STA2101 Assignment 3*

Ahmed Nawaz Amanullah †

January 2021

1

1.1 Problem

Let $X_1, ..., X_n$ be a random sample from a Bernoulli distribution with parameter θ.

1.1.1 (a)

Find the limiting distribution of

$$Z_n = 2\sqrt{n}\left(\sin^{-1}\sqrt{\bar{X}_n} - \sin^{-1}\sqrt{\theta}\right) \tag{1}$$

Hint: $\frac{d}{dx}\sin^{-1}(x) = \frac{1}{\sqrt{1-x^2}}$. The measurements are in radians, not degrees.

1.1.2 (b)

In a coffee taste test, 100 coffee drinkers tasted coffee made with two different blends of coffee beans, the old standard blend and a new blend. We will adopt a Bernoulli model for these data, with θ denoting the probability with

*This assignment was prepared by Jerry Brunner, Department of Statistical Sciences, University of Toronto. The LaTeX source code is available from the course website: http://www.utstat.toronto.edu/~brunner/oldclass/2101f19

†with help from the Overleaf team

a customer will prefer the new blend. Suppose 60 out of 100 consumers preferred the new blend of coffee beans. Using your answer to the first part of this question, test $H_0 : \theta = \frac{1}{2}$ using a variance-stabilized test statistic. Give the value of the test statistic (a number), and state whether the reject H_0 at the usual $\alpha = 0.05$ significance level. In plain, non-statistical language, what do you conclude? This is a statement about preference for types of coffee, and of course you will draw a directional conclusion if possible.

1.1.3 (c)

If the probability of an event is p, the *odds* of the event is (are?) defined $p/(1-p)$. Suppose again that $X_1, ..., X_n$ are a random sample from a Bernoulli distribution with parameter θ. In this case the *log odds* of $X_i = 1$ would be estimated by

$$Y_n = \log \frac{\bar{X}_n}{1 - \bar{X}_n}. \tag{2}$$

Naturally, that's the natural log. Find the approximate large-sample distribution (that is, the asymptotic distribution) of Y_n. It's normal, of course. Your job is to give the approximate (that is, asymptotic) mean and variance of Y_n.

1.1.4 (d)

Again using the Taste Test data, give a 95% confidence interval for the log odds of preferring the new brand. Your answer is a pair of numbers.

1.2 Solution

1.2.1 (a)

My humble idea is to make use of the delta method. According to our lecture slides, if $\sqrt{n}(T_n - \theta) \xrightarrow{d} T$ and $g''(x)$ is continuous in a neighbourhood of θ, then

$$\sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{d} g'(\theta)T \tag{3}$$

Now, in our case, we can say that, by the Central limit theorem, $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \sigma Z \sim N(0, \sigma^2)$, where $\mu = E(X_i) = \theta$ and $\sigma^2 = Var(X_i) = E(X_i^2) - E(X_i)^2 = \theta - \theta^2 = \theta(1-\theta)$. Note also that we are required to find the limiting distribution of $Z_n = 2\sqrt{n}\left(\sin^{-1}\sqrt{\bar{X}_n} - \sin^{-1}\sqrt{\theta}\right) = \sqrt{n}\left(2\sin^{-1}\sqrt{\bar{X}_n} - 2\sin^{-1}\sqrt{\theta}\right)$. So, choosing to apply the delta method, in our case, $T_n = \bar{X}_n$, $\theta = \theta$, $T \sim N(0, \theta(1-\theta))$ and $g(x) = 2\sin^{-1}(\sqrt{x})$ (and so, $g'(x) = 2\frac{1}{\sqrt{1-(\sqrt{x})^2}}\frac{d}{dx}(\sqrt{x}) = \frac{1}{\sqrt{x(1-x)}}$). As a result,

$$Z_n = 2\sqrt{n}\left(\sin^{-1}\sqrt{\bar{X}_n} - \sin^{-1}\sqrt{\theta}\right) \xrightarrow{d} \frac{1}{\sqrt{\theta(1-\theta)}}T \sim N\left(0, \theta(1-\theta)\left(\frac{1}{\sqrt{\theta(1-\theta)}}\right)^2\right) \tag{4}$$

$$Z_n = 2\sqrt{n}\left(\sin^{-1}\sqrt{\bar{X}_n} - \sin^{-1}\sqrt{\theta}\right) \xrightarrow{d} \frac{1}{\sqrt{\theta(1-\theta)}}T \sim N(0, 1) \tag{5}$$

In conclusion, the limiting distribution of $Z_n = 2\sqrt{n}\left(\sin^{-1}\sqrt{\bar{X}_n} - \sin^{-1}\sqrt{\theta}\right)$ is $N(0, 1)$, i.e. the standard normal distribution.

1.2.2 (b)

Presumably the variance-stabilized test statistic is what we were provided in the previous question, i.e. $Z_n = 2\sqrt{n}\left(\sin^{-1}\sqrt{\bar{X}_n} - \sin^{-1}\sqrt{\theta}\right)$. So, we insert the provided data into the above expression to get a **test statistic value of approximately 2.01**. Since this is greater than the critical value of 1.96 for the suggested significance level of $\alpha = 0.05$ we **reject** the null hypothesis $H_0 : \theta = \frac{1}{2}$ at the usual $\alpha = 0.05$ significance level. Thus, we conclude that **there is evidence to suggest customers have a higher chance of preferring the new blend than preferring the old standard blend**.

