Formulas.

If $X$ has the binomial distribution with parameters $n$ and $p$, then $\mathbb{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$, for $k = 0, \ldots, n$, $\mathbb{E}[X] = np$, $\text{Var}[X] = np(1-p)$.

If $X$ has the standard normal distribution, then its mean is zero, its variance is one, and its density equals

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R}.$$ 

True/false questions.

Problem 1.1. (2 points)

An observation is said to be influential for a least squares regression line if its deletion would have a large effect on the estimates of the slope and intercept parameters.

True or false?

Solution: TRUE

Free-response problems.

Problem 1.2. (8 points)

To write an article about Denver for a tourist magazine you would like to estimate the average nightly cost for a hotel room in the Denver area. You are willing to assume that the nightly cost of a room is normally distributed.

You open up the yellow pages and take a random sample of hotels. The sample of 16 hotels gives an average nightly cost of $55.98 and a sample standard deviation of $12. Estimate the mean nightly cost and include a 95% confidence interval.

Solution: The number of degrees of freedom of the $t$-distribution is $16 - 1 = 15$. The critical value associated with this distribution and the confidence level of 95% is 2.131. Since the sample standard deviation equals $12/\sqrt{16} = 3$. So, the confidence interval we are looking for is

$$\mu = 55.98 \pm 2.131 \times 3 = 55.98 \pm 6.393.$$ 

Problem 1.3. (7 points)

Consider an experiment designed to study the relation between the yield ($y$ in grams) of a chemical process and the temperature setting ($x$ in F) for an important reaction phase of the process. The following is the statistical software output of a simple linear regression.

```
. regress yield temp

Source | SS       df    MS
---------+------------------- Number of obs =  17
Model    | 6949.01454  1 6949.01454 F(  1,   15) = 64.08
Residual | 1626.56001 15 108.437334 Prob > F = 0.0000
---------+------------------------------ Adj R-squared = 0.7977
Total    | 8575.57455 16 535.97341 R-squared = 0.8103
       | Root MSE = 10.413
```

Course: M358K  Instructor: Milica Ćudina  Semester: Fall 2016
(i) (3 points) What is the equation of the least-squares regression line?

Solution:

\[ \hat{y} = 164.5075 - 0.4552782x \]

(ii) (2 points) What is the value of the coefficient of determination?

Solution:

\[ R = \sqrt{0.8103} = 0.9002 \]

(iii) (2 points) The sample mean (\( \bar{x} \)) and sample standard deviation (\( s_x \)) of the observed temperature values are given to be 113.8 and 45.8, respectively. What is the sample mean of the response random variable (\( \bar{y} \))?

Solution:

\[ \bar{y} = 164.5075 - 0.4552782\bar{x} = 164.5075 - 0.4552782 \times 113.8 = 112.6968. \]

Problem 1.4. (10 points)

Suppose that we have a random sample of size 25 from a normal population with an unknown mean \( \mu \) and a standard deviation of 4. Bertram Wooster was in charge of testing the hypothesis

\[ H_0 : \mu = 10 \quad \text{vs.} \quad H_a : \mu > 10. \]

Before he went on vacation, he obtained the rejection region (11.2, \( \infty \)). However, he forgot to tell anyone which significance level \( \alpha \) he used.

(i) (5 points) Calculate \( \alpha \).

Solution: The \( \alpha \) significance level needs to be consistent with the rejection region. So,

\[ 1 - \alpha = \Phi \left( \frac{11.2 - 10}{4/\sqrt{25}} \right) \Rightarrow \alpha = 1 - \Phi(1.5) = 0.0668. \]

(ii) (5 points) What is the power of the above test at the alternative mean \( \mu_a = 11 \)?

