Midterm Exercises

C. Durso

February 6, 2019

Instructions

Please work these problems on your own. You may use web searches, but not interactive methods such as asking others omline or in person.

Questions

1. Construct a reasonable model for rolling a fair (each face is equally likely) ten-sided die twice.

   What is the probability of single outcome (a, b) with? $a, b \in \{1, 2, 3, ...10\}$(5 points)

   There are 10×10 outcomes, eeach equally likely. Therefore the probability of any single outcome is $\frac{1}{100} = 0.01$

   What is the probability of the event $\{(a, b) \,|a \neq b\}$? (5 points)

   There are 10 outcomes with $a = b$, so there are 90 outcomes with $a \neq b$. Conclude that the probability of the event $\{(a, b) \,|a \neq b\}$ is $\frac{90}{100} = 0.9$.

2. The function defined by $f(x) = cx$ for $x \in [0, 1]$ and $f(x) = 0$ otherwise is the probability density function of a continuous random variable. What is the value of c? (10 points)

That f is a density requires that $\int_{-\infty}^{\infty} f(x)dx = \int_0^1 cxdx = 1$. Evaluation of the integral gives $\frac{1}{2}cx^2\Big|_0^1 = \frac{c}{2} = 1$. Conclude $c = 2$.

3. Construct a reasonable model for flipping a fair coin 4 times and recording the results in order. You don't have to explain the model, just provide the information requested below.

   Are the event that the number of heads is even and the event that the first flip is a head independent?(5 points)

   There are 16 equally likely outcomes. Of those $\left(\begin{array}{c} 4 \\ 0 \end{array}\right) = 1$ have 0 heads, $\left(\begin{array}{c} 4 \\ 2 \end{array}\right) = 6$ have 2 heads, $\left(\begin{array}{c} 4 \\ 4 \end{array}\right) = 1$ have 4 heads, so the probability of an even number of heads equals $\frac{1}{2}$. The probability that the first flip is a head is also $\frac{1}{2}$. The intersection of these events is that the first head is a flip and that there are 1 or 3 heads in the remaining 3. This has probability $\frac{1}{2}\left(\frac{1}{8}\right)\left[\left(\begin{array}{c} 3 \\ 1 \end{array}\right) + \left(\begin{array}{c} 3 \\ 3 \end{array}\right)\right]$, or $\frac{1}{4} = \frac{1}{2} \times \frac{1}{2}$ Conclude that the events are independent.

   Are the event that the number of heads is at least 3 and the event that the first flip is a head independent?(5 points)

   There are four sequences with 3 heads and one sequence with 4 heads, so the probability of at least three heads is $\frac{5}{16}$. The probability that the first flip is a head is $\frac{1}{2}$. The intersection consists of three sequences with 3 heads (omitting the one that starts with a tail) and one sequence with 4 heads. The probability of the intersection is $\frac{4}{16} \neq \frac{5}{16} \times \frac{1}{2}$. These events are not independent.

4. Consider a continuous random variable X with the probability density function defined by $f(x) = \frac{3}{4}\left(1 - x^2\right)$ for $x \in [-1, 1]$ and $f(x) = 0$ otherwise. What is the value of the cumulative distribution of X, $F(t)$, at $t = \frac{1}{2}$?

An antiderivative of $\frac{3}{4}\left(1-x^2\right)$ is $\frac{3}{4}\left(x-\frac{x^3}{3}\right)$. Evaluating this between -1 and $\frac{1}{2}$ gives $\frac{3}{4}\left(\frac{1}{2}-\frac{\frac{1}{2}^3}{3}\right)-\frac{3}{4}\left(-1-\frac{(-1)^3}{3}\right)=\frac{3}{2}$

5. If you model data in which there are 100 trials, of which 36 are successes, as a sample from the Binomial distribution with the size parameter equal to 100, what is the maximum likelihood estimate for the probability parameter?

We have previously derived that the maximum likelihood probability is $\frac{36}{100}=\frac{9}{25}$ or 0.36.

