Hierarchical Bayesian Models

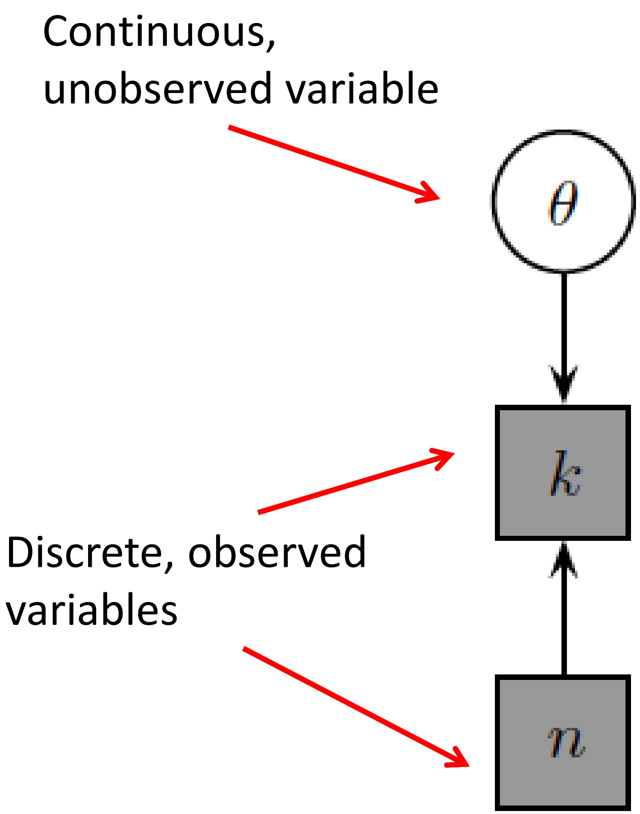
2022 CMMC Summer School
Laura Fontanesi
Chris Donkin
Michael Nunez

The Rate Problem

- Assume you are given a test that consist of 10 factual questions of equal difficulty. We want to estimate your ability, which we define as the rate θ with which you answer questions correctly
- Suppose you answer 9 out of 10 questions correctly.

Graphical Models

Graphical Model for the Rate Problem

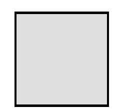
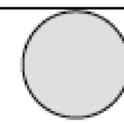
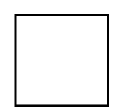
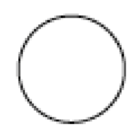
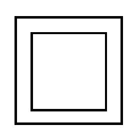
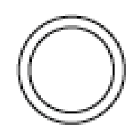


$$\theta \sim \mathrm{Beta}(1, 1)$$

$$k \sim \mathrm{Binomial}(\theta, n)$$

Graphical Model Notation

Table 9.1 *Notation for nodes used in graphical models*

	Type of Variable	
Status of Variable	Discrete	Continuous
Observed		
Unobserved		
Stochastic		
Deterministic		

