**Problem 1.** Suppose that $X_1, \ldots, X_n$ are an i.i.d. random sample of size $n$ with sample mean $\bar{X} = 12$ and sample variance $s^2 = 5$.

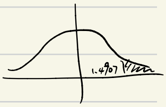
(a) Let $n = 5$ and suppose the samples are drawn from a Normal distribution with unknown mean $\mu$ and known variance $\sigma^2 = 9$. Let the null hypothesis be $H_0 : \mu = 10$ and the alternative hypothesis be $H_a : \mu \neq 10$. Calculate the relevant test statistic value and $p$-value. Determine the decision rule for $\alpha = 0.05$.

(b) Using the acceptance region of this test, construct a 95% confidence interval. (Hint: Think about how we can reverse our hypothesis test to construct the equivalent 95% confidence interval)

(a). $X_1, \cdots X_5 \sim N(\mu, \sigma^2)$ where $\sigma^2 = 9$

Assuming that $H_0: \mu = 10$ is true, the distribution of the outcome of $T$ is

$$T(X_1, \ldots, X_5) = \frac{\mu - 10}{\sigma/\sqrt{n}} = \frac{\mu - 10}{3/\sqrt{5}}$$




If $|T(X_1, \cdots X_5)| < Z_{\alpha/2}$, accept $H_0$

test statistic $= \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{12-10}{3/\sqrt{5}} = 1.4907$, $\alpha = 0.05$

p-value $= 2(1 - \varphi(t_{data})) = 0.135 > \alpha$, $\therefore$ accept $H_0$

(b). If $|T(X_1, \cdots, X_5)| < Z_{\alpha/2}$, accept $H_0$ $\quad Z_{0.025} = 1.96$

$$\left|\frac{\mu - 12}{\sigma/\sqrt{n}}\right| < Z_{\alpha/2} \Rightarrow |\mu - 12| < Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 1.96 \times \frac{3}{\sqrt{5}} \qquad \sigma^2 = 5$$

$\therefore$ the 95% CI is $(9.3704, 14.6296)$

**Problem 2.** Federal investigators identified a strong association between chemicals in drywall and electrical problems, and there is also strong evidence of respiratory difficulties due to emission of hydrogen sulfide gas. An extensive examination of 51 homes found that 41 had such problems. Suppose that 51 were randomly sampled from the population of all homes having drywall.

(a) Does the data provide strong evidence for concluding that more than 50% of all homes with drywall have electrical/environmental problems? Carry out a test of hypotheses using $\alpha = 0.01$.

$H_0: p > 0.5 \quad H_a: p \leq 0.5$

given that $\hat{p} = \frac{41}{51}$, the test statistic is $\frac{\hat{p} - p_0}{\sigma_0/\sqrt{n}} = \frac{\frac{41}{51} - 0.5}{\sqrt{\frac{0.5 \times 0.5}{51}}} \approx 4.34$

p-value $= 7 \times 10^{-6} < \alpha$, the corresponding p-value is less than 0.01 thus reject $H_0$ one-sided

the data does not provide strong evidence to conclude given statement

(b) Calculate a lower bound of a 99% confidence interval for the percentage of all such homes that have electrical/environmental problems.

$\hat{p} = \frac{41}{51}$, $Var(X) = p(1-p) = 0.25$

$Pr\left\{ \frac{-\beta}{6/\sqrt{n}} \le \frac{p-\hat{p}}{6/\sqrt{n}} \le \frac{\beta}{6/\sqrt{n}} \right\} = 1-\alpha$, $\alpha = 0.01$, $n=51$ $\quad Z_{0.01} = 2.33$

$\beta = Z_{\alpha/2} \frac{6}{\sqrt{n}} \approx 0.1631$

$\therefore$ 99% CI is $[0.6408, 0.9671]$

$\therefore$ lower bound is $0.6408$

**Problem 3.** Consider a random sample of size $n = 100$ with sample proportion $\hat{p} = 0.2$ from a population with a true unknown proportion $p$.