6. If you model the "total" column in the "animal_stats_compact.csv" as a sample from a Normal distribution, what are the maximum likelihood values of μ and σ^2? (Note that this may not be a very accurate model.)

```
dat<-read.csv("animal_stats_compact.csv")

(mu<-mean(dat$total))
```

```
## [1] 2074.696
```

```
(sigmasq<-sum((dat$total-mu)^2)/length(dat$total))
```

```
## [1] 3356423
```

7. In the data from #6, if you model the "dog_surg" variable as a linear function of the "dog" variable, what slope and intercept give the line with the least squares best fit?

```
model1<-lm(dog_surg~dog,data=dat)
(m<-model1$coefficients[2])
```

```
##       dog
## 0.5921053
```

```
(b<-model1$coefficients[1])
```

```
## (Intercept)
##    93.30204
```

```
# or

mx<-mean(dat$dog)
my<-mean(dat$dog_surg)
(m<-sum((dat$dog-mx)*(dat$dog_surg-my))/sum((dat$dog-mx)^2))
```

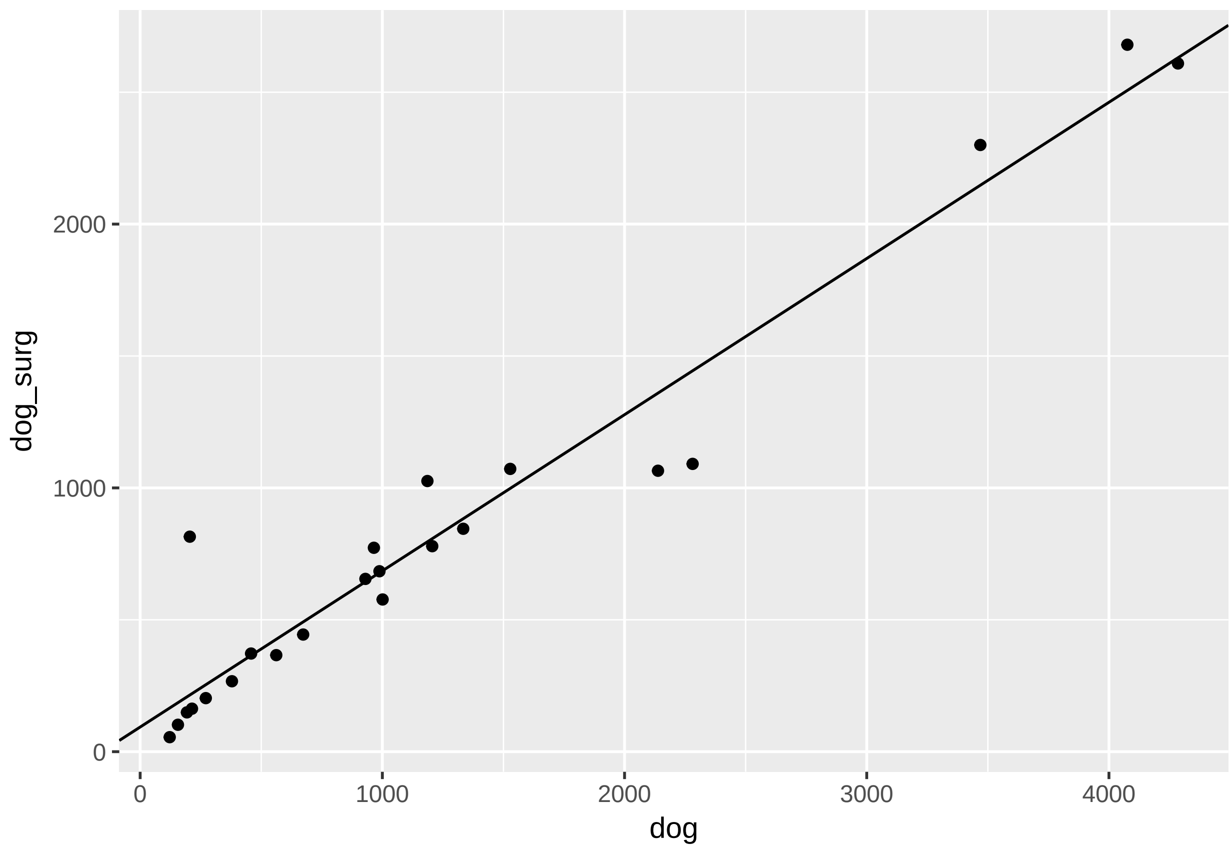
```
## [1] 0.5921053
```

```
(b<-my-m*mx)
```

```
## [1] 93.30204
```

8. Make a scatterplot with the "dog" variable on the horizontal axis and the "dog_surg" variable on the vertical axis. Add the line computed in #7.

```
g<-ggplot(dat, aes(x=dog,y=dog_surg))+geom_point()
g<-g+geom_abline(slope=m, intercept=b)
g
```



