

M347 TMA04 R2698663

Louis Botterill

08-April-2018

Question 1

1.a

1.a.i

$$\hat{Y}_0 = \hat{\alpha} + \hat{\beta}x_0 \sim N(\alpha + \beta x_0, \sigma^2 A)$$

is the sampling distribution for the estimator $\hat{\alpha} + \hat{\beta}x_0$

This sampling distribution arises due to the fact that we consider Y_1, Y_2, \dots, Y_n to be random and independently sampled from the regression model, which is as follows

$$Y_i = \alpha + \beta x_i + W_i$$

Where

$$W_i \propto N(0, \sigma^2)$$

for $i = 1, 2, 3, \dots, n$.

The general case arises because when we consider the i^{th} fitted value

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}x_i$$

And we can look at the mean and variance of this expression to find the parameters of the normal distribution

$$E[\hat{Y}_i] = E[\hat{\alpha} + \hat{\beta}x_i] = E[\hat{\alpha} + \hat{\beta}]x_i = \alpha + \beta x_i$$

which is as expected, the expected value of the i^{th} fitted value is the i^{th} fitted value on the regression line

Similarly for the variances we can use the expression for linear sums

$$V(aX + bY) = a^2V(X) + 2ab \text{Cov}(X, Y) + b^2V(Y)$$

$$V(\hat{Y}_i) = V(\hat{\alpha} + \hat{\beta}x_i)$$

Using

$$V(\hat{\alpha}) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)$$

$$\text{Cov}(\hat{\alpha}, \hat{\beta}) = -\sigma^2 \frac{\bar{x}}{S_{xx}}$$

$$V(\hat{\beta}) = \frac{\sigma^2}{S_{xx}}$$

$$= \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2 - 2\bar{x}x_i + x_i^2}{s_{xx}} \right)$$

We arrive at the linear combination of the two normally distributed variables $\hat{\alpha}$ and $\hat{\beta}$ such that

\hat{Y}_i is normally distributed with mean and variance as derived above, as follows

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}x_i \sim N\left(\alpha + \beta x_i, \sigma^2 \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}\right)\right)$$

Naturally for the zeroth value x_0 and its associated y_0 fitted value is trivially therefore

$$\hat{Y}_0 = \hat{\alpha} + \hat{\beta}x_0 \sim N\left(\alpha + \beta x_0, \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)\right)$$

\hat{Y}_i - the i^{th} fitted value, which is normally distributed

These values are constants:

- n - the sample size
- x_0 - the zeroth x value which we treat as a constant
- x_i - the i^{th} x value which we treat as a constant
- \bar{x} - the mean of the x values
- $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$ - the sum of squared difference of x_i 's from the mean of x

Where we can define A , a known (derived, calculated based known values of explanatory variables) quantity as

$$A = \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}$$

Allowing us to further simplify to

$$\hat{Y}_0 = \hat{\alpha} + \hat{\beta}x_0 \sim N(\alpha + \beta x_0, \sigma^2 A)$$

Constants:

- A - known, calculated value from known data
- x_0 - the zeroth x value which we treat as a constant
- α - an unknown parameter of the fitted regression line (intercept)
- β - an unknown parameter of the fitted regression line (gradient)

Random variables:

- \hat{Y}_i - the i^{th} fitted value, which is normally distributed, and hence so is \hat{Y}_0
- S^2 is the estimator for σ^2
- $\hat{\alpha}$ - an unbiased MLE of the unknown parameter of the fitted regression line (intercept) and has a normal sampling distribution
- $\hat{\beta}$ - an unbiased MLE of the unknown parameter parameter of the fitted regression line (gradient) and has a normal sampling distribution

1.a.ii

In the linear regression model with one explanatory variable, the estimated residual variance S^2 (an unbiased estimator of σ^2) is independent of $\hat{\alpha}$ and $\hat{\beta}$ and is distributed as

$$S^2 = \frac{R}{n-2}$$
$$R = \sum_{i=1}^n r_i^2$$

So R is the sum of squares of normally distributed residuals, leading to a χ^2 distribution

$$S^2 = \frac{\sigma^2 \chi^2(\nu = n-2)}{n-2}$$

and there are $n-2$ degrees of freedom because of the two unknown parameters, α and β and n known samples

