**University of California, Los Angeles**
**Department of Statistics**

**Statistics 100C** **Instructor: Nicolas Christou**

## Simple regression

1. Consider the following: Let $Y_1, \ldots, Y_n$ be i.i.d. random variables with $Y_i \sim N(\mu, \sigma)$.

   - Confirm that the statement above can be expressed using the following model equation:

     $$Y_i = \mu + \epsilon_i, \quad i = 1, \ldots, n$$

     Write a statement about $\epsilon_1, \ldots, \epsilon_n$.

   - Estimate $\mu$ and $\sigma^2$ using (a) the method of maximum likelihood and (b) the method of least squares.

   - Construct a $1 - \alpha$ confidence interval for $\mu$.

- Test the hypothesis $H_0 : \mu = \mu_0$ against $H_a : \mu \neq \mu_0$.

- Perform power analysis when $\mu = \mu_a$.

- Predict a new $Y$, say $Y_0$.

- Construct a prediction interval for $Y_0$.

2. Now consider the following two models:
$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \ldots, n$
$Y_i = \beta_1 x_i + \epsilon_i, i = 1, \ldots, n$
For both models assume that $x_1, \ldots, x_n$ are known and that $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. random variables with $E[\epsilon_i] = 0$ and $\text{var}[\epsilon_i] = \sigma^2$. These are called the Gauss-Markov conditions. Note: Later we will also assume that $\epsilon_i \sim N(0, \sigma)$.

We want to answer the same questions as in (1) above when we have used the model $Y_i = \mu + \epsilon_i$. The mean for the models in (2) depends on $x$, while the model in (1) has constant mean.

- Estimate $\beta_0, \beta_1, \sigma^2$.
- Construct a $1 - \alpha$ confidence interval for $\beta_0$ and $\beta_1$.
- Test hypotheses on $\beta_0, \beta_1$ or a linear combination of $\beta_0$ and $\beta_1$.
- Predict a new $Y$, say $Y_0$.
- Construct a confidence interval for $E[Y_0]$ for a given $x = x_0$.
- Construct a prediction interval for $Y_0$ for a given $x = x_0$.

Use the model $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \ldots, n$.
Estimate $\beta_0, \beta_1$.
Method of least squares:

$$\begin{aligned}
\min \quad Q &= \sum_{i=1}^{n} \epsilon_i^2 \\
\min \quad Q &= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2
\end{aligned}$$

Take derivative w.r.t. $\beta_0$ and $\beta_1$ set them equal to zero and solve:

$$\begin{aligned}
\frac{\partial Q}{\partial \beta_0} &= \\
\frac{\partial Q}{\partial \beta_1} &=
\end{aligned}$$

Rearrange the previous two equations so that (please complete the left side of the next two equations):

$$= \sum_{i=1}^{n} y_i \qquad (1)$$

$$= \sum_{i=1}^{n} x_i y_i \qquad (2)$$

These are called the "normal equations".
Now, divide (1) by $n$ and solve for $\hat{\beta}_0$ to get $\hat{\beta}_0 = \frac{\sum_{i=1}^{n} y_i}{n} - \hat{\beta}_1 \frac{\sum_{i=1}^{n} x_i}{n} = \bar{y} - \hat{\beta}_1 \bar{x}$. Finally, substitute $\hat{\beta}_0$ in (2) to get:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \\
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad \text{Why?}
\end{aligned}$$

Useful results:

$$\begin{aligned}
\sum_{i=1}^{n}(x_i-\bar{x}) &= 0.\\
\sum_{i=1}^{n}(x_i-\bar{x})^2 &= \sum_{i=1}^{n}x_i^2-\frac{(\sum_{i=1}^{n}x_i)^2}{n}=\sum_{i=1}^{n}x_i^2-n\bar{x}^2.\\
\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) &= \sum_{i=1}^{n}x_iy_i-\frac{1}{n}(\sum_{i=1}^{n}x_i)(\sum_{i=1}^{n}y_i)=\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}.\\
\sum_{i=1}^{n}(x_i-\bar{x})x_i &= \sum_{i=1}^{n}(x_i-\bar{x})^2.
\end{aligned}$$

Estimate $\beta_0, \beta_1, \sigma^2$.
Method of maximum likelihood:
$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \ldots, n$. Assume that $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. random variables with $\epsilon_i \sim N(0, \sigma)$.

* What is the distribution of $Y_i$?

* Write the pdf of $Y_i$.

* Write the likelihood function based on $Y_1, \ldots, Y_n$.

* Write the log-likelihood function.

* Maximize the log-likelihood function w.r.t. $\beta_0, \beta_1$ and $\sigma^2$. We observe that the estimators of $\beta_0$ and $\beta_1$ are the same with the least squares estimators (why?).

* Fitted regression equation, fitted values, and residuals:

$$\begin{aligned}
\hat{Y}_i &= \hat{\beta}_0+\hat{\beta}_1x_i\\
\hat{Y}_i &= \bar{Y}+\hat{\beta}_1(x_i-\bar{x}) \quad \text{why?}\\
e_i &= Y_i-\hat{Y}_i=Y_i-\bar{Y}-\hat{\beta}_1(x_i-\bar{x})
\end{aligned}$$

Properties of least squares for the model $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \ldots, n$.

$$\begin{aligned}
\sum_{i=1}^{n}e_i &= 0.\\
\sum_{i=1}^{n}e_ix_i &= 0.\\
\sum_{i=1}^{n}e_i\hat{y}_i &= 0.
\end{aligned}$$

Properties of least squares estimators:
Find the expected value and variance of $\hat{\beta}_1$ and $\hat{\beta}_0$.

It will be helpful if we expressed $\hat{\beta}_1$ and $\hat{\beta}_0$ as linear combinations of $Y_1, \ldots, Y_n$.

Recall that $\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$. Verify that $\hat{\beta}_1 = \sum_{i=1}^n k_i y_i$. What is $k_i$? Is it random or non random?

To find $E[\hat{\beta}_1]$ and var$[\hat{\beta}_1]$ it will be useful to find the following:
$\sum_{i=1}^n k_i =$

$\sum_{i=1}^n k_i x_i =$

$\sum_{i=1}^n k_i^2 =$

Now let's find $E[\hat{\beta}_1]$ and var$[\hat{\beta}_1]$. Remember that $E[Y_i] = \beta_0 + \beta_1 x_i$.
$E[\hat{\beta}_1] = E[\sum_{i=1}^n k_i y_i] =$

and

$\text{var}[\hat{\beta}_1] = \text{var}[\sum_{i=1}^n k_i y_i] =$
Note: $Y_1, \ldots, Y_n$ are independent with $\text{var}[Y_i] = \sigma^2$ and $k_1, \ldots, k_n$ are non random quantities.

