|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Review on LA | | Component wise op (x and y vector): cx, x±y, x\*y, x/y , x^y  Transpose: A' or AT  Matrix multiplication: A\*B = (cij)n x m , cij =  Vector multiplication: x\*y =  Inverse: A-1  Vector: c(1,2,3,4) / c(y,z) / A[,1] / A[1,]  Matrix: matrix(c(1,,2,3,4), ncol = 2, byrow = TRUE)  Dataframe: data.frame(x = c(1,2), y = c(2,3))  Get components of df: B$age / B[,2] | | | In R: c**x**, **x**±**y**, **x**\***y**, **x**/**y**  t(A)  A%\*%B  t(**x**)%\*%**y**  solve(A)  cbind(x,y), rbind(x,y)  data.frame(age = x1, income = x2)  read.table("path") / read.csv("path") |
| LI and vector space | | Let x1, ..., xp be p n-vectors. LC of these vectors = = c1**x**1 + ... + cp**x**p. If = 0 implies cj = 0 for all j. Then all p vectors are LI  All possible LC of x1, ..., xp will form a linear space. Space is spanned by these vectors. If x1, ..., xp are LI, linear space had dim p  If p < n, space spanned by x1, ..., xp is subspace of Rn | | | |
| Special Matrix | | A is symmetric is A = t(A)  A is idempotent if A = A2 -> eigenvalues of A are 1 or 0 & rk(A) = tr(A) | | Symmetric matrix A is +ve/(-ve) definite if for any vector x, xTAx > (<)0. ≥ (non -ve); ≤ (non +ve) definite  Max num of indep cols/rows vectors of A is its rank ( rk(A) )  If A is a sq matrix, sum of its diagonal elements is its trace ( tr(A) ) | |
| If A is a sq matrix and ATA = I, then A is orthogonal matrix -> rows/cols of A form orthonormal basis for Rn -> all rows/cols are pairwise orthogonal (u v = 0) & all rows/cols are unit vector | | | |
| Decom-position | | Cholesky decomposition: If A is non-negative definite, A = RTR, where R is an unique upper triang matrix w non -ve diag elems.  QR decomposition: For any n x p (n>p) matrix A, there is an orthogonal matrix Q s.t. A = QR, where R is a n x p matrix w all elems 0 below its diagonal, i.e. rij = 0, for i > j  Singular value decomposition: For any n x p matrix A, there is an n x n orthogonal matrix U and a p x p orthogonal matrix V s.t. UTAV = , if n ≥ p or UTAV = if n ≤ p. Where ∑ = diag(, , ..., ) where m = min{n, p} and ≥ ≥ ... ≥ > 0  Eigenvalue decomposition: For any non-negative definite matrix A, there is an orthogonal matrix Q s.t. A = QTDQ where D is a diagonal matrix whoses diagonal elems are eigenvalues of A arranged in order d11 ≥ d22 ≥ ... ≥ dnn | | | |
| More theorems | | | If AB and BA are both compatible -> tr(AB) = tr(BA) & AB and BA have the same non-zero eigenvalues  If A is non-negative definite, = (A), where denotes largest eigenvalue  If A is non-negative definite and B is non-singular (i.e. has inverse, det(B) = 0), = (AB-1) | | |
| Inverse of block matrix | | | Let ∑ = . This partition is a block matrix. If ∑ is non-singular and both and are sq matrices, then = where = OR  = where = | | |
| Review on Stats | | | Let X be a r.v.. cdf of r.v. X is F(x) = P(X≤x). If F is differentiable, pdf of X is f(x) = dF(x)/dx  (X1, ..., Xn) is called a random sample from distribution F, if Xi's are indep and identically distributed (iid) as F  Observation of random sample is denoted as (x1, ..., xn) | | |
| Descriptive quantities | | | Pop mean of X is = E(X) = ∫xf(x) dx (Riemann integral) = ∫xdF(x) (Lebesgue integral). Sample mean is =  For j = 1,2,..., the jth pop moment of X is mj = E(Xj) = . jth sample moment is mj = | | |
| Var of X is = E(X - E(X))2 = . Sample var is  Let Y be another r.v. w observed sample y1,..., yn. Covariance btw X and Y is Cov(X, Y) = E[(X - E(X))(Y - E(Y))]  If Cov(X, Y) = 0, then X and Y are un-correlated. Sample covariance is  Suppose X has a cts dist. The (1 - )-quantile (upper -quantile), is defined s.t. F() = 1 - or 1 - F() = | | |
| Mean: E = . Var: Var(X) = E(X2) - [E(X)]2  Var = in general; Var = if Xj's are un-correlated  Covariance: Cov(X, Y) = E(XY) - E(X)E(Y). Cov = | | |
| Sample var: - . Sample cov: = -  Empirical dist of observed sample x1, ..., xn is defined as Fn(x) = , where (indicator var) = 1 if xi ≤ x, 0 otherwise. Sample mean and sample var are indeed the mean and var of the empirical dist Fn | | |
| Matrix op on Exp and Var | | | Let A, B be constant matrics and **b, c** be vectors. Let X, Y be random vectors.  E(AX + **b**) = AE(X) + **b** where E(X) = (E(X1), ..., E(Xp))T  Var(AX + **b**) = AVar(X)AT where Var(X) = Cov(Xi, Xj) = E(X - E(X))(X - E(X))T. Note Cov(Xi, Xi) = Var(Xi)  Cov(AX + **b**, BX + **c**) = AVar(X)BT. Cov(AX + **b**, BY + **c**) = ACov(X, Y)BT where Cov(X, Y) = Cov(Xi, Yj) = E(X - E(X))(Y - E(Y))T | | |
| Uni-variate Normal dist | | r.v. X has uni-variate normal dist if pdf is f(y|, ) = where is the mean and is the var of the dist  Normal dist is denoted by N(, ). If = 0, = 1, it is the standard normal dist  Normal dist is symmetric about its mean, i.e. for any x, F(-x) = 1 - F(X) or for any c > 0, P(|X| ≥ c) = 2P(X ≥ c) = 2P(X ≤ -c) if = 0  If X ~ N(, ), then Z = (X - )/ ~ N(0,1) := standardisation | | | |
| Multi-variate normal dist | Let **Y** = (Y1, ..., Ym)T be random vector whose componenets are iid standard normal variables. Let A be q x m constant matrix and a constant vector. Dist of **X** = A**Y** + is a q-dimensional multivariable normal dist w mean E(**X**) = and var matrix ∑ = AAT, and denoted by N(, ∑)  ∑ = . Normal dist is uniquely determined by its mean and var matrix ∑  pdf of N(, ∑) is f(**x**|, ∑) =  If X is a multivariate normal vector, then for any constant matrix B, B**X** has a multivariate normal dist N(B, B∑BT)  constant vector **c**, the LC **c**T**X** has a univariate normal dist N(**c**T, **c**T∑**c**) and any component of **X** is an univariate normal var  If **X**~N(, ∑), then **Z** = ~N(**0**, I), i.e. components of **Z** are iid N(0, 1) variables | | | | |
| , t and F - dist | | Let **Z** = (Z1,..., Zm)T. Suppose Zj's are iid N(0,1) variables. Dist of **Z**T**Z** = is called the -dist w d.f. m and denoted by  Suppose Z~N(0,1), U~, Z and U are indep. The dist of Z/() is called the t-dist w d.f. m and denoted by tm  Suppose U~, V~, U and V are indep. Dist of is called the F-dist w d.f. m and n ande denoted by Fm,n  If **X**~N(, ∑), then W = (**X** - )T~  If Z~N(0,I), A is symmetric and idempotent, then **Z**TA**Z**~, where r = rk(A) = tr(A) | | | |
