1) If two events, A and B, are *statistically independent*, is there any sense in using observations of B to forecast or predict A? Why or why not? Use conditional probabilities to support your answer.
When A and B are independent, we have that $P[A|B] = P[A]$, and the conditional probability is equal to the unconditional. Thus, it is pointless to try and use B to help forecast A.

2) Suppose you flip two coins and roll a fair die. Every "random experiment" here is independent of the others. What is the probability that the number of heads you get is greater than or equal to the number from the dice roll?
Let N_H be the number of heads and D be the dice roll. Now:

$$\begin{aligned} P[N_H \geq D] &= P[N_H = 0] \cdot P[D \leq 0] + P[N_H = 1] \cdot P[D \leq 1] + P[N_H = 2] \cdot P[D \leq 2] \\ &= \frac{1}{4} \times 0 + \frac{1}{2} \times \frac{1}{6} + \frac{1}{4} \times \frac{2}{6} \\ &= \frac{1}{6} \end{aligned}$$

3) X can take the values $\{1, 2, 3\}$ with probabilities $\{0.6, 0.2, 0.2\}$. Compute $\mathbb{E}[X]$ and $\mathbb{V}[X]$.

$$\mathbb{E}[X] = 1 \times 0.6 + 2 \times 0.2 + 3 \times 0.2 = 1.6$$

$$\mathbb{V}[X] = 1 \times 0.6 + 4 \times 0.2 + 9 \times 0.2 - 1.6^2 = 3.2 - 1.6^2 = 0.64$$

4) Your friends, Rihanna and Beyonce, take turns giving you gifts for your birthday. You enjoy 30% of the gifts Rihanna gives you, and 50% of Beyonce's gifts. This year, they flip a fair coin to decide who will buy you two presents. Let $B \in \{0, 1\}$ indicate whether Beyonce bought your presents this year, and let Y be the number of gifts you enjoy.

1. Write down the joint distribution (as a table or a formula, whichever you prefer) of Y and B
When $B = 0$, Y is distributed as a binomial with $p = 0.3$, $N = 2$. When $B = 1$, Y is binomial with $p = 0.5$,$N = 2$. We can write the joint probabilities in the following table:

	Y=0	Y=1	Y=2
B = 0	0.49×0.5	0.42×0.5	0.09×0.5
B = 1	0.25×0.5	0.5×0.5	0.25×0.5

$=$

	Y=0	Y=1	Y=2
B = 0	0.245	0.21	0.045
B = 1	0.125	0.25	0.125

2. Calculate $\mathbb{E}[Y|B = 0]$, $\mathbb{E}[Y|B = 1]$, and $\mathbb{E}[Y]$. The conditional distributions are both binomials. Recall that the mean of a binomial is Np, so we get $\mathbb{E}[Y|B = 0] = 0.6$ and $\mathbb{E}[Y|B = 1] = 1$. Further, you can either note that the *marginal distribution* of Y is also binomial, with $p = 0.5 \times 0.3 + 0.5 \times 0.5 = 0.4$ which gives $\mathbb{E}[Y] = 0.8$, OR you can use the formula $\mathbb{E}[Y] = \mathbb{E}[Y|B =$

$0]P[B=0] + \mathbb{E}[Y|B=1]P[B=1]$ to get the same answer. OR, you can compute the marginal $P(y)$ and calculate the mean directly using the formula for expectation:

$$\begin{aligned}\mathbb{E}[Y] &= 0 \times (P(0,0) + P(1,0)) + 1 \times (P(0,1) + P(1,1)) + 2 \times (P(0,2) + P(1,2)) \\ &= 0 + 0.21 + 0.25 + 2 \times (0.045 + 0.125) \\ &= 0.46 + 0.34 = 0.8\end{aligned}$$

3. What is the distribution of the number of gifts you enjoy, *given* that Beyonce gave you the gifts ($B = 1$). Given that $B = 1$, Y is binomial with $p = 0.5$, $N = 2$.