(a) For the test $H_0 : p = 0.25$ versus $H_a : p < 0.25$, calculate the relevant test statistic value and $p$-value. Determine the decision rule for $\alpha = 0.05$ and $\alpha = 0.01$.

$Z_{0.05} = 1.645$

$Z_{0.01} = 2.326$

$n=100$, $\hat{p} = 0.2$, $6^2 = \hat{p}(1-\hat{p}) = 0.16$

$\therefore$ test statistic value $= \frac{\hat{p} - p_0}{\sqrt{\frac{6^2}{n}}} = \frac{0.2-0.25}{\sqrt{(0.25 \times 0.75)/100}} = -1.15$,

p-value $\approx 0.124 > \alpha$, if p-value (data) $> \alpha$, accept $H_0$

for $\alpha = 0.05$ and $\alpha = 0.01$, we accept hypothesis $H_0$

(b) For the test $H_0 : p = 0.25$ versus $H_a : p \neq 0.25$, calculate the relevant test statistic value and $p$-value. Determine the decision rule for $\alpha = 0.05$ and $\alpha = 0.01$.

$T(X_1, \cdots, X_n) = -1.15 \Leftarrow$ test statistic value.

$|T(X_1, \cdots X_n)| \underset{\text{accept } H_0}{\overset{\text{reject } H_0}{\gtrless}} Z_{\alpha/2}$

$\alpha = 0.05$, $Z_{\alpha/2} = Z_{0.025} = 1.96$

$\alpha = 0.01$, $Z_{\alpha/2} = Z_{0.005} = 2.576$

p-value $= 2(1 - \Phi_{t_{data}}) = 0.2485 > \alpha$

$\therefore$ we accept $H_0$ for $\alpha = 0.05$ and $\alpha = 0.01$

$\alpha = Pr\{\text{type I error}\} = Pr\{\text{reject } H_0 \mid H_0 \text{ is true}\}$

$\beta = Pr\{\text{type II error}\} = Pr\{\text{accept } H_0 \mid H_0 \text{ is false}\}$

power $= \{\text{reject } H_0 \mid H_0 \text{ is false}\}$

**Problem 4.** Recall that $\alpha$ represents the probability of a type I error. On the other hand, $\beta$ represents the probability of a type II error. The power of a test is the probability that the null hypothesis is rejected when it is false, and is therefore defined as $1 - \beta$. For this problem, you will explore how power depends on multiple factors.

power $\uparrow$

type I error $\downarrow$

(a) Suppose that $X_1, \ldots, X_n$ are an i.i.d. random sample of size $n = 15$ drawn from a Normal distribution with unknown mean $\mu$ and known variance $\sigma^2 = 4$. Using Python, plot the power of the test $H_0 : \mu = 0$ versus $H_a : \mu \neq 0$ as the true mean $\mu$ varies at level $\alpha = 0.05$. Make sure to display your graph clearly.

(b) Now using the same information from part (a) except for fixing $\mu$ at $\mu = 3$, plot the power of the test as $\alpha$ varies.

(c) Fix $\alpha$ back to $\alpha = 0.05$ and plot the power of the test as $n$ varies.

(d) Fix $n$ back to $n = 15$ and plot the power of the test as $\sigma^2$ varies.

(e) Compare and interpret your results.

