 0$$

Suppose N components are being observed for one year and n have lifetimes less than one year: $x_1, x_2, ..., x_n$.

The remaining $N - n > 0$ last longer than one year but, since the study stopped after one year, their exact lifetimes are unknown.

a) [10 marks] Derive the E-step of an EM algorithm to estimate θ.

b) [10 marks] Derive the M-step of the algorithm, and show how θ_{t+1} is a function of θ_t.

Question 2: (10 marks)

Consider a multiple regression model of the form

$$Y = X_1\beta_1 + X_2\beta_2 + \epsilon$$

where $\epsilon \sim N(0, \sigma^2 I)$ and X_1 and X_2 are matrices containing blocks of variables such that the matrix $[X_1 X_2]$ is of full column rank.

Prove the Frisch-Waugh-Lovell theorem: The least-squares coefficients for the regression of the (residuals of Y regressed on X_2) on the (residuals of X_1 regressed on X_2) are the same as the least-squares coefficients corresponding to X_1 in the multiple regression of Y on both $[X_1 X_2]$.

Question 3: (10 marks in 2 parts)

In a multiple regression of Y on two variables X_1 and X_2, you find that the p-value for $\hat{\beta}_2$ is 0.45 and you are considering dropping X_2 from the model.

Comment on the following two statements.

a) [5 marks] "Since X_2 is not significant, dropping it should not have much impact on inference for X_1 so, if our main interest concerns X_1, there's no problem dropping X_2." Discuss, referring to the geometry of confidence regions, if appropriate.

b) [5 marks] Are there situations where it would be important to keep X_2 in the model although its associated p-value does not achieve significance? Discuss, referring to the geometry of confidence regions and/or principles of causality, if appropriate.

Question 4: (10 marks)

Suppose you have observed a vector $\mathbf{x}$ of n observations that you wish to model with each component having an identical and independent exponential distribution with mean θ, that is:

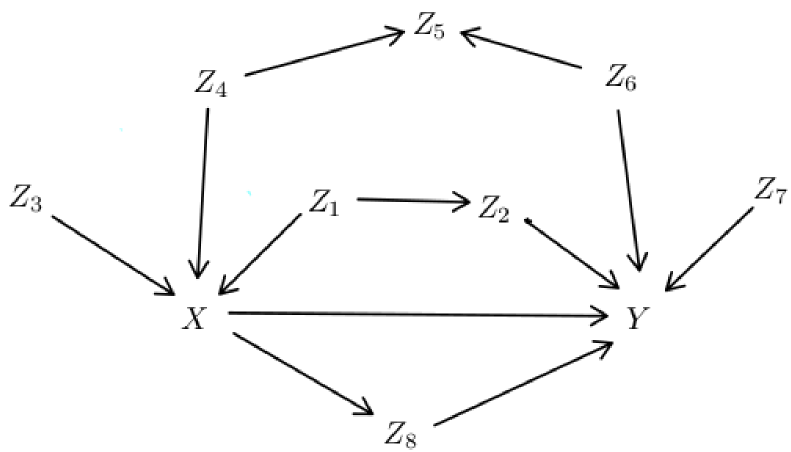
$$f(x_i;\theta) = \frac{1}{\theta}\exp -x/\theta \quad x_i, \theta > 0$$

.

Write a function in R that takes the vector of observations, $\mathbf{x}$, as input and plots the log-likelihood over a range of values for θ.

You may determine the range of values for θ from $\mathbf{x}$, or, for simplicity, use a range of values from 0.01 to 10.

Question 5: (20 marks in 2 parts)



Consider the linear causal DAG above and the following models:

1. `Y ~ X`
2. `Y ~ X + Z1 + Z5`
3. `Y ~ X + Z1`
4. `Y ~ X + Z8`
5. `Y ~ X + Z2`
6. `Y ~ X + Z2 + Z6`
7. `Y ~ X + Z2 + Z8`
8. `Y ~ X + Z2 + Z7`
9. `Y ~ X + Z1 + Z5 + Z6`

a) [10 marks] For each of these models discuss briefly **whether and why** fitting the model would produce, or not, an unbiased estimate of the causal effect of `X`.
b) [10 marks] Choose any two models that produce unbiased estimates of the causal effect of `X` and discuss in detail whether one can be determined to have a lower standard error for its estimate of β_X and why.

