tend to be gaussian. 1.1.2 The probability of getting 7 or more 3s in one roll with 20 dice is calculated by $1-\sum_{n=0}^6 C_{20}^n(\frac{1}{6})^n(\frac{5}{6})^{20-n} \approx 0.0371$ 1000 sample for N3 distribution N3 Frequency 0.1 10 Values With random sampling: Getting 7 or more 3s in a roll is 0.03714 1.2.1 The positive fraction of PCR is $\frac{2464}{103261} pprox 0.0239$, and the positive rate of AntiGen is $\frac{491}{26162} pprox 0.0188$ Using confidence interval between portion. Variance of p1-p2 is calculated by $rac{p1(1-p1)}{n1}+rac{p2(1-p2)}{n2}$ p1-p2 is 0.0051156826 The variance of p1-p2 is 0.0000009279Give 75% confidence interval, p1-p2 +/- 1.15*SE(p1-p2) where SE = sqrt(variance)Therefore, 75% confidence interval of p1-p2 should be 0.0047753 < 0.0051157 < 0.0069908So we have at least 75% confidence that they are statistically the same 1.2.2 Assuming that PCR test is totally accurate. Then the sample AntiGen use supposes to have $2464 imes 26192 \div 103261 pprox 625$ people infected. Therefore, the false negative rate is $(625-491) \div 625 = 21.44\%$ 1.2.3 First we have 0.02% false positive and 20% false negative, given: $47 = P(1 - 20\%) + (50000 - P) \times (0.02\%)$ where P is the positive in reality. Result in $P=37 \div 0.7998 \approx 46.2615654$. Therefore, the positive rate is P/50000, approximately 0.0925%1.3.1 Voltage Measurement 2000000 Voltage 1750000 **Npoints** 1000000 STD 2461.692 1500000 mean 16.793 1250000 1000000 750000 500000 250000 0 200000 400000 600000 800000 1000000 0 Num of measurements 1.3.2 In order to fit the peak in the diagram, first we need to seach all the peaks out, and then zoom in to see what the peaks looks like. The peaks is 541575 612736 644727 679105 755368 849590 872460 875504 915527 920077 983782] Voltage Measurement 2000000 Voltage 1750000 **Npoints** 1000000 STD 2461.692 16.793 mean 1500000 1250000 1000000 750000 500000 250000 200000 400000 800000 600000 1000000 0 Num of measurements Zoom in one peak to seek for the distribution. Notice that it follows a gaussian distribution 1200000 Voltage Voltage 1000000 800000 600000 400000 200000 0 28795 28800 28805 28815 88660 88670 28810 88665 88675 88680 Num of measurements Num of measurements 1200000 Voltage Voltage 1000000 800000 600000 400000 200000 173,205 173,210 173,215 173,220 173,225 440,210 440,215 440,220 440,225 Num of measurements Num of measurements No.0 Peal 200000 60000 50000 150000 40000 30000 20000 50000 1000 10,835 28,805 81,495 No.3 Peal No.4 Peak No.5 Peal 120000 60000 150000 20000 200000 88,670 163,305 173,210 173,215 173,220 No.7 Peak No.8 Peak No.6 Peak 200000 ---- Chi2 fit 150000 150000 125000 150000 100000 75000 75000 50000 25000 25000 375,640 440,215 483,655 No.9 Peak No.10 Peak No.11 Peak 60000 Chi2 fit 400000 150000 30000 100000 20000 50000 100000 10000 514.500 514.505 514.510 514.515 541.570 541.575 541.580 612,735 612,740 612,725 612,730 No.12 Peak No.13 Peak No.14 Peak ---- Chi2 fit ---- Chi2 fit Chi2 fit 120000 60000 100000 50000 60000 