

Statistics 341

Homework Assignment #6 Answers

Due Friday, November 7

115 points possible

1. (4 pts) Let the random variable Y have a Uniform distribution with $\theta_1 = 0$ and $\theta_2 = 1$. Show that

$$P(a \leq Y \leq a + b) \text{ for } a \geq 0, b \geq 0, a + b \leq 1$$

depends only on the value of b .

$$\begin{aligned} P(a \leq Y \leq a + b) &= P(Y \leq a + b) - P(Y \leq a) \\ &= F(a + b) - F(a) \\ &= \frac{a + b - \theta_1}{\theta_2 - \theta_1} - \frac{a - \theta_1}{\theta_2 - \theta_1} \\ &= a + b - a = b \end{aligned}$$

since $\theta_2 = 1$ and $\theta_1 = 0$.

2. (7 pts) The cycle time for trucks hauling concrete to a highway construction site is uniformly distributed over the interval from 50 to 70 minutes.
 - (a) (3 pts) What is the probability the cycle time will be greater than 65 minutes?

$$\begin{aligned} P(Y > 65) &= 1 - P(Y \leq 65) \\ &= 1 - F(65) \\ &= 1 - \frac{65 - \theta_1}{\theta_2 - \theta_1} \\ &= 1 - \frac{65 - 50}{70 - 50} \\ &= 1 - \frac{15}{20} = \frac{5}{20} = 0.25 \end{aligned}$$

Using R, you can calculate this probability with the following command

```
1 - punif(65, 50, 70)
```

- (b) (4 pts) Find the mean and variance of the cycle time.

$$E(Y) = \frac{\theta_1 + \theta_2}{2}$$

$$\begin{aligned}
&= \frac{50 + 70}{2} = 60 \\
V(Y) &= \frac{(\theta_2 - \theta_1)^2}{12} \\
&= \frac{(70 - 50)^2}{12} = 33.3333
\end{aligned}$$

3. (23 pts) As a measure of intelligence, mice are timed when going through a maze to reach a reward of food. The time (in seconds) required for any mouse is a random variable Y with a density function given by

$$f(y) = \begin{cases} \frac{60}{y^2} & y \geq 60 \\ 0 & \text{elsewhere} \end{cases}$$

where 60 seconds is the minimum possible time needed to traverse the maze.

- (a) (4 pts) Find the distribution function $F(y)$ for the random variable Y .

The p.d.f. of Y divides the real line into two regions: $y < 60$ and $y \geq 60$. For the first region,

$$F(y) = \int_{-\infty}^y f(t) dt = \int_{-\infty}^y 0 dt = 0$$

For the second region,

$$F(y) = \int_{-\infty}^y f(t) dt = \int_{-\infty}^{60} 0 dt + \int_{60}^y \frac{60}{t^2} dt = 0 + -60 \left(\frac{1}{t} \Big|_{60}^y \right) = -60 \left(\frac{1}{y} - \frac{1}{60} \right) = 1 - \frac{60}{y}$$

$$F(y) = \begin{cases} 0 & y < 60 \\ 1 - \frac{60}{y} & y \geq 60 \end{cases}$$

- (b) (6 pts) Use R to generate 10,000 observations of the random variable Y . Find the minimum and maximum values for these 10,000 observations. What do you notice about your values?

Let u be an observation from the uniform distribution with $\theta_1 = 0$ and $\theta_2 = 1$. Set u equal to the distribution function from part (a) and solve for y .

$$\begin{aligned}
u &= 1 - \frac{60}{y} \\
1 - u &= \frac{60}{y} \\
\frac{60}{1 - u} &= y
\end{aligned}$$

The following commands in R will generate 10,000 observations of the random variable Y and place them in the variable y .

```
u<- runif(10000,0,1)
y<- 60/(1-u)
```

To find the minimum and maximum values, the R commands are

```
min(y)
max(y)
```

Each person will get their own minimum and maximum values. Your minimum value is likely to be fairly close to 60 and your maximum value is likely to be very large.