Question 6: (30 marks in 3 parts)

Consider a model given by a family of densities $f(x, \theta), x, \theta \in \mathbb{R}$ with respect to Lebesgue measure on $\mathbb{R}$ for a random variable that is parameterized by $\theta \in \mathbb{R}$. Assuming a fixed support and making needed assumptions concerning the appropriateness of interchanging the order of differentiation and integration, and the existence of necessary integrals, show that:

a) [10 marks] The expected value of the score function evaluated at the true value of θ is 0.

b) [10 marks] The variance of the score is equal to the Fisher information.

c) [10 marks] Sketch a proof for the asymptotic distribution of the MLE.

Question 7: (20 marks)

Write an essay on variable selection strategies in statistical analyses. What are the major relevant factors? How do they influence the choice of strategy?

Question 8: (10 marks)

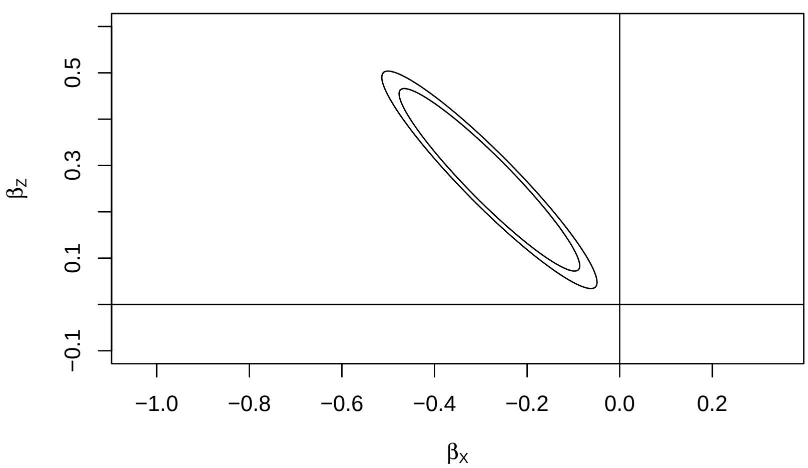
Explain clearly, with an appropriate sketch, the relationship between the 'within effect', the 'between effect' and the 'contextual effect' when working with clustered data.

Question 9: (20 marks)

Consider the following confidence ellipses for a linear model regressing Y on X and Z. Consider three possible models for a least-squares regression of Y on X and Z:

1. $E(Y) = \beta_0 + \beta_X X + \beta_Z Z$
2. $E(Y) = \gamma_{02} + \gamma_X X$
3. $E(Y) = \gamma_{03} + \gamma_Z Z$

The following are confidence ellipses for model 1. The outer ellipse is a joint 95% confidence ellipse for the vector (β_X, β_Z) and the inner ellipse is scaled so that its orthogonal projections onto the axes produces 95% confidence intervals.



Can you determine the outcome of the following tests? If so what would be the outcome of 5% tests? Discuss briefly why. (The alternative in each case is the negation of H_0). Show the basis of your reasoning using a diagram or other explanation.

a) $H_0 : \beta_X = \beta_Z = 0$

b) $H_0 : \beta_X = 0$

c) $H_0 : \beta_Z = 0$

d) $H_0 : \gamma_X = 0$

e) $H_0 : \gamma_Z = 0$

f) $H_0 : \beta_X = \beta_Z$

g) $H_0 : \beta_X + \beta_Z$ =0

Question 10: (10 marks)

State and explain the principle of marginality. Discuss how it is an example of a principle of invariance.

Question 11: (10 marks in 2 parts)

Suppose a test for mononucleosis (a disease) has a specificity and a sensitivity of 95%.

a. [5 marks] Does this mean that the test will be wrong 5% of the time? Prove or disprove.
b. [5 marks] If you take the test and the result is positive, does this mean that the probability that you have mononucleosis is 95%. Prove or disprove.

Question 12: (10 marks)

Explain clearly with a suitable sketch the relationship between the 'within effect', the 'between effect' and the 'contextual effect' when working with clustered data.

Question 13: (10 marks)

Suppose you are investigating the relationships between a variable Y and two possible predictors X and Z. Is it feasible for an observation to have relatively low leverage in each of the regressions on X and on Z, but to have high leverage in the multiple regression of Y on both X and Z? Using what you know about leverage and influence discuss either why this is not feasible, or, if it is feasible, under what conditions would it be expected to happen. If it can happen, how does this possibility inform your practice of regression?

Question 14: (10 marks)

Consider the following (now very familiar) model regressing income (in 1,000s of dollars) on years of education in three types of occupations: bc: blue collar, wc: white collar, and prof: professional.

