MATH 322 Probability & Statistics II  
Spring 2014  
Midterm II

1. We are interested in testing weather or not a coin is balanced based on number of heads Y on 10 tosses of the coin. Ho: \( p = 0.5 \) against Ha: \( p \neq 0.5 \). If we use the rejection region \( |Y - 5| \geq 2 \), what is \( n = 10 \)? [5]

\[ \alpha = P(\text{Type I error}) \]
\[ = P(\text{Reject Ho when Ho is true}) \]
\[ = P(10 \leq 5 \leq 2 \text{ when } p = 0.5) \]
\[ = P(Y = 0, 1, 2, 3, 7, 8, 9, 10 \text{ when } p = 0.5) \]
\[ = 1 - P(Y = 4, 5, 6 \text{ when } p = 0.5) \]
\[ = 1 - \left[ \binom{10}{4} \times 0.5^4 \times (0.5)^{10-4} + \binom{10}{5} \times (0.5)^{10-5} + \binom{10}{6} \times (0.5)^{10} \right] \]
\[ = 0.34375 \]

a. Probability of type I error (\( \alpha \))?

b. Probability of type II error (\( \beta \)) if \( p = 0.7 \)? [5]

\[ \beta (0.7) = P(\text{Type II error}) \]
\[ = P(\text{fail to reject Ho when Ho is false}) \]
\[ = P(10 \leq 5 \leq 2 \text{ when } p = 0.7) \]
\[ = P(Y = 4, 5, 6 \text{ when } p = 0.7) \]
\[ = \binom{10}{4} \times (0.7)^4 \times (0.3)^6 + \binom{10}{5} \times (0.7)^5 \times (0.3)^5 + \binom{10}{6} \times (0.7)^6 \]
\[ = 0.3398 \]

2. The output voltage for an electric circuit is specified to be 130. A sample of 40 independent readings on the voltage for this circuit gave a sample mean 128.6 and standard deviation 2.1. We wish to test the hypothesis that the average output voltage is 130 against the alternative that it is less than 130. What is the appropriate rejection region (in terms of sample mean) for an \( \alpha = 0.05 \) level test? [5]

\( H_0: \mu = 130 \)
\( H_a: \mu < 130 \)

\[ \alpha = P(\text{Reject } H_0 \text{ when } H_0 \text{ is true}) = 0.05 \]
\[ = P(\bar{X} \leq c \text{ when } \mu = 130) = 0.05 \]
\[ = P\left( \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{c - E(\bar{X})}{\sigma/\sqrt{n}} \right. \text{ when } \mu = 130 \) \]
\[ = P\left( Z \leq \frac{c - 130}{2.1/\sqrt{40}} \right) = 0.05 \]
\[ \Rightarrow c = \frac{129.3452}{2.1/\sqrt{40}} = 129.454 \]
3. You want to see if a redesign of the cover of a mail-order catalog will increase sales. A very large number of customers receive the catalog, and a random sample of customers will receive the one with the new cover. For planning purposes, you assume that sales from the new catalog will be approximately normal with \( \sigma = 60 \) and that the mean for the original catalog will be \( \mu = 40 \). You decide to use a sample of size \( n = 1000 \). You wish to test the following hypothesis:

\[
H_0: \mu = 40 \quad \text{vs} \quad H_A: \mu > 40
\]

You decide to reject \( H_0 \) if \( \bar{x} > 43.12 \) and fail to reject \( H_0 \) otherwise.

a. Find the probability of a Type I error. [5]

\[
\alpha = P(\text{Type I error})
\]

\[
= P(\text{Reject } H_0 \text{ when } H_0 \text{ is false})
\]

\[
= P(\bar{x} > 43.12 \text{ when } \mu = 40)
\]

\[
= P\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} > \frac{43.12 - 40}{60/\sqrt{1000}}\right)
\]

\[
= P(Z > 1.644) = 0.05
\]

b. Find the probability of a Type II error and the power of this test when in fact \( \mu = 45 \). [5]

\[
\text{Power} = 1 - \beta = P(\text{Reject } H_0 \text{ when } H_0 \text{ is false})
\]

\[
= 1 - \beta(45) = P(\bar{x} > 43.12 \text{ when } \mu = 45)
\]

\[
= P\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} > \frac{43.12 - 45}{60/\sqrt{1000}}\right)
\]

\[
= P(Z > -0.99) = 1 - P(Z \leq 0.99)
\]

\[
\Rightarrow \beta = 0.1611
\]

\[
\Rightarrow \text{Power} = 0.8389
\]

