Review Formulas

I. $\sum$ *operations*

$$\sum aX_i = a\sum X_i$$
$$\sum(X_i + Y_i) = \sum X_i + \sum Y_i$$
$$\sum\left(X_i - \bar{X}\right) = 0$$
$$\sum\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) = \sum\left(X_i - \bar{X}\right)Y_i$$
$$\sum\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) = \sum X_iY_i - n\bar{X}\bar{Y}$$
$$\sum\left(X_i - \bar{X}\right)^2 = \sum X_i^2 - n\bar{X}^2$$

II. Frequency Distribution

cumulative frequency distribution
mean, variance
joint frequency distribution
marginal frequency distribution
conditional frequency distribution

III. Simple Regression

conditional mean function
assumed functional form (eg. linear)
normal equations: least-squared error

$$\sum e_iX_i = 0$$
$$\sum e_i = 0$$

deviation notation $x_i = (X_i - \bar{X})$ $y_i = \left(Y_i - \bar{Y}\right)$

For $Y_i = a + b_{yx}X_i + e_i, \quad y_i = b_{yx}x_i + e_i$

$$b_{yx} = \frac{\sum y_ix_i}{\sum x_i^2} \qquad a = \bar{Y} - b\bar{X}$$

IV. Multiple Regression

$$Y_i = a + b_{y_{1.2}}X_{1i} + b_{y_{2.1}}X_{2i} + e_{y.12i}$$

normal equations

$$\sum e_ix_{1i} = 0$$
$$\sum e_ix_{2i} = 0$$

auxiliary regression $x_{1i} = b_{12}x_{2i} + e_{1.2i}$

$$b_{y_{1.2}} = \frac{\sum y_i e_{1.2i}}{\sum e_{1.2_i}^2}$$
$$b_{y_{1.2}} = b_{y_1} - b_{21} b_{y_{2.1}}$$
$$\sum y_i^2 = \sum \quad {}_i^2 + \sum \hat{e}_i^2$$

SST=SSR + SSE

$$R^2 = \frac{\sum \quad {}_i^2}{\sum y_i^2} = 1 - \frac{\sum e_i^2}{\sum y_i^2}$$

V. Dummy variable Regressions

$$Y_i = a + bD_i + e_i \qquad D_i = \begin{cases} 1 \text{ if member} \\ 0 \text{ if not member} \end{cases}$$

$a = \bar{Y}nonmember$

$b = \bar{Y}member - \bar{Y}nonmember$

Equations with several dummies

Equations with several dummies for categories

Equations with several dummies dummies and continuous variables

VI.Probability

Probability distribution

$\sum_{j=-\infty}^{\infty} P(x_j) = 1$

joint probability distribution p(Y,X)

marginal probability

conditional probability

$P(Y \mid X) = \frac{P(YandX)}{P(X)}$

independent random variables $P(XandY) = P(X) \bullet P(Y)$

VII.Expected Values

$E(g(x)) = \sum_j g(x_j) p(x_j)$

$E(y) = \sum_j Y_j p(y_j)$

Expected value of linear function $y = a + bX$

$E(Y) = a + bE(X)$

$\sigma_y^2 = E\left([Y_i - E(Y)]^2 \right) = b^2 \sigma_x^2$

Expected value of the sum of random variables

$E(Y_1 + Y_2 + Y_3) = E(Y_1) + E(Y_2) + E(Y_3)$

variance of sum of random variables

$\sigma^2_{y+x} = \sigma^2_y + \sigma^2_x + 2\sigma^2_{yx}$

$\sigma^2_{y-x} = \sigma^2_y + \sigma^2_x - 2\sigma^2_{yx}$

IIX. Binomial Distribution

$E(x) = n\pi$ x-success, n-trial, π –probability of success on single trial

$\sigma^2_x = n\pi(1-\pi)$

$\frac{x}{n} = proportion$

$E(\frac{x}{n}) = \pi$

$\sigma^2_{\frac{x}{n}} = \frac{\pi(1-\pi)}{n}$

IX.Normal Distribution

$\mu - mean$

$\sigma - s\tan darddeviation$

$\mu \pm \sigma = 68.3\% ofdistribution$

$\mu \pm 2\sigma = 95.4\% ofdistribution$

$\mu \pm 3\sigma = 99.7\% ofdistribution$

Standard Normal Distribution

$z = \frac{X_i - \mu_x}{\sigma_x}$ $\bar{z} = 0, \sigma_z = 0$

X.Frequency Distribution of $\bar{X}$

Central Limit Theorem

$E(\bar{X}) = \mu$

$\sigma^2_{\bar{x}} = \frac{\sigma^2_x}{n}$

Confidence Intervals and tests of significance

Hypothesis tests-H_o : *null hypothesis* H_a : *alternative hypothesis*

Point estimates and confidence intervals

Frequency distribution of $\bar{X}_1 - \bar{X}_2$

$\sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{\sigma^2_{x1}}{n_1} + \frac{\sigma^2_{x2}}{n_2}$ if X_1 and X_2 are independent

z=$\frac{\bar{X}_1 - \bar{X}_2 - \mu_0}{\sqrt{\frac{\sigma^2_{x1}}{n_1} + \frac{\sigma^2_{x2}}{n_2}}}$