| Mtd of Moment estimation (MME) | | | Let be param of interest. Suppose it can be expressed as a fn of the moments as = g(m1, m2,...) (1st moment, 2nd moment,...). Then MME of is obtained by replacing the theoretical moments in the fn w the corresponding sample moments, i.e. = g(, ,...)  For any dist, the var = m2 - m12 (E(X2) - [E(X)]2), its MME is given by = -  MME is not unique. E.g. mean and var of Poisson dist m1 = = and = m2 - m12, which yield 2 MME  = , = - | | |
| Maximum likelihood estimation (MLE) | | | If X has pdf f(x, ), given the observation x1,...,xn of a random sample, the log likelihood fn of is defined as () = . The MLE of is value of that maximizes the log-likehood fn.  E.g. let (x1,...,xn) be observation of random sample from N(, ).  Likelihood fn: ( is prediction here)  The log likelihood fn of (, ) is then (, ) = = – log(2π) –  MLE of and are obtained by maximizing (, ) and are given by = , = | | |
| Confidence intervals (CI) | | | A 100(1-)% CI for paramis an random interval [L(**X**), U(**X**)] s.t. P( [L(**X**), U(**X**)]) ≥ 1-  Most commonly, CI is [ - s.d(), + s.d()] where is an estimate of , s.d() is estimated s.d. of and is the upper quantile of the dist of ( - )/s.d.() | | |
| Neyman-Pearson framework of hypo testing | | | Null hypothesis H0 and alternative hypothesis H1. The 2 hypotheses are mutually exclusive  A test statistic, T(x) is used. If T(x) ≥ c for a predetermined constant c -> reject H0. If not -> don't reject H0  By doing this, might commit type I error (reject H0 when H0 true) or type II error (don't reject H0 when H0 false)  Rate of 2 errors cannot be controlled at the same time. Neyman-Pearson framework controls the type I error rate at a given level , i.e. choose c s.t. P(T(**X**) ≥ c|H0) ≤ . If true -> reject H0  H0 should be s.t. type I error is more serious. If still not clear, H0 should be a well-established theory | | |
| R fn related to dist | | | d\*, p\*, q\*, r\* computes pdf, cdf, quantiles and generation of random samples respectively  \* can be replaced by abbreviation of name of dist (e.g. norm, t, f, chisq for normal, t-, F-, and -dist)  E.g. dnorm(x, mean=0, sd=1). pnorm(q, mean, sd, lower.tail=TRUE). qnorm(p, mean, sd, lower.tail). rnorm(n, mean, sd)  If upper tail prob and quantiles are needed, then specify lower.tail = FALSE | | |
| R formula specifica-tion | | | lm(y~x1+x2): x1 + x2: c0 + c1x1 + c2x2. -1+x1+x2: c1x1 + c2x2  x1\*x2 (or x1+x2+x1:x2): c0 + c1x1 + c2x2 + dx1x2. x1\*x2+x1\*x3+x2\*x3: c0 + c1x1 + c2x2 + c3x3 + d12x1x2 + d13x1x3 + d23x2x3  offset(x): coefficient of x is fixed at 1  I(): e.g I(x1\*x2), it gives component-wise product of x1 and x2, it is different from the specification x1\*x2. | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Simple Linear Regres-sion | Correlation ≠ causation. Correlation coefficient can only be btw -1 and 1, i.e. -1 ≤ ≤ 1  Pearson's correlation (coefficient). Let X and Y be 2 r.v., the theoretical correlation is =  If (X,Y) has observations (xi, yi), i=1,...,n, the sample correlation is = = corr(X, Y)  If > 0: move in same dir. If < 0: move in opp dir. Both DO NOT imply a causal r/s  If = 0, X and Y have no LINEAR r/s. X and Y could have other kinds of r/s (e.g. quadratic) | | | | |
| Y is the *response variable*. X = *covariate*/predictor  A simple regression model (SRM) is Y = + X + , where , = *regression coefficients*/parameters and is a random error.  E(Y) = + X is called the regression function (E(Y|X) more technically correct way to write it)  W observations (xi, yi), observed simple LRM is yi = + xi + , i = 1,...,n. Error term is diff for ea observation | | | | |
| Assumptions of simple linear regression model (LRM):   |  |  | | --- | --- | | 1. xi and are indep. Note indep un-correlated, not necessary. (Unless normal var, then indep <-> un-correlated) | | | 2. have mean 0 | 3. are pairwise un-correlated, i.e. Cov(, ) = 0 | | 4. have common variance (Homogeneity) | 5. have a Normal dist (Normality) |   LRM is then called a Normal LRM. Assumptions can be simply stated, , iid ~N(0, ) in addition to (1) | | | | |
| From Y = + X + , 1) E(Y) = E( + X + ) = E() + E(X) + E() = + E(X) + 0 = + E(X)  2) Cov(X, Y) = Cov(X, + X + ) = Cov(X, ) + Cov(X, X) + Cov(X, ) = 0 + Cov(X, X) + 0 (since constant, and X, are indep) = Var(X)  So = = = and = - (exact values of , )  (, ) will then be the soln to (i.e. error term minimised)  So regression fn + X is best linear approximation of X to Y | | | | |
| Estima-tion of LRM | LSE of Property (, ) minimizes gives rise to the least square estimation (LSE), which estimate the params by minimizing the sum of squares, Q =  To find LSE of , , = -2, and = -2  Solving normal eqns = 0 and = 0  -2 = 0. n + - = 0. + - = 0 – (1)  -2 = 0. – (2)  Sub (1) into (2): = 0. = 0. = 0. = = (using a) and b))  a) = - = – = – (0)  b) = - = - =  Summary: = , = , where = and = | | | | |
| LSE of = Var() = E() - [E()]2 = E() (since E() = 0, assumption 2). =  Since not observable, we can estimate by  can be estimated using ei = yi . Since we are using estimates for , , i.e. there are only n-2 degree of freedom  SSE (sum of square of error) = = and estimated by = s2 = SSE/(n-2)  Estimator of is is called the residual standard error | | | | |
| Fitted/estimated regression fn: = + X  Fitted values: = , i = 1,...,n | | | Residuals: ei = yi - , i = 1,...,n  Value = + X\* is called the predicted value of a response at X\* | |
| LSE as Method of moments estimation (MME): just replace theoretical moments w sample moments  From = , = - LSE is then = and = - , where a qty w a hat = sample version of that qty | | | | |
| LSE as Maximum Likelihood estimation (MLE): Under the Normality assumption, the log-likelihood is  = MLE of is given by = = s2  Maximizing the log-likelihood to obtain the MLE of the regression coefficients is equivalent to minimizing Q | | | | |
