***Quiz***

* A web site for home pregnancy tests cites the following: "When the subjects using the test were women who collected + tested their own samples, overall sensitivity was 75%. Specificity was also low, in the range 52% to 75%." Suppose a subject has a negative test. Assume the lower bound for the specificity. What number is closest to the *multiplier* of the pre-test odds of pregnancy to obtain the post-test odds of pregnancy given a negative test result?
* **Sensitivity** = probability test is positive given the subject is pregnant (TP) = **P(+|D) = 75%**
* **Specificity** = probability test is negative given the subject is not pregnant (TN) = **P(-|Dc) = 52%**
* Bayes’ Rule:  and  so  🡺 **DLR+**
* **i.e.** post-test odds ofpregnant **=** DLR+ \* pre-test odds of having pregnant
* **DLR- == = P(-|D) / P(-|Dc) = (1-sensitivity) / specificity = (1-.75)/.52 = .4807 = ~.5**
* Assume the lower value for specificity. Suppose a subject has a negative test + that 30% of women taking pregnancy tests are actually pregnant. What number is closest to the probability of pregnancy given a negative test?
* Given P(Pr) = .3 (prevalence)
* 
* **P(Pr|-) = P(-|Pr)P(Pr) / [P(-|Pr)P(Pr) + P(-|PrC)P(PrC)]**
* **=** (1-.75)\*.3 / ((1-.75)\*.3 + .52\*(1-.3))
* **> p\_pr <- .3**
* **> p\_prc <- 1 - p\_pr**
* **> p\_pos\_given\_pr <- .75**
* **> p\_neg\_given\_pr <- 1 - (p\_pos\_given\_pr)**
* **> p\_neg\_given\_prc <- .52**
* **> p\_pos\_given\_prc <- 1 - (p\_neg\_given\_prc)**
* **> (p\_pr\_given\_neg <- (p\_neg\_given\_pr\*p\_pr) / ((p\_neg\_given\_pr\*p\_pr) + (p\_neg\_given\_prc\*p\_prc)))**
* **[1] 0.1708428 🡺 20%**
* Suppose hospital infection counts are models as Poisson with mean μ. Recall the Poisson mass function with mean μ is  for x=0, 1, …. 3 independent hospitals are observed for 1 year + their infection counts were 5, 4 and 6, respectively. What is the MLE for μ?
* **MLE =**  **= argument maximum over** ϴ of the likelihood, having plugged in data
*  🡺 Therefore, estimator  just = sample mean of the n sample observations
* makes intuitive sense b/c expected value of a Poisson random variable == its parameter (μ)
* **likelihood =**  **🡺 Taking logs, differentiating with respect to μ and solving for 0 yields the average** == 5
* Let X1,…,Xn be IID exponential(β). That is, having density = for x > 0. What is the MLE = for β?
* **likelihood is proportional to** 
* easier to *maximize* the **log-likelihood 🡺** 
* derive =  🡺 solve for 0 🡺 
* Let X be a geometric random variable that counts the # of coin flips until one obtains the 1st H . The mass function is  for x=1,2,…. What is the MLE for p if one observes a geometric random variable?
* **likelihood** is exactly the same as a sequence of *x* Bernoulli trials w/ 1 head and *x*−1 tails. So the ML estimate is the proportion of heads, **1/*x***
* Let X be a Poisson count with mean μ. Recall the Poisson mass function w/ mean μ =for x=0,1,…. What is the MLE for μ?
* **Poisson random variables** will have a **joint frequency function** that = a product of the marginal frequency functions
* log likelihood 🡺 
* find the maximum by finding the derivative 🡺 
* implies  as long as *l* is actually concave (making this a max)
* So, *OUR* log likelihood 🡺 , derived == , solving for 0 🡺 
* check 2nd derivative + case where x=0 to make sure concave
* Suppose diastolic BPs for men aged 35-44 are normally distributed w/ mean = 80 (mm Hg) + SD = 10. What is the probability that a random 35-44 year old has a DBP greater than 100 (mm Hg)?
* 100mmHG 🡺 (100 - 80) / 10 == 2SD above mean 🡺 95% of data are within 2 SD 🡺 5%/2 = **2.5%**
* **> mean <- 80**

**> sd <- 10**

**> DBP <- 100**

**> 1-pnorm(DBP,mean,sd) pnorm(bp,mean,sd,lower.tail = FALSE)**

**[1] 0.02275013 [1] 0.02275013**

* Brain volume for adult women is ~1,100 cc for women w/ a SD = 75 cc. About what brain volume represents the 10th percentile?
*  **🡺 normalize 🡺**  **🡺**  **= .1**
* -1.28 = the 10th percentile from the standard normal.
* Thus **1100 - 1.28 \* 75 == ~1000cc**.
* **qnorm(.1,mean,variance) = 1003.884**
* Now consider a sample mean of 144 random adult women from this population. Around what is the 10th percentile of the distribution of the distribution of sample means of 144 women?
* This refers to distribution of the sample mean which has a standard error = σ/**sqrt(n)**
* SD of the sample mean of 144 women is = SEmean = 75 / sqrrt(144) = 75/12 = 6.25 cc.
* Therefore the quantity in question 🡺 1100 - 1.28 \* 6.25 = 1092 cc
* **> mean <- 1100**