The code written in R to carry out these computations is disclosed below,

```
# 1 (b)
# test stat
z_n <- 2*sqrt(100)*(asin(sqrt(0.6))-asin(sqrt(0.5)))
# get the critical value
# significance level
```

```r
alpha <- 0.05
# critical value
z <- -qnorm(alpha/2)
reject <- 'no'
better <- 'no'
if(abs(z_n)>z){
        reject <- 'yes'
        if(z_n>0){
                better <- 'yes'
        }
}
# print the test stat and the critical value
sprintf('Z_n: %.2f, Z_crit: %.2f', z_n, z)
sprintf('Reject?: %s, Better?: %s', reject, better)
```

1.2.3 (c)

We will use the same approach that we utilized in (a). As stated earlier, by the Central limit theorem, $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \sigma Z \sim N(0, \sigma^2)$, where $\mu = E(X_i) = \theta$ and $\sigma^2 = Var(X_i) = E(X_i^2) - E(X_i)^2 = \theta - \theta^2 = \theta(1-\theta)$. So, choosing to apply the delta method, in our case, $T_n = \bar{X}_n$, $\theta = \theta$, $T \sim N(0, \theta(1-\theta))$ and $g(x) = 2\log\left(\frac{x}{1-x}\right)$ (and so, $g'(x) = \frac{1-x}{x}\frac{d}{dx}(\frac{x}{1-x}) = \frac{1-x}{x}\frac{(1-x)(1)-x(-1)}{(1-x)^2} = \frac{1}{x(1-x)}$). As a result,

$$\sqrt{n}\left(\log\frac{\bar{X}_n}{1-\bar{X}_n} - \log\frac{\theta}{1-\theta}\right) \xrightarrow{d} \frac{1}{\theta(1-\theta)}T \sim N\left(0, \theta(1-\theta)\left(\frac{1}{\theta(1-\theta)}\right)^2\right) \tag{6}$$

$$\sqrt{n}\left(\log\frac{\bar{X}_n}{1-\bar{X}_n} - \log\frac{\theta}{1-\theta}\right) \xrightarrow{d} \frac{1}{\theta(1-\theta)}T \sim N\left(0, \frac{1}{\theta(1-\theta)}\right) \tag{7}$$

We note that the approximate large sample distribution of $\sqrt{n}\left(\log\frac{\bar{X}_n}{1-\bar{X}_n} - \log\frac{\theta}{1-\theta}\right)$ is $N\left(0, \frac{1}{\theta(1-\theta)}\right)$ and thus,

$$\sqrt{n}\left(\log\frac{\bar{X}_n}{1-\bar{X}_n} - \log\frac{\theta}{1-\theta}\right) \dot{\sim} N\left(0, \frac{1}{\theta(1-\theta)}\right) \tag{8}$$

$$\left(\log \frac{\bar{X}_n}{1-\bar{X}_n} - \log \frac{\theta}{1-\theta}\right) \dot{\sim} N\left(0, \frac{1}{n\theta(1-\theta)}\right) \tag{9}$$

$$Y_n = \log \frac{\bar{X}_n}{1-\bar{X}_n} \dot{\sim} N\left(\log \frac{\theta}{1-\theta}, \frac{1}{n\theta(1-\theta)}\right) \tag{10}$$

So, the approximate large sample distribution of Y_n is $N\left(\log \frac{\theta}{1-\theta}, \frac{1}{n\theta(1-\theta)}\right)$ and the approximate mean and variance of Y_n is $\log \frac{\theta}{1-\theta}$ and $\frac{1}{n\theta(1-\theta)}$ respectively.

1.2.4 (d)

Essentially, we compute $\left(\log \frac{\bar{X}_n}{1-\bar{X}_n} - z_{0.975} \frac{1}{\sqrt{n\bar{X}_n(1-\bar{X}_n)}}, \log \frac{\bar{X}_n}{1-\bar{X}_n} + z_{0.975} \frac{1}{\sqrt{n\bar{X}_n(1-\bar{X}_n)}}\right)$ using the provided data. Carrying out this computation, we obtain an **approximate 95% confidence interval of (0.01, 0.81) for the log odds of preferring the new brand**.

The code written in R to carry out this computation is disclosed below,

```
# 1(d)
# provided sample mean and size
x_bar <- 60/100
n <- 100
# sample standard deviation
y_sd <- 1/sqrt(n*x_bar*(1-x_bar))
# critical value
alpha <- 0.05
z <- -qnorm(alpha/2)
# confidence interval bounds
y_bar <- log(x_bar/(1-x_bar))
lb <- y_bar - z*y_sd
ub <- y_bar + z*y_sd
# print bounds
sprintf('0.95 conf int: (%.2f, %.2f)', lb, ub)
```

2

2.1 Problem

The label on the peanut butter jar says peanuts, partially hydrogenated peanut oil, salt and sugar. But we all know there is other stuff in there too. There is very good reason to assume that the number of rat hairs in a jar of peanut butter has a Poisson distribution with mean λ, because it's easy to justify a Poisson process for how the hairs get into the jars. There is a government standard that says the true expected number rat hairs in a 500g jar may not exceed 8. A sample of thirty 500g jars yield $\bar{X} = 9.2$.

2.1.1 (a)

State the model for this problem.

2.1.2 (b)

What is the parameter space Θ?

2.1.3 (c)

State the null hypothesis in symbols. Because nothing will happen if the number of rat hairs is significantly *less* than the standard, I think we need a one-sided test here.

2.1.4 (d)

Find a variance-stabilizing transformation for the Poisson distribution. You may use the fact that a Poisson has expected value and variance both equal to the parameter λ.

2.1.5 (e)

Using the variance-stabilizing transformation, derive a test statistic that has an approximate normal distribution under H_0.

2.1.6 (f)

Calculate your test statistic for these data. Do you reject the null hypothesis one-sided at $\alpha = 0.05$? Answer Yes or No.

2.1.7 (g)

In plain, non-statistical language, what do you conclude? Your answer is something about peanut butter and rat hairs.

2.2 Solution

2.2.1 (a)

The measurements of number of rat hairs in the thirty 500g jars, $X_1, ..., X_{30}$ are a random sample from the distribution $Po(\lambda)$.