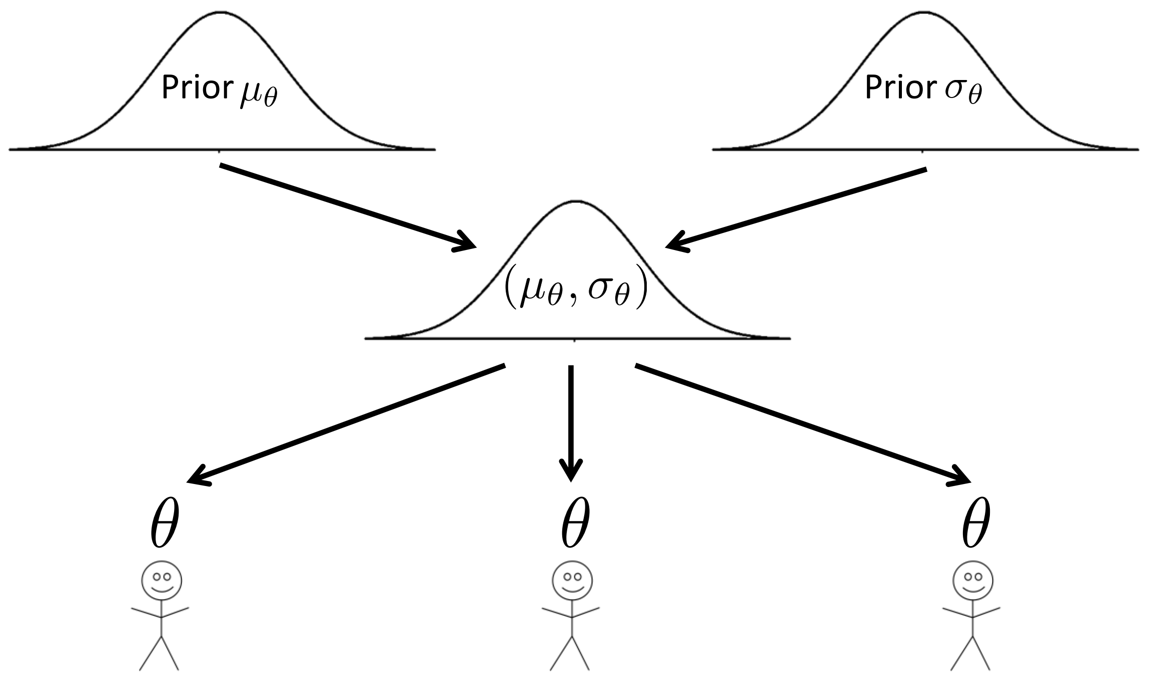
Individuals

- Now say there is a class of 30 students, all doing the same 10 question test.
- For each individual, get a score out of 10
  - So we have k=[5, 6, 8, 7, …]

Your Options

- No individual differences
  - assume participants have identical knowledge, θ
- Full individual differences
  - assume participants have their own knowledge, θ_i
- Structured individual differences
  - assume participants knowledge has some consistent structure, $\theta_i \sim \mathit{Dist}$

Hierarchical Modeling



Hierarchical Models

