Homework 7

Winter 2024

KEY

2024-03-05

Exercises

1.

a. $H_0 : \lambda_0 = .407$, $H_1 : \lambda_0 > .407$

b. We have that Zetterberg scored 44 goals in the 89 post season games. We see that more goals scored indicates more support for the alternative hypothesis. Thus our p-value, the probability of seeing our observed data or more extreme under the null, will be the $P(X \geq 44)$ when we have that $\lambda_0 = .407$. We can calculate this below:

```
null_lam <- .407
1-ppois(43,lambda = null_lam*89 )
```

```
## [1] 0.1153466
```

2.

a. We have that each Y is independent and identically distributed, meaning that each X_i is independent and has the same probability of equaling one for every X_i. We also have a fixed number of trials with a binary outcomes. Thus X will be a binomial distribution. Under the null we will have:

$$X \sim Binom(n = 19, \pi_0 = .4)$$

b. Since small values are in the direction of H_1, the P-value is the probability $P_{\pi_0=0.4}(X \leq x_{obs})$ where x_{obs} is the observed value. Since our decision rule is to reject H_0 when the P-value $\leq$ 0.05,we want to find the largest possible x_{obs} that will give us such a P-value.

```
binom_tib <- tibble(
  x = 0:19,
  pval = pbinom(x,19,.4),
  less_than_05 = pval <= .05
)
binom_tib
```

```
## # A tibble: 20 x 3
```

```
##         x      pval less_than_05
##     <int>     <dbl> <lgl>
##   1     0 0.0000609 TRUE
##   2     1 0.000833  TRUE
##   3     2 0.00546   TRUE
##   4     3 0.0230    TRUE
##   5     4 0.0696    FALSE
##   6     5 0.163     FALSE
##   7     6 0.308     FALSE
##   8     7 0.488     FALSE
##   9     8 0.667     FALSE
##  10     9 0.814     FALSE
##  11    10 0.912     FALSE
##  12    11 0.965     FALSE
##  13    12 0.988     FALSE
##  14    13 0.997     FALSE
##  15    14 0.999     FALSE
##  16    15 1.00      FALSE
##  17    16 1.00      FALSE
##  18    17 1.00      FALSE
##  19    18 1.00      FALSE
##  20    19 1         FALSE
```

Examining the table above, we would reject the null hypothesis for an $\alpha = 0.05$ level if we observed $X \leq 3$. Therefore $x_{obs} = 3$ is the largest value we would reject H_0 for.

Another way: Since we want to find x_{obs} that satisfies

$$\text{P-value} = P_{\pi=0.4}(X \leq x_{obs}) \leq 0.05$$

we could find the 5th percentile of the $Binom(n = 19, prob = 0.4)$ distribution.

```
qbinom(p = 0.05, size = 19, prob = 0.4)
```

```
## [1] 4
```

We can then check the P-value if we observed $x_{obs} = 4$. If it is exactly 0.05, then we reject for any $x_{obs} \leq 4$. If not, then we reject for $x_{obs} \leq 3$.

```
pbinom(q = 4, size = 19, prob = 0.4)
```

```
## [1] 0.06961371
```

Therefore, we reject for any $x_{obs} \leq 3$.

c. A Type I error occurs when we mistakenly reject H_0. It is given by the probability $P_{\pi_0=0.4}(X \leq 3)$.

```
pbinom(3, size = 19, prob = 0.4)
```

```
## [1] 0.02295932
```

d. A Type II error occurs when we should reject H_0 but we fail to do so. We know we will fail to reject the null when $X > 3$. Thus the type 2 error will be $P(X > 3)$ when we have that $\pi_0 = .2$. We have this as:

```
pbinom(3,19,.2,lower.tail=F)
```

```
## [1] 0.5449113
```

3.

a. We know from Problem 3, the null states $P(Y < 0) = .4$. Thus we must find a μ_0 so that the $P(Y < 0) = .4$ when $Y \sim N(\mu_0, 1)$. We may now find this μ_0:

$$P(Y < 0) = P(\frac{Y - \mu_0}{1} < \frac{0 - \mu_0}{1}) = .4$$

So we have that $-\mu_0$ is the .4 quantile of a N(0,1) distribution. Thus using qnorm we have that:

$\mu_0 = 0.2533471$.

We may test this value as well to check that the probability of $Y < 0$ is equal to .4:

```
mu_0_null <- -qnorm(.4)
pnorm(0,mean = mu_0_null, 1)
```

```
## [1] 0.4
```

For the alternative, we are testing $P(Y < 0) < .4$. This implies the Y's take larger values, and thus have a greater mean.

Thus we are testing the hypotheses:

$H_0 : \mu_0 = -qnorm(.4) = 0.2533$ and $H_1 : \mu_0 > -qnorm(.4) = 0.2533$

b. We have that each Y_i is independent and normally distributed. Thus using the properties of normal distributions which states that linear combinations of normal random variables is also normally distributed, we have:

$$\frac{1}{n}\sum_{i=1}^{n} Y_i = \bar{Y} \sim N(\mu_0, \sigma = \frac{1}{\sqrt{19}})$$

c. We reject the null for large values $\bar{Y}$. Hence the P-value will look like the probability $P_{\mu_0=0.2533}(Y \geq y_{obs})$. Since our decision rule is to reject H_0 when the P-value ≤ 0.05,we want to find the smallest possible y_{obs} that will give us such a P-value. This value is the 95th percentile of the distribution of Y. So we will reject when $\bar{y}_{obs}$ is greater than:

```
n <- 19
crit_val <- qnorm(.95, mean = -qnorm(.4),sd=1/sqrt(n))
crit_val
```

```
## [1] 0.6307024
```

d. For continuous distributions the type I error rate is equal to α. So in this case the Type 1 error rate is .05 .

e. We have that if $\pi_0 = .2$ then using the same process as above, we will have that under the alternative, $\mu_0 = -qnorm(.2) = 0.8416212$.

Thus the Type II error rate will be the probability we reject the null (our observed data is less than the critical value found above) when we have the alternative distribution with the mean we just calculated. We have this as:

```
alt_mean <- -qnorm(.2)
pnorm(crit_val, mean = alt_mean, sd = 1/sqrt(n))
```

```
## [1] 0.17895
```