Top of Form

1. Let us assume the Cancer Treatment Centers of America is using real-world data from patients to build a linear regression model to predict the right dosage amount for chemotherapy. Let this linear regression function be specified by the following equation:

Now let’s assume your female friend Lorri has recently been diagnosed with breast cancer. You two used to go to the gym regularly and thus you know her height to be 180 cm and weight to be 165 lb. You also recently gifted her a Fitbit on her 40-th birthday. The doctors are planning to go ahead with chemotherapy and have determined the dosage amount to 92 mg/m2.

**Predict** her blood type (,  or and diabetes type (type1 or type2) and show the intermediate steps for deriving the answer**.** **[15 points]**

1. Assume you are conducting a random response survey where you are trying to approximate - “What fraction of the population use drug?” To preserve user’s anonymity each participant is asked to respond in the following manner:

* Throw one fair dice (with six sides).
* If **dice value <= 4**, then respond truthfully
* If **dice value > 4**, then respond untruthfully

1. Draw the randomized response graph. **[5 points]**
2. If 60% of the observed participants answer ‘Yes’ to the question ‘Do you use drugs?’ then what is the **actual** fraction of the population using drugs? **[5 points]**
3. In the context of a Federated Learning based approach-
   1. How does a central server converge to a global model without accessing the sensitive data of each participating node? Describe the model updating process. **[5 points]**
   2. What is the basic difference between secure aggregation and differential privacy in the Federated Learning setting? **[5 points]**
4. In the context of ensuring fairness for machine learning algorithms-
   1. What does statistical parity mean? **[5 points]**
   2. Assume a company uses ML techniques to select candidates. In the past one year the hiring statistics look like the following:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Gets a job | Doesn’t get a job | Total |
| Female | 20 | 80 | 100 |
| Male | 40 | 80 | 120 |
| Total | 60 | 160 | 220 |

Calculate the conditional probability of landing a job given the candidate is a female versus male. Is the selection process biased towards one gender? **[5 points]**

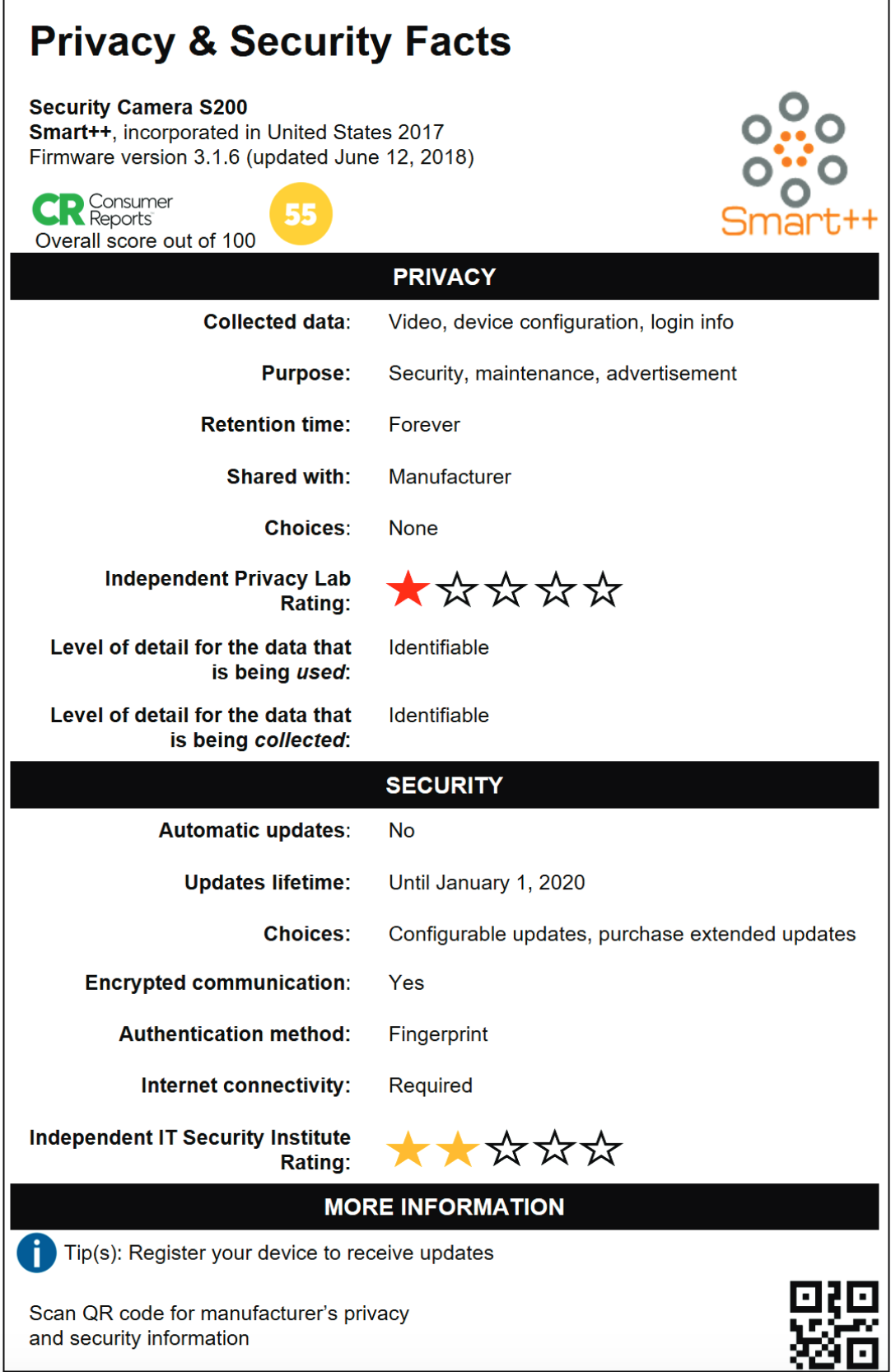
1. Compute the t-closeness for the following equivalence classes (P1, P2), where the original salary distribution is Q. Show the steps involved in the computation. **[15 points]**

Q= {3k,4k,5k,6k,7k,8k}

P1= {4k,7k}

P2= {3k,5k,6k,8k}

1. Briefly explain how bitcoin protects against –
   1. tampering of existing transaction records **[5 points]**
   2. double spending **[5 points]**
2. Answer the following questions:
   1. What are the main limitations of Westin’s privacy index study? **[5 points]**
   2. What are the different factors that influence users’ privacy behavior/attitude under real-world scenarios? **[5 points]**
3. How does P3P protocol work? How would you extend P3P protocol to support real-time privacy notice and choice for IoT devices? **[5+5 points]**
4. Consider the following privacy notice for an IoT device. Comment on what GDPR guidelines this privacy notice does or does not imply with. **[10 points]**



1. **[Bonus]** Imagine you want to develop a smart security and privacy assistant for your home. Assume it has access to all communications taking place between devices and those going out to the Internet (think of it like a smart router). What type of security and privacy capabilities would you recommend implementing and why? **[10 points]**