| Variation of Y is estimated by total sum of squares: SST = = + SSR + SSE  SSE is variation explained by X; regression sum of square  SSE is variation caused by random errors; residual sum of squares. Note yi = +  Proof: = = = + + 2 = + = SSR + SSE  Since 2 = 2 – 2 = 2 – 2 = 0 - 0 = 0 (1.20, 1.17 in book) | | | | |
| ANOVA table for SRM   |  |  |  |  |  | | --- | --- | --- | --- | --- | | source | df (deg of freedom) | SS (sum of sq) | MS (mean of sq) | F (f ratio) | | Regression | 1 | SSR | MSR | MSR/MSE | | Error | n-2 | SSE | MSE |  | | Total | n-1 | SST |  |  | | | | | |
| Text, letter  Description automatically generatedCoefficient of determination, R2 = = . R2 is proportion of variation in Y explained by X. Measures the strength of correlation btw Y and X  R2 = corr (Y, )2 = corr(Y, X)2 = = =  Using = , and var(Y) = var( + X + ) = 0 + var(X) + var() (since X, indep), so = + | | | | |
| Adjusted R2: Less biased estimate is given by = R2 - where p is num of predictors and equals 1 for simple LRM  R2 is strictly increasing as p increases. But does not necessarily incr as p incr | | | | |
| = + X. Using R2, can explain how much of variation of Y is due to X  is estimated amt of change in expectation of Y when X incr by an unit amt  Can predict Y\* w new observation X\* using + X\* | | | | |
| X <- c(); Y <- c()  plot(X, Y)  y.fit <- lm(Y~X)  abline(y.fit) # plot LRM on graph  summary(y.fit) # get summary statistics  residuals <- residuals(y.fit) OR y.fit.resi  fitted <- y.fit$fitted  cbind(X, Y, fitted, residuals)  To predict: Xnew <- c()  Ynew <- predict(y.fit, list(X=Xnew)  cbind(Xnew, Ynew) | | = and = , = , =  SST = , SSR = , SSE = SST - SSR, R2 =  Summary results explained:   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | |  | Estimate | | Std. Error | t value | | Pr(>|t|) | | | (Intercept) |  | | s() | = | | p-value for 2-sided test on | | | X |  | | s() | = | | p-value for 2-sided test on | | | Residual standard error: | | | |  | | | | | Multiple R-squared: | | R2 | | Adjusted R-squared: | | |  | | F-statistic: | | MSR/MSE | | p-value | p-value of the significant F test | | | | | |
| From summary results, MSE = , SSE = df (n-2) \* MSE, R2 = = 1 - , (MSR = SSR/df), F-statistic = ()2  Fitted regression function: = + X. We can say Y incr/decr as X incr. When X incr by 1 unit, Y incr/decr by  X explaines about (R2\*100)% of the variation in Y | | | | |
| Theoretical Properties of LSE | | In real life, both Y and X are r.v.  However, regression analysis is a conditional analysis, i.e. conditioning on the observed values of the covariate, i.e. Xi are treated as non-random variables. Thus, more strictly speaking, E(Y|X) = + X, but usually we just write E(Y) = + X | | | |
| Unbiasedness of and . LSE is unbiased, i.e. E() = . and E() =  Proof: = = – = – 0 = = + + = 0 + + = + = +  Hence, E() = E = E + E (since xi are non-random) = E (since E() = 0)  By summing up and divide by n yi = + xi + = + +  Using = , E() = E() = E( + + ) = E( + ) (since E() = ) = (since E() = 0) | | | |
| Variance of and :  var() = var = var (since fraction) = = =  var() = var() = var() + var() – 2cov(, ) = + – 2cov(, ) (since is constant) = + (since cov(, ) = cov(, ) = cov(, ) = ,) = ,) = = 0)  So, var() = , and var() = + | | | |
| Unbiasedness of s2 / as an estimator of :  (n-2)s2 = SSE = = = = = + – 2 = + – 2 = + – 2 = –  So, (n-2)E(s2) = E() – E() = E() – [var() + (E())2] = E() – = E() – –  Also, E() = E() = E() = E() = + E + = + E + 0 = + E – 2E + nE = + n – 2nE + nE = + n – nE = + n –  Hence, (n-2)E(s2) = E() – – = + n – – – = (n-2). Thus Es2 = | | | |
| Properties of estimated/fitted regression fn:  The estimated regression fn = + X is an unbiased estimator of EY = + X, i.e. E() = E(Y) since , are unbiased  var() = var() = var() + var() + 2Cov() = + + 0 = . OR using = + X, var() = var() + 2XCov() + X2var()  For the prediction at a new value X, the prediction mean square error is:  E(Y-)2 = var(Y-) + [E(Y-)]2 = var(Y) + var() + [E( + X + – – X)]2 = var( + X + ) + var() + [ + X + E – E – XE)]2 = var() + var() + [ + X + 0 – – X)]2 (since constant don't affect var) = var() + var() = | | | |
| Dist of LS estimators: Normality condition is assumed. , are LC of Y1, ..., Yn  By property of normal dist, , are normally distributed: ~N, ~N  (n-2) (where s2 = ) ~-dist w df n-2 & is indep from , (verification given later) | | | |
| Statistical Inference for simple LRM | | Significant test: test whether or not there is a linear regression r/s btw the response var & the covariate  Null hypothesis, H0: no such r/s. Alternative hypothesis, H1: r/s exist  test statistic is F-statistic, where F = (intuitively, if variation caused by covariate > variation caused by error r/s exist)  Under H0, F~F-dist with df 1 and n-2, since SSR and SSE indep and each follows a -dist  If F > upper quantile f1, n-1(), H0 is rejected at level , otherwise not rejected OR P(F > f1, n-1()) < = P(type 1 error) | | | |
| t-statistic for the inference on  Dist of cannot be used directly for inference, since its var involves which is unknown  Let s2() = . Define = = = = = t-dist ~ tn-2.  W/o normality assumption, has an asymptotic standard normal dist | | | |
| Two-sided test for : H0: = 0 and H1: ≠ 0  Test-statistic: = ~ tn-2 under H0  Significance level (usually taken as 0.05 or 0.01). If || > tn-2(/2), then reject H0. Otherwise, don't reject H0  OR p-value = p = 2P(|tn-2| > ||) (T is observed value here, above was r.v). If p < , reject H0; otherwise, don't reject H0  For simple LRM, the two-sided test for = p-value of F-statistic. F-statistic here = (t-statistic)2 | | | |
| One-sided test for  1) H0: ≤ 0, H1: > 0. (Equal sign w H0)  If > tn-2(), OR p = P(tn-2 > ): reject H0; otherwise don't  2) H0: ≥ 0, H1: < 0  If < tn-2(), OR p = P(tn-2 < ): reject H0; otherwise don't  For both 2-sided, and 1-sided test, value 0 in hypotheses can be replaced by any constant c, then test statistic = | | | |
