Classical and Bayesian Statistics

Problems 4

Problem 4.1

For the height of 18-20 year old men the mean value is 1.80 m with a standard deviation of 7.4 cm. The body height can be considered as normally distributed.

Make a sketch *by hand* for each of the following probabilities.

(∗∗) a) What is the probability that a randomly selected man in this age group is taller than 1.85 m?

(∗∗) b) What is the probability that a randomly selected man in this age group has a height between 1.70 m and 1.80 m?

(∗∗) c) In what symmetrical range around the mean are the heights of 50 % of the body heights?

(∗∗) d) How tall must a man be to be among the 5 % of the tallest men?

Problem 4.2

In one place there are several carp ponds. The mass of the carp is normally distributed with the expected value $\mu = 4$ kg and the standard deviation 1.25 kg.

(∗) a) What is the probability of catching a carp that is at most 2.5 kg?

(∗) b) What is the probability of catching a carp that weighs at least 5 kg?

(∗∗) c) What percentage of all carp weigh between 3 kg and 4.5 kg?

(∗∗) d) The Fishing Association wants to offer a prize for the heaviest carp.

What is the minimum weight required to have a probability of 2 % of getting the prize?

Problem 4.3

(∗∗) A cigarette manufacturer pretends that the nicotine content in a cigarette is on average 2.2 mg with standard deviation ofz 0.3 mg. However, for a sample of 100 randomly selected cigarettes, the sample mean is 3.1 mg.

If the cigarette manufacturer's statement is true, what is the probability that the sample mean reaches a value of 3.1 mg or more? Interpret your result.

Problem 4.4

The time a passenger spends at an airport check-in counter is a random variable with mean value 8.2 minutes and standard deviation 6 minutes. We randomly observe 36 passengers.

(∗) a) Calculate the probability that the average waiting time of these passengers is less than 10 minutes. Interpret your result.

(∗) b) Calculate the probability that the average waiting time of these passengers is between 5 and 10 minutes. Interpret your result.

(∗) c) Calculate the probability that the average waiting time of these passengers is more than 20 minutes. Interpret your result.

(∗ ∗ ∗) d) All of us have probably already had the experience of a longer waiting time at a check-in counter. Why is the probability of c) then so small?

(∗∗) e) Does the i.i.d. assumption hold here at all?

Problem 4.5

A lecturer knows from experience that the average score in an exam is 77 points with a standard deviation of 15 points. This semester the lecturer will teach two courses: one has 25 participants, the other 64.

(∗∗) a) What is the probability that the approximate average examination result in the course with 25 participants is between 72 and 82 points?

(∗) b) Repeat the calculation from part a) for the course with 64 participants. Compare and interpret the results a) and b).

Classical and Bayesian Statistic

Sample solution for Problems 4

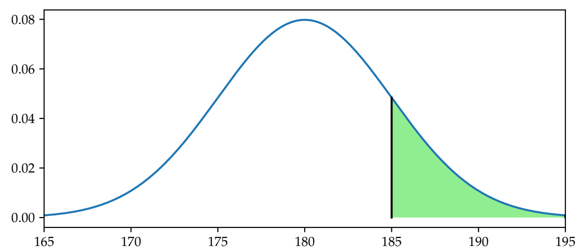
Solution 4.1

The random variable X denotes the body length of a randomly selected person. The distribution of X normally distributed:

$$X \sim \mathcal{N}(1.8, 0.074^2)$$

a) Sought is the probability $P(X \geq 1.85)$ and we obtain

$$P(X \geq 1.85) = 0.2496$$



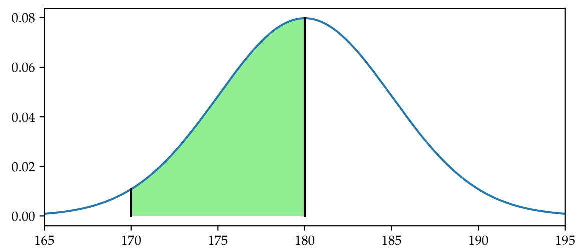
So about 25 % of the 18-20 year old men are taller than 1.85 m.

```
1 - pnorm(q = 1.85, mean = 1.80, sd = 0.074)

[1] 0.2496233
```

