Inference in Regression Analysis

Dr. Frank Wood
Inference in the Normal Error Regression Model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

- $Y_i$ value of the response variable in the $i^{th}$ trial
- $\beta_0$ and $\beta_1$ are parameters
- $X_i$ is a known constant, the value of the predictor variable in the $i^{th}$ trial
- $\epsilon_i \sim iid \ N(0, \sigma^2)$
- $i = 1, \ldots, n$
Inference concerning $\beta_1$

Tests concerning $\beta_1$ (the slope) are often of interest, particularly

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

the null hypothesis model

$$Y_i = \beta_0 + (0)X_i + \epsilon_i$$

implies that there is no relationship between $Y$ and $X$.

Note the means of all the $Y_i$’s are equal at all levels of $X_i$. 
Quick Review : Hypothesis Testing

- Elements of a statistical test
  - Null hypothesis, $H_0$
  - Alternative hypothesis, $H_a$
  - Test statistic
  - Rejection region
Quick Review: Hypothesis Testing - Errors

- **Errors**
  - A type I error is made if $H_0$ is rejected when $H_0$ is true. The probability of a type I error is denoted by $\alpha$. The value of $\alpha$ is called the level of the test.
  - A type II error is made if $H_0$ is accepted when $H_a$ is true. The probability of a type II error is denoted by $\beta$. 
P-value

The p-value, or attained significance level, is the smallest level of significance $\alpha$ for which the observed data indicate that the null hypothesis should be rejected.
Null Hypothesis

If the null hypothesis is that $\beta_1 = 0$ then $b_1$ should fall in the range around zero. The further it is from 0 the less likely the null hypothesis is to hold.
Alternative Hypothesis: Least Squares Fit

If we find that our estimated value of $b_1$ deviates from 0 then we have to determine whether or not that deviation would be surprising given the model and the sampling distribution of the estimator. If it is sufficiently (where we define what sufficient is by a confidence level) different then we reject the null hypothesis.
Testing This Hypothesis

- Only have a finite sample
- Different finite set of samples (from the same population/source) will (almost always) produce different point estimates of $\beta_0$ and $\beta_1$ ($b_0, b_1$) given the same estimation procedure
- Key point: $b_0$ and $b_1$ are random variables whose sampling distributions can be statistically characterized
- Hypothesis tests about $\beta_0$ and $\beta_1$ can be constructed using these distributions.
- The same techniques for deriving the sampling distribution of $b = [b_0, b_1]$ are used in multiple regression.
Sampling Dist. Of $b_1$

- The point estimator for $b_1$ is

\[
   b_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}
\]

- The sampling distribution for $b_1$ is the distribution of $b_1$ that arises from the variability of $b_1$ when the predictor variables $X_i$ are held fixed and the observed outputs are repeatedly sampled.

- Note that the sampling distribution we derive for $b_1$ will be highly dependent on our modeling assumptions.
For a normal error regression model the sampling distribution of $b_1$ is normal, with mean and variance given by

$$E(b_1) = \beta_1$$

$$\text{Var}(b_1) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}$$

To show this we need to go through a number of algebraic steps.
First step

To show

\[ \sum (X_i - \bar{X})(Y_i - \bar{Y}) = \sum (X_i - \bar{X})Y_i \]

we observe

\[
\sum (X_i - \bar{X})(Y_i - \bar{Y}) = \sum (X_i - \bar{X})Y_i - \sum (X_i - \bar{X}) \bar{Y} \\
= \sum (X_i - \bar{X})Y_i - \bar{Y} \sum (X_i - \bar{X}) \\
= \sum (X_i - \bar{X})Y_i - \bar{Y} \sum (X_i) + \bar{Y} n \frac{\sum X_i}{n} \\
= \sum (X_i - \bar{X})Y_i
\]
\( b_1 \) as convex combination of \( Y_i \)'s

\( b_1 \) can be expressed as a linear combination of the \( Y_i \)'s

\[
\begin{align*}
b_1 &= \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} \\
&= \frac{\sum(X_i - \bar{X})Y_i}{\sum(X_i - \bar{X})^2} \quad \text{from previous slide} \\
&= \sum k_i Y_i
\end{align*}
\]

where

\[
k_i = \frac{(X_i - \bar{X})}{\sum(X_i - \bar{X})^2}
\]
Properties of the $k'_i$s

It can be shown that

\[
\sum k_i = 0 \\
\sum k_i X_i = 1 \\
\sum k_i^2 = \frac{1}{\sum (X_i - \bar{X})^2}
\]

(possible homework). We will use these properties to prove various properties of the sampling distributions of $b_1$ and $b_0$. 
Normality of $b'_1$'s Sampling Distribution