Now let's do the same for $\hat{\beta}_0$. Recall that $\hat{\beta}_0 = \frac{\sum_{i=1}^n y_i}{n} - \hat{\beta}_1 \frac{\sum_{i=1}^n x_i}{n} = \bar{y} - \hat{\beta}_1 \bar{x}$. Verify that $\hat{\beta}_0 = \sum_{i=1}^n l_i y_i$. What is $l_i$? Is it random or non random?

To find $E[\hat{\beta}_0]$ and var$[\hat{\beta}_0]$ it will be useful to find the following:
$\sum_{i=1}^n l_i =$

$\sum_{i=1}^n l_i x_i =$

$\sum_{i=1}^n l_i^2 =$

Now let's find $E[\hat{\beta}_0]$ and var$[\hat{\beta}_0]$:
$E[\hat{\beta}_0] = E[\sum_{i=1}^n l_i y_i] =$

and

var$[\hat{\beta}_0] =$ var$[\sum_{i=1}^n l_i y_i] =$

Gauss-Markov theorem:
Show that $\hat{\beta}_1$ and $\hat{\beta}_0$ are BLUE (Best Linear Unbiased Estimators). We have seen that $\hat{\beta}_1 = \sum_{i=1}^n k_i y_i$ and that $E[\hat{\beta}_1] = \beta_1$ and $\text{var}[\hat{\beta}_1] = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$. Now let $b_1 = \sum_{i=1}^n a_i y_i$ be another unbiased estimator of $\beta_1$. Show that $\text{var}[\hat{\beta}_1] \leq \text{var}[b_1]$.

Proof:
For $b_1$ to be unbiased for all $\beta_0$ and $\beta_1$ we must have the following two conditions:
$\sum_{i=1}^n a_i =$

$\sum_{i=1}^n a_i x_i =$

It is also true that $a_i = k_i + d_i$, where $d_i$ is the difference between $k_i$ and $a_i$. Now let's find the variance of $b_1$:

$$\begin{aligned} \text{var}[b_1] &= \text{var}[\sum_{i=1}^n a_i y_i] \\ &= \sigma^2 \sum_{i=1}^n a_i^2 \text{ (why?) Replace } a_i = k_i + d_i \text{ and expand.} \end{aligned}$$

Similarly, by following the same steps as above, we can show that $\text{var}[\hat{\beta}_0] \leq \text{var}[b_0]$, where $b_0 = \sum_{i=1}^n c_i y_i$ is another linear unbiased estimator of $\beta_0$.

Back to the estimation of $\sigma^2$. We have seen that the MLE of $\sigma^2$ is $\hat{\sigma^2} = \frac{\sum_{i=1}^n e_i^2}{n}$. Is $\hat{\sigma^2}$ an unbiased estimator of $\sigma^2$?

$$\begin{aligned} E[\hat{\sigma^2}] = E\left[\frac{\sum_{i=1}^n e_i^2}{n}\right] &= \frac{1}{n}\sum_{i=1}^n E[e_i^2] \\ &= \frac{1}{n}\sum_{i=1}^n \left[\text{var}[e_i] + (E[e_i])^2\right] \qquad \text{Note: } E[e_i] = 0 \text{(why?)} \\ &= \frac{1}{n}\sum_{i=1}^n \text{var}[e_i] \end{aligned}$$

Therefore, we need to find var$[e_i]$. Recall that $e_i = Y_i - \hat{Y}_i = Y_i - \bar{Y} - \hat{\beta}_1(x_i - \bar{x})$.

Aside note: Let $Y_1, Y_2, Y_3$ be random variables and $a_1, a_2, a_3$ be known constants.

$$\begin{aligned} \text{var}[a_1Y_1 + a_2Y_2 + a_3Y_3] &= a_1^2\text{var}(Y_1) + a_2^2\text{var}(Y_2) + a_3^2\text{var}(Y_3) \\ &+ 2a_1a_2\text{cov}(Y_1, Y_2) + 2a_1a_3\text{cov}(Y_1, Y_3) + 2a_2a_3\text{cov}(Y_2, Y_3). \end{aligned}$$

Therefore

$$\begin{aligned} \text{var}[e_i] &= \text{var}\left[Y_i - \bar{Y} - \hat{\beta}_1(x_i - \bar{x})\right] \qquad \text{Please write the 6 terms that we need to consider} \\ &= \\ &+ \end{aligned}$$

Aside:

$$\begin{aligned} \text{var}[Y_i] &= \\ \text{var}[\bar{Y}] &= \\ \text{var}[\hat{\beta}_1] &= \\ \text{cov}[Y_i, \bar{Y}] &= \\ \text{cov}[Y_i, \hat{\beta}_1] &= \\ \text{cov}[\bar{Y}, \hat{\beta}_1] &= \end{aligned}$$

Now back to var$[e_i]$ to complete the question.

Distribution theory:
We have estimated the three parameters $\beta_0, \beta_1$ and $\sigma^2$ of the model $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \ldots, n$ with $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\sigma^2}$.

Under the normality assumption and because $\hat{\beta}_0, \hat{\beta}_1$ are linear combinations of independent $Y_1, \ldots, Y_n$ it follows that

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}\right)$$

$$\hat{\beta}_0 \sim N\left(\beta_0, \sigma\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}\right)$$

Theorem:
Let $S_e^2 = \frac{\sum_{i=1}^n e_i^2}{n-2}$. Then $\frac{(n-2)S_e^2}{\sigma^2} = \frac{\sum_{i=1}^n e_i^2}{\sigma^2} \sim \chi^2_{n-2}$.

Proof:
First we discuss the "centered" model.
We obtain the "centered model" from the model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ if we add and subtract on the right hand side the quantity $\beta_1 \bar{x}$ as follows:

$$\begin{aligned}
y_i &= \beta_0 + \beta_1 x_i + \epsilon_i + \beta_1 \bar{x} - \beta_1 \bar{x} \\
y_i &= \beta_0 + \beta_1 \bar{x} + \beta_1 (x_i - \bar{x}) + \epsilon_i \\
y_i &= \gamma_0 + \beta_1 (x_i - \bar{x}) + \epsilon_i \\
y_i &= \gamma_0 + \beta_1 z_i + \epsilon_i, \text{ where, } \gamma_0 = \beta_0 + \beta_1 \bar{x} \text{ and } z_i = x_i - \bar{x}.
\end{aligned}$$

The model $y_i = \gamma_0 + \beta_1 (x_i - \bar{x}) + \epsilon_i$ is the centered model. This model has the same functional form as the model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ and therefore the estimates of $\gamma_0$ and $\beta_1$ using least squares computed as follows:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (z_i - \bar{z}) y_i}{\sum_{i=1}^n (z_i - \bar{z})^2}, \text{ but, } \bar{z} = 0 \text{ (why?)}$$

Continue to show that it is the same estimator using the non-centered model

Also, verify that $\hat{\gamma}_0 = \bar{Y}$.