**> sd <- 75**

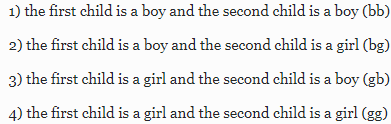
**> n <- 144**

**> SEmean <- sd/sqrt(n)**

**> q <- .1**

**> qnorm(q,SEmean,mean=mean)**

**[1] 1091.99**

* Consider 2 sample means, X¯1 and X¯2 from a the same normal population having mean μ + SD σ. X¯1 is based on n1=10k observations while X¯2 is based on n2=10(k+2) observations for some positive integer k. Take any percentile, say 100α, from the distribution of the means for each sample size excluding the median. What is the ratio of the distance of the 2 percentiles from the μ dividing distance 1 by distance 2?
* percentiles from the distributions 🡺  and 
*  == the relevant standard normal quantile
* ratio of the percentile differences from the mean == 
* If *n*1 is much bigger than *n*2, percentile for the mean based on *n*1’s observations will be much closer to *μ* than that of the mean based on *n*2 observations.
* 
* Thus, every time we increase the sample size by a factor of 100, you shrink the percentiles of the distributions of the sample mean toward the population mean by a factor of 10
* You flip a fair coin 6 times, which count of heads is most likely (0, 1, 2, 3, 4, 5, 6)?
* X ∼ Binomial(n,p) 🡺  🡺 6\*.5 = 3
* 
* 
* The respiratory disturbance index (RDI), a measure of sleep disturbance, for a specific population has a mean 15 (sleep events per hour) and SD 10. They are NOT normally distributed. Give your best estimate of the probability that a sample mean RDI of 100 people is above 17.
* SEmean = 10/**√100 = 1**
* **This 17 is 2 SEmeans above the mean 🡺 95% of data withing 🡺 above = 5%/2 = 2.5%**
* Consider a standard uniform density. The mean for this density = .5 and variance is 1/12. You sample 1,000 observations from this distribution and take the sample SD, what value would you expect it to be near?
* **Sample SD is consistent 🡺 SD = √variance = √1/√12 = 1/√12**
* You sample 1,000 sample means, each comprised of 100 observations. You take the mean of the 1,000 sample means. About what number would you expect it to be?
* **sample mean is unbiased 🡺 .5**
* You meet a person at the bus stop + strike up a convo. In the convo, the person gives the strange answer that “at least 1 of his 2 children is a girl”. Thinking on this, you decide to do a probability calculation. **Assuming** only that **genders are IID** with **50% probability each**, what is the chance of a 2 child family having 2 girls given the info that at least 1 is a girl?
*  🡺 **1/3**
* A web site for home pregnancy tests cites the following: "When the subjects using the test were women who collected + tested their own samples, overall sensitivity = 75%. Specificity was also low, in the range 52%-75%." Suppose a subject has a **positive** test. Assume the lower bound for the specificity. What number is closest to the multiplier of the pre-test odds of pregnancy to obtain the post-test odds of pregnancy given a positive test result?
* **Sensitivity** = probability test is positive given the subject is pregnant (TP) = **P(+|Preg) = 75%**
* **Specificity** = probability test is negative given the subject is not pregnant (TN) = **P(-|Preg) = 52%**
* Bayes’ Rule:  and  so  🡺 **DLR+**
* **i.e.** post-test odds ofpregnant **=** DLR+ \* pre-test odds of having pregnant
* **DLR+ == = P(+|Preg) / P(+|Preg c) = sensitivity / (1-specificity)= .75/(1-.52) = 1.5625**
* Now suppose a subject has a **positive** test + that 30% of women taking pregnancy tests are actually pregnant. What number is closest to the probability of pregnancy given the positive test?
* Given P(Pr) = .3 (prevalence)
* **> p\_pr <- .3**
* **> p\_prc <- 1 - p\_pr**
* **> p\_pos\_given\_pr <- .75**
* **> p\_neg\_given\_pr <- 1 - (p\_pos\_given\_pr)**
* **> p\_neg\_given\_prc <- .52**
* **> p\_pos\_given\_prc <- 1 - (p\_neg\_given\_prc)**
* **>**
* **> (p\_pr\_given\_pos <- (p\_pos\_given\_pr\*p\_pr) / ((p\_pos\_given\_pr\*p\_pr) + (p\_pos\_given\_prc\*p\_prc)))**
* **[1] 0.4010695**
* Let X1, …, Xk be independent Poisson counts with means for some known value . Recall the Poison mass function with mean μ is for x = 0,1,... What is the MLE for **λ**?
* log likelihood 🡺 
* find the maximum by finding the derivative 🡺 
* implies  as long as *l* is actually concave (making this a max)
* So, *OUR* log likelihood 🡺 , derived == , solving for 0 🡺 
* 
* Suppose a person is flipping a biased coin w/ success probability = p. She flips the coin 10 times yielding 1 head. Consider 2 possibilities: 1) person planned on flipping the coin ten times + got one head, 2) person planned to flip the coin until 1st head + it took ten times. What can you say about the likelihood in these two circumstances?
* For getting 1st H, **likelihood** = a sequence of *x* Bernoulli trials w/ 1 head and *x*−1 tails. So the ML estimate is the proportion of heads, 1/x 🡺 1/10
* For 10 flips, ending in 1 H, likelihood = ***L(p,0,0,0,0,0,0,0,0,0,1}*** *🡺* 
* 
* This likelihood *only* depends on the *total* number of H and the *total* number of T, not on order such as written in shorthand as 
* **The likelihood associated with p is identical (up to constants of proportionality) in either case.**
* Let X be a uniform random variable with support of length 1, but where we don't know where it starts. So that the **density is f(x) = 1** for x in (*θ*, *θ* + 1) and 0 otherwise. We observe a random variable from this distribution, say x1. What does the likelihood look like?
* **A horizontal line between x1 and x1 - 1.**
* Suppose diastolic BP’s for men aged 35-44 are normally distributed w/ mean 80 (mm Hg) + SD 10. What is the probability that a random 35-44 year old has a DBP less than 70?
* 100mmHG 🡺 (70 - 80) / 10 == 1SD below mean 🡺 68% of data are within 1 SD 🡺 32%/2 = **16%**
* **mu <- 80 sigma <- 10 val <- 70**

