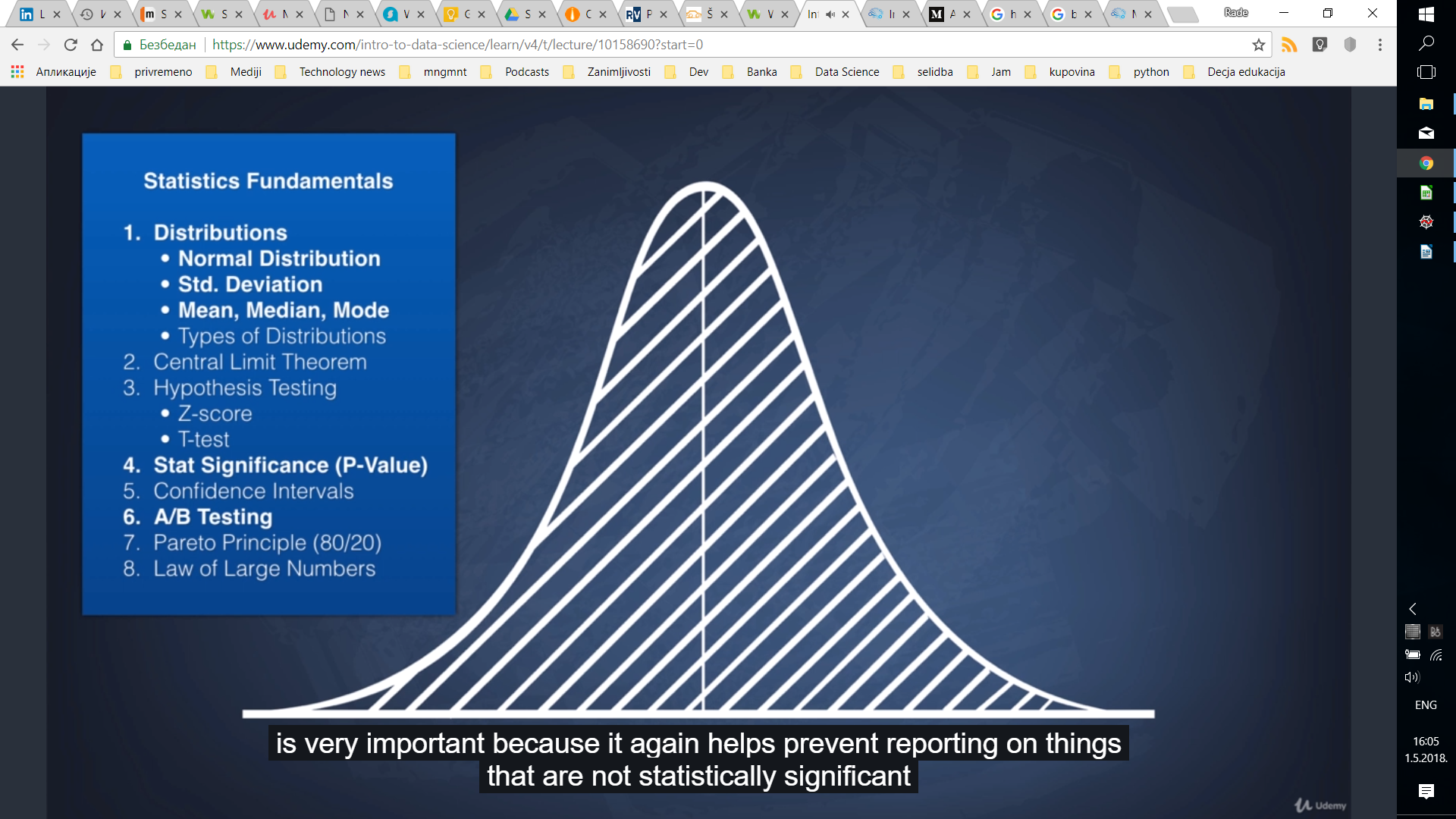
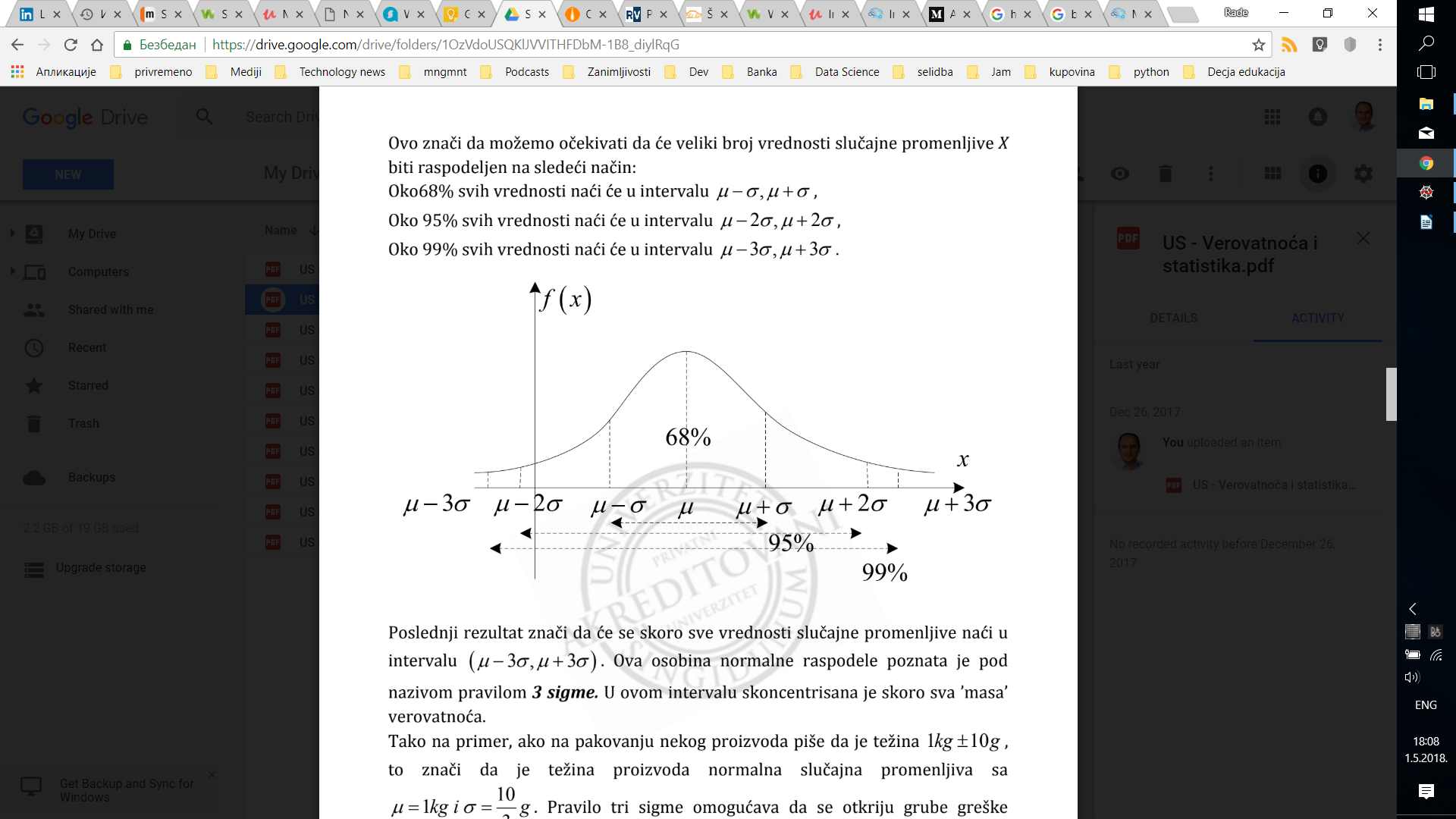
Data Science kurs