The coefficients have been rounded for ease of calculation.

```
library(car)
```

```
    Loading required package: carData
```

```
head(Prestige)
```

```
                        education income women prestige census type
    gov.administrators      13.11  12351 11.16     68.8   1113 prof
    general.managers        12.26  25879  4.02     69.1   1130 prof
    accountants             12.77   9271 15.70     63.4   1171 prof
    purchasing.officers     11.42   8865  9.11     56.8   1175 prof
    chemists                14.62   8403 11.68     73.5   2111 prof
    physicists              15.64  11030  5.13     77.6   2113 prof
```

```
d <- na.omit(Prestige)
d$type <- factor(d$type)
d$inc <- d$income/1000 # income in 1,000s of dollars
table(d$type)
```

```
      bc prof   wc
      44   31   23
```

```
fit <- lm(inc ~ type * education + I(education^2), d)
out <- summary(fit)
```

```r
out$coefficients <- round(out$coefficients)   # to make things easier
out$coefficients
```

```
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)              36         16       2        0
typeprof                 58         26       2        0
typewc                   28         13       2        0
education                -8          4      -2        0
I(education^2)            1          0       2        0
typeprof:education       -5          2      -2        0
typewc:education         -3          1      -2        0
```

The three types of occupations are 'blue collar' (bc), 'white collar' (wc), and professional (prof).

You are a statistical consultant discussing this analysis with a client who tells you that your results don't make sense.

The negative coefficient for education says that predicted income is lower as education increases and the negative coefficient for 'typeprof:education' says that the change in income associated with additional education is lower for professional occupations than it is for blue collar occupations.

Clearly explain the interpretation of this output for your client. Take into account that the average years of education required for professional occupations is greater than for 'white collar' and 'blue collar' occupations. (Continue your answer on the back of this page.)

Question 15: (10 marks)

Discuss the following statement: "To choose variables in a multiple regression, you can start by testing one variable at a time and only add the variables that are significant."

Question 16: (10 marks)

Consider a model regressing `Y` (e.g. math achievement) on `X` (e.g. SES) in J schools identified by a categorical variable `G`. Let `Xg` be a 'contextual variable' that is the mean of `X` within each school and let `Xd` be the 'centered-within-groups' version of `X`, i.e. `Xd = X - Xg`.

Consider the following two models:

1) $E(\texttt{Y}) = \beta_0 + \beta_1 \texttt{X} + \beta_2 \texttt{Xg}$
2) $E(\texttt{Y}) = \psi_0 + \psi_1 \texttt{Xd} + \psi_2 \texttt{Xg}$

Show that these models are equivalent.

Question 17: (10 marks)

With reference to the previous question, derive the relationship between the parameters β_1, β_2 and the parameters ψ_1, ψ_2.

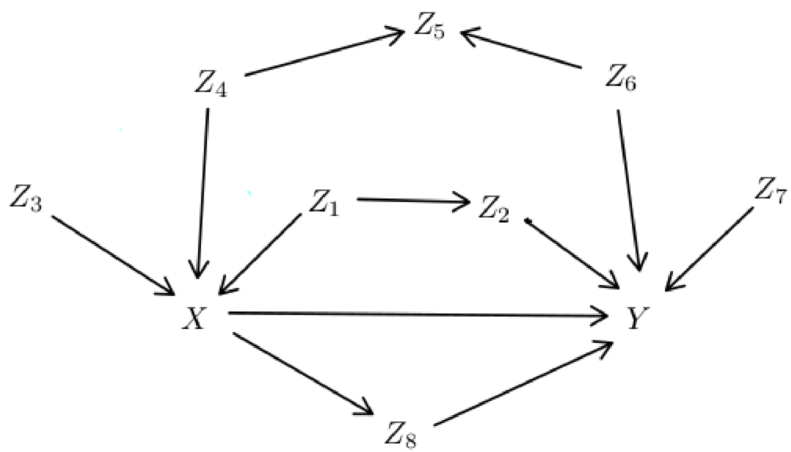
Question 18: (10 marks)

With reference to the previous question, discuss what considerations would lead you to choose to fit one model versus the other?

Question 19: (10 marks)

Discuss the taxonomy of outliers in regression presented in class. How do you identify the three archetypal kinds of outliers? What are the consequences of including a point of each type when the point is not, in fact, an observation from the target population?

Question 20: (30 marks in 3 parts)



Consider the linear causal DAG above and the following models: 1. `Y ~ X` 2. `Y ~ X + Z1` 3. `Y ~ X + Z2` 3. `Y ~ X + Z2 + Z3` 4. `Y ~ X + Z2 + Z6` 5. `Y ~ X + Z2 + Z4` 6. `Y ~ X + Z2 + Z4 + Z6` 7. `Y ~ X + Z2 + Z5` 8. `Y ~ X + Z2 + Z5 + Z6` 9. `Y ~ X + Z3 + Z8 + Z6` 10. `Y ~ X + Z2 + Z7 + Z8`

a) [10 marks] For each of these models discuss briefly **whether and why** fitting the model would produce, or not, an unbiased estimate of the causal effect of `X`.

b) [10 marks] List the the models that provide an unbiased estimate of the causal effect of `X` in the same order as they appear in the list above. Compare each model with the next model in the list, explaining, for each pair, whether you can determine which model provides the smaller expected standard error of $\hat{\beta}_X$, and, if so, which model does so and why. c) [10] In practice, how you would choose a model to estimate the causal effect of X in the example above, discussing which factors you would take into consideration.

