# MATH 640: Exam 1

Name:_____

## Instructions

1. Do not discuss this midterm with anyone other than Professor Meyer or the TA; and you may only ask clarifying questions.

2. Your responses to this exam must be typed.

3. This cover sheet *must* be the first page of your submission.

4. If you wish to cite a result or derivation from lecture, you may do so but be sure it is clearly cited (list the slide from the Note Set or the example) and relevant.

5. Please submit the exam to the assignment page on Canvas by **11:59pm on Sunday, March 14**.

6. Late submissions will be accepted up to 24 hours after the deadline, however they will be penalized: 4 points off for every six hours the submission is late.

| Portion | Question | Points | Score |
|---------|----------|--------|-------|
| Theory | 1 | 15 | |
| | 2 | 15 | |
| | 3 | 20 | |
| | 4 | 10 | |
| Computing | 1 | 20 | |
| | 2 | 20 | |
| Total | | 100 | |

# Midterm

## Jeff Gould

## 3/14/2021

## Theory

**1)**

**a)**

$\tau^2 = 1/\sigma^2 \rightarrow \sigma^2 = 1/\tau^2$

$$\left(\frac{2}{\pi\sigma^2}\right)^{1/2} \exp\left[-\frac{1}{2\sigma^2}s_i^2\right] = \left(\frac{2}{\pi(1/\tau^2)}\right)^{1/2} \exp\left[-\frac{1}{2(1/\tau^2)}s_i^2\right] = \left(\frac{2\tau^2}{\pi}\right)^{1/2} \exp\left(-\frac{\tau^2}{2}s_i^2\right) 1(s_i > 0)$$

Find Jeffrey's Prior:

$\log \mathcal{L} = \frac{1}{2}\log(2\tau^2) - \frac{1}{2}\log(\pi) - \frac{\tau^2}{2}s_i$

$\frac{\partial}{\partial\tau^2} = \frac{1}{2\tau^2} - \frac{s_i^2}{2}$

$\frac{\partial^2}{\partial(\tau^2)^2} = -(2\tau^2)^{-2}$

$-E[-(2\tau^2)^{-2}|\tau^2] = (2\tau^2)^{-2} \propto (\tau^2)^{-2}$

Take the square root to get Jeffrey's prior:

$\pi(\tau^2) \propto \frac{1}{\tau^2}$

**b)**

$$P(\tau^2|S) = \frac{1}{\tau^2} \prod \left(\frac{2\tau^2}{\pi}\right)^{1/2} \exp\left(-\frac{\tau^2}{2}s_i^2\right)$$

$$= (\tau^2)^{-1} \left(\frac{2\tau^2}{\pi}\right)^{n/2} \exp\left[-\frac{\tau^2}{2}\sum s_i^2\right]$$

$$\propto (\tau^2)^{n/2-1} \exp\left[-\frac{\tau^2}{2}\sum s_i^2\right]$$

Which we recognize as the kernel for a gamma distribution with parameters $\alpha = \frac{n}{2}$ and $\beta = \frac{1}{2}\sum s_i^2$

$\tau^2 \sim Gamma(\frac{n}{2}, \frac{1}{2}\sum s_i^2)$

**c)**

Since $\sigma^2$ has an Inverse-Gamma distribution, the kernel for it's distribution is

$(1/\sigma^2)^{\alpha+1} \exp[-\beta/\sigma^2] = (\tau^2)^{\alpha+1} \exp[-\beta\tau^2]$

Which is the kernel for a Gamma distribution. So, since the identity replacement within the kernel transforms us from IG to Gamma, placing the prior on the precision is the same as placing it on th variation since we could simply reverse the equality to get the distribution kernel for the variance

**2)**

**a)**

Define $\bar{x} = \frac{1}{n} \sum \log(x_i)$ and define $s^2 = \frac{1}{n-1} \sum (\log(x_i) - \bar{x})^2$

$$\pi(\mu,\sigma^2)\mathcal{L}(x_i|\mu,\sigma^2) = (\sigma^2)^{-1} \prod \frac{1}{x_i\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(\log(x_i)-\mu)^2\right]$$

$$\propto (\sigma^2)^{-1}(\sigma^2)^{-n/2} \exp\left[-\frac{1}{2}\sigma^2 \sum(\log(x_i)-\mu)^2\right]$$

$$\propto (\sigma^2)^{-(n/2+1)} \exp\left[-\frac{1}{2\sigma^2} \sum(\log(x_i)-\bar{x}+\bar{x}-\mu)^2\right]$$

