STAT331

Austin Xia

April 5, 2021

Contents

1 review **3**
- 1.1 variance . . . 3
- 1.2 corvariance . . . 3
- 1.3 Chi-sqaure . . . 3
- 1.4 t-Distribution . . . 3
- 1.5 correlation . . . 4
- 1.6 simple linear regression . . . 4

2 inference **4**
- 2.1 standard error . . . 4
- 2.2 Hypothesis test . . . 5
- 2.3 CI . . . 5

3 Random Vector and Multivariable Normal **5**
- 3.1 inference . . . 6

4 **6**

5 ANOVA, R-squared **7**
- 5.1 Multicolinearity . . . 7

6 Model Building Principles **7**
- 6.1 interpretability . . . 7
- 6.2 Parsimony . . . 7
- 6.3 Goodness of fit . . . 7
- 6.4 Predictive accuracy . . . 8

7 assumption **8**
- 7.1 independence . . . 8
- 7.2 normality . . . 8
- 7.3 heteroskadastcity . . . 8
- 7.4 linearity . . . 8
- 7.5 outliers . . . 8

List of Figures

List of Tables

1 review

1.1 variance

$Var(aY + b) = a^2Var(Y)$
$Var(X + Y) = Var(x) + Var(Y)$

$$Var(y) = \frac{1}{n-1}\sum_{i}^{n}(y_i - \bar{y})^2$$

1.2 corvariance

$cov(X, X) = Var(X)$
$cov(aY + c, bX + d) = abcov(X, Y)$
$cov(U + V, X + Y) = cov(U, X) + cov(U, Y) + cov(V, X) + cov(V, Y)$
For (x_i, y_i), sample corvariance is

$$\frac{1}{n-1}\sum_{i}^{n}(y_i - \bar{y})(x_i - \bar{x})$$

1.3 Chi-sqaure

$X \sim \chi^2_\nu$ with ν degrees of freedom

$$E[x] = \nu$$
$$Var(X) = 2\nu$$

For $Z_i \sim^{iid} N(0, 1)$:

$$X = \sum_{i=1}^{n} Z_i^2 \sim \chi^2_n$$

1.4 t-Distribution

$$Y \sim t_\nu$$

with ν degrees of freedom

$$E[y] = 0$$
$$Var(Y) = 2\nu$$

For independent $Z \sim N(0, 1)$ and $X \sim \chi^2_\nu$

$$\frac{Z}{\sqrt{X/\nu}} \sim t_\nu$$

1.5 correlation

sample correlation

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{\sum_i^n (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum^n (y_i - \bar{y})^2}\sqrt{\sum (x_i - \bar{x})^2}}$$

This is about how closely dots are clustered to a line.
Not about the line itself

1.6 simple linear regression

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \epsilon_i \sim^{iid} N(0, \alpha^2)$$

or

$$y_i \sim^{indep} N(\beta_0 + \beta_1 x_i, \sigma^2)$$

$\beta_1 = \frac{S_{xy}}{S_{xx}}$
$\beta_0 = \bar{y} - \bar{x} * \beta_1$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n e_i^2}{n-2}$$

2 inference

$\hat{\beta}_1 = \sum_{i=1}^n w_i y_i$
so

$$\hat{\beta}_1 \sim N(\sum_{i=1}^n w_i(\beta_0 + \beta_1 x_i), \sigma^2 \sum_{i=1}^n w_i^2)$$

$w = \frac{x_i - \bar{x}}{S_{xx}}$
$E[\hat{\beta}_1] = \beta_1$
$Var(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma/\sqrt{S_{xx}}} \sim N(0,1)$$

$0.95 = P(-1.96 < Z < 1.96)$
when σ is known, we can then calculate a 95% CI for β_1 is $\hat{\beta}_1 \pm 1.96\frac{\sigma}{\sqrt{S_{xx}}}$
the random interval will cover β 95% of time
In practice, σ is not known. from complicated steps (slide 13 lec3) we obtain that

$$\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}/\sqrt{S_{xx}}} \sim t_{(n-2)}$$