- (c) (6 pts: 3 pts for the histogram and 3 pts for the explanation) Make a probability histogram for your 10,000 observations. Print this out and include it with your assignment. How would you describe this histogram?

While each person will get a slightly different histogram from their 10,000 observations, the general pattern will be the same. There will be one large bar starting at 0 and no other visible bars on the histogram. However, the values on the horizontal axis will extend out to the maximum observation value you found in part(b). This bar is so small that it doesn't show up on the histogram.

- (d) (7 pts: 2 pts each for the mean and variance and 3 pts for the explanation) Find the mean and variance of these 10,000 observations. How do these observed values compare to the theoretical values of the mean and variance from Homework Assignment #6? Explain your answer.

The R commands to get the mean and variance are

```
mean(y)
var(y)
```

In homework #5, the theoretical mean and variance were both undefined. The integrals diverged. Because in this case we have actual observations, we will get values for the observed mean and variance. Each person in class will get their own mean and variance. However, just like with the histogram, there will be a pattern. The mean will be somewhere in the 500-800 range, although someone could get a value outside of this range. Almost everyone should get a **very** large variance.

This shows you why the mean and variance will diverge when you have an infinite number of observations. Almost all of the observation for this random variable will be in a range fairly close to 60. However, you will occasionally get a very large observation for this random variable. These very large observations push both the mean and variance of the random variable out on the real line. In the limit, there are too many of these very large observations to get a finite mean or variance for this random variable.

4. (6 pts) Wires manufactured for use in a certain computer system are specified to have resistances between 0.12 and 0.14 ohms. The actual measured resistances of the wires produced by Company A have a normal distribution with a mean of 0.13 ohm and a standard deviation of 0.005 ohm.

- (a) (3 pts) What is the probability that a randomly selected wire from Company A's production will meet the specifications?

Let Y be the measured resistance of the randomly selected wire. We need to find $P(0.12 \leq Y \leq 0.14) = P(Y \leq 0.14) - P(Y \leq 0.12)$. Using R, the command is

```
pnorm(0.14,0.13,0.005) - pnorm(0.12,0.13,0.005)
```

The correct answer is 0.9545.

- (b) (3 pts) If four such wires are used in the system and all are from Company A, what is the probability that all four will meet the specifications?

Let the random variable X be the number of the 4 wires in the system from Company A that meet the specifications. Since the 4 wires are independent and have the same probability of meeting specifications (0.9545), the random variable X has a binomial distribution with parameters $n = 4$ and $p = 0.9545$. We need to find the probability $P(X = 4)$. This is $p^4 = 0.9545^4 = 0.8300$. Alternatively, using R, the command is

```
dbinom(4,4,0.9545)
```

which gives the answer 0.8300

5. (8 pts) A soft drink machine can be regulated so that it discharges an average of μ ounces per cup. If the ounces of fill are normally distributed with standard deviation equal to 0.3 ounce, give the setting for μ so that 8-ounce cups will overflow only 1% of the time.

Let Y be the amount of soft drink that is dispensed into an 8-ounce cup. We need to find the value μ so that $P(Y > 8) = 0.01$. This is equivalent to finding the value of μ so that $P(Y \leq 8) = 0.99$. Transform Y into Z .

$$P(Y \leq 8) = P\left(\frac{Y - \mu}{\sigma} \leq \frac{8 - \mu}{0.3}\right) = P\left(Z \leq \frac{8 - \mu}{0.3}\right) = 0.99$$

Use R to find the value of $z = \frac{8 - \mu}{0.3}$ that solves this equation. The command in R is

```
qnorm(0.99,0,1)
```

The value of $z = 2.326348$. Solve for the value of μ .

$$\begin{aligned} z &= \frac{8 - \mu}{0.3} \\ 2.326348 &= \frac{8 - \mu}{0.3} \\ 0.6979044 &= 8 - \mu \\ \mu &= 7.3020956 \end{aligned}$$

So the value of μ is 7.3021 ounces.