4. You enjoyed both gifts this year, would you reject the Null Hypothesis that Rihanna chose the gifts when conducting a test of size 10%. What do you notice about this case that is different from the usual set up of Hypothesis Testing? The probability that you enjoy both gifts from Rihanna is 0.09. If my decision rule was to reject the Null whenever I enjoy both gifts, then I would only commit Type I error with probability 0.09. Since this is less than 0.1, such a test test satisfies my requirements on size, and I would reject the Null in this case. Usually, hypothesis testing is conducted without any notion of probabilities over the truthfulness of the hypotheses. Here, the initial probability of the null being true is 0.5 (coming from the coin flip), so it's a bit weird to conduct a hypothesis test in this setting.

5) Define the following jointly distributed random variables X and Y according to the joint distribution:

$$P(x,y) = \begin{cases} 1/3 & \text{for } (x,y) = (-1,1) \\ 1/3 & \text{for } (x,y) = (0,0) \\ 1/3 & \text{for } (x,y) = (1,1) \end{cases}$$

Part (a): show that $\mathbb{C}(X,Y) = 0$. Part (b): show that X and Y are not independent.
Part (a): First we work out that $\mathbb{E}[X] = 0$ and $\mathbb{E}[Y] = 2/3$. Then, calculate the covariance:

$$\mathbb{C}(X,Y) = (-1-0) \times (1-2/3) \times 1/3 + 0 \times (0-2/3) \times 1/3 + (1-0) \times (1-2/3) \times 1/3 = -1/9 + 0 + 1/9 = 0$$

Part (b): notice that $P[Y=0] = P[X=0] = 1/3$, so $P(0,0) = 1/3 \neq P[X=0] \times P[Y=0]$.

6) Beyonce runs a Lemonade stand in Houston, where her revenue every hour (in dollars) is distributed as a normal with mean 20 and variance 16.

- What is the probability that her revenue is greater than \$18?

$$P[R \geq 18] = P[Z \geq (18-20)/4] = P[Z \geq -0.5] = P[Z \leq 0.5] \approx 0.691$$

- What is the probability that her revenue is between \$19 and \$24?

$$P[19 \leq R \leq 24] = P[-0.25 \leq Z \leq 1] = P[Z \leq 1] - P[Z \leq -0.25] = 0.841 - (1 - 0.598) = 0.439$$

- If she sells glasses at 50c per glass, what is her expected number of glass sold in an hour? Remember that $revenue = price \times items\ sold$. Is the number of glasses sold in an hour also normal? Explain. Let X be no. of items sold. We have $R = 0.5 \times X$ so $X = 2R$. Since R is normal, we know that X is normal also with $\mathbb{E}[X] = 2\mathbb{E}[R] = 40$.

CDF values for a standard normal, Z:

z	.0	0.25	0.5	0.75	1.0
$F(z)$	0.5	0.598	0.691	0.773	0.841

7) Let some data $X_1, X_2, ..., X_N$ be drawn from a normal distribution with mean μ_X and variance σ_X^2.

1. How is the sample mean, $\overline{X}_N$, distributed?
   See your notes

2. Define $S = X_1 + X_2$. Is S a statistic? How is S distributed?
   In class we saw that the sum of two normals is normal. Here, $\mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = 2\mu_X$, and since the observations are independent of each other, $\mathbb{V}[X_1 + X_2] = \mathbb{V}[X_1] + \mathbb{V}[X_2] = 2\sigma_X^2$. Thus, S is normal with mean $2\mu_X$ and variance $2\sigma_X^2$.

3. Define $U = X_1 + X_2 + X_3$. Is U a statistic? How is U distributed?
   Similar to above, we have that U is normal with mean $3\mu_X$ and variance $3\sigma_X^2$.

4. Suppose that the observations, X_i, are not drawn from a normal distribution. What idea can we use to approximate the distribution of $\overline{X}_N$? What will this approximation be? We can use the central limit theorem and use a normal with mean μ_X and variance σ_X^2/N to approximate the distribution of $\overline{X}_N$.

