LOT2020 Bayesian course (Vasishth/Nicenboim)

Statistics Quiz 2020-01-14 **ID 00001**

1. (a)						0	.	5	4	1
2. (a)						0	.	4	1	7
(b)						1	.	2	2	8
3. (a)						3	.	6	9	5
(b)				1	4	7	.	0	0	0
(c)						1	.	9	7	6
(d)				1	1	2	.	5	2	3
(e)				1	2	7	.	1	2	5
4. (a)						0	.	0	9	1
(b)						0	.	9	0	9
(c)						0	.	0	0	0
5. (a)						0	.	3	0	9
(b)						0	.	2	3	6
(c)						0	.	4	6	0
6. (a)					5	4	.	6	0	6
(b)					5	6	.	6	1	6
7. (a)						0	.	0	7	9

	(b)						**0**	.	**0**	**7**	**5**
	(c)						**0**	.	**0**	**6**	**8**
	(d)						**0**	.	**0**	**6**	**0**
8.	(a)				**5**	**0**	**5**	.	**4**	**0**	**0**
	(b)				**-**	**4**	**8**	.	**3**	**7**	**2**
	(c)				**-**	**4**	**9**	.	**1**	**7**	**2**
	(d)				**5**	**0**	**5**	.	**4**	**0**	**0**

1. **Problem**

   [**Please give your answer as a number with three decimal places. Example: 0.010.**]

   Given a normal distribution with mean 190 and standard deviation 101, use the pnorm function to calculate:

   (a) the probability of obtaining values between 202 and -62 from this distribution.

   Solution

```
> q1

[1] 202

> q2

[1] -62

> if(q1>q2){
+   upper<-q1
+   lower<-q2
+ }  else {
+   upper<-q2
+   lower<-q1
+ }
> upperprob<-pnorm(upper,
+              mean=abs(mean.val),
+              sd=abs(sd.val))
> lowerprob<-pnorm(lower,mean=abs(mean.val),
+           sd=abs(sd.val))
> sol<-round(upperprob-lowerprob,3)
```

   The probability is therefore: 0.541.

2. **Problem**
   Consider a normal distribution with mean 1 and standard deviation 1.

   Compute, to three decimal places, the lower and upper boundaries such that:

   (a) the area (the probability) to the left of the lower boundary is 0.28

   (b) the area (the probability) to the left of the upper boundary is 0.59.

Solution

You have to use the function `qnorm`.

The commands are

```
> ## lower bound:
> round(qnorm(prob1,mean=1,sd=1),digits=3)

[1] 0.417

> ## upper bound:
> round(qnorm(prob2,mean=1,sd=1),digits=3)

[1] 1.228
```

3. **Problem**

[Give answers up to three decimal places for each case.]

Take an independent random sample of size 148 from a normal distribution with mean 122, and standard deviation 41. Next, we are going to pretend we don't know the population parameters (the mean and standard deviation). We compute the MLEs of the mean and standard deviation using the data and get the sample mean 119.824 and the sample standard deviation 44.953. Compute:

(a) the estimated standard error using the sample standard deviation provided above.
(b) What are your degrees of freedom for the relevant t-distribution?
(c) Calculate the **absolute** critical t-value for a 95% confidence interval using the relevant degrees of freedom you just wrote above.
(d) Next, compute the lower bound of the 95% confidence interval using the estimated standard error and the critical t-value.
(e) Finally, compute the upper bound of the 95% confidence interval using the estimated standard error and the critical t-value.

Solution

```
> n

[1] 148
```

```
> sample.sd<-round(sd(x),3)
> sample.mean<-round(mean(x),digits=3)
> estimated.se<- round(sample.sd/sqrt(n),3)
> crit.t<-abs(round(qt(0.025,df=n-1),3))
> lower<-round(sample.mean-crit.t*estimated.se,digits=3)
> upper<-round(sample.mean+crit.t*estimated.se,digits=3)
```

4. **Problem**

   [Give answers up to three decimal places for each case. Example: 0.123.]

   Calculate the following probabilities:

   Given a normal distribution with mean 48 and standard deviation 3, what is the probability of getting

   (a) a score of 44 or less
   (b) a score of 44 or more
   (c) a score of 60 or more

   Solution

```
> p1<-round(pnorm(q,mean=mu,sd=sigma),digits=3)
> p2<-round(1-pnorm(q,mean=mu,sd=sigma),digits=3)
> p3<-round(1-pnorm(q1,mean=mu,sd=sigma),digits=3)
```

   Answer: 0.091,0.909, 0

5. **Problem**

   [Give answers up to three decimal places for each case. Example: 0.123.]

