

Problem Set 1

Clinton Tepper

Part I

1. Definition of expectation:

$$E[X] = \int_{s \in \mathfrak{R}} p(s) s ds$$

2. Definition of variance:

$$\begin{aligned} V[X] &= E[X^2] - E[X]^2 \\ &= \int_{s \in \mathfrak{R}} p(s) s^2 ds - \left(\int_{s \in \mathfrak{R}} p(s) s ds \right)^2 \end{aligned}$$

3. Definition of $E[Y|X]$:

$$E[Y|X = x] = \frac{\int_{s \in \mathfrak{R}} p(x, s) s ds}{p_X(x)}$$

4. Independence

$$\begin{aligned} p(X, Y) &= p(X) p(Y) \\ E[Y|X = x] &= \frac{\int_{s \in \mathfrak{R}} p(x, s) s ds}{p_X(x)} \\ &= \frac{\int_{s \in \mathfrak{R}} p_X(x) p_Y(s) s ds}{p_X(x)} \\ &= \int_{s \in \mathfrak{R}} p_Y(s) s ds \\ &= E[Y] \end{aligned}$$

5. Scalar expectations:

$$\begin{aligned} E[aX_1 + bX_2 + cX_3] &= (a + b + c) E[X] \\ V[aX_1 + bX_2 + cX_3] &= (a + b + c)^2 V[X] \end{aligned}$$

6. Yes!

$$\begin{aligned} E\left[\frac{1}{N} \sum_i X_i\right] &= \frac{1}{N} E\left[\sum_i X_i\right] \\ &= \frac{1}{N} \sum_i E[X_i] \\ &= E[X] \end{aligned}$$

7. Variance of \bar{X} :

$$\begin{aligned}
 V[\bar{X}] &= E[\bar{X}^2] - E[\bar{X}]^2 \quad (\text{by q2 definition}) \\
 &= E[\bar{X}^2] - E[X]^2 \quad (\text{by q6}) \\
 &= \frac{1}{N^2} E\left[\left(\sum_i X_i\right)^2\right] - E[X]^2 \\
 &= \frac{1}{N^2} E\left[\sum_i X_i^2 + \sum_{i \neq j} X_i X_j\right] - E[X]^2 \\
 &= \frac{1}{N^2} \sum_i E[X_i^2] + \frac{1}{N^2} \sum_{i \neq j} E[X_i] E[X_j] - E[X]^2 \quad (\text{linearity \& independence}) \\
 &= \frac{1}{N} E[X^2] + \frac{N-1}{N} E[X]^2 - E[X]^2 \quad (\text{iid \& linearity}) \\
 &= \frac{1}{N} (E[X^2] - E[X]^2) \\
 &= \frac{1}{N} V[X] \\
 \lim_{n \rightarrow \infty} \frac{1}{N} V[X] &= 0
 \end{aligned}$$

Part II

8. Start with FOC:

$$\begin{aligned}
 0 &= -2X'Y + 2X'X\beta \\
 \beta &= [X'X]^{-1} [X'Y] \quad (\text{premultiply by } [X'X]^{-1})
 \end{aligned}$$

9. See below code:

```
p1q1 = function(N = 100) {

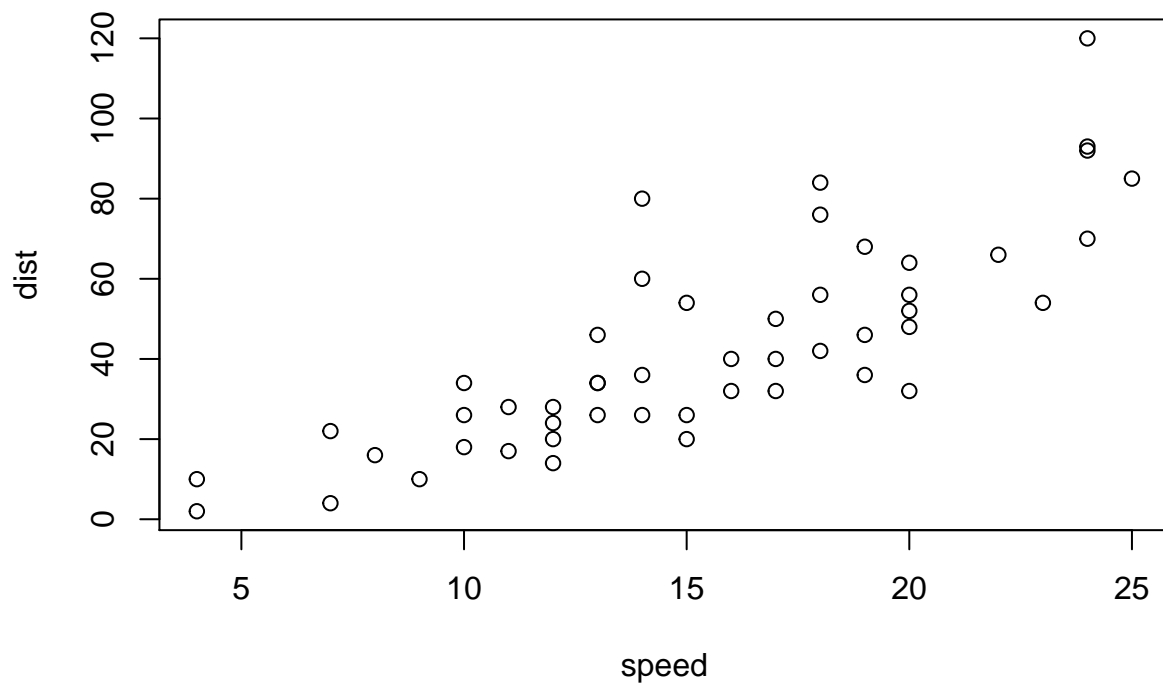
  X = matrix(c(rep(1, N), rnorm(N), rnorm(N)), nrow = N, ncol = 3)
  ep = rnorm(N)
  B = c(1, 2, 3)
  Y = X %*% B + ep

  Xqr = qr(X)
  Q = qr.Q(Xqr)
  R = qr.R(Xqr)
  Best = solve(R) %*% t(Q) %*% Y
  cat("Our estimate: ", Best, "\n")
  cat("lm estimate:\n")
  print(coef(lm(Y ~ X + 0)))
}

p1q1(N=100)
```

```
## Our estimate:  1.103211 2.142035 2.816583
## lm estimate:
##      X1      X2      X3
## 1.103211 2.142035 2.816583
```

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.