EC220 - Suggested list of formulae

May 7, 2013

This list is in not meant to be a comprehensive summary of the course. It is a list of formulae that I feel useful to remember, as they can come in handy in an exam context or in your future studies. Some of them should be known from your first year statistics course, others are specific to this course. There are many additional formulae and proof from the course that you should also know, so do not rely solely on this document. In addition, if you feel that you are comfortable solving the exercises and understanding the course material without knowing this list, do not spend extra time trying to memorise it. Different people have different learning styles and all can be perfectly valid. If you find any typo, or have any questions or suggestions, please feel free to email me. Good luck!

A. Statistics review

For X, Y, V and W random variables and b a non-random number:

1. $E(X+Y) = E(X) + E(Y)$
2. $E(bX) = bE(X)$
3. $E(b) = b$
4. $VAR(X) = E[(X - E(X))^2]$
5. $VAR(X+Y) = VAR(X) + VAR(Y) + 2COV(X,Y)$
6. $VAR(bX) = b^2 VAR(X)$
7. $VAR(b) = 0$
8. $COV(X,Y) = E[(X - E(X))(Y - E(Y))]$
9. $COV(X, V+W) = COV(X,V) + COV(X,W)$

10. $COV(X, bY) = bCOV(X, Y)$

11. $COV(X, b) = 0$

12. $Corr(X, Y) = \frac{COV(X,Y)}{\sqrt{VAR(X)VAR(Y)}}$

13. Sample mean: $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$

14. Sample variance: $S^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$

15. Sample covariance: $S_{XY} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{n-1}$

    If $\hat{\theta}$ is an estimator for a true parameter θ,

16. Error: $\hat{\theta} - \theta$

17. Bias: $E[\hat{\theta}] - \theta$

18. Mean Squared Error (MSE): $E[(\hat{\theta} - \theta)^2]$

19. $\hat{\theta}$ is consistent if (1) $plim(\hat{\theta})$ exists and (2) $plim(\hat{\theta}) = \theta$

20. $plim(X + Y) = plim(X) + plim(Y)$

21. $plim(bX) = b\ plim(X)$

22. $plim(XY) = plim(X)plim(Y)$

23. $plim(\frac{X}{Y}) = \frac{plim(X)}{plim(Y)}$, provided both *plims* exist and $plim(Y) \neq 0$

24. Type I error: reject the null hypothesis when it is true

25. Type II error: fail to reject the null hypothesis when it is false

26. Power of a test = 1 - Pr(Type II error)

B. Simple Regression analysis:

Given a regression model: $Y = \beta_1 + \beta_2 X_2 + u$

1. Regression residual: $e_i = Y_i - \hat{Y}_i$

2. OLS estimator of β_1: $b_1 = \bar{Y} - b_2\bar{X}$

3. OLS estimator of β_2: $b_2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$

4. If the regression contains an intercept: $b_1 \neq 0$: (1) $\bar{e} = 0$, (2) $\bar{\hat{Y}} = \bar{Y}$, (3) $\sum_{i=1}^{n} X_i e_i = 0$, (4) $\sum_{i=1}^{n} e_i \hat{Y}_i = 0$

5. TSS $= \sum_{i=1}^{n}(Y_i - \bar{Y})^2$

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

7. RSS $= \sum_{i=1}^{n} e_i^2$

8. TSS = RSS + ESS

9. $R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}$

10. $VAR(b_2) = \frac{VAR(u)}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$

11. $s.e.(b_2) = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{(n-2)\sum_{i=1}^{n}(X_i-\bar{X})^2}}$

12. t statistic: $t = \frac{b_2-\beta_2^0}{s.e.(b_2)}$ where β_2^0 is the null hypothesis.

C. Multiple Regression analysis (also generally valid for simple regression analysis):

Given a regression model: $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_k X_k + u$

1. t statistic: $t = \frac{b_2-\beta_2^0}{s.e.(b_2)}$ where β_2^0 is the null hypothesis.