Sta je vazno iz statistike?



najvazniji su 1, 4 i 6

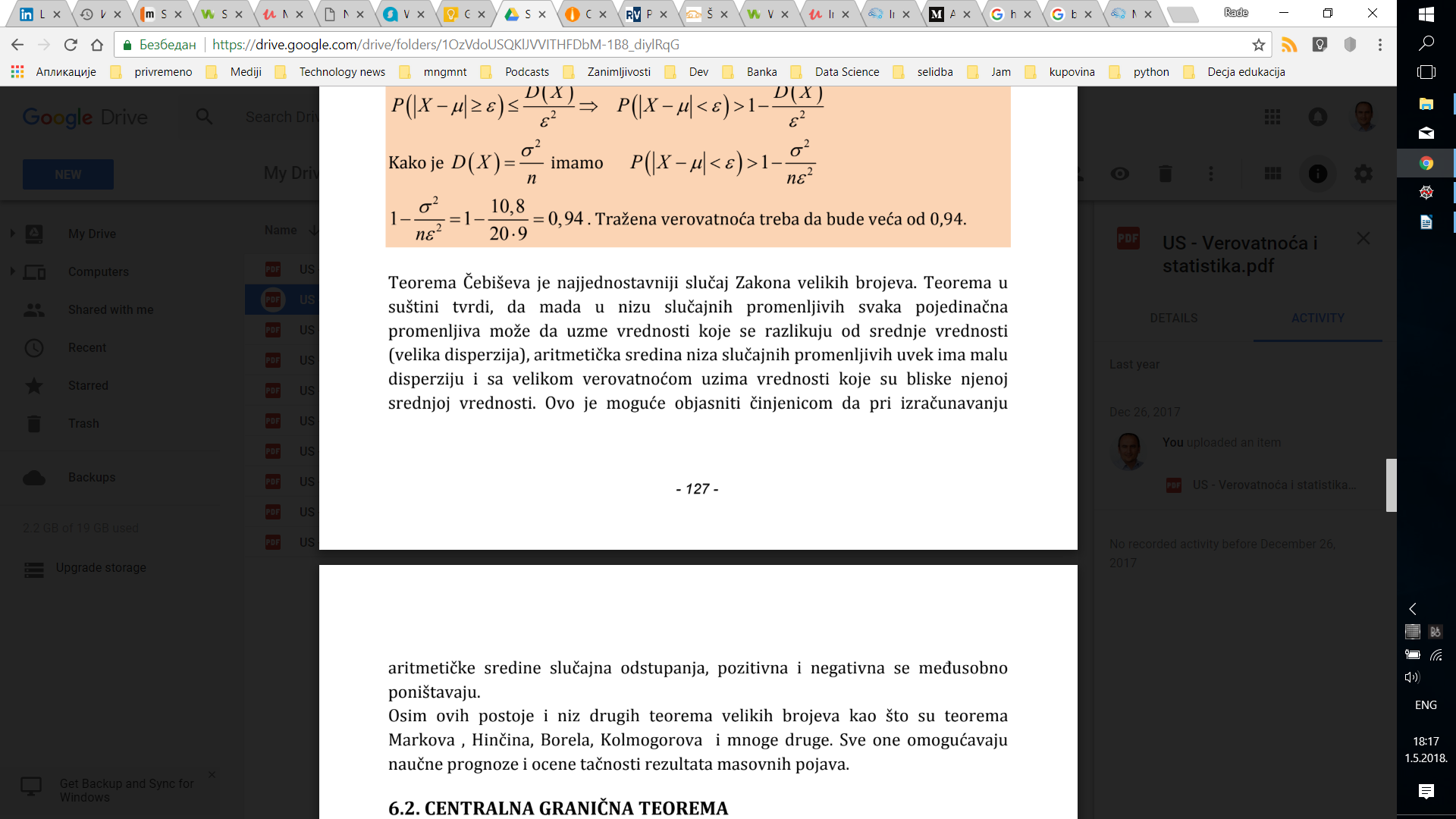
Normalna raspodela



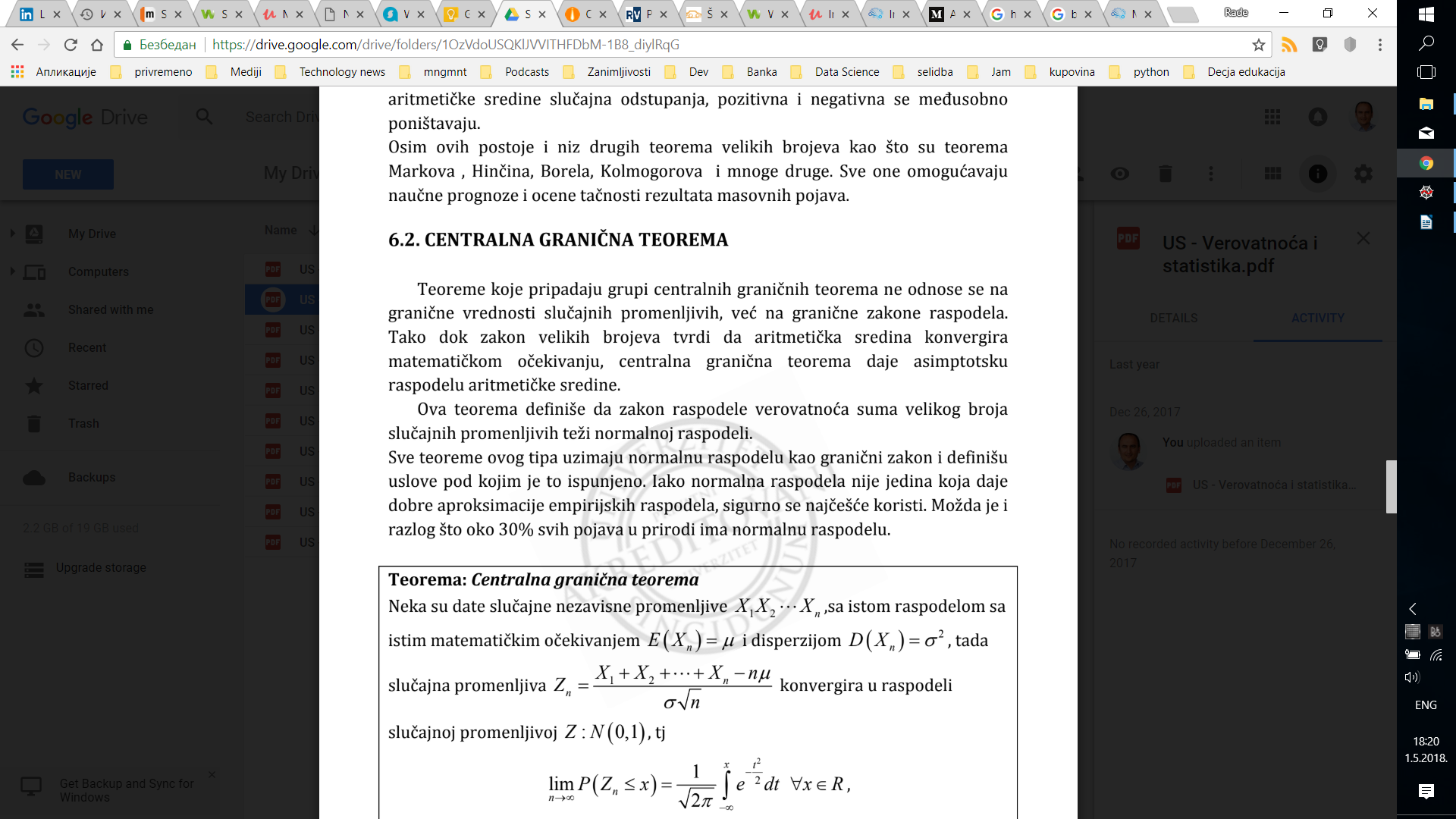