| Confidence intervals for  From dist of , P = 1 - , where tn-2(/2) is upper /2 quantile of tn-2 dist  From , the 100(1 - )% confidence interval (CI) for is [ - tn-2(/2)s(), + tn-2(/2)s()]. CI are used for 2-sided tests | | | |
| Confidence bounds for  100(1-)% lower confidence bound: ≥ - tn-2()s(). This corresponds to 1-sided test (H0: ≤ 0, H1: > 0)  100(1-)% upper confidence bound: ≤ + tn-2()s(). This corresponds to 1-sided test (H0: ≥ 0, H1: < 0) | | | |
| MM example: guess muscle mass tends to decrease as ppl ages ( < 0). So H0: ≥ 0, H1: < 0  p-value for one sided test can just divide Pr(>|t|) from R summary table by 2 = P(t14 ≤ ) OR pt(, n-2)  Since p-value smaller than any conventional size (i.e. 0.05 or 0.01), H0 is rejected. guess is confirmed  95% upper confidence bound for is ≤ + tn-2(0.05)s(). (i.e. W prob 0.95, ≤ + tn-2(0.05)s())  tn-2() can be calculated using qt(0.05, 14, lower.tail=FALSE) OR qt(0.95, 14) (which gets the right side instead) | | | |
| The inference on  = , where s2() =  Dist of also tn-2. Inference is similar to . E.g. 95% CI for is ± tn-2(/2)s() | | | |
| Prediction  CI of E(Y) when predictor value is xh: where is the predicted value for xh  Prediction interval of Ynew when predictor value is xh:  predict(model, newdata, interval = c("none", "confidence", "prediction"), level = 0.95) | | | |
| Birth Rate example  BR.dat <- read.table(data,y=,x=); plot(x,y)  BR.fit<-lm(y~x); summary(BR.fit)  b <- BR.fit$coef; V <- vcov(BR.fit) #estimated var-cov matrix  confint(BR.fit, level=0.95) #CI for params  Ynew <- predict(BR.fit, list(x=30), interval="prediction") | | | (i) Scatter plot shows linear r/s btw x and y  Significant F-test can be conducted to answer whether r/s is statistically significant  F = and p = . Since p-value is almost 0, H0 of no r/s is rejected, hence r/s is significant  Proportion of variation in y explained by x is R2 = |
| (ii) Want to determine if > 0  H0: ≤ 0, H1 > 0.  test statistic from sumary results, T = . The p-value is (Pn-2 > T) = ..., (half of p value for 2-sided test)  Since p-value almost 0, H0 rejected, confirmed that > 0 | | | (iii) 95% CI for  Can compute manually ± tn-2()s() OR extract from confint(BR.fit, level=0.95) |
| (iv) Prediction interval given by where var(Ynew) =  var(Ynew) = var() + , where = + xh  Manual computation: var() = var() + var() + 2xhcov(, ). values and can be found from summary results  OR using R: predict(BR.fit, list(x=30), interval="prediction", level=0.95) | | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Multiple Linear Regression | | Matrix form of linear fns: = a1x1 + ... + anxn = **a**T**x**, where **a** = (a1,...,an)T, **x** = (x1,...,xn)T  Matrix form of quadratic fns: = **x**TA**x**, where A =  Matrix form of differentiation: Let f(**x**) be a multivariate fn. Define = ,  then = **a**, and = 2A**x** if A is symmetric, (A + AT)**x** otherwise | | | |
| In multiple LRM, deal w r/s btw response var Y and p predictor variables/covariates **X** = (X1,..., Xp)  Then multiple LRM is Y = + X1 + ... + Xp + , where is random error w mean 0  Then EY = + X1 + ... + Xp is the multiple linear regression fn. (Or more strictly speaking EY is E(Y|X)  For n observations, yi = + + ... + + , i = 1,...,n | | | |
| Let X = , , **y** = ,  **=** .  Then = OR **y** = X+ . X aka design matrix | | | |
| |  |  | | --- | --- | | Assumptions for multiple LRM (similar to simple LRM) | 1. X1,...,Xp and are indep. | | 2. have mean 0 | 3. are pairwise un-correlated, i.e. Cov(, ) = 0 | | 4. have common variance (Homogeneity) | 5. have a Normal dist (Normality) |   Let denote mean of Y, denote mean vector of X = (X1,...,Xp), : covariance vector btw Y and X, : covariance matrix of X and = (,..., ). By assumption1, = , = - | | | |
| Least sq estimate for multiple LRM | | Estimate , , ..., by minimizing Q = = (norm)  = (**y**-X)T(**y**-X) = **y**T**y** - 2**y**TX+ TXTX. Set = 0. - 2XT**y** + 2XTX= 0. (using = **a**, and = 2A**x**). Hence XTX= XT**y**  LSE of is = (XTX)-1XT**y**  = = = | | | |
| Hat matrix & properties | | H = X(XTX)-1XT is the hat matrix of X. It is the projection matrix of the linear space spanned by the cols of X  H = HT, HX = X, H2 = H, (I-H)2 = I-H, XTH = XT  Let **1** = (1,...,1)T, **x**j = (x1j,...,xnj)T. Then HX = X implies H**1** = **1**, H**x**j = **x**j, j = 1,...,p. (since X = (**1** **x**1 ... **x**p)  Residual vector: **e** = **y** - = (I-H)**y**. Vector of fitted values: = H**y**  Hence **e**T**1** = 0 (by assumption 2), **e**T**x**j = 0 (by assumption 1), mean of = | | | |
| Decompo-sition of Sum of Squares | | SST = . Let **y**T = = **y** - **1** = (I – )**y** HT**y**, where = . Thus, SST = **y**T = **y**THT**y**  SSR = . Let **y**R = = - **1** = [X(XTX)-1XT – ]**y** HR**y**. Thus, SSR = **y**R = **y**THR**y**  SSE = . Consider **e** = **y** - = (I – X(XTX)-1XT)**y**  HE**y**. Thus, SSE = **e**T**e** = **y**THE**y**  Matrices HT, HR and HE are all symmetric and idempotent (i.e. HTH = H)  Hence SST = **y**THT**y** = **y**T(I – )**y**. SSR = **y**THR**y** = **y**T[X(XTX)-1XT – ]**y**. SSE = **y**THE**y** = **y**T[I – X(XTX)-1XT]**y**. SST = SSR + SSE | | | |
| Degree of freedom of sum of squares | | SST, SSR, SSE are the squared norms of the n-vectors **y**T, **y**R and **e** respectively  Since **y**T = HT**y**, **y**R = HR**y** and **e** = HE**y**, **y**T R(HT), **y**R R(HR), **e** R(HE), where R(A) denotes subspace of Rn spanned by cols of A  Thus, HT, HR and HE have ranks n-1, p and n-p-1 respectively.  For a subspace R1 of Rn w dimension k, there is a subspace R2 w dimension n-k which is orthogonal to R1. Hence there are n-k indep vectors in R2, say **vj**, j = 1,...,n-k which are perpendicular to vectors in R1. Thus, for any vector **u** in R1, **u**T**v**j = 0, j = 1,...,n-k, thus only k components of **u** can vary freely. So dimension k aka degress of freedom  Thus, degrees of freedom of SST, SSR and SSE are n-1, p and n-1-p respectively | | | |
