# [STAT409] Homework 1

1. Use `nycflights13` package and the `flights` data to answer the following questions.

   (a) What month had the highest proportion of cancelled flights? What month had the lowest? Interpret any seasonal patterns.

   (b) What plane (specifed by the `tailnum` variable) traveld the most times from NY city airports in 2013? Plot the number of trips per week over the year.

2. For a random variable $X$, define the following function, $g(a)$:

$$g(a) = E(X - a)^2.$$

   Show that the minimizer of $g(a)$ is $E(X)$ i.e., $g(a)$ is minimized at $a = E(X)$. (*Hint: $g(a)$ is convex and differentiable function*)

3. Suppose $X_1, X_2, \cdots, X_n$ are iid sample from a distribution $F$ with $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$, then for any given $\epsilon > 0$ we have

$$\lim_{n\to\infty} P(|\bar{X}_n - E(X)| > \epsilon) = 0$$

   where $\bar{X}_n = \sum_{i=1}^n X_i/n$. (*Hint: Use Chevy-Scheff Inequality learned at STAT232 - Mathematical Statistics*)

4. Projection matrix $\mathbf{P_A}$ is defined as

$$\mathbf{P_A} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$$

   (a) Show that $\mathbf{P_A}$ is idempotent matrix, i.e., $\mathbf{P_A}\mathbf{P_A} = \mathbf{P_A}$.

   (b) Prove that $\mathbf{z} := \mathbf{P_A}\mathbf{v}$ lies on col($\mathbf{A}$) (i.e., $\mathbf{z} \in$ col($\mathbf{A}$)) by showing that $\mathbf{P_A}\mathbf{z} = \mathbf{z}$.

   (c) Show that

$$(\mathbf{v} - \mathbf{u})^T(\mathbf{v} - \mathbf{u}) \geq (\mathbf{v} - \mathbf{z})^T(\mathbf{v} - \mathbf{z})$$

   for all $\mathbf{u} \in$ col($\mathbf{A}$) satisfying $\mathbf{P_A}\mathbf{u} = \mathbf{u}$.