hi-kvadratna raspodela



Zakon Velikih brojeva



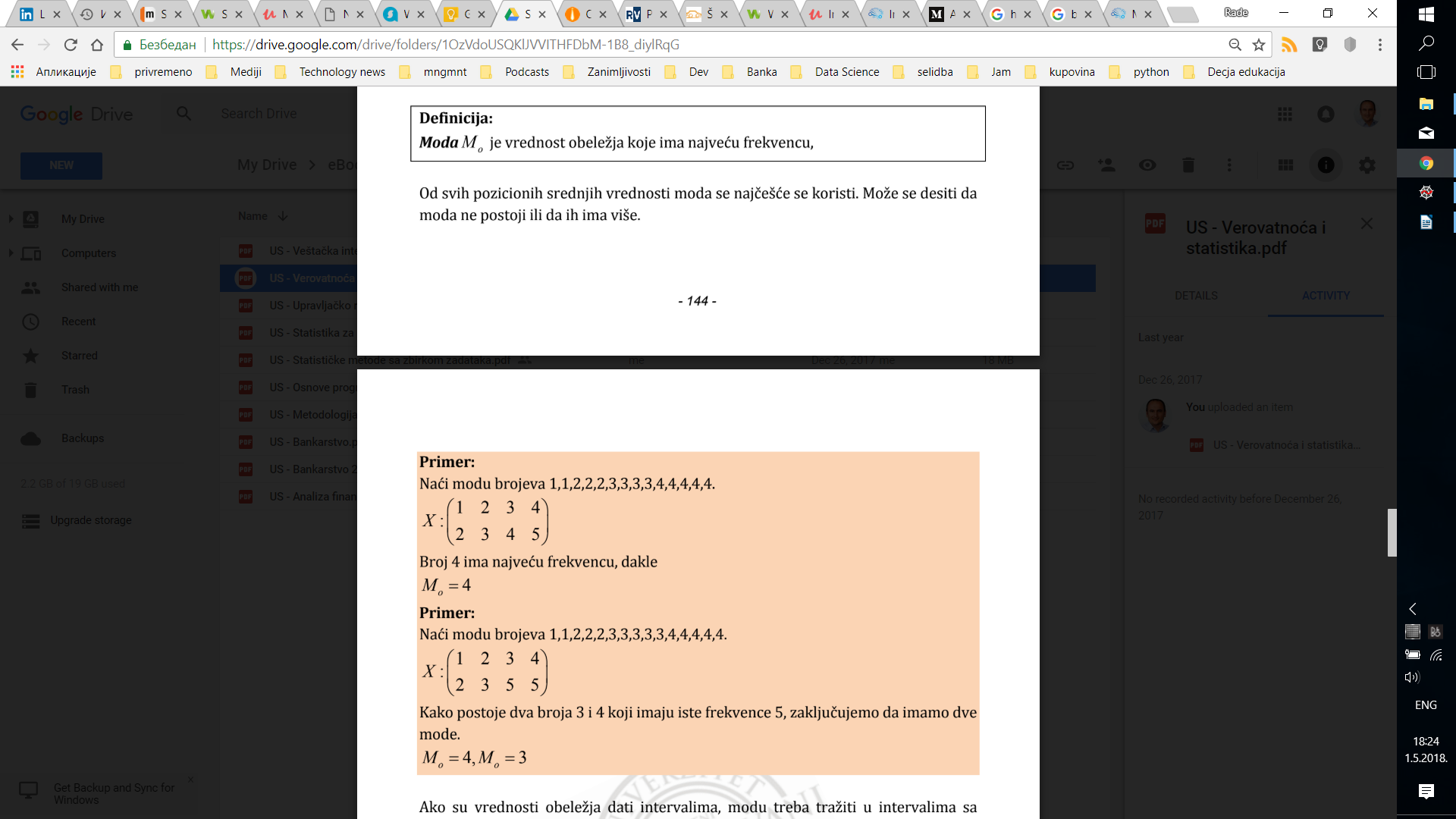
Centralna granicna teorema



