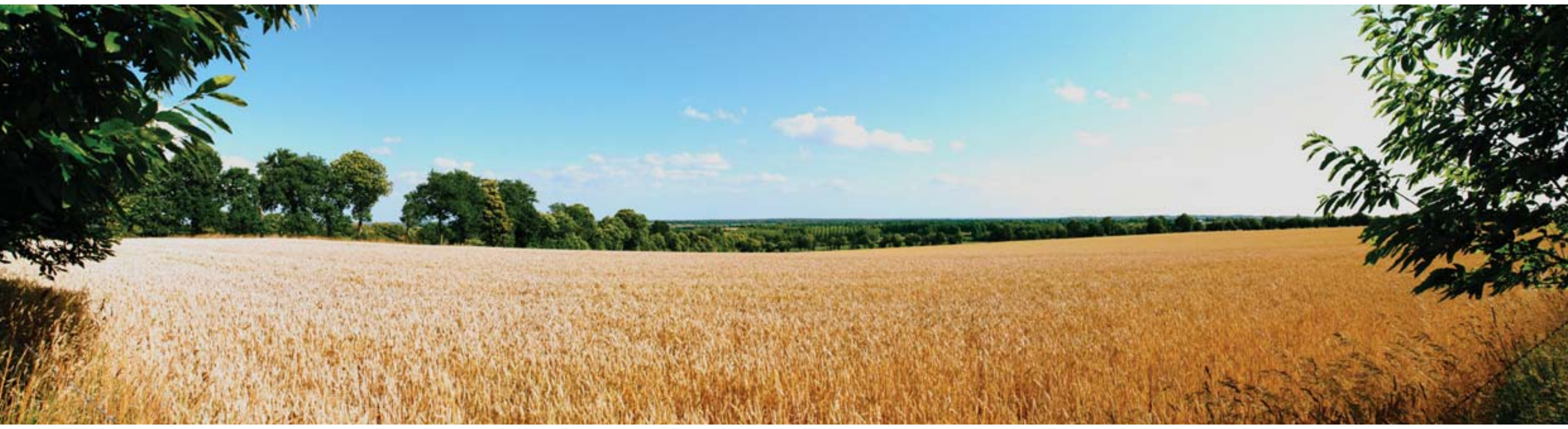

5. Interval Estimation and Hypothesis Testing

Interval Estimation

5.1 Confidence Intervals for Regression Coefficients β_1 and β_2

In Sum

A $100(1-\alpha)$ percent **confidence interval** for β_2 can be defined as:

$$\hat{\beta}_2 \pm t_{\alpha/2}\mathrm{se}(\hat{\beta}_2)$$

or

$$\Pr[\hat{\beta}_2 - t_{\alpha/2}\mathrm{se}(\hat{\beta}_2) \leq \beta_2 \leq \hat{\beta}_2 + t_{\alpha/2}\mathrm{se}(\hat{\beta}_2)] = 1-\alpha$$

Analogously, we can define $100(1-\alpha)$ percent **confidence interval** for β_1 as:

$$\hat{\beta}_1 \pm t_{\alpha/2}\mathrm{se}(\hat{\beta}_1)$$

or

$$\Pr[\hat{\beta}_1 - t_{\alpha/2}\mathrm{se}(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2}\mathrm{se}(\hat{\beta}_1)] = 1-\alpha$$

Example

Table 5.1: Estimating the expenditure of the household with income

Family (i)	Actual Y_i	Income X_i	$X-\bar{X}$	$Y-\bar{Y}$	$(X-\bar{X})(Y-\bar{Y})$	$(X-\bar{X})^2$
1	390	500	-250	-153.17	38291.67	62500
2	425	600	-150	-118.17	17725.00	22500
3	560	700	-50	16.83	-841.67	2500
4	575	800	50	31.83	1591.67	2500
5	630	900	150	86.83	13025.00	22500
6	679	1000	250	135.83	33958.33	62500
Sum	3259	4500	0	0	103750	175000

