**Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Midterm Exam 01**(Part 1 in-class: open book, notes, laptop, calculators. It is closed internet, wireless, other people.)

(Part 2 take-home solo submission: open everything and everyone worldwide. Cite your sources.)

**Instructions**: To help you budget your time, questions are marked with \*s, with more stars indicating more difficulty. If you can’t solve a question completely, solve it as far as you can. Describe in detail what you could do to solve the problem. **Show your work on all problems.**

**Round final solutions to three decimal places unless specified otherwise.**

**Question setup**: The games Pass the Pigs™ and Toss Up!™ are often compared due to their similar press your luck style of play. So I thought, why not combine them. Call it Toss the Pigs™ by me. Begin your turn by rolling 10 pigs. You may collect any of the pigs that score (land in a position other than on their side). Then you have the choice to "bank" your points or to roll the remaining dice (the ones that had landed on their side). You proceed until you choose to stop and bank your total points, or until you make a roll where none of the pigs score and you get 0 points for your turn (this game's version of a "pig out"). The strategy is knowing when to stop. For example, rolling 9 or 10 pigs is pretty safe (small chance that no pigs score), but rolling 1 or 2 pigs is risky (moderate chance that no pigs score). A risk can pay off though. If during your turn, all 10 of your pigs score, then you can choose to keep that turn going, starting with rolling 10 more pigs and racking up more points for that turn. So in theory, there is no limit to how many points you can score in one turn. The game ends when someone banks 100 or more points (the sudden death variation where players get one last shot to beat the leader is also an option).

|  |  |  |
| --- | --- | --- |
| http://averweij.web.cern.ch/averweij/random%5Crandom_08.JPGimage used without permission from: http://averweij.web.cern.ch/averweij/random%5Crandom\_08.JPG | **Possible positions of a pig die:** | **Points for that pig die** |
| Dot side | 0 |
| No dot side | 0 |
| Razorback | 5 |
| Trotter | 5 |
| Snouter | 10 |
| Leaning jowler | 15 |
| Getting 0 points for a roll results in 0 points for the whole turn, just like a "pig out" or "running a red light" in the aforementioned games. |

**\*1 - 4 pts)** Let X = the points scored for a TURN of Toss the Pigs™, i.e. the points score from one or more rolls of the pigs. What is the sampling space for X?

**\*2 - 4 pts)** Assuming that each pig die lands with the following probabilities and is independent of the outcome of the other pig dice, what is the probability of a "pig out" when rolling just one die, i.e.
P(rolling no points | rolling one die)?

|  |  |  |
| --- | --- | --- |
| Side with no dot: 0.35 | On its back: 0.22 | On its snout: 0.03 |
| Side with a dot: 0.30 | On its feet: 0.09 | Leaning on its jowl: 0.01 |

**\*3 - 4 pts)** Referring to question 2, what is the probability of a "pig out" when rolling six dice, i.e.
P(rolling no points | rolling six dice)?

**\*4 - 6 pts)** Referring to X in question 1 (X = points scored in one turn), I spent the weekend rolling 1900 turns of Toss the Pigs™. The mean of the 1900 turns was 30 and the standard deviation was 20. Assume this is a representative sample of my game play, e.g. my strategy of when to bank my points was and is stable. Calculate an approximate **90%** confidence interval for the true mean of one of my turns of Toss the Pigs™. **Show the "by hand" calculations and write a brief sentence to justify the method you used.**

**\*\*5 - 10 pts)** Thinking back to question 4, you decide to design a simulation to test the true coverage rate of an asymptotic Normal 95% CI for the mean points scored per turn given a sample size N = 1900. The exact sampling distribution of the sample mean would be very hard to calculate, so you decide to (1) simulate the scores from 1900 turns given my personal behavior pattern, (2) calculate the CI using that sample's mean and sd, and (3) check to see if the CI worked, i.e. check whether it contains the true parameter. You can get the "true" parameter by taking the mean of a sample of one billion turns. You repeat steps 1-3 many times and use the proportion of times the CI worked as your estimate for the true coverage rate. Call the number of times you repeat steps 1-3 the "SimulationLoops". In order to ensure your estimate of the true coverage rate is accurate to two decimal places with near certainty, you want a **99.7%** asymptotic Normal CI for the true coverage rate to have a **half width of 0.001**. What is the minimum number of SimulationLoops that will ensure this?