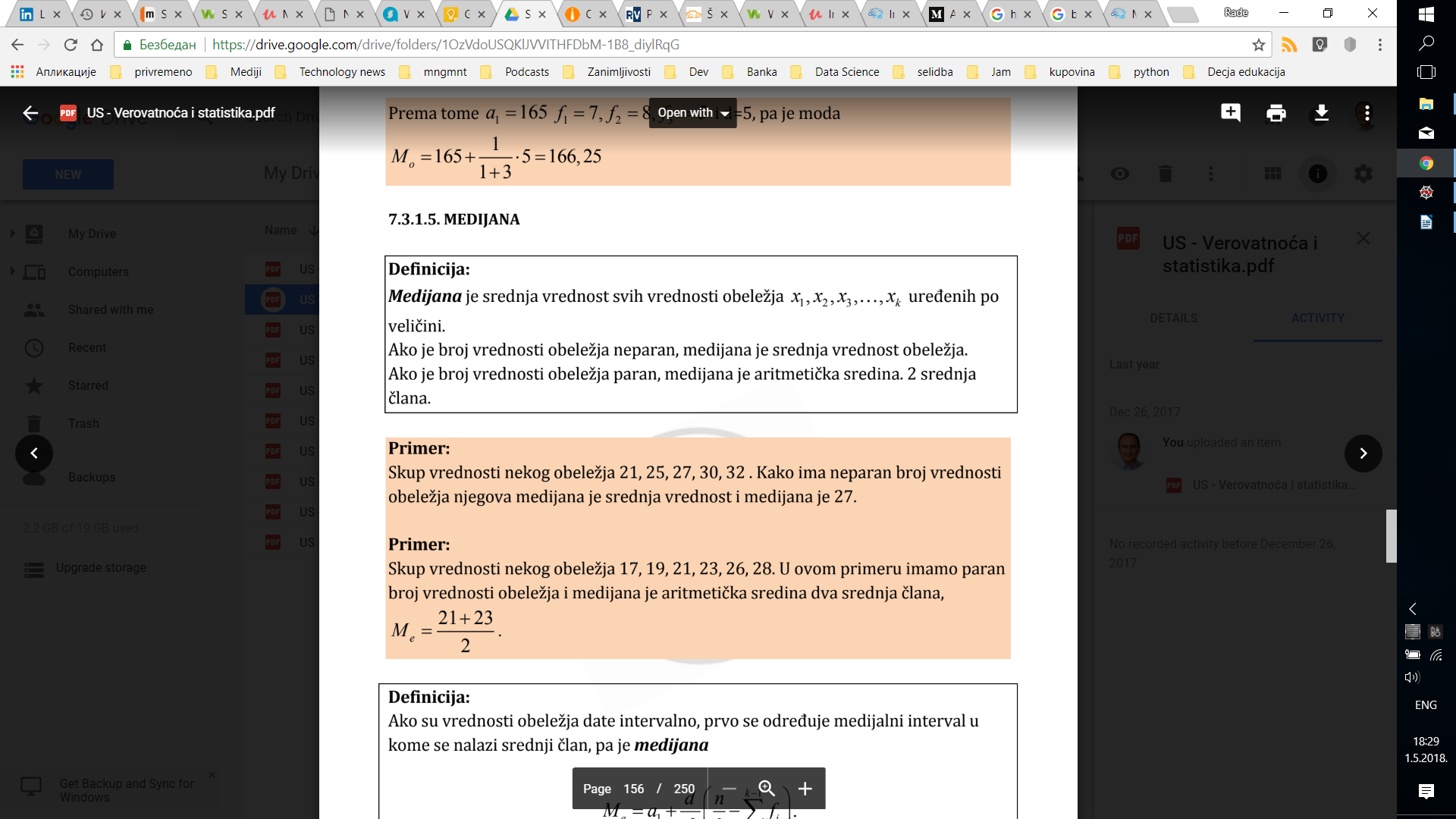
Srednja vrednost

Average

Moda (ako se neki elemenat skupa pojavljuje najvise puta onda je on moda)