- Share information
- Better at predicting new data
- Protect against outliers

- Standard method is ‘no individual differences’
  - Each individual treated as equally informative
  - Hierarchical models underweight outliers

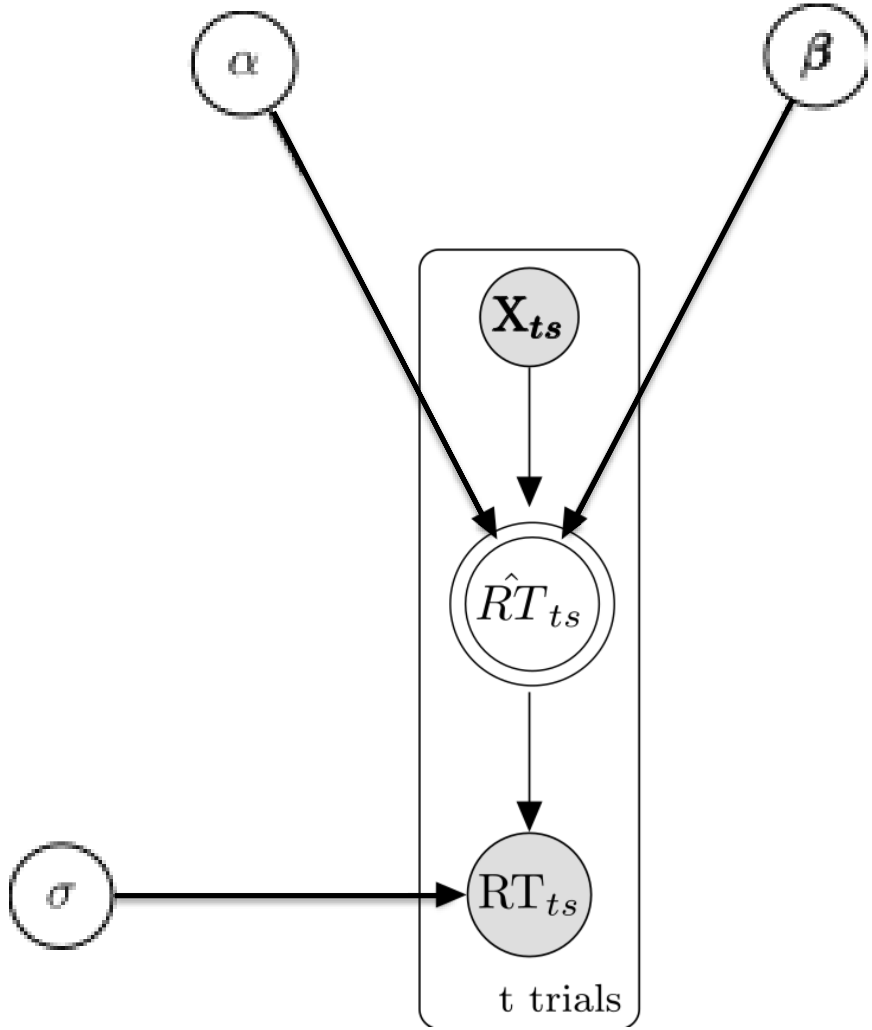
JAGS code for Simple Regression

```
model {
    ## Priors ##
    alpha ~ dnorm(0, 1)
    beta ~ dnorm(0, 1)
    sigma ~ dnorm(0, pow(5, -2)) T(0,) # Half-normal

    ## Likelihood ##
    for (t in 1:trials) {
        RT_hat[t] <- x[t]*beta + alpha
        RT[t] ~ dnorm(RT_hat[t], pow(sigma, -2))
    }
}
```