Starting with

$$\hat{\mu} \sim N(\mu, \sigma^2 A)$$

Where A is a known quantity and is defined as

$$A = \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}$$

Using the standard relationship, if

$$X \sim N(\mu, \sigma^2)$$

then

$$\frac{X - \mu}{\sigma} = Z \sim N(0, 1)$$

So that

$$Z = \frac{\hat{\mu} - \mu}{\sigma \sqrt{A}} \sim N(0, 1)$$

And rearranging

$$S^2 = \frac{\sigma^2 \chi^2(\nu = n-2)}{n-2}$$

To give

$$U = S^2 \frac{n-2}{\sigma^2} = \chi^2(\nu = n-2)$$

Using the definition of the student's t distribution

$$T = \sqrt{v} \frac{Z}{\sqrt{\chi^2(\nu)}}$$

and substituting our expressions for Z and χ

$$T = \sqrt{n-2} \frac{(\hat{\mu} - \mu)}{\sigma \sqrt{A}} \times \frac{\sigma}{S \sqrt{n-2}} = \frac{\hat{\mu} - \mu}{\sqrt{AS}} \sim t(n-2)$$

Using this result

$$\frac{\hat{\mu} - \mu}{S\sqrt{A}} \sim t(n-2)$$

where

$$\hat{\mu} = \hat{\alpha} + \hat{\beta}x_0 \text{ and } \mu = \alpha + \beta x_0$$

So the confidence interval is

$$\left(\hat{\alpha} + \hat{\beta}x_0 - c, \hat{\alpha} + \hat{\beta}x_0 + c \right)$$

Where c is

$$t_{1-(p/2)}S\sqrt{A}$$

which is the $(1 - (p/2))$ -quantile of the $t(n-2)$ distribution

and A is defined previously.

So we end up with a statistic that also depends on S. The student's t distribution arises because of the (ratio of the) Normal distribution of $\hat{\mu}$ and the χ^2 distribution of S^2 . Therefore the confidence interval (of $\alpha + \beta x_0$) is given by the percentile of the $t(n-2)$ distribution.

1.b

The chosen prior is a joint uniform distribution

$$f(\alpha, \beta, \tau) \propto \tau^{-1}$$

which is an improper and uninformative prior, leading to Bayesian analysis of the linear regression model with one explanatory variable to have the same results as the classical (frequentist) analysis does.

An outline of why this distribution arises is as follows

Posterior mean of the regression line is

$$E(\alpha + \beta x_0 | \tau, \mathbf{y}) = E(\alpha | \tau, \mathbf{y}) + E(\beta | \tau, \mathbf{y})x_0 = \hat{\alpha} + \hat{\beta}x_0$$

Using $V(X + Y) = V(X) + 2Cov(X, Y) + V(Y)$ for non-independent random variables, the posterior variance for $\alpha + \beta x_0$ conditional on τ is given by:

$$V(\alpha + \beta x_0 | \tau, \mathbf{y}) = V(\alpha | \tau, y) + 2x_0 Cov(\alpha, \beta | \tau, y) + x_0^2 V(\beta | \tau, y)$$

$$V(\alpha + \beta x_0 | \tau, y) = \frac{1}{\tau} \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) - 2x_0 \frac{\bar{x}}{\tau S_{xx}} + x_0^2 \frac{1}{\tau S_{xx}}$$

$$V(\alpha + \beta x_0 | \tau, y) = \frac{1}{\tau} \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)$$

So given that the distribution for the regression line $\alpha + \beta x_0$ (conditional on τ) is a linear combination of normally distributed random variables, it is also a normal distribution with mean and variance as calculated above.

$$\alpha + \beta x_0 | \tau, \mathbf{y} \sim N \left(\hat{\alpha} + \hat{\beta} x_0, \frac{1}{\tau} \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right) \right)$$

Recalling joint conjugacy posterior for normal mean and precision

$$f(\mu, \tau) = f(\mu, \tau) f(\tau)$$

$$\mu | \tau \sim N(a, b/\tau)$$

$$\tau | y \sim \text{Gamma}(c, d)$$

$$\mu, \tau | \mathbf{y} \sim \text{NGamma}(a, b, c, d)$$

The posterior distribution for $\tau | y$ is a gamma distribution with parameters $(n-2)/2$ and $R/2$.

$$\tau | \mathbf{y} \sim \text{Gamma} \left(\frac{n-2}{2}, \frac{R}{2} \right)$$

Therefore the joint posterior distribution of $(\alpha + \beta x_0, \tau) | \mathbf{y}$ is normal-gamma.