2.2.2 (b)

$$\{(\lambda) : \lambda \geq 0\}. \tag{11}$$

2.2.3 (c)

$$H_0 : \lambda \leq 8. \tag{12}$$

2.2.4 (d)

According to the CLT, we know $\sqrt{n}(\bar{X}_n - \lambda) \xrightarrow{d} T \sim N(0, \lambda)$. Using the delta method we have, $\sqrt{n}(g(\bar{X}_n) - g(\lambda)) \xrightarrow{d} g'(\lambda)T \sim N(0, (g'(\lambda))^2\lambda)$. We note that for our transformation to be variance-stabilizing we want,

$$(g'(\lambda))^2\lambda = 1 \tag{13}$$

$$g'(\lambda) = \frac{1}{\sqrt{\lambda}} \tag{14}$$

That is, we want our transformation to satisfy the differential equation $g'(x) = \frac{1}{\sqrt{x}}$. Proceeding to solving this differential equation,

$$g'(x) = \frac{1}{\sqrt{x}} \tag{15}$$

$$\int g'(x)dx = \int \frac{1}{\sqrt{x}}dx \tag{16}$$

$$g(x) = 2\sqrt{x} + c \tag{17}$$

We obtain the variance-stabilizing transformation as $g(x) = 2\sqrt{x}$, as required.

2.2.5 (e)

This is essentially the delta method result from the previous question, using the transformation derived in the previous problem,

$$\sqrt{n}(g(\bar{X}_n) - g(\lambda)) \xrightarrow{d} g'(\lambda)T \sim N(0, (g'(\lambda))^2\lambda) \tag{18}$$

$$\sqrt{n}(2\sqrt{\bar{X}_n} - 2\sqrt{\lambda}) \xrightarrow{d} \frac{1}{\sqrt{\lambda}}T \sim N(0,1) \tag{19}$$

$$2\sqrt{n}(\sqrt{\bar{X}_n} - \sqrt{\lambda}) \xrightarrow{d} \frac{1}{\sqrt{\lambda}}T \sim N(0,1) \tag{20}$$

Under H_0, $\lambda = 8$, **so the required test statistic that has an approximate standard normal distribution under H_0 is $Y_n = 2\sqrt{n}(\sqrt{\bar{X}_n} - \sqrt{8})$.**

2.2.6 (f)

Plugging in our provided data into our derived test statistic, $Y_n = 2\sqrt{n}(\sqrt{\bar{X}_n} - \sqrt{8})$, we get $Y_n = 2.24$. Since the test statistic value exceeds the critical value of 1.64 at significance level $\alpha = 0.05$, **yes**, we reject the null hypothesis one-sided at $\alpha = 0.05$.

The code written in R to carry out this computation is disclosed below,

```
# 2 (f)
# Provided sample mean
X_bar <- 9.2
# sample size
n <- 30
```

```
# null hypothesis
lambda_o <- 8
# test statistic
z_t <- sqrt(n)*(2*sqrt(X_bar)-2*sqrt(lambda_o))
# critical value
alpha <- 0.05
z_c <- -qnorm(alpha) # note that this is 1-sided
# display computation and test results
reject <- 'no'
if(z_t>z_c){
        reject <- 'yes'
}
# print the test stat and the critical value
sprintf('z_t: %.2f, z_c: %.2f', z_t, z_c)
sprintf('Reject?: %s', reject)
```

2.2.7 (g)

There is evidence to suggest that the true expected number of rat hairs in a 500g peanut butter jar exceeds 8.

3

3.1 Problem

If the $p \times 1$ random vector $\mathbf{x}$ has variance-covariance matrix $\mathbf{\Sigma}$ and $\mathbf{A}$ is an $m \times p$ matrix of constants, prove that the variance-covariance matrix of $\mathbf{Ax}$ is $\mathbf{A\Sigma A}^T$. Start with the definition of a variance-covariance matrix:

$$cov(\mathbf{Z}) = E(\mathbf{Z} - \boldsymbol{\mu}_z)(\mathbf{Z} - \boldsymbol{\mu}_z)^T. \tag{21}$$

3.2 Solution

$$cov(\mathbf{Ax}) = E[(\mathbf{Ax} - E(\mathbf{Ax}))(\mathbf{Ax} - E(\mathbf{Ax}))^T] \tag{22}$$

$$cov(\mathbf{Ax}) = E[(\mathbf{Ax} - \mathbf{A}E(\mathbf{x}))(\mathbf{Ax} - \mathbf{A}E(\mathbf{x}))^T] \tag{23}$$

$$cov(\mathbf{Ax}) = E[(\mathbf{A}(\mathbf{x} - E(\mathbf{x})))(\mathbf{A}(\mathbf{x} - E(\mathbf{x})))^T] \tag{24}$$
$$cov(\mathbf{Ax}) = E[\mathbf{A}(\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))^T \mathbf{A}^T] \tag{25}$$
$$cov(\mathbf{Ax}) = \mathbf{A}E[(\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))^T]\mathbf{A}^T \tag{26}$$
$$cov(\mathbf{Ax}) = \mathbf{A\Sigma A}^T \tag{27}$$

N.B. In our proof we used the formula $(AB)^T = B^T A^T$.

4

4.1 Problem

If the $p \times 1$ random vector $\mathbf{x}$ has mean $\boldsymbol{\mu}$ and variance-covariance matrix $\mathbf{\Sigma}$, show $\mathbf{\Sigma} = E(\mathbf{xx}^T) - \boldsymbol{\mu\mu}^T$.

4.2 Solution

$$\mathbf{\Sigma} = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T] \tag{28}$$
$$\mathbf{\Sigma} = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x}^T - \boldsymbol{\mu}^T)] \tag{29}$$
$$\mathbf{\Sigma} = E(\mathbf{xx}^T - \mathbf{x}\boldsymbol{\mu}^T - \boldsymbol{\mu}\mathbf{x}^T + \boldsymbol{\mu\mu}^T) \tag{30}$$
$$\mathbf{\Sigma} = E(\mathbf{xx}^T) - E(\mathbf{x})\boldsymbol{\mu}^T - \boldsymbol{\mu}E(\mathbf{x}^T) + \boldsymbol{\mu\mu}^T \tag{31}$$
$$\mathbf{\Sigma} = E(\mathbf{xx}^T) - \boldsymbol{\mu\mu}^T - \boldsymbol{\mu\mu}^T + \boldsymbol{\mu\mu}^T \tag{32}$$
$$\mathbf{\Sigma} = E(\mathbf{xx}^T) - \boldsymbol{\mu\mu}^T \tag{33}$$

5

5.1 Problem

Let the $p \times 1$ random vector $\mathbf{x}$ have mean $\boldsymbol{\mu}$ and variance-covariance matrix $\mathbf{\Sigma}$, and let $\mathbf{c}$ be a $p \times 1$ vector of constants. Find $cov(\mathbf{x} + \mathbf{c})$. Show your work.