| Dist of Sum of Squares | Under assumptions of multiple regression models, (Note var(AY) = Avar(Y)AT, cov also similar)  - SSR and SSE are indep, since cov(**y**R, **e**) = cov(HR**y**, HE**y**) = HRvar(**y**) = [X(XTX)-1XT – ][I – X(XTX)-1XT] (since HE symmetric) = [X(XTX)-1XT – X(XTX)-1XT X(XTX)-1XT – + X(XTX)-1XT] = [X(XTX)-1XT – X (XTX)-1XT – + X(X-1(XT)-1)XT] = 0  ~ . Since = **y**T[I – X(XTX)-1XT]**y** = (X+ )T[I – X(XTX)-1XT](X+ ) = [I – X(XTX)-1XT](expand out and can see (X)T \* middle term \* (X) will cancel out) = [I – X(XTX)-1XT], where = / ~N(0,1)  Under hypothesis = ...= = 0, SSR/ ~ (using same argument as above) | | | | |
| ANOVA Table | | |  |  |  |  |  | | --- | --- | --- | --- | --- | | Source of variation | SS | df | MS | F-statistic | | Regression | SSR | p | MSR=SSR/p | MSR/MSE | | Error | SSE | n-p-1 | MSE=SSE/(n-p-1) |  | | Total | SST | n-1 |  |  | | | | |
| Coefficient of multiple determi-nation | | coefficient of multiple determination for multiple LRM is R2 = = = corr(**y**, )2  corr(**y**, ) = and = = **y**T[X(XTX)-1XT - ]**y** = **y**T[X(XTX)-1XT - ](I - )**y** = SSR \* SST = (**y** – ) =  corr(**y**, )2 = = = =  Adjusted R2: = R2 – (1 – R2) | | | |
| Multiple correlation coefficient | | Correlation of response var Y w scalar covariate is measured by the Pearson's correlation coefficient  The multiple correlation coefficient btw response var Y and vector **z** of covariates (X1,...,Xp)T is MCORR(Y, **z**) = where **a** is a vector of constants, (i.e. **a**T**z** is LC of **z**)  = , where = (Cov(Y, X1),...,Cov(Y,Xp)), = variance matrix of **z** and = , = var(Y) | | | |
| Let **y** be vector of n observations of y, Z be the matrix of observed **z** w ith row given by **z**i.  =  In multiple LRM w vector of p covariates **z** = (X1,...,Xp)T, R2 = | | | |
| Partial correlation coefficient | | Let X(-j) denote the sub-matrix of the design matrix X obtained by deleting col **x**j of X  Let be the residual of **y** regressed on X(-j), is residual of **x**j regressed on X(-j)  Correlation btw and is called partial correlation btw **y** and **x**j adjusting for effects of X(-j), given by  CORR(, ) = = where H-j is the hat matrix of X(-j)  Justification: Since = 0, = 0, then = 0, = 0 and  = = = **y**T(I – H-j)**x**j,  = **y**T(I – H-j)**y**, =  Since residual of **y** regressed on X(-j) is part of **y** unexplained by X(-j), the squared partial correlation btw **y** and **x**j adjusting for effects of X(-j) is the proportion of unexplained variation of **y** which is explained by **x**j | | | |
| Explicit expression of | | Explicit expression of can be obtained by considering the minimization of : 1st) minimize wrt with fixed, then minimize wrt , where is the sub-vector of eliminating | | | |
| Minimizing w fixed : = = = (Let Z = , be new response var, w design matrix X(-j). So LSE of is (XTX)-1XT**y**)  = = = = | | | |
| Minimizing wrt : = | | | |
| var() = | | | |
| (Standardization) = / =  Note = CORR(, ) = = = some constant \*  So just by looking at in summary table of R code, can tell which covariate contribute more to variation in y | | | |
| Example | df <- read.table(); attach(df); y <- col1; x1 <- col2; x2 <- col3  df.fit <- lm(y~x1+x2); summary(df.fit), anova(df.fit)  SSx1 = variation of y explained by x1 (ignoring effect of x2)  SSx2 = variation of y explained by x2 (eliminating effect of x1)  SST = SSR + SSE | | | In ANOVA table, the sum of squares (SS) generated are sequential SS, ie. the SS associated w each covariate is the one after adjusting for the effects of the covariate preceding it  E.g. For y~x1+x2: - SS for x1 is SSR of y~x1  - but SS for x2 is ry~rx2, where ry is residual of y regressed on x1 and rx2 is residual of x2 regressed on x1  SSR of y~x1+x2 is sum of all sequential SS | |
| Properties of LSE | | Dist of is normal w mean = , var = (XTX)-1  E = E{(XTX)-1XT**y**} = E{(XTX)-1XT**y**} = E{(XTX)-1XT(X+ )} = E{(XTX)-1(XTX)} = E{} =  var() = var{(XTX)-1XT**y**} = (XTX)-1XTvar(**y**)X(XTX)-1 [since var(AX) = Avar(X)AT)] = (XTX)-1(XTX)(XTX)-1 = (XTX)-1 | | | |
| (n-p-1) ~ , since  (n-p-1) = **y**T(I – H)**y**/ = (X+ )T(I – H)(X+ )/ = T(I – H)/ (some magic happening here)  Since (I-H)2 = I-H, rank of I-H = Tr(I-H) = n-p-1 / ~N(**0**, **1**) | | | |
| and are indep, since  cov(, ) = cov{(XTX)-1XT**y**, [I – X(XTX)-1XT]**y**} = {(XTX)-1XT[I – X(XTX)-1XT]} = 0 (more magic) | | | |
| Let = (XTX)-1. Let denote the jth diagonal elems of (estimated var of )  Note = cjj, where cjj is the jth diagonal elem of (XTX)-1  Then ~ tn-p-1; & For any constant vector **c**, ~ tn-p-1.  (to proof t-dist, just prove numerator is Z, denom is , and both indep) | | | |
| MSR and MSE are indep, since SSR and SSE are indep as shown previously (??) | | | |
| MSR/MSE follows F-dist w df p and n-p-1 under hypothesis that (,...,) = 0, since ~ , ~ | | | |
| Signifi-cance F-test | | Hypothesis: H0: = ... = = 0 vs H1: ≠ 0 for at least one of j = 1,...,p.  Test statistic: F = MSR/MSE. Under H0, F ~ Fp, n-p-1  For a significance level , reject H0 if F ≥ fp,n-p-1() or the p-value P(Fp,n-p-1 ≥ F) < ; otherwise, do not reject H0 | | | |
| Wald test statistic | | Let be a vector of parameters, be its estimator, and the estimated var matrix of  Wald statistic for testing H0: = 0 is given by  Thus Wald test statistic for significance test H0: = 0, where = (,...,)T is given by W = , where is estimator of and is estimated variance matrix of  In context of multiple LRM, F = W/p, where p is dimension of | | | |
| Example2 | | df <- read.table(); attach(df); y <- col1; x1 <- col2; x2 <- col3; x3 <- col4  par(mfrow=c(2,2))  plot(x1,y); plot(x2,y); plot(x3,y)  df.fit <- lm(y~x1+x2+x3); summary(df.fit), anova(df.fit)  r <- df.fit$resid; yhat <- df.fit$fitted; bb <- df.fit$coef; VV <- vcov(df.fit)  b <- bb[-1]; V <- VV[-1,-1]  W <- t(b)%\*%solve(V)%\*%b; F <- W/3 | | | #arrange plots in 2 by 2 format  #vcov for variance matrix  #b: estimates for , V: variance matrix w/o intercept, W: Wald statistic |
