Exercise Set 9-1

1) This statement is false. In frequentist statistics, θ is not random, so we cannot make probability statements about it. In Bayesian statistics, the value of θ that maximizes $L(\theta)$ is not generally the most probable value of θ given the data, but sometimes it is (see chapter 10). A better statement is "*The value of θ that maximizes $L(\theta)$ is the one that maximizes the probability of obtaining the observed data*." This statement is correct for discrete random variables; for continuous random variables, it could be modified to "*The value of θ that maximizes $L(\theta)$ is the one that maximizes the joint probability density associated with the observed data*."

2) a) The density is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

b) The log-likelihood is the log of the density,

$$l(\mu) = \ln\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right).$$

This expression can by simplified. First, notice that it is a product and that the log of a product is the sum of the logs of the terms being multiplied,

$$l(\mu) = \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) + \ln\left(e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right)$$

The second term contains the log of an exponent. Raising e to a power and taking a natural log are inverse operations, so

$$l(\mu) = \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(x-\mu)^2}{2\sigma^2}.$$

We could simplify the first term further, but this is enough for us.

c) Because the two random variables are independent, the joint density function is the product of the two marginal density functions,

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} * \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_2-\mu)^2}{2\sigma^2}}.$$

This expression simplifies to

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{\sigma^2 2\pi} e^{-\frac{(x_1-\mu)^2+(x_2-\mu)^2}{2\sigma^2}}.$$

d) The log-likelihood for two observations is the log of the density of two observations. We could either take the log of the expression in part (c) directly, or we could start with part (b), remembering that the log of a product is the sum of the logs of the terms being multiplied. Either way, we obtain an expression equivalent to

$$l(\mu) = 2\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(x_1-\mu)^2}{2\sigma^2} - \frac{(x_2-\mu)^2}{2\sigma^2}.$$

e) Extending the result in part (d), we have

$$l(\mu) = n\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}.$$

Exercise Set 9-2

1) a) The likelihood function is equal to the joint probability mass function. Because the observations are assumed to be independent, their joint probability mass function is the product of their individual probability mass functions. Thus, it is

$$L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}.$$

b) The log-likelihood is the natural log of the likelihood function, which in this case is

$$l(p) = \ln[L(p)] = \ln\left[\prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}\right] = \sum_{i=1}^{n}\ln[p^{x_i}(1-p)^{1-x_i}]$$
$$= \sum_{i=1}^{n} x_i\ln(p) + (1-x_i)\ln(1-p).$$

c) Here is some `R` code that draws the requested samples and plots the likelihood and log-likelihood functions.

```
#Draw a sample
n <- 10
p <- 0.6
x <- rbinom(n,1,p)

#Given a vector of values for p and a vector of Bernoulli trials
#x, this function computes the
#likelihood for each value of p.
Ln.Bern <- function(p, x){
  k <- sum(x)
  n <- length(x)
```

```
  Ln <- numeric(length(p))
  for(i in 1:length(p)){
    Ln[i] <- prod(p[i]^k * (1-p[i])^(n-k))
  }
  return(Ln)
}

#a set of values of p to plot
p <- seq(0.001, 0.999, length.out = 999)

#The likelihood.
Ln <- Ln.Bern(p, x)
#The log-likelihood
ln <- log(Ln)
plot(p, Ln, type = "l")
plot(p, ln, type = "l")
```

d) The maximum-likelihood estimate of p is the mean of the sample, or the proportion of "1" outcomes.

2) a) The likelihood function, remembering that the observations are independent and identically distributed, is

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x_i-\theta)^2}{2\sigma^2}} = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2}.$$

The log-likelihood function is then

$$\begin{aligned} l(\theta) = \ln[L(\theta)] &= n\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2 \\ &= n\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i^2 - 2x_i\theta + \theta^2). \end{aligned}$$