2. F statistic: $F(k-1, n-k) = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{ESS/(k-1)}{RSS/(n-k)}$

3. "Generalised" F statistic:

$$F(\text{number of restrictions}, d.f.^* \text{ in unrestricted model}) = \frac{(RSS_R - RSS_U)/(\text{number of restrictions})}{RSS_U/(d.f. \text{ in unrestricted model})}$$

** degrees of freedom*

4. For a regression model with two variables $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$, $s.e.(b_2) = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{(n-3)\sum_{i=1}^{n}(X_i-\bar{X})^2} \times \frac{1}{1-corr(X_2,X_3)^2}}$

D. Non-linear model:

1. Double logarithmic model:

$$Y = \beta_1 X_2^{\beta_2} v \Rightarrow log(Y) = log(\beta_1) + \beta_2 log(X_2) + log(v)$$

2. Semi-logarithmic model:

$$Y = \beta_1 e_2^{\beta_2 X_2} v \Rightarrow log(Y) = log(\beta_1) + \beta_2 X_2 + log(v)$$

E. Instrumental variables:

1. If Z is an instrument for X, then $b_2^{IV} = \frac{\sum_{i=1}^{n}(Z_i-\bar{Z})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(Z_i-\bar{Z})(X_i-\bar{X})}$
2. $VAR(b_2^{IV}) = \frac{VAR(u)}{\sum_{i=1}^{n}(X_i-\bar{X})^2} \times \frac{1}{corr(X,Z)^2}$
3. Two Stage Least Square:
   (a) Regress the reduced form equation
   (b) Use the fitted values as an instrument
   (c) $b_2^{2SLS} = \frac{\sum_{i=1}^{n}(\hat{X}_i-\bar{\hat{X}})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(\hat{X}_i-\bar{\hat{X}})(X_i-\bar{X})}$

F. Time series:

1. Adaptive expectation model:

$$Y_t = \beta_1 + \beta_2 X_{t+1}^e + u_t$$
$$X_{t+1}^e - X_t^e = \lambda(X_t - X_t^e)$$

2. Partial adjustment model:

$$Y_t^* = \beta_1 + \beta_2 X_t + u_t$$
$$Y_t - Y_{t-1} = \lambda(Y_t^* - Y_{t-1})$$

3. Prediction error: $f_{T+p} = Y_{T+p} - \hat{Y}_{T+p}$
4. Autoregressive AR(n) process: $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_n u_{t-n} + \varepsilon_t$
5. Moving average MA(n) process: $u_t = \lambda_0 \varepsilon_t + \lambda_1 \varepsilon_{t-1} + \ldots + \lambda_n \varepsilon_{t-n}$
6. Durbin-Watson d statistic: $d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$
7. Durbin-Watson d statistic: $h = \hat{\rho} \times \sqrt{\frac{n}{1-ns^2_{b_{y(t-1)}}}}$, where $\hat{\rho} = 1 - 0.5d$ and $s^2_{b_{y(t-1)}}$ is the estimator of the variance of the coefficient of the lagged variable.
8. An **endogenous variable** is a variable whose value is determined interactively within the model.
9. An **exogenous variable** is a variable whose value is determined outside the model.
10. A time series process is stationary if $E(x_t)$, $VAR(X_t)$ and $COV(X_t, X_{t-s})$ are independent of t.

11. Random walk: time series process of the form $X_t = X_{t-1} + \varepsilon_t$
12. Random walk with drift: $X_t = \beta_1 + \beta_2 X_{t-1} + \varepsilon_t$
13. Deterministic trend $X_t = \beta_1 + \beta_2 t + \varepsilon_t$
14. X_t is integrated of order 1 (I(1)) if $(X_t - X_{t-1})$ is stationary but X_t is non-stationary.
15. X_t is trend stationary if $(X_t - \hat{X}_t)$ is stationary where $\hat{X}_t$ is a trend.