Stat 342 Homework 1 Solution

April 19, 2019

Solution 1. **(a)** We denote the weights by $y_1, \ldots, y_n$ where $n = 40$. We let y_i i.i.d. $N(\mu, \sigma^2)$. Given that $\sigma = 0.22$ pounds. We perform the test $H_0 : \mu = 1.6$ vs $H_1 : \mu > 1.6$. The rejection region is given in the problem as : $\{\bar{y} \geq 1.67\}$. Note that $\bar{y} \sim N(\mu, \frac{\sigma^2}{n})$.

Therefore, significance level $= P_{H_0}(\bar{y} \geq 1.67)$. Under H_0, $\mu = 1.6$. So, significance level $=$

$$P(\bar{y} \geq 1.67) = 0.02209,$$

noting that under H_0, $\bar{y} \sim N(1.6, \frac{0.22^2}{40}) = N(1.6, 0.00121)$.

(b) When $\mu = 1.68$, $\bar{y} \sim N(1.68, 0.00121)$ so probability of Type II error

$$= P(\bar{y} < 1.67) = 0.38687$$

under that distribution.

(c) At level $\alpha = 0.05$, power function at $\mu = 1.68$ is $P(\bar{y} \geq 1.67) = 0.9$ (given) for $\bar{y} \sim N(1.68, \frac{0.22^2}{n})$. So, $P(\bar{y} \leq 1.67) = P(\bar{y} < 1.67) = 0.1$, since $\bar{y}$ has a continuous distribution. For

$$n = 788,$$

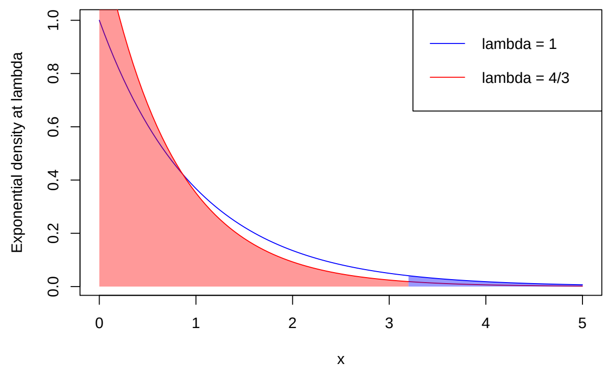
we note that, $P(\bar{y} < 1.67) \approx 0.10098$. Also, $P(\bar{y} \geq 1.67) \approx 1 \Rightarrow P(\bar{y} \leq 1.67) \approx 0 < 0.05$ for $\bar{y} \sim N(1.6, \frac{0.22^2}{788})$ and so it is a level 0.05 test.

Solution 2. **(a)** Here $Y \sim f_\lambda(y) = (1/\lambda)e^{-y/\lambda}, y > 0$. Rejection region is given to be $\{Y \geq 3.20\}$.

Probability of Type I error $= P(Y \geq 3.20)$, under $H_0 : \lambda = 1$ when $Y \sim e^{-y}, y > 0$, which equals 0.04076.

(b) Probability of Type II error $= P(Y < 3.20)$, under $\lambda = 4/3$, which equals 0.909.

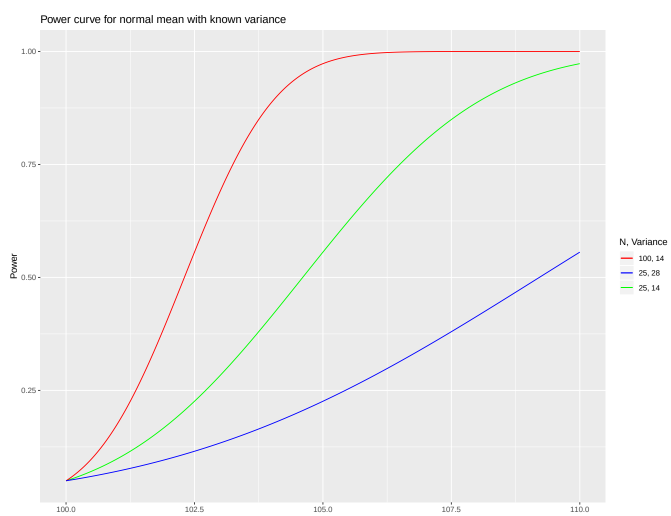
(c) The following figure plots the exponential curves for $\lambda = 1, 4/3$, and shades the probability of Type I and Type II errors with colors blue and red respectively.



Solution 3. $\beta(\lambda) = P(y \geq \ln 10)$ at

$$\lambda = 1 - e^{-\lambda \ln 10} = 1 - e^{-\lambda \ln 10} = 1 - 10^{-\lambda}$$

Solution 4. **(a)** (Zoom in to see legends; code provided below)



(b) With fixed σ, increase in n increased power. With fixed n, decrease in σ increased value of power function.

Solution 5. **(a)** Denoting f_Y and F_Y as the density and distribution functions of Y_i's respectively, density of Y_3' – the largest order statistic – is

$$f_{Y_3'}(y) = 3 f_Y(y) F_Y(y)^2, 0 \leq y \leq \theta$$

Now, for $0 \leq y \leq \theta$, $F_Y = \int_0^y \frac{2t}{\theta^2} dt = \frac{y^2}{\theta^2}$. So, we have

$$f_{Y_3'}(y) = \frac{6y^5}{\theta^6}, 0 \leq y \leq \theta$$

It follows that, $\frac{Y_3'}{\theta^{6/5}}$ is a pivot with density $\sim 6y^5, 0 \leq y \leq \theta$.

