MA Quant II Midterm Revision[a]

Felipe Balcazar

NYU

March 17, 2021

[a]Thanks to Pedro L. Rodríguez

What are we doing?

- Observe the world.
- Intuit a relationship between Y and $X = (X_1, X_2, ..., X_p)$.

$$Y = f(X) + \epsilon$$

- Y = *systematic component* (S) + *random error term* (E)
- S: systematic information that X provides about Y.
- E: random error term with mean 0 and independent of X.
- Task: estimate f using random samples from the population.

Why are we doing it?

1. Prediction:
    - $\hat{Y} = \hat{f}(X)$
    - Note: $\hat{f}$ is an estimate of f.
    - $Y - \hat{Y} = f(x) + \epsilon - \hat{f}(X) = (f(X) - \hat{f}(X)) + \epsilon$
    - prediction error = reducible error + irreducible error
2. Inference:
    - Draw conclusions about the true model from observed data.
    - Understand relationship between Y and $X = (X_1, X_2, ..., X_p)$.
    - Ex. How does Y change if we vary X_1.

How do we estimate f?

There are many methods. All boil down to:

1. Sample data (preferably random).
2. Use sampled data to estimate f.

Two approaches:

1. Parametric:
   - Make assumptions on the functional form of f.
   - Estimate parameters using training data.
2. Non-Parametric:
   - Don't rely on assumptions about f.
   - Often times more accurate but less interpretable.

Ordinary Least Squares (OLS)

- Parametric.
- Functional form: linear.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi} + \epsilon_i$$

- Estimation method: minimize residual sum of squares (RSS).

$$RSS = e_1^2 + e_2^2 + ... + e_n^2$$

where $e_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + ... + \hat{\beta}_p x_{pi})$

True model vs. Fitted model

- Fitted model is an *approximation* of the true model.
- Formally: $f(x) \neq \hat{f}(x)$ but hopefully $f(x) \approx \hat{f}(x)$
- Moreover: $\epsilon \neq e$
- English: the error term $\neq$ residual.
- Recall: $Y - \hat{Y} = f(x) + \epsilon - \hat{f}(X) = (f(X) - \hat{f}(X)) + \epsilon$
- prediction error $=$ reducible error $+$ irreducible error
- residual $=$ reducible error $+ \epsilon$
- Also important: estimator $\neq$ estimate.
- Estimator $=$ method of estimation (ex. OLS).
- Estimate $=$ result of applying an estimator to a given sample.

Fitted values, residuals and coefficients

- We believe: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$
- We estimate: $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$
- Fitted values: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$
- Residual: $Y - \hat{Y}$

How does a unit change in X_1 affect Y (ceteris paribus)?

$$(\hat{\beta}_0 + \hat{\beta}_1 X_{1_2} + \hat{\beta}_2 X_2) - (\hat{\beta}_0 + \hat{\beta}_1 X_{1_1} + \hat{\beta}_2 X_2)$$

$$= \hat{\beta}_1 X_{1_2} - \hat{\beta}_1 X_{1_1} = \hat{\beta}_1 (X_{1_2} - X_{1_1}) = \hat{\beta}_1 * 1$$

Properties of OLS estimators

These are algebraic facts of OLS, not assumptions.

1. $\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}$
2. $\bar{\hat{Y}} = \bar{Y}$
3. $\bar{e}_i = 0$
4. $\rho(e_i, \hat{Y}_i) = 0$
5. $\rho(e_i, \hat{X}_i) = 0$

English:

1. Regression line passes through means of sample values $(\bar{X}, \bar{Y})$
2. Mean of the predicted values = mean of the observed values
3. Sample mean of the residuals = 0
4. Correlation between residuals and predicted values of Y =0
5. Correlation between residuals and observed values of X = 0

Gaus Markov Assumptions

Assumptions about the true model necessary to make inferences.