Question 21: (10 marks)

1. [10] Consider a multiple regression of the form $Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$ where $\varepsilon \sim N(0, \sigma^2 I)$ and X_1, X_2 are matrices of observed values on groups of variables such that the matrix $[X_1 X_2]$ is of full column rank. The first column of the X_1 matrix is a column of 1's.
   Prove the following parts of the Frisch-Waugh-Lovell theorem:

   1. The 'added-variable-plot regression' consisting of the regression of the residuals of the regression of Y on X_1 regressed on the residuals of the regression of X_2 on X_1 has the same vector of least-squares coefficients as the least-squares coefficients for the predictors in X_2 in the multiple regression of Y on X_1 and X_2, and that

   2. the residuals of the added-variable-plot regression are the same as those of the multiple regression.

Question 22: (10 marks)

2. [10] Consider a normal linear regression of a response Y on two predictors X_1 and X_2. Discuss whether it is possible for the regression of Y on each of X_1 and X_2 individually to fail to reach significance (at the 0.05 level, for example) although the multiple regression of Y on both X_1 and X_2 achieves significance for both parameters. Justify your answer with appropriate diagrams in 'predictor data space' and/or 'beta' space. If this is possible provide a plausible example and discuss the implications for forward and backward stepwise regression.

Question 23: (20 marks in 2 parts)

The following questions are based on the 'Vocab' data set used in an assignment in the course. vocabulary scores obtained in samples of U.S. residents during the years 1974 to 2016, categorized by binary gender (Male and Female) and education (in years).

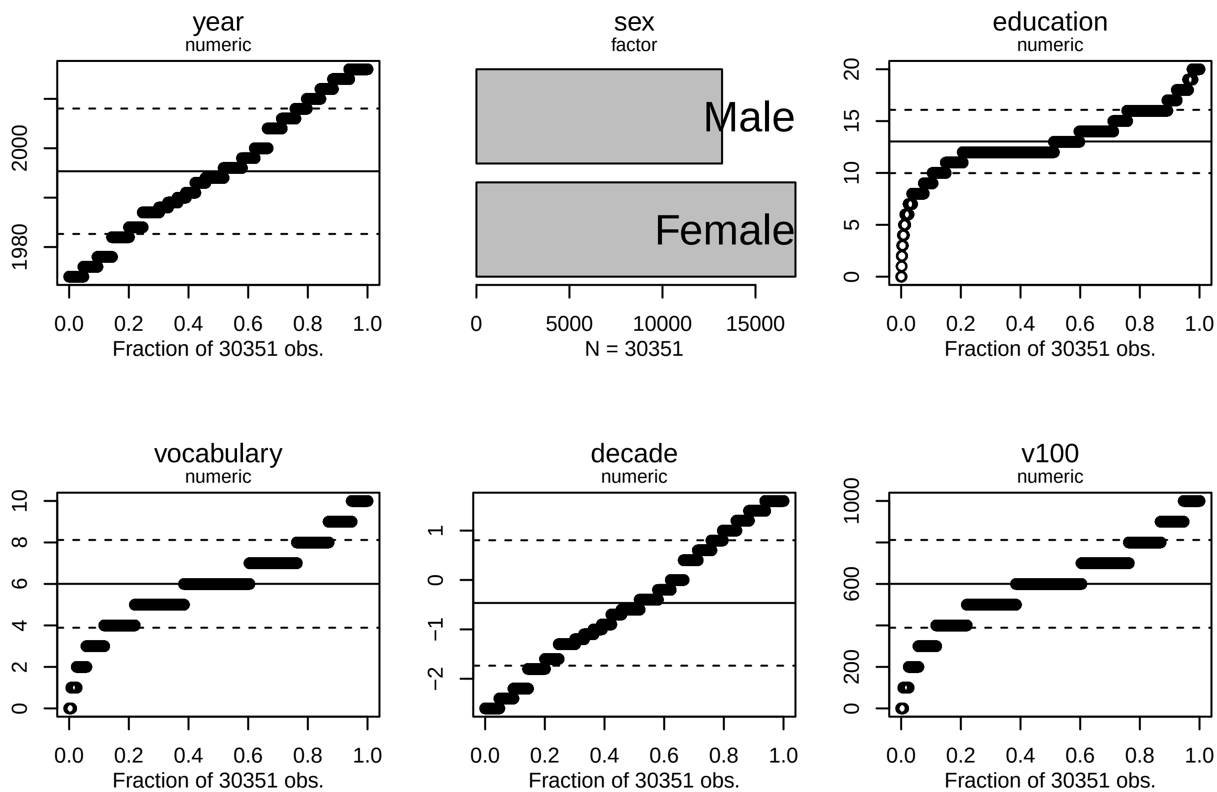
Recall that it records a vocabulary score for over 30,000 subjects tested over the years between 1974 and 2016. The questions below refer to the following output in R.