8) Let each observation of some dataset, $X_1, X_2, ..., X_N$, be drawn from a uniform distribution with lower bound 10 and upper bound 20.

1. For each observation X_i, what is $\mathbb{E}[X_i]$? What is $\mathbb{V}[X_i]$? You may refer to your lecture notes for the right formulae.
   $\mathbb{E}[X_i] = (a+b)/2 = (10+20)/2 = 15$ and $\mathbb{V} = (b-a)^2/12 = (20-10)^2/12 = 100/12 (\approx 8.33)$

2. Based on your answer above, what is $\mathbb{E}[\overline{X}_N]$? What is $\mathbb{V}[\overline{X}_N]$?
   We have that $\mathbb{E}[\overline{X}_N] = \mathbb{E}[X_i] = 15$, and $\mathbb{V}[\overline{X}_N] = \mathbb{V}[X_i]/N \approx 8.33/N$

3. If N was from a large sample, how could you approximate the distribution of $\overline{X}_N$?
   $\overline{X}_N$ would be approximately normal with mean and variance given by the previous question

9) Let $X_1, X_2, ..., X_N$ be data drawn from a normal distribution with mean 10 and standard deviation 4.

1. Define $T = X_1 + X_2 + X_3$. Is T a statistic? What is the distribution of T?
   T is a function of the data, so it is a statistic. Since the sum of two normal random variables is a normal random variable, and so it goes for three, four, etc, T is a normal random variable with mean $\mu_T = 3\mu_X = 30$ and variance $\sigma_T^2 = 3\sigma_X^2 = 3 \times 4^2 = 48$.

2. Write the lower and upper bounds of a 95% acceptance interval for T, using that $P[Z \leq -1.96] = 0.025$, where Z is a standard normal (i.e. $z_{.025} = 1.96$).
   The acceptance interval is:

$$\mu_T \pm \sigma_T \times 1.96 = [30 - \sqrt{48} \times 1.96, 30 + \sqrt{48} \times 1.96] = [16.42, 43.58]$$

10) (hard) Let $X_1, X_2, ..., X_8$ be data drawn from a normal distribution with mean 5 and standard deviation 1. Define the following:

$$T_8 = X_1 - X_2 + X_3 - X_4 + X_5 - X_6 + X_7 - X_8$$

1. Is T_8 a statistic? What is its sampling distribution?
   T_8 is a function of the data and therefore a statistic. First, note that if X is normal with mean μ_X and standard deviation σ_X, then $-X$ is also normal with mean $-\mu_X$ and standard deviation σ_X. Separating T_8 into added and subtracted variables, gives:

   $$T_8 = \underbrace{X_1 + X_3 + X_5 + X_7}_{=Y_1} - \underbrace{(X_2 + X_4 + X_6 + X_8)}_{=Y_2} = Y_1 - Y_2$$

   Using the same logic as we used in Question 1, it must be that Y_1 is normal with mean 20 and variance 4, and Y_2 is also normal with mean 20 and variance 4. To finish then, T_8 must itself be normal with mean 0 and variance 8.

2. Following the above logic, for a sample of size N, where N is any even number, define:

   $$T_N = \sum_{i \text{ is odd}} X_i - \sum_{i \text{ is even}} X_i = X_1 - X_2 + X_3 - ...X_{N-1} - X_N$$

   What is the sampling distribution of T_N?
   If N is even, there are $N/2$ odd variables and $N/2$ even variables. Following the logic of above and of question 1, let us say that:

   $$T_N = \sum_{i \text{ is odd}} X_i - \sum_{i \text{ is even}} X_i = Y_{\text{odd}} - Y_{\text{even}}$$

   As before, the sum of the odd observations, Y_{odd} must be normal with mean $5 \times N/2$ and variance $N/2$. The sum of the even observations, Y_{even} must be normal with mean $5 \times N/2$ and variance $N/2$. Therefore $T_N = Y_{\text{odd}} - Y_{\text{even}}$ must also be normal with mean 0 and variance N.

