ENV 710

inference

Duke | NICHOLAS SCHOOL OF THE ENVIRONMENT

roadmap

- quick summary of ‘where we are’, sampling and any questions?
- pod work – activities 1-3

descriptive statistics

↓

discrete probability/distributions

↓

continuous probability/distributions

↓

inference

sampling, central limit theorem, confidence intervals, t-distribution, p-values

0 – summary

Discuss the following terms/subjects (10 min.)

- explain the relationship between the population, a sample, and a sampling distribution
- what is the value of the standard deviation of a single sample?
- why is the standard deviation of the sampling distribution smaller than the standard deviation of the population?
- what are the tenets of the Central Limit Theorem?

- are the following statements true or false? why?

1. samples are used for making inferences about the population from which they were drawn
2. μ is a random variable (varying from sample to sample) and is an unbiased estimator of the sample mean, $\bar{x}$
3. if we double the sample size, we halve the standard error of the sample mean, $\bar{x}$

0 – summary

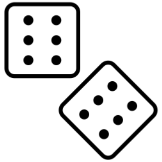
Discuss the following terms/subjects (10 min.)

- explain the relationship between the population, a sample, and a sampling distribution
- what is the value of the standard deviation of a single sample? – similar to σ
- why is the standard deviation of the sampling distribution smaller than the standard deviation of the population?
- what are the tenets of the Central Limit Theorem?

- are the following statements true or false? why?

1. samples are used for making inferences about the population from which they were drawn - true
2. μ is a random variable (varying from sample to sample) and is an unbiased estimator of the sample mean, $\bar{x}$
   - false, μ is fixed
3. if we double the sample size, we halve the standard error of the sample mean, $\bar{x}$
   - false, need to quadruple the sample size

1 – CLT



1. simulate rolling a die
2. make a histogram from 6 rolls of the die
   - what does it look like?
   - what is the mean of the 6 rolls?
   - what is the expected mean (population mean) for a six-sided die?

1 – CLT

1. simulate rolling a die
2. make a histogram from 6 rolls of the die

3. simulate 6 rolls of a die, 100 times
4. make a histogram of all your samples
   – what does it look like?
   – what is the mean of your samples?

5. repeat the above simulating 10 and 30 rolls, each 100 times

Rolling 1 die

```
roll <- runif(n=6, min=1, max=6)

ggplot(data.frame(roll), aes(x=roll)) +
          geom_histogram() +
            theme_bw()
```

Rolling 100 dice

```
roll2 <- c()
 for(i in 1:100){roll2[i]<-
     mean(runif(n=30, min=1, max=6))}

ggplot(data.frame(roll2), aes(x=roll2)) +
       geom_histogram(colour = "white") +
            theme_bw()
```

2 – gypsy moths

The gypsy moth is a serious threat to oak and aspen trees. The forestry department has placed millions of traps throughout the state to determine a population mass of 0.5 kilograms/hectare with a standard deviation of 0.7.

1. what is the mean and standard deviation of the average moth mass in 50 traps?
2. from your sample of 50 traps, what is the probability that the population mass is greater than or equal to 0.6?

2 – gypsy moths

The gypsy moth is a serious threat to oak and aspen trees. The forestry department has placed millions of traps throughout the state to determine a population mass of 0.5 kilograms/hectare with a standard deviation of 0.7.

- what is the mean and standard deviation of the average moth mass in 50 traps?

$$\bar{x} = 0.5 \qquad se = \frac{0.7}{\sqrt{50}} = 0.099$$

2 – gypsy moths

The gypsy moth is a serious threat to oak and aspen trees. The forestry department has placed millions of traps throughout the state to determine a population mass of 0.5 kilograms/hectare with a standard deviation of 0.7.

- from your sample of 50 traps, what is the probability that the population mass is greater than or equal to 0.6?

$$P(X \geq 0.6) = 0.156$$

```
1 - pnorm(q=0.6, mean=0.5, sd=0.099)

pnorm(q=0.6, mean=0.5, sd=0.099, lower.tail = F)
```

3 – student IQ’s

Suppose that the IQ’s of Duke University students can be described by a Normal distribution with mean of 130 and standard deviation of 8 points (population standard deviation)

1. select one Duke student at random. what is the probability that this student’s IQ is greater than 140?

2. select 5 Duke students at random. what is the probability that their *average* IQ is greater than 140?