To maximize the log-likelihood, we take the derivative with respect to θ:

$$\frac{\partial l(\theta)}{\partial\theta} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(\theta - x_i) = -\frac{n}{\sigma^2}(\theta - \bar{x}),$$

where $\bar{x}$ is the sample mean, $\frac{1}{n}\sum_{i=1}^{n} x_i$. Because n and σ^2 are both positive, the only value of θ that sets the derivative equal to 0 is $\theta = \bar{x}$. To check that this solution maximizes, rather than minimizes, the log-likelihood function, we confirm that the second derivative is negative (exercise 2 of exercise set A-1):

$$\frac{\partial^2 l(\theta)}{\partial \theta^2} = -\frac{n}{\sigma^2}.$$

Setting $\theta = \bar{x}$ maximizes the log-likelihood function, and thus the likelihood function. The maximum-likelihood estimator of θ is therefore $\hat{\theta} = \bar{x}$, the sample mean.

b) The natural logs of the observations have a normal distribution, $\ln(Y_i) \sim \text{Normal}(\theta, \sigma^2)$ for all i. Thus, by part a, the maximum-likelihood estimator of θ is

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} \ln(Y_i).$$

By the functional invariance property of maximum-likelihood estimators, the maximum likelihood estimator of e^{θ} is $e^{\hat{\theta}}$. Notice that this is different from the method-of-moments estimator, which would estimate $e^{\theta} = \mathrm{E}(Y)$ as the sample mean, $\frac{1}{n}\sum_{i=1}^{n} Y_i$.

Exercise Set 9-3

1) To find the expectation of $\hat{\beta}$, start by making the substitution indicated in the problem's hint:

$$\begin{aligned}\mathrm{E}(\hat{\beta}) &= \mathrm{E}\left(\frac{\sum_{i=1}^{n} x_i Y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}\right) \\ &= \mathrm{E}\left(\frac{\sum_{i=1}^{n} x_i(\alpha + \beta x_i + \epsilon_i) - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n}(\alpha + \beta x_i + \epsilon_i)}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}\right).\end{aligned}$$

Now, expanding the terms in the numerator, using the linearity of expectation (equation 5.4), and noticing that the only random variables are the disturbance terms, we rewrite as

$$\begin{aligned}&\mathrm{E}(\hat{\beta}) \\ &= \frac{\alpha\sum_{i=1}^{n} x_i + \beta\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} x_i\,\mathrm{E}(\epsilon_i) - n\alpha\frac{1}{n}\sum_{i=1}^{n} x_i - \beta\frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 + \frac{1}{n}\sum_{i=1}^{n} x_i\,\mathrm{E}(\epsilon_i)}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}.\end{aligned}$$

From here, we can make several simplifications. The two α terms in the numerator exactly cancel each other, and the two terms with the expectation of the disturbances ϵ_i disappear because $\mathrm{E}(\epsilon_i) = 0$, leaving

$$\mathrm{E}(\hat{\beta}) = \frac{\beta\sum_{i=1}^{n} x_i^2 - \beta\frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} = \beta\,\frac{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} = \beta.$$

Because $\mathrm{E}(\hat{\beta}) = \beta$, $\hat{\beta}$ is an unbiased estimator of β.

We find the expectation of $\hat{\alpha}$ similarly, beginning by making the substitution in the hint,

$$\mathrm{E}(\hat{\alpha}) = \mathrm{E}\left(\frac{\sum_{i=1}^{n} Y_i - \hat{\beta}\ \sum_{i=1}^{n} x_i}{n}\right) = \mathrm{E}\left(\frac{\sum_{i=1}^{n}(\alpha + \beta x_i + \epsilon_i) - \hat{\beta}\ \sum_{i=1}^{n} x_i}{n}\right).$$

Applying equation 5.4 (linearity of expectation) gives

$$\mathrm{E}(\hat{\alpha}) = \frac{n\alpha + \beta\ \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \mathrm{E}(\epsilon_i) - \mathrm{E}(\hat{\beta})\ \sum_{i=1}^{n} x_i}{n}.$$

Remembering that $\mathrm{E}(\epsilon_i) = 0$ and $\mathrm{E}(\hat{\beta}) = \ \beta$, the expression simplifies to

$$\mathrm{E}(\tilde{\alpha}) = \frac{n\alpha}{n} = \alpha.$$

Because $\mathrm{E}(\hat{\alpha}) = \alpha$, $\hat{\alpha}$ is an unbiased estimator of α.

