Week 1 Exercises

I created some notes that review the concepts required for these exercises, but you should draw on familiar textbooks and notes where possible.

And this is an ambitious set of exercises. Make a serious effort to freshen your mathematical toolkit. Your effort will make for a smoother semester. If you find yourself struggling to finish all the problems, then focus on 4(f), 4(g), 5, 7, 8, 9, 10, 13(1), 14, 15, 20, 21, 22, 24(c), 24(g), 25(b), 25(e), 26(c), and 27. Use the other exercises as needed to brush up on weaknesses.

1 Fractions, Logarithms, and Exponents

Exercise 1 Some Practice with Fractions

Simplify each of the following:

a. $\frac{x}{y} + \frac{z}{y}$

b. $\frac{m}{n} - \frac{r}{n}$

c. $\frac{p+q}{r} + \frac{s}{r}$

d. $\frac{2a}{b} + \frac{3a}{b}$

e. $\frac{x}{2} + \frac{y}{3}$

f. $\frac{1}{a} - \frac{1}{b}$ (write as a single fraction)

Solutions

a. $\frac{x}{y} + \frac{z}{y} = \frac{x+z}{y}$

b. $\dfrac{m}{n} - \dfrac{r}{n} = \dfrac{m-r}{n}$

c. $\dfrac{p+q}{r} + \dfrac{s}{r} = \dfrac{p+q+s}{r}$

d. $\dfrac{2a}{b} + \dfrac{3a}{b} = \dfrac{2a+3a}{b} = \dfrac{5a}{b}$

e. $\dfrac{x}{2} + \dfrac{y}{3} = \dfrac{3x+2y}{6}$

f. $\dfrac{1}{a} - \dfrac{1}{b} = \dfrac{b-a}{ab}$

Exercise 2 Some Practice with Logarithms

You can assume arguments to logs are always positive, for the sake of these exercises.

Simplify each of the following:

a. $\log_{10}(10^5)$

b. $\log(e^3)$ (note: $\log(x)$ refers to the natural log in our class)

c. $\log(3x) + \log(4)$

d. $\log(x^3y^2)$

e. $\log\left(\dfrac{x^2}{y^3}\right)$

f. $\log\left(\prod_{i=1}^{n} x_i^2\right)$

Solutions

a. $\log_{10}(10^5) = 5$

b. $\log(e^3) = 3$

c. $\log(3x \cdot 4) = \log(12x)$

d. $3\log(x) + 2\log(y)$

e. $2\log(x) - 3\log(y)$

f. $\sum_{i=1}^{n} 2\log(x_i)$

Exercise 3 Some Practice with Exponents

Simplify each of the following:

a. $e^2 \cdot e^5$

b. $\left(e^3\right)^2$

c. $\dfrac{e^x}{e^4}$

d. $\left(\dfrac{1}{e}\right)^x$

e. $10^x \cdot 10^{-x}$

f. $\left(ab^2\right)^3$

Solutions

a. e^7

b. e^6

c. e^{x-4}

d. e^{-x}

e. 1

f. a^3b^6

Exercise 4 Some Practice Combining Fractions, Logarithms, and Exponents

Simplify each of the following:

a. $\dfrac{x^2+x}{x}$

b. $\dfrac{1}{2}\log(x^2)$

c. $\frac{e^{x+1}}{e}$

d. $\log\left(\dfrac{e^2}{x}\right)$

e. $\dfrac{\log(x) + \log(y)}{2}$

f. $\log\left(\dfrac{a^3 b^2}{c}\right)$

g. $\log\left(\prod_{i=1}^{n} \pi^{y_i} (1 - \pi)^{(1-y_i)}\right)$

Solutions

a. $x + 1$

b. $\log(x)$

c. e^x

d. $2 - \log(x)$

e. $\dfrac{\log(xy)}{2}$

f. $3\log(a) + 2\log(b) - \log(c)$

g. $\left(\sum_{i=1}^{n} y_i\right) \log(\pi) \ + \ \left(n - \sum_{i=1}^{n} y_i\right) \log(1 - \pi)$

Exercise 5 Inverse Logit

Show that the two common forms of the inverse-logit function

$$f_1(x) \ = \ \frac{1}{1 + e^{-x}} \quad \text{and} \quad f_2(x) \ = \ \frac{e^x}{1 + e^x}$$

are algebraically equivalent for every real number x.

Solution

Start with $f_1(x) = \frac{1}{1+e^{-x}}$. Multiply the numerator and denominator by the strictly positive quantity e^x (which leaves the fraction's value unchanged). After a bit of algebra, we obtain f_2.

$$f_1(x) \ = \ \frac{1 \cdot e^x}{(1 + e^{-x}) \cdot e^x} \ = \ \frac{e^x}{e^x + 1} \ = \ \frac{e^x}{1 + e^x} \ = \ f_2(x).$$

Hence $f_1(x) = f_2(x)$, so the two expressions are equivalent for all x.