Key result of the centered model: The two estimators are $\hat{\beta}_1$ and $\bar{Y}$ and because $\text{cov}(\bar{Y}, \hat{\beta}_1) = 0$ under normality they are independent.

Under normality and using the centered model we obtain the following distributions:

$$y_i \sim$$

$$\hat{\beta}_1 \sim$$

$$\hat{\gamma}_0 \sim$$

What is the distribution of $\frac{\sum_{i=1}^n [y_i - \gamma_0 - \beta_1 (x_i - \bar{x})]^2}{\sigma^2}$?

Now begin with $\sum_{i=1}^{n}[y_i - \gamma_0 - \beta_1(x_i - \bar{x})]^2$ and add/subtract $\hat{\gamma}_0$ and $\hat{\beta}_1(x_i - \bar{x})$ and expand.

$$\sum_{i=1}^{n}[y_i - \gamma_0 - \beta_1(x_i - \bar{x}) \pm \hat{\gamma}_0 \pm \hat{\beta}_1(x_i - \bar{x})]^2 = \sum_{i=1}^{n}[e_i + (\hat{\gamma}_0 - \gamma_0) + (\hat{\beta}_1 - \beta_1)(x_i - \bar{x})]^2$$

Expand to get 6 terms.

Three of the terms are equal to zero.

Now divide both sides by $\sigma^2$ and using the distributions from page 9 we see that all the previous terms follow $\chi^2$ distributions.

So far we have:

$$\begin{array}{ccccccc} \sum_{i=1}^{n}\left[\frac{y_i-\gamma_0-\beta_1(x_i-\bar{x})}{\sigma}\right]^2 & = & \frac{(n-2)S_e^2}{\sigma^2} & + & \left(\frac{\hat{\gamma}_0-\gamma}{\frac{\sigma}{\sqrt{n}}}\right)^2 & + & \left(\frac{\hat{\beta}_1-\beta_1}{\frac{\sigma}{\sqrt{\sum_{i=1}^{n}(x_i-x)^2}}}\right)^2 \\ Q & = & Q_1 & + & Q_2 & + & Q_3 \end{array}$$

Distributions:

$$\begin{array}{ll} Q \sim & \text{Moment generating function } M_Q(t) = \\ Q_2 \sim & \text{Moment generating function } M_{Q2}(t) = \\ Q_3 \sim & \text{Moment generating function } M_{Q3}(t) = \end{array}$$

It is given that $Q_1, Q_2, Q_3$ are independent (we will show this next). Find the moment generating function of $Q$.

$$M_Q(t) = M_{Q_1+Q_2+Q_3}(t)$$

$M_Q(t) =$ (Use properties of moment generating functions.)

Solve for $M_{Q1}(t)$

$$M_{Q1}(t) =$$

We mentioned above that $Q_1, Q_2, Q_3$ are independent. Show that the covariances are zero and therefore under normality they are independent:

1. We have seen already that $\text{cov}(\bar{Y}, \hat{\beta}_1) = 0$.
2. Show that the covariance between $\bar{Y}$ and each residual is zero, $\text{cov}(\bar{Y}, e_i) = 0$.
   Recall that $e_i = Y_i - \hat{Y}_i = Y_i - \bar{Y} - \hat{\beta}_1(x_i - \bar{x})$. Therefore,

   $$\text{cov}\left(\bar{Y}, Y_i - \bar{Y} - \hat{\beta}_1(x_i - \bar{x})\right) =$$

3. Show that the covariance between $\hat{\beta}_1$ and each residual is zero, $\text{cov}(e_i, \hat{\beta}_1) = 0$. Therefore,

   $$\text{cov}\left(Y_i - \bar{Y} - \hat{\beta}_1(x_i - \bar{x}), \hat{\beta}_1\right) =$$

Variance and covariance operations in simple regression using summations.

A general result: The covariance operation is additive.

$$\operatorname{cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \operatorname{cov}(X_i, Y_j).$$

Proof:
Let $E(X_i) = \mu_i$ and $E(Y_j) = v_j$. Then

$$E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mu_i \text{ and } E(\sum_{j=1}^{m} Y_j) = \sum_{j=1}^{m} v_j.$$

Therefore using the definition of covariance,

$$\begin{aligned}
\operatorname{cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) &= E\left[\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i\right)\left(\sum_{j=1}^{m} Y_j - \sum_{j=1}^{m} v_j\right)\right] \\
&= E\left[\sum_{i=1}^{n}(X_i - \mu_i)\sum_{j=1}^{m}(Y_j - v_j)\right] = E\left[\sum_{i=1}^{n}\sum_{j=1}^{m}(X_i - \mu_i)(Y_j - v_j)\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} E(X_i - \mu_i)(Y_j - v_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} \operatorname{cov}(X_i, Y_j).
\end{aligned}$$

Similarly, find the covariance between $\sum_{i=1}^{n} a_i Y_i$ and $\sum_{j=1}^{n} b_j Y_j$:

$$\operatorname{cov}\left(\sum_{i=1}^{n} a_i Y_i, \sum_{j=1}^{n} b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j \operatorname{cov}(Y_i, Y_j).$$

**Result**
If $Y_1, \ldots, Y_n$ are independent (e.g. one of the Gauss-Markov conditions in regression) then, when $i \neq j$ we have $\operatorname{cov}(Y_i, Y_j) = 0$ and therefore:

$$\begin{aligned}
\operatorname{cov}\left(\sum_{i=1}^{n} a_i Y_i, \sum_{j=1}^{n} b_j Y_j\right) &= a_1 b_1 \operatorname{cov}(Y_1, Y_1) + a_1 b_2 \operatorname{cov}(Y_1, Y_2) + \ldots + a_1 b_n \operatorname{cov}(Y_1, Y_n) \\
&+ a_2 b_1 \operatorname{cov}(Y_2, Y_1) + a_2 b_2 \operatorname{cov}(Y_2, Y_2) + \ldots + a_2 b_n \operatorname{cov}(Y_2, Y_n) \\
&+ \\
&\vdots \\
&+ \\
&+ a_n b_1 \operatorname{cov}(Y_n, Y_1) + a_n b_2 \operatorname{cov}(Y_n, Y_2) + \ldots + a_n b_n \operatorname{cov}(Y_n, Y_n) \\
&= \sum_{i=1}^{n} a_i b_i \operatorname{var}(Y_i).
\end{aligned}$$

(Because when $i = j$ we have $cov(Y_i, Y_i) = var(Y_i)$).