b) Sought is the probability $P(1.70 \leq X \leq 1.80)$ and we get

$$P(1.70 \leq X \leq 1.80) = 0.4117$$

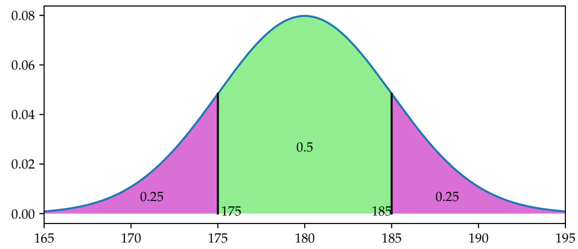


So about 41 % of the 18-20 year old men are between 1.70 m and 1.80 m.

```
pnorm(q = 1.80, mean = 1.80, sd = 0.074) - pnorm(1.70, 1.80, 0.074)

[1] 0.4117085
```

c) Sought are the quantiles $q_{0.25}$ and $q_{0.75}$ (these are just the lower and upper quartiles):

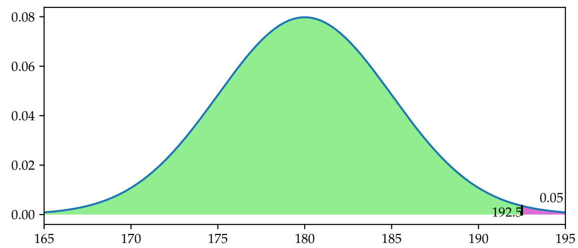



```r
qnorm(p = c(0.25,0.75), mean = 1.80, sd = 0.074)

[1] 1.750088 1.849912
```

That is, 50 % of the men are between 1.75 m and 1.85 m tall.

d) Sought is the quantile $q_{0.95}$:




```r
qnorm(p = 0.95, mean = 1.80, sd = 0.074)

[1] 1.921719
```

That is, 5 % of the men are taller than 1.92 m.

Solution 4.2

The random variable X denotes the weight of the carp. X is distributed as follows

$$X \sim \mathcal{N}(4, 1.25^2)$$

a) Sought is the probability $P(X \leq 2.5)$ and we obtain

$$P(X \leq 2.5) = 0.115$$

About 11 % of the carp weigh less than 2.5 kg.

```r
pnorm(q = 2.5, mean = 4, sd = 1.25)

[1] 0.1150697
```

b) Sought is the probability $P(X \geq 5)$ and we get

$$P(X \geq 5) = 0.212$$

About 21 % of the carp weigh more than 5 kg.

```
1 - pnorm(q = 5, mean = 4, sd = 1.25)

[1] 0.2118554
```

c) Sought is the probability

$$P(3 \leq X \leq 4.5) = 0.4436$$

About 44 % of the carp weigh between 3 kg and 4.5 kg.

```
pnorm(q = 4.5, mean = 4, sd = 1.25) - pnorm(3, 4, 1.25)

[1] 0.4435663
```

d) The quantile $q_{0.98}$ is sought

```
qnorm(p = 0.98, mean = 4, sd = 1.25)

[1] 6.567186
```

At 2 % of winning the prize, you have to catch a carp that weighs 6.57 kg or more.

Solution 4.3

Let X_i denote the random variable of the nicotine content in the i-th cigarette. We know that $\mu = 2.2$ and $\sigma_X = 0.3$.

We consider the average nicotine content $\overline{X}_{100}$, which according to the Central Limit Theorem (CLT) is approximately normally distributed as follows:

$$\overline{X}_{100} \sim \mathcal{N}\left(\mu, \frac{\sigma_X^2}{n}\right) = \mathcal{N}\left(2.2, \frac{0.3^2}{100}\right) = \mathcal{N}(2.2, 0.0009)$$

We are looking for $P(\overline{X}_{100} \geq 3.1)$ and get

$$P(\overline{X}_{100} \geq 3.1) \approx 0$$

```
1 - pnorm(q = 3.1, mean = 2.2, sd = 0.3/sqrt(100))

[1] 0
```

This probability is practically 0, that means that the mean 3.1 mg is extremely unlikely, *assuming* that the manufacturer's information of the average 2.2 mg is true. That indicates that something with the value 2.2 mg is dubious.