2 Derivatives

Exercise 6 Some Practice with Derivatives

Differentiate each of the following:

a. Let $f(x) = 5$

b. Let $f(x) = x^4$

c. Let $f(x) = 3x^2 - 4x + 7$

d. Let $f(x) = e^{2x}$

e. Let $g(t) = \log(t^3)$

f. Let $h(y) = y^3 \log(y)$

g. Let $f(x) = (x^2 + 1)^5$

h. Let $f(x) = \dfrac{e^x}{x}$

i. Let $q(u) = u^2 e^{u^3}$

j. Let $r(z) = \dfrac{\log(z^2 + 1)}{e^z}$

Solutions

a. 0. Constant rule.

b. $4x^3$. Power rule (nx^{n-1}).

c. $6x - 4$. Term-by-term power rule.

d. $2e^{2x}$. Chain rule on e^{kx}.

e. $\dfrac{3}{t}$. Rewrite $\log(t^3) = 3\log(t)$, then logarithm rule.

f. $3y^2 \log(y) + y^2$. Product rule on $y^3 \cdot \log(y)$.

g. $10x\,(x^2 + 1)^4$. Chain rule on an outer fifth power.

h. $e^x \dfrac{x-1}{x^2}$. Rewrite $e^x x^{-1}$, then product rule.

i. $e^{u^3}(2u + 3u^4)$. Product rule; inner derivative $d/du(e^{u^3}) = e^{u^3} \cdot 3u^2$.

j. $\dfrac{\frac{2z}{z^2+1} - \log(z^2+1)}{e^z}$. Quotient rule with a chain inside $\log(z^2 + 1)$.

Exercise 7 An Important Preview

Let $\ell(\pi) = S \log(\pi) + (n - S) \log(1 - \pi)$ for $0 < \pi < 1$, where S and n (with $0 \le S \le n$) are fixed numerical constants. Find $\frac{d\ell(\pi)}{d\pi}$ and $\frac{d^2\ell(\pi)}{d\pi^2}$.

Solution

- **First derivative**

$$\frac{d\ell}{d\pi} = \frac{S}{\pi} - \frac{n-S}{1-\pi}.$$

(Quotient-inside-log derivatives: $d/d\pi[\log(\pi)] = 1/\pi$, $d/d\pi[\log(1-\pi)] = -1/(1-\pi)$.)

- **Second derivative**

$$\frac{d^2\ell}{d\pi^2} = -\frac{S}{\pi^2} - \frac{n-S}{(1-\pi)^2}.$$

Each term differentiates again via the power rule.
The result is negative for all $\pi \in (0, 1)$, so $\ell(\pi)$ is concave.

Exercise 8 Inverse Logit

Define $p(\theta) = \frac{1}{1+e^{-\theta}}$ for $\theta \in \mathbb{R}$. Find $p'(\theta)$ and $p''(\theta)$.

Solution

- **First derivative**

  Rewrite $p(\theta) = (1 + e^{-\theta})^{-1}$ and apply the chain rule:

$$p'(\theta) = -(1 + e^{-\theta})^{-2} \cdot (-e^{-\theta}) = \frac{e^{-\theta}}{(1 + e^{-\theta})^2} = p(\theta)[1 - p(\theta)].$$

- **Second derivative**

  Differentiate $p'(\theta) = p(1 - p)$ using the product rule:

$$p''(\theta) = p'(1 - p) - p\,p' = p(\theta)[1 - p(\theta)][1 - 2p(\theta)].$$

  The factor $1 - 2p(\theta)$ shows the curve's slope is steepest (and p'' is zero) when $p(\theta) = 0.5$.

Exercise 9 Brambor, Clark, and Golder (2006)

Let $E[y \mid x_1, x_2] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$, where $\beta_0, \dots, \beta_3$ are constants. Find the marginal effect of x_1 on y. That is, compute $\dfrac{\partial E[y \mid x_1, x_2]}{\partial x_1}$.

Solution

Differentiate with respect to x_1 while treating x_2 as constant, so that

$$\frac{\partial E[y \mid x_1, x_2]}{\partial x_1} = \beta_1 + \beta_3 x_2.$$

The marginal effect of x_1 depends on x_2; when $x_2 = 0$, the effect is β_1, and it changes by β_3 units for every unit increase in x_2.

Exercise 10 Brambor, Clark, and Golder (2006) for Cubic Polynomial

Let $E[y \mid x] = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$, where $\beta_0, \dots, \beta_3$ are constants. Find the marginal effect of x on y. That is, compute $\dfrac{\partial E[y \mid x]}{\partial x}$.

Solution

Apply the power rule term-by-term so that

$$\frac{\partial E[y \mid x]}{\partial x} = \beta_1 + 2\beta_2 x + 3\beta_3 x^2.$$

The marginal effect varies with x in a quadratic fashion; the sign and magnitude are governed by the coefficients on the polynomial terms.

3 Integrals

Exercise 11 Some Practice Problems for Integrals

Compute each integral.

a. $\int 7\,dx$

b. $\int_0^2 x^6\,dx$

c. $\int (3x^2 - 4)\,dx$

d. $\int_0^1 e^{3x}\,dx$

e. $\int \frac{1}{x}\,dx$

f. $\int 4x\,e^{x^2}\,dx$ *Hint:* let $u = x^2$.

g. $\int x\,e^x\,dx$ *Hint:* integration by parts with $u = x$.

h. $\int_0^1 x^2\,dx$

Solutions

a. $7x + C$

b. $\dfrac{128}{7}$ (power rule $\frac{x^7}{7}$ evaluated from 0 to 2).

c. $x^3 - 4x + C$

d. $\dfrac{e^3 - 1}{3}$ (antiderivative $e^{3x}/3$ then plug in 1 and 0).

e. $\log|x| + C$

f. $2e^{x^2} + C$ (substitute $u = x^2 \Rightarrow 4x\,dx = 2\,du$).

g. $e^x(x-1)+C$ (integration by parts: $u=x,\ dv=e^x dx$).

h. $\frac{1}{3}$ (antiderivative $x^3/3$ evaluated from 0 to 1).