5.2 Solution

$$cov(\mathbf{x} + \mathbf{c}) = E[(\mathbf{x} + \mathbf{c} - E(\mathbf{x} + \mathbf{c}))(\mathbf{x} + \mathbf{c} - E(\mathbf{x} + \mathbf{c}))^T] \tag{34}$$

$$cov(\mathbf{x}+\mathbf{c}) = E[(\mathbf{x}-E(\mathbf{x}))(\mathbf{x}-E(\mathbf{x}))^T] \tag{35}$$

$$cov(\mathbf{x}+\mathbf{c}) = cov(\mathbf{x}) = \mathbf{\Sigma} \tag{36}$$

6

6.1 Problem

Let the $p \times 1$ random vector $\mathbf{x}$ have mean $\boldsymbol{\mu}$ and variance-covariance matrix $\mathbf{\Sigma}$; let $\mathbf{A}$ be a $q \times p$ matrix of constants and let $\mathbf{B}$ be an $r \times p$ matrix of constants. Derive a nice simple formula for $cov(\mathbf{Ax}, \mathbf{Bx})$.

6.2 Solution

$$cov(\mathbf{Ax}, \mathbf{Bx}) = E[(\mathbf{Ax}-E(\mathbf{Ax}))(\mathbf{Bx}-E(\mathbf{Bx}))^T] \tag{37}$$

$$cov(\mathbf{Ax}, \mathbf{Bx}) = E[(\mathbf{A}(\mathbf{x}-E(\mathbf{x})))(\mathbf{B}(\mathbf{x}-E(\mathbf{x})))^T] \tag{38}$$

$$cov(\mathbf{Ax}, \mathbf{Bx}) = E[\mathbf{A}(\mathbf{x}-E(\mathbf{x}))(\mathbf{x}-E(\mathbf{x}))^T\mathbf{B}^T] \tag{39}$$

$$cov(\mathbf{Ax}, \mathbf{Bx}) = \mathbf{A}E[(\mathbf{x}-E(\mathbf{x}))(\mathbf{x}-E(\mathbf{x}))^T]\mathbf{B}^T \tag{40}$$

$$cov(\mathbf{Ax}, \mathbf{Bx}) = \mathbf{A\Sigma B}^T \tag{41}$$

N.B. In our proof we used the formula $(AB)^T = B^T A^T$.

7

7.1 Problem

Let $\mathbf{x}$ be a $p \times 1$ random vector with mean $\boldsymbol{\mu}_x$ and variance-covariance matrix $\mathbf{\Sigma}_x$, and let $\mathbf{y}$ be a $q \times 1$ random vector with mean $\boldsymbol{\mu}_y$ and variance-covariance matrix $\mathbf{\Sigma}_y$. Let $\mathbf{\Sigma}_{xy}$ denote the $p \times q$ matrix $cov(\mathbf{x}, \mathbf{y}) = E((\mathbf{x}-E(\mathbf{x}))(\mathbf{y} - E(\mathbf{y}))^T)$.

7.1.1 (a)

What is the (i, j) element of $\mathbf{\Sigma}_{xy}$? You don't to know any work; just write down the answer.

7.1.2 (b)

Find an expression for $cov(\mathbf{x}+\mathbf{y})$ in terms of $\mathbf{\Sigma}_x$, $\mathbf{\Sigma}_y$ and $\mathbf{\Sigma}_{xy}$. Show your work.

7.1.3 (c)

Simplify further for the special case where $Cov(X_i, Y_j) = 0$ for all i and j.

7.1.4 (d)

Let $\mathbf{c}$ be a $p \times 1$ vector of constants and $\mathbf{d}$ be a $q \times 1$ vector of constants. Find $cov(\mathbf{x}+\mathbf{c}, \mathbf{y}+\mathbf{d})$. Show your work.

7.2 Solution

7.2.1 (a)

$$E((x_i - \mu_{x,i})(y_j - \mu_{y,j})). \tag{42}$$

7.2.2 (b)

$$cov(\mathbf{x}+\mathbf{y}) = E[(\mathbf{x}+\mathbf{y} - E(\mathbf{x}+\mathbf{y}))(\mathbf{x}+\mathbf{y} - E(\mathbf{x}+\mathbf{y}))^T] \tag{43}$$
$$cov(\mathbf{x}+\mathbf{y}) = E[(\mathbf{x}+\mathbf{y} - E(\mathbf{x}) - E(\mathbf{y}))(\mathbf{x}^T+\mathbf{y}^T - E(\mathbf{x})^T - E(\mathbf{y})^T)] \tag{44}$$
$$cov(\mathbf{x}+\mathbf{y}) = E[(\mathbf{x}+\mathbf{y} - \boldsymbol{\mu}_x - \boldsymbol{\mu}_y)(\mathbf{x}^T+\mathbf{y}^T - \boldsymbol{\mu}_x^T - \boldsymbol{\mu}_y^T)] \tag{45}$$
$$cov(\mathbf{x}+\mathbf{y}) = E[((\mathbf{x} - \boldsymbol{\mu}_x) + (\mathbf{y} - \boldsymbol{\mu}_y))((\mathbf{x} - \boldsymbol{\mu}_x)^T + (\mathbf{y} - \boldsymbol{\mu}_y)^T)] \tag{46}$$
$$cov(\mathbf{x}+\mathbf{y}) = \mathbf{\Sigma}_x + \mathbf{\Sigma}_y + E[(\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{y} - \boldsymbol{\mu}_y)^T] + E[(\mathbf{y} - \boldsymbol{\mu}_y)(\mathbf{x} - \boldsymbol{\mu}_x)^T] \tag{47}$$
$$cov(\mathbf{x}+\mathbf{y}) = \mathbf{\Sigma}_x + \mathbf{\Sigma}_y + \mathbf{\Sigma}_{xy} + \mathbf{\Sigma}_{xy}^T \tag{48}$$