$$\alpha + \beta x_0 | \tau, \mathbf{y} \sim \text{NGamma}(a, b, c, d)$$

where

$$\tau | \mathbf{y} \sim \text{Gamma}(c, d)$$

Using the relationship with a relocated and rescaled t-distribution

$$\mu | y \sim t \left(2c; a, \frac{b}{E(\tau | y)} \right)$$

$$2c = (n-2)/2 * 2 = n-2$$

$$a = \hat{\alpha} + \hat{\beta} x_0$$

$$\mu = \alpha + \beta x_0$$

Therefore the marginal distribution for $\alpha + \beta x_0 | \mathbf{y}$ is given by

$$\alpha + \beta x_0 | \mathbf{y} \sim t \left(n-2; \hat{\alpha} + \hat{\beta} x_0, \frac{1}{E(\tau | y)} \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right) \right)$$

where $E(\tau | \mathbf{y}) = \frac{1}{S^2}$

The width of the credible interval for the term $\alpha + \beta x_0$ can be shown to be based on the variance terms which arises in the following distribution:

$$\alpha + \beta x_0 \mid \mathbf{y} \sim t\left(n - 2; \hat{\alpha} + \hat{\beta}x_0, S^2\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)\right)$$

This is a posterior distribution for $\alpha + \beta x_0$ given the data $y = y_1, y_2, \dots, y_n$, a predictive distribution for a future Y_0 at a given x_0 after observing data for y .

In the Bayesian context, we have:

α, β and σ^2 are considered to be random variables.

These random variables have corresponding (point) estimators $\hat{\alpha}, \hat{\beta}$ and S^2 respectively. These are considered as functions of the sample data and so are considered as fixed constants of the posterior distribution.

The point estimators $\hat{\alpha}, \hat{\beta}$ and $1/S^2$ are the same as the estimators of α, β and τ in the classical statistical case.

The meaning of x_0, \bar{x}, n, S_{xx} are the same as per the classical case.

n is the sample size, a known constant

x_0 is the zeroth explanatory variable and hence is a constant

\bar{x} is the mean of the explanatory variable and hence is a constant

As with all Bayesian inference, the distribution of $\alpha + \beta x_0$ arises by combining the (improper) prior distributions with the likelihood (of observed data) using Bayes' Theorem, to give the posterior distribution.

The likelihood function is based on the regression model, the regression model is the same as used for the classical case.

It should be noted the Bayesian posterior using the uniform uninformative prior gives the same conclusion as the classical case, however in practice using informative priors this would not be the case and the Bayesian posterior would combine the prior beliefs with observations to form the most informative conclusion possible from the sources of data available.

Question 2

2.a

$$E(\widehat{\alpha_2 - \alpha_3}) = (a_1 + a_4)\alpha_1 + (a_3 + a_5)\alpha_2 + (a_2 + a_6)\alpha_3 + (a_1 + a_2)\beta_1 + (a_3 + a_4)\beta_2 + (a_5 + a_6)\beta_3$$

This is an unbiased when

$$E(\widehat{\alpha_2 - \alpha_3}) = \alpha_2 - \alpha_3$$

$$E(\widehat{\alpha_2 - \alpha_3}) = (a_1 + a_4)\alpha_1 + (a_3 + a_5)\alpha_2 + (a_2 + a_6)\alpha_3 + (a_1 + a_2)\beta_1 + (a_3 + a_4)\beta_2 + (a_5 + a_6)\beta_3 = \alpha_2 - \alpha_3$$

$$E(\widehat{\alpha_2 - \alpha_3}) = \alpha_2 - \alpha_3 = 0 \times \alpha_1 + 1 \times \alpha_2 + -1 \times \alpha_3 + 0 \times \beta_1 + 0 \times \beta_2 + 0 \times \beta_3$$

The requirements for unbiasedness are found by equating terms

$$a_3 + a_5 = 1$$

$$a_2 + a_6 = -1$$

The other terms are 0

$$a_1 + a_4 = 0, a_1 + a_2 = 0, a_3 + a_4 = 0, a_5 + a_6 = 0$$

2.b

Letting $a_1 = c$ for some arbitrary value of c

$$a_1 + a_4 = 0$$

$$c + a_4 = 0$$

$$a_4 = -c$$

$$a_1 + a_2 = 0$$

$$c + a_2 = 0$$

$$a_2 = -c$$

$$a_2 + a_6 = -1$$

$$-c + a_6 = -1$$

$$a_6 = -1 + c$$

$$a_3 + a_4 = 0$$

$$a_3 - c = 0$$

$$a_3 = c$$

$$a_3 + a_5 = 1$$

$$c + a_5 = 1$$

$$a_5 = 1 - c$$

Hence, for any value of c

$$\widehat{\alpha_2 - \alpha_3} = cY_1 - cY_2 + cY_3 - cY_4 + (1 - c)Y_5 - (1 - c)Y_6$$

is an unbiased estimator of $\alpha_2 - \alpha_3$.