| Individual t-test | | Answers qn: given other vars in model, does a particular predictor have a significant effect?  Hypotheses: H0 : = 0 vs H1: ≠ 0.  Test statistic: T = , where is the estimated SD of . Under H0, T ~ tn-p-1  For a significance level , reject H0 if |T| ≥ tn-p-1(/2) or p-value 2P(tn-p-1 ≥ |T|) < ; otherwise do not reject H0  - p-value better than test statistic as not only can reject H0, if p-value very small -> evidence supporting H1 is strong  - 1-sided test also same way as in simple LRM | | | |
| Testing general linear hypothesis | | General linear hypothesis : H0 : = 0  For testing linear hypothesis, test statistic is T = . Under H0, t ~ tn-p-1  If only a few components of **c** are non-zero, T can be simplified.  E.g. **c** = (c1,c2,0,...,0)T, the var becomes var() + var() + 2c1c2cov(, ) and becomes c1 + c2  For a significance level , reject H0 if |T| ≥ tn-p-1(/2) or p-value 2P(tn-p-1 ≥ |T|) < ; otherwise do not reject H0 | | | |
| CI and confidence bound | | A 100(1 - )% CI for is [ – tn-p-1(/2), + tn-p-1(/2)]  A 100(1 - )% CI for is  The 100(1 - )% confidence bounds for are | | | |
| Prediction | | Given a new observation **x**0 = (1, x01,..., x0p)T, predicted value for both Ey0 and y0 is =  Estimated variance of fitted value is = = (XTX)-1  Estimated prediction error variance is = = +  Note y0 = + . But Ey0 = . Thats why prediction have extra error term  CI for Ey0 is ±  Prediction interval for y0 is ± | | | |
| Example2 | | For = 0.05, the upper /2 quantile is qt(0.975, 19). Then can manually compute CI OR confint(df.fit, level = 0.95)  Comparing test statistic: For H1: >. If T statistic > qt(1-, n-p-1) OR pt(T statistic, n-p-1, lower.tail=FALSE), then reject H0  For H1: <. If T statistic < qt(, n-p-1) OR pt(T statistic, n-p-1), then reject H0  Upper 95% confidence bound for  (XTX)-1 is extracted as VV previously  For prediction: xnew <- list(x1=.., x2=.., x3=..) OR xnew <- list(x1=c(), x2=c(), x3=c()) #if have multiple predictor values  predict(df.fit, newdata=xnew, interval='confidence', level=.95) # CI for Ey0  predict(df.fit, newdata-xnew, interval='prediction', level=.95) #PI for y0 | | | |
| Example 3 | data <- read.csv("", header=TRUE)  X <- as.matrix(data[,1:6]); Y <- data[,7]  x1 <- X[,1]; x2 <- X[,2]; ... x6 <- X[,6]  par(mfrow=c(2,3)); plot(x1, Y); ... plot(x6,Y)  data.fit <- lm(Y~x1+x2+...+x6)  summary(data.fit); confint(data.fit, level=0.95)  b <- data.fit$coef  V <- vcov(data.fit)  new\_x <- c(1,2,3,4,5,6)  y\_pred <- t(new\_x)%\*%b  sF <- sqrt(t(new\_x)%\*%V%\*%new\_x)  sP <- sqrt(t(new\_x)%\*%V%\*%new\_x + )  t <- qt(0.975, 78)  CI <- c(y\_pred-sF\*t, y\_pred+sF\*t)  PI <- c(y\_pred-sP\*t, y\_pred+sP\*t) | | #get scatter plots to see r/s  #CI for all coefficients (If CI contains 0, then likely covariate not significant)  #get all estimated coefficients,  #covariance matrix,  #new data  #prediction for new data  #estimated fitted value SD  #estimated prediction error SD  #CI for Ey  #PI for y | | |
| Overall regression is strongly sig, w p-value for F-test ...  Predictors account for roughly (R2) of the variability in Y  x1, x3 are highly sig, w p-values ...  x2 is marginally sig w p-value (close to 0.05) | | | | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Regression models for ANOVA | | Designed experiment: subjects randomly assigned to the diff treatment grps  Observational study: subjects randomly selected from diff treatment grps (could have unbalanced data)  Treatment are given values of a factor variable (called levels of the factor, no meaning e.g. Chinese: 1, Malay:2, Indian:3) | | | | | |
| One-Way ANOVA | One-Way ANOVA model deals w only one factor, A having a levels.  , i = 1,...,a, j=1,...,ni, where yij is the value of Y for the jth member of the ith grp,  and are unknown params, are iid random errors, and ni is num of observations in ith grp  Let be mean response to ith grp. = . For identifiability, we restrict = 0.  Then represents the overall mean and represents effect of ith treatment | | | | | | |
| Answers qn whether factor A has any effect (i.e. any diff btw its levels) by analyzing the components of the variation in Y  SST = SSA + SSE. Total sum of squares = sum of squares attributable to factor A and sum of squares attributable to errors  SST = , SSA = , SSE =  = , = , n = | | | | | | |
| One-way ANOVA table is:   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Source | df | SS | MS | F-test | | A | a-1 | SSA | MSA | MSA/MSE | | Error | n-a | SSE | MSE |  | | Total | n-1 | SST |  |  |   Effect of the factor is tested by the F-statistic, F = MSA/MSE  Under hypothesis = ... = = 0, F~Fa-1,n-a  p-value for F-test use pf(F-statistic, a-1, n-a, lower.tail=FALSE) | | | | | | |
| Two-way ANOVA | Allows analysis of 2 factors at the same time & analysis of interaction of 2 factors  Analyse effect of 2 factors A and B on response variable Y. Suppose A has a levels, B has b levels  , i = 1,...,a, j = 1,...,b, k = 1,...,nij, (gamma: interaction btw the 2 factors) where  = 0, = 0, = = 0  Two-way ANOVA investigates whether there is interaction btw the 2 factors, whether the main effect (ave effects of 1 factor over all the levels of the other factor) is significant | | | | | | |
| SST = SSA + SSB + SSE + SSAB (sum of squares due to interaction of factor A and B)  SST = , SSA = , SSB = ,  SSAB = , SSE =  = , = , = , = ,  = , n = , = | | | | | | |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | Source | df | SS | MS | F-test | | A | a-1 | SSA | MSA | MSA/MSE | | B | b-1 | SSB | MSB | MSB/MSE | | AB | (a-1)(b-1) | SSAB | MSAB | MSAB/MSE | | Error | n-ab | SSE | MSE |  | | Total | n-1 | SST |  |  |   Test for interaction: F = MSAB/MSE. Under hypothesis of no interaction F~F(a-1)(b-1), n-ab  Test for main effect: F1 = MSA/MSE, F2 = MSB/MSE. Under hypothesis of zero main effects, F1~Fa-1, n-ab, F2~Fb-1, n-ab  Factor w zero main effect factor has no effect, unless the interaction effect DNE | | | | | | |
