Exploratory Data Analysis (S670):Assignment 6

Created by **Krishna Mahajan,0003572903**

Q1

Show that the jackknife estimate and the jackknife SE are exactly the same as sample mean $\bar{x}$ and standard error $se(\bar{x})$= $s/\sqrt{n}$ n when $\hat{\theta} = \bar{x}$ (sample mean).
(sol)
We know that jacknife estimate is similar to the bootstrap in that it involves resampling, but instead of sampling with replacement, the method samples without replacement.
Defination
The i_{th} jacknife replication of $\hat{\theta_{(i)}}$ of the statistic $\hat{\theta} = s(x)$ is $\hat{\theta_{(i)}} = s(x_i)$, $\forall i = 1, ...n$
Jacknife estimation of mean
$s(x_{(i)}) = \frac{1}{n-1} \sum_{j \neq i} x_j$
$= \frac{n\bar{x}-x_i}{n-1}$
$= \bar{x_i}$
now,
$\bar{x_{(.)}} = \frac{1}{n} \sum_{i=1}^{n} \bar{x_i} = \bar{x}$

Thus Jacknife estimate of sample mean is same as $\bar{x}$.

Jacknife estimate of the standard error of the mean

For $\hat{\theta} = \bar{x}$,it is easy to show that:

$\bar{x_i}$= $\frac{1}{n-1} \sum_{j \neq i} x_j$
$= \frac{n\bar{x}-x_i}{n-1}$
$= \bar{x_i}$ &

$\bar{x_{(.)}} = \frac{1}{n} \sum_{i=1}^{n} \bar{x_i} = \bar{x}$

Therefore:
$\hat{se_{jack}}(\bar{x})$= $(\sum_{i=1}^{n} \frac{x_i - \bar{x}^2}{(n-1)n})^{1/2}$
$=\frac{\bar{\sigma}}{\sqrt(n)}$
$=\frac{s}{\sqrt(n)}$

Q2

Show that the standard error of the pseudo-values (jackknife SE) is the same as the standard deviation of the " leave-out-one" values multiplied by $(n-1)/\sqrt(n)$ i.e. that the standard error of the pseudo-values (jackknife SE) is the same as the standard error of the "leave-out-one" values multiplied by (n -1).
(sol)
To Prove : $SE(PV's) = \frac{(n-1)}{\sqrt(n)}.SD(y_{(-i)})$ Here,
$y_{(-i)} = mean(xwithoutx_i) = \sum_{k \neq i} x_k/19$
$PV's = n.mean(y) - (n-1)y_{(-i)}$ (as Taught in Class)
so,
$var(PV's) = var(n.mean(y) - (n-1)y_{(-i)})$
$var(PV's) = var(n.mean(y)) + var((n-1)y_{(-i)})$
$var(PV's) = n^2 var(mean(y)) + (n-1)^2 var(y_{(-i)})$
$Sd(PV's) = n * sd(mean(y)) + (n-1) * sd(y_{(-i)})$ (i,e Taking root on both the sides to get standard deviation)

$SE(Pv's) = \frac{sd(mean(y))}{\sqrt(n)} + \frac{n-1sd(y_{(-i)})}{\sqrt(n)}$ (i,e dividing by $n^1/2$ on both sides to get SE)
Now, $\frac{sd(mean(y))}{\sqrt(n)} = 0$ as mean(y) is true population mean and it wont have any variance.
Thus,
$SE(Pv's) = \frac{n-1sd(y_{(-i)})}{\sqrt(n)}$