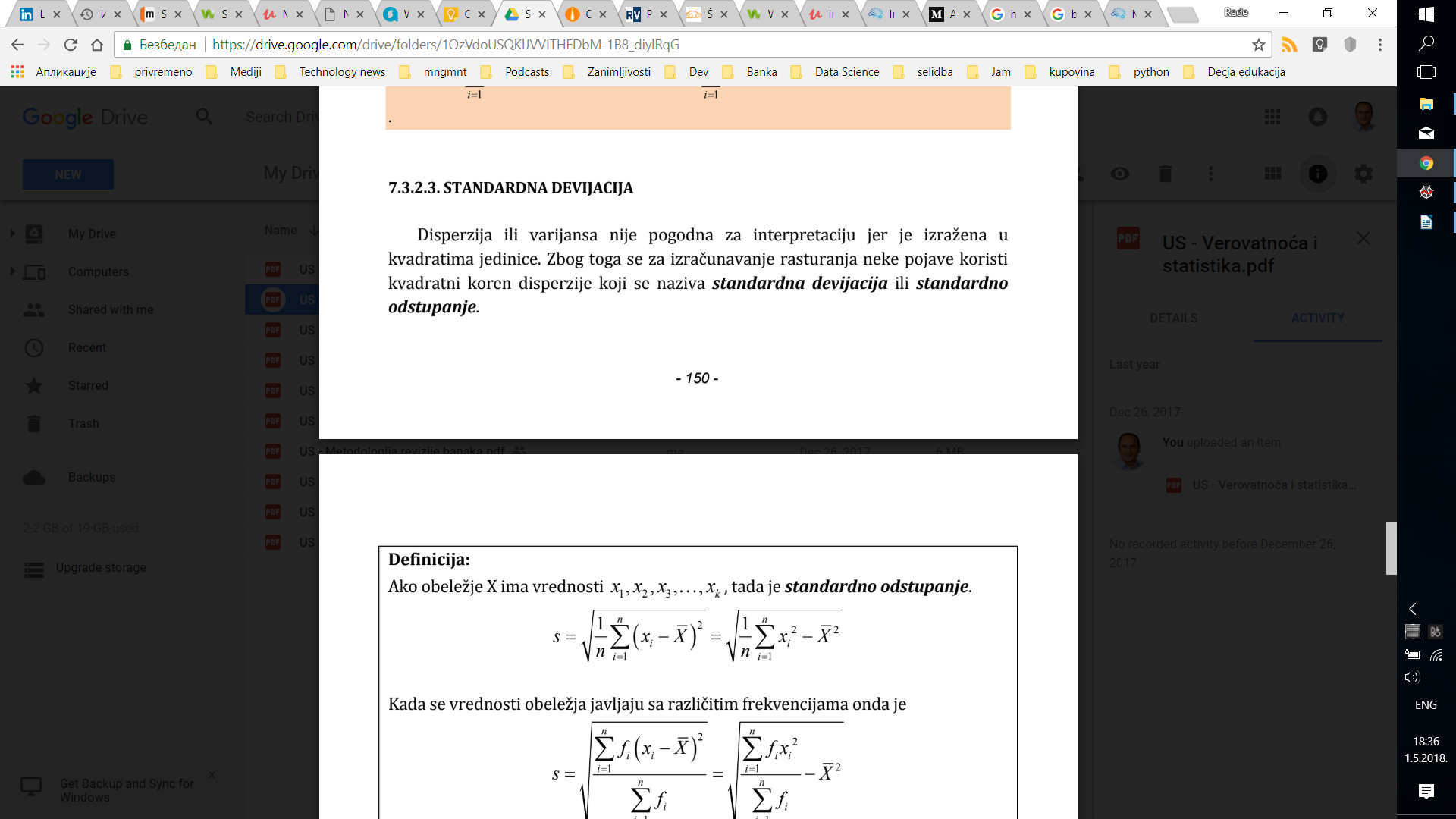
Medijana (ako slozim ceo skup po velicini pogledam koji je srednji elemenat i to je medijana)