30000 100000 644,725 644,730 644,735 679,100 679,105 679,110 679,115 755,365 755,370 755,375 175000 80000 150000 125000 20000 849,590 849,595 872,460 872,465 875,505 200000 50000 80000 100000 20000 10000 The fit is given relative uncertainty 10% to obtain a better fitting. Nevertheless, some of the fitting is still problematic because the peak is so sharp that gaussian distribution is hard to catch up. Almost every peak lasts only for one frame. **II** – Error propagation: 2.1.1 $x=1.96\pm0.03$, therefore $\sigma_x=0.03$, $\sigma_y^2=(rac{\partial y}{\partial x})^2\sigma_x^2$, $\sigma_z^2=(rac{\partial z}{\partial x})^2\sigma_x^2$. $rac{\partial y}{\partial x}=rac{-2x}{(1+x^2)^2}$ $\frac{\partial z}{\partial x} = \frac{2}{(1-x)^3}$ The value of y is 0.207, and the uncertainty(sigma y) is 0.0050The value of z is 1.085, and the uncertainty (sigma z) is 0.06782.1.2 The value of y is 0.520, and the uncertainty(sigma y) is 0.0156 The value of z is 625.000, and the uncertainty(sigma z) is 937.5000 2.2.1 Weighted mean is $rac{\sum x_i/\sigma^2}{\sum 1/\sigma^2}$, and the related SD is $\sqrt{rac{1}{\sum 1/\sigma^2}}$ Directly look at the percentage uncertainty Group 1: 1.572% Group 2: 1.068% Group 3: 1.098% Group 4: 0.811% Group 5: 1.403% Group 6: 0.609% Group 7: 0.304% Group 8: 1.325% Group 9: 0.409% While looking at the percentage uncertainty, the best group is group 7 with $g = 9.86 \text{ m/s}^2$, and sigma = 0.03 m/s² Meanwhile, they also own the best absolute uncertainty The best estimation from me is weighted average The calculation result of weighted mean of g is 9.824 m/s^2 And the mean of uncertainty 0.020 m/s^2 2.2.2 $\chi^2=32.396$ and p-value = 0. From the graph, we can say that the measurement group 1 and 2, 3, 5 are outside of $3\sigma_g$. Besides, although group 5 is also out of range, its uncertainty is large enough to touch the range of our estimation. Hence, group 1, 2, 3s' measurements are unlikely. weighted mean 10.0 9.8 9.6 9.824 +/- 0.020 Chi2 32.396 9.4 ndof 0.000 Prob 2 3 5 Ó 4 2.2.3 Yes! The value $9.8158\pm0.001m/s^2$ is less than one sigma away from my best estimate $9.824\pm0.020m/s^2$! So they agree with each other. III - Monte Carlo 3.1.1 First, inspect the PDF. We found that the highest probability is 1.25. Then, we sample with accept and reject, with t = [0, 10] (Since ft(10) = 4.66 * 1e-6, small enough!), and a random number = [0, 1.25]. Text(0.5, 0, 't') PDF of f(t) 1.25 1.00 Probability 0.75 0.50 0.250.00 20 40 60 80 100 0 t Histogram of u with gaussian fit 0.25 Npoints 1000 STD 1.717 0.20 mean 3.906 Frequency 0.15 0.10 0.05 0.00 12 10 14 Value 3.1.2 Assume that the bin count is Poisson distributed. The p-value is zero, which makes sense because the plot has a long tail. Therefore, the distribution is not normal distribution Histogram of u with gaussian fit 200 Chi2 / ndof 2239.1 / 12 Prob 0.000 177.037 +/- 0.624 175 3.615 +/- 0.007mu 1.713 +/- 0.007sigma 150 125 Frequency 100 75 50 25 6 10 12 14 Value 3.1.3 I will try to fit with some