4. 4 out of 10 doctors knew the generic name for the drug methadone. We want to test if fewer than half of all doctors know the generic name for methadone. Perform the hypothesis test using a significance level of 0.2. [5]

\[
\alpha = 0.2
\]

\[
P(\text{Type I error}) = 0.2
\]

\[
P(\text{Reject } H_0 \text{ when } H_0 \text{ is true}) = 0.2
\]

\[
P(\bar{X} \leq 4 \text{ when } \bar{p} = 0.5) = 0.2
\]

\[
\Rightarrow c = 3
\]

Since \( x = 4 > c = 3 \), we fail to reject \( H_0 \)

\[
p \text{-value} = P(X \leq 4, X \sim \text{bin}(10, 0.5)) = 0.337
\]

\[
p \text{-value} > \alpha \text{-value}, \text{we fail to reject } H_0
\]
5. A package-delivery service advertises that at least 80% of all packages brought to its office by 9 A.M. for delivery in the same city are delivered by noon that day. Let \( p \) denote the true proportion of such packages that are delivered as advertised and consider the hypotheses \( H_0: p = 0.80 \) versus \( H_a: p < 0.80 \).

a. If the truth is that only 75% of the packages are delivered as advertised, how likely is it that a level 0.05 test based on \( n = 220 \) packages fails to detect such a departure from \( H_0 \)? [5]

\[
P(0.75) = 1 - \Phi \left( \frac{0.75 - 0.80}{\sqrt{0.80(1 - 0.80)/220}} \right)
= 1 - \Phi \left( 0.19 \right)
= 0.4247
\]

Power of the test \( 1 - \beta = 0.5753 \)

b. What should the sample size be to ensure that \( \beta(0.75) = 0.05 \)? [5]

The sample size \( n \) for which the level \( \alpha \) test also satisfies \( P(0.75) = 0.05 \) is

\[
n = \left[ \frac{Z_\alpha \sqrt{p_0(1-p_0)} + Z_{0.05} \sqrt{0.75(1-0.75)}}{0.75 - 0.80} \right]^2
= 752
\]

Note: \( \alpha = 0.05 \Rightarrow Z_\alpha = 1.645 \) and \( \beta(0.75) = 0.05 \Rightarrow Z_\beta = 1.645 \)

6. 51% of the children born in the United States are male. Researchers chose a sample of 200 babies born last week.

a. Characterize the sampling distribution of \( \hat{p} \), the proportion of the sample that was males. [2]

\[
N \cdot P_0 = 102 > 10 \quad \text{and} \quad N \cdot (1 - P_0) = 98 > 10
\]

\[
\hat{p} \sim \text{N}(P_0, \sqrt{P_0(1-P_0)/n}) \quad (i.e.) \quad \hat{p} \sim \text{N}(0.51, 0.035)
\]

b. What is the probability that more than 55% of the babies in the sample were male? [5]

\[
P(\hat{p} > 0.55 \text{ when } p = 0.51) = P \left( \frac{\hat{p} - \mu(\hat{p})}{\text{s.d}(\hat{p})} > \frac{0.55 - E(\hat{p})}{\text{s.d}(\hat{p})} \right)
= P \left( Z > \frac{0.55 - 0.51}{\sqrt{(0.51)(0.49)/200}} \right)
= 1 - \Phi(1.13) = 0.1292
\]
7. You are analyzing a sample of 50 values drawn from a normal distribution with mean 4 and variance 16. $S^2$ is the unbiased sample variance. Find the critical value $c$ such that $Pr(S^2 \leq c) = 0.9$. [5]

$$P \left( S^2 \leq c \right) = P \left( \frac{(n-1)S^2}{\sigma^2} \leq \frac{(n-1)c}{\sigma^2} \right) = P \left( \chi^2 \leq \frac{49c}{16} \right) = 0.9$$

$$\frac{49c}{16} = 62.0117 \Rightarrow c = 20.25$$

8. A machine packs cereal into boxes, and you don’t want too much variation from box to box. You decide on a standard deviation of no more than five grams (about 1/6 ounce). To determine whether the machine is operating within specification, you randomly select 45 boxes. Here are the weights of the boxes, in grams:

<table>
<thead>
<tr>
<th>386</th>
<th>388</th>
<th>381</th>
<th>395</th>
<th>392</th>
<th>383</th>
<th>389</th>
<th>383</th>
<th>370</th>
</tr>
</thead>
<tbody>
<tr>
<td>379</td>
<td>382</td>
<td>388</td>
<td>390</td>
<td>386</td>
<td>393</td>
<td>374</td>
<td>381</td>
<td>386</td>
</tr>
<tr>
<td>391</td>
<td>384</td>
<td>390</td>
<td>374</td>
<td>386</td>
<td>393</td>
<td>384</td>
<td>381</td>
<td>386</td>
</tr>
<tr>
<td>386</td>
<td>374</td>
<td>393</td>
<td>385</td>
<td>388</td>
<td>384</td>
<td>385</td>
<td>388</td>
<td>392</td>
</tr>
<tr>
<td>400</td>
<td>377</td>
<td>378</td>
<td>392</td>
<td>380</td>
<td>380</td>
<td>395</td>
<td>393</td>
<td>387</td>
</tr>
</tbody>
</table>

Find the p-value to test the hypothesis at 1% [5]

$H_0: \sigma = 5$, the machine is within spec  
$Ha: \sigma > 5$, the machine is not working right

$$H_0 : \sigma = 5 \quad \text{and same as} \quad H_0 : \sigma^2 = 5^2$$

$$H_a : \sigma^2 > 5^2$$

Test statistic: $\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{(44)(41.25)}{5^2} = 72.59$

$P\text{-value} = P(\chi^2 \geq \chi^2)$

$$= P(\chi^2 > 72.59) = 1 - P(\chi^2 \leq 72.59)$$

$$= 1 - P(\text{chisq}(72.59, 44))$$

$$= 0.00427$$

Since $P\text{-value} < \alpha\text{-value}$, reject $H_0$ at 1%.
9. The GPA for the \( m = 100 \) students who do not own cars had a sample average equal to 2.70 (assume the GPA for these students has known population variance of 0.36). The \( n = 100 \) car owners had a sample average GPA of 2.54 (assume the GPA for these students has known population variance of 0.40). Suppose the truth is that \( \mu_1 - \mu_2 = 0.2 \). What is the probability of detecting such a departure from \( H_0 \) with the current data set?

\[
H_0: \mu_1 - \mu_2 = 0 \\
H_1: \mu_1 - \mu_2 > 0 \\
\text{Pick } \alpha = 0.01
\]

\[
\beta(0.2) = \Phi\left(2.33 - \frac{0.2}{\sqrt{\frac{0.36}{100} + \frac{0.40}{100}}}\right) = \Phi\left(2.33 - 2.298\right) = \Phi(0.032) = 0.6255
\]

\[
\text{Power } = 1 - \beta = 1 - 0.6255 = 0.3745
\]

10. Low back pain (LBP) is a serious health problem in many industrial settings. The article YYY reported the accompanying summary data on lateral range of motion (degrees) for a sample of workers without a history of LBP and a history of this malady. Assume normality where needed.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Sample size</th>
<th>Sample mean</th>
<th>Sample SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>No LBP</td>
<td>28</td>
<td>91.5</td>
<td>5.5</td>
</tr>
<tr>
<td>LBP</td>
<td>31</td>
<td>88.3</td>
<td>7.8</td>
</tr>
</tbody>
</table>

Calculate a 90% confidence interval for the difference between population mean extent of lateral motion for the 2 conditions. Does the interval suggest that the population mean lateral motion differs for the 2 conditions? [5]

\[
d \cdot \bar{t} = S_d.
\]

\[
(91.5 - 88.3) \pm 1.676 \sqrt{\frac{5.5^2}{28} + \frac{7.8^2}{31}}
\]

\[
(0.276, 6.123)
\]

Yes, since "0" is not in the interval suggest that \( H_1: \mu_1 \neq \mu_2 \)

11. An herbal medicine is tested on 4 randomly chosen patients with sleep disorders. Each patient has the amount of sleep (in hours) for one night with the herbal medicine and for one night without the herbal medicine. Estimate the 90% confidence interval for the difference in the mean amount of sleep achieved with and without the herbal. What is the assumption you need to make? [5]

\[
\begin{array}{c|cccc}
\text{Patient} & 1 & 2 & 3 & 4 \\
\hline
\text{Without herbal} & 1.8 & 5.1 & 4.2 & 4.0 \\
\text{With herbal} & 3.0 & 6.6 & 5.6 & 6.2 \\
\end{array}
\]

\[
\bar{d} = \bar{x} - \bar{y} = -1.2 -1.5 -1.4 -2.2 \\
\bar{d} = -1.575 \quad \text{and} \quad \bar{s} = 0.435
\]
\[ x = 0.1 \Rightarrow x/2 = 0.05, \quad d.f = 3 \]
\[ t = 2.353 \]
\[ -1.575 \pm 2.353 \cdot 0.435 = (-2.087, -1.043) \]

All pairs are independent and the difference is normally distributed.