To make the coefficients easier to manipulate, 'year' has been changed to 'decade' relative to the year 2000 and the vocabulary rating has been multiplied by 100. The estimated coefficients are rounded to one decimal place.

```
library(car)
Vocab <- within(
  Vocab,
  {
    v100 <- vocabulary * 100
    decade <- (year - 2000)/10
  }
)
summary(Vocab)
```

```
      year           sex          education       vocabulary
 Min.   :1974   Female:17148   Min.   : 0.00   Min.   : 0.000
 1st Qu.:1987   Male  :13203   1st Qu.:12.00   1st Qu.: 5.000
 Median :1994                  Median :12.00   Median : 6.000
 Mean   :1995                  Mean   :13.03   Mean   : 6.004
 3rd Qu.:2006                  3rd Qu.:15.00   3rd Qu.: 7.000
 Max.   :2016                  Max.   :20.00   Max.   :10.000
     decade             v100
 Min.   :-2.6000   Min.   :   0.0
 1st Qu.:-1.3000   1st Qu.: 500.0
 Median :-0.6000   Median : 600.0
 Mean   :-0.4655   Mean   : 600.4
 3rd Qu.: 0.6000   3rd Qu.: 700.0
 Max.   : 1.6000   Max.   :1000.0
```

```
spida2::xqplot(Vocab)
```




```r
head(Vocab)
```

```
             year    sex education vocabulary decade v100
19740001 1974   Male        14          9   -2.6  900
19740002 1974   Male        16          9   -2.6  900
19740003 1974 Female        10          9   -2.6  900
19740004 1974 Female        10          5   -2.6  500
19740005 1974 Female        12          8   -2.6  800
19740006 1974   Male        16          8   -2.6  800
```

```r
fit <- lm(v100 ~ sex * education * decade, Vocab)
smry <- summary(fit)
smry$coefficients[,"Estimate"] <- round(smry$coefficients[,"Estimate"],1)
smry
```

```
Call:
lm(formula = v100 ~ sex * education * decade, data = Vocab)

Residuals:
    Min      1Q  Median      3Q     Max
-824.84 -112.05    8.04  125.51  858.66

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)              170.1000     7.1210  23.886  < 2e-16 ***
```

```
sexMale                    -21.0000    10.4758  -2.001 0.045350 *
education                   33.2000     0.5223  63.507  < 2e-16 ***
decade                      19.4000     4.9965   3.884 0.000103 ***
sexMale:education            0.4000     0.7623   0.544 0.586519
sexMale:decade             -11.6000     7.2237  -1.608 0.107738
education:decade            -2.7000     0.3799  -7.162 8.12e-13 ***
sexMale:education:decade     1.3000     0.5432   2.438 0.014766 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 184.6 on 30343 degrees of freedom
Multiple R-squared:  0.2384,   Adjusted R-squared:  0.2382
F-statistic:  1357 on 7 and 30343 DF,  p-value: < 2.2e-16
```

a) [10 marks] Using this model, what is the estimated 'gender gap' (Female - Male) in v100' in the year 2000 for individuals with 20 years of education?

b) [10 marks] Using this model is the gender gap in the year 1990 for individuals with 20 years of education getting narrower or getting wider? By how much per decade?