6. (16 pts) Many variables that measure physical characteristics, such as height, are considered to be normally distributed. For the past several semesters, students enrolled in Statistics 101 have filled out a questionnaire on the web. One of the questions asks for their height in inches. On the course web page, find and click on the link for the Stat 101 Female Height Data. Then copy and paste the data into R using the `scan` command in R. Use these data to answer the following questions.

- (a) (2 pts) Find the mean and standard deviation of the heights of the Statistics 101 Females. Here is the R code for finding the mean and standard deviation of the heights with the correct answers.

```
> mean(heights)
[1] 65.76545
> sqrt(var(heights))
[1] 2.952565
```

- (b) (2 pts) Find the maximum and minimum height of the Statistics 101 Females. Here is the R code for finding the minimum and maximum height with the correct answers.

```
> min(heights)
[1] 56
> max(heights)
[1] 84
```

- (c) (6 pts: 3 pts for the plot and 3 pts for the explanation) Use R to make a normal quantile plot of the height values. Print this plot and include it with your assignment. According to the plot, do you think the height of Statistics 101 Females has a normal distribution? Explain your answer.

The normal quantile plot shows the height values following the normal distribution throughout a large range of the data. However, there appears to be large heights in this data set that do not follow the probabilities of the normal distribution. These large heights are occurring more frequently than we would expect if the distribution were truly normal. It looks like a symmetric distribution with a bell-shaped curve, but with more probability in the tails of the distribution (fatter tails) would fit these data better than a normal distribution.

- (d) (4 pts) Regardless of your answer to the previous question, let's assume the heights of Stat 101 Females do have a normal distribution. What do you think are reasonable values for μ and σ for this distribution? Explain your answer.

It would be reasonable to use the observed values of the mean and standard deviation for the theoretical mean and standard deviation of the distribution. These values are $\mu = 65.76545$ and $\sigma = 2.952565$. We know these values are not exactly the correct ones for μ and σ , but they shouldn't be too far off from the true values.

- (e) (2 pts) The height data for this problem are for females only. Explain why I did not include the males in these data.

The distribution for the heights of males would be centered about 4 to 5 inches higher than the distribution for the heights of females. This results in a bimodal distribution which does not follow a normal distribution. When looking at physical characteristics, such as height, you want to look at a homogeneous group.

7. (15 pts) The weekly amount of down time Y (in hours) for an industrial machine has a gamma distribution with $\alpha = 3$ and $\beta = 2$.

- (a) (3 pts) Find the probability the machine will be down for more than 3 hours.

This is $P(Y > 3) = 1 - P(Y \leq 3) = 1 - F(3)$. Using R, the command is

```
1 - pgamma(3, shape = 3, scale = 2)
```

The answer is 0.8088.

- (b) (3 pts) For 95% of all weeks, the machine will be down less than how many hours?

We need to find the value of y so that $P(Y \leq y) = 0.95$. Using R, the command is

```
qgamma(0.95, shape = 3, scale = 2)
```

The answer is 12.5916 hours.

- (c) (9 pts) The loss (in dollars) to the industrial operation as a result of down time is the function $L = 30Y + 2Y^2$. Find the expected loss.

For the gamma distribution, $E(Y) = \alpha\beta = 3(2) = 6$ and $V(Y) = \alpha\beta^2 = 3(2^2) = 12$.

The expected loss is then

$$\begin{aligned} E(L) &= E(30Y + 2Y^2) \\ &= 30E(Y) + 2E(Y^2) \\ &= 30E(Y) + 2[V(Y) + (E(Y))^2] \\ &= 30(6) + 2(12 + 6^2) \\ &= 276 \end{aligned}$$

8. (10 pts) In class, we learned that the parameters for the gamma distribution are the shape parameter α and the scale parameter β . Use R to generate 10,000 values from each of four gamma distributions. The four gamma distributions will all have $\beta = 2$, but will have $\alpha = 5, 20, 50, 100$, respectively. For each set of observed gamma values, calculate the mean and variance and obtain a histogram. Then write a paragraph explaining the effect of α on the shape, mean and variance of the gamma distribution.

Attached to this homework assignment is the graph of the histograms of the observed values of these 4 distributions. The table below gives the mean and variance of the observed values.