$$\propto (\sigma^2)^{-(n/2+1)} \exp\left[-\frac{1}{2\sigma^2} \sum\{(\log(x_i)-\bar{x})^2 + 2(\log(x_i)-\bar{x})(\bar{x}-\mu) + (\bar{x}-\mu)^2\}\right]$$

$$\propto (\sigma^2)^{-(n/2+1)} \exp\left[-\frac{1}{2\sigma^2}\{\sum(\log(x_i)-\bar{x})^2 + \sum 2(\log(x_i)-\bar{x})(\bar{x}-\mu) + \sum(\bar{x}-\mu)^2\}\right]$$

$$\propto (\sigma^2)^{-(n/2+1)} \exp\left[-\frac{1}{2\sigma^2}\{\sum(\log(x_i)-\bar{x})^2 + n(\bar{x}-\mu)^2\}\right]$$

$$\propto (\sigma^2)^{-(n/2+1)} \exp\left[-\frac{1}{2\sigma^2}\{(n-1)s^2 + n(\bar{x}-\mu)^2\}\right]$$

$$\propto (\sigma^2)^{-(n/2+1)} \exp\left[-\frac{1}{2\sigma^2}(n-1)s^2\right] \exp\left[-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2\right]$$

So $P(\mu|\sigma^2, X) \propto \exp\left[-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2\right] = \exp\left[-\frac{1}{2(\sigma^2/n)}(\bar{x}-\mu)^2\right]$

$P(\mu|\sigma^2, X) \sim \mathcal{N}(\bar{x}, \sigma^2/2) = \mathcal{N}(\frac{1}{n}\sum \log(x_i), \frac{\sigma^2}{n})$

**b)**

Note the support for $\mu$ is given as all real numbers, so $\mu \in (-\infty, \infty)$

2

$$P(\sigma^2|X) \propto \int_{-\infty}^{\infty} (\sigma^2)^{-n/2+1} \exp\left[-\frac{1}{2\sigma^2}(n-1)s^2\right] \exp\left[-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2\right] d\mu$$

$$\propto (\sigma^2)^{-n/2+1} \exp\left[-\frac{1}{2\sigma^2}(n-1)s^2\right] \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2(\sigma^2/n)}(\bar{x}-\mu)^2\right] d\mu$$

$$\propto (\sigma^2)^{-n/2+1} \exp\left[-\frac{1}{2\sigma^2}(n-1)s^2\right] \int_{-\infty}^{\infty} \frac{\sqrt{2\pi\sigma^2/n}}{2\pi\sigma^2/n} \exp\left[-\frac{1}{2(\sigma^2/n)}(\bar{x}-\mu)^2\right] d\mu$$

$$\propto (\sigma^2)^{-n/2+1} \exp\left[-\frac{1}{2\sigma^2}(n-1)s^2\right] \sqrt{2\pi\sigma^2/n} \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma^2/n} \exp\left[-\frac{1}{2(\sigma^2/n)}(\bar{x}-\mu)^2\right] d\mu$$

$$\propto (\sigma^2)^{-n/2+1} \exp\left[-\frac{1}{2\sigma^2}(n-1)s^2\right] \sqrt{2\pi\sigma^2/n}$$

$$\propto (\sigma^2)^{-n/2+1} \exp\left[-\frac{1}{2\sigma^2}(n-1)s^2\right]$$

Which we recogniaze as the kernel for an Inverse-Gamma distribution

$$\sigma^2 \sim IG(\tfrac{n-1}{2}, \tfrac{n-1}{2}s^2) = IG(\tfrac{n-1}{2}, \tfrac{1}{2}\sum(\log(x_i)-\bar{x})^2)$$

**3)**

**a)**

$$\mathcal{L} = \prod_{i=1}^{n} b_0\eta \exp[b_0 t_i + \eta] \exp[-\eta e^{bt_i}]$$

$$\propto (b_0\eta)^n \exp\left[\sum(b_0 t_i + \eta)\right] \exp\left[-\sum \eta e^{b_0 t_i}\right]$$

$$\propto \eta^n \exp\left[\sum(b_0 t_i) + n\eta\right] \exp\left[-\sum \eta e^{b_0 t_i}\right]$$

$$\propto \eta^n \exp[\sum b_0 t_i] \exp[n\eta] \exp[-\eta \sum e^{b_0 t_i}]$$

$$\propto \eta^n \exp[n\eta] \exp[-\eta \sum e^{b_0 t_i}]$$

$\pi(\eta) = \eta^{-1}$

So, $P(\eta) \propto \eta^{n-1} \exp[n\eta] \exp[-\eta \sum e^{b_0 t_i}]$

**b)**

$$\log P = \log[\eta^{n-1} \exp[n\eta] \exp[-\eta \sum e^{b_0 t_i}]] = (n-1)\log(\eta) + n\eta - \eta \sum e^{b_0 t_i}$$

$$\frac{\partial \log P}{\partial \eta} = \frac{n-1}{\eta} + n - \sum e^{b_0 t_i} = 0$$

$$\frac{n-1}{\eta} = \sum e^{b_0 t_i} - n$$

$$\hat{\eta} = \frac{n-1}{\sum e^{b_0 t_i} - n}$$

## c) Observed Information

$$\frac{\partial^2 \log P}{\partial \eta^2} = -(n-1)(\eta)^{-2} + 0 + 0$$

$$I = (n-1)(\hat{\eta})^{-2} = \frac{n-1}{\left(\frac{n-1}{\sum e^{b_0 t_i} - n}\right)^2} = \frac{(\sum e^{b_0 t_i} - n)^2}{n-1}$$

## d)

Since we have the mode and the observed information, we can sample $\eta | t_1, \ldots, t_n, b_0$ using the approximation to the normal:

$$Y \sim \mathcal{N}(\hat{\theta}, I[\hat{\theta}]^{-1})$$

This gives us

$$\eta \sim \mathcal{N}\left(\frac{n-1}{\sum e^{b_0 t_i}}, \frac{n-1}{(\sum e^{b_0 t_i})^2}\right)$$

## 4)

Since we only have one observation, the likelihood is simply the pmf

$$\mathcal{L} = \Gamma(\sum x_i) \frac{\theta_0^{x_0}}{\Gamma(x_0)} \prod_{i=1}^{3} \frac{\theta_i^{x_i}}{x_i!} \propto \theta_0^{x_0} \theta_1^{x_1} \theta_2^{x_2} \theta_3^{x_3}$$

Which we recognize as the Kernel for a Dirichlet distribution

The non-informative prior for a Dirichlet distribution is $\alpha_0 = \alpha_1 = \alpha_2 = \alpha_3 = 1$

# Analysis

## a)

$$f(w_i) = \frac{w_i}{\theta^2} \exp(-\frac{w_i^2}{2\theta^2})$$

$$\begin{aligned}
\mathcal{L} &= \prod \frac{w_i}{\theta^2} \exp(-\frac{w_i^2}{2\theta^2}) \\
&= (\frac{1}{\theta^2})^n \exp(-\frac{\sum w_i^2}{2\theta^2}) \prod w_i \\
&\propto (\theta^2)^{-n} \exp(-\frac{\sum w_i^2}{2\theta^2})
\end{aligned}$$

Which is the kernel for an Inverse Gamma distribution

So choose a non-informative IG prior: $\pi(\theta^2) \propto (\theta^2)^{-1}$

This is an improper prior, but also wht we would get with an IG distribution with both parameters set to 0. The smaller both parameters are in an IG distribution, the less information in the prior, with the prior becoming completely non-informative when both parameters are equal to 0.

$$P(\theta|W) \propto (\theta^2)^{-1}(\theta^2)^{-n} \exp(-\frac{\sum w_i^2}{2\theta^2})$$

$$= (\theta^2)^{-n-1} \exp\left(-\frac{\sum w_i^2}{2\theta^2}\right)$$

The kernel for an inverse-gamma distribution with parameters $\alpha = n$ and $\beta = \frac{\sum w_i^2}{2}$

$$\theta^2 \sim IG(n, \frac{\sum w_i^2}{2})$$

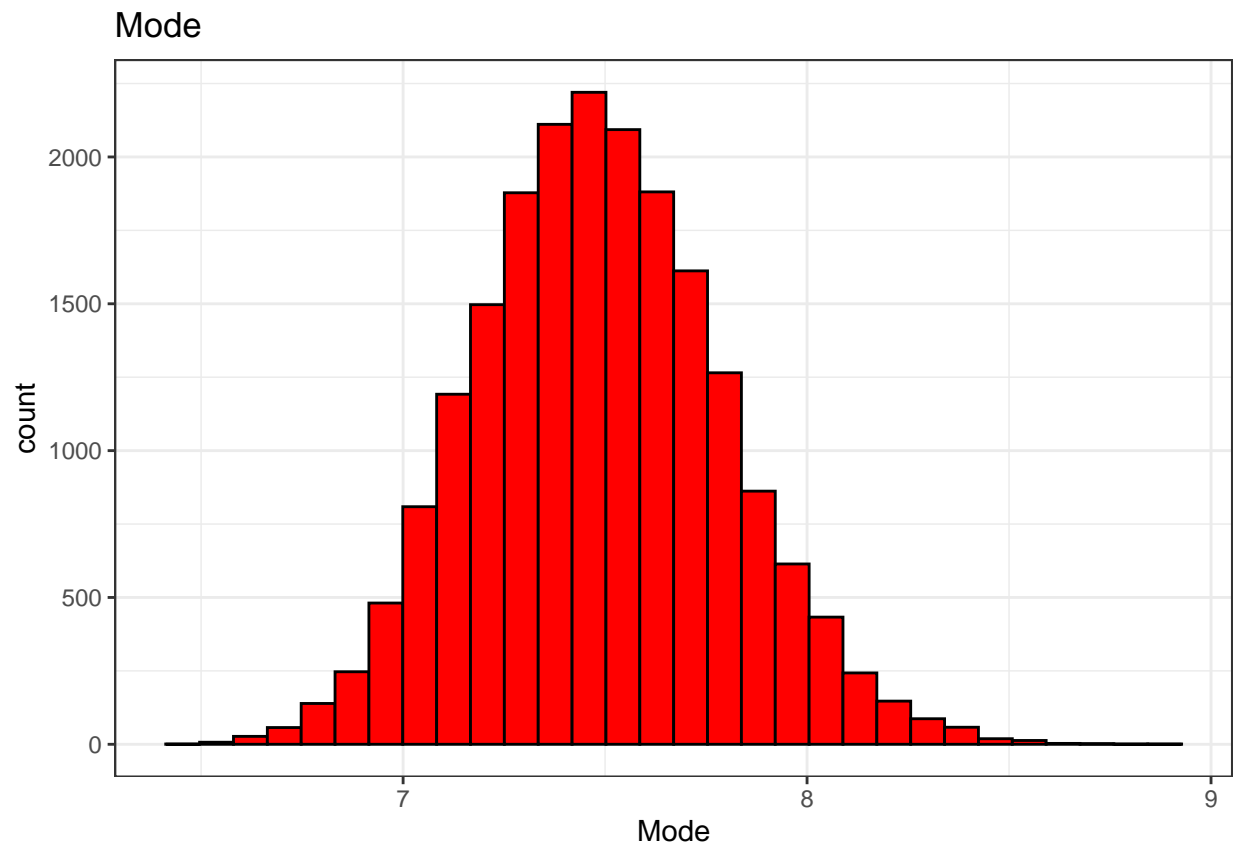```r
wind <- read.delim("wind.txt", header = F)

alpha <- nrow(wind)
beta <- sum(wind^2) / 2


set.seed(6302)
theta <- MCMCpack::rinvgamma(n = 20000, shape = alpha, scale = beta)


theta.df <- data.frame(Theta2 = theta) %>%
  dplyr::mutate(Mode = sqrt(Theta2),
                Median = Mode*sqrt(2 * log(2)),
                Mean = Mode * sqrt(pi / 2))


ggplot(theta.df, aes(x = Mode)) +
  geom_histogram(col = "black", fill = "red") +
  theme_bw() +
  labs(title = "Mode")
```
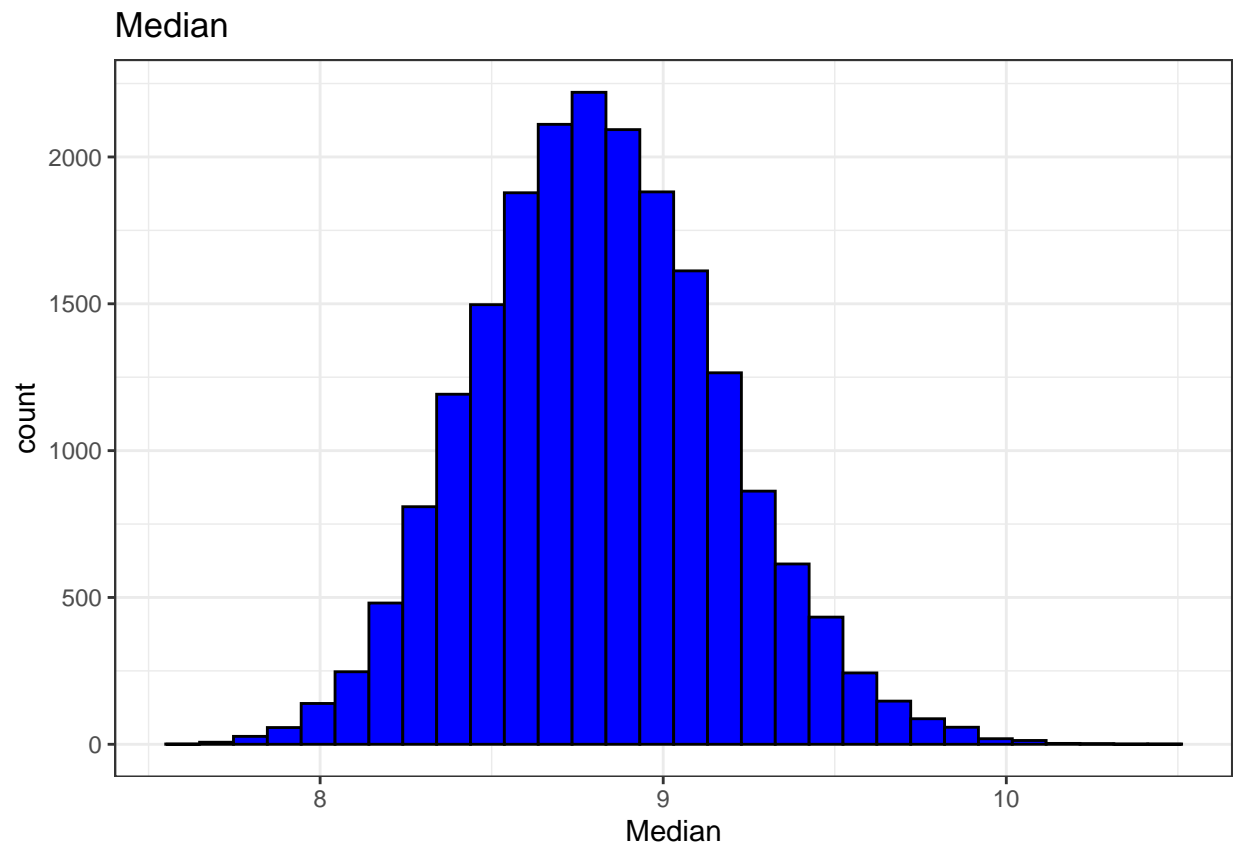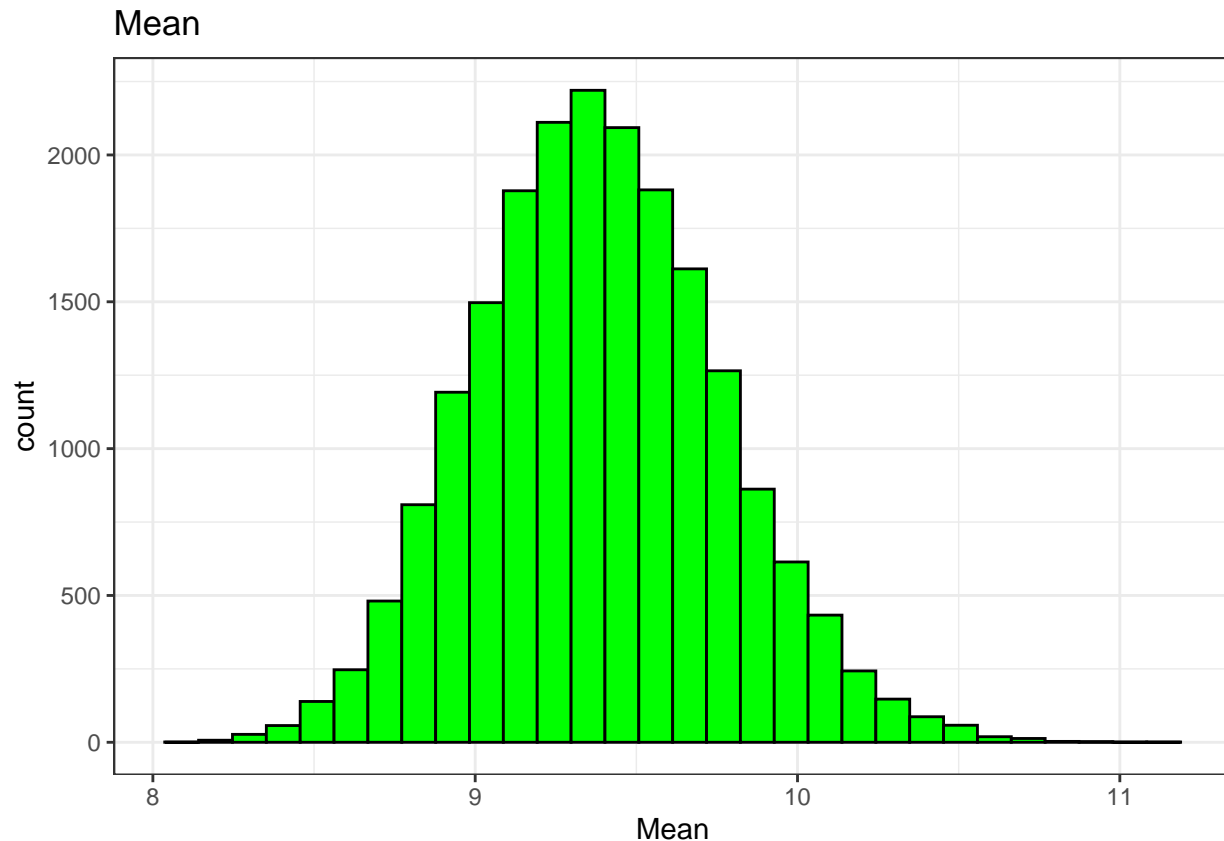
# Mode



```r
ggplot(theta.df, aes(x = Median)) +
  geom_histogram(col = "black", fill = "blue") +
  theme_bw() +
  labs(title = "Median")
```

## Median



```r
ggplot(theta.df, aes(x = Mean)) +
  geom_histogram(col = "black", fill = "green") +
  theme_bw() +
  labs(title = "Mean")
```

Mean

```
round(
apply(theta.df[,2:4], 2, quantile, probs = c(0.5, 0.025, 0.975)), 3
)
```

```
##          Mode Median   Mean
## 50%     7.478  8.804  9.372
## 2.5%    6.922  8.150  8.675
## 97.5%   8.115  9.555 10.171
```

To find the probabilty of a day having wind speeds in excess of 15 mph, we could either evaluate at our average $\theta^2$ value, or evaluate the integral for all $\theta^2$ values and then average (It is the same value). The CDF for the distribution is $1 - \exp[-x^2/(2\theta^2)]$, so given a $\theta^2$, the orobabilty of wind speeds exceeding 15 mph is simply $\exp[-(15^2)/(2\theta^2)]$

The median probabilty is about 13.37% that wind speeds could exceed 15 mph, however with the higher values of $\theta^2$ that we drew, we see an upper-bound of an 18.11% chance that wind speeds could be that high.
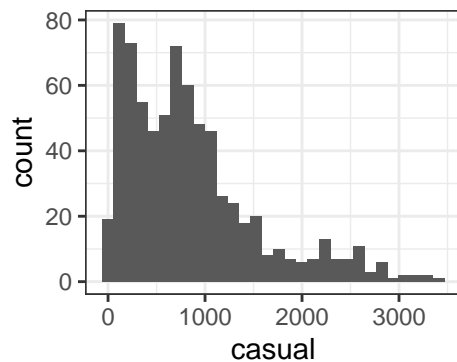
```
wind_probs <- exp(-(15)^2 / (2 * theta))
quantile(wind_probs, c(0.5, 0.025, 0.975))
```

```
##         50%        2.5%       97.5%
## 0.13372408 0.09554032 0.18116432
```
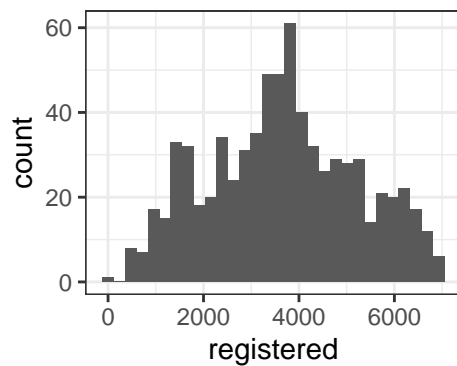
**b)**

```
bikeShare <- read.delim("day.txt", sep = " ")

ggplot(bikeShare, aes(x = casual)) +
  geom_histogram() +
  theme_bw()
```



```
ggplot(bikeShare, aes(x = registered)) +
  geom_histogram() +
  theme_bw()
```



It looks like the number of registered bikers in a day is relatively normally distributed. However, the number of casual bikers in a day is not normally distributed, and has a right-tail. Therefore we will log-transform the number of casual bikers

Registered Biker model:

```
B     <- 20000
n     <- nrow(bikeShare)
X     <- model.matrix(
  registered ~ yr + holiday + workingday+ temp + atemp + hum + windspeed,
  data = bikeShare
  )
Y     <- bikeShare$registered
K     <- ncol(X)
```

```
bhat     <- c(solve(t(X)%*%X)%*%(t(X)%*%Y))
SSY <- t(Y - X%*%bhat)%*%(Y - X%*%bhat)
XtXi     <- solve(t(X)%*%X)
rbeta    <- matrix(0, nrow = B, ncol = K)

set.seed(726)
rsig     <- rinvgamma(B, (n-K)/2, (1/2)*SSY)
for(i in 1:B){
  CovX      <- rsig[i]*XtXi
  rbeta[i,] <- c(rmvnorm(1, mean = bhat, sigma = CovX))
}

rbMat      <- apply(rbeta, 2, quantile, probs = c(0.5, 0.025, 0.975))
postp    <- apply(rbeta > 0, 2, mean)
outMat   <- cbind(t(rbMat), postp)
rownames(outMat) <- colnames(X)
colnames(outMat) <- c(rownames(rbMat), 'P(b > 0)')
round(outMat,4)
```

```
##                    50%        2.5%       97.5% P(b > 0)
## (Intercept)  1492.2842  1068.4875  1909.3768   1.0000
## yr           1736.4844  1618.6585  1855.6594   1.0000
## holiday      -277.4642  -644.3167    87.0738   0.0679
## workingday    917.3964   785.6478  1048.8944   1.0000
## temp          510.8692 -2049.0673  3038.0597   0.6562
## atemp        4151.7529  1310.2404  7040.4651   0.9978
## hum         -1535.4455 -1976.5875 -1100.1794   0.0000
## windspeed   -3049.2338 -3863.0461 -2259.5336   0.0000
```

Building an initial model on all the possible covariates, we see that temperature and holiday are not statistically significant predictors. Let's now remove them and build a new model

```
B      <- 20000
n      <- nrow(bikeShare)
X      <- model.matrix(registered ~ yr  + workingday + atemp + hum + windspeed,
                   data = bikeShare)
Y    <- bikeShare$registered
K    <- ncol(X)

bhat     <- c(solve(t(X)%*%X)%*%(t(X)%*%Y))
SSY <- t(Y - X%*%bhat)%*%(Y - X%*%bhat)
XtXi     <- solve(t(X)%*%X)
rbeta    <- matrix(0, nrow = B, ncol = K)

set.seed(726)
rsig     <- rinvgamma(B, (n-K)/2, (1/2)*SSY)
for(i in 1:B){
  CovX      <- rsig[i]*XtXi
  rbeta[i,] <- c(rmvnorm(1, mean = bhat, sigma = CovX))
}

rbMat      <- apply(rbeta, 2, quantile, probs = c(0.5, 0.025, 0.975))
postp    <- apply(rbeta > 0, 2, mean)
```

```
outMat     <- cbind(t(rbMat), postp)
rownames(outMat) <- colnames(X)
colnames(outMat) <- c(rownames(rbMat), 'P(b > 0)')
round(outMat,4)
```

```
##                   50%        2.5%      97.5% P(b > 0)
## (Intercept)  1438.6355  1037.3529  1848.109        1
## yr           1736.0399  1616.8166  1854.465        1
## workingday    942.2605   814.0901  1069.614        1
## atemp        4734.7868  4365.1382  5105.350        1
## hum         -1539.0372 -1970.6135 -1101.620        0
## windspeed   -3011.4854 -3810.1584 -2218.712        0
```

```
registered_summary <- outMat
registered_model <- rbeta
```

Now we have a model with all statistically significant variables

And predict the number of riders on February 29, 2012:

```
theta <- matrix(c(1, 1, 1, 0.34847, 0.804783, 0.179117), ncol = 1)
predicted <- data.frame(Registered = registered_model %*% theta)
```

Now Casual riders:

```
B       <- 20000
n       <- nrow(bikeShare)
bikeShare$logCasual <- log(bikeShare$casual)

X       <- model.matrix(
  logCasual ~ yr + holiday + workingday+ temp + atemp + hum + windspeed,
  data = bikeShare
  )
Y       <- bikeShare$logCasual
K       <- ncol(X)

bhat    <- c(solve(t(X)%*%X)%*%(t(X)%*%Y))
SSY <- t(Y - X%*%bhat)%*%(Y - X%*%bhat)
XtXi    <- solve(t(X)%*%X)
rbeta   <- matrix(0, nrow = B, ncol = K)

set.seed(97)
rsig    <- rinvgamma(B, (n-K)/2, (1/2)*SSY)
for(i in 1:B){
  CovX      <- rsig[i]*XtXi
  rbeta[i,] <- c(rmvnorm(1, mean = bhat, sigma = CovX))
}

rbMat     <- apply(rbeta, 2, quantile, probs = c(0.5, 0.025, 0.975))
postp     <- apply(rbeta > 0, 2, mean)
outMat    <- cbind(t(rbMat), postp)
rownames(outMat) <- colnames(X)
colnames(outMat) <- c(rownames(rbMat), 'P(b > 0)')
round(outMat,4)
```

```
##              50%    2.5%   97.5% P(b > 0)
## (Intercept)  5.9222  5.6340  6.2080   1.0000
## yr           0.3747  0.2922  0.4568   1.0000
## holiday     -0.2440 -0.4962  0.0060   0.0281
## workingday  -0.8972 -0.9878 -0.8077   0.0000
## temp         0.3857 -1.3762  2.1307   0.6671
## atemp        3.8582  1.8860  5.8469   1.0000
## hum         -1.3196 -1.6160 -1.0235   0.0000
## windspeed   -1.7366 -2.3008 -1.1842   0.0000
```

Again, temperature and holiday are not statistically significant predictors. Remove them from model and build new one

```r
B      <- 20000
n      <- nrow(bikeShare)
bikeShare$logCasual <- log(bikeShare$casual)

X      <- model.matrix(logCasual ~ yr + workingday + atemp + hum + windspeed,
                        data = bikeShare)
Y      <- bikeShare$logCasual
K      <- ncol(X)

bhat    <- c(solve(t(X)%*%X)%*%(t(X)%*%Y))
SSY <- t(Y - X%*%bhat)%*%(Y - X%*%bhat)
XtXi    <- solve(t(X)%*%X)
rbeta   <- matrix(0, nrow = B, ncol = K)

set.seed(97)
rsig    <- rinvgamma(B, (n-K)/2, (1/2)*SSY)
for(i in 1:B){
  CovX      <- rsig[i]*XtXi
  rbeta[i,] <- c(rmvnorm(1, mean = bhat, sigma = CovX))
}

rbMat      <- apply(rbeta, 2, quantile, probs = c(0.5, 0.025, 0.975))
postp    <- apply(rbeta > 0, 2, mean)
outMat   <- cbind(t(rbMat), postp)
rownames(outMat) <- colnames(X)
colnames(outMat) <- c(rownames(rbMat), 'P(b > 0)')
round(outMat,4)
```

```
##              50%    2.5%   97.5% P(b > 0)
## (Intercept)  5.8795  5.6049  6.1553        1
## yr           0.3738  0.2906  0.4562        1
## workingday  -0.8746 -0.9623 -0.7877        0
## atemp        4.2966  4.0403  4.5484        1
## hum         -1.3215 -1.6200 -1.0268        0
## windspeed   -1.7144 -2.2724 -1.1637        0
```

```r
casual_summary <- outMat
casual_model <- rbeta
```

We end up seeing the same inputs are used for both models. Neither one uses `holiday` or `temp`. This makes sense as most of the necessary signal in `holiday` lies simply in them not being a `workingday`. And then

most of the information in `temp` can be found in `atemp`, and the other factors that change `temp` from `atemp` are mostly found in `hum` and `windspeed`

Directionally, all the variables are the same excpet for `workingday`. For our registered user model, there is a positive relationship between `workingday` and number of riders. In our casual model, `workingday` has a negative impact on the number of riders. This makes some sense as many registered users likely use the bikes as their method of transportation to and from work, while casual riders are probably using the bikes to explore the city on weekends.

And here is our projection for number of riders on February 29, 2012:

```r
theta <- matrix(c(1, 1, 1, 0.34847, 0.804783, 0.179117), ncol = 1)
predicted$Casual <- exp(casual_model %*% theta)

predRiders  <- apply(predicted, 2, quantile, probs = c(0.5, 0.025, 0.975))
meanPred    <- apply(predicted , 2, mean)
outMat   <- cbind(t(predRiders), meanPred)
outMat
```

```
##                   50%      2.5%     97.5%  meanPred
## Registered 3989.5648 3852.3489 4126.5145 3989.4965
## Casual      245.7582  223.4443  270.3632  246.0946
```