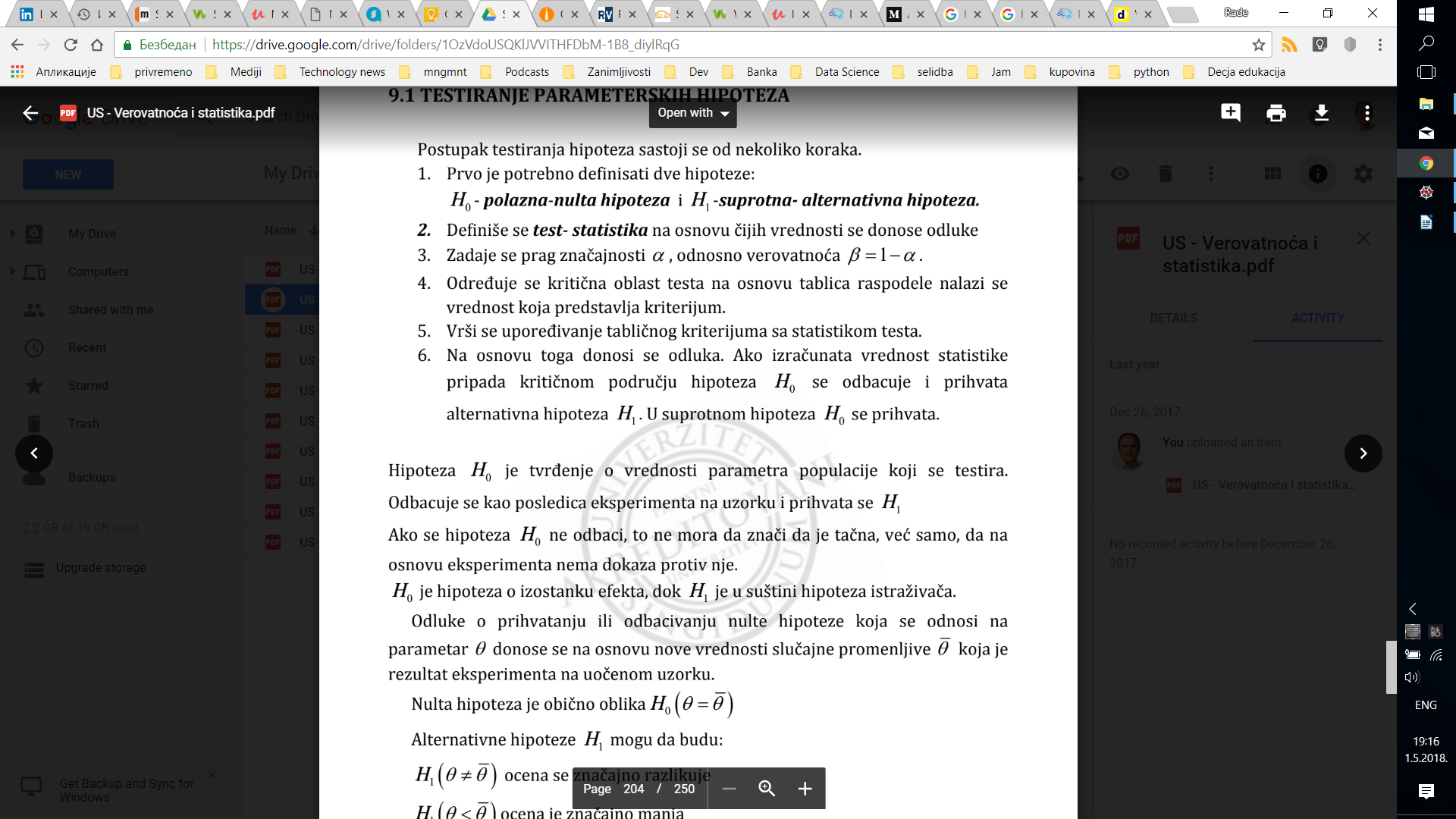
Varijansa ili disperzija (koliko je prosecno odstupanje podataka od srednje vrednosti svih podataka)