Also note that the CDF of Y_3' is

$$F_{Y_3'}(y) = \int_0^y \frac{6y^5}{\theta^6} dy = (\frac{y}{\theta})^6, 0 \leq y \leq \theta$$

It is also noted from the density of this pivot or Y_3' that for large θ, Y_3' is distributed towards larger values and vice-versa. So, for the one-sided alternative $H_1 : \theta > 5$, we reject $H_0 : \theta = 5$ if $\frac{Y_3'}{5^{6/5}} > c$ for some cut-off $c \in \mathbb{R}$.

Now, under H_0, $\frac{Y_3'}{5^{6/5}} \sim 6y^5$. So, for level $\alpha = 0.05$ we have, under $H_0 : \theta = 5$,

$$P(\frac{Y_3'}{5^{6/5}} > c) = 0.05$$

That is, denoting $F_{Y_3';\theta}$ as the CDF of Y_3' at θ, the above equation is same as writing $F_{Y_3';5}(5^{6/5} \cdot c) = 1 - P(Y_3' > 5^{6/5} \cdot c) = 0.95 \Rightarrow (\frac{(5^{6/5}c)}{5})^6 = 5^{6/5} \cdot c^6 = 0.95 \Rightarrow$

$$c = (5^{-6/5} \times 0.95)^{1/6} \approx 0.71861$$

.

Therefore we reject H_0 if $Y_3' > k$, where $k = 5^{6/5} \cdot c = 4.95744$.

(b) Type II error probability $= P(\frac{Y_3'}{5^{6/5}} \leq c)$ when $\theta = 7$. It is same as writing it as, $F_{Y_3';7}(5^{6/5} \cdot c)$ which equals $(\frac{5^{6/5} \cdot c}{7})^6 = 0.12617$.

Solution 6. **(a)** There is a placebo and drug pair corresponding to each subject. So comparison can be more accurate as other not-accounted-for physiological variables are fixed upon fixing the subject. In contrast, if the placebo and drug corresponded to different subjects many such variables might have taken different values for different subject. This difference in the physiological variable values could have affected the outcome but all of them would then be needed to be taken into account for greater accuracy.

(b) Knowledge of group assignment may lead to a psychological effect on the subject which itself may affect the outcome. Using a placebo and hiding which one is real drug and which is placebo can help alleviate this psychological effect on the health outcome.

(c) We denote A as 1 and P as 0. We denote by X the random variable modeling what the left eye receives among placebo and drug. That is $X = 1$, if left-eye receives a drug, and $X = 2$, if the left eye receives a placebo. That is X is a Bernoulli random variable, with parameter $p = P(X = 1)$, say, where $p \in [0, 1]$. We consider for the observations $X_1, \ldots, X_{10}$ for 10 subjects to be independent and identically distributed as X.

The question motivates to test the hypothesis $H_0 : p = 1/2$ vs $H_1 : p > 1/2$ based on the observations $X_1, \ldots, X_{10}$. We use the test statistic $T = \sum_{i=1}^{10} X_i$ and use a one-sided test. That is we reject H_0 if $T > c$ for some $c \in \mathbb{R}$. $T \sim$ Binomial$(10, 1/2)$ under H_0. At level α, $P(T > c) = \alpha$ under H_0. Since Binomial distribution is symmetric, we have that $c = (1 - \alpha)$th quantile of Binomial$(10, 1/2)$. We get that $c = 8$ for $\alpha = 0.0545$. Thus a size $\alpha = 0.0545$ test would reject H_0 if $T \geq 8$.

In the given data, T takes the value 7. So we state that there is not enough evidence to reject H_0 by this testing approach. In other words, we fail to reject the hypothesis that the assignment is random.

R code

```
#Q 2.(c)
curve(dexp, xlim=c(0,5), col="blue", ylab="Exponential density at lambda")
curve(dexp2, xlim=c(0,5), add=T, col="red")
polygon(cord.x2, cord.y2, col=rgb(red = 1, green=0, blue=0, alpha=0.4),
                                                        border=NA)
polygon(cord.x, cord.y1, col=rgb(red = 0, green=0, blue=1, alpha=0.4)
                                                        , border=NA)
legend("topright", legend=c("lambda = 1", "lambda = 4/3"),
                                    col=c("blue", "red"), lty=1)

#Q 3.
x=seq(from=0, to =1, by =0.001)
y=1- e^(-x*log(10))
plot(x, y, type="l", xlab="lambda", ylab="beta")

#Q 4.(a)
# Only code for the plot has been modified:
powercurve <- ggplot()+
  geom_line(data=powercomp, aes(x=mu, y=powerfn1, color="red"))+
  geom_line(data=powercomp, aes(x=mu, y=powerfn2, color="blue"))+
  geom_line(data=powercomp, aes(x=mu, y=powerfn3, color="green"))+
  ggtitle("Power curve for normal mean with known variance")+
  labs(title="Power curve for normal mean with known variance",
       x="Mean", y="Power", colour="N, Variance")+
  scale_color_manual(labels = c("100, 14", "25, 28", "25, 14"), values = c("red
powercurve

# Q 6.(a)
qbinom(p=1-0.05, size=10, prob=0.5)
```