Hypothesis tests and confidence intervals for means and proportions

t sample=$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S^2_{x1}}{n_1} + \frac{S^2_{x2}}{n_2}}}$ (if greater then t critical, reject $H_0 : \mu_{x1} - \mu_{x2} = 0$)

t sample=$\frac{P_1 - P_2}{\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}}$(if greater then t critical, reject $H_0 : \bar{P}_1 - \bar{P}_2 = 0$)

$\left(\bar{X}_1 - \bar{X}_2\right)$ −t critical$\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \leq \mu_{x1-}\mu_{x2} \leq \left(\bar{X}_1 - \bar{X}_2\right)$ +t critical$\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

$(P_1 - P_2)$ −t critical$\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}} \leq \bar{P}_1 - \bar{P}_2 \leq (P_1 - P_2)$ −t critical$\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}$

XI. Variance of the Regression Coefficient

Derived via dummy variable regression

$E[(b_{y_1} - E(b_{y_1})]^2)$

$\sigma^2_{by_1} = \frac{\sigma^2_{by_1}}{\sum x_i^2}$

$\sigma^2_{by_{1.2}} = \frac{\sigma^2_{by\cdot 12}}{\sum e^2_{1.2i}}$

t distribution and t tests

t=$\frac{b_{y_1} - B_{y_0}}{\sigma_{by_1}}$

relation of t and z

Hypothesis test regarding $by_{1.2}$ $u\sin g$ t.

t sample=$\frac{by_{1.2} - H_0}{\sigma_{by_{1.2}}}$ (if greater than t critical reject H_0)

Confidence Intervals

$by_{1.2} - tcritical\sigma_{by_{1.2}} \leq \beta y_{1.2} \leq by_{1.2} + tcritical\sigma_{by_{1.2}}$

XII. F Statistic

Using F to test restricted model vs.unrestricted model

test on regression as a whole

test on subset of regression coefficient

test on equivalence across data sets (e.g. time periods)

$F_{(Q-K,N-Q)} = \frac{R_u^2 - R_R^2}{1 - R_u^2} \bullet \frac{N-Q}{Q-K}$

XIIV. Decision Theory

State of Nature

decision criteria: min-max, max-max, max expected value

$$\sum_i priorprobabilityofSoN_j x(payoff \mid A_j andSoN_j) = E(payoff \mid A_j)$$

posterior probabilities and Bayer rule

prob $(SoN_j \mid sample_j) = \frac{prob(sample_j \mid SoN_j)priorprob(SoN_j)}{prob(sample_j)}$

expected value of certain prediction

expected value of perfect information

XIV. Assumptions of Classical Linear Regression

$E(\epsilon_i) = 0$
$E(\epsilon_i^2) = 0$
$E(\epsilon_i \epsilon_j) = 0$
$E(\epsilon_i X_i) = 0$

$$\left\{ \begin{matrix} E(b) \\ var(b) \end{matrix} \right\} \text{use classical assumptions to reduce to} \left\{ \begin{matrix} E(b) = \beta \\ var(b) = \frac{\sigma_\epsilon^2}{\sum x_i^2} \end{matrix} \right\}$$

XV. Properties of Estimators

unbiasedness $\mathrm{E}\left(\hat{B}\right) = \beta + E\left[\frac{\sum \epsilon_i X_i}{\sum x_i^2}\right]$ $var(\hat{\beta}) = E\left[\frac{\sum \epsilon_i^2 X_i^2}{\left(\sum x_i^2\right)^2}\right] + 2E\left[\frac{\sum_{i<j} \epsilon_i \epsilon_j X_i X_j}{\left(\sum x_i^2\right)^2}\right]$

efficiency

XVI. Non-Standard Analysis

problems in data: effects on ordinary least squares
omitted variable bias, inclusion of irrelevant variable
multi-colinearity
measurement error
problems in the population
violations of classical assumptions: effect on ordinary least squares and fixups
heteroskedasticity
auto-correlation
simultaneous equations bias