α	5	20	50	100
Mean	9.99277	39.90140	100.0563	199.9412
Variance	20.34551	79.81464	198.5942	403.9936

As you can see from this histogram, the effect of the parameter α on the distribution is to change the shape. Smaller values of α been the distribution will be fairly skewed right, while for larger values of α , the distribution will be more symmetric. Since the mean and variance of a gamma distribution is $\alpha\beta$ and $\alpha\beta^2$, the mean and variance of the observed values are close to these values. For example, when α doubles from 50 to 100, the values of the mean and variance are also about double.

9. (18 pts) Explosive devices used in mining operations produce nearly circular craters when detonated. The radii of these craters are exponentially distributed with a mean of 10 feet.

- (a) (3 pts) Find the probability that one of these explosive devices will produce a radii between 6 and 12 feet.

We need to find $P(6 \leq Y \leq 12)$. You could do this in one of three ways. First, you could integrate the density function of the exponential distribution between the values 6 and 12.

$$P(6 \leq Y \leq 12) = \int_6^{12} \frac{1}{10} e^{-y/10} dy$$

Or, you could use the distribution function to do the calculations.

$$P(6 \leq Y \leq 12) = P(Y \leq 12) - P(Y \leq 6) = F(12) - F(6) = (1 - e^{-12/10}) - (1 - e^{-6/10})$$

Or you could use the commands in R

```
pexp(12,1/10) - pexp(6,1/10)
```

Either one of the three ways will get you the answer of 0.2476.

- (b) (10 pts) Find the mean and variance of the areas produced by these explosive devices. For an exponential distribution, $E(Y) = \beta = 10$ and $V(Y) = \beta^2 = 10^2 = 100$. The area of a circle is $A = \pi Y^2$ where Y is the radius of the circle.

$$\begin{aligned} E(A) &= E(\pi Y^2) \\ &= \pi E(Y^2) \\ &= \pi [2!10^2] \\ &= 200\pi \end{aligned}$$

$$\begin{aligned} V(A) &= V(\pi Y^2) \\ &= \pi^2 V(Y^2) \\ &= \pi^2 [E(Y^4) - (E(Y^2))^2] \\ &= \pi^2 [4!10^4 - (2!10^2)^2] \\ &= 200000\pi^2 \end{aligned}$$

- (c) (5 pts) Find the probability that an explosive device will produce an area larger than 64π ft².

For the area to be larger than 64π , the radius will need to be larger than 8. $P(Y > 8) = 1 - P(Y \leq 8)$. Using R, the command is

```
1 - pexp(8,1/10)
```

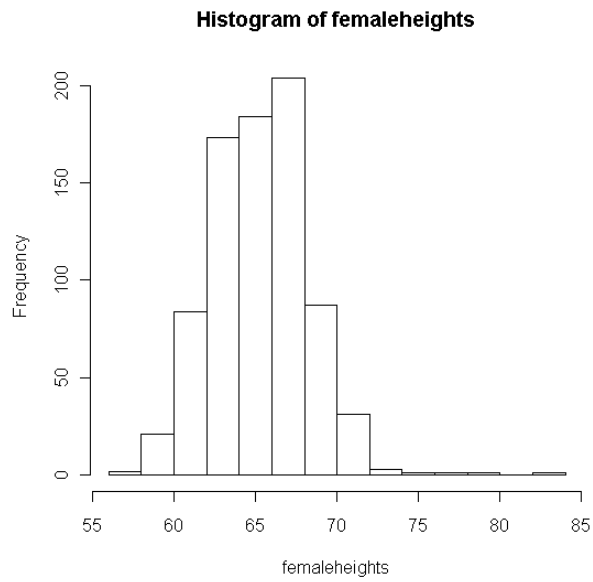
The answer is 0.4493

10. (8 pts) Let the random variable Y have an exponential distribution with mean β . Find $E(Y^r)$ for an positive integer r .