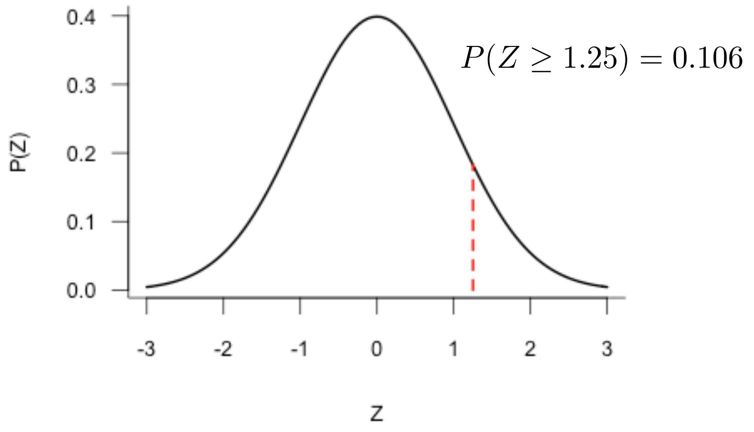
3 – student IQ's

Suppose that the IQ's of Duke University students can be described by a Normal distribution with mean of 130 and standard deviation of 8 points (population standard deviation)

1. select one Duke student at random. what is the probability that this student's IQ is greater than 140?

$$Z = \frac{x - \mu}{\sigma} = \frac{140 - 130}{8} = 1.25$$

```
1-pnorm(q=1.25, mean=0, sd=1)
```




```
1-pnorm(q=140, mean=130, sd=8)
```

Table entry for z is the area under the standard normal curve to the left of z.

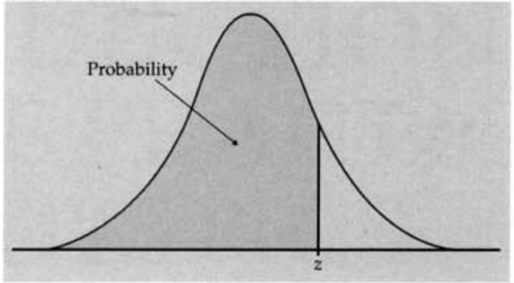


TABLE A Standard normal probabilities (*continued*)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	$\boxed{.8944}$	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

table gives the probability to the left of *z*

3 – student IQ's

Suppose that the IQ's of Duke University students can be described by a Normal distribution with mean of 130 and standard deviation of 8 points (population standard deviation)

2. select 5 Duke students at random. what is the probability that their *average* IQ is greater than 140?

$$Z = \frac{x - \mu}{\sigma/\sqrt{n}} = \frac{140 - 130}{8/\sqrt{5}} = 2.8$$

```
1-pnorm(q=2.8, mean=0, sd=1)
```

$$P(Z \geq 2.8) = 0.003$$

```
1-pnorm(q=140, mean=130, sd=8/sqrt(5))
```

3 – student IQ's

Suppose that the IQ's of Duke University students can be described by a Normal distribution with mean of 130 and standard deviation of 8 points (population standard deviation)

2. select 5 Duke students at random. what is the probability that their *average* IQ is greater than 140?

probability that the average IQ of the five students is greater than 140 is much smaller than the probability of one student having an IQ greater than 140

4 – trout

Researchers sample 40 rainbow trout, calculating a mean weight of at least 1.17 kg. The trout weights are normally distributed and come from a population with a mean of 1.1 kg and a standard deviation of 0.18.

1. calculate the Z-score for this problem
2. what does the Z-score signify or mean?
3. what is the probability of sampling 40 fish with a mean weight of 1.17 given the population mean of 1.1 kg and standard deviation of 0.18?

4 – trout

Researchers sample 40 rainbow trout, calculating a mean weight of at least 1.17 kg. The trout weights are normally distributed and come from a population with a mean of 1.1 kg and a standard deviation of 0.18.

1. calculate the Z-score for this problem

```
Z <- (1.17-1.1)/(0.18/sqrt(40))
[1] 2.459549
```

2. what does the Z-score signify or mean? – the number of standard deviations from the mean

4 – trout

Researchers sample 40 rainbow trout, calculating a mean weight of at least 1.17 kg. The trout weights are normally distributed and come from a population with a mean of 1.1 kg and a standard deviation of 0.18.

3. what is the probability of sampling 40 or more fish with a mean weight of 1.17 given the population mean of 1.1 kg and standard deviation of 0.18?

```
1-pnorm(q = 1.17, mean = 1.1, sd = 0.18/sqrt(40))

1-pnorm(q = (1.17-1.1)/(0.18/sqrt(40)))
```

Questions?