The unbiasedness of the least-squares estimators is not guaranteed by their status as method-of-moments or maximum-likelihood estimators, so we had to show it directly. Notice also that we did not rely on the assumptions of normality of disturbances or constant variance of the disturbances. We did not even invoke independence of the disturbance terms—we just used $\mathrm{E}(\epsilon_i) = 0$ for all x. Thus, the least-squares estimator is also unbiased under the weaker assumptions used in chapter 8.

2) To identify the variance, start with the substitution indicated in the problem's hint:

$$\begin{aligned}\mathrm{Var}(\hat{\beta}) &= \mathrm{Var}\left(\frac{\sum_{i=1}^{n} x_i Y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2}\right) \\ &= \mathrm{Var}\left(\frac{\sum_{i=1}^{n} x_i(\alpha + \beta x_i + \epsilon_i) - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n}(\alpha + \beta x_i + \epsilon_i)}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2}\right).\end{aligned}$$

Because the x_i are not random, many of these terms are constants and thus do not influence the variance (equation 5.8). Dropping the constants leaves

$$\mathrm{Var}(\hat{\beta}) = \mathrm{Var}\left(\frac{\sum_{i=1}^{n} x_i \epsilon_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} \epsilon_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2}\right).$$

In the numerator, we have two functions of the disturbances. Making the substitution $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and moving $\bar{x}$ inside the sum lets us combine them,

$$\mathrm{Var}(\hat{\beta}) = \mathrm{Var}\left(\frac{\sum_{i=1}^{n} x_i\epsilon_i - \bar{x}\sum_{i=1}^{n}\epsilon_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}\right) = \mathrm{Var}\left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})\epsilon_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}\right).$$

The variance of $\hat{\beta}$ is thus the sum of functions of the variance of the disturbances. Because the disturbances are independent, the variance of the sum is the sum of the variances of the individual terms (equation 5.9). Applying this insight and equation 5.8 ($\mathrm{Var}(a + cX) = c^2\mathrm{Var}(X)$) gives

$$\mathrm{Var}(\hat{\beta}) = \frac{1}{\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]^2}\sum_{i=1}^{n}(x_i - \bar{x})^2\mathrm{Var}(\epsilon_i).$$

For all i, $\mathrm{Var}(\epsilon_i) = \sigma^2$ by assumption. We can therefore pull it out of the sum, leaving

$$\mathrm{Var}(\hat{\beta}) = \sigma^2\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]^2}.$$

The last step is to notice that, by the identity given in the hint, $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2$, the denominator of the fraction on the right is the square of the numerator. Applying this insight yields two equivalent simplified expressions for the variance:

$$\mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

The second form makes clear that the variance of $\hat{\beta}$ decreases as the number of observations increases. In this proof, we relied on the independence and constant variance of the disturbances, but we did not invoke their normality.

3) Start by writing the likelihood:

$$L(\sigma^2) = \prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(y_i - \alpha - \beta x_i)^2}{2\sigma^2}}.$$

We define $v = \sigma^2$ and make the appropriate substitutions, as suggested in the hint:

$$L(v) = \prod_{i=1}^{n} \frac{1}{\sqrt{v}\sqrt{2\pi}} e^{\frac{-(y_i-\alpha-\beta x_i)^2}{2v}}.$$

The log of this expression is the log-likelihood,

$$l(v) = n \ln\left(\frac{1}{\sqrt{v}\sqrt{2\pi}}\right) - \frac{1}{2v}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2.$$

We need to take the derivative of the log-likelihood with respect to v. For the first term, notice that by equation 9.3,

$$n \ln\left(\frac{1}{\sqrt{v}\sqrt{2\pi}}\right) = n \ln\left(\frac{1}{\sqrt{v}} * \frac{1}{\sqrt{2\pi}}\right) = n\left[\ln\left(\frac{1}{\sqrt{v}}\right) + \ln\left(\frac{1}{\sqrt{2\pi}}\right)\right] = -n \ln(\sqrt{v}) - n \ln(\sqrt{2\pi}).$$

We make this substitution to take the derivative,

$$\begin{aligned}\frac{\partial}{\partial v} l(v) &= \frac{\partial}{\partial v}\left[-n \ln(\sqrt{v}) - n \ln(\sqrt{2\pi}) - \frac{1}{2v}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2\right] \\ &= -\frac{n}{2v} + \frac{1}{2v^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2.\end{aligned}$$