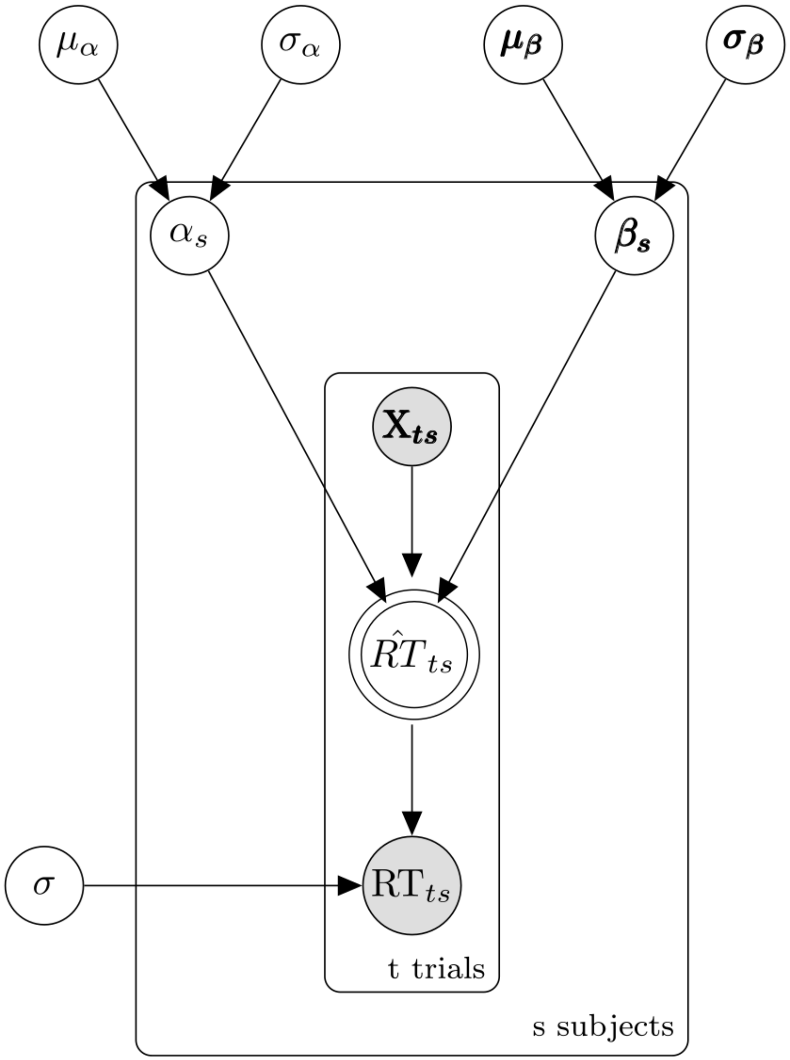
What is the graphical model?