| p-value for F1 = pf(F1, a-1, n-ab, lower.tail=FALSE)  p-value for F2 = pf(F2, b-1, n-ab, lower.tail=FALSE)  p-value for F = pf(F, (a-1)(b-1), n-ab, lower.tail=FALSE)  A and B have a significant interaction effect on Y; the main effect of A is significant, but main effect of B is not | | | | | | |
| Main effect & Contrasts | | If change in 1 var elicits a change in another var, then the var has an effect on the other var  Effect of a factor is the differences of the expected response it causes among its levels  Effect can be measure by contrasts. Let denote expected response at level k.  Contrast is defined as , w = 0, i.e. **c**T**1** = 0, where **c** = (c1,...,ck) is called a contrast vector | | | | | |
| For a factor of a levels, all the contrast form a subspace in the a-dimensional Euclidean space Rawhich is perpendiculat to the vector **1** = (1,...,1)T (since dot product = 0). Denote this subspace by . Any a-1 indep contrast form a basis for  Basis could be: **c**1 = (-1,1,0,...,0)T, ..., **c**a-1 = (-1,0,...,0,1)T, corresponding to , ...,  If factor has no effect, then all contrasts are 0, i.e. , for all **c** | | | | | |
| Interaction Effect | | If effect of factor A is diff when factor B is fixed at diff levels, or equivalently, effect of B is diff when A is fixed at diff levels, then it is said that the 2 factors have an interaction effect  E.g. If A and B only have 2 levels, the effects of A at the 2 levels of B are: , .  If both not same -> there is interaction btw A and B.  Interaction effect is measured by | | | | | |
| In general, effect of factor A at a fixed level, i, of B is measured by any contrasts . If there is at least one pair (i,j) and at least one contrast **c**, s.t. ≠ , then A and B have interaction effect  Interaction effect is measured by interaction contrasts: (i.e. contrast of the contrast)  = , where , | | | | | |
| Interaction contrast vector is of the form (c1d1,...,c1db,...,cad1,...,cadb)  Components of the vector have restrictions: , j=1,...,b; , k=1,...,b;  Among the 1st b restrictions, only b-1 are indep, among 2nd a restrictions, only a-1 are indep, and altogether there are a+b-1 indep restrictions. Thus num of indep interaction contrasts is ab-a-b+1 = (a-1)(b-1) | | | | | |
| Illustration of main and interaction effects | | | Chart, line chart  Description automatically generated | | | | |
| Limitation of ANOVA | | Not convenient for detailed analysis of effects  In two-way ANOVA model, if nij (group sizes) are not the same, the SSAB does not measure the interaction effect, i.e.  in SSAB, the term is not the estimate of an interaction contrast | | | | | |
| ANOVA by LRM | | Factor predictor can be represented by dummy vars. If factor has a levels, it can be represented by a-1 dummy vars  Dummy vars: | | | | | |
| LRM for one-way ANOVA  Model: , expressed in dummy variables uk:  In matrix form: **y** = X +, where **y** = , = , X = . Alternatively, model can be expressed as  Regression params and mean response at ea level have relation: , k=2,...,a  Hence param (which is a contrast) is the diff btw expected values at level k annd level1 | | | | | |
| Example for one-way ANOVA:  y = c(y11,...,y1n,y21,...,y2n,y31,...,y3n,y41,...,y4n), u2 = c(0,...,0,1,...,1,0,...,0,0,...,0), u3 = c(0,...,0,0,...,0,1,...,1,0,...,0), u4 = c(0,...,0,0,...,0,0,...,0,1,...,1); OR package <- factor(c(1,...,1,2,...,2,3,...,3,4,...,4)  reg.fit <- lm(y~u2+u3+u4); OR reg.fit <- lm(y~package)  summary(reg.fit); anova(reg.fit)  For estimates from summary table, note estimate for , ,  Sum of squares of estimator = SSA  By looking at estimates from summary table and p=values, can claim that is significantly larger/smaller than | | | | | |
| LRM for two-way ANOVA  2 cases: (i) no repeated observations at ea level combination, i.e. nij = 1; (ii) nij > 1  (i): only main effects can be analyzed; (ii) both main effects and interaction effects can be analyzed | | | | | |
| Main effect models:  Dummy var for 2 factors: ,  Main effect model:  Design matrix: e.g. let a=b=3, X = = (, u2, u3, v2, v3)  Expectation of the main effect model at diff levels:  = : diff of effect of A at level i and 1 when B is fixed at level 1, which is the same as , the diff of effects of A at level k and 1 when B is fixed at level j. i.e. diff does not depend on j  = , diff of effects of B at level j and 1 when A is fixed at level 1, which is the same as , the diff of effects of B at level j and 1 when A is fixed at level i. i.e. diff does not depend on i | | | | | |
| Interaction model:  Design matrix: e.g. a=b=3: X = ,  Expectation of the main effect model at diff levels:  = : diff of effect of A at level i and 1 when B is fixed at level 1,  = , diff of effects of B at level j and 1 when A is fixed at level 1,  = : diff of effect of A at level i and 1 when B is fixed at level j,  = , diff of effects of B at level j and 1 when A is fixed at level i,  = () – () = , interaction contrast (diff of effect of A at level i and 1 when B is at level j and 1)  So main effect caused by level i and level 1 of A is the average over all the levels of B, i.e. | | | | | |
| Inference on interaction effect | | For testing whether there is a significant overall interaction effect, under the regression model,  H0: = 0, i=2,...,a; j=2,...,b; vs H1: at least one of ≠ 0, where are the basis interaction contrasts  Use F-test statistic to test hypothesis  1) F-test statistic from table produced by anova  2) F-test statistic from computation of Wald statistic: W = , where is vector of estimated 's and is the estimated covariance matrix of . F-statistic is F =  Under H0 F~F(a-1)(b-1), n-ab. Reject H0 at level , if F > F(a-1)(b-1), n-ab () or p-value pf(F, (a-1)(b-1), n-ab, lower.tail=FALSE) < | | | | | |