The first term of the derivative comes from the problem hint. To set the derivative equal to zero, we need to solve the equation

$$0 = -\frac{n}{2v} + \frac{1}{2v^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2.$$

Adding $n/(2v)$ to both sides and then multiplying both sides by $2v^2/n$ gives the unique solution,

$$v = \frac{1}{n}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2.$$

We omit the step of showing that the second derivative of the log-likelihood is negative. The maximum-likelihood estimator of σ^2 is

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \alpha - \beta x_i)^2,$$

This is what we would do if we *knew* α and β already and only needed to estimate σ^2. This situation is rare in practice. When α and β are unknown, we replace them with their maximum-likelihood estimators, and the maximum-likelihood estimate of σ^2 becomes

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{\alpha} - \hat{\beta}x_i\right)^2,$$

or the average of the squared line errors. However, if α and β are unknown and must be estimated, then the expectation of the maximum-likelihood estimator of the disturbance variance is

$$\mathrm{E}\left(\widehat{\sigma^2}\right) = \frac{n-2}{n}\sigma^2.$$

Thus, the maximum likelihood estimator is biased downward. One way to understand this bias is to notice that the maximum-likelihood estimator of σ^2 is directly proportional to the sum of the squared line errors, and $\hat{\alpha}$ and $\hat{\beta}$ are chosen to make the sum of the squared line errors *as small as possible*. (They are the *least-squares* estimates!) Thus, to the extent that our estimates of α and β err, they will err in ways that make the sum of the squared line errors smaller than they would be if the true values of α and β were known.

One unbiased estimator of the variance of the disturbances is

$$\widetilde{\sigma^2} = \frac{n}{n-2}\widehat{\sigma^2} = \frac{1}{n-2}\sum_{i=1}^{n}\left(Y_i - \hat{\alpha} - \hat{\beta}x_i\right)^2.$$

When n is large, $n/(n-2) \approx 1$, and the two estimators are nearly identical. In practice, the unbiased estimator is used more often.

4) Starting from the definition of $\hat{\beta}$ and making the substitution $Y_i = \alpha + \beta x_i + \epsilon_i$ gives

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i(\alpha + \beta x_i + \epsilon_i) - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n}(\alpha + \beta x_i + \epsilon_i)}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}.$$

Making the substitution $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and collecting the sums in the numerator gives

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\alpha + \beta x_i + \epsilon_i)}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}.$$

We split this sum into two components, a random one and a non-random one,

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\alpha + \beta x_i)}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2} + \frac{\sum_{i=1}^{n}(x_i - \bar{x})\,\epsilon_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2}.$$

The first term is fixed, and by (i), it does not affect whether $\hat{\beta}$ is normally distributed. We label it c and ignore it.

$$\hat{\beta} = c + \frac{\sum_{i=1}^{n}(x_i - \bar{x})\,\epsilon_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2}.$$

The second term on the right is the sum of n independent random variables, each of which is the product of a non-random term and ϵ_i. By (i), each of the individual random variables is normally distributed. Because the individual random variables are independent and normally distributed, (ii) guarantees that their sum is also normally distributed. Thus, $\hat{\beta}$ is normally distributed. In combination, you have proven in exercises 1, 2, and 4 that given the assumptions in this section,

$$\hat{\beta} \sim \text{Normal}\left(\beta, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right).$$

Exercise Set 9-4

1) a) The likelihood function is

$$L(\theta) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!}.$$

The log-likelihood is

$$l(\theta) = \sum_{i=1}^{n}[-\theta + x_i \ln\theta - \ln(x_i!)] = -n\theta + n\bar{x}\ln\theta - \sum_{i=1}^{n}\ln(x_i!),$$

where $\bar{x}$ is the sample mean. Taking the derivative of the log-likelihood with respect to θ gives

$$\frac{\partial}{\partial\theta} l(\theta) = -n + \frac{n\bar{x}}{\theta}.$$

Setting the log-likelihood to 0 gives

$$0 = n\left(\frac{\bar{x}}{\theta} - 1\right) \Rightarrow \; \theta = \bar{x}.$$