Graphical Model for Simple Regression



What is the graphical model for a <u>hierarchical</u> linear regression over participants?

Graphical Model for Hierarchical Regression



$$\mu_\alpha \sim \mathcal{N}(0,5); \sigma_\alpha \sim \mathcal{HN}(0,5)$$

$$\boldsymbol{\mu_\beta} \sim \mathcal{N}(0,5); \boldsymbol{\sigma_\beta} \sim \mathcal{HN}(0,5)$$

$$\sigma \sim \mathcal{HN}(0,5)$$

$$\alpha_s \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha); \beta_s \sim \mathcal{N}(\mu_\beta, \sigma_\beta)$$

$$\hat{\mathrm{RT}}_{t,s} = \mathbf{X}_{\boldsymbol{t,s}} \boldsymbol{\beta_s} + \alpha_s; \mathrm{RT}_{t,s} \sim \mathcal{N}(\hat{\mathrm{RT}}_{t,s}, \sigma)$$

JAGS code for Hierarchical Regression

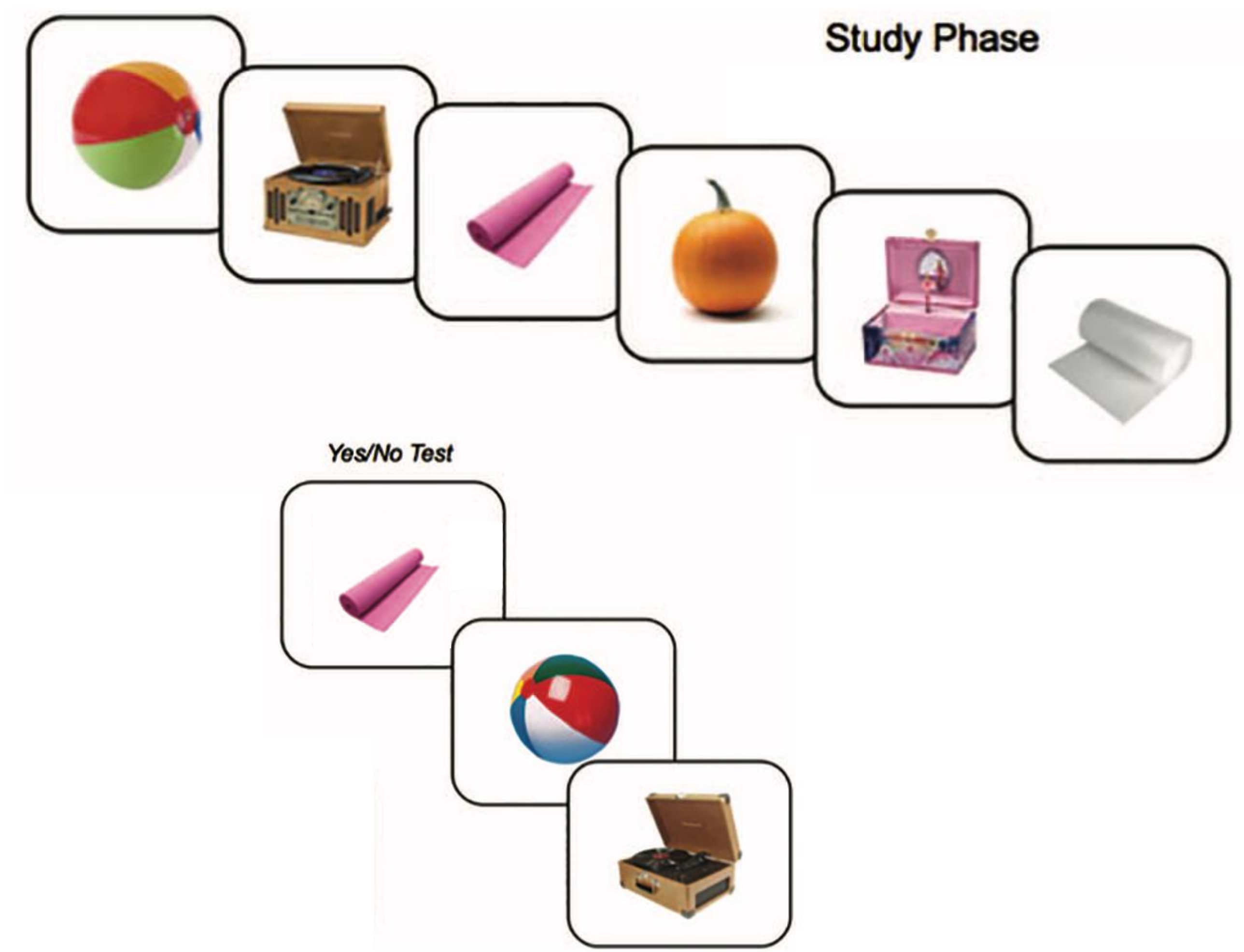
```
model {
	## Hyperpriors ##
	mu_alpha ~ dnorm(0, pow(5, -2))
	sigma_alpha ~ dnorm(0, pow(5, -2)) T(0,) # Half-normal
	mu_beta ~ dnorm(0, pow(5, -2))
	sigma_beta ~ dnorm(0, pow(5, -2)) T(0,) # Half-normal

	## Priors ##
	sigma ~ dnorm(0, pow(5, -2)) T(0,) # Half-normal
	for (s in 1:nparticipants) {
		alpha[s] ~ dnorm(mu_alpha, pow(sigma_alpha, -2))
		beta[s] ~ dnorm(mu_beta, pow(sigma_beta, -2))

		## Likelihood ##
		for (t in 1:ntrials) {
			RT_hat[t,s] <- x[t,s]*beta[s] + alpha[s]
			RT[t,s] ~ dnorm(RT_hat[t,s], pow(sigma, -2))
		}
	}
}
```