7.2.3 (c)

If $Cov(X_i, Y_j) = 0$ for all i and j, then $\mathbf{\Sigma}_{xy} = \mathbf{0}$. So,

$$cov(\mathbf{x}+\mathbf{y}) = \mathbf{\Sigma}_x + \mathbf{\Sigma}_y + \mathbf{\Sigma}_{xy} + \mathbf{\Sigma}_{xy}^T \tag{49}$$
$$cov(\mathbf{x}+\mathbf{y}) = \mathbf{\Sigma}_x + \mathbf{\Sigma}_y + \mathbf{0} + \mathbf{0} \tag{50}$$
$$cov(\mathbf{x}+\mathbf{y}) = \mathbf{\Sigma}_x + \mathbf{\Sigma}_y \tag{51}$$

7.2.4 (d)

$$cov(\mathbf{x}+\mathbf{c}, \mathbf{y}+\mathbf{d}) = E[(\mathbf{x}+\mathbf{c}-E(\mathbf{x}+\mathbf{c}))(\mathbf{y}+\mathbf{d}-E(\mathbf{y}+\mathbf{d}))^T] \tag{52}$$

$$cov(\mathbf{x}+\mathbf{c}, \mathbf{y}+\mathbf{d}) = E[(\mathbf{x}-E(\mathbf{x}))(\mathbf{y}-E(\mathbf{y}))^T] = \mathbf{\Sigma}_{xy} \tag{53}$$

8

8.1 Problem

Let $\mathbf{x} = (X_1, X_2, X_3)^T$ be multivariate normal with

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 0 \\ 6 \end{pmatrix} \text{ and } \mathbf{\Sigma} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{54}$$

Let $Y_1 = X_1 + X_2$ and $Y_2 = X_2 + X_3$. Find the joint distribution of Y_1 and Y_2.

8.2 Solution

According to the properties of multivariate normal distribution (particularly linear combinations of multivariate normals being multivariate normal) we know that if $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma})$ and $\mathbf{A}$ is a matrix of constants, then $\mathbf{Ax} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A\Sigma A}^T)$. We have been provided the linear combinations,

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 + X_2 \\ X_2 + X_3 \end{pmatrix} \tag{55}$$

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \tag{56}$$

So plugging in $\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$ into what we have above, we get a joint distribution of,

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 6 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}\right) \tag{57}$$

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 1 \\ 6 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 0 \\ 0 & 2 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}\right) \tag{58}$$

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 1 \\ 6 \end{pmatrix}, \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}\right) \tag{59}$$

9

9.1 Problem

Let X_1 be Normal(μ_1, σ_1^2) and X_2 be Normal(μ_2, σ_2^2), independent of X_1. What is the joint distribution of $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$? What is required for Y_1 and Y_2 to be independent? Hint: Use matrices.

9.2 Solution

X_1 and X_2 being independent normals, we can set them up as a multivariate normal with zero covariance (making use of the multivariate normal property of zero covariance implying independence),

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right) \tag{60}$$

We have been provided with the linear combinations,

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 + X_2 \\ X_1 - X_2 \end{pmatrix} \tag{61}$$

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \tag{62}$$

Now, as we did in the previous question, making use of the multivariate normal property of linear combinations of multivariate normals being multivariate normal, we get the required joint distribution,

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\right) \tag{63}$$

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 + \mu_2 \\ \mu_1 - \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_2^2 \\ \sigma_1^2 & -\sigma_2^2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\right) \tag{64}$$

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 + \mu_2 \\ \mu_1 - \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 + \sigma_2^2 & \sigma_1^2 - \sigma_2^2 \\ \sigma_1^2 - \sigma_2^2 & \sigma_1^2 + \sigma_2^2 \end{pmatrix}\right) \tag{65}$$

Again making use of multivariate normal property of zero covariance implying independence, we want the covariance of Y_1 and Y_2 to be zero for them to be independent, i.e. $cov(Y_1, Y_2) = \sigma_1^2 - \sigma_2^2 = 0$ or $\sigma_1^2 = \sigma_2^2$.

10

10.1 Problem

Show that if $\mathbf{w} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}$ positive definite, $Y = (\mathbf{w}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{w}-\boldsymbol{\mu})$ has a chi-squared distribution with p degrees of freedom.

10.2 Solution

First, recall *spectral decomposition* for a symmetric matrix (which $\boldsymbol{\Sigma}$ is) from assignment 2, that is, $\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^T$, where $\mathbf{P}$ is a matrix whose columns are the orthonormal eigenvectors of $\boldsymbol{\Sigma}$, $\boldsymbol{\Lambda}$ is a diagonal matrix of the corresponding eigenvalues, and $\mathbf{P}^T\mathbf{P} = \mathbf{P}\mathbf{P}^T = \mathbf{I}$. Next, we recall the use of square root matrices in the same assignment to have,

$$\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^T \tag{66}$$

$$\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\mathbf{P}^T \tag{67}$$

$$\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}^{1/2}\mathbf{I}\boldsymbol{\Lambda}^{1/2}\mathbf{P}^T \tag{68}$$

$$\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}^{1/2}\mathbf{P}^T\mathbf{P}\boldsymbol{\Lambda}^{1/2}\mathbf{P}^T \tag{69}$$