Setting $\theta = \bar{x}$ maximizes the likelihood; thus $\hat{\theta} = \bar{x}$ is the maximum-likelihood estimator of θ.

b) $\hat{\theta} = \bar{x}$, the sample mean, is the maximum-likelihood estimator of θ. Each observation averaged in $\bar{x}$ is independent, and because they are distributed as $\text{Poisson}(\theta)$, their variance is θ. By equations 5.8 and 5.9,

$$\text{Var}(\bar{x}) = \text{Var}\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{n\theta}{n^2} = \frac{\theta}{n}.$$

We could estimate the variance of $\hat{\theta}$ by plugging the estimator $\hat{\theta}$ in for θ.

c) Picking up where we left off in part (a), the first derivative of the log-likelihood is

$$\frac{\partial}{\partial\theta} l(\theta) = -n + \frac{n\bar{x}}{\theta}.$$

The second derivative is then

$$\frac{\partial^2}{\partial\theta^2} l(\theta) = -\frac{n\bar{x}}{\theta^2}.$$

The Fisher Information is the negative expectation of

$$-\frac{n\bar{X}}{\theta^2},$$

where $\bar{X}$ is a random variable representing the mean of a sample of independent, identically distributed $\text{Poisson}(\theta)$ random variables. The expectation of $\bar{X}$ is θ, so

$$\mathcal{I}(\theta) = \frac{n}{\theta}.$$

By equation 9.17, the asymptotic variance of $\hat{\theta}$ is

$$\text{Var}(\hat{\theta}) = \frac{\theta}{n},$$

which we would estimate by plugging in $\hat{\theta}$ for θ. The Fisher Information method only gives us the asymptotic variance of $\hat{\theta}$, but we know from the direct method (part b) that this is also the small-sample variance.

Exercise Set 9-5

1) a) The Wald test statistic (equation 9.20) is

$$W \approx W^* = \frac{\hat{\beta} - \beta_0}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta})}}.$$

The maximum-likelihood estimate $\hat{\beta}$ is the least-squares slope, which we have already computed as $\hat{\beta} = 0.5$. (See, for example, chapter 3.) Because of the null hypothesis specified in the problem, $\beta_0 = 0$. The estimated variance of $\hat{\beta}$ is

$$\widehat{\mathrm{Var}}(\hat{\beta}) = \frac{\widetilde{\sigma^2}}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

where $\widetilde{\sigma^2}$ is as given in equation 9.14b. Computing this value for the example data gives a result of approximately 0.014. The following `R` function computes the test statistic:

```
#Function to compute Wald statistic for slope in simple
#linear regression.
wald.stat.slr <- function(x, y, B0 = 0){
  n <- length(x)
  #compute MLEs of beta and alpha
  B.hat <- (sum(x*y)-sum(x)*sum(y)/n)/( sum(x^2) - sum(x)^2/n)
  A.hat <- (sum(y) - B.hat*sum(x))/n
  #Compute estimated variance of MLE of beta
  vhat.dists <- sum((y - A.hat - B.hat*x)^2)/(n-2)
  vhat.Bhat <- vhat.dists/sum((x - mean(x))^2)
  #Wald statistic
  wald <- (B.hat - B0)/sqrt(vhat.Bhat)
  return(wald)
}
```

The `lm()` function also computes the Wald statistic, but it labels it "t." Applying the function to the agricultural data in the running example gives $W^* \approx 4.24$. By equation 9.21, $p = 2\varphi(-|W^*|) \approx 0.00002$. Comparing the test statistic against the appropriate t distribution gives a p value of 0.002 (using the t distribution with 9 degrees of freedom, because there are 11 data points minus two parameters being estimated), which is in close agreement with the permutation test from chapter 8.

2) Under the null hypothesis, the Wald test statistic is distributed as $\mathrm{Normal}(0,1)$. The square of the Wald statistic is thus distributed as the square of a $\mathrm{Normal}(0,1)$ random variable—in other words, it is distributed as $\chi^2(1)$.

