1. (Confidence intervals and Hypothesis testing)

(a) TRUE/FALSE: Confidence intervals are random intervals.

(b) What is the maximum confidence level for \((0, \infty)\) CI covering the population variance?

(c) Suppose we have the following confidence interval for the population mean

\[
\bar{X} \pm t_{1-\alpha/2, n-1} \frac{\hat{\sigma}}{\sqrt{n}} = [10, 15]
\]

If my hypotheses were \(H_0 : \mu - 12 = 0\) and \(H_0 : \mu - 12 \neq 0\), would I reject the null under this sample?

(d) What is the minimum value that power can take for any hypothesis testing procedure?

(e) What is the power of a testing procedure that always rejects \(H_0\), regardless of what the sample is? What is the Type I error? What is the Type II error? Briefly explain your answers.

2. Remember the Gmail spam example at the very first lecture? In these series of short questions, we'll answer various questions about characteristics of spam mail by considering a sample of 4,601 e-mails, of which 1813 are considered spam.

(a) Suppose I believe that spam mail tend to contain more capitalized letters than non-spam mail. I collected the following data on the number of capital letters betwen spam mail and non-spam mail.

<table>
<thead>
<tr>
<th></th>
<th>Sample mean</th>
<th>Sample SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spam mail</td>
<td>471 capitalized letters</td>
<td>825</td>
</tr>
<tr>
<td>Regular mail</td>
<td>161 capitalized letters</td>
<td>356</td>
</tr>
</tbody>
</table>

Table 1: Table of mean capitalized letters in the sample along with its standard deviation

Conduct a hypothesis test by setting up \(H_0\) and \(H_a\), using an appropriate test statistic, and choosing a reasonable sampling distribution. Make sure you state your assumptions about the sampling distribution. Do we reject the null at the \(\alpha = 0.05\) level?
(b) Construct a $1 - \alpha$ CI for the difference in the average number of capital letters in spam mail and non-spam mail. *You must show mathematical justification!*

(c) Suppose I believe there is more variability in the number of capitalized letters for spam mail than non-spam mail. Conduct a hypothesis test by setting up $H_0$ and $H_a$, using an appropriate test statistic, and choosing a reasonable sampling distribution. Make sure you state your assumptions about the sampling distribution. Do we reject the null at the $\alpha = 0.05$ level?