# Introduction to Bayesian Statistics

Greg Snow

Intermountain Healthcare
Office of Research

September 6, 2018

# Probability

In Bayesian Statistics, probability represents our uncertainty or lack of knowledge.

# Coin example

Flip a fair coin, look at it, but don't show me. Ask me what is the probability that it shows heads:

**Frequentist** Either 100% or 0%, just don't know which

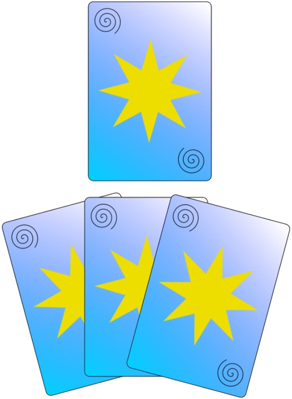
**Bayesian** 50%

# Card Example

We start with 4 cards, 2 red, 2 black.

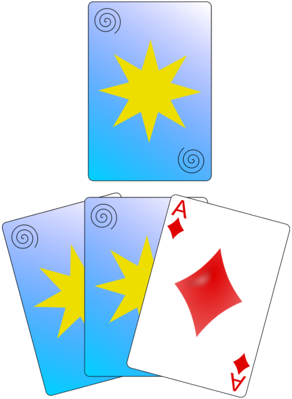
# Card Example

The cards are shuffled and one is drawn at random and placed face down on the table.



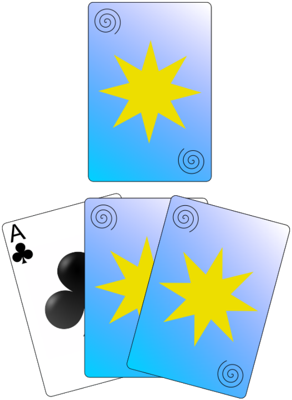
# Card Example

Person A is not shown any cards, person B is shown that one of the remaining cards is Red.



# Card Example

Person C is shown that one of the remaining cards is Black.



# Card Example

What is the probability that the original card is Red/Black?

Person A: $P(Red) = \frac{2}{4} = \frac{1}{2}$
Person B: $P(Red) = \frac{1}{3}$
Person C: $P(Red) = \frac{2}{3}$

# Bayesian Statistics

- Develop a model of how the data are related to parameters of interest (and nuisance parameters).
- Choose prior distributions for the parameter(s) (Include prior information if any).
- Compute posterior distributions on the parameters using the model and data.

# Bayesian Computation



To compute the posterior you need:

- Simple and specific model and priors
  or
- Complicated math that ranges from very difficult to **impossible**
  or
- A computer and lots of calculations to approximate the math.

# Prior Distribution



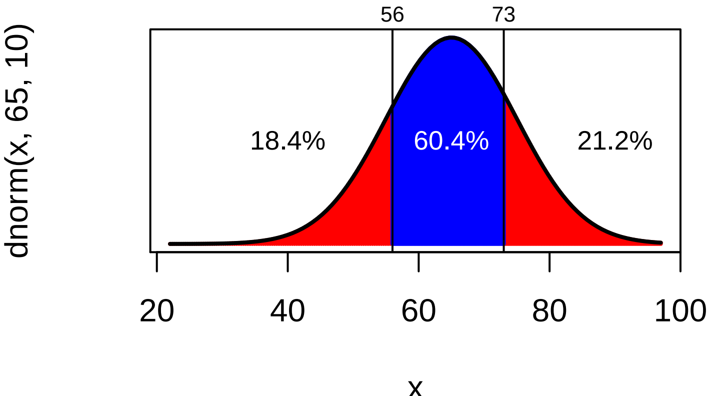
A new island in the Pacific has been discovered and it is our job to estimate the mean heights of the Men/Women that live there. What should our prior on the mean be?

# Prior Distribution

Mean heights range from 56 inches (Bolivia Females) to 73 inches (Dinaric Alps Males) [Wikipedia article]. With world records for individuals of 22 inches and 97 inches.