2.1 standard error

$$SD(\hat{\beta}_1) = \frac{\sigma}{\sqrt{S_{xx}}}$$

$$SE(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{S_{xx}}}$$

2.2 Hypothesis test

$H_0 : \beta_1 = \theta_0$ against $H_1 : \beta_1 \neq \theta_0$
under H_0 what's probablity of observing a more extreme outcome

2.3 CI

CI for y

$$\frac{\hat{\mu}_0 - \mu_0}{\hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}} \sim t_{n-2}$$

CI for new y

$$\frac{\hat{\mu}_0 - \mu_0}{\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}} \sim t_{n-2}$$

3 Random Vector and Multivariable Normal

Calculus review

- $z = f(y_1...y_k)$, $y = (y_1...y_k)^T$ $\frac{dz}{dy} = \begin{bmatrix} \frac{dz}{dy_1} \\ ... \\ \frac{dz}{dy_k} \end{bmatrix}$
- $z = y^T Ay$ then
$$\frac{dz}{dy} = Ay + A^T y$$

$var(y) = E[(y - \mu)(y - \mu)^T]$ var cov matrix is positive semidefinite

$$a^t V a \leq 0$$

$$Var(Ay) = AVar(y)A^T$$

suppose $z = (z_1...z_n)^T$ iid, $z_i \sim^{iid} N(0, 1)$
then $y = Ay+\mu$ has multivariate normal distribution $y \sim MVN(\mu, \sum)$ where $E[y] = \mu, Var(y) = \sum = AA^T$
property:

- linearity $u = Cy + d \sim MVN(C\mu + d, C \sum C^T)$
- marginal distribution: if $\tilde{y}$ is vector subset of y, then it is MVN-distributed with part of y's varMatrix
- conditional distribution If $u = (y_1, y_2)$ $y_1|y_2$ is MVN-distributed
- independence if $\sum_{ij} = 0$, then y_i and y_j are independent

$$y \sim MVN(X\beta, \sigma^2 I)$$
$$\hat{\beta} = (X^T X)^{-1} X^T y$$
$$E[\hat{\beta}] = \beta$$
$$var[\hat{\beta}] = \sigma^2 (X^T X)^{-1}$$

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 V_{jj})$$
$$V = (X^T X)^{-1}$$
$$\hat{y} = [X(X^T X)^{-1} X^T]y = Hy$$

$$e = (I - H)y \quad \text{with some easy albegra } X^T e = 0$$
$$\hat{\sigma}^2 = \frac{1}{n-p-1} e^T e$$
$$E[e] = 0 \quad e = y - \hat{y}$$

- $e \sim N(0, \sigma^2(I - H))$
- $\hat{\beta}$ and e are independent

if we can show $\frac{1}{\sigma^2} e^T e \sim \chi^2_{n-p-1}$ and it is independent of $\hat{\beta}$

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 V_{jj}}} \sim t_{n-p-1}$$

$$\hat{\beta}_j \pm t_{n-p-1,1-\alpha/2} \hat{\sigma} \sqrt{V_{ii}}$$

note H is symetric and projective(idenpodent)

3.1 inference

$E[\hat{\mu}_0] = x_0 \beta$
$Var[\hat{\mu}_0] = x_0 \sigma^2 (X^T X)^{-1} x_0^T$

$$\frac{\hat{\mu}_0 - \mu_0}{\hat{\sigma}\sqrt{x_0 (X^T X)^{-1} x_0^T}} \sim t_{n-p-1}$$

CI is

$$\hat{\mu}_0 \pm t_{n-p-1} \hat{\sigma} \sqrt{x_0 (X^T X)^{-1} x_0^T}$$

prdiction CI is

$$\frac{\hat{\mu}_{new} - \mu_{new}}{\hat{\sigma}\sqrt{1 + x_0 (X^T X)^{-1} x_0^T}} \sim t_{n-p-1}$$