11) The length of Taylor Swift's grudges follows an exponential distribution, and her average grudge lasts 5 years.

1. Exactly 3 years ago, at a *Taylor's Version* listening party, she overhears you making fun of her cats. What is the probability that she still holds a grudge against you?
   Let T be the length of her grudge, which is an exponential with $\lambda = 1/5$. Thus $P[T \geq 3] = \exp(-1/5 \times 3) \approx 0.548$

2. You are going to collect a dataset consisting of each of her 1000 grudges, and the length of time each grudge lasts. Given this large sample size, what distribution will the sample mean from this dataset approximately follow? (Hint: the variance of the exponential is $\frac{1}{\lambda^2}$).
   Let $\overline{T}_{1000}$ be the sample mean. Since $\mathbb{E}[T] = 5$, we know that $\mathbb{E}[\overline{T}_{1000}] = 5$. Similary, since $\mathbb{V}[T] = 1/\lambda^2 = 25$, we know that $\mathbb{V}[\overline{T}_{1000}] = \mathbb{V}[T]/N = 25/1000 = 0.025$. Finally, the central limit theorem tells us that the distribution will be approximately normal, so $\overline{T} \sim \mathcal{N}(5, 0.025)$ approximately.

12) You are asked to estimate the population mean and variance (μ, σ^2) of workers' wages given a random sample of wages, $W_1, W_2, ..., W_N$, from the population. You have $N = 50$ observations. Suppose you get that $\overline{W} = \$10/hour$, with $s^2 = 9$.

1. Write a 95% confidence interval for μ that can be computed from this information (one of the critical values below).
   Though $N = 50$ seems small, let's go ahead and use the central limit theorem. The confidence interval is:
   $$\overline{X} \pm z_{0.025} \times s/\sqrt{50} = 10 \pm 1.96 \times 3/\sqrt{50} = [9.17, 10.83]$$

2. Test the null hypothesis that $\mu = 12$ against the alternative that it is smaller than 12. Use size $\alpha = 0.05$, and use one of: $z_{0.025} = 1.96$,$z_{0.05} = 1.64$.
   This is a one-sided test. We reject the null here if
   $$\frac{\overline{X} - 12}{\sqrt{s^2/50}} < -z_{0.05}$$
   . We get a value ≈ -4.71 so we reject the null.

13) You are asked to estimate the unemployment rate, using a sample of $N = 100$ workers, who report their employment status (either employed or unemployed).

1. Describe how you would use this data to estimate the unemployment rate. Describe how you would compute a $(1 - \alpha) \times 100\%$ confidence interval.
   Let $U_i \in \{0, 1\}$ indicate whether the worker is unemployed. Note that the population unemployment rate is $\mathbb{E}[U_i]$. The sample mean is simply the proportion of workers who report being unemployed, $\hat{p}$. If $\mathbb{E}[U_i] = p$, then $\mathbb{V}[U_i] = p(1 - p)]$ and the sample variance will turn out to be $\hat{p}(1 - \hat{p})$. Now we can use the confidence interval formula:
   $$\hat{p} \pm z_{\alpha/2} \times \sqrt{\hat{p}(1 - \hat{p})/N}$$

2. You estimate a sample proportion of unemployed workers of 7.6%. Give the 95% confidence interval for the true unemployment rate.
   Since $z_{0.025} = 1.96$, we use the formula to get a confidence interval $[0.024, 0.128]$.