4 Matrices

Exercise 12 Some Practice with Transposes

Find the transpose of each of the following matrices:

a. $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$

b. $B = \begin{bmatrix} 0 & -1 & 2 \\ 5 & 3 & 1 \end{bmatrix}$

c. $C = \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}$

Solutions

a. $A' = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$

b. $B' = \begin{bmatrix} 0 & 5 \\ -1 & 3 \\ 2 & 1 \end{bmatrix}$

c. $C' = \begin{bmatrix} 7 & 8 & 9 \end{bmatrix}$

Exercise 13 Some Practice with Matrix Multiplication

Problem 1

Multiply A and B using paper-and-pencil. Check your work with R.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \qquad B = \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix}.$$

Problem 2

Let

$$A = \begin{bmatrix} 2 & 1 & 0 \\ -1 & 3 & 4 \end{bmatrix}.$$

Compute $A'A$ (i.e., $A^{\top}A$) using paper-and-pencil. Check your work with R.

Problem 1 Solution

Check conformability: A is 2×3 and B is 3×2, so AB is 2×2.

$$AB = \begin{bmatrix} (1)(7) + (2)(9) + (3)(11) & (1)(8) + (2)(10) + (3)(12) \\ (4)(7) + (5)(9) + (6)(11) & (4)(8) + (5)(10) + (6)(12) \end{bmatrix} = \begin{bmatrix} 58 & 64 \\ 139 & 154 \end{bmatrix}.$$

```r
A <- matrix(c(1, 2, 3,
              4, 5, 6),
            nrow = 2, byrow = TRUE)

B <- matrix(c( 7,  8,
               9, 10,
              11, 12),
            nrow = 3, byrow = TRUE)

A %*% B   # confirm work
```

```
     [,1] [,2]
[1,]   58   64
[2,]  139  154
```

Problem 2 Solution

1. **Conformability**
   A is 2×3, so $A'A$ will be 3×3.
2. **Form** A'

$$A' = \begin{bmatrix} 2 & -1 \\ 1 & 3 \\ 0 & 4 \end{bmatrix}.$$

3. **Multiply entry-by-entry**

$$A'A = \begin{bmatrix} (2)(2)+(-1)(-1) & (2)(1)+(-1)(3) & (2)(0)+(-1)(4) \\ (1)(2)+(3)(-1) & (1)(1)+(3)(3) & (1)(0)+(3)(4) \\ (0)(2)+(4)(-1) & (0)(1)+(4)(3) & (0)(0)+(4)(4) \end{bmatrix} = \begin{bmatrix} 5 & -1 & -4 \\ -1 & 10 & 12 \\ -4 & 12 & 16 \end{bmatrix}.$$

```r
A <- matrix(c( 2, 1, 0,
              -1, 3, 4),
            nrow = 2, byrow = TRUE)

t(A) %*% A   # confirm work
```

```
     [,1] [,2] [,3]
[1,]    5   -1   -4
[2,]   -1   10   12
[3,]   -4   12   16
```

Exercise 14 Matrices, OLS, and R

For the linear model $y = X\beta + \varepsilon$, where X is an $n \times k$ full-rank design matrix (including a column of ones), we define

- OLS estimator: $\hat{\beta} = (X'X)^{-1}X'y$.
- Residual vector: $e = y - X\hat{\beta}$.
- Predicted values: $\hat{y} = X\hat{\beta}$.
- Classical variance–covariance matrix: $\mathrm{Var}_{\mathrm{cl}}(\hat{\beta}) = \hat{\sigma}^2\,(X'X)^{-1}$, where $\hat{\sigma}^2 = \dfrac{e'e}{n-k}$.

1. Write an R function `ols(X, y)` that returns a list with the coefficient estimates `beta_hat`, the classical variance matrix `Var_cl`, the residuals `e`, and the predicted values `y_hat`.
2. Use the `penguins` dataset in the {palmerpenguins} package to test your function. The following starter code prepares X and y.

```r
# load packages
library(palmerpenguins)

# 1. drop NAs
penguins_complete <- na.omit(penguins)

# 2. make response vector (y)
y <- penguins_complete$body_mass_g # numeric vector (n × 1)
```

```r
## 3. make design matrix (X)
X_predictors <- as.matrix(penguins_complete[ ,
                   c("bill_length_mm",
                     "bill_depth_mm",
                     "flipper_length_mm")])

X <- cbind("(Intercept)" = 1, X_predictors)
```

