Part VI: Actual Statistics

Probability distributions

```
pnorm(0)
pnorm(1)
pnorm(2)


dnorm(1)


pnorm(0.5)
pnorm(0.5, mean=0.0, sd=10)
```

Probability distributions II

- Many distributions: Chi-squared, exponential, gamma, normal, log-normal, Poission, uniform…
- For each distribution, there is:
  - A PDF function (‘dxxx’)
  - A CDF function (‘pxxx’)
  - A quantile function (‘qxxx’)
- e.g. exponential PDF function is ‘dexp’

Random samples from a distribution

```
rnorm(10, mean=0.5, sd=0.01)


rexp(10)
```

Kolmogorov-Smirnov

```
rvals.norm <- rnorm(1000)
rvals.unif <- runif(1000)


ks.test(rvals.norm, “pnorm”)
ks.test(rvals.norm, rnorm(1000, mean=0.2))
ks.test(rvals.unif, “pnorm”)
```

Student's t-test

```
t.test(rvals.norm, rnorm(1000, mean=0.1))
```

Chi-squared test

```
x <- c(5,4,6,3,4,5,6,4)
y <- c(3,5,4,6,3,5,3,6)

chisq.test(x,y)
```

Linear regression

```
glm.of.xy <- glm(x~y)
plot(glm.of.xy)
```

ANOVA

```
aov.T <- aov(Premium ~ Institution +
  CornellEducated, data=T)


summary(aov.T)
```

ACTIVITY #9

- Explore the relationship between a physician's current institution and whether or not they are Cornell-educated using the Malpractice data.
- Explore the relationships between a physician's age and no. of suits and their premium using the Malpractice data.

Things we covered in part VI

- Probability distributions
  - PDF, CDF, quantile functions
- Random sampling from a distribution
- Standard statistical tests
  - Kolmogorov-Smirnov, t-test, Chi-square
- Simple linear regression
- Simple ANOVA

Part VII: Existing libraries

- R commander (Rcmdr)
- Bootstrapping (boot)
- Classification (class)
- Clustering (cluster)
- Survival analysis (survival)
- Bioconductor (www.bioconductor.org)

Loading a library

```
library(“MASS”)
library(“boot”)
```

Installing a library

- Use the “Package Installer” from the “Packages & Data” menu
- For BioConductor packages, use the biocLite function from their install instructions:

```
source(“http://www.bioconductor.org/biocLite.R”)
biocLite(“pkgName”)
```

ACTIVITY #10

- Install the “fortunes” library.
- Load the library and use the help system to find out about the “fortune” function.
- Call the “fortune” function a few times; see if anything remotely funny or fortune-tellish appears.

Things we covered in part VII

- Loading libraries
- Installing libraries
- Installing BioConductor and its libraries

Special Bonus: 1-Slide Functions

```
add2 <- function(X) X + 2
add2(3)
add2(c(1,2,3))

foo <- function(X) {
   Y <- -1 * X;
   print “one thing”;
   print “two things”;
   Y
}
foo(3)
Y
```