14) Suppose you have data on the annual earnings of $N = 500$ households, $E_1, E_2, ..., E_{500}$. This is a random sample from the population. The policy group you work for deems a household to be living in poverty if their earnings fall below the poverty line, \$10,000/year. Describe how you would use this data to estimate the poverty rate (i.e. the population proportion of households living in poverty). Give an estimate as well as a formula for the $(1 - \alpha) \times 100\%$ confidence interval.
Define p to be the population poverty rate, and let $\hat{p}$, the sample proportion, be our proposed estimator. Define $Y_i = 1$ if $E_i \leq 10,000$ and 0 otherwise. Our estimate of the poverty rate is the sample proportion of households earnings less than 10,000, $\hat{p}$. If you want to use the data Y, we know that $\hat{p} = \overline{Y}$. The confidence interval is given by:
$$\hat{p} \pm z_{\alpha/2} \times \sqrt{\hat{p}(1 - \hat{p})/N}$$

15) Calculate the bias and the variance of the following estimators.

1. $\theta = \overline{X} + a$ where a is a constant, as an estimator for μ. $\mathbb{E}[\theta] = \mathbb{E}[\overline{X} + a] = \mu + a$ so the bias is a. We can also calculate

$$\mathbb{E}[(\theta - \mu - a)^2] = \mathbb{E}[(\overline{X} - \mu)^2] = \mathbb{V}(\overline{X}) = \sigma^2/N.$$

   The variance is σ^2/N.

2. $\theta = \overline{X} + \epsilon$ where ϵ is independent of $\overline{X}$ and $\mathbb{E}[\epsilon] = 0$, as an estimator for μ. $\mathbb{E}[\theta] = \mathbb{E}[\overline{X}] + \mathbb{E}[\epsilon] = \mu$, so the estimator is unbiased. Also:

$$\mathbb{E}[(\theta - \mu)^2] = \mathbb{E}[(\overline{X} + \epsilon - \mu)^2] = \sigma^2/N + \mathbb{V}(\epsilon)$$

16) You have data from a random sample of firms in the market for widgets. In particular, you have output Y_i, and costs C_i for each firm i. Your data contains this information for a random sample of N firms. You know that every firm sells their widgets at the market price of \$2, and so profit for each firm can be defined as:

$$\Pi_i = 2Y_i - C_i$$

Assume that Y_i and C_i are each *normally distributed*, with *unknown* variance.

1. What kind of distribution will Π have? You can use the parameters $\mathbb{E}[\Pi] = \mu_\pi$ and $\mathbb{V}[\Pi] = \sigma_\pi^2$ to describe this distribution.
   Π will be distributed normally with mean μ_π and variance σ_π^2.

2. How would you estimate a 95% confidence interval for μ_π?
   Since we don't know σ_π^2, we can calculate the sample variance s_π^2 and use the t-distribution:

$$\overline{\Pi} \pm t_{N-1,\alpha/2} s_\pi / \sqrt{N}$$

3. Do you need to assume that Y_i and C_i are independent at a given firm for this to work?
   Nope! As long as Π is normal, this method will work. Of course, we still need observations of the data *across* firms to be independent, otherwise our sampling distributions won't be correct.

17) Two sample means $\overline{X}_1$ and $\overline{X}_2$ are calculated from random samples from the *same* population. The only difference is the sample size $N_1 < N_2$. Which is the more efficient estimator (i.e. which estimator has smaller variance) of the population mean, μ? Why?
We will get variances of the sample mean equal to $\sigma^2/N_1 > \sigma^2/N_2$ so $\overline{X}_2$ is more efficient. Greater sample size gives more precision.

18) Two sample means $\overline{X}_1$ and $\overline{X}_2$ are calculated from random samples from *different* populations with the *same* population mean, μ. The population variances are different, though, with $\sigma_1^2 < \sigma_2^2$. Which sample mean is a more efficient estimator for μ? Why?
We will get variances of the sample mean equal to $\sigma_1^2/N < \sigma_2^2/N$ so $\overline{X}_1$ is more efficient.

19) Let X and Y both be *independent* random variables. Both are *normally distributed* with means μ_X, μ_Y and variances σ_X^2, σ_Y^2. Let $\overline{X}$ and $\overline{Y}$ be *sample means* of these random variables, from samples of size N_X and N_Y. Your task is to write down the *sampling distribution* of the following statistics.