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2} \tag{70}$$

So, we have $\boldsymbol{\Sigma}^{1/2} = \mathbf{P}\boldsymbol{\Lambda}^{1/2}\mathbf{P}^T$, with $\boldsymbol{\Lambda}^{1/2}$ being a diagonal matrix of the square root of the corresponding eigenvalues. Through a similar way, we come up with the square root of the inverse of $\mathbf{\Sigma}$ (which exists because $\mathbf{\Sigma}$ is positive definite and has positive eigenvalues as a result). Firstly, we note that $\mathbf{\Sigma}^{-1} = \mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}^T$ (where $\mathbf{\Lambda}^{-1}$ is a diagonal matrix of the reciprocals of the corresponding eigenvalues),

$$\mathbf{\Sigma}\mathbf{\Sigma}^{-1} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^T\mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}^T = \mathbf{P}\mathbf{\Lambda}\mathbf{I}\mathbf{\Lambda}^{-1}\mathbf{P}^T = \mathbf{P}\mathbf{I}\mathbf{P}^T = \mathbf{I} \tag{71}$$

Next we observe that,

$$\mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{P}^T\mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{P}^T = \mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{I}\mathbf{\Lambda}^{-1/2}\mathbf{P}^T = \mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}^T = \mathbf{\Sigma}^{-1} \tag{72}$$

From which we get our square root of $\mathbf{\Sigma}^{-1}$, $(\mathbf{\Sigma}^{-1})^{1/2} = \mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{P}^T$, where $\mathbf{\Lambda}^{-1/2}$ is a diagonal matrix of the reciprocals of the square roots of the corresponding eigenvalues. We also note that,

$$\mathbf{\Sigma}^{1/2}(\mathbf{\Sigma}^{-1})^{1/2} = \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}^T\mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{P}^T = \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{I}\mathbf{\Lambda}^{-1/2}\mathbf{P}^T = \mathbf{P}\mathbf{I}\mathbf{P}^T = \mathbf{I} \tag{73}$$

Which makes $(\mathbf{\Sigma}^{-1})^{1/2}$ the inverse of $\mathbf{\Sigma}^{1/2}$, justifying the notation $\mathbf{\Sigma}^{-1/2}$ for $(\mathbf{\Sigma}^{-1})^{1/2}$. Now, making use of the multivariate normal property of linear combinations of multivariate normals being multivariate normal,

$$\mathbf{x} = \mathbf{w} - \boldsymbol{\mu} \sim N_p(\mathbf{0}, \mathbf{\Sigma}) \tag{74}$$

$$\mathbf{z} = \mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{x} \sim N_p(\mathbf{0}, \mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{\Sigma}(\mathbf{\Sigma}^{-\frac{1}{2}})^T) = N_p(\mathbf{0}, \mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{\Sigma}^{-\frac{1}{2}}) = N_p(\mathbf{0}, \mathbf{I}) \tag{75}$$

Using the multivariate normal property of zero covariance implying independence, we know that this makes $\mathbf{z}$ a vector of p independent standard normals, and thus,

$$Y = (\mathbf{w} - \boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{w} - \boldsymbol{\mu}) = \mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x} = \mathbf{z}^T\mathbf{z} = \sum_{j=1}^{p} z_i^2 \sim \chi^2(p) \tag{76}$$

as required (the last part of our proof uses the information that the sum of squares of p independent standard normal random variables has a chi-squared distribution with p degrees of freedom).

11

11.1 Problem

You know that if $\mathbf{w} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\mathbf{A}\mathbf{w} + \mathbf{c} \sim N_r(\mathbf{A}\boldsymbol{\mu} + \mathbf{c}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$. Use this result to obtain the distribution of the sample mean under normal random sampling. That is, let $X_1, ..., X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. Find the distribution of $\bar{X}$. You might want to use $\mathbf{1}$ to represent an $n \times 1$ column of ones.

11.2 Solution

Using the multivariate normal property of zero covariance implying independence, we can setup a $n \times 1$ random vector $\mathbf{x} = (X_1, ..., X_n)^T$ such that $\mathbf{x} \sim N_n(\mu\mathbf{1}, \sigma^2\mathbf{I})$. Since $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i = \frac{1}{n}\mathbf{1}^T\mathbf{x}$, we thus have,

$$\bar{X} = \frac{1}{n}\mathbf{1}^T\mathbf{x} \sim N\left(\frac{1}{n}\mathbf{1}^T(\mu\mathbf{1}), \left(\frac{1}{n}\mathbf{1}^T\right)(\sigma^2\mathbf{I})\left(\frac{1}{n}\mathbf{1}^T\right)^T\right) = N\left(\mu, \frac{\sigma^2}{n}\right) \quad (77)$$

as required.

12

12.1 Problem

Let $X_1, ..., X_n$ be independent and identically distributed random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$.

12.1.1 (a)

Show $Cov(\bar{X}, (X_j - \bar{X})) = 0$ for $j = 1, ..., n$.

12.1.2 (b)

Why does this imply if $X_1, ..., X_n$ are normal, $\bar{X}$ and S^2 are independent?

12.2 Solution

12.2.1 (a)

$$Cov(\bar{X}, (X_j - \bar{X})) = E[(\bar{X} - E(\bar{X}))((X_j - \bar{X}) - E(X_j - \bar{X}))] \tag{78}$$

$$Cov(\bar{X}, (X_j - \bar{X})) = E[(\bar{X} - E(\bar{X}))(X_j - E(X_j)) - (\bar{X} - E(\bar{X}))(\bar{X} - E(\bar{X}))] \tag{79}$$

$$Cov(\bar{X}, (X_j - \bar{X})) = Cov(\bar{X}, X_j) - Var(\bar{X}) \tag{80}$$

$$Cov(\bar{X}, (X_j - \bar{X})) = \frac{1}{n}\sum_{i=1}^{n} Cov(X_i, X_j) - Var(\bar{X}) \tag{81}$$

$$Cov(\bar{X}, (X_j - \bar{X})) = \frac{1}{n}\sum_{i=1}^{n} Cov(X_i, X_j) - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j) \tag{82}$$

$$Cov(\bar{X}, (X_j - \bar{X})) = \frac{1}{n}\sigma^2 - \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) \tag{83}$$

$$Cov(\bar{X}, (X_j - \bar{X})) = \frac{1}{n}\sigma^2 - \frac{1}{n^2}n\sigma^2 \tag{84}$$

$$Cov(\bar{X}, (X_j - \bar{X})) = \frac{1}{n}\sigma^2 - \frac{1}{n}\sigma^2 = 0 \tag{85}$$

N.B. The expansions in the fourth and fifth lines of our proofs are the result of a similar procedure as was employed in the first three lines of the proof to obtain $Cov(\bar{X}, X_j) - Var(\bar{X})$ from $Cov(\bar{X}, (X_j - \bar{X}))$. Additionally the simplification in the sixth line comes from the provided information that $X_1, ..., X_n$ are independent.

