MEMORIAL UNIVERSITY

Lecture Notes for

ENGI 4421
Probability & Statistics

by

Dr. G.H. George
Associate Professor,
Faculty of Engineering and Applied Science

Seventh Edition, reprinted
2021 Spring

http://www.engr.mun.ca/~ggeorge/4421/

Table of Contents

List of Symbols:

$\subset$ proper subset
$\subseteq$ subset
$\bigcap$, $\wedge$, $\times$ intersection (same as logical AND)
$\bigcup$, $\vee$ union = logical inclusive OR
Ø empty set = null set = {}
α probability of type I error
$\mathrm{b}(x; n, p)$ binomial p.m.f.
$\mathrm{B}(x; n, p)$ binomial c.d.f.
β probability of type II error
$\beta(\mu_1)$ probability of type II error if $\mu = \mu_1$
β_0 intercept of regression line
$\hat{\beta}_0$ estimate of β_0
β_1 slope of regression line
$\hat{\beta}_1$ estimate of β_1
CI confidence interval
c boundary of rejection region or CI (one-sided)
c_L lower boundary of two-sided rejection region or CI
c_U upper boundary of two-sided rejection region or CI
${}^n\mathrm{C}_r$ number of combinations of r objects from n
$\mathrm{Cov}[X, Y]$ covariance
χ^2_ν chi-square distribution with ν degrees of freedom
$\chi^2_{\alpha,\nu}$ c for which $\mathrm{P}\left[\chi^2_\nu > c\right] = \alpha$
E an event (value 1 if true, else 0)
$\tilde{E}$ or ~E or not-E or E^c or E' or E^* or $\overline{E}$ the complement of event E
$\mathrm{E}[X]$ expected value of X $(= \mu)$

e_i residual for i^{th} point (in SLR) **or**
e_i number expected in cell i if the null hypothesis is true (χ^2 tests)
e_{ij} number expected in cell (i, j) if the null hypothesis is true
ε_i true error for i^{th} point (in SLR)
f value of F statistic = MSR/MSE
f_i frequency (number of observations in i^{th} interval)
$f(x)$ probability density function (pdf)
$F(x)$ cumulative distribution function (cdf) $= \mathrm{P}[X \le x]$
$\Gamma(x)$ gamma function (= (x–1)!)
$\Gamma(r, \lambda)$ gamma distribution
$\mathcal{H}_o$ null hypothesis
$\mathcal{H}_A$ or $\mathcal{H}_a$ or $\mathcal{H}_1$ alternative hypothesis
IQR interquartile range
λ occurrence rate (in Poisson and exponential distributions)
$\tilde{x}$ sample median
$\tilde{\mu}$ population median
MAD mean absolute deviation from the mean
MSE mean square error $= s^2$
MSR mean square regression
$\overline{X}$ sample mean (estimator)
$\overline{x}$ sample mean (estimate)
μ population mean
μ^* posterior estimate of μ (Bayesian)
μ_o prior estimate of μ (Bayesian)
or value of μ if null hypothesis true
n sample size

List of Symbols (continued)

N population size

$\mathrm{N}(\mu, \sigma^2)$ normal distribution with mean μ, variance σ^2 ν number of degrees of freedom

o_i number observed in cell i

o_{ij} number observed in cell (i, j)

$o_{i\bullet}$ number observed in row i

$o_{\bullet j}$ number observed in column j

$o_{\bullet\bullet}$ total number observed = n

p population proportion

$p = \mathrm{P}[E]$ probability that event E occurs

$\mathrm{P}[A \mid B]$ conditional probability (that event A occurs given that event B has occurred)

$\mathrm{P}[A \cap B]$ joint probability (that both events A and B occur); it is also $\mathrm{P}[A \wedge B] = \mathrm{P}[A \times B] = \mathrm{P}[A \text{ and } B] = \mathrm{P}[AB]$

$\mathrm{P}[A \cup B]$ probability that at least one of events A and B occurs; it is also $\mathrm{P}[A \vee B] = \mathrm{P}[A \text{ or } B]$

$p(x) = \mathrm{P}[X = x]$ probability mass function (pmf) for discrete x

$p(x, y) = \mathrm{P}[(X = x) \cap (Y = y)]$

$p_X(x) = \sum_y p(x, y)$ marginal pmf

$p_{Y|X}(y \mid x) = p(x, y) / p_X(x)$

P sample proportion (estimator)

$\hat{p}$ sample proportion (estimate)

p^* adjusted $\hat{p}$ (Agresti-Coull CI)

${}^n\mathrm{P}_r$ number of permutations of r objects from n

PI prediction interval (SLR)

$\phi(z)$ standard normal p.d.f.

$\Phi(z)$ standard normal c.d.f.

$q = \tilde{p} = 1 - p$ complement of the probability p

$Q = 1 - P$ complement of P

$\hat{q} = 1 - \hat{p}$ complement of $\hat{p}$

q_L or x_L lower quartile

q_U or x_U upper quartile

q^* adjusted $\hat{q}$ (Agresti-Coull CI)

r sample correlation **or**

$r = \dfrac{p}{1-p}$ odds that an event with probability p occurs.

$r_i = f_i / n$ relative frequency

ρ population correlation

S sample space = universal set = possibility space (set of all possible outcomes)

s sample standard deviation

s^2 sample variance

σ population standard deviation

σ^2 population variance

σ_o^2 prior variance (Bayesian)

$(\sigma^2)^*$ posterior estimate of σ^2 (Bayesian)

$S_{xy} = n\sum xy - \sum x \cdot \sum y$

$S_{xx} = n\sum x^2 - \left(\sum x\right)^2$

$S_{yy} = n\sum y^2 - \left(\sum y\right)^2$

$SIQR$ semi-interquartile range

SLR simple linear regression

$s.e.$ standard error $= \sigma / \sqrt{n}$

SSE sum of squares due to error

SSR sum of squares due to regression

SST total sum of squares

T_ν random quantity following t - distribution with ν degrees of freedom

$t_{\alpha,\nu}$ t for which $\mathrm{P}\left[T_\nu > t\right] = \alpha$

t_{obs} observed value of t

$\mathrm{V}\left[X\right]$ variance of X $= \sigma^2$

w_{o} prior weight (Bayesian)

w_d data weight (Bayesian)

$\hat{y}$ value of y predicted from regression line

Z standard normal random quantity

z_α z for which $\mathrm{P}\left[Z > z\right] = \alpha$

z_{obs} observed value of z

[Replace this sheet by the Course Outline handout that is posted on the course web site during the first week of classes in this course]

[Replace this sheet by the Course Outline handout that is posted on the course web site during the first week of classes in this course]

[Replace this sheet by the Course Outline handout that is posted on the course web site during the first week of classes in this course]

[Replace this sheet by the Course Outline handout that is posted on the course web site during the first week of classes in this course]

[Replace this sheet by the Course Outline handout that is posted on the course web site during the first week of classes in this course]