1. $\overline{X}$ and $\overline{Y}$
   $\overline{X} \sim \mathcal{N}(\mu_X, \sigma^2/N_X)$, $\overline{Y} \sim \mathcal{N}(\mu_Y, \sigma^2/N_Y)$

2. $T = X + a$ where a is a constant.
   $\mathbb{E}[T] = \mu_X + a$, $\mathbb{V}[T] = \sigma_X^2$, $T \sim \mathcal{N}(\mu_X + a, \sigma_X^2)$ (since a is constant)

3. $T = \overline{X} + Y$
   T will be normal with mean $\mathbb{E}[T] = \mathbb{E}[\overline{X} + Y] = \mu_X + \mu_Y$, and variance $\mathbb{V}[T] = \mathbb{V}[\overline{X} + Y] = \sigma_X^2/N_X + \sigma_Y^2$

4. $T = \overline{Y} + X$
   T will be normal with mean $\mathbb{E}[T] = \mathbb{E}[\overline{Y} + X] = \mu_Y + \mu_X$, and variance $\mathbb{V}[T] = \mathbb{V}[\overline{Y} + X] = \sigma_Y^2/N_Y + \sigma_X^2$

5. $T = aX + bY$ where a and b are constants.
   T will be normal with mean $\mathbb{E}[T] = \mathbb{E}[aX + bY] = a\mu_X + b\mu_Y$, and variance $\mathbb{V}[T] = \mathbb{V}[aX + bY] = a^2\sigma_X^2 + b^2\sigma_Y^2$

6. $T = \overline{X} - \overline{Y}$
   T will be normal with mean $\mathbb{E}[T] = \mathbb{E}[\overline{X} - \overline{Y}] = \mu_X - \mu_Y$, and variance $\mathbb{V}[T] = \mathbb{V}[aX + bY] = \sigma_X^2/N_X + \sigma_Y^2/N_Y$

20) Suppose we are interested in estimating the mean wage, μ, for a population. We collect a random sample of wages, $W_1, W_2, ..., W_n$ from this population. **However**, suppose that people who earn less than \$8/hour do not report their earnings, and are dropped from the sample.

1. What is the *effective* probability distribution we are drawing from, if we cannot observe wages of $w < 8$? (Hint: the answer involves a conditional distribution).
   Since we are discarding all the observations $W < 8$, an otherwise random sample will be drawn from the distribution:

   $$P[W \mid W \geq 8]$$

   Similarly, we can say that this distribution has a density $f(w|w \geq 8)$.

2. Given this new distribution, what is the expected value of a single (non-missing) observation W taken from the population?
   We can compute the expectation as:

   $$\mathbb{E}[W \mid W \geq 8] = \int w f(w|w > 8) dw$$

   Since we know that obersvations are now drawn from this truncated distribution.

3. What is the expected value of $\overline{W}$, the sample mean from this truncated sample?
   Our logic is identical to when we derived the expectation of the sample mean from the proper random sample:

   $$\mathbb{E}[\overline{W}| \underbrace{W_1, W_2, ..., W_N > 8}_{\text{truncated sample}}] = N \times \mathbb{E}[W|\ W > 8]/N = \mathbb{E}[W|W > 8]$$

4. Consider the following rule: $\mathbb{E}[X|X > c] > \mathbb{E}[X]$ for any constant c. What does this imply for $\overline{W}$ from this sample as an estimator for μ?
If this holds, then we know that $\mathbb{E}[\overline{W} \mid \text{truncated sample}] = \mathbb{E}[W \mid W > 8] > \mathbb{E}[W]$. Thus, the sample mean from the truncated sample is a *biased* estimator of the population mean of wages.

