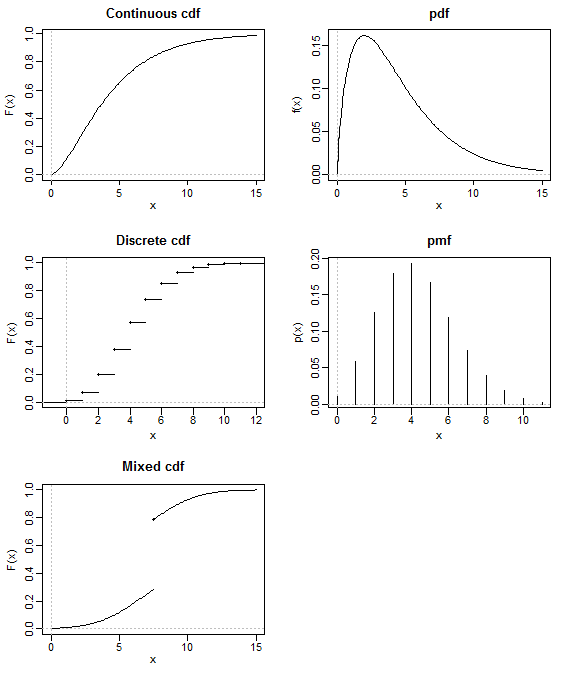
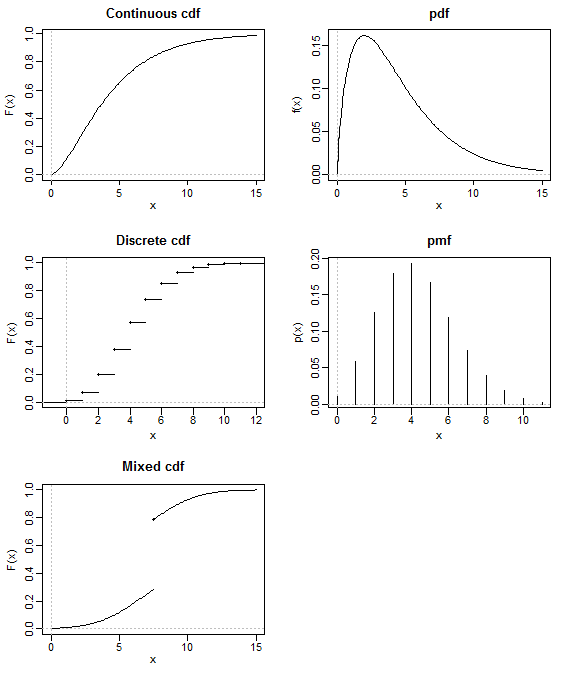
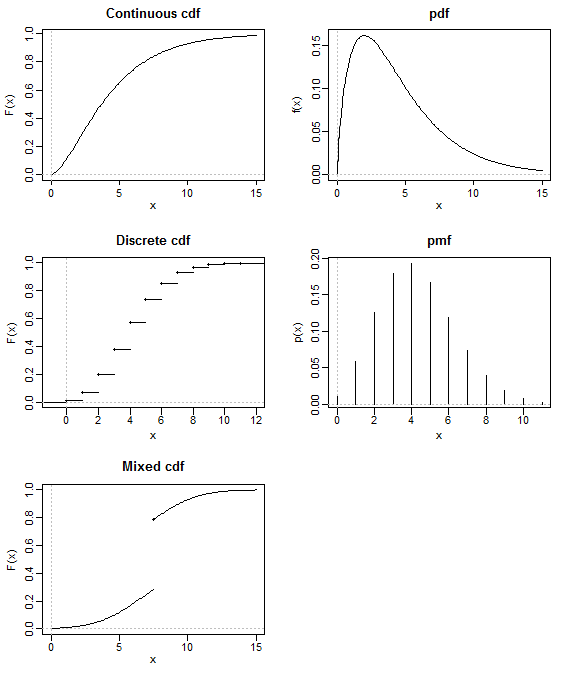
**Things you should probably know:**

Simple probability problems – Homework 2 and 3

* Multiplication of probabilities and being able to extract probability densities using functions, e.g. dbinom() and pbinome().

The difference between a PDF, PMF, and a CDF

* PDF = Probability density function
  + Applies to continuous random variable
* PMF = Probability mass function
  + Only applies to discrete random variables
* CDF = Cumulative distribution function
  + Cumulative probability that increases as you move along the curve

Central Limits Theorem

* This theorem state: estimates of a population mean when repeated will be normally distributed even when the population that you are looking at is not normally distributed.

Definition of a Random Variable

* A variable whose values are not known for certain before a sample is taken, i.e. the observed values of a random variable are the results of a random experiment (the sampling process).
  + Random variables may be either discrete or continuous

Measures of location (mean, median, mode)

* Mean
  + Sum of all values divided by number of values summed
  + Assumptions:
    - Observations are taken from a random sample
    - All observations are independent
    - Population is normally distributed
* Median
  + Middle number in a set of numbers
  + Most non-parametric statistic rely upon the median
* Mode
  + Most common value OR peak of density curve
  + Values may be qualitative, e.g. color
  + Can have multiple
* Geometric mean
  + A.k.a. the back transformed mean
  + Different from the arithmetic mean
  + Useful when dealing with multiplicative, e.g. population growth (exponential)
* Harmonic mean
  + A.k.a. the reciprocal mean
  + Useful for total average speed problems
    - E.g. 100 Km distance, traveled 10 Km/hr there and 20 Km/hr back. What is the total average speed?
* Regarding means:

Measures of spread (standard deviation, range)

* Standard deviation (σ)
  + Measures spread by looking at deviation of distribution from the center or measure of location
    - Average of sum of squares (sum of squared deviations from the mean)
    - Sum of squares:
  + Variance (σ2) is the standard deviation squared
    - Parameter (works great if n = ∞)
    - Statistic (estimate of parameter)
* Range

The parameters of the following distributions, whether they are discrete or continuous, and an example of what type of data might fit these distributions:

Binomial

Poisson

Uniform

Normal

What is a standard normal distribution?

Should be able to calculate and explain

Standard Error

95% Confidence Interval

Coefficient of Variation

Standard Deviation and Variance

Mean

Be familiar with the following statistical distributions and how they are related:

t

F

χ2

Basic data transformations, and when and why they should be applied.

Covariance vs. Correlation

What is Fisher’s r to z transformation? When and why is it used?

What is a linear model?

What is ‘dummy’ coding?

How to calculate the coefficient of determination.

How is the slope calculated in a simple linear regression?

How is the intercept calculated in a simple linear regression?

What is meant by “model selection”?

Methods for model selection

sequential vs. marginal sums of squares.

What is collinearity, why is it a problem, how can we test for it?

What is Maximum Likelihood?

What is a likelihood ratio test, and what determines the number of degrees of freedom?

The statistical hypothesis of a paired t-test

The statistical hypothesis of a t-test

Blocked experimental design

Post-hoc tests following an ANOVA

What does a significant interaction term indicate? What should be done if it is significant?

What is meant by ‘power’?

Beta-coefficients

Odds ratios and the logit transformation

Logistic Regression

General linear model vs. generalized linear model

AIC and BIC

Be able to interpret loadings and PC scores

Why p-values should, or should not, be worshipped.