**\*6 - 4 pts)** To examine how often I would "pig out" (get 0 points on a turn), I played 100 turns in which I recorded 11 pig outs. Using the asymptotically Normal (Wald) interval, calculate a **95%** confidence interval for my true probability of pigging out on a turn. **Show the "by hand" calculations**.

**\*7 - 4 pts)** Referring to question 6, calculate a 95% **"exact"** confidence interval for my true probability of pigging out on a turn. **Show the R command(s) or "by hand" calculations**. A single R command is fine.

**\*8 - 4 pts)** Referring to question 6, calculate a 95% **"percentile bootstrap"** confidence interval for my true probability of pigging out on a turn. **Show the R command(s) or "by hand" calculations**.

**\*\*9 - 8 pts)** When making up the rules, I took the number of points needed for a win in Toss the Pigs™ straight from Pass the Pigs™ (100 pts). I'm worried this wasn't a good choice. If the game goes too long or ends too quickly, it won't be fun. For example, if the true mean number of turns to win is just 4, the game will go by too fast. To test this out, I play one game and reach 100 pts in 7 turns. Calculate an exact 95% CI for the true mean number of turns to win. **Show the R command(s) or "by hand" calculations**. **In a brief sentence, describe whether or not I can stop worrying based on this data.** Hint: Notice I didn't say what sample statistic to use. You decide what could work well here.

**\*10 - 2 pts)** Referring to question 9, I decide to collect more data. I play 80 games and find the mean number of turns to win is 4.5. In standard mathematical notation, write the null and alternative hypothesis for a two-sided test of my concern that the true mean is 4 turns to win.

**\*\*11 - 10 pts)** Propose and calculate a test statistic for the hypothesis test in question 10. Write a brief sentence justifying your choice. Hint 1: 80 games is a pretty big sample size in this setting. Hint 2: Notice I didn't give you the sample's standard deviation. Think about all the consequences of assuming the null hypothesis is true and decide if you still need the sample's sd.

**\*12 - 2 pts)** Calculate the two-sided p-value for the test statistic in question 11. Write a brief sentence interpreting the results at a 5% significance level.

**\*\*13 - 4 pts)** Hemoglobin A1c (mmHg) is a physiological measure of blood glucose. In a certain population with diabetes, the distribution of A1c can be approximated well using a chi-square distribution. Let X ~ chisq(df = 6). Let A1c = X/2 + 4. What is E[A1c]? **Show the R command(s) or "by hand" calculations**.

**\*\*14 - 4 pts)** Referring to question 13, what is the standard deviation, SD[A1c]? **Show the R command(s) or "by hand" calculations**.

**\*\*\*15 - 10 pts)** Referring to question 13, what's the smallest N where the true coverage rate of the 95% asymptotic Normal CI for mean A1c, rounded to two decimal places, is 0.95? Remember for part 1, describing how you would solve the problem is a good answer. For parts 2 and 3, an exact solution is needed.

**\*\*\*16 - 10 pts)** Referring to question 13, suppose you wanted to take a random sample of N=10 from this population with diabetes and perform a two-sided hypothesis at a 5% significance level. You consider two options: (1) assume the observed sample sd is an excellent estimate of the true sd and perform a one-sample Z test, and (2) account for the fact that you're using the sample sd and perform a one-sample t-test. Calculate and discuss what the impact is of using the t-test over the Z-test in terms of Type I and Type II error. For Type II error, consider the case of testing that the true mean is 6 mmHg.

**\*\*\*17 - 10 pts)** Referring to question 16, answer the same question assuming A1c had the same mean and standard deviation as in question 16, but was Normally distributed.