21) Suppose that unemployment durations (in weeks) are distributed according to an exponential (that is, the wait times until the arrival of a job offer are exponentially distributed) with a rate parameter λ. Let T be the unemployment duration of one individual. Recall that:

$$\mathbb{E}[T] = \frac{1}{\lambda}, \ \mathbb{V}[T] = \frac{1}{\lambda^2}$$

1. Let $\overline{T}$ be the sample mean from a random sample of the population of unemployed workers. Calculate $\mathbb{E}[\overline{T}]$ and $\mathbb{V}[\overline{T}]$.

   We know that:
   $$\mathbb{E}[\overline{T}] = N \times \mathbb{E}[T]/N = \frac{1}{\lambda}$$
   And
   $$\mathbb{V}[\overline{T}] = N \times \mathbb{V}[T/N] = N \times \mathbb{V}[T]/N^2 = \frac{1}{\lambda^2 N}$$

2. What does the central limit theorem say about $\overline{T}$ as our sample size grows?
   The CLT says that
   $$\overline{T} \sim \mathcal{N}(\mathbb{E}[T], \mathbb{V}[T]/N) = \mathcal{N}\left(\frac{1}{\lambda}, \frac{1}{\lambda^2 N}\right)$$

3. Propose an estimator for λ based on your answer to part (a).
   Since $\overline{T}$ converges around $\frac{1}{\lambda}$ with smaller and smaller variance, I propose:
   $$\hat{\lambda} = \frac{1}{\overline{T}}$$
   as an estimator for λ. Remember that $\overline{T}$ is the sample mean of all the unemployment durations we observe in the data.

4. Combine parts (2) and (3) to propose an approximate 95% confidence interval for λ.
   We know, by the CLT, that:
   $$\frac{\overline{T} - \frac{1}{\lambda}}{1/(\lambda\sqrt{N})} \sim \mathcal{N}(0,1)$$
   So this means that:
   $$P\left[-z_{\alpha/2} < \frac{\overline{T} - \frac{1}{\lambda}}{1/(\lambda\sqrt{N})} < z_{\alpha/2}\right] = 1 - \alpha$$
   We can rearrange this to:
   $$P[-z_{\alpha/2} < \sqrt{N}\lambda\overline{T} - \sqrt{N} < z_{\alpha/2}] = (1 - \alpha)$$
   And from there get to:
   $$P\left[\frac{1}{\overline{T}} - z_{\alpha/2}\frac{1}{\sqrt{N\overline{T}}} < \lambda < \frac{1}{\overline{T}} + z_{\alpha/2}\frac{1}{\sqrt{N\overline{T}}}\right] = P\left[\hat{\lambda} - z_{\alpha/2}\frac{1}{\sqrt{N\overline{T}}} < \lambda < \hat{\lambda} + z_{\alpha/2}\frac{1}{\sqrt{N\overline{T}}}\right] = 1 - \alpha$$
   Notice the similarity to previous confidence intervals. This one is also centered around our estimator, $\hat{\lambda}$.

22) Let p be the true proportion of UMN students who like pineapple on pizza. Let $\overline{p}$ be the fraction of students from a random, iid, sample of size N that report that they like pineapple on pizza.

1. What is the approximate sampling distribution of $\overline{p}$? First, Let $X_i \in \{0, 1\}$ be the data for person i, equal to 1 if they report liking pineapple on pizza. The fraction $\overline{p}$ is the sample mean of X. Since X_i satisfies $\mathbb{E}[X_i] = p$ and $\mathbb{V}[X] = p(1-p)$, we have that $\hat{p} \sim \mathcal{N}(p, p(1-p)/N)$ approximately by the central limit theorem.

2. Construct a two-sided hypothesis test (with size α) of the null hypothesis that $p = p_0$, where p is the unknown population proportion. Accept the null only if:

$$-z_{\alpha/2} < \frac{\hat{p} - p_0}{\sqrt{\hat{p}/(1-\hat{p})}} < z_{\alpha/2} \qquad \text{OR} \qquad p_0 - z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/N} < \hat{p} < p_0 + z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/N}$$

(these rules are the same) and reject otherwise.