Solution

The function.

```r
ols <- function(X, y) {
  ## 1. Components of (X'X) and its inverse
  X_transpose <- t(X)                    # X'
  XtX         <- X_transpose %*% X       # X'X
  XtX_inv     <- solve(XtX)              # (X'X)^{-1}

  ## 2. OLS estimator beta_hat
  Xty      <- X_transpose %*% y          # X'y
  beta_hat <- XtX_inv %*% Xty            # (X'X)^{-1} X'y

  ## 3. Fitted values (y_hat) and residuals (e)
  y_hat <- X %*% beta_hat
  e     <- y - y_hat

  ## 4. Classical variance-covariance matrix (Var_cl)
  n <- nrow(X)
  k <- ncol(X)

  RSS         <- t(e) %*% e                          # e'e
  sigma2_hat  <- as.numeric(RSS) / (n - k)           # \hat\sigma^2
  Var_cl      <- sigma2_hat * XtX_inv                # \hat\sigma^2 (X'X)^{-1}

  ## 5. Return list matching notation
  list(
    beta_hat = beta_hat,
    Var_cl   = Var_cl,
    e        = e,
    y_hat    = y_hat
  )
}
```

Demonstration with the `penguins` dataset.

```r
# load packages
library(palmerpenguins)

# 1. drop NAs
penguins_complete <- na.omit(penguins)

# 2. make response vector (y)
y <- penguins_complete$body_mass_g # numeric vector (n × 1)

## 3. make design matrix (X)
X_predictors <- as.matrix(penguins_complete[ ,
                   c("bill_length_mm",
                     "bill_depth_mm",
                     "flipper_length_mm")])

X <- cbind("(Intercept)" = 1, X_predictors)

# compute quantities
results <- ols(X, y)

# print  each component
results$beta_hat      # beta-hat
```

```
                          [,1]
(Intercept)       -6445.476043
bill_length_mm        3.292863
bill_depth_mm        17.836391
flipper_length_mm    50.762132
```

```r
results$Var_cl        # classical VCOV
```

```
                  (Intercept) bill_length_mm bill_depth_mm flipper_length_mm
(Intercept)       320503.2363     805.525086   -6528.69599      -1211.206658
bill_length_mm       805.5251      28.793240     -17.85211         -8.786474
bill_depth_mm      -6528.6960     -17.852112     191.15678         20.067377
flipper_length_mm  -1211.2067      -8.786474      20.06738          6.236320
```

```r
head(results$e)       # first few residuals
```

```
          [,1]
[1,]  545.23877
[2,]  363.29828
[3,] -656.89703
[4,] -366.70577
[5,]  -46.16813
[6,]  436.95010
```

```r
head(results$y_hat)   # first few fitted values
```

```
        [,1]
[1,] 3204.761
[2,] 3436.702
[3,] 3906.897
[4,] 3816.706
[5,] 3696.168
[6,] 3188.050
```

Exercise 15 Gradients and Hessians

Consider the function of two variables $f(x, y) = 3x^2y + 2e^{xy} - y^3$. Compute (1) the gradient vector $\nabla f(x, y)$ and (2) the Hessian matrix $H_f(x, y)$.

Gradient

The gradient collects the first-order partial derivatives.

$$\frac{\partial f}{\partial x} = 6xy + 2y\, e^{xy},$$
$$\frac{\partial f}{\partial y} = 3x^2 + 2x\, e^{xy} - 3y^2.$$

Hence

$$\nabla f(x, y) \;=\; \begin{bmatrix} 6xy + 2y\, e^{xy} \\ 3x^2 + 2x\, e^{xy} - 3y^2 \end{bmatrix}.$$

2. **Hessian**

The Hessian collects the second-order partial derivatives.

$$\begin{aligned}
\frac{\partial^2 f}{\partial x^2} &= 6y + 2y^2 e^{xy}, \\
\frac{\partial^2 f}{\partial x\,\partial y} &= 6x + 2e^{xy} + 2xy\,e^{xy}, \\
\frac{\partial^2 f}{\partial y^2} &= -6y + 2x^2 e^{xy}.
\end{aligned}$$

Notice that the mixed partials are equal ($\partial^2 f/\partial x\,\partial y = \partial^2 f/\partial y\,\partial x$).

The Hessian is

$$H_f(x, y) \;=\; \begin{bmatrix} 6y + 2y^2 e^{xy} & 6x + 2e^{xy} + 2xy\,e^{xy} \\ 6x + 2e^{xy} + 2xy\,e^{xy} & -6y + 2x^2 e^{xy} \end{bmatrix}.$$

5 Probability Theory

Exercise 16 Some Results of the Axioms and Definition of Probability

Prove the following results:

a. $\Pr(\emptyset) = 0$.
b. If event $A \subseteq B$, then $\Pr(A) \le \Pr(B)$.
c. For event A, $0 \le \Pr(A) \le 1$.
d. For any event A, $\Pr(A^c) = 1 - \Pr(A)$.

Hints

a. Use Axiom 3.
b. Notice that $B = A \cup (B \cap A^c)$. Then use the Additional Rule for Two Disjoint Events.
c. Axiom 1 establishes that $0 \le \Pr(A)$. Now show that $\Pr(A) \le 1$. To do this, the result from (b).
d. Notice that $\Pr(S) = \Pr(A) + \Pr(A^c)$ and follow this forward.