| For testing of particular interaction effect: perform individual test on 's and on linear combination of 's  E.g. 1) Test whether diff btw level j and level 1 of Factor B is same at level i and level 1 of Factor A  i.e. H0 : () – () = = 0  2) Test whether diff btw level j and level l of Factor B is same at level i and level 1 of Factor A  i.e. H0 : () – () = – = 0  3) Test whether diff btw level j and level k of Factor B is same at level i and level l of Factor A  i.e. H0 : () – () = – – + = 0  General rule of thumb: if subscript of contains 1: ignore; else convert to  Since any particular interaction contrast is a LC of 's, i.e. **c**Twhere is the vector of . Let be the vector of estimated 's and is the estimated covariance matrix of  Test statistic for **c**T= 0 is = . Under H0: **c**T= 0, ~tn-ab  E.g. for a particular interaction contrast, above formula can be simplified. H0: – – + = 0. Only vector = and its covariance matrix are needed. And corresponding **c** reduces to **b** = (1, -1, -1, 1)T | | | | | |
| Example for balanced data & interaction model | | data <- read.table("", header = TRUE)  data$col1 <- factor(data$col1); data$col2 <- factor(data$col2)  model.int <- lm(y~col1\*col2, data = data)  summary(model.int); anova(model.int)  b <- model.int$coef; V <- vcov(model.int)  b.x <- b[s:t]; v.x <- V[s:t, s:t]  W <- t(b.x)%\*%solve(v.x)%\*%b.x; F <- W/4 | | | #convert to factor  #\*: to indicate main and iteraction effect  #only s to t are related to iteraction effect | | |
| Example overall interaction | | The F-test for the interaction effect has a p-value of ... < 0.05, indicating sig of interaction effect btw Factor A and B  E.g. the interaction contrast (correspond to factorA3:factorB2 in summary table) is sig w p-value.... i.e. the diff btw level 2 and 1 of factorB is significantly diff btw level 3 and 1 of factor A | | | | | |
| Example particular interaction | | Test diff btw level 3 and 2 of factor B is sig diff btw level 3 and 2 of factorA  H0: – – + = 0. Since in summary table arranged in , corresponding **c** = (1,-1,-1,1)T  c <- c(1,-1,-1,1), T.c <- t(c) %\*% b.x / sqrt(t(c)%\*%v.x%\*%c), where b.x and v.x are interaction effect obtained above  p-value = 2pt(T.c, n-ab, lower.tail=FALSE) | | | | | |
| Example main effect | | If data is balanced, significance of main effects can be tested using ANOVA table from interaction model  If data not balanced, main effect SS's in anova depends on the order of the factors in lm(). Only last main effect SS is correct. To get the correct SS and corresponding F-values, must fit model twice, switching order. i.e. lm(y~x1\*x2): only correct inference can be made on main effect for x2  Since data balanced in this e.g., order don't matter (both are correct). So can just look at anova table for p-value of F-test | | | | | |
| Example main effect contrast | | Main effect contrasts cannot be conveniently calculated using iteraction model: main effect diff btw level i and 1 of factor A is  Main effect model has correct estimates for main-effect contrasts. But estimate from main-effect model not correct estimate of error variance. SSEM = SSEI + SSAB (error caused by iteraction + interaction effect). Instead, from iteraction model is the correct estimate of error variance | | | | | |
|  | | Let Xm denote design matrix of main-effect model. The estimated var matrix from main-effect model is V.m =  It shld be adjusted to V = = V.m  Test statistic for main-effect contrast can be computed using estimated regression coefficient from main-effect model and the adjusted estimated var matrix | | | | | |
| sigma.I <- summary(model.int)$sigma  model.main <- lm(y~col1+col2, data=data)  V.m <- vcov(model.main)  sigma.M <- summary(model.int)$sigma  V <- V.m\*sigma.I^2/sigma.M^2  b <- model.main$coeff; s.b <- sqrt(diag(V))  t <- b/s.b; p <- 2\*pt(abs(t), n-ab, lower.tail=FALSE)  data.frame(Estimate=b, Std.Error = s.b, t.value=t, p.value=p) | | #  #main effect model  #adjusted estimated var matrix  #getting adjusted parameters  #adjusted summary table | | | |
| Estimate show diff btw level i and level 1 of factor A (if looking at estimate for level i of factor A in adjusted summary table)  So can draw conclusion on the contrast based on the p-value from adjusted summary table  For contrast of level i and level j, i,j ≠ 1, need to compute test statistics using adjusted var matrix  E.g. H0: = 0. Then beta <- b[u:v]; V.b <- V[u:v, u:v]; d <- c(1,-1) (u,v and d depend on how level are arranged in table)  T <- t(d)%\*%beta/sqrt(t(D)%\*%V.b%\*%d); p.b <- 2\*pt(abs(T), n-ab, lower.tail=FALSE)  Then can draw conclusion for 2-sided test for main effect contrast | | | | | |
| Remarks | | Insignificance of the main effect does not imply it has no effect if its iteraction w another factor is significant  When interaction is significant, levels of a factor shld be compared at ea level of the other factor  In general, inference on main effect when iteraction is significant is not very relevant | | | | | |
| Example2 w unbalanced data on iteraction effect | | data0 <- read.csv("", header=TRUE); data <- data0[,c(1,2,3)]  is.na(data); data <- na.omit(data)  factorA <- data$col1;  factorA[factorA=="1"] <- "3"; factorA[factorA=='2'] <- '3'  data$col1 <- factorA  data <- data[col2 != '']  data.int <- lm(y~x1\*x2, data = data)  summary(data.int); anova(data.int)  From p-value in anova table, significance of iteraction is ... | | | | #extracting relevant columns  #check if there is missing values  #grouping some data tgt  #putting col back to df  #removing unwanted rows | |
| Example 2 on main effect | | lm(y~x1\*x2). From anova table, x2 is sig w p-value...  lm(y~x2\*x1). From anova table, x1 is not sig w p-value ...  data.main <- lm(y~x1+x2, data=data)  sigma.I <- summary(data.int)$sigma; sigma.M <- summary(data.main)$sigma  V.m <- vcov(data.main); V <- sigma.I^2/sigma.M^2\*V.m  b <- data.main$coef; s.b <- sqrt(diag(V)); t <- b/s.b; p <- 2\*(pt(abs(t), n-ab, lower.tail=FALSE)  data.frame(Estimate=b, Std.Error=s.b, t.value=t, P.value=p) | | | | | #main effect model  #adjustment of var matrix |
| From looking at data.frame, can tell if level i of factor is sig diff from level 1 of factor from p-value | | | | | |