12.2.2 (b)

Using the multivariate normal property of zero covariance implying independence, we have,

$$\mathbf{x} = (X_1, ..., X_n)^T \sim N(\mu\mathbf{1}, \sigma^2 I) \tag{86}$$

Next, consider $\mathbf{y} = (X_1 - \bar{X}, X_2 - \bar{X}, ..., X_{n-1} - \bar{X}, \bar{X})^T = \mathbf{Ax}$, where

$$\mathbf{A} = \begin{pmatrix} 1-\frac{1}{n} & -\frac{1}{n} & ... & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & ... & -\frac{1}{n} \\ \vdots & \vdots & \vdots & \vdots \\ -\frac{1}{n} & ... & 1-\frac{1}{n} & -\frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \frac{1}{n} \end{pmatrix} \tag{87}$$

$\mathbf{y}$ is multivariate normal because of the multivariate normal property of linear combinations of multivariate normals being multivariate normal. Next, partition $\mathbf{y} = \mathbf{Ax} = (X_1 - \bar{X}, X_2 - \bar{X}, ..., X_{n-1} - \bar{X}|\bar{X})^T = (\mathbf{y}_2|\bar{X})^T$, and using our result from the previous question (i.e. $Cov(\bar{X}, (X_j - \bar{X})) = 0$ for $j = 1, ..., n$) we know that $\bar{X}$ and $\mathbf{y}_2$ are independent because of the multivariate normal property of zero covariance implying independence. We also note that

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{88}$$

$$S^2 = \frac{1}{n}\left(\sum_{i=1}^{n-1}(X_i - \bar{X})^2 + (X_n - \bar{X})^2\right) \tag{89}$$

$$S^2 = \frac{1}{n}(\mathbf{y}_2^T\mathbf{y}_2 + (X_n - \bar{X})^2) \tag{90}$$

$$S^2 = \frac{1}{n}\left(\mathbf{y}_2^T\mathbf{y}_2 + \left(X_n + \sum_{i=1}^{n-1}X_i - \sum_{i=1}^{n-1}X_i - \bar{X}\right)^2\right) \tag{91}$$

$$S^2 = \frac{1}{n}\left(\mathbf{y}_2^T\mathbf{y}_2 + \left(X_n + (n\bar{X} - X_n) - \sum_{i=1}^{n-1}X_i - \bar{X}\right)^2\right) \tag{92}$$

$$S^2 = \frac{1}{n}\left(\mathbf{y}_2^T\mathbf{y}_2 + \left((n-1)\bar{X} - \sum_{i=1}^{n-1}X_i\right)^2\right) \tag{93}$$

$$S^2 = \frac{1}{n}\left(\mathbf{y}_2^T\mathbf{y}_2 + \left(-\mathbf{1}^T\mathbf{y}_2\right)^2\right) \tag{94}$$

$$S^2 = \frac{1}{n}\left(\mathbf{y}_2^T\mathbf{y}_2 + \left(\mathbf{1}^T\mathbf{y}_2\right)^2\right) = g(\mathbf{y}_2) \tag{95}$$

Since $\bar{X}$ and $\mathbf{y}_2$ are independent, $\bar{X}$ and $S^2 = g(\mathbf{y}_2)$ are independent. Thus we have shown that our result from (a) imply that if $X_1, ..., X_n$ are normal, $\bar{X}$ and S^2 are independent, as required.

13

13.1 Problem

Consider the usual multiple regression model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{X}$ is an $n \times p$ matrix of known constants with linearly independent columns, β is a $p \times 1$ vector of unknown constants, and ϵ is multivariate normal with mean zero and covariance matrix $\sigma^2\mathbf{I}_n$. The constant $\sigma^2 > 0$ is unknown. We have $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$, $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ and $\mathbf{e} = (\mathbf{y} - \hat{\mathbf{y}})$

13.1.1 (a)

Show $\mathbf{X}^T\mathbf{e} = \mathbf{0}$.

13.1.2 (b)

If the model has an intercept, why does this last result show that the sum of residuals equals zero?

13.1.3 (c)

Let $\mathbf{1}$ denote an $n \times 1$ columns of ones, and let $\bar{\mathbf{x}} = \frac{1}{n}\mathbf{X}^T\mathbf{1}$ **.**

i. What are the dimensions of the matrix $\bar{\mathbf{x}}$?

ii. If the model has an intercept, what is the first element of $\bar{\mathbf{x}}$?

iii. What is the second element of $\bar{\mathbf{x}}$?

13.1.4 (d)

We are interested in the predicted value of y (height of the least-squares regression plane) when all the explanatory variables are set to their sample mean values. Express this quantity in terms of $\bar{\mathbf{x}}$ and $\hat{\boldsymbol{\beta}}$.

13.1.5 (e)

Assuming the model has an intercept, simplify your answer to the last question. What do you get?