Solution

a. $\Pr(\emptyset) = \Pr\Big(\bigcup_{i=1}^{\infty} \emptyset\Big) = \sum_{i=1}^{\infty} \Pr(\emptyset)$. This equality can hold only if $\Pr(\emptyset) = 0$.
b. $B = A \cup (B \cap A^c)$. By the addition rule for disjoint events, $\Pr(B) = \Pr(A) + \Pr(B \cap A^c)$. By Axiom 1, $\Pr(B \cap A^c) \ge 0$, so $\Pr(A) \le \Pr(B)$.

c. By Axiom 1, $\Pr(A) \geq 0$. Since $A \subseteq S$ (the sample space), by monotonicity we have $\Pr(A) \leq \Pr(S) = 1$. Thus $0 \leq \Pr(A) \leq 1$.
d. $\Pr(S) = \Pr(A) + \Pr(A^c)$. Then $1 = \Pr(A) + \Pr(A^c)$, so $\Pr(A^c) = 1 - \Pr(A)$.

Exercise 17 Simplifying the Multiplication Rule

Simplify $\Pr(A \mid B)$ for the following scenarios.

a. $A \subset B$ and $\Pr(B) > 0$.
b. A and B are disjoint and $\Pr(B) > 0$.
c. B is the empty set (tricky!).
d. B is the sample space S.

Solution

We simplify $\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$ in each case:

a. If $A \subset B$ and $\Pr(B) > 0$, then $A \cap B = A$. So $\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\Pr(A)}{\Pr(B)}$.
b. If A and B are disjoint and $\Pr(B) > 0$, then $A \cap B = \emptyset$, and $\Pr(\emptyset) = 0$. So $\Pr(A \mid B) = \frac{0}{\Pr(B)} = 0$.
c. If $B = \emptyset$, then $\Pr(B) = 0$, and the expression $\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$ is undefined. Conditional probability is only defined when the probability of the conditioning event is positive.
d. If $B = S$ (the sample space), then $A \cap S = A$ and $\Pr(S) = 1$. So $\Pr(A \mid S) = \frac{\Pr(A)}{1} = \Pr(A)$.

Exercise 18

Suppose A and B are independent and $\Pr(B) < 1$. Find $\Pr(A^c|B^c)$ in terms of A and B. Prove that A^c and B^c are independent.

Solution

First, show that A^c and B^c are independent.

$$
\begin{aligned}
\Pr(A^c \cap B^c) &= \Pr([A \cup B]^c) \\
&= 1 - \Pr(A \cup B) \\
&= 1 - [\Pr(A) + \Pr(B) - \Pr(A \cap B)] \\
&= 1 - \Pr(A) - \Pr(B) + \Pr(A)\Pr(B) \\
&= [1 - \Pr(A)] \times [1 - \Pr(B)] \\
&= \Pr(A^c) \times \Pr(B^c)
\end{aligned}
$$

Then, by independence, we know that $\Pr(A^c|B^c) = \Pr(A^c)$ and that $\Pr(A^c) = 1 - \Pr(A)$.

Exercise 19 Independence when $\Pr(B) = 0$

Suppose A and B are events and $\Pr(B) = 0$. (A is any event.) Find $\Pr(A \cap B)$. Prove that A and B are independent.

Solution

Since $\Pr(B) = 0$, we have $\Pr(A \cap B) \leq \Pr(B) = 0$, hence $\Pr(A \cap B) = 0$. Also $\Pr(A)\Pr(B) = \Pr(A) \cdot 0 = 0$. Therefore $\Pr(A \cap B) = \Pr(A)\Pr(B)$, so A and B are independent.

Exercise 20 Sixes

Suppose a six-sided die is rolled 10 times. What's the probability of…

a. all sixes?
b. not all-sixes?
c. all not-sixes?

Solution

a. $(1/6)^{10}$
b. $1 - (1/6)^{10}$
c. $(5/6)^{10}$

Exercise 21 Drug Testing (Bayes' Rule)

A drug test is used to detect the presence of a banned substance in professional athletes. Suppose that 2 in every 1,000 athletes use the substance. The test correctly identifies users 98% of the time. However, it also produces a false positive 1% of the time for non-users. You are randomly selected for testing, and your result comes back positive. Use Bayes' rule to compute the probability that you actually use the substance.

Solution

Let U be the event that you use the banned substance and T be the event that the test result is positive.

We are given the following values:

- $\Pr(U) = 0.002$
- $\Pr(U^c) = 0.998$
- $\Pr(T \mid U) = 0.98$
- $\Pr(T \mid U^c) = 0.01$

We want to compute $\Pr(U \mid T)$ using Bayes' rule $\Pr(U \mid T) = \frac{\Pr(T|U)\Pr(U)}{\Pr(T)}$.

To compute $\Pr(T)$, we apply the law of total probability $\Pr(T) = \Pr(T \mid U)\Pr(U) + \Pr(T \mid U^c)\Pr(U^c)$.

Substitute the known values, so that $\Pr(T) = (0.98)(0.002) + (0.01)(0.998) = 0.00196 + 0.00998 = 0.01194$.

Now apply Bayes' rule, so that $\Pr(U \mid T) = \frac{(0.98)(0.002)}{0.01194} \approx \frac{0.00196}{0.01194} \approx 0.1642$.

So the chance that you actually use the substance, given a positive test result, is about 16%.

Exercise 22 Bernoulli Distribution

Let $X \sim \text{Bernoulli}(p)$ with pmf

$$f(x;p) = \begin{cases} p, & x = 1, \\ 1-p, & x = 0, \\ 0, & \text{otherwise}, \end{cases} \qquad 0 < p < 1.$$

Compute

1. the cumulative distribution function $F(x) = \Pr(X \leq x)$,

2. the expected value $\mathrm{E}[X]$, and

3. the variance $\mathrm{Var}(X)$.