Note that even though `R` returns 0, the probability is *never* exactly 0.

Solution 4.4

Let X_i denote the random variable of the waiting time for the i-th passenger (in minutes). We know that $\mu = 8.2$ and $\sigma_X = 6$.

We consider the average waiting time $\overline{X}_{36}$, which according to the CLT is approximately distributed as

$$\overline{X}_{36} \sim \mathcal{N}\left(\mu, \frac{\sigma_X^2}{n}\right) = \mathcal{N}\left(8.2, \frac{6^2}{36}\right) = \mathcal{N}(8.2, 1)$$

a) Sought is $P(\overline{X}_{36} \leq 10)$ and we get

$$P(\overline{X}_{36} \leq 10) = 0.9640697$$

```
pnorm(q = 10, mean = 8.2, sd = 6/sqrt(36))

[1] 0.9640697
```

The probability of the average waiting time of less than 10 minutes for the these 36 passengers is quite high. That means this group has to wait for *more* than 10 minutes is quite low. So, they unlikely have to wait for more than 10 minutes *on average*.

b) Sought is $P(5 \leq \overline{X}_{36} \leq 10)$ and we calculate

$$P(5 \leq \overline{X}_{36} \leq 10) = 0.9633825$$

```
pnorm(q = 10, mean = 8.2, sd = 6/sqrt(36)) - pnorm(q = 5, mean = 8.2, sd = 6/sqrt(36))

[1] 0.9633825
```

The difference to b) is very small. This means that the probability that the average waiting time is below 5 minutes is very small.

c) Sought is $P(\overline{X}_{36} \geq 20)$ and we get

$$P(\overline{X}_{36} \geq 20) \approx 0$$

```
1 - pnorm(q = 20, mean = 8.2, sd = 1)

[1] 0
```

The probability that the average waiting time is longer than 20 minutes is *very* small. It is almost impossible to wait more than 20 minutes on average.

d) The probability that *you* can wait more than 20 minutes is of course much higher.

The probability in c) describes the probability that 36 randomly chosen people waited *on average* more than 20 minutes and that probability is almost 0.

The probability that a lot of people have to wait more than 20 minutes is *on average* less than the probability that *one* person has to wait more than 20 minutes.

e) This is a tricky one. If you consider a large airport and choose passengers randomly from *any* check-in counter at *any* time of the day, then the i.i.d. assumption seems to be justified.

If you pick *one* check-in counter at random and choose 36 passengers from *that* counter then the assumption is not justified. If there is already a long queue then *most* passenger wait longer than average.

The same applies if you choose a *specific* time during the day, when the airport is unusually busy. The waiting time for *most* passengers would be longer.

Solution 4.5

Let X_i denote the random variable for the score of the i-th student. We know that $\mu = 77$ and $\sigma_X = 15$.

a) We consider the average number of points $\overline{X}_{25}$, which is according to the CLT approximately distributed as

$$\overline{X}_{25} \sim \mathcal{N}\left(\mu, \frac{\sigma_X^2}{n}\right) = \mathcal{N}\left(77, \frac{15^2}{25}\right) = \mathcal{N}(7.7, 9)$$

We are looking for $P(72 \leq \overline{X}_{25} \leq 82)$ and obtain

$$P(72 \leq \overline{X}_{25} \leq 82) = 0.9044193$$

```
pnorm(q = 82, mean = 77, sd = 15/sqrt(25)) - pnorm(72, 77, 15/sqrt(25))

[1] 0.9044193
```

b) We consider the average number of points $\overline{X}_{64}$, which is according to the CLT approximately distributed as

$$\overline{X}_{64} \sim \mathcal{N}\left(\mu, \frac{\sigma_X^2}{n}\right) = \mathcal{N}\left(77, \frac{15^2}{64}\right) = \mathcal{N}(7.7, 225/64)$$

We are looking for $P(72 \leq \overline{X}_{64} \leq 82)$ and get

$$P(72 \leq \overline{X}_{64} \leq 82) = 0.9923392$$

```
pnorm(q = 82, mean = 77, sd = 15/sqrt(64)) - pnorm(72, 77, 15/sqrt(64))

[1] 0.9923392
```

This probability is larger then in a). The reason is that there are more students in the class and the larger sample reduces randomness *on average*.