9. What is the total number of dogs in each city?

```
summarize(group_by(dat,city), sum(dog))
```

```
## # A tibble: 4 x 2
##   city        `sum(dog)`
##   <fct>            <int>
## 1 Atlanta           6166
## 2 Chicago           7013
## 3 Los Angeles      10941
## 4 Philly            4500
```

10. What is the probability that a Normal random variable with $\mu = 10$ and $\sigma^2 = 4$ lies between 11 and 12?

```
pnorm(12,10,2)-pnorm(11,10,2)
```

```
## [1] 0.1498823
```

11. At what value is the cumulative distribution of a Normal random variable with $\mu = 10$ and $\sigma^2 = 4$ equal to 0.80?

```
qnorm(.8,10,2)
```

```
## [1] 11.68324
```

12. Set the random seed equal to 12345. Sample the Binomial distribution with size equal to 12 and probability of success equal to 0.5 ten times. Display the results.

```
set.seed(12345)
rbinom(10,12,.5)
```

```
## [1]  7  8  7  8  6  4  5  6  7 10
```

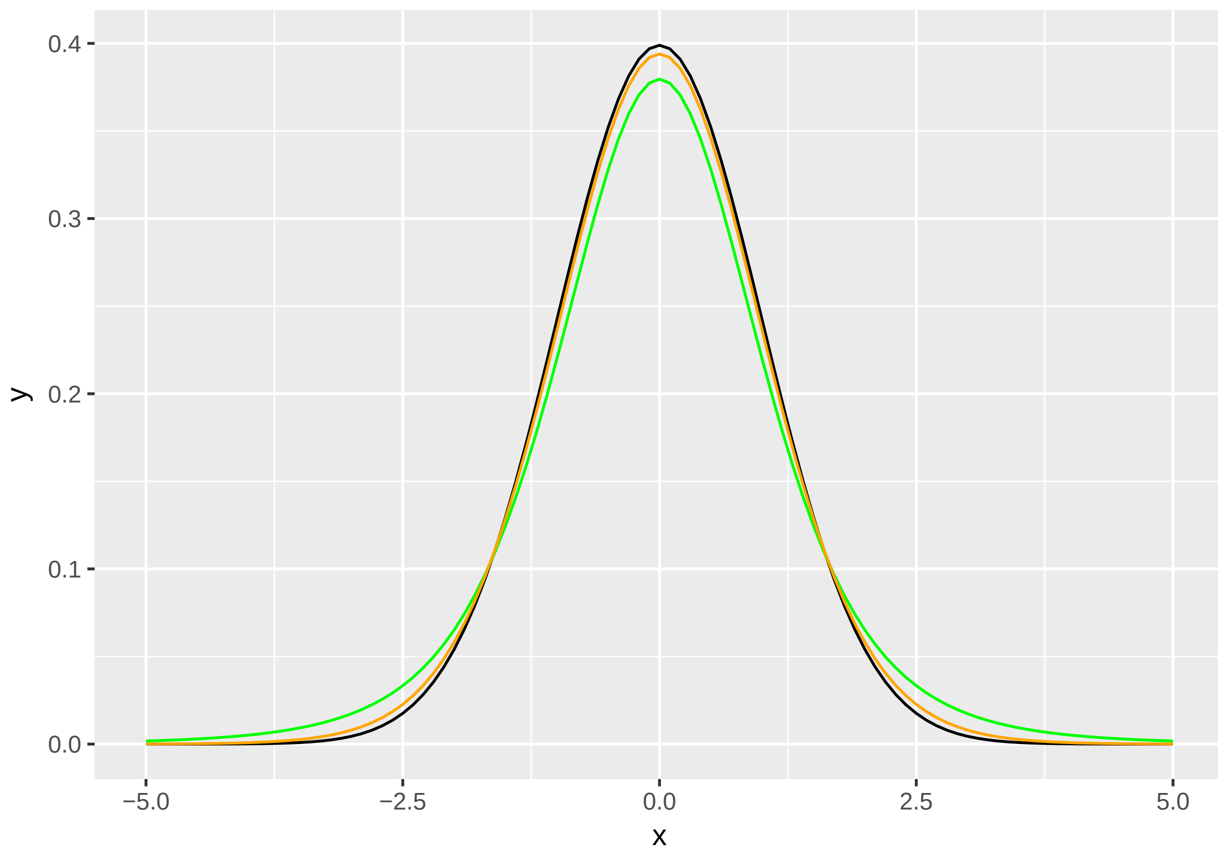
13. Again, set the random seed to 12345. Sample the integers between 1 and 50 ten times without replacement. What is the last value in the sample?

```r
set.seed(12345)
s<-sample(1:50,10)
s[10]
```

```
## [1] 41
```

