MA 354: Data Analysis I – Fall 2019
Formula Sheet:

R/LaTeX Sweave notes – this should be all that you need.

- To run R and print the output.

```
<<>>=
        #Rcode goes here
        #Output is automatically printed in the .pdf
@
```

Remark: All R chunks must have no spaces preceding the <<>>= or @ syntax.

- Provide R code for plot and place the plot into our document.

```
<<plotName,eval=FALSE>>=
        #Rcode for plot
        #We will call this later so make sure it has a unique name
@
\begin{figure}[H]
        \centering
        <<fig=TRUE,echo=FALSE>>=
        library("graphics")
        <<plotName>>
        @
        \caption{Some information about our plot} \label{Fig:plot1}
\end{figure}
```

You can then reference a graph in latex using `\ref{Fig:plot1}`.
Remark: All R chunks must have no spaces preceding the <<>>= or @ syntax.

- If you wanted a one line equation that is centered like this,

$$\widehat{y_i} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon$$

you can use this LaTeX.

```
\[\widehat{y_i} = \beta_0 + \beta_1 x_{1i}+ \beta_2 x_{2i} + \epsilon\]
```

- If you wanted a multiple line equation that is centered like this,

$$\begin{aligned} f_X(x) &= 90x^8(1-x)\\ &= 90x^8 - 90x^9 \end{aligned}$$

you can use this LaTeX.

```
\begin{align*}
        f_X(x) &= 90 x^8(1-x)\\
                    &= 90x^8 - 90x^9\\
\end{align*}
```

Help: You can ask for information about any of the following functions that we've used by asking R. For example, if I wanted help with the lm() function I would run ?lm() in the R console. Note that if you're asking a question about a function, its library must be loaded.

- Stock R functions
  - which()
  - subset()
  - summary()
  - names()
  - cumsum()
  - apply()
  - lapply()
  - sapply()
  - tapply()
  - table()
  - prop.table()
  - pie()
  - barplot()
  - hist()
  - density()
  - boxplot()
  - lines()
  - points()
  - jitter()
  - legend()
  - optim()
  - prop.test()
  - t.test()
  - var.test()
  - aov()
  - lm()
  - anova()
  - tukeyHSD()
  - p.adjust()
  - fisher.test()
  - chisq.test()
  - cor()
  - cor.test()
- stringr Package
  - str_split()
- extraDistr Package
  - dmnom()
- nleqslv Package
  - nleqslv()
- ggplot2 Package Plotting
  - ggplot()
  - geom_bar()
  - coord_polar()
  - geom_hline()
  - geom_text()
  - geom_histogram()
  - geom_density()
  - geom_freqpoly()
  - geom_boxplot()
  - geom_jitter()
  - geom_violin()
  - geom_point()
  - geom_line()
  - facet_grid()
  - coord_flip()
  - theme_bw()
  - xlab()
  - ylab()
  - ggtitle()
- Probability Distribution
  - dbinom()
  - dhyper()
  - dnbinom()
  - dpois()
  - dunif()
  - dnorm()
  - dlnorm()
  - dchisq()
  - dt()
  - df()
- gridExtra Package
  - grid.arrange()
- qqplotr Package
  - stat_qq_band()
  - stat_qq_line()
  - stat_qq_point()
- boot Package
  - boot()
  - boot.ci()
- BSDA Package
  - SIGN.test()
- simpleboot Package
  - two.boot()
- RVAideMemoire Package
  - mood.medtest()
  - cramer.test()
- rcompanion Package
  - pairwiseMedianTest()
  - cldList()
  - phi()
  - cramerV()
- multcomp Package
  - glht()
  - cld()
- FSA Package
  - dunnTest()
- DescTools Package
  - StuartTauC()

- Bernoulli Distribution

$$f_X(x|p) = p^x(1-p)^{1-x} I(x \in \{0,1\})$$
[PMF]

$$E(X) = p \qquad \textbf{[Expected Value]}$$
$$var(X) = p(1-p) \qquad \textbf{[Variance]}$$