distributions here. c:\users\gaozheming\appdata\local\programs\python\python37\lib\site-packages\ipy kernel_launcher.py:5: RuntimeWarning: divide by zero encountered in log c:\users\gaozheming\appdata\local\programs\python\python37\lib\site-packages\ipy kernel_launcher.py:5: RuntimeWarning: invalid value encountered in true_divide Histogram of u with lognormal fit 160 Chi2 / ndof 21.8 / 12 Prob 0.039 Ν 597.307 +/- 19.219 140 1.280 +/- 0.016mu sigma 0.484 +/- 0.015120 100 Frequency Chi2 fit model result Data, with Poisson errors 80 60 40 20 12 8 10 Value Seems it works well, try the skewed normal distribution Histogram of u with skewed normal fit Chi2 / ndof 236.5 / 12 175 Prob 0.000 457.925 +/- 16.577 N -0.000 +/- 0.077 150 loc 3.349 +/- 0.092125 Frequency 001 Chi2 fit model result Data, with poisson errors 75 50 25 12 Value Not good, try double normal distribution. Histogram of u with double gaussian fit Chi2 / ndof 15.1 / 9 Prob 0.088 175 N1 91.419 +/- 26.523 Ν2 130.327 +/- 29.446 150 mu1 4.763 + / - 0.444mu2 2.934 +/- 0.222sigma1 1.756 +/- 0.135125 sigma2 1.065 +/- 0.157Frequency 001 Chi2 fit model result Data, with poisson errors 75 50 25 2 6 8 10 12 14 Value The double gaussian fit gives a p-value of 0.09, which is rather good! Histogram of u with exponormal fit 200 22.8 / 12 Chi2 / ndof Prob 0.029 Κ 1.564 +/- 0.069 175 2.376 +/- 0.067 745.106 +/- 23.847 loc 150 125 Frequency 001 Chi2 fit model result Data, with expnorsson errors 75 50 12 14 8 Value After χ^2 fit of the exponential normal distribution, the p-value is 0.03, which also make sense. Therefore, double gaussian fit and exponormal fit might be options. 3.2.1 The F(x) of PDF is $-Ce^{-x}(x+1)$. Since $-e^{-x}(x+1)|_0^\infty$ is already nomalized, C can be 1.Take a look at the PDF distribution, the largest value of f(x) is no higher than 0.4. The PDF of f(x)x*exp(-x)0.3 $\widehat{\mathbf{z}}^{\,0.2}$ 0.1 0.0 2 8 10 6 4 Х The value of x=20 is 0.00000459Therefore, the upper boundary can be set to 20. Next, generate random numbers with accept and reject method. Plot one of the experiment with histogram. No 0 's experiment x median value is 1.66781 No 1 's experiment x median value is 1.62487 No 2 's experiment x median value is 1.72595 No 3 's experiment x median value is 1.66640 No 4 's experiment x median value is 1.57138 No 5 's experiment x median value is 1.65675 No 6 's experiment x median value is 1.63958 No 7 's experiment x median value is 1.71817 No 8 's experiment x median value is 1.60462 No 9 's experiment x median value is 1.66333 Histogram of 1000 values of x 200 count histogram frequency 100 50 2 0 4 8 x value From sampling, the median of x is around 1.6. Anallytically, when $CDF = 0.5, x \approx 1.67835$. IV - Statistical tests 4.1.1 Use fisher exact test to seek the significant level of the vaccine experiment. The p-value of Fisher exact test is 1.5664e-38 Which is much lower than 5% Since the number is large, chi2 test in contingency is also a choice The p-value of Chi2 test is 6.6419e-32 Which