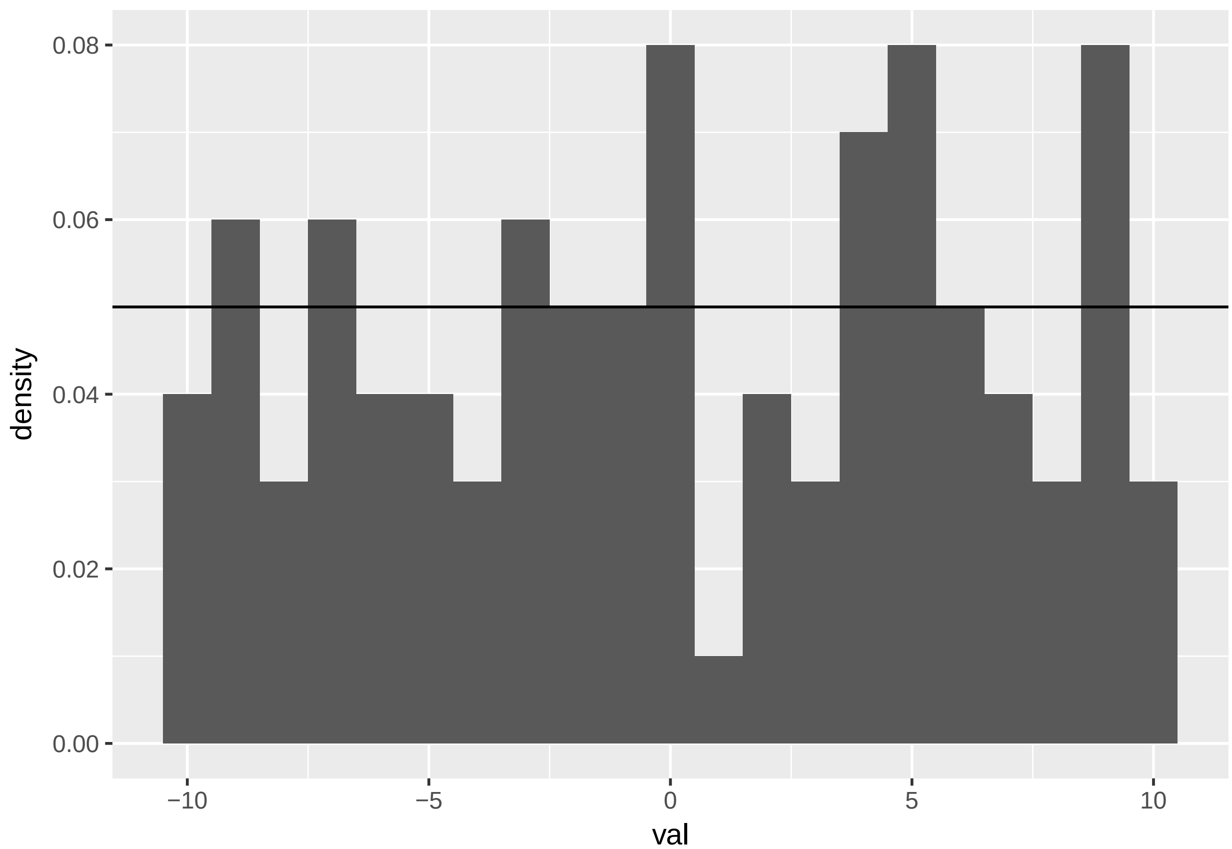
14. Plot the standard Normal density curve, the Student's t density curve (the function dt) for df=5 in green, and the Student's t density curve for df=20 in orange.

```r
mini<-data.frame(x=seq(-5,5,by=.1))
g<-ggplot(data=mini,aes(x=x))+stat_function(fun=dnorm)+stat_function(fun=dt, args=5, color="green")+sta
g
```



15. Set the random seed to 12345 again. Generate a sample of size 100 from the uniform distribution on [-10,10] and plot the histogram using the binwidth=1 and the "..density.." aes. On this, plot the horizontal line at height 1/20. We haven't used the geom for a horizontal line. You will need to identify it and figure out the syntax yourself.) (5 points)

```r
set.seed(12345)
samp<-data.frame(val=runif(100,-10,10))
ggplot(data=samp,aes(x=val))+geom_histogram(aes(y=..density..),binwidth =1)+
  geom_hline(yintercept=1/20)
```


density
0.08
0.06
0.04
0.02
0.00
−10
−5
0
5
10
val