12. Chemotherapy versus a combination of chemotherapy and radiation for breast cancer patients were studied. Of 154 patients who had chemotherapy only, 76 survived at least 15 years. Among 164 patients who had the combination, 98 survived at least 15 years.

a. Test whether there is a significant difference between the proportion of survivors using chemotherapy only and that of survivors using the combination?

Use \( \alpha = 0.05 \). [5+5]

\[ \hat{p}_1 = \frac{76}{154} = 0.4935, \quad \hat{p}_2 = \frac{98}{164} = 0.5976 \]

Combined \[ \hat{p} = \frac{76 + 98}{154 + 164} = 0.5474 \]

\[ Z^* = \frac{0.4935 - 0.5976}{\sqrt{0.5474(0.4526)\left[\frac{1}{154} + \frac{1}{164}\right]}} = -1.871 \]

Reject \( H_0 \) if \( Z^* \neq Z_{d/2} \) (or) \( Z^* \leq -Z_{d/2} \)

b. Construct a 99% confidence interval for the difference between two proportions.

\[ x = 0.01 \Rightarrow Z_{d/2} = Z_{0.005} = 2.575 \]

\[ (0.4935 - 0.5976) \pm 2.575 \sqrt{\frac{0.4935(0.5065) + 0.5976(0.4034)}{154 + 164}} \]

\[ (-0.2472, 0.0386) \]

13. Let \( P_1 \) and \( P_2 \) be the probabilities of a child getting paralytic polio for the control and treatment group conditions. The objective was to test

\[ H_0: P_1 - P_2 = 0 \text{ versus } H_a: P_1 - P_2 > 0. \]

Supposing the true value of \( P_1 \) is 0.0003, and \( P_2 = 0.00015 \). Using a level \( \alpha = 0.05 \) test, what would be reasonable to ask for sample sizes for \( \beta = 0.1 \) by assuming the sample sizes are equal?

\[ n = 171,073 \]
14. An experiment was planned to compare the mean time (in days) required to recover from a common cold for persons given a daily dose of 4 milligrams of vitamin C versus those who were not given a vitamin supplement. Suppose that 45 adults were randomly selected for each treatment category and the mean recovery time and standard deviations for the two groups were as follows:

<table>
<thead>
<tr>
<th></th>
<th>No vitamin</th>
<th>4mg vitamin C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>Sample mean</td>
<td>6.9</td>
<td>5.8</td>
</tr>
<tr>
<td>Sample sd</td>
<td>2.9</td>
<td>1.2</td>
</tr>
</tbody>
</table>

The researchers' hypothesis is that the use of vitamin C reduces the mean time required to recover from a common cold. [10]

a. If the no vitamin group is population 1, what are the appropriate null and alternative hypotheses here? Which test is appropriate to perform this hypothesis?

\[ H_0 : \mu_1 - \mu_2 = 0 \quad \text{since } n \geq 30, \text{ we use large sample } Z\text{-test.} \]

\[ H_a : \mu_1 - \mu_2 > 0 \]

b. Is the variation in recovery times between two populations different? Use \( \alpha = 0.02 \). Find P-value. What assumptions are necessary to carry out a test of hypothesis?

\[
\begin{align*}
H_0 : \sigma_1^2 &= \sigma_2^2 & \left[ \frac{\sigma_1^2}{\sigma_2^2} = 1 \right] \\
H_a : \sigma_1^2 &\neq \sigma_2^2 & \left[ \frac{\sigma_1^2}{\sigma_2^2} \neq 1 \right] \\
F^* &= \frac{S_1^2}{S_2^2} = \frac{(2.9)^2}{(1.2)^2} = 5.840 \\
R\text{-code:} & 2\times(1 - \Phi(5.840, 44, 44)) = 0.000 \\
\end{align*}
\]

Since \( P\text{-value} \approx 0 < 0.02 \), we reject \( H_0 \).

\[
P\text{-value} = 2\times P(F \geq 5.840) \\
\]

\[
\text{c. Will it be reasonable to use a pooled t-test to compare the mean recovery times?}
\]

\[
\text{NO, since } \sigma_1^2 \neq \sigma_2^2.
\]