Homework 2 for Stat Inference

Brian Caffo

About these slides

- These are some practice problems for Statistical Inference Quiz 1
- They were created using slidify interactive which you will learn in Creating Data Products
- Please help improve this with pull requests here (https://github.com/bcaffo/courses) runif(1)

The probability that a manuscript gets accepted to a journal is 12% (say). However, given that a revision is asked for, the probability that it gets accepted is 90%. Is it possible that the probability that a manuscript has a revision asked for is 20%?

1. Yeah, that's totally possible.
2. *No, it's not possible.*
3. It's not possible to answer this question.

*** .hint $A = accepted$, $B = revision$. $P(A) = .12$, $P(A|B) = .90$. $P(B) = .20$

*** .explanation $P(A \cap B) = P(A|B) * P(B) = .9 \times .2 = .18$ this is larger than $P(A) = .12$, which is not possible since $A \cap B \subset A$.

Suppose that the number of web hits to a particular site are approximately normally distributed with a mean of 100 hits per day and a standard deviation of 10 hits per day. What's the probability that a given day has fewer than 93 hits per day expressed as a percentage to the nearest percentage point?

1. 76%
2. *24%*
3. 47%
4. 94%

*** .hint Let X be the number of hits per day. We want $P(X \leq 93)$ given that X is $N(100, 10^2)$.

*** .explanation

```
round(pnorm(93, mean = 100, sd = 10) * 100)
```

```
## [1] 24
```

Suppose 5% of housing projects have issues with asbestos. The sensitivity of a test for asbestos is 93% and the specificity is 88%. What is the probability that a housing project has no asbestos given a negative test expressed as a percentage to the nearest percentage point?

1. 0%
2. 5%
3. 10%
4. 20%
5. 50%
6. *100%*

*** .hint $A = asbestos$, $T_+ = testspositive$, $T_- = testsnegative$. $P(T_+|A) = .93$, $P(T_-|A^c) = .88$, $P(A) = .05$.

*** .explanation We want

$$P(A^c|T_-) = \frac{P(T_-|A^c)P(A^c)}{P(T_-|A^c)P(A^c) + P(T_-|A)P(A)}$$

```
(.88 * .95) / (.88 * .95 + .07 * .05)
```

```
## [1] 0.9958309
```

Suppose that the number of web hits to a particular site are approximately normally distributed with a mean of 100 hits per day and a standard deviation of 10 hits per day.

1. What number of web hits per day represents the number so that only 5% of days have more hits? Express your answer to 3 decimal places.

*** .hint Let X be the number of hits per day. We want $P(X \leq 93)$ given that X is $N(100, 10^2)$.

*** .explanation 116.449

```
round(qnorm(.95, mean = 100, sd = 10), 3)
```

```
## [1] 116.449
```

```
round(qnorm(.05, mean = 100, sd = 10, lower.tail = FALSE), 3)
```

```
## [1] 116.449
```

Suppose that the number of web hits to a particular site are approximately normally distributed with a mean of 100 hits per day and a standard deviation of 10 hits per day. Imagine taking a random sample of 50 days.

1. What number of web hits would be the point so that only 5% of averages of 50 days of web traffic have more hits? Express your answer to 3 decimal places.

*** .hint Let $\bar X$ be the average number of hits per day for 50 randomly sampled days. X is $N(100, 10^2/50)$.

*** .explanation 102.326

```
round(qnorm(.95, mean = 100, sd = 10 / sqrt(50) ), 3)
```

```
## [1] 102.326
```

```
round(qnorm(.05, mean = 100, sd = 10 / sqrt(50), lower.tail = FALSE), 3)
```

```
## [1] 102.326
```

You don't believe that your friend can discern good wine from cheap. Assuming that you're right, in a blind test where you randomize 6 paired varieties (Merlot, Chianti, ...) of cheap and expensive wines

1. what is the change that she gets 5 or 6 right expressed as a percentage to one decimal place?

*** .hint Let $p = .5$ and X be binomial

*** .explanation

89.1

```
round(pbinom(4, prob = .5, size = 6, lower.tail = TRUE) * 100, 1)
```

```
## [1] 89.1
```

Consider a uniform distribution. If we were to sample 100 draws from a a uniform distribution (which has mean 0.5, and variance 1/12) and take their mean, $\bar{X}$

1. what is the approximate probability of getting as large as 0.51 or larger expressed to 3 decimal places?

*** .hint Use the central limit theorem that says $\bar{X} \sim N(\mu, \sigma^2/n)$

*** .explanation

0.365

```
round(pnorm(.51, mean = 0.5, sd = sqrt(1 / 12 / 100), lower.tail = FALSE), 3)
```

```
## [1] 0.365
```

If you roll ten standard dice, take their average, then repeat this process over and over and construct a histogram,

1. what would it be centered at?

*** .hint $E[X_i] = E[\bar{X}]$ where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

*** .explanation

The answer will be 3.5 since the mean of the sampling distribution of iid draws will be the population mean that the individual draws were taken from.

If you roll ten standard dice, take their average, then repeat this process over and over and construct a histogram,

1. what would be its variance expressed to 3 decimal places?

*** .hint

$$Var(\bar{X}) = \sigma^2/n$$

*** .explanation The answer will be 0 since the variance of the sampling distribution of the mean is $\sigma^2/12$ and the variance of a die roll is

```
mean((1 : 6 - 3.5)^2)
```

```
## [1] 2.916667
```

The number of web hits to a site is Poisson with mean 16.5 per day.

1. What is the probability of getting 20 or fewer in 2 days expressed as a percentage to one decimal place?

*** .hint Let X be the number of hits in 2 days then $X \sim Poisson(2\lambda)$

*** .explanation 1

```
round(ppois(20, lambda = 16.5 * 2) * 100, 1)
```

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