1. $E[(u_i|X_i)] = 0$
2. $Cov(u_i, u_j) = 0 \; \forall i \neq j$
3. $Var(u_i|X_i) = \sigma^2 \; \forall i$
4. $Cov(u_i, X_i) = 0 \; \forall i$
5. $Y = \beta X + \epsilon$

English:

1. For any value of X the disturbances average out to 0
2. The disturbance term is independent across observations
3. The variance of the disturbance term is the same for all i
4. The disturbances are exogenous
5. The regression model is properly specified

Normality assumption

To make inferences on the coefficient estimates we assume:

$$\epsilon_i \sim N(0, \sigma^2)$$

From this assumption it follows that:

$$\hat{\beta}_p \sim N(\beta_p, \sigma_{\hat{\beta}_p})$$

This assumption allows us to perform hypotheses test.

Hypotheses testing

Elements of a statistical test:

1. Null hypothesis, H_0.
   - Ex. $H_0 : \beta_1 = 0$
   - English: there is no relationship between Y and X_1.
2. Alternative hypothesis, H_1.
   - Ex. $H_1 : \beta_1 \neq 0$
   - English: there is a relationship between Y and X_1.
3. Test statistic (TS).
   - Ex. *t-statistic*: $t = \frac{\hat{\beta}_1 - \beta_0}{SE(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)}$
   - Measures # of std. deviations that $\hat{\beta}_1$ is away from the null.
4. Rejection region (RR).
   - Values of TS for which H_0 will be rejected in favor of H_1.
   - Ex. RR = *t-stat* $> k$ for some choice of k.

Hypotheses testing

Two types of errors we want to minimize:

1. *Type I error*: we reject H_0 when H_0 is true.
   - Probability of *Type I error* $= \alpha$.
2. *Type II error*: we accept H_0 when H_1 is true.
   - Probability of *Type II error* $= \beta$.

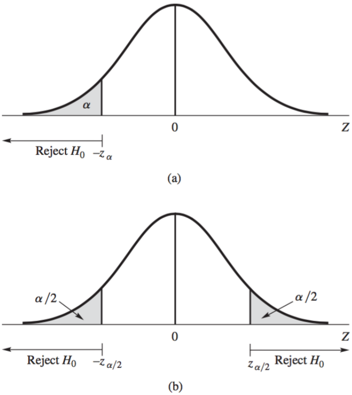
We use α to define the rejection region.

- RR $=$ *t-stat* $> k$ for some choice of k.
- k determined by fixing α and choosing corresponding t-value.
- For small samples use the t-distribution (recall normality ass.)
- *p-value* smallest level of α for which the observed data indicate H_0 should be rejected.

Hypotheses testing

Figure 1: Hypothesis Testing

FIGURE **10.4** Rejection regions for testing $H_0 : \theta = \theta_0$ versus (a) $H_a : \theta < \theta_0$ and (b) $H_a : \theta \neq \theta_0$, based on $Z = \dfrac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$



Source: Mathematical Statistics with Applications (7th Edition).

Hypotheses testing

Figure 2: Hypothesis Testing

A Small-Sample Test for μ

Assumptions: $Y_1, Y_2, \ldots, Y_n$ constitute a random sample from a normal distribution with $E(Y_i) = \mu$.

$H_0 : \mu = \mu_0$.

$$H_a : \begin{cases} \mu > \mu_0 & \text{(upper-tail alternative).} \\ \mu < \mu_0 & \text{(lower-tail alternative).} \\ \mu \neq \mu_0 & \text{(two-tailed alternative).} \end{cases}$$

Test statistic: $T = \dfrac{\overline{Y} - \mu_0}{S/\sqrt{n}}$.

$$\text{Rejection region: } \begin{cases} t > t_\alpha & \text{(upper-tail RR).} \\ t < -t_\alpha & \text{(lower-tail RR).} \\ |t| > t_{\alpha/2} & \text{(two-tailed RR).} \end{cases}$$

(See Table 5, Appendix 3, for values of t_α, with $\nu = n - 1$ df.)