Signal Detection Theory

Recognition Memory



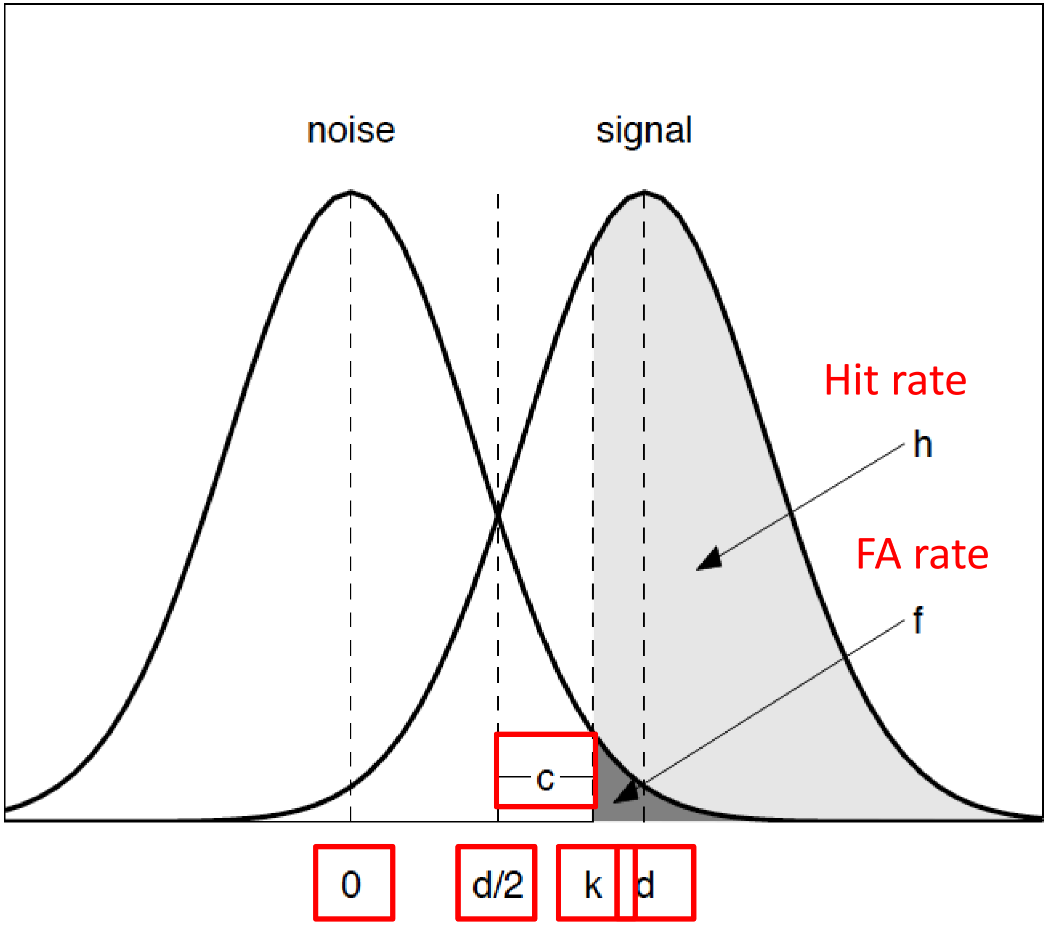
Signal Detection Theory

- Applicable to 2AFC experiments or any situation that can be conceived as a 2 x 2 table of counts
- There are ‘signal’ trials and ‘noise’ trials, and ‘yes’ responses and ‘no’ responses

	Signal Trial	Noise Trial
Yes Response	Hit	False Alarm
No Response	Miss	Correct Rejection

- The data for SDT analysis are just counts of hits, false alarms, misses and correct rejections

Equal-variance Gaussian Signal Detection Theory Framework



mean of noise = 0

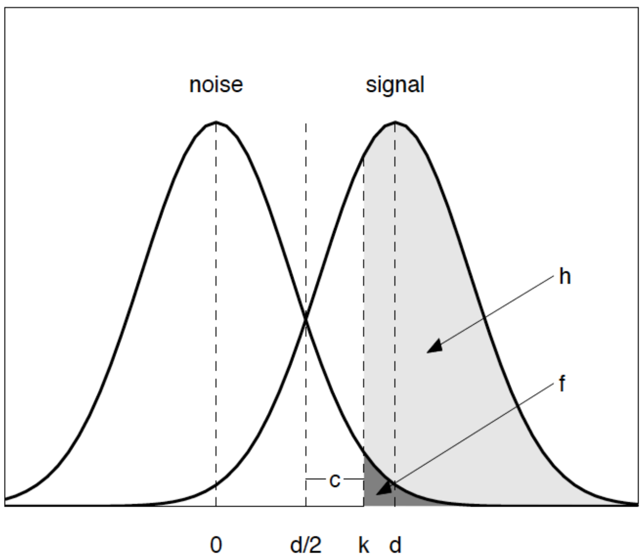
d = discriminability (the distance between the means of the signal and noise distributions)

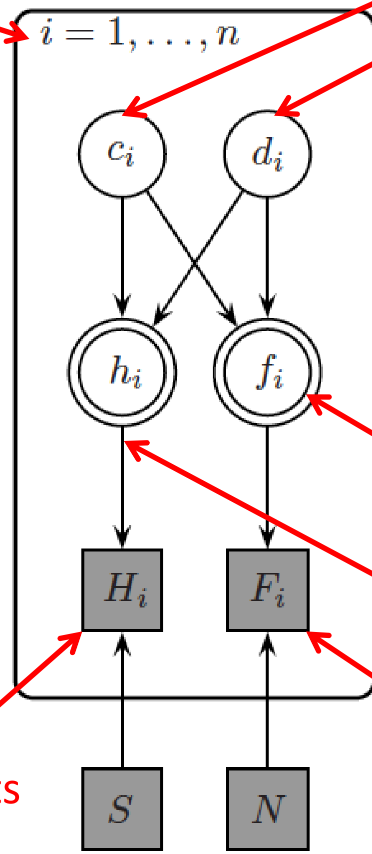
d/2 = criterion value at which both signal and noise distributions are equally likely