- Useful fact:
  - A linear combination of independent normal random variables is normally distributed
  - More formally: when $Y_1, \ldots, Y_n$ are independent normal random variables, the linear combination $a_1 Y_1 + a_2 Y_2 + \ldots + a_n Y_n$ is normally distributed, with mean $\sum a_i \mathbb{E}(Y_i)$ and variance $\sum a_i^2 \text{Var}(Y_i)$
Normality of $b_1'$s Sampling Distribution

Since $b_1$ is a linear combination of the $Y_i'$s and each $Y_i$ is an independent normal random variable, then $b_1$ is distributed normally as well

$$b_1 = \sum k_i Y_i, \quad k_i = \frac{(X_i - \bar{X})}{\sum(X_i - \bar{X})^2}$$

From previous slide

$$\mathbb{E}(b_1) = \sum k_i \mathbb{E}(Y_i), \quad \text{Var}(b_1) = \sum k_i^2 \text{Var}(Y_i)$$
$b_1$ is an unbiased estimator

This can be seen using two of the properties

\[
\mathbb{E}(b_1) = \mathbb{E}\left(\sum k_i Y_i\right) \\
= \sum k_i \mathbb{E}(Y_i) \\
= \sum k_i (\beta_0 + \beta_1 X_i) \\
= \beta_0 \sum k_i + \beta_1 \sum k_i X_i \\
= \beta_0(0) + \beta_1(1) \\
= \beta_1
\]
Variance of $b_1$

Since the $Y_i$ are independent random variables with variance $\sigma^2$ and the $k_i's$ are constants we get

$$\text{Var}(b_1) = \text{Var}(\sum k_i Y_i) = \sum k_i^2 \text{Var}(Y_i)$$

$$= \sum k_i^2 \sigma^2 = \sigma^2 \sum k_i^2$$

$$= \sigma^2 \frac{1}{\sum(X_i - \bar{X})^2}$$

note that this assumes that we know $\sigma^2$. Can we?
Estimated variance of $b_1$

- When we don't know $\sigma^2$ then we have to replace it with the MSE estimate

- Let

\[ s^2 = MSE = \frac{SSE}{n - 2} \]

where

\[ SSE = \sum e_i^2 \]

and

\[ e_i = Y_i - \hat{Y}_i \]

plugging in we get

\[
\begin{align*}
\text{Var}(b_1) &= \frac{\sigma^2}{\sum(X_i - \bar{X})^2} \\
\hat{\text{Var}}(b_1) &= \frac{s^2}{\sum(X_i - \bar{X})^2}
\end{align*}
\]
Recap

- We now have an expression for the sampling distribution of $b_1$ when $\sigma^2$ is known

\[ b_1 \sim \mathcal{N}(\beta_1, \frac{\sigma^2}{\sum(X_i - \bar{X})^2}) \] (1)

when $\sigma^2$ is known.

- When $\sigma^2$ is unknown we have an unbiased point estimator of $\sigma^2$

\[ \hat{\text{Var}}(b_1) = \frac{s^2}{\sum(X_i - \bar{X})^2} \]

- As $n \to \infty$ (i.e. the number of observations grows large) $\hat{\text{Var}}(b_1) \to \text{Var}(b_1)$ and we can use Eqn. 1.

- Questions
  - When is $n$ big enough?
  - What if $n$ isn’t big enough?
Digression : Gauss-Markov Theorem

In a regression model where $\mathbb{E}(\epsilon_i) = 0$ and variance $\text{Var}(\epsilon_i) = \sigma^2 < \infty$ and $\epsilon_i$ and $\epsilon_j$ are uncorrelated for all $i$ and $j$ the least squares estimators $b_0$ and $b_1$ are unbiased and have minimum variance among all unbiased linear estimators.

Remember

\[
\begin{align*}
  b_1 &= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \\
  b_0 &= \bar{Y} - b_1 \bar{X}
\end{align*}
\]
Proof

- The theorem states that $b_1$ as minimum variance among all unbiased linear estimators of the form

$$\hat{\beta}_1 = \sum c_i Y_i$$

- As this estimator must be unbiased we have

$$\mathbb{E}(\hat{\beta}_1) = \sum c_i \mathbb{E}(Y_i) = \beta_1$$

$$= \sum c_i(\beta_0 + \beta_1 X_i) = \beta_0 \sum c_i + \beta_1 \sum c_i X_i = \beta_1$$
Proof cont.

Given these constraints

\[ \beta_0 \sum c_i + \beta_1 \sum c_i X_i = \beta_1 \]

Clearly it must be the case that \( \sum c_i = 0 \) and \( \sum c_i X_i = 1 \)

The variance of this estimator is

\[ \text{Var}(\hat{\beta}_1) = \sum c_i^2 \text{Var}(Y_i) = \sigma^2 \sum c_i^2 \]
Proof cont.