Solution

1. cdf

Since X is discrete, its cdf is a step function.

- For $x < 0$, $F(x) = 0$

- For $0 \leq x < 1$,

$$F(x) = \Pr(X \leq x) = \Pr(X = 0) = 1 - p$$

- For $x \geq 1$,

$$F(x) = \Pr(X \leq x) = \Pr(X = 0) + \Pr(X = 1) = 1 - p + p = 1$$

So the cdf is:

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - p, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

2. expected value

Using the definition of expected value for a discrete random variable,

$$\mathrm{E}[X] = \sum_x x \cdot \Pr(X = x) = 0 \cdot (1 - p) + 1 \cdot p = p.$$

3. variance

We use the formula:

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - \left(\mathrm{E}[X]\right)^2$$

Since X only takes the values 0 and 1, $X^2 = X$, so $\mathrm{E}[X^2] = \mathrm{E}[X] = p$.

Then:

$$\begin{aligned} \mathrm{Var}(X) &= p - p^2 \\ &= p(1 - p). \end{aligned}$$

Summary

- $F(x) = 1 - p$ for $0 \leq x < 1$, and $F(x) = 1$ for $x \geq 1$
- $\mathrm{E}[X] = p$
- $\mathrm{Var}(X) = p(1-p)$

Exercise 23 Exponential Distribution (Challenging)

Note: This is a very challenging exercise that requires you to evaluate an increasingly difficult set of integrals. It's acceptable to use a symbolic solver for these questions, but it's also good to see what "simple" integrals can look like in "simple" real problems.

Let $X \sim \text{Exponential}(\lambda)$ with probability density function

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0, \end{cases} \qquad \lambda > 0.$$

Compute

1. the cumulative distribution function $F(x) = \Pr(X \leq x)$,
2. the expected value $\mathrm{E}[X]$, and
3. the variance $\mathrm{Var}(X)$.

Hints:

- Use integration by parts to compute $\mathrm{E}[X]$. Set $u = x$ and $dv = \lambda e^{-\lambda x}\,dx$.
- To compute $\mathrm{E}[X^2]$, apply integration by parts **twice**. First let $u = x^2$ and $dv = \lambda e^{-\lambda x} dx$. This will leave an integral of the form $\int x e^{-\lambda x}\,dx$, which you already solved when computing $\mathrm{E}[X]$.
- Use the formula $\mathrm{Var}(X) = \mathrm{E}[X^2] - \left(\mathrm{E}[X]\right)^2$.

Solution

1. cdf

To find $F(x)$, integrate the pdf from 0 to x, so that

$$
\begin{aligned}
F(x) &= \int_0^x \lambda e^{-\lambda t}\,dt \\
&= \left[-e^{-\lambda t}\right]_0^x \\
&= -e^{-\lambda x} + e^{-\lambda \cdot 0} \\
&= 1 - e^{-\lambda x}.
\end{aligned}
$$

Then the cdf is

$$
F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}
$$

2. expected value

We need to to compute

$$
\mathrm{E}[X] = \int_0^\infty x \cdot \lambda e^{-\lambda x}\,dx.
$$

Following the hint, this requires integration by parts.

Let - $u = x$ so that $du = dx$,
- $dv = \lambda e^{-\lambda x}\,dx$ so that $v = -e^{-\lambda x}$.

Then

$$
\begin{aligned}
\mathrm{E}[X] &= \left[-xe^{-\lambda x}\right]_0^\infty + \int_0^\infty e^{-\lambda x}\,dx \\
&= 0 + \left[-\frac{1}{\lambda}e^{-\lambda x}\right]_0^\infty \\
&= 0 + \left(0 - \left(-\frac{1}{\lambda}\right)\right) \\
&= \frac{1}{\lambda}.
\end{aligned}
$$

3. variance

We nee to use the identity $\mathrm{Var}(X) = \mathrm{E}[X^2] - \left(\mathrm{E}[X]\right)^2$.

To compute $\mathrm{E}[X^2]$, compute

$$\mathrm{E}[X^2] = \int_0^\infty x^2 \cdot \lambda e^{-\lambda x}\, dx.$$

Step 1: first integration by parts

Let - $u = x^2$ so that $du = 2x\, dx$,
- $dv = \lambda e^{-\lambda x}\, dx$ so that $v = -e^{-\lambda x}$.

Then

$$\begin{aligned}
\mathrm{E}[X^2] &= \left[-x^2 e^{-\lambda x}\right]_0^\infty + \int_0^\infty 2x e^{-\lambda x}\, dx \\
&= 0 + 2\int_0^\infty x e^{-\lambda x}\, dx.
\end{aligned}$$

Step 2: second integration by parts

Now evaluate $\int_0^\infty x e^{-\lambda x}\, dx$, which is the same type of integral we solved for $\mathrm{E}[X]$, but without the constant λ out front.

Let - $u = x$ so that $du = dx$,
- $dv = e^{-\lambda x}\, dx$ so that $v = -\frac{1}{\lambda} e^{-\lambda x}$.