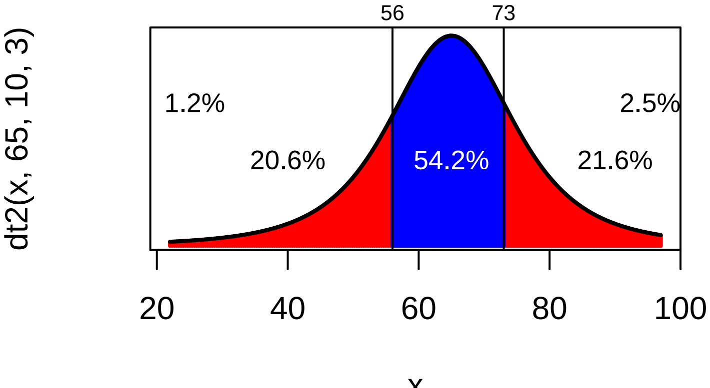
# Prior Distribution

A Prior Distribution of a Normal with mean 65 inches and standard deviation of 10 inches:



# Prior Distribution

A shifted and scaled t-distribution with 3 degrees of freedom gives a little more tail area:



# Live Demonstration

One of our deep space probes has discovered an inhabitable planet. We are tasked with estimating the proportion of the surface covered with Land vs. Water to help prepare our ~~invasion force~~ exploratory teams.

# Live Demonstration

If the globe comes to you, catch it, and look at where your **Left Index** finger is touching the globe. Determine if it is touching Land or Water (if both, or an animal, or a symbol, then figure out which is most appropriate). Yell out "Land" or "Water", then give the globe a spin and toss it (gently) to someone else.

# Live Demonstration

Do Demo

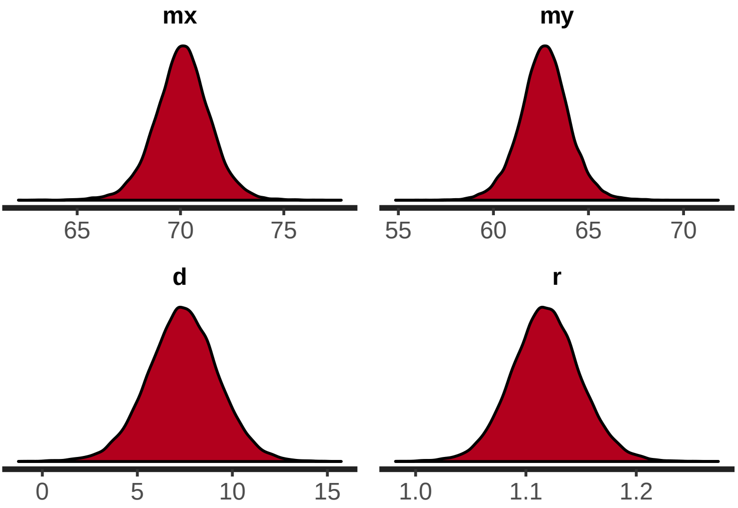
# Compare to T-test



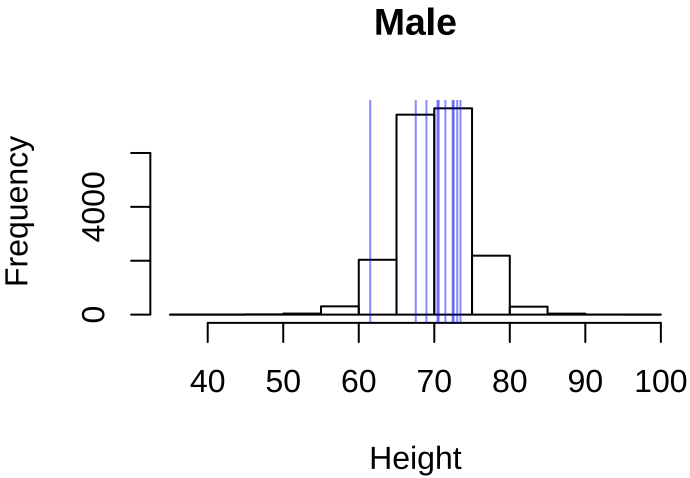
10/1,000 samples of heights of men/women from our island (actually from the NHANES data).

| n | Males | Females | Difference |
|---|---|---|---|
| n=10 | | | |
| T-test | 70.2 (67.6-72.8) | 62.7 (60.1-65.2) | 7.6 (4.2-10.9) |
| Bayes | 70.1 (67.4-72.8) | 62.7 (60.1-65.4) | 7.4 (3.6-11.2) |
| n=1,000 | | | |
| T-test | 69.2 (69.0-69.4) | 63.6 (63.4-63.7) | 5.6 (5.4-5.9) |
| Bayes | 69.2 (69.0-69.4) | 63.6 (63.4-63.7) | 5.6 (5.4-5.9) |

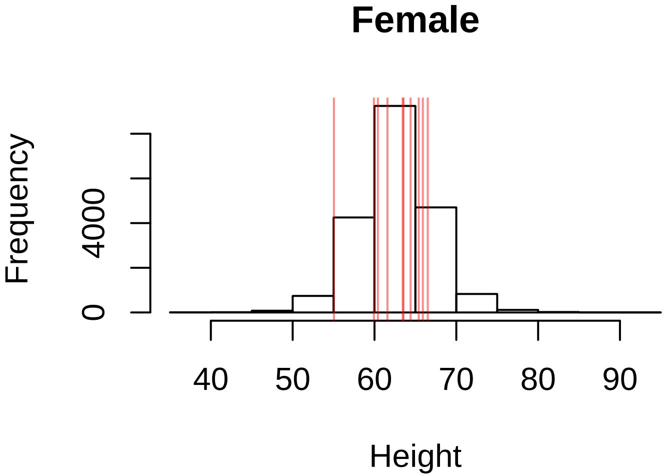
# Posterior Distribution



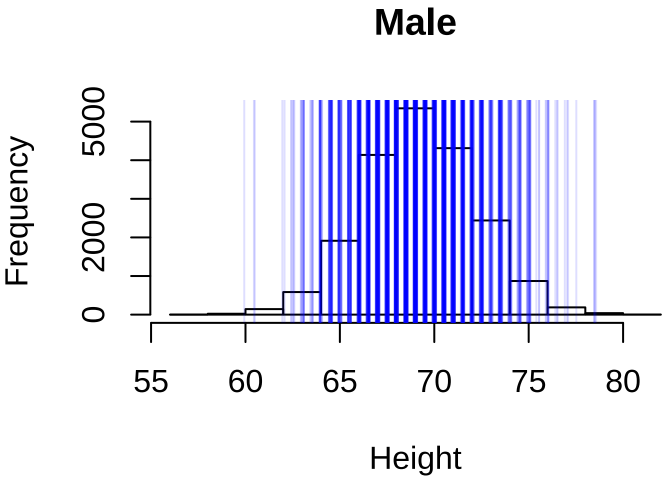
# Posterior Predictive distribution



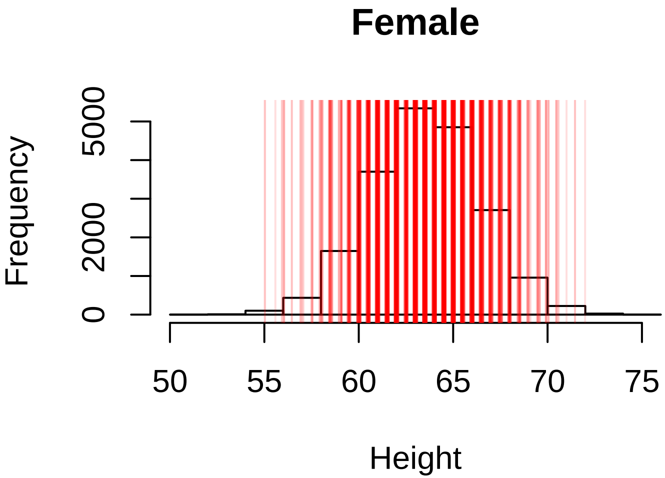
# Posterior Predictive distribution



# Posterior Predictive distribution



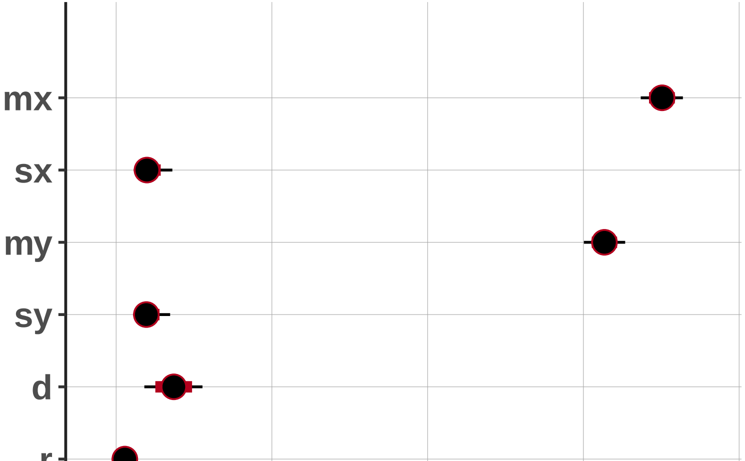
# Posterior Predictive distribution



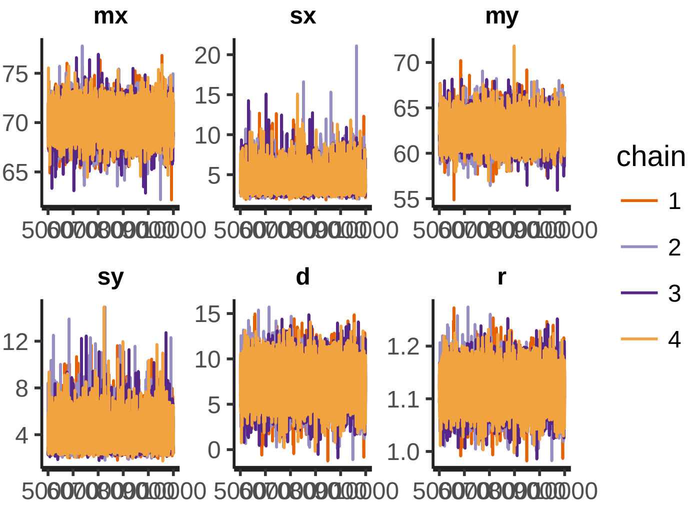
# Additional Plots

`ci_level: 0.8 (80% intervals)`

`outer_level: 0.95 (95% intervals)`



# Additional Plots



# Additional Output

```
Inference for Stan model: e578df46739f2ea996ab65c2b0ad77f3.
4 chains, each with iter=10000; warmup=5000; thin=1;
post-warmup draws per chain=5000, total post-warmup draws=200

        mean se_mean   sd   2.5%    25%    50%
mx     70.10    0.01 1.36  67.37  69.27  70.11
sx      4.19    0.01 1.23   2.56   3.36   3.96
my     62.70    0.01 1.32  60.06  61.88  62.70
sy      4.09    0.01 1.16   2.51   3.28   3.88
d       7.39    0.02 1.89   3.62   6.19   7.41
r       1.12    0.00 0.03   1.06   1.10   1.12
lp__ -34.14    0.02 1.56 -38.04 -34.91 -33.78
         75%  97.5% n_eff Rhat
mx     70.94  72.78 14444    1
sx      4.74   7.22 13097    1
my     63.52  65.36 15716    1
```

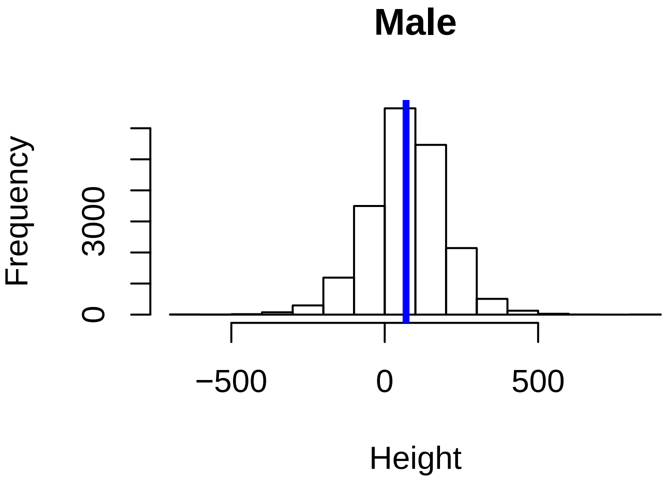
# Wrong Scale

What if we accidentally used height in cm, while prior is in inches?

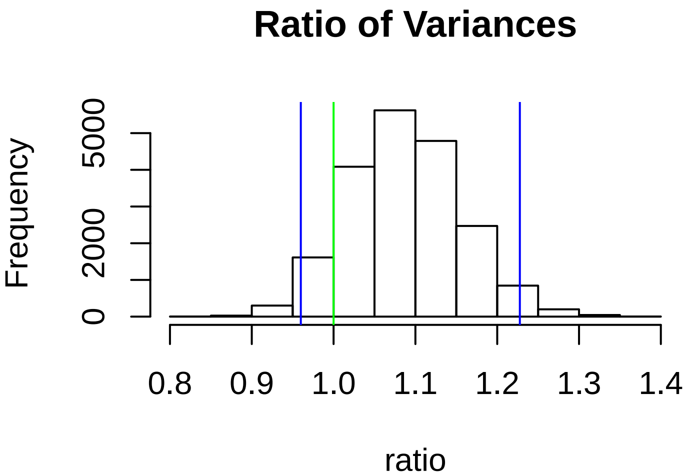
| n=10 | Male |
|---|---|
| T-Test | 178.4 (172.0-184.8) |
| Bayes | 73.7 (53.0-94.7) |

Bayes estimate of SD = 120, SD of data = 3.6

# Posterior Predictive distribution



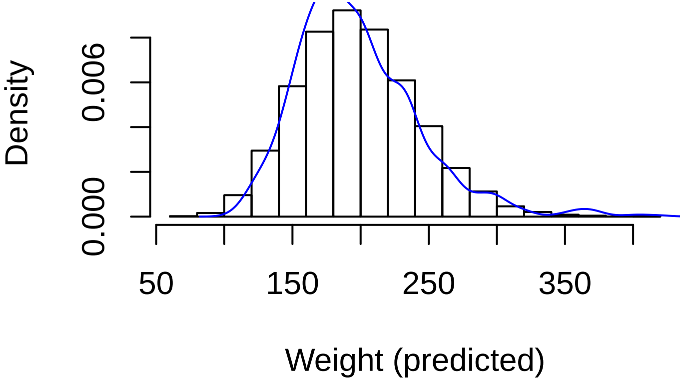
# Unequal Variances



# Non-normal Likelihood

```
             mean       se_mean          sd
alpha     21.0756732 0.01514169895 0.926085385
beta       0.1067342 0.00007759381 0.004751169
mn       197.4692361 0.00978816888 1.363454862
lp__   -5168.7387699 0.01479934633 0.985483201
                2.5%           25%            50%
alpha     19.30968951    20.4351054     21.0628790
beta       0.09775751     0.1034473      0.1066889
mn       194.80272080   196.5541499    197.4684955
lp__   -5171.41769041 -5169.1276888 -5168.4325244
               75%         97.5% n_eff      Rhat
alpha     21.694822    22.9514038  3741 1.000346
beta       0.109917     0.1163881  3749 1.000405
mn       198.371862   200.1701892 19403 1.000025
lp__   -5168.029995 -5167.7712281  4434 1.000006
```

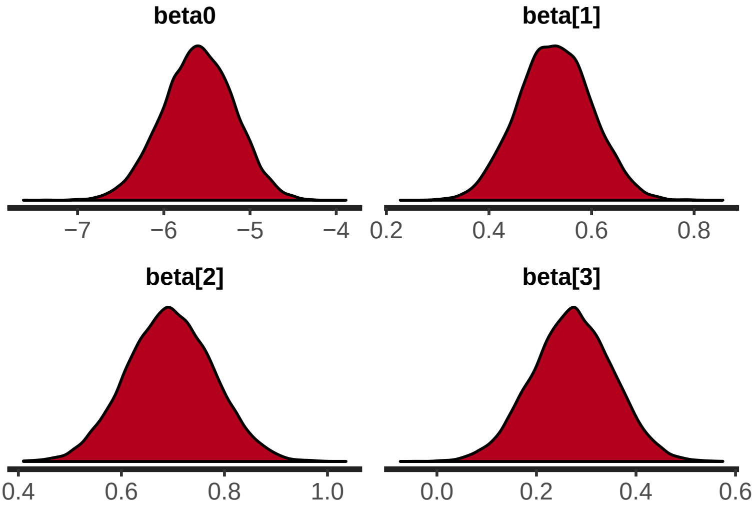
# Non-normal Posterior Prediction



# Logistic Regression

```
            mean  se_mean     sd     2.5%
beta0     -5.600 0.004574 0.4140   -6.422
beta[1]    0.529 0.000726 0.0717    0.390
beta[2]    0.695 0.000801 0.0834    0.534
beta[3]    0.273 0.000749 0.0812    0.113
lp__    -462.070 0.015459 1.4116 -465.661
             25%      50%      75%    97.5% n_eff
beta0     -5.879   -5.596   -5.314   -4.806  8193
beta[1]    0.480    0.528    0.577    0.669  9757
beta[2]    0.638    0.694    0.752    0.859 10838
beta[3]    0.219    0.273    0.328    0.431 11772
lp__    -462.767 -461.747 -461.040 -460.304  8338
        Rhat
beta0      1
beta[1]    1
beta[2]    1
```

# Logistic Regression



# Muliple Comparisons

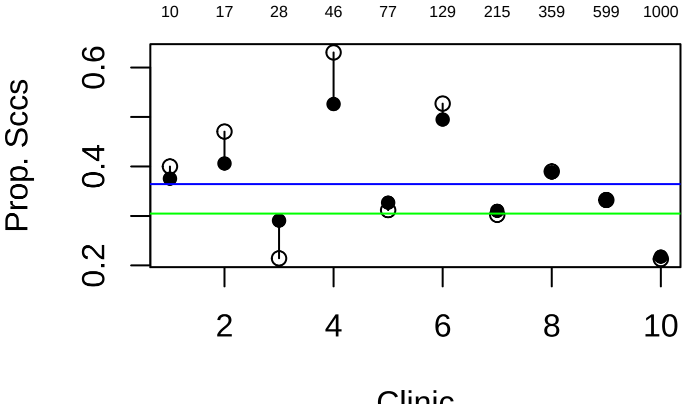


Simulate 10 groups with n=30, mean=100, sd=5 and look at all pairwise comparisons. Look at biggest difference. Compare to Bayesian Hierarchical model.

| n=30 | Biggest Difference |
|---|---|
| Uncorrected | 102.51 - 98.82 = 3.69 (1.07-6.30) |
| Tukey | 102.51 - 98.82 = 3.69 (-0.35-7.72) |
| Bayes Hier | 101.27 - 99.83 = 1.44 (-0.25-4.03) |

# Partial Pooling

We are looking at the success rate for a particular procedure at 10 different clinics (whose success rates probably vary).



# Advantages / Disadvantages



- Flexibility
- Incorporating prior information
- Using different distributions
- Requires more thought
- Not as standard (more explanation/justification)
- Longer analysis time (sometimes shorter)

# Want to Learn More

- The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy by Sharon Bertsch McGrayne

- Statistical Rethinking by Richard McElreath
  `https://xcelab.net/rm/statistical-rethinking/`

# R packages

- Bayesian Task View on CRAN
- rethinking – package to accompany the book Stastical Rethinking
- BRUGS, R2WinBUGS – Interface packages to OpenBUGS and WinBUGS
- rstan – Interface with Stan program (compiles code to C++)
- nimble – R package to create/compile/run C++
- brms, rstanarm – common models and interfaces, uses rstan to do the computations
- loo – Leave One Out cross-validation for Bayesian models