Now define $c_i = k_i + d_i$ where the $k_i$ are the constants we already defined and the $d_i$ are arbitrary constants. Let's look at the variance of the estimator

$$
\text{Var}(\hat{\beta}_1) = \sum c_i^2 \text{Var}(Y_i) = \sigma^2 \sum (k_i + d_i)^2
$$

$$
= \sigma^2 \left( \sum k_i^2 + \sum d_i^2 + 2 \sum k_i d_i \right)
$$

Note we just demonstrated that

$$
\sigma^2 \sum k_i^2 = \text{Var}(b_1)
$$
Proof cont.

Now by showing that \( \sum k_i d_i = 0 \) we’re almost done

\[
\sum k_i d_i = \sum k_i (c_i - k_i) \\
= \sum k_i (c_i - k_i) \\
= \sum k_i c_i - \sum k_i^2 \\
= \sum c_i \left( \frac{X_i - \bar{X}}{\sum(X_i - \bar{X})^2} \right) - \frac{1}{\sum(X_i - \bar{X})^2} \\
= \frac{\sum c_i X_i - \bar{X} \sum c_i}{\sum(X_i - \bar{X})^2} - \frac{1}{\sum(X_i - \bar{X})^2} = 0
\]
So we are left with

\[ \text{Var}(\hat{\beta}_1) = \sigma^2 \left( \sum k_i^2 + \sum d_i^2 \right) \]

\[ = \text{Var}(b_1) + \sigma^2 \left( \sum d_i^2 \right) \]

which is minimized when the \( d_i \)'s = 0. This means that the least squares estimator \( b_1 \) has minimum variance among all unbiased linear estimators.
Sampling Distribution of \((b_1 - \beta_1)/S(b_1)\)

- \(b_1\) is normally distributed so \((b_1 - \beta_1)/\sqrt{\text{Var}(b_1)}\) is a standard normal variable.
- We don’t know \(\text{Var}(b_1)\) so it must be estimated from data. We have already denoted it’s estimate.
- If using the estimate \(\hat{\text{V}}(b_1)\) it can be shown that

\[
\frac{b_1 - \beta_1}{\hat{S}(b_1)} \sim t(n - 2)
\]

\[
\hat{S}(b_1) = \sqrt{\hat{\text{V}}(b_1)}
\]
For now we need to rely upon the following theorem

For the normal error regression model

\[ \frac{SSE}{\sigma^2} = \frac{\sum(Y_i - \hat{Y}_i)^2}{\sigma^2} \sim \chi^2(n - 2) \]

and is independent of \( b_0 \) and \( b_1 \)

Intuitively this follows the standard result for the sum of squared normal random variables.

Here there are two linear constraints imposed by the regression parameter estimation that each reduce the number of degrees of freedom by one.

We will revisit this subject soon.
Another useful fact: t distributed random variables

Let $z$ and $\chi^2(\nu)$ be independent random variables (standard normal and $\chi^2$ respectively). The following random variable is a t-distributed random variable:

$$t(\nu) = \frac{z}{\sqrt{\frac{\chi^2(\nu)}{\nu}}}$$

This version of the t distribution has one parameter, the degrees of freedom $\nu$
Distribution of the studentized statistic

To derive the distribution of this statistic, first we do the following rewrite

\[
\frac{b_1 - \beta_1}{\hat{S}(b_1)} = \frac{b_1 - \beta_1}{\hat{S}(b_1)} \frac{\hat{S}(b_1)}{S(b_1)}
\]

\[
\frac{\hat{S}(b_1)}{S(b_1)} = \sqrt{\frac{\hat{V}(b_1)}{\text{Var}(b_1)}}
\]
Studentized statistic cont.

And note the following

\[
\frac{\hat{V}(b_1)}{\text{Var}(b_1)} = \frac{\frac{MSE}{\sum(X_i - \bar{X})^2}}{\frac{\sigma^2}{\sum(X_i - \bar{X})^2}} = \frac{MSE}{\sigma^2} = \frac{SSE}{\sigma^2(n - 2)}
\]

where we know (by the given theorem) the distribution of the last term is \( \chi^2 \) and indep. of \( b_1 \) and \( b_0 \)

\[
\frac{SSE}{\sigma^2(n - 2)} \sim \frac{\chi^2(n - 2)}{n - 2}
\]
Studentized statistic final

But by the given definition of the t distribution we have our result

$$\frac{b_1 - \beta_1}{\hat{S}(b_1)} \sim t(n - 2)$$

because putting everything together we can see that

$$\frac{b_1 - \beta_1}{\hat{S}(b_1)} \sim \frac{z}{\sqrt{\frac{\chi^2(n-2)}{n-2}}}$$
Confidence Intervals and Hypothesis Tests

Now that we know the sampling distribution of $b_1$ (t with n-2 degrees of freedom) we can construct confidence intervals and hypothesis tests easily. Things to think about

- What does the t-distribution look like?
- Why is the estimator distributed according to a t-distribution rather than a normal distribution?
- When performing tests why does this matter?
- When is it safe to cheat and use a normal approximation?