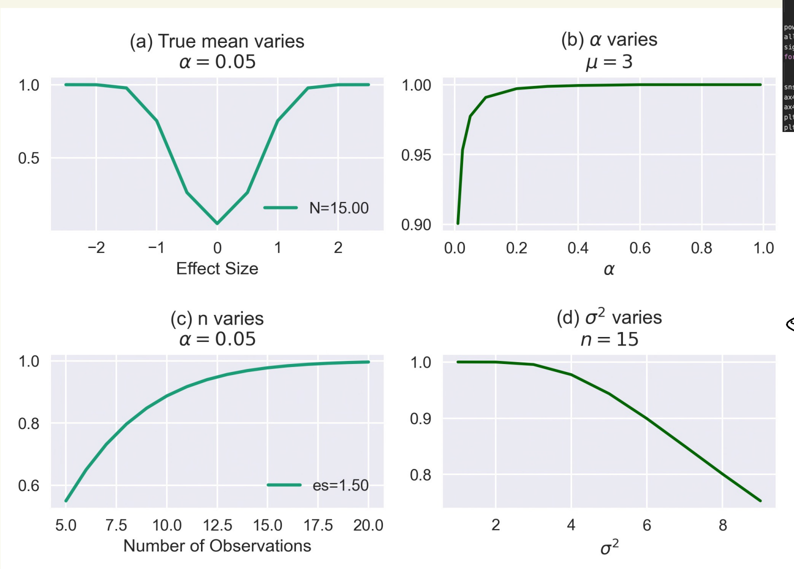
```python
fig, ((ax1, ax2), (ax3, ax4))= plt.subplots(2,2,figsize = (7,5))
true_mean = np.array([(i-5) for i in range(11)])
TTestIndPower().plot_power(dep_var='effect_size',
                           nobs= np.array([15]),
                           effect_size=(true_mean-0)/2,
                           alpha=0.05,
                           ax=ax1, title='(a) True mean varies' + '\n' + r'$\alpha = 0.05$')

power_analysis = TTestIndPower()
all_p=[]
alpha = [0.01,0.025,0.05,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,0.95,0.99]
for a in alpha:
    p = power_analysis.power(effect_size=1.5,nobs1=15, alpha = a, ratio = 1)
    all_p.append(p)

sns.lineplot(x = alpha, y=all_p, ax = ax2, color = 'darkgreen')
ax2.set_title(r'(b) $\alpha$'+' varies' + '\n' + r'$\mu = 3$')
ax2.set_xlabel(r'$\alpha$')
TTestIndPower().plot_power(dep_var='nobs',
                           nobs= np.array([i for i in range(5,21)]),
                           effect_size=[1.5],
                           alpha=0.05,
                           ax=ax3, title='(c) n varies' + '\n' + r'$\alpha = 0.05$')

power_analysis = TTestIndPower()
all_p=[]
sigmas_ = [i for i in range(1,10)]
for s in sigmas_:
    p = power_analysis.power(effect_size=(3-0)/math.sqrt(s),nobs1=15, alpha = 0.05, ratio = 1)
    all_p.append(p)
sns.lineplot(x = sigmas_, y=all_p, ax = ax4, color = 'darkgreen')
ax4.set_title(r'(d) $\sigma^2$'+' varies' + '\n' + r'$n = 15$')
ax4.set_xlabel(r'$\sigma^2$')
plt.tight_layout()
plt.savefig('542hw3_q4.jpg',dpi=300)
```

Code




y-axis is $1-\beta$.

(e).
- power decreases first as the true mean $\mu$ increases, and after $\mu$ is larger than 0, the power then increases.
- When $\mu$ is 3, power increases when $\alpha$ increases, the probability of type II error approaches 0 quickly when $\alpha$ is small
- When $\alpha$ is fixed, as $n$ increases, power increases and approach 1.
- For fixed $n$, the larger $\sigma$ is, the smaller power will be, and the type II error is more likely to happen.

**Problem 5.** Suppose that $X_1, \ldots, X_n$ are an i.i.d. random sample of size $n$ drawn from a Poisson distribution with unknown parameter $\lambda$. Find the likelihood ratio for testing $H_0 : \lambda = \lambda_0$ versus $H_a : \lambda = \lambda_1$ where $\lambda_1 > \lambda_0$. Use the fact that the sum of independent Poisson random variables follows a Poisson distribution to explain how to determine a rejection region for a test at significance level $\alpha_0$. for $X_i \sim Pois(\lambda)$, $\mu = \lambda$ $\quad \sigma^2 = \lambda$

$$\frac{lik(\lambda_0)}{lik(\lambda_1)} = T(X_1, \cdots, X_n) \quad \text{, for poisson distribution, } f(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$lik(\lambda) = f(x_1, \cdots x_n|\lambda) = \prod_{i=1}^{n} f(x_i|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

$$\log \text{ of } lik(\lambda) = \sum_{i=1}^{n} [x_i \log(\lambda) + (-\lambda) - \log(x_i!)] = -n\lambda - \sum_{i=1}^{n} \log(X_i!) + \log(\lambda) \sum_{i=1}^{n} x_i$$