- Binomial Distribution

$$f_X(x|n,p) = \binom{n}{x} p^x(1-p)^{n-x} I(x \in \{0,1,\dots n\})$$
[PMF]

$$E(X) = np \qquad \textbf{[Expected Value]}$$
$$var(X) np(1-p) \qquad \textbf{[Variance]}$$

- Hypergeometric Distribution

$$f_X(x|N,n,m,k) = \frac{\binom{m}{x}\binom{n}{(k-x)}}{\binom{N}{k}} I(x \in \mathcal{X})$$
[PMF]

$$E(X) = \frac{km}{m+n}$$
[Expected Value]

$$var(X) = \frac{km}{m+n} \; \frac{-n}{m+n} \; \frac{m+n-k}{m+n-1}$$
[Variance]

- Negative Binomial Distribution

$$f_X(x|n,p) = \binom{n+x-1}{x} p^n(1-p)^x I(x \in \{0,1,\dots\})$$
[PMF]

$$E(X) = \frac{n(1-p)}{p} \qquad \textbf{[Expected Value]}$$
$$var(X) = \frac{n(1-p)}{p^2} \qquad \textbf{[Variance]}$$

- Poisson Distribution

$$f_X(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \; I(x \in \{0,1,\dots\}) \qquad \textbf{[PMF]}$$
$$E(X) = \lambda$$
$$var(X) = \lambda$$

- Uniform Distribution

$$f_X(x|a,b) = \frac{1}{b-a} \; I(x \in [a,b]) \qquad \textbf{[PDF]}$$
$$E(X) = \frac{a+b}{2} \qquad \textbf{[Expected Value]}$$
$$var(X) = \frac{(b-a)^2}{12} \qquad \textbf{[Variance]}$$

- Gaussian Distribution

$$f_X(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \; I(x \in \mathbb{R}) \qquad \textbf{[PDF]}$$
$$E(X) = \mu \qquad \textbf{[Expected Value]}$$
$$var(X) = \sigma^2 \qquad \textbf{[Variance]}$$

- Log-Normal Distribution

$$f_X(x|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{(ln(x)-\mu)^2}{2\sigma^2}} \; I(x \in (0,\infty))$$
[PDF]

$$E(X) = e^{\mu+\sigma^2/2} \qquad \textbf{[Expected Value]}$$
$$var(X) = e^{2\mu+\sigma^2} e^{\sigma^2 - 1} \qquad \textbf{[Variance]}$$

- Chi-squared Distribution

$$f_X(x) = \frac{1}{\Gamma\left(\frac{v}{2}\right) 2^{v/2}} x^{\frac{v}{2}-1} e^{\frac{-x}{2}} \qquad \textbf{[PDF]}$$
$$E(X) = v \qquad \textbf{[Expected Value]}$$
$$var(X) = 2v \qquad \textbf{[Variance]}$$

- Student T distribution

$$f_T(t) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{\pi}\;\Gamma(v/2)} \left(1 + \frac{t^2}{2}\right)^{-(v+1)/2}$$
[PDF]

$$E(X) = 0 \qquad \textbf{[Expected Value for } v > 1\textbf{]}$$
$$var(X) = \frac{v}{v-2} \qquad \textbf{[Variance for } v > 2\textbf{]}$$

- F distribution

$$f_W(w) = \frac{\Gamma(\frac{u+v}{2})}{\Gamma(\frac{u}{2})\Gamma(\frac{v}{2})} \left(\frac{u}{v}\right)^{u/2} \frac{w^{\frac{u}{2}-1}}{[1+(\frac{u}{v})w]^{(u+v)/2}} \; I(w > 0)$$
[PDF]

$$E(W) = \frac{v}{v-2}$$
(**[Expected Value for** $v > 2$**]**)

$$var(W) = \left(\frac{u-2}{u}\right)\left(\frac{v}{v+2}\right) \qquad (\textbf{[Variance]})$$