4

$MeHG|fishpart = N \sim N(\gamma, \sigma^2_N)$

$$\frac{\hat{\beta}_M - \hat{\beta}_{MW}}{SE(\hat{\beta}_M - \hat{\beta}_{MW})} \sim f_{n-p-1}$$

where $var(\hat{\beta}_M - \hat{\beta}_{MW}) = \sigma^2 (V_{3,3} + V_{4,4} - 2Cov(3,4))$
if null hypothese is 3 β are 0,
we construct

$$F = \frac{\hat{\beta}_*^T (V_*)^{-1} (\hat{\beta}^*)}{q\hat{\sigma}^2} \sim F(q, n-p-1)$$

q is number of beta we hypoed, p is number of parameter n is number of data

5 ANOVA, R-squared

$$SSTotal = \sum_{i=1}^{n}(y_i - \bar{y})^2 \quad df = n-1$$

$$SSReg = \sum_{i=1}^{n}(\hat{y} - \bar{y})^2 \quad df = p$$

$$SSRes = \sum_{i=1}^{n}(y_i - \hat{y})^2 \quad df = n-(p+1)$$

SStoal = SSRes+ SSReg
assume beta is 0,

$$F = \frac{SSReg/p}{SSRes/n-(P+1)} \sim F_{p,n-(p+1)}$$

To do F test for a group of corvariates:

$$\frac{(SSreg(Full) - SSReg(Reduced))/q}{SSres/(n-p-1)} \sim F_{q,n-(p+1)}$$

5.1 Multicolinearity

regressing x_j on X_j: $r^2_{y,\hat{y}} = R^2 = \frac{SSreg}{SStotal}$
in simple linear regression,

$$Var(\hat{\beta}) = \frac{\sigma^2}{\sum(x_{ij} - \bar{x})^2}$$

in multiple linear regression,

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{\sum(X_{ij} - \bar{x}_j)^2} * \frac{1}{1-R_j^2}$$

we call $VIF_j = \frac{1}{1-R_j^2}$ Variance Inflation Factor
Intepretation:
$VIF_j = \frac{Var(\hat{\beta}_j)}{Var(\hat{\beta}^*)}$ where β^* is the coefficient vector for a model with an idealized design matrix X^* such that:

- $X_j^* = X_j$
- column space of X^* is same as that of X, hence we'd get same $\hat{\sigma}^2$
- X_J^* is uncorrelated with other elements of X^*

6 Model Building Principles

6.1 interpretability

6.2 Parsimony

6.3 Goodness of fit

- R square
  always increasing, favor large model

- Adjusted R sqaure
  $1 - \frac{SSres/(n-p-1)}{SStot/(n-1)}$
  prefer larger

- Mean sqaure error here is $\hat{\sigma}$, $SSres/(n-p-1)$

- AIC
  $$-2logL(\hat{\theta}) + 2k$$
  prefer lower AIC
  BIC:
  $$-2logL(\hat{\theta}) + 2klog(n)$$

$$R^2_{adj} = R^2 - \frac{p}{n-p-1}(1-R^2)$$

6.4 Predictive accuracy

7 assumption

7.1 independence

7.2 normality

access with: noraml qq plot / studentize residuals

7.3 heteroskadastcity

access with: residuals vs fitted
cure: WLS

7.4 linearity

access with: residuals vs X
or added variable plot res y vs res other should look linear
cure: transformation

7.5 outliers

x outliers: leverage
y outliers: studentized residuals sneaky LOO residuals jackknife residuals has mean near 0 and var near $\frac{1}{(n-p-1)-1}\sum_{n}^{i=1} r^2_{(-i)}$
influence of a single point over our data:

- dffits
- cook's distance
- dfbetas