3) Modifying the function for simulating permutation-test p values from exercise set 8-5, problem 2, the below function will return Wald-test p values from simulated datasets with independent, normally distributed disturbances (by default). The x values are here drawn from a normal distribution by default:

```
sim.Wald.B <- function(a, b, B0 = 0, n.sim = 1000, var.eps = 1, 
n = 50,
                       mu.x = 8, var.x = 4, rdist = rnorm, rx = 
rnorm, pfun = pnorm, ...){
  #Initialize variables.
  ps <- numeric(n.sim)
  for(i in 1:n.sim){
    #Simulate data and compute p value.
    dat <- sim.lm(a, b, var.eps, n, mu.x, var.x, rdist = rdist, 
rx = rx)
    x <- dat[,1]
    y <- dat[,2]
    #compute MLEs of beta and alpha
    B.hat <- (sum(x*y)-sum(x)*sum(y)/n)/( sum(x^2) - sum(x)^2/n)
    A.hat <- (sum(y) - B.hat*sum(x))/n
    #Compute estimated variance of MLE of beta
    vhat.dists <- sum((y - A.hat - B.hat*x)^2)/(n-2)
    vhat.Bhat <- vhat.dists/sum((x - mean(x))^2)
    #Wald statistic
    wald <- (B.hat - B0)/sqrt(vhat.Bhat)
    ps[i] <- 2*pfun(-abs(wald), ...)
  }
  #Return the p values
  return(ps)
}
```

To call the function, assess the rate at which the null hypothesis is rejected with a significance level of 0.05, and plot a histogram of the permutation p values for $n = 10$ and $\beta = 0$, use the following commands:

```
> ps <- sim.Wald.B(0, 0, n = 10)
> mean(ps < 0.05)
> hist(ps)
```

When I ran these simulations, I arrived at the following results. Your exact results may differ slightly.

	$n = 10$	$n = 50$	$n = 100$
$\beta = 0$	0.088	0.060	0.039
$\beta = 0.1$	0.127	0.317	0.530
$\beta = 0.2$	0.257	0.789	0.956

The top row of the Wald test table, in which $\beta = 0$, is somewhat unsettling. When $n = 10$, the Wald test produces $p < 0.05$ in 9% of the simulations. This problem is attributable to the fact that the variance of the disturbances is unknown and has to be estimated. As mentioned in the main text, the normal null distribution for the Wald statistic depends on the assumption that the variance of the disturbances is known. Increasing n allows for better estimation of the unknown variance, and the problem is ameliorated. One solution—the one adopted by `R`'s `lm()` function—is to compare the Wald statistic to an appropriate t distribution rather than a standard normal distribution. (You can simulate the type I error rate when the t distribution is used for comparison with, for example,

```
mean(sim.Wald.B(0, 0, n = 10, pfun = pt, df = 8) < .05)
```

The `df` parameter should be set to two less than `n`. When comparing with the t distribution, the table looks much like the one for the permutation test.)

For comparison, here are the results I obtained with the permutation test:

	$n = 10$	$n = 50$	$n = 100$
$\beta = 0$	.034	0.046	0.052
$\beta = 0.1$	0.086	0.300	0.486
$\beta = 0.2$	0.182	0.760	0.978

In these simulations, the power of the Wald test is similar to that of the permutation test for $n = 50$ and $n = 100$. At $n = 10$, the Wald test has greater power than the permutation test, but this is of little use—the Wald test is untrustworthy for $n = 10$ (because of the high type I error rate).

Exercise Set 9-6

1) a) There are several ways to show this, but here is one. Recall that the log-likelihood is

$$l(\alpha, \beta) = n \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2.$$

Under the null hypothesis that $\beta = 0$, the log-likelihood becomes

$$l(\alpha) = n \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \alpha)^2 = n \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i^2 - 2y_i\alpha - \alpha^2).$$

The derivative with respect to α is

$$\frac{\partial l(\alpha)}{\partial \alpha} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(2y_i - 2\alpha) = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \alpha) = -\frac{n}{\sigma^2}(\bar{y} - \alpha),$$

Where the last step follows because $\bar{y} = \sum_{i=1}^{n} y_i / n$. Setting the derivative of the log-likelihood to zero maximizes the log-likelihood and is achieved by setting $\alpha = \bar{y}$. Thus, when β is constrained to be zero, $\hat{\alpha} = \bar{y}$.