is also much lower than 5% Therefore, probability that the vaccine has worse effect is less than 1e-30 On the other hand, a binomial test can also offer a result Given the affect rate 0.0075, assume the distribution is binomial, then do the binomial distribution test to see if we need to accept that BNT162b2 has no effect P-value is 3.3355309441616695e-58 Therefore, the probability of vaccine has no effect is low At last, we do a null hypothesis test 68% confidence interval of p1-p2 should be 0.0064895 < 0.0070875 < 0.0076854So we have at least 68% confidence that the vaccine is useful But less than 75% confidence (75% overflows) 4.1.2 Use risk ratio $\mathrm{RR}=p_{vaccine}/p_{placebo}$, then the 68% confidence interval of the efficacy is: $(1-\operatorname{RR}\exp(1\cdot\sqrt{1/N_v+1/N_p}),1-\operatorname{RR}\exp(-1\cdot\sqrt{1/N_v+1/N_p}))$ Therefore, the interval is (0.92904, 0.96561) We see the BNY162b2 efficacy is in 68 confident interval: 0.92904 < 0.95062 < 0.965614.1.3 Assume that vaccine is useless, do a binomial test for the severe cases. Assuming the p=0.5, and do two sides test. That the probability that BNT162b2 had no effect is 0.02148 4.2.1 The distribution of the number of aces follows binomial distribution B(4, 1/13). Number of 0 has probability $C_4^0(\frac{12}{13})^4$, 1 has $C_4^1(\frac{12}{13})^3(\frac{1}{13})^1$ and so on. Check if the card is intact 4 4 4 4 4 4 4 4 4 4 4 4 4 Chance of getting 3 aces or more is 0.0017156 Ace number probability with replacement Probability 0.4 0.0 4 Ace number 4.2.2 Without replacement, probability of drawing more than 3 aces is $C_4^1(\frac{4}{52})(\frac{3}{51})(\frac{2}{50})(\frac{48}{49}) + (\frac{4}{52})(\frac{3}{51})(\frac{2}{50})(\frac{1}{49}) = 0.000713$ 4.2.3 Test the number of odd after even and so on The number of odd even, odd odd, even even, even odd is 15 9 13 14 Correlation coefficient with suit is -0.09053914] [-0.09053914 1.]] Correlation coefficient with value is -0.17567809] [-0.17567809 1. Hist2D of suit 4.0 3.5 -3.0 2.5 2.0 1.5 1.0 20 10 30 40 50 Number Hist2D of value 14 12 10 8 6 4 2 10 20 30 40 50 Number Looks like the randomness of the cards is ok, although there might be weak relat ion between value and the sequence of cards. So I can say it is not really well shuffled, but ok shuffled. V – Fitting data: 5.1.1 Price as function of cum capacity price 100 **Npoints** 44 STD 21.271 80 13.092 mean price / \$/W 60 40 20 0 2000000 500000 1000000 1500000 2500000 Cumulative solar power capacity / MW 5.1.2 c:\users\gaozheming\appdata\local\programs\python\python37\lib\site-packages\ipy kernel_launcher.py:2: RuntimeWarning: divide by zero encountered in power Fit with cumulative data 150 Chi2 fit 125 100 75 price wei avg 0.670 +/- 0.027 price_avg 13.092 86.598 Chi2 50 ndof 42 Prob 0.000 72.891 +/- 3.345 25 b 0.381 +/- 0.0060 200000 300000 100000 400000 500000 600000 Cumulative solar power capacity / MW The best χ^2 fit still can't get a decent p value. so the power law fit might not be a good fit.