13.2 Solution

13.2.1 (a)

$$\mathbf{X}^T\mathbf{e} = \mathbf{X}^T(\mathbf{y} - \hat{\mathbf{y}}) \tag{96}$$
$$\mathbf{X}^T\mathbf{e} = \mathbf{X}^T\mathbf{y} - \mathbf{X}^T\hat{\mathbf{y}} \tag{97}$$
$$\mathbf{X}^T\mathbf{e} = \mathbf{X}^T\mathbf{y} - \mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} \tag{98}$$
$$\mathbf{X}^T\mathbf{e} = \mathbf{X}^T\mathbf{y} - \mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{99}$$
$$\mathbf{X}^T\mathbf{e} = \mathbf{X}^T\mathbf{y} - \mathbf{I}\mathbf{X}^T\mathbf{y} \tag{100}$$
$$\mathbf{X}^T\mathbf{e} = \mathbf{X}^T\mathbf{y} - \mathbf{X}^T\mathbf{y} = \mathbf{0} \tag{101}$$

13.2.2 (b)

If the model has an intercept, then the first element of $\boldsymbol{\beta}$ is the intercept term presumably, and consequently the first column of $\mathbf{X}$ is $\mathbf{1}$. If this is so, then the first row of $\mathbf{X}^T$ is $\mathbf{1}^T$. Then, since we have shown $\mathbf{X}^T\mathbf{e} = \mathbf{0}$ in our previous question, we would have $\mathbf{1}^T\mathbf{e} = \sum_{i=1}^n e_i = 0$, that is the sum of residuals equals zero. This is why if the model has an intercept, our last result show that the sum of residuals equals zero.

13.2.3 (c)

i. $p \times 1$.

ii. As mentioned earlier, the first row of $\mathbf{X}^T$ is $\mathbf{1}^T$, so the first element of $\bar{\mathbf{x}}$ is $\frac{1}{n}\mathbf{1}^T\mathbf{1} = \frac{1}{n}\sum_{i=1}^n 1 = \frac{1}{n}n = 1$.

iii. It is the mean value of the first independent variable, i.e. since the second row of $\mathbf{X}^T$ consist of the values of the first independent variable, the second element of $\bar{\mathbf{x}}$ is $\bar{x}_1 = \frac{1}{n}(X_{1,1}, X_{2,1}, ..., X_{n,1})\mathbf{1} = \frac{1}{n}\sum_{i=1}^n X_{i,1}$.

13.2.4 (d)

This quantity is $\bar{\mathbf{x}}^T\hat{\boldsymbol{\beta}}$.

13.2.5 (e)

$$\bar{\mathbf{x}}^T\hat{\boldsymbol{\beta}} = \frac{1}{n}\mathbf{1}^T\mathbf{X}\hat{\boldsymbol{\beta}} \tag{102}$$

$$\bar{\mathbf{x}}^T\hat{\boldsymbol{\beta}} = \frac{1}{n}\mathbf{1}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{103}$$

$$\bar{\mathbf{x}}^T\hat{\boldsymbol{\beta}} = \frac{1}{n}\mathbf{1}^T\mathbf{X}\mathbf{X}^{-1}(\mathbf{X}^T)^{-1}\mathbf{X}^T\mathbf{y} \tag{104}$$

$$\bar{\mathbf{x}}^T\hat{\boldsymbol{\beta}} = \frac{1}{n}\mathbf{1}^T\mathbf{I}\mathbf{I}\mathbf{y} \tag{105}$$

$$\bar{\mathbf{x}}^T\hat{\boldsymbol{\beta}} = \frac{1}{n}\mathbf{1}^T\mathbf{y} \tag{106}$$

$$\bar{\mathbf{x}}^T\hat{\boldsymbol{\beta}} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y} \tag{107}$$

I get the mean value of the independent variable, $\bar{y}$.

14

14.1 Problem

High School History classes from across Ontario are randomly assigned to either a discovery-oriented or a memory-oriented curriculum in Canadian history. At the end of the year, the students are given a standardized test and the median score of each class is recorded. Please consider a regression model with these variables:

X_1 Equals 1 if the class uses the discovery-oriented curriculum, and equals 0 if the class uses the memory-oriented curriculum.

X_2 Average parents' education for the classroom.

X_3 Average family income for the classroom.

X_4 Number of university History courses taken by the teacher.

X_5 Teacher's final cumulative university grade point average.

Y Class median score on standardized history test.

The full regression model (as opposed to the reduced models for various null hypotheses) implies

$$E[Y|X] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5. \tag{108}$$

For each question below, please give

- The null hypothesis in terms of β values.
- $E[Y|X]$ for the reduced model you would use to answer the question. Don't re-number the variables.

14.1.1 (a)

If you allow for parents' education and income and for teacher's university background, does curriculum type affect test scores? (And why is it okay to use the word "affect?")

14.1.2 (b)

Controlling for parents' education and income and for curriculum type, is teacher's university background (two variables) related to their students' test performance?

14.1.3 (c)

Correcting for teacher's university background and for curriculum type, are parents' education and family income (considered simultaneously) related to students' performance?

14.1.4 (d)

Taking curriculum type, teacher's university background and parents' education into consideration, is parents' income related to students' test performance?

14.1.5 (e)

Here is one final question. Assuming that $X_1, ..., X_5$ are random variables (and I hope you agree that they are),

i. Would you expect X_1 to be related to the other explanatory variables?

ii. Would you expect the other explanatory variables to be related to each other?

14.2 Solution

14.2.1 (a)

$$H_0 : \beta_1 = 0 \tag{109}$$

$$E[Y|X] = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 \tag{110}$$

Assuming there is a relationship, curriculum type ”affecting” test scores is the only logical way to describe the relationship (as classes are assigned randomly to a curriculum type before the test takes place, so it does not make sense to suggest test scores affect curriculum type).

14.2.2 (b)

$$H_0 : \beta_4 = \beta_5 = 0 \tag{111}$$

$$E[Y|X] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \tag{112}$$

14.2.3 (c)

$$H_0 : \beta_2 = \beta_3 = 0 \tag{113}$$

$$E[Y|X] = \beta_0 + \beta_1 X_1 + \beta_4 X_4 + \beta_5 X_5 \tag{114}$$

14.2.4 (d)

$$H_0 : \beta_3 = 0 \tag{115}$$

$$E[Y|X] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_4 X_4 + \beta_5 X_5 \tag{116}$$

14.2.5 (e)

i. No, because classes are assigned randomly to a curriculum type.

ii. Yes. For example, there may be possible relationship between X_2 and X_3 (parents' education and income as higher income means more resources to pursue higher education and higher education means more qualification to pursue higher paid jobs).