c) The likelihood-ratio test statistic (equation 9.20) is

$$\Lambda = 2\ln\left(\frac{L(\hat{\theta})}{L(\widehat{\theta_0})}\right) = 2\left(l(\hat{\theta}) - l(\widehat{\theta_0})\right).$$

Here, $l(\hat{\theta})$ is the value of the log-likelihood at its maximum when all parameters are free, and $l(\widehat{\theta_0})$ is the value of the log-likelihood at its maximum assuming that $\beta = 0$. In part (a), you showed that if $\beta = 0$, then $\hat{\alpha} = \bar{y}$. Thus, Λ becomes

$$\Lambda = 2\left[\left(n\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \hat{\alpha} - \hat{\beta}x_i\right)^2\right) - \left(n\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \bar{y})^2\right)\right].$$

Noticing that the natural log terms cancel, as do the 2's in front and in the denominator of the remaining terms, this simplifies to

$$\Lambda = \frac{1}{\sigma^2}\left[\sum_{i=1}^{n}(y_i - \bar{y})^2 - \sum_{i=1}^{n}\left(y_i - \hat{\alpha} - \hat{\beta}x_i\right)^2\right],$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the maximum-likelihood estimates of α and β (equations 9.10 and 9.11). The expression can be simplified further, as in part b of problem 2 below. But this version is easy enough to calculate. The below `R` code computes the likelihood-ratio statistic, substituting an estimate of σ^2 (the one given by the estimator in equation 9.14b) for σ^2 itself.

```
lr.stat.slr <- function(x, y){
  n <- length(x)
  #compute MLEs of beta and alpha
  B.hat <- (sum(x*y)-sum(x)*sum(y)/n)/( sum(x^2) - sum(x)^2/n)
  A.hat <- (sum(y) - B.hat*sum(x))/n
  #Compute estimated variance of MLE of beta
  vhat <- sum((y - A.hat - B.hat*x)^2)/(n-2)
  #likelihood-ratio statistic
  lr <- (sum((y - mean(y))^2) - sum((y - A.hat - B.hat*x)^2))
/vhat
  return(lr)
}
```

Applying the function to the agricultural data in the running example gives $\Lambda^* \approx 17.99$. (We add the asterisk to indicate that this value of Λ has been calculated using an estimate of σ^2.) Because we held exactly one parameter constant—namely, β—we compare Λ^* to a $\chi^2(1)$ distribution. The p value of 0.00002 is found using `pchisq(17.99, 1)` in `R`.

c) The p values from part (b) and from problem 1 of Exercise Set 9-5 are identical. Moreover, the test statistic from part (b) is the square of the test statistic from problem 1 of Exercise Set 9-5. This relationship explains the identity of the p values. The test statistic in part (b) is compared to a $\chi^2(1)$ distribution, which is the distribution of the square of a single draw from the standard normal distribution, and we compare the test statistic from problem 1 of Exercise Set 9-5 to a standard normal distribution. The agreement between the two tests is not a coincidence—in the simple linear regression case, the Wald test and the likelihood-ratio test are equivalent, even for small sample sizes.

2) a) Starting from the hint, we write

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}\left(\left[Y_i - \hat{Y}_i\right] + [\hat{Y}_i - \bar{Y}]\right)^2.$$

Expanding the squared term on the right gives

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + 2\sum_{i=1}^{n}(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}).$$

Notice that this statement implies that the claim we want to show is true if and only if

$$2\sum_{i=1}^{n}(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) = 0.$$

We now set out to prove that, in fact, $2\sum_{i=1}^{n}(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) = 0$. Start by replacing $\hat{Y}_i$ with its definition, $\hat{Y}_i = \hat{\alpha} + \hat{\beta}x_i$, giving

$$2\sum_{i=1}^{n}(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) = 2\sum_{i=1}^{n}(Y_i - \hat{\alpha} - \hat{\beta}x_i)(\hat{\alpha} + \hat{\beta}x_i - \bar{Y})$$