23) Suppose that you have the data on some health measure, Y, for a group of treated individuals and a group of untreated individuals. Let $\overline{Y}_T$ be the sample mean of the treated, and let $\overline{Y}_N$ be the sample mean of the untreated. The *effectiveness* of the drug in the population can be written as $\mu = \mathbb{E}[Y_T] - \mathbb{E}[Y_N]$. When the population variance across the two populations is *equal, but unknown*, describe how you would construct a hypothesis test (with significance α) of the following:

1. The null that the drug has *no effect* agains the alternative that it has *some effect*
$H_0 : \mu = 0$, $H_1 : \mu \neq 0$. Let s^2 be the pooled variance from the data. We do this because we are assuming the variances are equal. If $\mu = 0$ then we know that:

$$P[-z_{\alpha/2} < \frac{\overline{Y}_T - \overline{Y}_N}{\sqrt{s^2/N_T + s^2/N_N}} < z_{\alpha/2}] = 1 - \alpha$$

So we reject the null if $\frac{\overline{Y}_T - \overline{Y}_N}{\sqrt{s^2/N_T + s^2/N_N}}$ lies outside the critical values. Equivalently we know

$$P[-z_{\alpha/2}\sqrt{s^2/N_T + s^2/N_N} < \overline{Y}_T - \overline{Y}_N < z_{\alpha/2}\sqrt{s^2/N_T + s^2/N_N}] = 1 - \alpha$$

so we reject the null if the difference in sample means lies outside the above critical values.

2. The null that the drug has *no effect* agains the alternative that it has a *positive effect* $H_0 : \mu = 0$, $H_1 : \mu > 0$. We reject the null if:

$$z_\alpha < \frac{\overline{Y}_T - \overline{Y}_N}{\sqrt{s^2/N_T + s^2/N_N}} \qquad \text{OR} \qquad z_\alpha\sqrt{s^2/N_T + s^2/N_N} < \overline{Y}_T - \overline{Y}_N$$

24) Suppose that a statistic (it could be anything) is distributed such that:

$$P[-\frac{a_\alpha}{\sqrt{N}} < T - \gamma < \frac{a_\alpha}{\sqrt{N}}] = 1 - \alpha$$

Where γ is a population parameter of interest. Use this fact to:

- Write a hypothesis test, with Type I error probability $1-\alpha$, of the Null hypothesis that $\gamma = 0$ against the alternative $\gamma \neq 0$
  If $\gamma = 0$,
  $$P[-\frac{a_\alpha}{\sqrt{N}} < T < \frac{a_\alpha}{\sqrt{N}}] = 1-\alpha$$
  So reject the null if we observe a value of T outside of this range.

- Write a hypothesis test, with Type I error probability $1-\alpha$, of the Null hypothesis that $\gamma = 2$ against the alternative $\gamma \neq 2$ Similarly to before, we accept the null only if:
  $$2 - \frac{a_\alpha}{\sqrt{N}} < T < 2 + \frac{a_\alpha}{\sqrt{N}}$$

- Calculate the power of your first test above to correctly reject the Null when $\gamma = 2$. Write your answer in terms of the CDF of T (i.e. in terms of a function F such that $P[T \leq t] = F(t)$) We accept the null (and hence commit type II error) if we see:
  $$-\frac{a_\alpha}{\sqrt{N}} < T < \frac{a_\alpha}{\sqrt{N}}$$
  This occurs with probability $\beta = F(a_\alpha/\sqrt{N}) - F(-a_\alpha/\sqrt{N})$ so we get
  $$\text{Power} = 1 - F(a_\alpha/\sqrt{N}) + F(-a_\alpha/\sqrt{N})$$

- What happens to the power of your test as N increases? Why?
  As $\sqrt{N}$ gets larger and larger, the critical values $(a_\alpha/\sqrt{N})$ of the test get smaller and smaller, and there is less chance that T lies between them. Thus, the probability of type II error decreases and power increases.