Table 5.2: Estimating the expenditure of the household with income

Family (i)	Actual Y_i	Income X_i	Regression Estimate $\hat{Y}$	Residual $Y-\hat{Y}$	Residual squared $(Y-\hat{Y})^2$
1	390	500	394.95	-4.95	24.53
2	425	600	454.24	-29.24	854.87
3	560	700	513.52	46.48	2160.04
4	575	800	572.81	2.19	4.80
5	630	900	632.10	-2.10	4.39
6	679	1000	691.38	-12.38	153.29
Sum	3259	4500	0	0	3201.90

Confidence Interval for β_2

Confidence Interval for β_1

5.2 Confidence Interval for σ^2

5.3 Hypothesis Testing: The Confidence-Interval Approach

Based on our sample data, the estimated marginal propensity to consume (MPC), $\hat{\beta}_2$ is 0.593. Suppose we postulate that

$$H_0 : \beta_2 = 0.6$$

$$H_1 : \beta_2 \neq 0.6$$

5.4 Hypothesis Testing: The Test of Significance Approach

5.4.1 Two-Tail Test

Based on the sample data, the estimated marginal propensity to consume (MPC), $\hat{\beta}_2$ is 0.593. Suppose we postulate that

$$H_0 : \beta_2 = 0.6$$

$$H_1 : \beta_2 \neq 0.6$$

5.4.2 One-Tail Test

Based on the sample data, the estimated marginal propensity to consume (MPC), $\hat{\beta}_2$ is 0.593. Suppose we postulate that

$$H_0 : \beta_2 \leq 0.6$$

$$H_1 : \beta_2 > 0.6$$

We can summarize the decision rules for the t test as follow:

Figure 5.1 The t test of Significance: Decision rules

Type of hypothesis	H_0: the null hypothesis	H_1: the alternative hypothesis	Decision rule: reject H_0 if
Two-tail	$\beta_2 = \beta_2^*$	$\beta_2 \neq \beta_2^*$	$\lvert t \rvert > t_{\alpha/2,\mathrm{df}}$
Right-tail	$\beta_2 \leq \beta_2^*$	$\beta_2 > \beta_2^*$	$t > t_{\alpha,\mathrm{df}}$
Left-tail	$\beta_2 \geq \beta_2^*$	$\beta_2 < \beta_2^*$	$t < -t_{\alpha,\mathrm{df}}$

Notes: β_2^* is the hypothesized numerical value of β_2.
$\lvert t \rvert$ means the absolute value of t.
t_α or $t_{\alpha/2}$ means the critical t value at the α or $\alpha/2$ level of significance.
df: degrees of freedom, $(n - 2)$ for the two-variable model, $(n - 3)$ for the three-variable model, and so on.
The same procedure holds to test hypotheses about β_1.

5.4.3 Testing the significance of σ^2: The χ^2 test

Figure 5.2 The χ^2 Test : Decision rules

H_0: the null hypothesis	H_1: the alternative hypothesis	Critical region: reject H_0 if
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\frac{\text{df}(\hat{\sigma}^2)}{\sigma_0^2} > \chi^2_{\alpha,\text{df}}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\frac{\text{df}(\hat{\sigma}^2)}{\sigma_0^2} < \chi^2_{(1-\alpha),\text{df}}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\frac{\text{df}(\hat{\sigma}^2)}{\sigma_0^2} > \chi^2_{\alpha/2,\text{df}}$ or $< \chi^2_{(1-\alpha/2),\text{df}}$

Note: σ_0^2 is the value of σ^2 under the null hypothesis. The first subscript on χ^2 in the last column is the level of significance, and the second subscript is the degrees of freedom. These are critical chi-square values. Note that df is $(n-2)$ for the two-variable regression model, $(n-3)$ for the three-variable regression model, and so on.