Now replace $\hat{\alpha}$ with its definition, $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x}$, giving

$$2\sum_{i=1}^{n}(Y_i - [\bar{Y} - \hat{\beta}\bar{x}] - \hat{\beta}x_i)([\bar{Y} - \hat{\beta}\bar{x}] + \hat{\beta}x_i - \bar{Y}).$$

Cancelling and grouping terms as appropriate gives

$$2\sum_{i=1}^{n}(Y_i - \bar{Y} - \hat{\beta}[x_i - \bar{x}])(\hat{\beta}[x_i - \bar{x}])$$

Split this into two sums by distributing the term on the right to give

$$2\hat{\beta}\sum_{i=1}^{n}(Y_i - \bar{Y})(x_i - \bar{x}) - 2\hat{\beta}^2\sum_{i=1}^{n}(x_i - \bar{x})^2$$

We're getting close. Now replace $\hat{\beta}$ with its definition, $\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(Y_i-\bar{Y})}{\sum_{i=1}^{n}[(x_i-\bar{x})^2]}$, to give

$$2\frac{\sum_{i=1}^{n}(x_i-\bar{x})(Y_i-\bar{Y})}{\sum_{i=1}^{n}[(x_i-\bar{x})^2]}\sum_{i=1}^{n}(Y_i-\bar{Y})(x_i-\bar{x}) - 2\left[\frac{\sum_{i=1}^{n}(x_i-\bar{x})(Y_i-\bar{Y})}{\sum_{i=1}^{n}[(x_i-\bar{x})^2]}\right]^2\sum_{i=1}^{n}(x_i-\bar{x})^2$$

We're home. Carrying through the multiplications gives

$$2\frac{[\sum_{i=1}^{n}(x_i-\bar{x})(Y_i-\bar{Y})]^2}{\sum_{i=1}^{n}[(x_i-\bar{x})^2]}(1-1) = 0.$$

Thus, by showing that

$$2\sum_{i=1}^{n}(Y_i-\hat{Y}_i)(\hat{Y}_i-\bar{Y}) = 0,$$

We have also proven the claim we set out to show,

$$\sum_{i=1}^{n}(Y_i-\bar{Y})^2 = \sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2 + \sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2.$$

b) Start by writing Λ for simple linear regression. We are considering Λ as a random variable, so we use capital Y,

$$\Lambda = 2\lambda(\beta) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i-\alpha_0-\beta_0 x_i)^2 - \left(Y_i-\hat{\alpha}-\hat{\beta}x_i\right)^2$$

Under the null hypothesis given in the problem, $\beta_0 = 0$ and $\alpha_0 = \bar{Y}$ (see exercise 1, part b). So the statistic becomes

$$\Lambda = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i-\bar{Y})^2 - \left(Y_i-\hat{\alpha}-\hat{\beta}x_i\right)^2.$$

Now notice that the terms in this expression appear in the statement you proved in part (a), the ANOVA identity. Specifically, $\hat{Y}_i = \hat{\alpha} + \hat{\beta}x_i$, and so by rearranging the ANOVA identity,

$$\Lambda = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i-\bar{Y})^2 - \left(Y_i-\hat{Y}_i\right)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(\hat{Y}_i-\bar{Y}\right)^2.$$

Now switch $\hat{Y}_i$ back to its definition, $\hat{Y}_i = \hat{\alpha} + \hat{\beta}x_i$, and recall that $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x}$ to obtain

$$\Lambda = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(\hat{\alpha} + \hat{\beta}x_i - \bar{Y}\right)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(\bar{Y} - \hat{\beta}\bar{x} + \hat{\beta}x_i - \bar{Y}\right)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(\hat{\beta}[x_i - \bar{x}]\right)^2.$$

Pulling $\hat{\beta}^2$ outside the sum gives

$$\Lambda = \frac{\hat{\beta}^2}{\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

We already know that if $\beta = 0$, then $\hat{\beta} \sim \text{Normal}(0, \sigma^2/\sum_{i=1}^{n}(x_i - \bar{x})^2)$ (see equation 9.13). Thus,

$$\frac{\hat{\beta}\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}{\sigma} \sim \text{Normal}(0,1),$$

and the square of this quantity—in other words, Λ—therefore has a $\chi^2(1)$ distribution.