Then

$$\begin{aligned}
\int_0^\infty x e^{-\lambda x}\, dx &= \left[-\frac{x}{\lambda} e^{-\lambda x}\right]_0^\infty + \int_0^\infty \frac{1}{\lambda} e^{-\lambda x}\, dx \\
&= 0 + \frac{1}{\lambda}\left[-\frac{1}{\lambda} e^{-\lambda x}\right]_0^\infty \\
&= \frac{1}{\lambda^2}.
\end{aligned}$$

So,

$$\mathrm{E}[X^2] = 2 \cdot \frac{1}{\lambda^2} = \frac{2}{\lambda^2}.$$

Now use the variance formula

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}[X^2] - \left(\mathrm{E}[X]\right)^2 \\ &= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 \\ &= \frac{1}{\lambda^2}.\end{aligned}$$

Summary

- $F(x) = 1 - e^{-\lambda x}$ for $x \geq 0$
- $\mathrm{E}[X] = \frac{1}{\lambda}$
- $\mathrm{Var}(X) = \frac{1}{\lambda^2}$

Exercise 24 Some Practice with Expectations

Use the stated rules to simplify each expression.

a. $E[\,2X + 3Y\,]$

b. $E[\,5X - 7Y\,]$

c. $E[\,4X + Y - 2Z\,]$

d. $E[\,10\,]$

e. $E[\,\pi\,]$

f. $E[\,2XY\,]$ (assume X and Y independent)

g. Let X take values $1, 2, 3$ each with probability $1/3$. Compute $E[X^2]$.

h. Let X have pdf $f_X(x) = 2x$ on $[0, 1]$. Compute $E[X^3]$.

i. Let X take values $0, 1$ with $P(X = 1) = p$. Compute $E[\log(1 + X)]$.

Solutions
a. $2E[X] + 3E[Y]$
b. $5E[X] - 7E[Y]$
c. $4E[X] + E[Y] - 2E[Z]$

d. 10
e. π
f. $2E[X]\,E[Y]$
g. $E[X^2] = \frac{1^2+2^2+3^2}{3} = \frac{14}{3}$
h. $E[X^3] = \int_0^1 x^3 \cdot 2x\,dx = \int_0^1 2x^4\,dx = \frac{2}{5}$
i. $E[\log(1+X)] = (1-p)\log(1) + p\log(2) = p\log(2)$

Exercise 25 Some Practice with Variances

Use the stated rules to simplify each expression.

a. $\mathrm{Var}(X)$ in terms of $E[X^2]$ and $E[X]$

b. $\mathrm{Var}(3X)$ (express in terms of $E[X^2]$ and $E[X]$)

c. $\mathrm{Var}(X+Y)$ in terms of $\mathrm{Var}(X)$, $\mathrm{Var}(Y)$, and $\mathrm{Cov}(X,Y)$

d. $\mathrm{Var}(X+Y)$ if X and Y are independent

e. $\mathrm{Var}(2X-3Y)$ (assume independence)

Solutions

a. $E[X^2] - (E[X])^2$

b. $E[9X^2] - (3E[X])^2 = 9E[X^2] - 9(E[X])^2$

c. $\mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$

d. $\mathrm{Var}(X) + \mathrm{Var}(Y)$

e. $4\,\mathrm{Var}(X) + 9\,\mathrm{Var}(Y)$

Exercise 26 Some Practice with Covariances

Use the stated rules to simplify each expression.

a. $\mathrm{Cov}(2X, 3Y)$

b. $\text{Cov}(-X, 4Y)$

c. $\text{Cov}(2X - 3Y, Z)$

d. $\text{Cov}(2X, Y)$ if X and Y are independent

Solutions

a. $6\,\text{Cov}(X, Y)$

b. $-4\,\text{Cov}(X, Y)$

c. $2\,\text{Cov}(X, Z) - 3\,\text{Cov}(Y, Z)$

d. $2\,\text{Cov}(X, Y) = 0$

Exercise 27 OLS, Interactions, and R

A Theoretical Result

For this portion, it will be helpful to have Brambor, Clark, and Golder (2006) handy.

Consider the linear interaction model $Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + \varepsilon$ estimated by OLS. For a given value $Z = z$, the marginal effect of X on $E(Y)$ is $\beta_1 + \beta_3 z$, with plug-in estimate $\widehat{\text{ME}}_X(z) = \hat{\beta}_1 + \hat{\beta}_3 z$ (i.e., see eq. 13 in Brambor, Clark, and Golder 2006).

Let $\widehat{\text{Var}}(\hat{\beta}_1)$, $\widehat{\text{Var}}(\hat{\beta}_3)$, and $\widehat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_3)$ denote the corresponding elements from the estimated variance–covariance matrix of $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$. Treat z as fixed. Re-derive the standard error of $\widehat{\text{ME}}_X(z)$ shown in eq. 8 of Brambor, Clark, and Golder (2006). In other words, show that

$$\widehat{\text{SE}}\{\widehat{\text{ME}}_X(z)\} = \sqrt{\widehat{\text{Var}}\left(\hat{\beta}_1 + \hat{\beta}_3 z\right)} = \sqrt{\widehat{\text{Var}}(\hat{\beta}_1) + z^2\,\widehat{\text{Var}}(\hat{\beta}_3) + 2z\,\widehat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_3)}\,.$$

Solution

First, we have

$$\text{Var}\left(\hat{\beta}_1 + \hat{\beta}_3 z\right) = \text{Var}\left(\hat{\beta}_1\right) + \text{Var}\left(\hat{\beta}_3 z\right) + 2\text{Cov}\left(\hat{\beta}_1, \hat{\beta}_3 z\right)$$

by the rules for the variance of a sum. Then we have

$$= \text{Var}\left(\hat{\beta}_1\right) + z^2\text{Var}\left(\hat{\beta}_3\right) + 2z\text{Cov}\left(\hat{\beta}_1, \hat{\beta}_3\right)$$

by the rules for variance of aX and covariance of X and aY. The standard error is simply the square root of the variance.