Application in simple regression:
Let $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$. The Gauss-Markov conditions hold. Use the previous result to find $\text{cov}(\bar{Y}, \hat{\beta}_1)$, $\text{cov}(\hat{\beta}_0, \hat{\beta}_1)$. and $\text{cov}(\hat{Y}_i, \hat{Y}_j)$. Hint: Express $\bar{Y}, \hat{\beta}_0, \hat{\beta}_1, \hat{Y}_i$ as linear combinations of $Y_1, \dots, Y_n$.

- $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i = \sum_{i=1}^n a_i Y_i$, where $a_i = \frac{1}{n}$.

- $\hat{\beta}_1 = \sum_{i=1}^n k_i Y_i$. What is $k_i$?

- $\hat{\beta}_0 = \sum_{i=1}^n l_i Y_i$. What is $l_i$?

- $\hat{Y}_i = \sum_{l=1}^n c_l Y_l$ and $\hat{Y}_j = \sum_{r=1}^n d_r Y_r$. What are $c_l, d_r$?

Therefore,

1. $\text{cov}(\bar{Y}, \hat{\beta}_1) = \text{cov}(\sum_{i=1}^n a_i Y_i,\ \sum_{j=1}^n k_j Y_j)=$

2. $\text{cov}(\hat{\beta}_0, \hat{\beta}_1) = \text{cov}(\sum_{i=1}^n l_i Y_i,\ \sum_{j=1}^n k_j Y_j)=$

3. $\text{cov}(\hat{Y}_i, \hat{Y}_j) = \text{cov}(\sum_{i=1}^n c_l Y_l,\ \sum_{j=1}^n d_r Y_r)=$

Variance of a linear combination of random variables:
Using the result above find $\text{var}(Y_1 + \ldots + Y_n)$:

$$\begin{aligned}
\text{var}\left(\sum_{i=1}^{n} Y_i\right) &= \text{cov}\left(\sum_{i=1}^{n} Y_i, \sum_{j=1}^{n} Y_j\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \text{cov}(Y_i, Y_j) \\
&= \sum_{i=1}^{n} \text{var}(Y_i) + \sum_{i=1}^{n}\sum_{j \neq i}^{n} \text{cov}(Y_i, Y_j) \\
&= \sum_{i=1}^{n} \text{var}(Y_i) + 2\sum_{i=1}^{n-1}\sum_{j>i}^{n} \text{cov}(Y_i, Y_j)
\end{aligned}$$

The previous result can be extended to the more general case:

Let $Y_1, Y_2, \ldots, Y_n$ be random variables, and $a_1, a_2, \ldots, a_n$ be constants. Find the variance of the linear combination $Q = a_1Y_1 + a_2Y_2 + \ldots + a_nY_n$.

$$\begin{aligned}
\text{var}\left(\sum_{i=1}^{n} a_iY_i\right) &= \text{cov}\left(\sum_{i=1}^{n} a_iY_i, \sum_{j=1}^{n} a_jY_j\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_ia_j\text{cov}(Y_i, Y_j) \\
&= \sum_{i=1}^{n} a_i^2\text{var}(Y_i) + \sum_{i=1}^{n}\sum_{j \neq i}^{n} a_ia_j\text{cov}(Y_i, Y_j) \\
&= \sum_{i=1}^{n} a_i^2\text{var}(Y_i) + 2\sum_{i=1}^{n-1}\sum_{j>i}^{n} a_ia_j\text{cov}(Y_i, Y_j)
\end{aligned}$$

Application in simple regression:
When the Gauss-Markov condition of independence holds, the previous expression is simplified considerably, because $\text{cov}(Y_i, Y_j) = 0$. Let $Y_i = \beta_0 + \beta_1x_i + \epsilon_i$.

Use the previous result to find $\text{var}(\hat{Y}_i)$. First, express $\hat{Y}_i$ as a linear combination of $Y_1, \ldots, Y_n$.

Find the variance of $\text{var}(e_i)$ by expressing $e_i = \sum_{l=1}^{n} a_lY_l$

Show that $\sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$.

Note:

$$\sum_{i=1}^{n}(y_i-\bar{y})^2 \qquad \text{Total sum of squares } (SST)$$
$$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \qquad \text{Error sum of squares } (SSE)$$
$$\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 \qquad \text{Regression sum of squares } (SSR)$$

Proof:
Begin with $\sum_{i=1}^{n}(y_i-\bar{y})^2$. What do you need to add/subtract to get $e_i$? And then expand.

Distribution of $S_e^2$
Use $\frac{(n-2)S_e^2}{\sigma^2} \sim \chi^2_{n-2}$ to find the distribution of $S_e^2$.

Aside notes:
If $Y \sim \Gamma(\alpha,\beta)$ then the moment generating function of $Y$ is $M_Y(t) = (1-\beta t)^{-\alpha}$ and one of the properties of moment generating functions is $M_{cY} = M_Y(ct)$ ($c$ is a constant).

Let $Q = \frac{(n-2)S_e^2}{\sigma^2} \sim \chi^2_{n-2}$. Solve for $S_e^2$ and find the moment generating function of $S_e^2$ using the notes above.

Construct a $1-\alpha$ Confidence interval for $\beta_1, \beta_0$ and for a linear combination of $\beta_1$ and $\beta_0$, for example, $\beta_0 - 2\beta_1$. We want a statement like this $P[L < \theta < U] = 1-\alpha$.

Confidence interval for $\beta_1$.
We have seen that $\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}})$. Suppose $\sigma$ is known. Standardized $\hat{\beta}_1$ to get a "pivotal" expression and then use it to construct a $1-\alpha$ confidence interval for $\beta_1$.

Repeat for a $1-\alpha$ confidence interval for $\beta_0$ assuming $\sigma$ is known.

However $\sigma$ is not known. Therefore, use $\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}})$ and $\frac{(n-2)S_e^2}{\sigma^2} \sim \chi^2_{n-2}$ to construct a $t$ ratio ($\sigma$ now is eliminated.) What are the degrees of freedom?

Use the previous $t$ ratio as a pivotal expression to construct a $1-\alpha$ confidence interval for $\beta_1$.

Repeat for a $1-\alpha$ confidence interval for $\beta_0$.

Construct a confidence interval for $\beta_0 - 2\beta_1$.
What is the estimator of $\beta_0 - 2\beta_1$?

Find $E[\hat{\beta}_0 - 2\hat{\beta}_1]$.

Find $\text{var}[\hat{\beta}_0 - 2\hat{\beta}_1]$.

What is the distribution of $\hat{\beta}_0 - 2\hat{\beta}_1$?