Source: Mathematical Statistics with Applications (7th Edition).

Confidence intervals

Assuming $\hat{\beta}_p \sim N(\beta_p, \sigma_{\hat{\beta}_p})$...

The two-sided CI for β with confidence coefficient $1 - \alpha$ is:

$$\hat{\beta} \pm z_{\frac{\alpha}{2}} SE(\hat{\beta})$$

Relationship to hypothesis testing: recall two-tailed test...

$$-z_{\alpha/2} \leq \frac{\hat{\beta} - \beta_0}{\sigma_{\hat{\beta}}} \leq z_{\alpha/2}$$

$$\Rightarrow \hat{\beta} - z_{\alpha/2}\sigma_{\hat{\beta}} \leq \beta_0 \leq \hat{\beta} + z_{\alpha/2}\sigma_{\hat{\beta}}$$

Usually we have $\beta_0 = 0$.

Properties of estimators

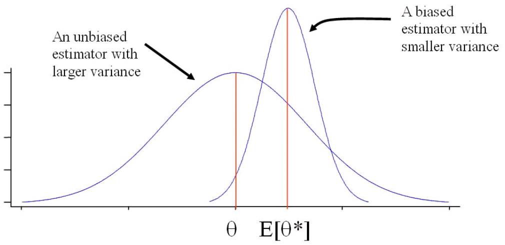
1. Bias

   Systematic deviation of estimates from true parameter.

2. Variance

   Spread of the estimates of an estimator.

Figure 3: Bias-Variance Tradeoff



Variance of OLS estimators

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1-R_j^2)'} \text{ for } j = 1, 2, ..., k$$

- where $SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$ is the total variation in x_j.
- $R_j^2 = R^2$ from regressing x_j on all other x's (inc. intercept).
- We don't usually know σ^2 (variance of disturbance term).
- Use sample estimate:

$$\hat{\sigma}^2 = \frac{1}{(n-k-1)} \sum_{i=1}^{n} \hat{u}_i^2 = \frac{1}{(n-k-1)} \sum_{i=1}^{n} e_i^2 = \frac{RSS}{(n-k-1)}$$

Assessing model accuracy

Two common measures:

1. residual standard error.
2. R^2 statistic.

Residual standard error (RSE):

- Recall: $Y = f(X) + \epsilon$.
- $Y =$ *systematic component* (S) $+$ *random error term* (E)
- Even if we knew $f(X)$ we would not perfectly predict Y.
- RSE is an estimate of the standard deviation of ϵ.
- $\approx$ avg. amount response will deviate from true regression line.

$$RSE = \sqrt{\frac{1}{n-2}RSS} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i^2)}$$

- Absolute measure of lack of fit.
- Not always clear what a good RSE is.

Assessing model accuracy

R^2 **statistic**:

- A proportion (hence a value between 0 and 1).
- Easier to interpret.
- Proportion of variability in Y that *can be explained* using X.

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

TSS = total sum of squares and RSS = residual sum of squares.

F-test 1

- Test-statistic for multiple regression setting.
- $H_0 : \beta_1 = \beta_2 = \cdots \beta_p = 0$
- H_1 : at least one $\beta_j \neq 0$

$$F = \frac{(TSS-RSS)/p}{RSS/(n-k-1)}$$

Where $k =$ number of parameters and $n =$ number of observations.

F-test 2

- We might want to test a particular subset q of the coefficients.
- $H_0 : \beta_{p-q+1} = \beta_{p-q+2} = \cdots \beta_p = 0$
- H_1 : at least one $(\beta_1, ..., \beta_q) \neq 0$

$$F = \frac{(RSS_0 - RSS)/q}{RSS/(n-k-1)}$$

$RRS_0 =$ RRS for the model excluding the q parameters.

$RRS =$ RRS for the full (unrestricted) model.

Reading regression output

- Recommend: `https://stats.idre.ucla.edu/stata/output/regression-analysis/`