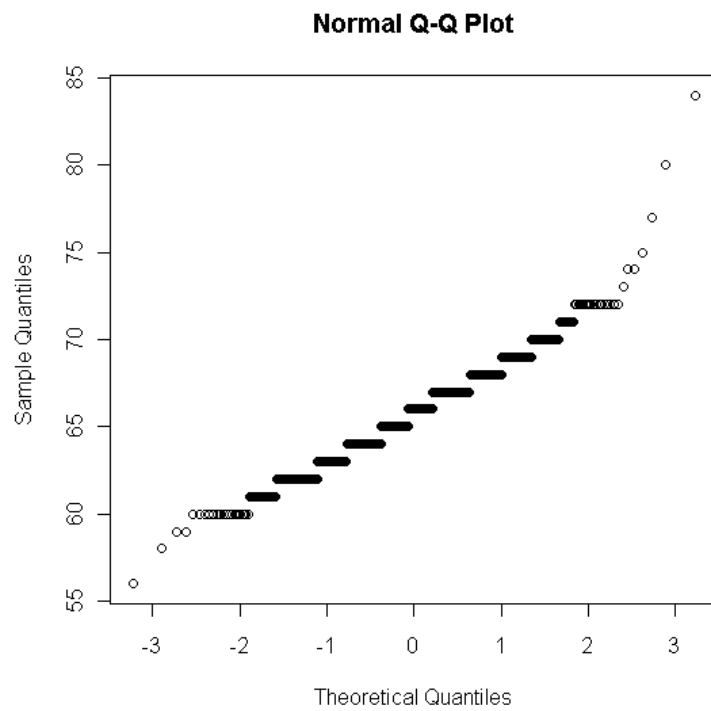
$$\begin{aligned}
E(Y^r) &= \int_0^\infty y^r \frac{1}{\beta} e^{-y/\beta} \\
&= \frac{1}{\beta} \int_0^\infty y^r e^{-y/\beta} \\
&= \frac{\beta^{r+1} \Gamma(r+1)}{\beta} \int_0^\infty \frac{y^r e^{-y/\beta}}{\beta^{r+1} \Gamma(r+1)} \\
&= \frac{\beta^{r+1} \Gamma(r+1)}{\beta} (1) \\
&= r! \beta^r
\end{aligned}$$

Two explanations for the above: The integrand $\frac{y^r e^{-y/\beta}}{\beta^{r+1} \Gamma(r+1)}$ is the p.d.f. of a Gamma distribution with $\alpha = r + 1$ and β . The integral of a p.d.f. over the range of the random variable is 1. The function $\Gamma(r + 1) = r!$ for positive integer values for $r + 1$.

Histogram of female heights.

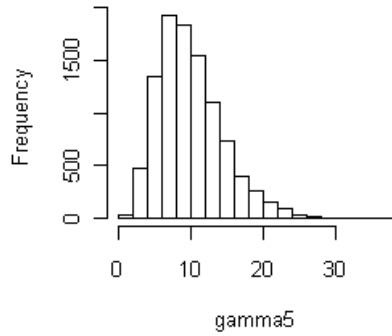


Normal quantile plot for female heights.

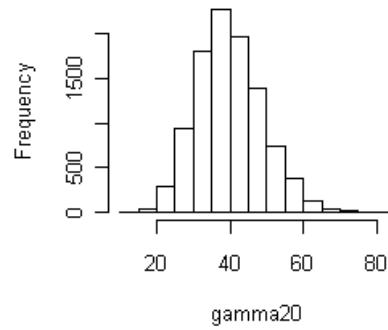


Histograms for observed Gamma values from problem #8.

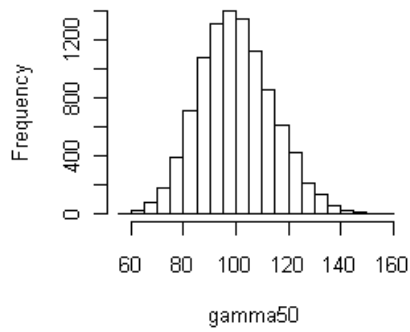
Histogram of gamma5



Histogram of gamma20



Histogram of gamma50



Histogram of gamma100