2.c

By setting $c = 0$ in the general form of the unbiased estimator

$$\widehat{\alpha_2 - \alpha_3} = cY_1 - cY_2 + cY_3 - cY_4 + (1 - c)Y_5 - (1 - c)Y_6$$

we get

$$\widehat{\alpha_2 - \alpha_3} = 0Y_1 - 0Y_2 + 0Y_3 - 0Y_4 + 1Y_5 - 1Y_6$$

$$\widehat{\alpha_2 - \alpha_3} = Y_5 - Y_6$$

which therefore confirms that the unbiased estimator $\widehat{\alpha_2 - \alpha_3} = Y_5 - Y_6$, or in other words that $Y_5 - Y_6$ is an unbiased estimator of the difference between treatments B and C, represented by $\alpha_2 - \alpha_3$.

In terms of the experimental design of the treatments and locations we can see that treatment B producing Y_5 and treatment C producing yield Y_6 are both conducted at the same location 3, this eliminating the effects of different locations and tests the effects of the treatments under similar conditions.

The disadvantage of this however is it restricts the analysis to this one location and ignores the other 2 locations and the rest of the yield data which includes the application of treatment C in location 1 producing yield Y_2 and the application of treatment B in location 2 producing yield Y_3 . So not all of the available data is used to provide the best statistical results.

2.d

$$\widehat{\alpha_2 - \alpha_3} = cY_1 - cY_2 + cY_3 - cY_4 + (1 - c)Y_5 - (1 - c)Y_6$$

Using the general formula for uncorrelated variables

$$V\left(\sum_{i=1}^d a_i X_i\right) = \sum_{i=1}^d a_i^2 V(X_i)$$

$$V(\widehat{\alpha_2 - \alpha_3}) = c^2 V(Y_1) + (-c)^2 V(Y_2) + c^2 V(Y_3) + (-c)^2 V(Y_4) + (1 - c)^2 V(Y_5) + (-(1 - c))^2 V(Y_6)$$

Using the variance of Y_i , which is σ by definition and collecting up terms

$$V(\widehat{\alpha_2 - \alpha_3}) = (c^2 + c^2 + c^2 + c^2 + (1 - 2c + c^2) + (c^2 - 2c + 1))\sigma^2$$

$$V(\widehat{\alpha_2 - \alpha_3}) = (6c^2 - 4c + 2)\sigma^2$$

Therefore we have

$$V(\widehat{\alpha_2 - \alpha_3}) = 2(3c^2 - 2c + 1)\sigma^2$$

as required.

2.e

Minimising $v(c) = V(\widehat{\alpha_2 - \alpha_3}) = 2(3c^2 - 2c + 1)\sigma^2$

$$v'(c) = \frac{d}{dc}v(c) = 2(6c - 2) = 12c - 4$$

Where the stationary point at 0 gives

$$12c - 4 = 0 \text{ such that } 12c = 4 \text{ or } c = 4/12 = 1/3$$

To check that this a minimum, we take the 2nd derivative of $v(c), v''(c)$

$$v''(c) = 12$$

which is positive constant (for all c), so that $v(c) = 2(3c^2 - 2c + 1)\sigma^2$ at $c = 1/3$ is a minimum, as required.

The minimum variance estimator would therefore have a variance of

$$2(3(1/3)^2 - 2/3 + 1)\sigma^2$$

$$2(1/3 - 2/3 + 1)\sigma^2$$

$$\frac{2}{3}(1 - 2 + 3)\sigma^2$$

$$\frac{4}{3}\sigma^2$$

The estimator corresponding to this minimum variance is therefore the estimator with $c = 1/3$, which gives

$$\widehat{\alpha_2 - \alpha_3} = (1/3)Y_1 - (1/3)Y_2 + (1/3)Y_3 - (1/3)Y_4 + (1 - 1/3)Y_5 - (1 - 1/3)Y_6$$

$$\widehat{\alpha_2 - \alpha_3} = \frac{1}{3}(Y_1 - Y_2 + Y_3 - Y_4 + 2Y_5 - 2Y_6)$$