# STA437/2005 Methods for Multivariate Data
# Practice Problem set #1

**Problem 1.** Consider the following seven observations of three variables:

| V1 | V2 | V3 |
|---|---|---|
| 5.1 | 3.3 | 1.7 |
| 6.8 | 2.8 | 4.8 |
| 5.8 | 2.7 | 3.9 |
| 6.9 | 3.1 | 4.9 |
| 5.7 | 2.5 | 5.0 |
| 5.8 | 2.8 | 5.1 |
| 6.4 | 3.2 | 4.5 |

(a) Compute the sample mean
(b) Compute the sample variance
(c) Make $z$-transformation on V3 and investigate whether or not there are any anomalous observations. [Hint: Standardize and compare to 3 or 3.5 or any appropriate number]
(d) Compute the $\chi^2$-statistic of the first observation and determine whether or not it is anomalously big. [Hint: Pick an appropriate high quantile point]
(e) Assess a hypothesis $H_0 : \mu = O$. [Hint: Use Hotelling's $T^2$-statistic]
(f) Assess whether or not V3 is normally distribute. [Hint: Interval test check up like $\#\{x_i : x_i \in \bar{x} \pm k\hat{\sigma}\} \approx \text{Binomial}(n, \Phi(k) - \Phi(-k))$]

**Problem 2.** $X = (X_1, X_2, X_3)^\top \sim N_3(\mu, \Sigma)$ where $\mu = (1, 2, 3)^\top$ and $\Sigma = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix}$
(a) Find the distribution of $2X_1 - 3X_2 + X_3$. [Hint: Linear sums of multivariate normal distributions are also normal]
(b) Solve $a$ and $b$ so that $X_1$ and $aX_2 + bX_3$ are independent. [Hint: Joint multivariate normal random variables are independent if correlations are zero]
(c) Solve $a$ and $b$ so that $X_1$ and $aX_2 + bX_3$ are independent as well as $X_2$ and $aX_1 + bX_3$ are independent. [Hint: Joint multivariate normal random variables are independent if correlations are zero]
(d) Find the conditional distribution of $X_3$ given $(X_1, X_2)$

**Problem 3.** A set of two dimensional data is observed which is assumed to be $\mathbf{x}_j \sim N_2(\mu, \Sigma)$. Sample statistics are given by $n = 40$, $\bar{\mathbf{x}} = (0.7, 0.4)^\top$ and $S = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$.
(a) Write the equation of 95% confidence region for $\mu$ and draw it roughly. [Hint: Solve a quadratic equation]
(b) Assess a hypothesis $H_0 : \mu = (1, 0)^\top$. [Hint: Consider Hotelling's $T^2$-statistic]
(c) Assess a hypothesis $H_0 : \mu_1 = 2\mu_2$. [Hint: Set $\psi = \mu_1 - 2\mu_2$ and test $H_0 : \psi = 0$]
(d) Find a 95% confidence interval for $\mu_2$. [Hint: Use $t$-distribution or asymptotic normal distribution]
(e) Assess whether or not $X_1$ and $X_2$ are independent. [Hint: Fisher's $z$-transformation or Pearson's independence test for categorical data]

**Problem 4.** A $n \times p$ random matrix $\mathbf{X}$ and a random vector $Y \in \mathbb{R}^p$ has a relationship $Y = \mathbf{X}\beta + \boldsymbol{\epsilon}$ where $\beta \in \mathbb{R}^p$ is an unknown parameter, $\boldsymbol{\epsilon} \sim N_n(O, I_n)$ and $\mathbf{X}$ and $\boldsymbol{\epsilon}$ are independent.
(a) Find the least square estimator $\hat{\beta}$ of $\beta$. [Hint: $\hat{\beta} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top Y$]
(b) Show the consistency of $\hat{\beta}$. [Hint: Use Chebyschev's inequality being provided by $(\mathbf{X}^\top\mathbf{X})^{-1}$ converges to zero.]
(c) Show the asymptotic normality of $\hat{\beta}$. [Hint: It is normally distributed if $\mathbf{X}$ is given.]
(d) Assess a hypothesis $H_0 : \beta = O$. [Hint: Use asymptotic distribution in part (c)]
(e) Write the required assumptions for (a)-(d). [Hint: For part (d) you need $\mathbf{X}^\top\mathbf{X}/n \to \Sigma_0$ in probability]

**Problem 5.** (a) Describe a procedure of detecting outliers.
(b) Describe a procedure of assessing normality.
(c) Describe Box-Cox transformation.