Question 24: (20 marks in 2 parts)

With reference to the same data as that of the preceding question, but a different model, answer the following questions.

```r
library(car)
summary(Vocab)
```

```
      year           sex          education       vocabulary
 Min.   :1974   Female:17148   Min.   : 0.00   Min.   : 0.000
 1st Qu.:1987   Male  :13203   1st Qu.:12.00   1st Qu.: 5.000
 Median :1994                  Median :12.00   Median : 6.000
 Mean   :1995                  Mean   :13.03   Mean   : 6.004
 3rd Qu.:2006                  3rd Qu.:15.00   3rd Qu.: 7.000
 Max.   :2016                  Max.   :20.00   Max.   :10.000
     decade              v100
 Min.   :-2.6000   Min.   :   0.0
 1st Qu.:-1.3000   1st Qu.: 500.0
 Median :-0.6000   Median : 600.0
 Mean   :-0.4655   Mean   : 600.4
 3rd Qu.: 0.6000   3rd Qu.: 700.0
 Max.   : 1.6000   Max.   :1000.0
```

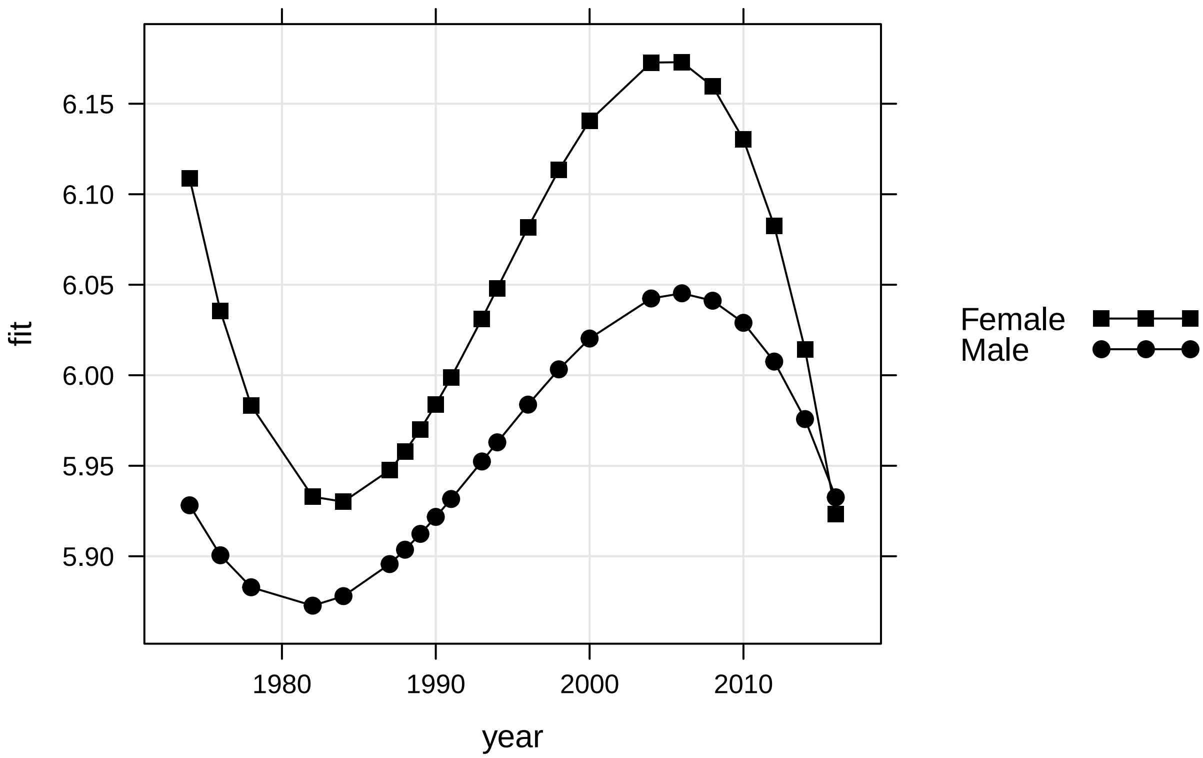
```r
data <- within(Vocab,
               {
                 Year <- year - 2000
                 Year2 <- Year^2
                 Year3 <- Year^3
               }
)
fit <- lm(vocabulary ~ (Year + Year2 + Year3) * sex, data = data)
summary(fit)
```

```
Call:
lm(formula = vocabulary ~ (Year + Year2 + Year3) * sex, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-6.1730 -1.0826  0.0162  1.0994  4.1273

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    6.141e+00  2.790e-02 220.123  < 2e-16 ***
Year           1.215e-02  2.793e-03   4.351 1.36e-05 ***
Year2         -8.358e-04  1.880e-04  -4.445 8.83e-06 ***
Year3         -4.831e-05  1.017e-05  -4.749 2.06e-06 ***
sexMale       -1.203e-01  4.226e-02  -2.847  0.00441 **
Year:sexMale  -4.446e-03  4.230e-03  -1.051  0.29331
Year2:sexMale  3.865e-04  2.842e-04   1.360  0.17376
Year3:sexMale  2.488e-05  1.533e-05   1.623  0.10461
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.114 on 30343 degrees of freedom
Multiple R-squared:  0.001568, Adjusted R-squared:  0.001337
F-statistic: 6.806 on 7 and 30343 DF,  p-value: 4.252e-08
```

```
pred <- within(
    with(data, expand.grid(Year = unique(Year), sex = unique(sex))),
    {
      Year2 <- Year ^ 2
      Year3 <- Year ^ 3
      year <- Year + 2000
    }
  )

pred$fit <- predict(fit, newdata = pred)
library(latticeExtra)
```

```
Loading required package: lattice
```

```
trellis.par.set(
  superpose.symbol = list(pch = c(15, 19), col = 'black', lty = 1, cex =1.0),
  superpose.line = list(col = 'black', lty = 1)
  )
xyplot(fit ~ year, pred, groups = sex, type = 'b',
       auto.key = T) +
  layer_(panel.grid(h = -1, v = -1))
```



