10 años Plan Ceibal

Problem 1

Since $Cov(X) = E((X-E(X))(X-E(X))^T)$

$Y = AX$ $\Rightarrow$

$$\Rightarrow Cov(AX) = E(AX-E(AX))(AX-E(AX))^T)$$
$$= E(A(X-E(X))((X-E(X))^T A^T)$$
$$= A\underbrace{E[(X-E(X))(X-E(X))^T]}_{=\Sigma_X}A^T$$
$$= A\Sigma_X A^T$$

Problem 2

a) $Cov(Z,X_1) = E(ZX_1^T) - E(Z)E(X_1)^T$ → X_1 has mean $0 \Rightarrow E(X_1)^T$ vector row of 0

covariance definition

$$= E(ZX_1^T) = E((X_2-\alpha X_1)X_1^T)$$

$Z = X_2 - \alpha X_1$

$$= E(X_2X_1^T) - \alpha E(X_1X_1^T) = 0$$

$Cov(Z,X_1)$

$$\Rightarrow \alpha = \frac{E(X_2X_1^T)}{E(X_1X_1^T)} = \frac{\sigma_{21}}{\sigma_{11}}$$

b) Suppose $v = \begin{pmatrix} Z \\ X_1 \end{pmatrix}$ $\Sigma = \begin{pmatrix} \Sigma_{ZZ} & \Sigma_{ZX_1} \\ \Sigma_{X_1Z} & \Sigma_{X_1X_1} \end{pmatrix}$ $Cov(Z,X_1)=0$

$\mu = \begin{pmatrix} \mu_Z \\ \mu_{X_1} \end{pmatrix}$ $Cov(Z,X_1)=0 \,\|\, 0$

$$f_v(v) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\{-1/2 (x-\mu)' \Sigma^{-1} (x-\mu)\}$$

$$\Rightarrow \begin{pmatrix} v_1^T - \mu_1^T & v_2^T - \mu_2^T \end{pmatrix} \begin{pmatrix} \Sigma_{ZZ} & 0 \\ 0 & \Sigma_{X_1X_1} \end{pmatrix}^{-1} \begin{pmatrix} v_1 - \mu_1 \\ v_2 - \mu_2 \end{pmatrix} =$$

Cov matrix diagonal

$$= \begin{pmatrix} v_1^T - \mu_1^T & v_2^T - \mu_2^T \end{pmatrix} \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix} \begin{pmatrix} v_1 - \mu_1 \\ v_2 - \mu_2 \end{pmatrix} =$$

$$= (v_1 - \mu_1)^T \Sigma_{11}^{-1} (v_1 - \mu_1) + (v_2 - \mu_2)^T \Sigma_{22}^{-1} (v_2 - \mu_2)$$

This can be factored because it is within exponential function

So, the joint pdf $f_v(x) = g_Z(v_1) h_{X_1}(v_2)$ shows that Z and X_1 are independent

$$Var(Z) = Var(X_2 - \alpha X_1) = Var(X_2) + \alpha^2 (Var(X_1)) - 2\alpha Cov(X_2, X_1)$$

$$= \sigma_{22} + (\sigma_{21}\sigma_{11}^{-1})^2 \sigma_{11} - 2\sigma_{21}\sigma_{11}^{-1}\sigma_{21}$$

$$= \sigma_{22} + \sigma_{21}\sigma_{11}^{-1}\sigma_{21} - 2\sigma_{21}\sigma_{11}^{-1}\sigma_{21}$$

$$= \sigma_{22} - \sigma_{21}\sigma_{11}^{-1}\sigma_{21}$$

c) $E(X_2|X_1) = \alpha x + E(Z) = E(X_2|X_1) = \alpha x + E(X_2) - \alpha E(X_1) = \sigma_{21}\sigma_{11}^{-1} X$

$$Var(X_2|X_1) = Var(\alpha X|X_1) + Var(Z|X_1) = Var(Z) = \sigma_{22} - \sigma_{21}\sigma_{11}^{-1}\sigma_{21}$$

Problem 3

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$P = I_n - \frac{1}{n} 1_n 1_n'$ → symmetric and idempotent matrix (HW1)

a) $(n-1)s^2 = X'PX$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{n-1}(PX)^T PX$$

$$= \frac{1}{n-1} X^T P^T P X$$

$$= \frac{1}{n-1} X^T P P X$$

P symmetric

$$= \frac{1}{n-1} X^T P X$$

P idempotent

$$\Rightarrow s^2(n-1) = X^T P X$$

b) $E\left(\sum(X_i-\mu)^2\right) = \sum(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2$

$$= \sum\left((x_i-\bar{x}) + (\bar{x}-\mu)\right)^2$$

$$= \sum\left((x_i-\bar{x})^2 + 2(x_i-\bar{x})(\bar{x}-\mu) + (\bar{x}-\mu)^2\right)$$

$$= \sum(x_i-\bar{x})^2 + 2(\bar{x}-\mu)\underbrace{\cancel{\sum(x_i-\bar{x})}}_{0} + n(\bar{x}-\mu)^2$$

$$= \sum(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2$$

$$E(s^2) = E\left(\frac{\sum(x_i-\bar{x})^2}{n-1}\right) = \frac{1}{n-1} E\left(\sum(x_i-\mu)^2 - n((\bar{x}-\mu)^2)\right)$$

$$= \frac{1}{n-1}\left(\sum E((x_i-\mu)^2) - nE((\bar{x}-\mu)^2)\right) = \frac{1}{n-1}\left(\sum\sigma^2 - n\,Var(\bar{x})\right)$$

$$= \frac{1}{n-1}\left(n\sigma^2 - \frac{n\sigma^2}{n}\right) = \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2$$

c) $Var(s^2)$ $\quad$ Var normal $\to \sigma^2$

$$Var\left(\frac{(n-1)s^2}{\sigma^2}\right) = \frac{(n-1)^2}{\sigma^4} Var(s^2) = 2(n-1)$$

$$Var(s^2) = \frac{2\sigma^4}{n-1}$$

d) $$\underbrace{\sum_{i=1}^{n} (x_i-\mu)^2}_{1} = \sum_{i=1}^{n} (x_i - \bar{x} + \bar{x} - \mu)^2 = \sum_{i=1}^{n} (x_i-\bar{x})^2 + 2(x_i-\bar{x})(\bar{x}-\mu) + (\bar{x}-\mu)^2$$

$$= \sum_{i=1}^{n} (x_i-\bar{x})^2 + 2(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\bar{x}) + n(\bar{x}-\mu)^2$$

$$= \underbrace{\sum_{i=1}^{n} (x_i-\bar{x})^2}_{2} + \underbrace{n(\bar{x}-\mu)^2}_{3}$$

1 $\to$ $\dfrac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2}$ has a chi squared distribution w/ n degrees of freedom because it is the sum of n independent standard normals

2 $\to$ $\dfrac{n(\bar{x}-\mu)^2}{\sigma^2}$ has chi-squared distribution w/ 1 degree of freedom because it is the square of a standard normal.

e) Since $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$ is a function of $x_i - \bar{x}$, $i=1 \ldots n$

it follows that s^2 is independent of $\bar{x}$

Problem 5

a)

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 22.55565   17.19680   1.312   0.1968
sex        -22.11833    8.21111  -2.694   0.0101 *
status       0.05223    0.28111   0.186   0.8535
income       4.96198    1.02539   4.839 1.79e-05 ***
verbal      -2.95949    2.17215  -1.362   0.1803
---
```

Sex and income are statistically significant at the 5% level.

```
> confint(lm, level=0.95)
                  2.5 %     97.5 %
(Intercept) -12.1489038 57.2602050
sex         -38.6890301 -5.5476301
status       -0.5150722  0.6195399
income        2.8926538  7.0313047
verbal       -7.3430703  1.4240833
```

b) Sex is just a dummy variable representing the categorical data of sex. 1 represents female and 0 represents male.

c)

```
Analysis of Variance Table

Model 1: gamble ~ income
Model 2: gamble ~ sex + status + income + verbal
  Res.Df   RSS Df Sum of Sq      F  Pr(>F)
1     45 28009
2     42 21624  3    6384.8 4.1338 0.01177 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Since the p-value is very small the null hypothesis is rejected.

d)

```
Model 1: gamble ~ 1
Model 2: gamble ~ sex + status + income + verbal
  Res.Df   RSS Df Sum of Sq      F    Pr(>F)
1     46 45689
2     42 21624  4     24066 11.686 1.815e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

10 años Plan Ceibal

Problem 6

$$t\ value = \frac{Estimate}{Std.Error} \Rightarrow t\ value = \frac{0.7190}{0.1535} = 4.684$$

$$p\text{-value} \rightarrow 2 \cdot pt(-abs(t), df=28) = 6.5776 \times 10^{-5}$$

$$R^2 = 1 - \frac{RSS}{TSS}$$

$$F = \frac{\frac{TSS-RSS}{r-1}}{\frac{RSS}{n-r}} = 21.94 \quad , r=2$$

$$RSE = \sqrt{\frac{RSS}{n-r}} \Rightarrow 0.8662 = \sqrt{\frac{RSS}{28}} \Rightarrow RSS = 21.008$$

$$F = \frac{TSS-21}{\frac{21}{28}} = 21.94 \Rightarrow TSS = 37.455$$

$$\Rightarrow R^2 = 1 - \frac{21}{37.455} = 0.5606$$

Appendix

```
# title: "MSDS596 - HW2"
# author: "Diego Sarachaga"
# date: "10/02/2018"

library(faraway)
data(teengamb)


lm <- lm(gamble ~ sex + status + income + verbal, teengamb)
summary(lm)

#a
# Coefficients:
#             Estimate  Std. Error  t value Pr(>|t|)
# (Intercept)  22.55565   17.19680   1.312   0.1968
# sex        -22.11833    8.21111  -2.694   0.0101 *
#   status       0.05223    0.28111   0.186   0.8535
# income       4.96198    1.02539   4.839 1.79e-05 ***
#   verbal      -2.95949    2.17215  -1.362   0.1803
# ---
#   Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
#
# Residual standard error: 22.69 on 42 degrees of freedom
# Multiple R-squared:  0.5267,         Adjusted R-squared:  0.4816
# F-statistic: 11.69 on 4 and 42 DF,  p-value: 1.815e-06

#Sex and income are statistically significant at the 5% level
confint(lm, level=0.95)

#Confident intervals
#sex        -38.6890301 -5.5476301
#income       2.8926538  7.0313047



#b
#Sex is a dummy variable representing the categorical data of sex. A 1
represents female and a 0 represents male.

#c
lmi <- lm(gamble ~ income, teengamb)
summary(lmi)

anova(lmi, lm)
# Analysis of Variance Table
```

```
#
# Model 1: gamble ~ income
# Model 2: gamble ~ sex + status + income + verbal
# Res.Df   RSS Df Sum of Sq      F  Pr(>F)
# 1     45 28009
# 2     42 21624  3    6384.8 4.1338 0.01177 *
#   ---
#   Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#Since the p-value is so small the null hypothesis is rejected.

#d
nullmod <- lm(gamble ~ 1, teengamb)
anova(nullmod, lm)
# Analysis of Variance Table
#
# Model 1: gamble ~ 1
# Model 2: gamble ~ sex + status + income + verbal
# Res.Df   RSS Df Sum of Sq      F    Pr(>F)
# 1     46 45689
# 2     42 21624  4     24066 11.686 1.815e-06 ***
#   ---
#   Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```