Why do we say “we cannot reject the null hypothesis?" instead of “We accept the null hypothesis"

The Level of Significance: α

Type I error

Type II error

The Exact Level of Significance: The p Value

5.4.1 Regression Analysis and Analysis of Variance

Table 5.3: ANOVA Table for the two-variable regression model

Source of variation	Sum of Square **SS**	**df**	Mean Sum of Square **MSS**
Due to regression (ESS)			
Due to residuals (RSS)			
TSS			

Table 5.4: Estimating the expenditure of the household

Family Number (i)	Actual	Estimate	**Error in Estimation**	**Errors Squared**
	Y_i	$\hat{Y}_i = \bar{Y}$	$Y_i - \bar{Y}$	$(Y_i - \bar{Y})^2$
1	390	543	-153	23460.03
2	425	543	-118	13963.36
3	560	543	17	283.36
4	575	543	32	1013.36
5	630	543	87	7540.03
6	679	543	136	18450.69
Sum	3259	3259	0	64710.83

Table 5.5: Estimating the expenditure of the household with income

Family (i)	Actual	Income	Regression Estimate	Residual	Residual squared
	Y_i	X_i	$\hat{Y}$	$Y - \hat{Y}$	$(Y - \hat{Y})^2$
1	390	500	394.95	-4.95	24.53
2	425	600	454.24	-29.24	854.87
3	560	700	513.52	46.48	2160.04
4	575	800	572.81	2.19	4.80
5	630	900	632.10	-2.10	4.39
6	679	1000	691.38	-12.38	153.29
Sum	3259	4500	0	0	3201.90

Table 5.6: ANOVA Table: Estimating the expenditure of the household with income

Source of variation	Sum of Square **SS**	**df**	Mean Sum of Square **MSS**
Due to regression (ESS)			
Due to residuals (RSS)			
TSS			

(After MID-TERM)

5.5 The problem of prediction

Based on our sample data, we have the following sample regression:

We can use the above regression to “Predict” or “Forecast” the future consumption expenditure Y corresponding to some given level of income X

There are two kinds of predictions which are:

[1] **Mean prediction** We will predict the conditional mean value of Y corresponding to a chosen X (i.e X_0)

[2] **Individual Prediction** We will predict an individual Y value corresponding to (i.e X_0)

Mean Prediction

We know that

$$\hat{Y}_0 \sim N(E(\hat{Y}_0), \text{var}(\hat{Y}_0))$$

where

$$E(\hat{Y}_0) = \beta_1 + \beta_2 X_0$$

and

$$\text{var}(\hat{Y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right]$$

If we replace the unknown σ^2 by the unbiased estimator $\hat{\sigma}^2$ we can get

$$t = \frac{\hat{Y}_0 - (\beta_1 + \beta_2 X_0)}{se(\hat{Y}_0)}$$

which is the t distribution with n-2 df.

Therefore, we can derive the confidence interval for the true $E(Y_0|X_0)$ as following:

$$Pr[\hat{\beta}_1 + \hat{\beta}_2 X_0 - t_{\frac{\alpha}{2}} se(\hat{Y}_0) \leq \beta_1 + \beta_2 X_0 \leq \hat{\beta}_1 + \hat{\beta}_2 X_0 + t_{\frac{\alpha}{2}} se(\hat{Y}_0)] = 1 - \alpha$$

Example

Individual Prediction
We can prediction an individual Y value, Y_0 corresponding to a given X value (X_0) :

but the variance in this case is:

$$\operatorname{var}(Y_0 - \hat{Y}_0) = E[Y_0 - \hat{Y}_0]^2 = \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right]$$

We can show that Y_0 follows the normal distribution:

$$Y_0 \sim N(\hat{Y}_0, \operatorname{var}(Y_0 - \hat{Y}_0))$$

Therefore, we can construct the confidence interval for the Y_0 as well.

From our example: