Name ____

Spring 2017 **STAT 401D** **Exam II**

(100 points)

Instructions:

- Full credit will be given only if you show your work.
- The questions are not necessarily ordered from easiest to hardest.
- You are allowed to use any resource except aid from another individual.
- Aid from another individual, will automatically earn you a 0.

1. Please answer the following questions using your own words. If I find that you have copied your answers from the internet, you will receive a 0.

   (a) What is the likelihood? (5 points)

   Answer: The likelihood is the joint probability mass or density function thought of as a function of the parameters.

   (b) What does a pvalue measure? (5 points)

   Answer: The pvalue measures how incompatible the data are with the model associated with the null hypothesis. It is also a measure of how extreme a test statistic could have been had you observed different data.

   (c) What does "95% confidence" mean in a 95% confidence interval? (5 points)

   Answer: A 95% confidence interval means that if you were to repeat the procedure that constructed the confident interval on different data sets, then on average 95% of these confidence intervals would cover the truth.

   (d) What does "95% credible" mean in a 95% credible interval? (5 points)

   Answer: A 95% credible interval measures your personal degree of belief about where a parameter is. Thus you are more sure the parameter is contained within a 95% credible interval than a 90% credible interval and less sure than a 99% credible interval.

2. SpaceX has successfully completed 28 of its 30 Falcon 9 launches. SpaceX claims it is "right in the ballpark" to the industry standard of 95% of launches being successful. For the following assume $Y \sim Bin(n, \theta)$ where Y is the number of successful launches, n is the number of attempted launches, and θ is the unknown probability of success. Also assume $\theta \sim Unif(0, 1)$.

   (a) State the posterior distribution for θ. (5 points)
   Answer:

```
y <- 28
n <- 30
```

$$\theta|y \sim Be(29, 3)$$

   (b) State the Bayes estimate for θ. (5 points)
   Answer:

```
(1+y)/(2+n)

## [1] 0.90625
```

   (c) Find a 95% equal-tail credible interval for θ. (5 points)
   Answer:

```
qbeta(c(.025,.975), 1+y, 1+n-y)

## [1] 0.7857838 0.9795801
```

   (d) Find the probability that θ is less than the industry standard of 0.95. (5 points)
   Answer:

```
pbeta(.95, 1+y, 1+n-y)

## [1] 0.799247
```

3. A manufacturing process for maple syrup has a target viscosity for the maple syrup of 5,000 cP @ 25^oC. A random sample of 25 lots of maple syrup had an average viscosity of 4,956 cP with a sample standard deviation of 100 cP. Assume $Y_i \stackrel{ind}{\sim} N(\mu, \sigma^2)$ where Y_i is the ith viscosity measurement.

   (a) Calculate a pvalue for $H_0 : \mu = 5,000$ vs $H_A : \mu \neq 5,000$. (5 points)
   Answer:

```
n = 25; ybar = 4956; s = 100; mu0 = 5000
2*pt(-abs((ybar-mu0)/(s/sqrt(n))), df=n-1)

## [1] 0.03767682
```

   (b) Calculate a 95% confidence interval for μ. (5 points)
   Answer:

```
ybar + c(-1,1)*qt(.975,df=n-1)*s/sqrt(n)

## [1] 4914.722 4997.278
```

   (c) If you were to calculate a 99% confidence interval for μ, would the interval be wider or narrower than 95% confidence interval for μ? Why? (5 points)
   Answer: Wider, as can be seen below because if we must cover the truth more often, then the interval will necessarily be wider.

```
ybar + c(-1,1)*qt(.995,df=n-1)*s/sqrt(n)

## [1] 4900.061 5011.939
```

   (d) The predictive distribution for a new observation $\tilde{Y}$ is $\tilde{Y} \sim t_{n-1}(\overline{y}, s^2(1+1/n))$. Calculate the probability a new observation is over 5,000 cP. (5 points)
   Answer:

$$
\begin{aligned}
P(\tilde{Y} > 5000) &= P\left(\frac{\tilde{Y}-\overline{y}}{s\sqrt{1+1/n}} > \frac{5000-4956}{100\sqrt{1+1/25}}\right) \\
&= P(T_{n-1} > 0.4314555) \\
&= 1 - P(T_{n-1} < 0.4314555)
\end{aligned}
$$

```
1-pt((5000-ybar)/(s*sqrt(1+1/n)), df = n-1)

## [1] 0.3349941
```

4. A random sample of homes in Ames found that 35 out of 100 use "smart thermostats" while a random sample of homes in Des Moines found that 245 out of 923 use "smart thermostats".

   (a) State an appropriate model to use for these data. (10 points)

   Answer: Let Y_i be the number of homes in our sample from region i (where $i = 1$ indicates Ames and $i = 2$ indicates Des Moines) that have smart thermostats and n_i is the number samples taken from region i. Assume

$$Y_i \stackrel{ind}{\sim} Bin(n_i, \theta_i).$$

   (b) Calculate a pvalue for the null hypothesis that the proportion of homes in Ames that use smart thermostats is less than or equal to the proportion in Des Moines versus the alternative that the proportion is higher in Ames. (5 points)

   Answer:

```r
y <- c(35,245)
n <- c(100,923)
prop.test(y, n, alternative = "greater")$p.value

## [1] 0.04614426
```

   (c) Calculate a 90% equal-tail confidence interval for the difference in the true proportions. (5 points)

   Answer:

```r
prop.test(y, n, conf.level = 0.9)$conf.int

## [1] -0.002996652  0.172119079
## attr(,"conf.level")
## [1] 0.9
```

5. The file `skin.csv` contains data on a study where participants were randomly assigned one arm to receive skin lotion while the other arm served as a control. After using the lotion for one week, the dryness on each arm was measured.

   (a) State an appropriate model to use for these data. (10 points)

   Answer: Let D_i be the difference in dryness for the lotion arm minus the control arm. Assume

   $$D_i \stackrel{ind}{\sim} N(\mu, \sigma^2).$$

   You could have defined the difference in the opposite direction and then your answers below would have been the negative of the answers here.

   (b) Calculate the Bayes estimate for the difference in skin dryness. (5 points)

   Answer:

```r
skin = read.csv("skin.csv")
t.test(skin$Lotion, skin$NoLotion, paired=TRUE)$estimate

## mean of the differences
##                   -2.94
```

   (c) Calculate a 95% equal-tail credible interval for the difference in skin dryness. (5 points)

   Answer:

```r
t.test(skin$Lotion, skin$NoLotion, paired=TRUE)$conf.int

## [1] -4.562948 -1.317052
## attr(,"conf.level")
## [1] 0.95
```