Construct a $t$ ratio using the distribution of $\hat{\beta}_0 - 2\hat{\beta}_1$ and $\frac{(n-2)S_e^2}{\sigma^2} \sim \chi^2_{n-2}$.

Finally, use the previous $t$ distribution to construct a $1 - \alpha$ confidence interval for $\beta_0 - 2\beta_1$.

Suppose now we want to predict $Y_0$ (a new value of $y$) when $x = x_0$ and construct a prediction interval.

What is the expression for the error of prediction?

Find the expected value of the error of prediction.

Find the variance of the error of prediction.

What is the distribution of the error of prediction?

Construct a $t$ ratio using the distribution of the error of prediction and $\frac{(n-2)S_e^2}{\sigma^2} \sim \chi^2_{n-2}$.

Finally, use the previous $t$ distribution to construct a $1 - \alpha$ prediction interval for $Y_0$.

Hypothesis testing
Consider the following tests.

1. $H_0 : \beta_1 = 0$
   $H_a : \beta_1 \neq 0$

2. $H_0 : \beta_0 = 0$
   $H_a : \beta_0 \neq 0$

3. $H_0 : \beta_1 - 2\beta_0 = 0$
   $H_a : \beta_1 - 2\beta_0 \neq 0$

4. $H_0 : \beta_1 = \beta_1^*, \beta_0 = \beta_0^*$
   $H_a :$ Not true.

Test for linear association between $Y$ and $x$:
$H_0 : \beta_1 = 0$
$H_a : \beta_1 \neq 0$
Construct a $t$ ratio with $n-2$ degrees of freedom. This is the same $t$ ratio we used for a $1-\alpha$ confidence interval for $\beta_1$. Here, under $H_0 : \beta_1 = 0$. Therefore, $t = \frac{\hat{\beta}_1}{\frac{S_e}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}} = \frac{\hat{\beta}_1 \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}{S_e} \sim t_{n-2}$. Reject $H_0$ if $t > t_{1-\frac{\alpha}{2};n-2}$ or $t < -t_{\frac{\alpha}{2};n-2}$.

Can we test this hypothesis using the $F$ statistic?
To form a ratio that follows the $F$ distribution we need two independent $\chi^2$ random variables. Here they are:
From $\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}})$ find a $\chi^2_1$.

And together with $\frac{(n-2)S_e^2}{\sigma^2} \sim \chi^2_{n-2}$ find a test statistic that follows $F_{1,n-2}$.

Is there a connection between the $t$ and the $F$ statistics?

Likelihood ratio test

A general note: Reject $H_0$ is $\Lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} < k$, where $L(\hat{\omega})$ is the maximized likelihood function under $H_0$ and $L(\hat{\Omega})$ is the maximized likelihood function under no restrictions.

We are testing $H_0 : \beta_1 = 0$. Therefore under $H_0$ the model is:

Find the MLEs of the parameters of the model under $H_0$.

Under no restriction we have the MLEs $\hat{\beta}_0$ and $\hat{\beta}_1$ and $\hat{\sigma^2} = \frac{\sum_{i=1}^{n} e_i^2}{n}$.

Begin the likelihood ratio test $\Lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} < k$ and show that it is equivalent to the $F$ statistic above.

This class activity will demonstrate step-by-step the procedure for finding a test statistic using the $t$, $F$, and likelihood ratio test.

Consider the simple regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$. The Gauss-Markov conditions hold and in addition $\epsilon_i \sim N(0, \sigma)$. We wish to test the hypothesis $H_0 : \beta_1 = 1$ against $H_a : \beta_1 \neq 1$. This test can be performed using the following methods:

A. $t$ statistic.

B. $F$ statistic.

C. Likelihood ratio test.

Let's begin...

A. $t$ statistic:

1. What is the distribution of $\hat{\beta}_1$ under $H_0$.
   $\hat{\beta}_1 \sim N( \quad , \quad )$.

2. What is the practical difficult in using (1) for inference on $\beta_1$?

3. Now obtain a $N(0, 1)$ distribution using (1).

4. Which other distribution do we need to use in order to obtain a $t$ statistic?

5. Construct a $t$ ratio using (3) and (4). What are the degrees of freedom for this $t$ statistic?

B. $F$ statistic:

1. What is the distribution of the square of A-3?

2. Use B-1 and A-4 to construct the $F$ statistic. What are the degrees of freedom?

C. Likelihood ratio test:

A general note: Reject $H_0$ is $\Lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} < k$, where $L(\hat{\omega})$ is the maximized likelihood function under $H_0$ and $L(\hat{\Omega})$ is the maximized likelihood function under no restrictions.

1. Estimate the parameters of the model under $H_0$.

2. Estimate the parameters of the model under no restrictions.

3. Begin the likelihood ratio test $\Lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} < k$ and show that it is equivalent to the $F$ statistic (same as in $B$).

Power analysis
Consider the simple regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, ..., n$. The Gauss-Markov conditions hold and also $\epsilon_i \sim N(0, \sigma)$. We want to test the hypothesis $H_0 : \beta_1 = 0$ against the alternative $H_a : \beta_1 \neq 0$. The test can be performed either using the $t$ statistic or the $F$ statistic. Under $H_0$ these two statistics follow the central $t_{n-2}$ and the central $F_{1,n-2}$.

$$t = \frac{\hat{\beta}_1}{\frac{S_e}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}} \quad \text{or} \quad F = \frac{\hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}{S_e^2}.$$

If $H_0$ is not true then these two statistics follow the non-central $t$ distribution with non-centrality parameter $\frac{\beta_1 \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}{\sigma}$ and the non-central $F$ distribution with non-centrality parameter $\frac{\beta_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2}$. Why?

Let's explore...

Power analysis using the noncentral $t$ distribution.
A note on the noncentral $t$ distribution:
Let $Z \sim N(\delta, 1)$ and independently of $Z$ let $U \sim \chi^2_n$ then $\frac{Z}{\sqrt{\frac{U}{n}}} \sim t_n(NCP = \delta)$. Therefore the idea is that we need to transform the normal random variable so it has standard deviation 1. Apply this to the test $H_0 : \beta_1 = 0$.

Under $H_0$ the numerator follows    If $H_0$ is not true the numerator follows

$$\frac{\frac{\frac{\hat{\beta}_1}{\sigma}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}}{\sqrt{\frac{\frac{(n-2)S_e^2}{\sigma^2}}{n-2}}}$$

The ratio under $H_0$ follows    If $H_0$ is not true the ratio follows

Recall that the power of a test is the following:
$1 - \beta = P$[rejecting $H_0$ when $H_0$ is false]. Therefore,
$1 - \beta = P[t_{n-2;NCP} > t_{\frac{\alpha}{2};n-2}] + P[t_{n-2;NCP} < -t_{\frac{\alpha}{2};n-2}]$.