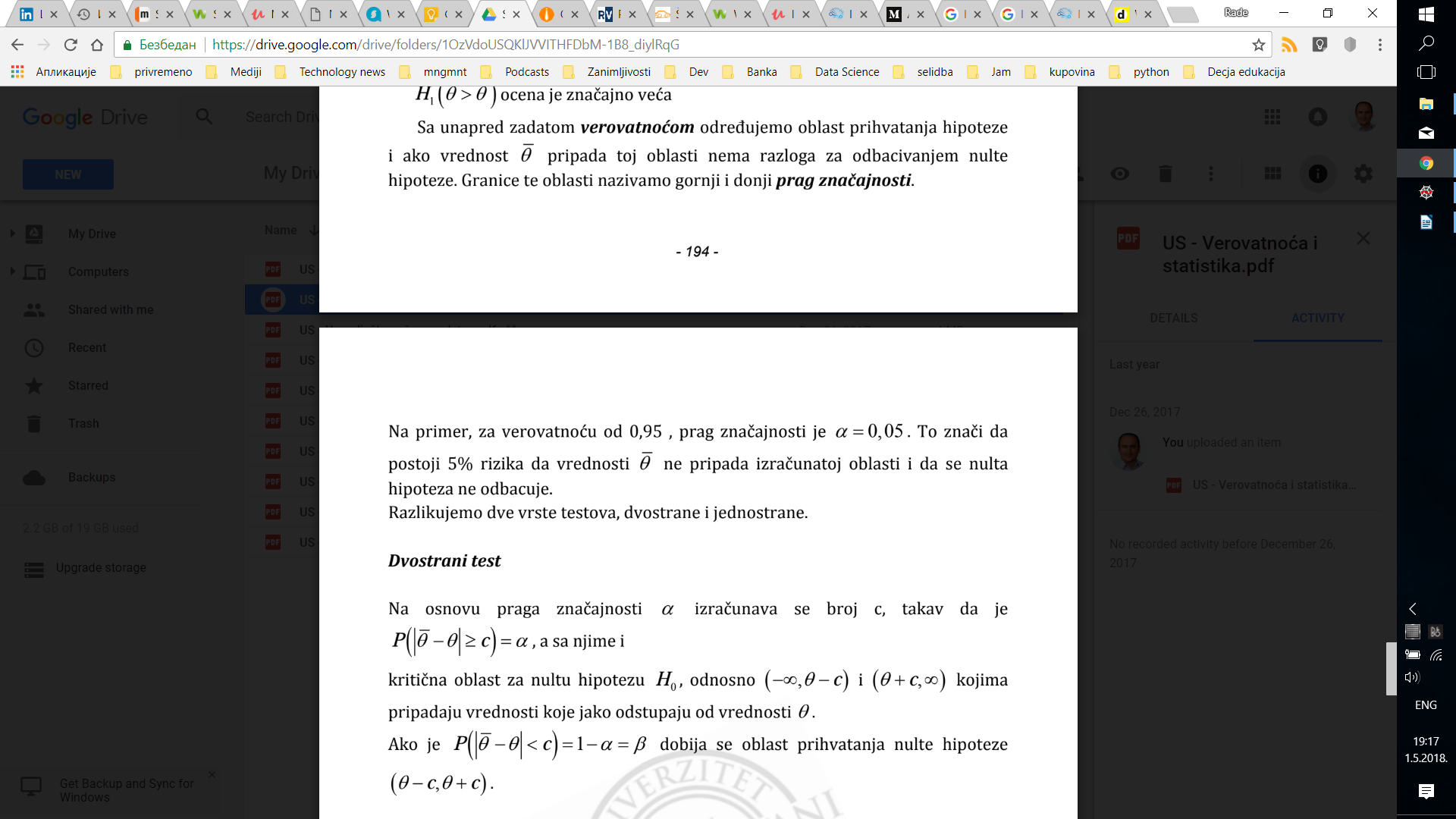


Standardna devijacija



p-Value i testiranje hipoteza (p value u nasoj literaturi je alfa)





**Pytorch je najbolji reference za ML i Deep learning**

Dva najevca referentna uzora Geofrey Hinton i Andrew Ng

Computer vision modeli

Open CV nije deep learning i nije ni najbolja biblioteka ali je mozda najpoznatija

SSD je veoma mocan

YOLO v2 (joseph redmon) je mozda i mocniji od SSDa

GANs za kreiranje slika (Cycle Gans jos bolji). Radi na nivou pixela

Deep NLP (i ovde postoji midel koji nije deep learning a zove se bag of words)

Sentiment analysis (input je tekst a output je klasifikacija npr positive ili neka druga klasa kao englez ili drama)

Machine translation modelom Seq2Seq (koristi se i za pravljenje chatbotova). Na githubu je mnogo ovih translator modela.

Chatbot se pravi istim Seq2Seq kao kod prevodjenja (input je neka recenica a output je neka recenica samo kod prevodjenje je input recenica na engleskom a kod chatbota je mozda forma pitanja). Feedujem chatbota velikom kolicinom tekstova i on onda nauci sta treba da odgovori. Npr. uzmes sve twitter diskusije ili stotine konverzacija iz filmova

Deep reinforcment learning (AI)

Knjiga Reinforcement learning za teoriju a gym.openai.com je mesto gde covek moze da vezba deep reinforcment.

Q-learning nije deep learning tj nema neurona… deep qlearning je vec uvodjenje neurona. A sa ocima koje nesto vide pravi se deep convolutional q-learning. Uvodenje memorije I onda imamo deep convolutional lstm q-learning. A3C je kada dodamo I kriticki mozak na prethodnu pricu.

Visualization

Tableau

Bank marketing data