Kian Gao

shp593

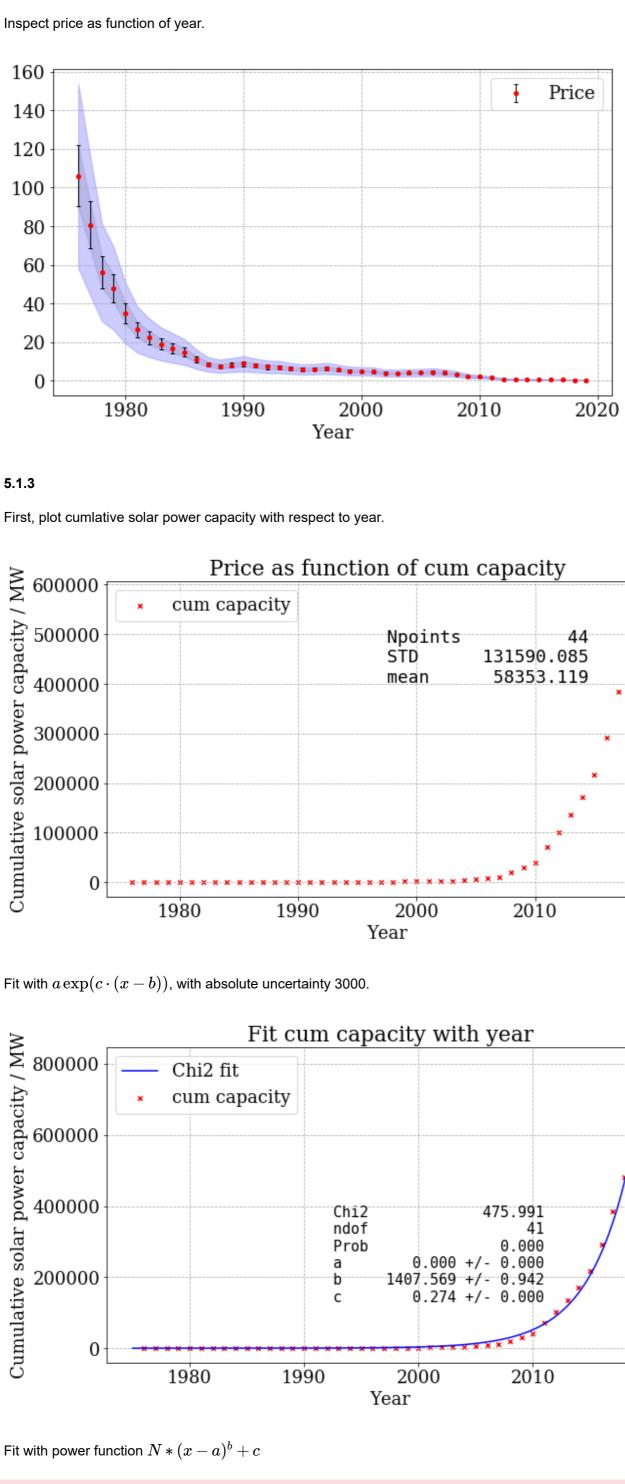
1.1.1

Final exam on AppStat2020

I – Distributions and probabilities:

The probability of N_3 follows binomial distribution. First, for each roll, the N3 distribution is binomial distribution N(20, 1/6).

Another property is that according to central limit theorem, if number of dice is large enough, the final distribution of N3 will

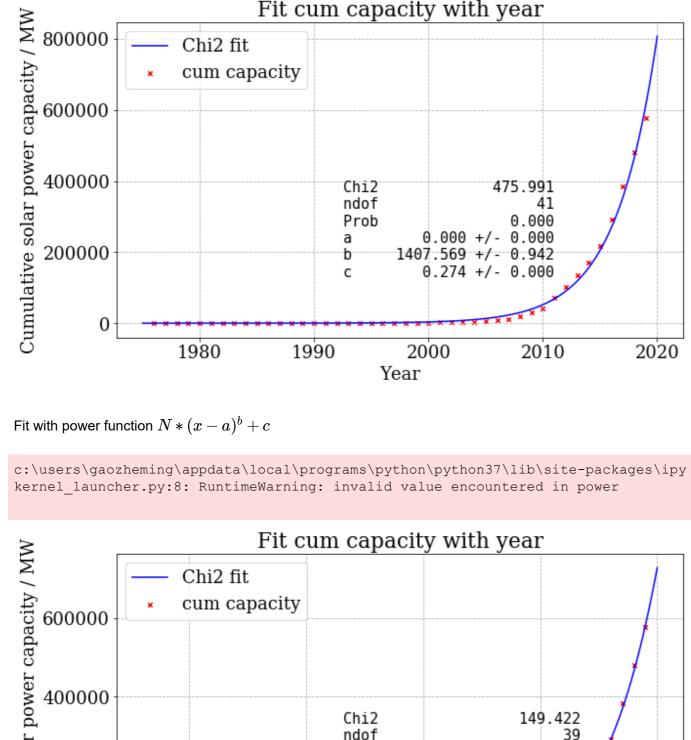


2020

100

80

5.1.3



Cumulative solar power capacity / MW composity / MW composition of the Chi2 ndof 149.422 39 Prob 0.000 1975.982 +/- 0.215 9.647 +/- 0.016 -4230.903 +/- 622.650 0.000 +/- 0.000 b c N 1990 1980 2000 2010 2020 Year

Fit with 7 order polynomial function, set year 1976 as 0 point (Only shown in code). Fit cum capacity with year Chi2 fit 700000 cum capacity 600000 500000 Chi2 ndof Prob 44.142 400000 а