Note: To compute the noncentrality parameter we need values for $\beta_1$ and $\sigma$.

Power analysis using the noncentral $F$ distribution.
A note on the noncentral $\chi^2$ distribution:
Let $Y \sim N(\mu, 1)$. Then $Y^2 \sim \chi^2_1(NCP = \mu^2)$. Therefore, if $Y \sim N(\mu, \sigma)$ then $\frac{Y}{\sigma} \sim N(\frac{\mu}{\sigma}, 1)$ and $\frac{Y^2}{\sigma^2} \sim \chi^2_1(NCP = (\frac{\mu}{\sigma})^2)$. If $Y_1, \ldots, Y_n$ are i.i.d. $N(\mu, \sigma)$ then $\sum_{i=1}^n \frac{Y_i^2}{\sigma^2} \sim \chi^2_n(NCP = n(\frac{\mu}{\sigma})^2)$.

A note on the noncentral $F$ distribution:
Let $U \sim \chi^2_n(NCP = \theta)$ and independently of $U$ let $V \sim \chi^2_m$. Then $\frac{\frac{U}{n}}{\frac{V}{m}} \sim F_{n,m}(NCP = \theta)$. Apply this to the test $H_0 : \beta_1 = 0$.

$\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}})$. Find a noncentral $\chi^2$ random variable. What is the noncentrality parameter $\theta$?

Now the $F$ statistic:

| | Under $H_0$ the numerator follows | If $H_0$ is not true the numerator follows |
|---|---|---|
| $\frac{\frac{\hat{\beta}_1^2 \sum_{i=1}^n (x_i - x)^2}{\sigma^2}}{\frac{\frac{(n-2)S_e^2}{\sigma^2}}{n-2}}$ | | |
| | The ratio under $H_0$ follows | If $H_0$ is not true the ratio follows |

Example
We will use the following data:

```
#Read the data:
a <- read.table("http://www.stat.ucla.edu/~nchristo/statistics100C/soil.txt",
header=TRUE)
#Use only n=7:
a1 <- a[3:9,]
```

We run the regression `lead` on `zinc`

```
#Run simple regression of lead on zinc:
q <- lm(a1$lead ~ a1$zinc)

#Summary of the regression:
summary(q)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 65.12581    8.31592   7.831 0.000545
a1$zinc      0.20743    0.02166   9.577 0.000210

#Using alpha=0.05 and degrees of freedom 5 we obtain
the critical t value= 2.570582:
qt(0.975, 5)
```

We see that $t = 9.577$. The critical value using $\alpha = 0.05$ is $t_{0.975,5} = 2.570582$, therefore $H_0$ is rejected. We can also use the p-value=0.000210. Since the p-value $< \alpha = 0.05$ we reject $H_0$.

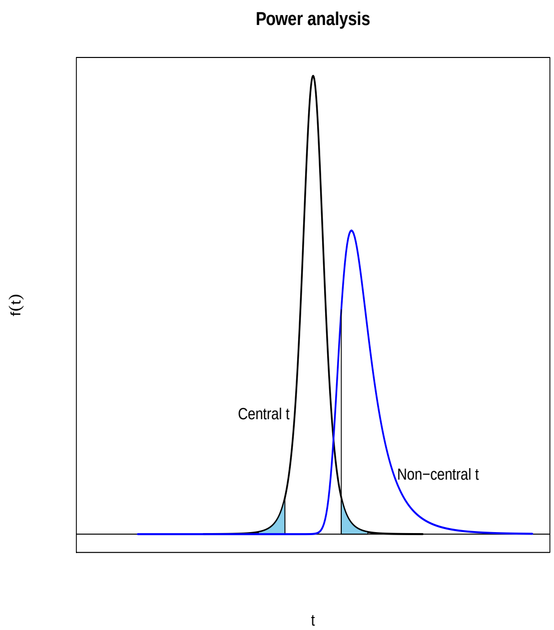
Now suppose we want to compute the power of the test if $\beta_1 = 0.3$ and $\sigma^2 = 625$. We will use either the non-central $t$ or the non-central $F$ distributions.

A. Using the non-central $t$ distribution. The `R` calculations follow:

```
#We need to compute the non-centrality parameter:
beta1 <- 0.3
sigma <- 25
ncp1 <- beta1*sqrt(6*var(a1$zinc)) / sigma

#Compute the power:
pt(2.570582, 5, ncp=ncp1, lower.tail=FALSE) + pt(-2.570582, 5, ncp=ncpar1)

#It is equal to 0.8729043.
```



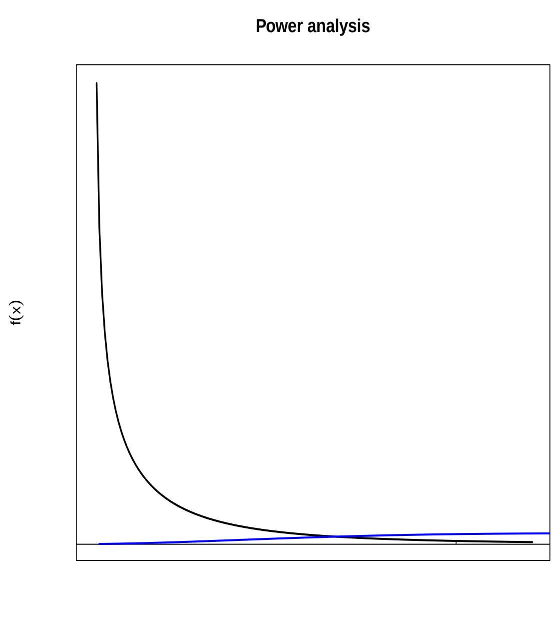

B. Using the non-central $F$ distribution. The `R` calculations follow:

```
#Critical value using the F distribution is equal to 6.607891
obtained as follows:
qf(0.95, 1,5)

#We need the non-centrality parameter for the chi-squared distribution:
ncp2 <- beta1^2*6*var(a1$zinc)/sigma^2

#Compute the power:
1-pf(6.607891, 1,5, ncp=ncp2)

#It is equal to 0.8729043
```



Hypothesis testing using the extra sum of squares principle
Under the null hypothesis we have a constrained (restricted) least squares problem. We estimate the parameters of the model under $H_0$ and we compute the error sum of squares. For example, suppose we are testing $H_0 : \beta_1 = 0$, then under this hypothesis the model is $y_i = \beta_0 + \epsilon_i$. What is the error sum of squares under this hypothesis? It is $SSE_c =$
We then compute the error sum of squares under the full model (no restrictions). The error sum of squares is $SSE_f =$

The $F$ statistic using the extra sum of squares method is defined as

$$\frac{\frac{SSE_c - SSE_f}{df_c - df_f}}{\frac{SSE_f}{df_f}},$$

where $df_c$ and $df_f$ denote the degrees of freedom under the restricted model and under no restrictions respectively.
What is $df_c$?
What is $df_f$?