$$T(X_1, \cdots, X_n) = -n\lambda_0 - \sum_{i=1}^{n} \log(X_i!) + \log(\lambda_0) \sum_{i=1}^{n} X_i - \left(-n\lambda_1 - \sum_{i=1}^{n} \log(X_i!) + \log(\lambda_1) \sum_{i=1}^{n} X_i\right)$$

$$= -n(\lambda_0 - \lambda_1) + [\log(\lambda_0) - \log(\lambda_1)] \sum_{i=1}^{n} X_i \underset{H_a}{\overset{H_0}{\gtrless}} \log k \quad , \lambda_1 > \lambda_0 \quad \log(\lambda_0/\lambda_1) < 0$$

$$\sum_{i=1}^{n} X_i \underset{H_a}{\overset{H_0}{\lessgtr}} \frac{\log k + n(\lambda_0 - \lambda_1)}{\log(\lambda_0/\lambda_1)} \quad (*), \qquad (*) = k'$$

To find $k'$, $Pr\{\text{type I error}\} = \alpha_0 \Rightarrow Pr\{H_0 \text{ is rejected} \mid H_0 \text{ is true}\} = \alpha_0$

$$Pr\{\sum_{i=1}^{n} X_i > k' \mid x_i \sim \text{poisson}(\lambda_0)\} = \alpha_0$$

$$Pr\{\sum_{i=1}^{n} X_i > k' \mid \Sigma X_i \sim \text{Poisson}(n\lambda_0)\} = \alpha_0$$

$$\alpha = Pr\{\sum_{i=1}^{n} X_i > k' \mid \lambda = \lambda_0\}$$

$$= 1 - F(y) \quad , F = \text{CDF of Poisson distribution}$$

$E[X] = \frac{1}{\lambda}$
$Var[X] = \frac{1}{\lambda^2}$

**Problem 6.** Let $X_1, \ldots, X_n$ be a random sample from an exponential distribution with the density function $f(x|\theta) = \theta \exp(-\theta x)$. Derive a likelihood ratio test of $H_0 : \theta = \theta_0$ verses $H_A : \theta \neq \theta_0$, and show that the rejection region is of the form $\bar{X} \exp(\theta_0 \bar{X}) \leq c$.

$\bar{x} \exp(-\theta_0 \bar{x}) \leq c$

$X_i \overset{iid}{\sim} \exp\left(\frac{1}{\theta}, \frac{1}{\theta^2}\right)$

the numerator of likelihood ratio statistic is

$f(x_1, \cdots x_n | \theta_0) = \theta_0^n \, e^{\sum_{i=1}^n -\theta_0 x_i}$

the denominator is the likelihood of $x_1, \cdots x_n$ evaluated with $\theta = \hat{\theta}$

$$\hat{\theta}^n e^{\sum_{i=1}^n -\hat{\theta} x_i} = \left(\frac{1}{\bar{x}}\right)^n e^{\sum_{i=1}^n -\frac{1}{\bar{x}} x_i}$$

$$T(x_1, \cdots x_n) = \left(\frac{\theta_0}{\frac{1}{\bar{x}}}\right)^n \exp\left\{\sum_{i=1}^n -\theta_0 x_i - \sum_{i=1}^{1} -\frac{1}{\bar{x}} x_i\right\} \underset{H_a}{\overset{H_0}{\gtrless}} k$$

$$(\theta_0 \bar{x})^n \exp\{n(-\theta_0 \bar{x} + 1)\} \underset{H_a}{\overset{H_0}{\gtrless}} k$$

rejection region is $\left(\theta_0 \bar{x} \exp\{-\theta_0 \bar{x} + 1\}\right)^n \leq c''$

$\theta_0 \bar{x} \exp\{-\theta_0 \bar{x} + 1\} \leq c'$, where $c' = c''^{\frac{1}{n}}$

$\bar{x} \exp\{-\theta_0 \bar{x}\} \leq c$, where $c = \frac{1}{\theta_0 e} c'$