**pnorm(val,mu,sigma,lower.tail = T) 🡺 0.1586553 🡺 ~16%**

* For these same men , what's the probability that in a random sample of 5 subjects, 4 or more have DBPs more than 90?
* 100mmHG 🡺 (90 - 80) / 10 == 1SD above mean 🡺 68% of data w/in 1 SD 🡺 32%/2 = **16% = p**
* **p <- .16**
* **n <- 5**
* **choose(n,4)\*((p^4)\*(1-p)^(n-4)) 🡺 0.002752512 == 0.29 %**
* Brain volume for adult women is normally distributed with a mean of about 1,100 cc for women with a standard deviation of 75 cc. About what brain volume represents the 95th percentile?
*  **🡺 normalize 🡺**  **🡺**  **= .95**
* 1.64 = the 95th percentile from the standard normal
* Thus **1100 + 1.64 \* 75 == ~1220 cc**.
* **qnorm(.1,mean,sd) = 1223.364 🡺 1220**
* Consider the sample mean of 100 random adult women from this population. Around what is the 95th percentile of the distribution of that sample mean?
* This refers to distribution of the **sample mean** which has a **standard error** = σ/**sqrt(n)**
* **SD of the sample mean** of 100 women is = SEmean = 75 / sqrrt(100) = 75/10 = 7.5 cc.
* Therefore the quantity in question 🡺 1100 + 1.64\*7.5 = **1112.3 cc**
* **> mu <- 1100**
* **> sd <- 75**
* **> n <- 100**
* **> SE <- sd/sqrt(n)**
* **> qnorm(.95,SE,mean=mu)**
* **[1] 1112.336**
* Consider two sample meansfrom a the same normal population having mean = 0 and SD = δ. The 1st sample mean is based on n1 observations + the 2nd is based on n2 observations. Take any percentile, say **100α**, for the distribution of sample means based on the 2 sample sizes excluding the median. What is the ratio of the 2 percentiles (percentile of the 1st group divided by the 2nd)?
* percentiles from the distributions 🡺  and 
*  == the relevant standard normal quantile
* ratio of the percentile differences from the mean == 
* If *n*1 is much bigger than *n*2, percentile for the mean based on *n*1’s observations will be much closer to *μ* than that of the mean based on *n*2 observations.
* You flip a fair coin 5 times, what's the probability of getting 4 or 5 heads?
* X ∼ Binomial(n,p) 🡺  🡺 5\*.5 = 2.5
* 
* **> p <- .5**
* **> n <- 5**
* **> choose(n,5)\*((p^5)\*(1-p)^(n-5)) + choose(n,4)\*((p^4)\*(1-p)^(n-4))**
* **[1] 0.1875 == ~19%**
* The respiratory disturbance index (RDI), a measure of sleep disturbance, for a specific population has a mean of 15 (sleep events per hour) and a SD of 10. They are NOT normally distributed. Give your best estimate of probability a sample mean RDI of 100 people is between 14-16 events/hr.?
* SEmean = 10/**√100 = 1**
* **14 and 16 are is 1 SEmeans below + among the mean 🡺 68% of data within**
* Consider a standard uniform density. The mean for this density is .5 and the variance is 1 / 12. You sample 1,000 observations from this distribution and take the sample mean, what value would you expect it to be near?
* **sample mean is unbiased 🡺 .5**
* Consider a standard uniform density. The mean for this density is .5 and the variance is 1 / 12. You sample 1,000 sample means where each sample mean is comprised of 100 observations. You take the SD of the 1,000 sample means. About what number would you expect it to be?
* 
* Now sample 1,000 sample variances where each sample variance is comprised of 100 observations. You take the average of the 1,000 sample variances. What number do you expect that to be near?
* **Sample variance is consistent 🡺 1/12**