Therefore,

$$\frac{\frac{SSE_c - SSE_f}{df_c - df_f}}{\frac{SSE_f}{df_f}} = \frac{SSE_c - SSE_f}{S_e^2}.$$

Now substitute $SSE_c$ and $SSE_f$ to show that this is exactly the $F$ statistic that we found earlier, on page 20.

Another example:
Suppose we are testing the hypothesis that the arm span is approximately equal to height, so we are testing $\beta_1 = 1$. Under the null hypothesis we have a restriction which makes the model $y_i = \beta_0 + x_i + \epsilon_i$ or $y_i - x_i = \beta_0 + \epsilon_i$ or $y_i^* = \beta_0 + \epsilon_i$. The estimate of $\beta_0$ using this model is $\hat{\beta}_{0c} = \bar{y}^* = \bar{y} - \bar{x}$. Therefore the error sum of squares of this model is $SSE_c = \sum_{i=1}^{n} (y_i^* - \bar{y}^*)^2$. Show that it is equal to $\sum_{i=1}^{n} e_i^2 + (\hat{\beta}_1 - 1)^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$.

What is $df_c$?
What is $df_f$?

Finally apply the extra sum of squares $F$ statistic

$$\frac{\frac{SSE_c - SSE_f}{df_c - df_f}}{\frac{SSE_f}{df_f}}$$

to show that this is exactly the $F$ statistic that we found earlier, on page 22.

Constrained least squares - Lagrange multiplier method

In the previous section we have seen the $F$ test using the extra sum of squares principle. In doing so we solve for the constraint to obtain the restricted model and from there after estimation we computed the constrained error sum of squares. Another method for computing the constrained error sum of squares is to minimize the error sum of squares subject to the constraint which involves the use of the Lagrange multiplier method. Suppose we are testing $H_0 : \beta_0 + \beta_1 = 1$. The set up of the problem is the following.

$$\begin{aligned} \min Q &= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \\ \text{subjet to} &\quad \beta_0 + \beta_1 = 1 \text{ And using a Lagrange multiplier we minimize} \\ \min Q &= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 - 2\lambda(\beta_1 + \beta_0 - 1) \end{aligned}$$

This is a minimization problem that involves three unknowns, $\beta_0, \beta_1$, and $\lambda$. After taking partial derivatives we get

$$\begin{aligned} \frac{\partial Q}{\partial \beta_0} &= -2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) - 2\lambda = 0 \\ \frac{\partial Q}{\partial \beta_1} &= -2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) x_i - 2\lambda = 0 \\ \frac{\partial Q}{\partial \lambda} &= \beta_0 + \beta_1 - 1 = 0 \end{aligned}$$

The solution is left as an exercise and it is equal to

$$\begin{aligned} \hat{\beta}_{1c} &= \frac{\sum_{i=1}^{n} (x_i - 1)(y_i - 1)}{\sum_{i=1}^{n} (x_i - 1)^2} \\ \hat{\beta}_{0c} &= 1 - \hat{\beta}_{1c} \end{aligned}$$

The error sum of squares in this case is equal to $SSE_c = \sum_{i=1}^{n} (y_i - \hat{\beta}_{0c} - \hat{\beta}_{1c} x_i)^2$. This error sum of squares can also be obtained using the extra sum of squares principle of the previous section. How? We will have to substitute in the modes $\beta_0 = 1 - \beta_1$ to get

$$\begin{aligned} y_i &= 1 - \beta_1 + \beta_1 x_i + \epsilon_i \\ y_i - 1 &= \beta_1 (x_i - 1) + \epsilon_i \end{aligned}$$

And therefore, the estimate of $\beta_1$ is $\hat{\beta}_{1c} = \frac{\sum_{i=1}^{n} (x_i - 1)(y_i - 1)}{\sum_{i=1}^{n} (x_i - 1)^2}$, the one obtained using the Lagrange multiplier method. We should also emphasize that the error sum of squares is given by $SSE_c = \sum_{i=1}^{n} (y_i - 1 - \hat{\beta}_{1c}(x_i - 1))^2$, which is exactly the same as the error sum of squares obtained using the method of Lagrange multiplier. Again, once the error sum of squares for the constrained model is computed we can compute the $F$ statistic using the extra sum of squares principle as was presented in the previous section.

The $F$ statistic for testing more than one restrictions is simple regression.
So far we have discussed tests that involve on parameter (either $\beta_1$ or $\beta_0$) or a linear combination for $\beta_1$ or $\beta_0$. In this section we discuss the case when testing for more than one parameter. Suppose we are testing the following hypothesis:
$H_0 : \beta_0 = \beta_0^*$ and $\beta_1 = \beta_1^*$, (where $\beta_0^*, \beta_1^*$ are given values)
$H_a$ : Not true.
We will test this hypothesis using the likelihood ratio test. We will find first the MLEs under restrictions imposed by the null hypothesis and under no restrictions.

a. Find the maximum likelihood estimate of $\sigma^2$ under $H_0$.
Since $\beta_0$ and $\beta_1$ are specified under the null hypothesis the only parameter that needs estimation is $\sigma^2$. Therefore, $\hat{\sigma_0^2} = \frac{\sum_{i=1}^n (y_i - \beta_0^* - \beta_1^* x_i)^2}{n}$.

b. Find the maximum likelihood estimate of $\sigma^2$ under no restrictions.
The MLEs of $\beta_0, \beta_1$, and $\sigma^2$ are the same as the ones that were previously computed: $\hat{\beta}_1, \hat{\beta}_0, \hat{\sigma^2} = \frac{\sum_{i=1}^n e_i^2}{n}$.

c. We construct now the likelihood ratio test and show that the test statistic follows the $F_{2,n-2}$ distribution.

$$\begin{aligned}
\Lambda &= \frac{L(\hat{\omega})}{L(\hat{\Omega})} < k \\
&= \frac{(2\pi\hat{\sigma}_0^2)^{-\frac{n}{2}} exp(-\frac{1}{2\hat{\sigma}_0^2}\frac{\sum_{i=1}^n (y_i - \beta_0^* - \beta_1^* x_i)^2}{n})}{(2\pi\hat{\sigma}_1^2)^{-\frac{n}{2}} exp(-\frac{1}{2\hat{\sigma}_1^2}\frac{\sum_{i=1}^n e_i^2}{n})} < k \quad \text{after simplification we get} \\
&= \frac{\hat{\sigma_1^2}}{\hat{\sigma_0^2}} < k^{\frac{2}{n}} \\
&= \frac{\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{\sum_{i=1}^n (y_i - \beta_0^* - \beta_1^* x_i)^2} < k^{\frac{2}{n}}
\end{aligned}$$