Cumulative solar power capacity / MW 36 0.165 -0.000 +/- 0.000 b 0.002 +/- 0.000 300000 -0.081 +/- 0.000 1.033 +/- 0.005 d 200000 10.615 +/- 0.170 -294.869 +/- 5.982 1540.968 +/- 146.185 -1312.190 +/- 1310.945 f g h 100000 0 1980 1990 2000 2010 2020 Year Predict cum capacity 4000000 Predict by polynomial fit 3500000 3000000 2500000

Decent fit, using the polynomial fit to estimate the price $% \left(1\right) =\left(1\right) \left(1\right)$ Do a price predict in the next 10 years Cumulative solar power capacity / MW 2000000 1500000 1000000 500000 2022 2013 2016 2019 2025 2028 2010 Year The cumulative capacity is over a million MW when at 2022. When the price is 0.37533 /W.Approximately equals to the real price around 2019. 5.2.1

2031 Positive Positive rate 2500 100000 2250 2.2 ₹ 2.0 90000 2000 Positive Percentage / 1.6 Npositives 1500 80000 70000 1250 60000 1000 1.0 50000 750 8 10 12 14 2021 January Dates 8 10 12 14 2021 January Dates The average number of tests from 4th to 18th is approximately 83090.5.2.2 Do the χ_2 fit.

Fit Scaled positives Chi2 fit SP Definition SP 2500 Scaled bositives 1500 Chi2 1472894.241 ndof 12 0.000 Prob 2701.151 +/- 0.000 sp0 0.617 +/- 0.000 2.714 +/- 0.000 r 1000 500 6 8 14 16 4 10 12 18 Date in January 2021

5.2.3

Chi2 fit with t0-t SP Definition SP 3000 Chi2 ndof Prob Scaled positives 1000 26.223 12 0.010 sp0 2701.271 +/- 0.000 0.616 +/- 0.000 2.715 +/- 0.000 t0 0 6 10 12 4 8 14 16 18 Date in January 2021 By test, when the absolute systematic uncertainty touch 237, the P-value will be reasonable, which is by now 0.01. 5.2.4 Analytically, seperate R out we get: $R=(rac{ ext{SP}(t)}{ ext{SP}_0})^{(t_0-t)/t_g}$ Using this, and assume that $SP_0, t_0, SP(t), t$ are super accurate (without uncertainty), we can in\text{SP}ect the contribution when t_G has ± 1.0 days. $egin{aligned} \sigma_R^2 &= (rac{\partial R}{\partial t_G})^2 \cdot \sigma_{t_G}^2 \ \sigma_R^2 &= (rac{ ext{SP}(t)}{ ext{SP}_0})^{(t_0-t)/t_g} \cdot ext{ln}[rac{ ext{SP}(t)}{ ext{SP}_0}] \cdot rac{(t-t_0)}{t_g^2})^2 imes 1 \end{aligned}$ Now plug in fit Sp0, t0, tg, and t, Sp. On Day $\,$ 4 , The uncertainty on R is 0.02982 On Day $\,$ 5 , The uncertainty on R is 0.08520 On Day $\,$ 6 , The uncertainty on R is 0.132737 , The uncertainty on R is 0.131688 , The uncertainty on R is 0.10190 On Day 9 , The uncertainty on R is 0.03716

On Day 10 , The uncertainty on R is 0.01402On Day 11 , The uncertainty on R is 0.06247

Fit Scaled positives

On Day 12 , The uncertainty on R is 0.04577On Day 13 , The uncertainty on R is 0.02100On Day 14 , The uncertainty on R is 0.02420On Day 15 , The uncertainty on R is 0.00780 On Day 16 , The uncertainty on R is 0.00110 On Day 17 , The uncertainty on R is 0.00044On Day 18 , The uncertainty on R is $0.00445\,$ The largest uncertainty contribution from tg is around 0.133. $[{\tt NbConvertApp}] \ {\tt Converting \ notebook \ FinalExam2020_KianGao.ipynb \ to \ {\tt PDFviaHTML}}$ [NbConvertApp] Writing 3913682 bytes to FinalExam2020_KianGao.pdf