a) [10 marks] Using the model `fit`, estimate the rate and direction in which the gender gap is changing in the year 2010. Is it changing faster in 2010 than it did in 1990? Describe how you would you go about obtaining a confidence interval for the difference in the rate of change between 2010 and 1990.

b) [10 marks] Describe the precise interpretations in terms of the behaviour of the response function as a function of year and sex, of the estimated coefficients for `sexMale`, for `Year:sexMale`, and for `Year2` in the summary output for `fit`.

Question 25: (5 marks)

Write a function in R that takes a matrix as input and returns the index of the column that has the largest sum of squares.

Question 26: (10 marks)

Describe how you would simulate an observation from a distribution with density $f(x) = 2\exp\{-2x\}, x > 0$ using the inverse CDF. Write code in R or in pseudocode for a **function** that would generate n observations from this distribution using `runif` as a uniform random number generator.

Question 27: (10 marks)

When is it important to include a term in a regression model although its coefficient is not statistically significant? Discuss at least one situation in a context in which the goal of the analysis is causal inference using observational data and in a context in which the goal is predictive inference using observational data.

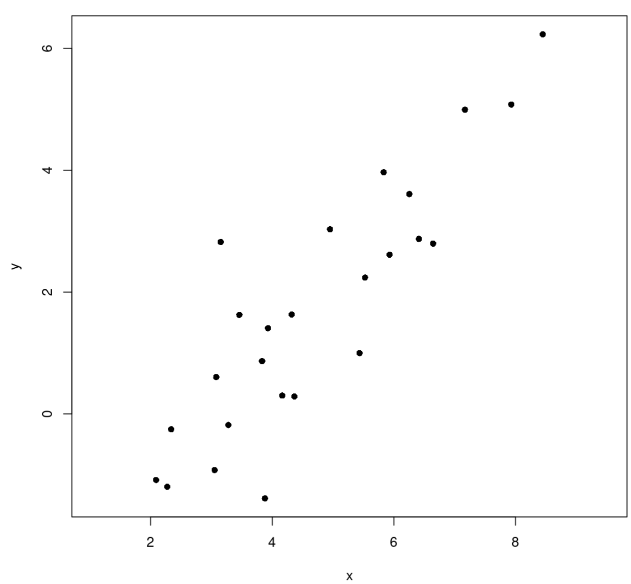
Question 28: (10 marks)

Describe the Metropolis-Hastings algorithm. Prove that if the 'target function', $h(x)$ is equal to a density, $f(x)$, multiplied by a positive constant that might be unknown, i.e. $h(x) = cf(x)$, then the distribution with density $f(x)$ is a stationary distribution for the MCMC step of the Metropolis-Hastings algorithm, i.e. if X_t has a distribution with density f, then so does X_{t+1}.

Question 29: (10 marks)

Many researchers who find that a hypothesis test of a particular null hypothesis has achieved a p-value of 0.04 have the impression that there is strong evidence against the null hypothesis and it is 'very unlikely to be correct.' Discuss whether this is a correct interpretation of the p-value.

Question 30: (10 marks in 2 parts)



Consider the scatterplot above displaying the values for 25 observations from the variables 'x' and 'y'.

a. [5] Construct an approximate 95% confidence interval for the least-squares estimate of the slope of a linear relationship regressing y on x.
b. [5] What can you say about the p-value for a test of the hypothesis that the population slope is zero against the alternative that it is not 0.

Question 31: (20 marks in 4 parts)

Answer the following questions.

a) [5 marks] Briefly discuss the following statement. Is the statement true, false, partly true and partly false? If its validity depend on other conditions, how does it do so?