Masovna komunikacija je jeftina ali neefikasna. Personalna komunikacija je vec mnogo bolja ali zog kompleksnosti proizvoda vrlo je vazno da agent (CC Ili u expozituri) razgovara sa klijentom. Vec sledeci put korisnik ce znati o proizvodu I mozda nije neophodno prodavati na taj nacin (sto se pokazalo na sheetu previous campaign poutome). Orocena stednja mozda nije pravi primer ali recimo prodaja Payday loan-a odnosno kratkorocne pozajmice je nesto sto korisnicima treba vise puta tokom godine ali im ne treba objasnjavati svaki put. Dakle, treba resiti problem kome prodavati I na koga trositi agentsko vreme.

Poenta je da su podaci nekakvo sredstvo pomocu koga mozemo da pravimo bolje odluke bilo ka ustedama bilo ka zaradama bilo ka optimalnim resenjima kao u ovom slucaju OVO MOZE DA BUDE ZAKLJUCAK.

Isprobamo svaku varijablu kao prediktor za y (zavisnu varijablu) I na taj nacin sprovodimo “vizuelno istrazivanje” (visualization for data mining).

GRETL alat koji moze da odredi koje varijable su znacajne za linearne modele (gradient boosting I deep learning onda ne moze da radi)

AWS upuststvo u pdf-u…

mada, obzirom na moj komp verovatno nije potrebno jer 4CPU I 16GB linux virtuelna masina na AWS-u kosta 0.2 USD/hour (1CPU na 4 GB kosta 0,05 USD/hour)

osim AWS-a I drugi su dobri a neobican provajder je paperspace

Nevezano za kurs:

na ovoj adresi je jedan ceo postupak Data Science-a gde je ML samo deo postupka i polazi od Exploaratory Data analize u pythonu na stvarno comprehensive nacin ima i github ceo sa postupkom

<https://towardsdatascience.com/bayesian-linear-regression-in-python-using-machine-learning-to-predict-student-grades-part-1-7d0ad817fca5>

[https://github.com/WillKoehrsen/Data-Analysis/blob/master/bayesian\_lr/Bayesian%20Linear%20Regression%20Project.ipynb](https://github.com/WillKoehrsen/Data-Analysis/blob/master/bayesian_lr/Bayesian Linear Regression Project.ipynb)