The denominator, after $\pm\hat{\beta}_0 \pm \hat{\beta}_1 x_i$ and divide by $\sigma^2$ can be expressed as

$$\frac{\sum_{i=1}^n e_i^2}{\sigma^2} + \frac{n(\hat{\beta}_0 - \beta_0^*)^2}{\sigma^2} + \frac{(\hat{\beta}_1 - \beta_1^*)^2 \sum_{i=1}^n x_i^2}{\sigma^2} + \frac{2(\hat{\beta}_0 - \beta_0^*)(\hat{\beta}_1 - \beta_1^*) \sum_{i=1}^n x_i}{\sigma^2},$$

which algebraically is also is equal to

$$\frac{\sum_{i=1}^n e_i^2}{\sigma^2} + \frac{(\hat{\beta}_1 - \beta_1^*)^2}{\text{var}(\hat{\beta}_1)} + \frac{(C - (\beta_0^* + \beta_1^* \bar{x}))^2}{\text{var}(C)},$$

where, $C = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y}$. Each one of the last two terms is $\chi_1^2$ and because of independence it follows that

$$\frac{(\hat{\beta}_1 - \beta_1^*)^2}{\text{var}(\hat{\beta}_1)} + \frac{(C - (\beta_0^* + \beta_1^* \bar{x}))^2}{\text{var}(C)} \sim \chi_2^2.$$

The likelihood ratio test, after dividing by $\sum_{i=1}^n e_i^2$, can be now expressed as

$$\frac{\frac{(\hat{\beta}_1 - \beta_1^*)^2}{\text{var}(\hat{\beta}_1)} + \frac{(C - (\beta_0^* + \beta_1^* \bar{x}))^2}{\text{var}(C)}}{\frac{(n-2)S_e^2}{\sigma^2}} > k^{-\frac{2}{n}} - 1 = k'.$$

The expression on the left follows $F_{2,n-2}$. Finally, to find $k'$ we will choose a significance level $\alpha$ so that

$$P(F_{2,n-2} > k' | H_0) = \alpha.$$

Therefore, the null hypothesis $H_0 : \beta_0 = \beta_0^*, \beta_1 = \beta_1^*$ is rejected when

$$\frac{(\hat{\beta}_1 - \beta_1^*)^2 \sum_{i=1}^n (x_i - \bar{x})^2 + (\bar{y} - (\beta_0^* + \beta_1^* \bar{x}))^2}{2S_e^2} > F_{1-\alpha;2,n-2}.$$

Efficiency of least squares estimators
In the one parameter case we say that an estimator $\hat{\theta}$ is an efficient estimator for $\theta$ if

1. $E[\hat{\theta}] = \theta$.
2. The variance of $\hat{\theta}$ is equal to the Cramér-Rao lower bound for the variance of an unbiased estimator of $\theta$: $\text{var}[\hat{\theta}] = \frac{1}{I_n(\theta)}$, where $I_n(\theta)$ is the information in the sample.

The information in the sample can be computed using $I_n(\theta) = -E\left[\frac{\partial^2 lnL}{\partial \theta^2}\right]$, where $lnL$ is the log likelihood function.

Example:
Let $Y_1, \ldots, Y_n$ be i.i.d. random variables with $Y_i \sim N(\mu, \sigma)$. Show that $\bar{Y}$ is an efficient estimator of $\mu$.

Multi parameter case
Let $\hat{\boldsymbol{\theta}}$ be the estimator of $\boldsymbol{\theta}$ ($p \times 1$ vector). For example, in the model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ we have $\boldsymbol{\theta} = \begin{pmatrix} \\ \\ \\ \end{pmatrix}$. We say that $\hat{\boldsymbol{\theta}}$ is an efficient estimator of $\boldsymbol{\theta}$ if

1. $E[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$.
2. $\text{var}[\hat{\boldsymbol{\theta}}] = \mathbf{I}^{-1}(\boldsymbol{\theta})$, where $\mathbf{I}(\boldsymbol{\theta})$ is the information matrix computed as follows:

$$\mathbf{I}(\boldsymbol{\theta})\text{=-E}\left[\frac{\partial^2 lnL}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta'}}\right] = -E \begin{bmatrix} \frac{\partial^2 \text{lnL}}{\partial \theta_1^2} & \frac{\partial^2 \text{lnL}}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 \text{lnL}}{\partial \theta_1 \partial \theta_p} \\ \frac{\partial^2 \text{lnL}}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \text{lnL}}{\partial \theta_2^2} & \cdots & \frac{\partial^2 \text{lnL}}{\partial \theta_2 \partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \text{lnL}}{\partial \theta_p \partial \theta_1} & \frac{\partial^2 \text{lnL}}{\partial \theta_p \partial \theta_2} & \cdots & \frac{\partial^2 \text{lnL}}{\partial \theta_p^2}. \end{bmatrix}.$$

Write $\mathbf{I}(\boldsymbol{\theta})$ for the model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$.

Write the likelihood function based on the normality assumption

$lnL =$

Find the following

$$\frac{\partial lnL}{\partial \beta_0} =$$

$$\frac{\partial^2 lnL}{\partial \beta_0^2} =$$

$$\frac{\partial^2 lnL}{\partial \beta_0 \partial \beta_1} =$$

$$\frac{\partial^2 lnL}{\partial \beta_0 \partial \sigma^2} =$$

$$\frac{\partial lnL}{\partial \beta_1} =$$

$$\frac{\partial^2 lnL}{\partial \beta_1^2} =$$

$$\frac{\partial^2 lnL}{\partial \beta_1 \partial \sigma^2} =$$

$$\frac{\partial lnL}{\partial \sigma^2} =$$

$$\frac{\partial^2 lnL}{\partial {\sigma^2}^{(2)}} =$$

Find the information matrix: $-E\left[\frac{\partial^2 lnL}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta'}}\right]$.

Note the information matrix is block diagonal. Therefore, find the inverse of the upper left $2 \times 2$ matrix of the information matrix $\mathbf{I}(\boldsymbol{\theta})$ and compare the elements of the $2 \times 2$ matrix with $\text{var}[\hat{\beta}_1] = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$, $\text{var}[\hat{\beta}_0] = \sigma^2[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}]$, and $\text{cov}[\hat{\beta}_0, \hat{\beta}_1] = -\frac{\bar{x}\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$.