   Given a normal distribution with mean 48 and standard deviation 10, what is the probability of getting

   (a) a score of 43 or less
   (b) a score between 45 and 51
   (c) a score of 49 or more

   Solution

```
> mu
```

```
[1] 48

> sigma

[1] 10

> p1<-round(pnorm(mu-5,mean=mu,
+                 sd=sigma),digits=3)
> p2<-round(pnorm(mu+3,mean=mu,
+                 sd=sigma)-
+             pnorm(mu-3,mean=mu,
+                   sd=sigma),digits=3)
> p3<-round(1-pnorm(mu+1,mean=mu,
+                   sd=sigma),digits=3)
> p1

[1] 0.309

> p2

[1] 0.236

> p3

[1] 0.46
```

6. **Problem**

   Given a normal distribution with mean 55.611 and standard deviation 0.784. There exist two quantiles, the lower quantile q1 and the upper quantile q2, that are equidistant from the mean 55.611, such that the area under the curve of the Normal probability between q1 and q2 is 80%. Find q1 and q2.

   Give your answer to three decimal places.

   (a) lower bound:
   (b) upper bound:

   Solution

   Let 80 percent be the total probability that we want the bounds for. The sample mean is 55.611 and the standard deviation is 0.784. The commands are:

```
> mu
```

```
[1] 55.611

> sigma

[1] 0.784

> tailprob<-(1-prob)/2
> q1<-round(qnorm(tailprob,mean=mu,sd=sigma),digits=3)
> q2<-round(qnorm(tailprob,mean=mu,sd=sigma,lower.tail=FALSE),digits=3)
> q1

[1] 54.606

> q2

[1] 56.616
```

7. **Problem**

   [**Please give your answer as a number with three decimal places. Example: 0.010.**]

   Given the data point 12.822. The function dnorm gives the likelihood given a data point (or multiple data points) and a value for the mean and the standard deviation (sd). Using dnorm, compute

   (a) the likelihood of the data point 12.822 assuming a mean of 12 and standard deviation 5.
   (b) the likelihood of the data point 12.822 assuming a mean of 11 and standard deviation 5.
   (c) the likelihood of the data point 12.822 assuming a mean of 10 and standard deviation 5.
   (d) the likelihood of the data point 12.822 assuming a mean of 9 and standard deviation 5.

   Solution

```
> ## means:
> mn

[1]  9 10 11 12

> round(dnorm(x,mean=mn[4],sd=5),3)
```

```
[1] 0.079

> round(dnorm(x,mean=mn[3],sd=5),3)

[1] 0.075

> round(dnorm(x,mean=mn[2],sd=5),3)

[1] 0.068

> round(dnorm(x,mean=mn[1],sd=5),3)

[1] 0.06
```

8. **Problem**

   [**Please give each answer as a number with three decimal places. Example: 0.010.**]

   You are given 10 independent and identically distributed data points that are assumed to come from a Normal distribution with unknown mean and unknown standard deviation:

```
> x

 [1] 509 505 510 510 509 510 505 482 491 523
```

   The function dnorm gives the likelihood given multiple data points and a value for the mean and the standard deviation (sd). The log-likelihood can be computed by typing dnorm(...,log=TRUE).

   The product of the likelihoods for two independent data points can be computed like this: Suppose we have two independent and identically distributed data points 5 and 10. Then, assuming that the Normal distribution they come from has mean 10 and sd 2, the joint likelihood of these is:

```
> dnorm(5,mean=10,sd=2)*dnorm(10,mean=10,sd=2)

[1] 0.0017482
```

   It is easier to do this on the log scale, because then one can add instead of multiplying. This is because $\log(x \times y) = \log(x) + \log(y)$. For example:

```
> log(2*3)

[1] 1.7918
```

```
> log(2) + log(3)
```

```
[1] 1.7918
```

So the joint log likelihood of the two data points is:

```
> dnorm(5,mean=10,sd=2,log=TRUE)+dnorm(10,mean=10,sd=2,log=TRUE)
```

```
[1] -6.3492
```

Even more compactly:

```
> sum(dnorm(c(5,10),mean=10,sd=2,log=TRUE))
```

```
[1] -6.3492
```

Compute the following quantities:

(a) Given the 10 data points above, calculate the maximum likelihood estimate (MLE) of the expectation.
(b) The sum of the log-likelihoods of the data-points x, using as the mean the MLE from the sample, and standard deviation 5.
(c) What is the sum of the log-likelihood if the mean used to compute the log-likelihood is 503.4?
(d) Which value for the mean, the MLE or 503.4, gives the higher log-likelihood? As your answer, write either the MLE (the actual number, not the words MLE!) or 503.4.

Solution

```
> x
```

```
 [1] 509 505 510 510 509 510 505 482 491 523
```

```
> mn<-round(mean(x),3)
> mn
```

```
[1] 505.4
```

```
> loglik<-round(sum(dnorm(x,mean=mn,sd=5,log=TRUE)),3)
> loglik
```

```
[1] -48.372
```

```
> loglik2<-round(sum(dnorm(x,mean=mn-2,sd=5,log=TRUE)),3)
> loglik2

[1] -49.172

> loglik>loglik2

[1] TRUE
```

Clearly the log likelihood using the sample mean gives the higher log-likelihood. This is as it should be, because the mean from the sample is the maximum likelihood estimate.