Thus, eq. 8 in Brambor, Clark, and Golder follows directly from the application of the rules for variances and covariances.

A Computational Result

For this portion, it will be helpful to have Clark and Golder (2006) handy.

The code below reproduces Clark and Golder's (2006) 1946-2000 Established Democracies model in Table 2 on p. 698. See also eq. 4 on p. 695 for their model specification.

```
# load packages
library(sandwich)  # for robust SEs

# load Clark and Golder's data
cg <- crdata::cg2006  # from my data package

# reproduce regression
f <- enep ~ eneg*log(average_magnitude) + eneg*upper_tier + en_pres*proximity
fit <- lm(f, data = cg)

# grab estimated coefs and var matrix
beta_hat <- coef(fit)
V_hat <- vcovCL(fit, cluster = ~ country, type = "HC1")

# print table (shows we reproduce!)
# modelsummary::modelsummary(fit, vcov = V_hat, fmt = 2)
```

1. Derive the formula for the marginal effect of `eneg` and its SE. *Hint: It depends on both `average_magnitude` and `upper_tier`.
2. Use R to compute the marginal effect of `eneg` for the maximum value of `log(average_magnitude)` and it's SE. (You can fix `upper_tier` to 0, which simplifies the calculation.)
3. Use R to compute the marginal effect of `eneg` and SE for a *range* of values from the minimum value of `log(average_magnitude)`. Compute a 90% confidence interval and create a marginal effect plot like those in Figure 1 of Clark and Golder (2006). *Note: you should have the log of `average_magnitude` (not `average_magnitude` itself) along the x-axis.*

Solution

Question 1

Clark and Golder fit the model

$$
\begin{aligned}
\text{ElectoralParties} = \beta_0 &+ \beta_1 \,\text{Ethnic} + \beta_2 \ln(\text{Magnitude}) + \beta_3 \,\text{UppertierSeats} \\
&+ \beta_4 \,\text{PresidentCandidates} + \beta_5 \,\text{Proximity} \\
&+ \beta_6 \,\text{Ethnic} \times \ln(\text{Magnitude}) + \beta_7 \,\text{Ethnic} \times \text{UppertierSeats} \\
&+ \beta_8 \,\text{PresidentCandidates} \times \text{Proximity} + \epsilon.
\end{aligned}
$$

Taking the derivative w.r.t. Ethnic, we have

$$
\frac{\partial \,\text{ElectoralParties}}{\partial \,\text{Ethnic}} = \beta_1 + \beta_6 \ln(\text{Magnitude}) + \beta_7 \,\text{UppertierSeats}.
$$

Question 2

```r
# load packages
library(tinytable)

# value of z = log(average_magnitude) at its maximum
z <- max(log(cg$average_magnitude), na.rm = TRUE)

# point estimate of marginal effect of eneg
me <- beta_hat["eneg"] + z * beta_hat["eneg:log(average_magnitude)"]

# variance components
v1  <- V_hat["eneg", "eneg"]
v3  <- V_hat["eneg:log(average_magnitude)", "eneg:log(average_magnitude)"]
c13 <- V_hat["eneg", "eneg:log(average_magnitude)"]

# standard error from the formula
se <- sqrt(v1 + z^2 * v3 + 2 * z * c13)

# 90% confidence interval
ci <- me + c(-1, 1)*1.64 * se

data.frame(
  log_mag = z,
  me,
  se,
  ci_lower = ci[1],
  ci_upper = ci[2]
)
```

```
      log_mag       me        se   ci_lower ci_upper
eneg 5.010635 1.432714 0.8246638 0.08026526 2.785162
```

Question 3

```r
# load packages
library(ggplot2)
```

```

Attaching package: 'ggplot2'


The following object is masked from 'package:tinytable':

    theme_void
```

```r
# grid over observed log(average_magnitude)
z<- seq(min(log(cg$average_magnitude), na.rm = TRUE),
             max(log(cg$average_magnitude), na.rm = TRUE),
             length.out = 200)

# point estimate of marginal effect of eneg
me <- beta_hat["eneg"] + z * beta_hat["eneg:log(average_magnitude)"]

# standard error from the formula
se <- sqrt(v1 + z^2 * v3 + 2 * z * c13)

# 90% confidence interval
ci_lower <- me - 1.64 * se
ci_upper <- me + 1.64 * se


# combine qis into data frame
gg_data <- data.frame(
  log_mag = z,
  me,
  se,
  ci_lower,
  ci_upper
)

# draw plot
ggplot(gg_data, aes(x = log_mag, y = me,
                     ymin = ci_lower,
                     ymax = ci_upper)) +
```

```
  geom_ribbon(fill = "grey70") +
  geom_line()
```