Solution: The power of the test is the probability that the null hypothesis is rejected under a particular alternative. In this case, we have

\[ 1 - \Phi \left( \frac{11.2 - 11}{0.8} \right) = 0.4013. \]

Problem 1.5. (15 points)

A casino game involves rolling three dice. The winnings are proportional to the total number of sixes rolled. Suppose a gambler plays the game 150 times, with the following observed counts:

<table>
<thead>
<tr>
<th>Number of sixes</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>72</td>
</tr>
<tr>
<td>1</td>
<td>51</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
</tr>
</tbody>
</table>

Assuming that the die rolls are independent, test the null hypothesis that the dice are all fair.
Solution: If the dice were all fair, we would have the following probabilities of the events that a particular number of sixes was rolled:

\[ P[0 \text{ sixes were rolled}] = \left(\frac{5}{6}\right)^3, \]
\[ P[1 \text{ six was rolled}] = 3\left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right), \]
\[ P[2 \text{ sixes were rolled}] = 3\left(\frac{5}{6}\right)\left(\frac{1}{6}\right)^2, \]
\[ P[3 \text{ sixes were rolled}] = \left(\frac{1}{6}\right)^3. \]

So, the expected numbers \( E_i \) of times that \( i \) sixes in 150 rolls occur (for \( i = 0, 1, 2, 3 \)) are

\[ E_0 = 150 \times \frac{125}{6^3} = 86.80556, \]
\[ E_3 = 150 \times \frac{25}{6^3} = 52.08333, \]
\[ E_0 = 150 \times \frac{15}{6^3} = 10.41667, \]
\[ E_0 = 150 \times \frac{125}{6^3} = 0.6944. \]

The observed value of the \( \chi^2 \)-statistic is

\[ \frac{(86.80556 - 72)^2}{86.80556} + \frac{(52.08333 - 51)^2}{52.08333} + \frac{(10.41667 - 21)^2}{10.41667} + \frac{(0.6944 - 6)^2}{0.6944} = 53.83487. \]

With the number of degrees of freedom is \( 4 - 1 = 3 \), we see that the observed value of the \( \chi^2 \)-statistic exceeds even the critical value 17.73 at the upper-tail probability of 0.0005.

Problem 1.6. (10 points)
A simple random sample of 200 students is selected from a large university. In this sample, there are 35 minority students. A simple random sample of 80 students is selected from the community college in the same town. In this sample, there are 28 minority students. What is the standard error of the difference in sample proportions of minority students?

Solution:
In our usual notation, we have

\[ \hat{p}_1 = \frac{35}{200} = 0.175, \quad \hat{p}_2 = \frac{28}{80} = 0.35. \]

So, the standard error is

\[ SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} = \sqrt{\frac{0.175 \times 0.825}{200} + \frac{0.35 \times 0.65}{80}} = 0.0597. \]

Problem 1.7. (15 points) Source: Ramachandran-Tsokos.

A study of two kinds of machine failures shows that 58 failures of the first kind took on the average 79.7 minutes to repair with a sample standard deviation of 18.4 minutes, whereas 71 failures of the second kind took on average 87.3 minutes to repair with a sample standard deviation of 19.5 minutes. Find a 99% confidence interval for the difference between the true average amounts of time it takes to repair failures of the two kinds of machines.

Solution: The standard error of the difference in sample means is

\[ \sqrt{\frac{18.4^2}{58} + \frac{19.5^2}{71}} = 3.3456. \]

So, the 99%-confidence interval for \( \mu_1 - \mu_2 \) is

\[ (79.7 - 87.3) \pm 2.575(3.3456) = (-16.2149, 1.01486). \]
Multiple-choice problems.

Problem 1.8. (5 points) Source: Ramachandran-Tsokos.

A dendritic tree is a branched formation that originates from a nerve cell. In order to study brain development, researchers want to examine the brain tissues from adult guinea pigs. At least how many cells must the researchers select (randomly) so as to be 95% sure that the sample mean is within 3.4 cells of the population mean? Assume that a previous study has shown that the cells have the standard deviation of exactly 10 dendrites.

(a) 28  
(b) 33  
(c) 34  
(d) 35  
(e) None of the above.

Solution: (c) We require that the margin of error at the 95% confidence be at most 3.4. So, the sample size $n$ must satisfy

$$1.96 \times \frac{10}{\sqrt{n}} \leq 3.4 \Rightarrow n \geq \left( \frac{1.96 \times 10}{3.4} \right)^2 = 33.23183.$$