   "Since the predicted Z-score for the intelligence of the child of an unusually intelligent parent is less than that of the parent, and the predicted z-score for the child of a parent of below average intelligence is greater than that of the parent, and, furthermore, since this is true for each subsequent generation, it must be true that, asymptotically, the predicted intelligence of a distant descendant must converge to the average of the population, and the population as a whole will therefore become concentrated at the average."

b) [5 marks] You are analyzing observational data on the relationship between health and daily coffee consumption to investigate whether there is evidence that coffee consumption is harmful to health.

Suppose you want to control for a possible confounding factor 'stress'. In this kind of study discuss whether it is more important to make sure that you measure coffee consumption accurately than it is to make sure that you measure the confounding factor 'stress' accurately? What are the consequences of measurement error in coffee consumption? What are the consequences of measurement error in stress? Which consequences are more consequential?

c) [5 marks] A random survey of Canadian families yielded average 'equity' (i.e. total owned in real estate, bonds, stocks, etc. minus total owed) of \$48,000. Aggregate government data of the total equity held by individuals in the Canadian population shows that this figure must be much larger, in fact more than three times as large. You are testifying before a government committee and asked to explain likely reasons why there would be such a discrepancy.
Some committee members believe that the discrepancy proves that respondents lie about their equity, others believe that the survey could not have been truly random. As a statistician, using the common understanding that the distribution of personal equity is highly skewed, can you explain why the discrepancy is expected to occur even in the absence of the two explanations profferred by committee members. Hence the discrepancy is not suggestive of either conclusion?

d) [5 marks] In a multiple regression, if you drop a predictor whose effect is not significant (in the multiple regression) is it true that the p-values of the other predictors should not change very much. If this statement is not always true describe the circumstances under which you expect it to be true, or not to be true.

Question 32: (10 marks)

Suppose you have access to a uniform random number generator. Describe how you could use the Metropolis-Hastings algorithm to generate a Markov chain with a Normal distribution with mean 0 and unit standard deviation as a stationary distribution. Write R code for a function that generates the next observation from this chain.

Question 33: (20 marks in 2 parts)

a) [10 marks] Let h be a function on $\mathbb{R}$ that is proportional to a probability density function f, i.e. $h(x) = cf(x)$ with $c > 0$ unknown. Let g be a probability density on $\mathbb{R}$ that is symmetric around 0, from which you can generate random observations. Describe how you would use random observations with pdf g to generate a Markov chain with f as a stationary distribution.

b) [10 marks] Prove that the process in the previous question has the distribution given by f as stationary distribution.

Question 34: (20 marks in 2 parts)

Consider estimating the average lifetime of some electronic component, a 'gizmo', whose lifetime has an exponential(ψ) distribution:

$$f(x|\psi) = \psi e^{-x\psi} \quad x > 0, \psi > 0$$

Suppose N gizmos are observed for a year and n $(1 < n < N)$ burn out during the year, with lifetimes recorded as: $x_1, x_2, ..., x_n$ where x is measured in years (fractional, of course). The remaining $N - n$ are still functioning at the end of the year.

a) [10 marks] Set up the E step of the EM algorithm to estimate ψ. Recall that exponential lifetimes are 'memoryless', i.e. the distribution of the *additional life* of a gizmo that survives one year is exponential($1/\psi$).
b) [10 marks] Derive the M step, i.e. how do you obtain ψ_{t+1} as a function of ψ_t?

Question 35: (10 marks)

3. Prove that each step in the EM algorithm cannot reduce the log-likelihood. Assume that the densities in the model, $f(x|\theta)$ are positive with the same support.

Question 36: (10 marks)

Describe the Newton-Raphson method for finding the MLE of a statistical model. Give a sketch of a proof, using Taylor series expansions, that it converges to a maximum if the starting value is sufficiently close to a value that produces a maximum.

Question 37: (10 marks)

Describe the Fisher Scoring method for finding MLEs and discuss some of its advantages or disadvantages in comparison with the Newton-Raphson method.

Question 38: (10 marks)

Describe the Nelder-Mead algorithm in detail. (You'll forget the details after the exam but it's worth having to learn it at some point in your career.) Be ready to apply it for one or two steps in a simple example.

Question 39: (10 marks)

An insurance company has N clients. It assumes that the number of claims filed by client $k, k = 1, ..., N$, in a year has a Poisson(θ_k) distribution and that the number of claims filed by each client is statistically independent of the number of claims filed by other clients.

Let $Y_1, Y_2, ..., Y_k, ..., Y_N$ represent the number of claims filed by the N clients in one year, and let $g(\theta)$ represent the density for the distribution of the values of θ among clients.

Under these conditions, is it reasonable to assume that the distribution of the number of claims, $Y_1, Y_2, ..., Y_k, ..., Y_N$, is Poisson? Discuss.

files a number of claims each