k = actual criterion used for responding

c = distance between k and d/2

Graphical Model





$$d_i \sim \text{Gaussian}(0, 1/2)$$

$$c_i \sim \text{Gaussian}(0, 2)$$

$$\boxed{\begin{aligned} h_i &= \Phi(\tfrac{1}{2} d_i - c_i) \\ f_i &= \Phi(-\tfrac{1}{2} d_i - c_i) \end{aligned}}$$

$$H_i \sim \text{Binomial}(h_i, S)$$

$$F_i \sim \text{Binomial}(f_i, N)$$

Code

```
# Signal Detection Theory
model{
 for (j in 1:k){
 # Observed counts
  h[j] ~ dbin(thetah[j],s[j])
  f[j] ~ dbin(thetaf[j],n[j])
  # Reparameterization Using Equal-Variance Gaussian SDT
  thetah[j] <- phi(d[j]/2-c[j])
  thetaf[j] <- phi(-d[j]/2-c[j])
  # These Priors over Discriminability and Bias Correspond
  # to Uniform Priors over the Hit and False Alarm Rates
  d[j] ~ dnorm(0,0.5)
  c[j] ~ dnorm(0,2)
 }
}
```

Use hit and FA rates to model hit and FA counts

Hit and FA rates from SDT

Priors on SDT parameters

Exercise

- Consider the results from the following recognition memory experiment with odors:

Table 9.2: Recognition memory for odors reported by Lehrner et al. (1995).

	Control Group		Group I		Group II	
	Old Odor	New Odor	Old Odor	New Odor	Old Odor	New Odor
Old Resp.	148	29	150	40	150	51
New Resp.	32	151	30	140	40	139

- What conclusions can you draw?

Individual Differences

- The data from this recognition memory experiment with odors came from individuals
- We can estimate the parameters for these individuals separately

Exercise

- Open ind_SDT.txt and edit it to estimate different parameters for each individual
- Use ind_SDT_jags.r to
  - make the data
  - edit the code so that it estimates parameters
  - look at parameters for individuals
  - Anything odd?

Code

```
# Signal Detection Theory
model{
 for (j in 1:k){
  for (i in 1:ns){
  # Observed counts
   h[i,j] ~ dbin(thetah[i,j],s[i,j])
   f[i,j] ~ dbin(thetaf[i,j],n[i,j])
   # Reparameterization Using Equal-Variance Gaussian SDT
   thetah[i,j] <- phi(d[i,j]/2-c[i,j])
   thetaf[i,j] <- phi(-d[i,j]/2-c[i,j])
   # These Priors over Discriminability and Bias Correspond
   # to Uniform Priors over the Hit and False Alarm Rates
   d[i,j] ~ dnorm(0,0.5)
   c[i,j] ~ dnorm(0,2)
  }
 }
}
```

Use hit and FA rates to model hit and FA counts

Hit and FA rates from SDT

Priors on SDT parameters

Exercise

- Open hier_SDT.txt and edit it so that each individual’s parameters come from population-level Normal distributions
- Use hier_SDT_jags.r to
  - What do the population-level parameters look like?

Signal Detection Code

```
model{
  for (j in 1:k){
    for (i in 1:ns){
      h[i,j] ~ dbin(thetah[i,j], s[i,j])
      f[i,j] ~ dbin(thetaf[i,j], n[i,j])

      thetah[i,j] <- phi(d[i,j]/2 - c[i,j])
      thetaf[i,j] <- phi(-d[i,j]/2 - c[i,j])

      d[i,j] ~ dnorm(D[j], precD[j])
      c[i,j] ~ dnorm(C[j], precC[j])
    }
    D[j] ~ dnorm(2, 1)
    precD[j] ~ dgamma(0.001, 0.001)
    C[j] ~ dnorm(0, 2)
    precC[j] ~ dgamma(0.001, 0.001)
  }
}
```

What about Participant 5?

- The weakness of the full individual differences model is evident in its predictions for Subject 5
- Because each subject is assumed to have their own parameters, the only information the model has about the new subject are the priors
- Intuitively, we might predict that Subject 5 will have parameters represented by some sort of average of Subjects 1-4