Stat 202  
Fall 2004  
Exam 1 Solution Procedures  

Read this first!  

There are multiple versions of this exam, and different versions have different numbers in the problems. I’m showing the solutions for a specific version, and am showing you the process used in each problem.  

For problems that you missed, your graded exam should show you the correct answer. It would be great practice to work out each problem you missed on your exam to verify that you can match that solution!  

Let $Z$ represent the standard normal random variable. Calculate:  

1. $P\{Z < 1.54\}$ Report your answer to four decimal places!  

Draw a sketch: you’d want the area to the left of 1.54. That’s 0.4382 + 0.5000 = 0.9382  

2. $P\{-0.37 < Z < +1.82\}$  

Draw a sketch: you want the area between these two numbers. Solution is 0.1443 + 0.4656 = 0.6099.  

Suppose that the times that customers wait to be served at the Department of Motor Vehicles follow a normal distribution with a mean of 150 minutes and a standard deviation of 60 minutes.  

3. What is the probability that a randomly selected customer has to wait for more than 170 minutes?  

Draw a sketch: it’s an upper tail probability. Z-value is $(170 – 150)/60 = 0.33$. Solution is 0.5000 – 0.1293 = 0.3707.  

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.1587</td>
<td>0.2514</td>
<td>0.3085</td>
<td>0.3446</td>
<td>0.3707</td>
<td>0.6293</td>
<td>0.6554</td>
<td>0.6915</td>
<td>0.7486</td>
<td>0.8413</td>
<td>1.0</td>
<td></td>
</tr>
</tbody>
</table>

4. What is the 65th percentile waiting time? Draw a sketch: you want the cutoff that has area 0.65 to the left, which puts 0.15 between our cutoff and the mean. The corresponding z-value is 0.39, so you need to be 0.39 standard deviations above the mean. Solution is $150 + 0.39(60) = 173.4$.  

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>90.8</td>
<td>101.2</td>
<td>111.6</td>
<td>122.0</td>
<td>132.4</td>
<td>157.8</td>
<td>161.7</td>
<td>165.6</td>
<td>169.5</td>
<td>173.4</td>
<td></td>
</tr>
</tbody>
</table>
5. What is the probability that a randomly selected customer has to wait for exactly 125 minutes?

*The probability that a continuous RV is exactly equal to any value is always zero. Only intervals have non-zero probabilities.*

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>0.0</strong></td>
<td>0.1056</td>
<td>0.2033</td>
<td>0.2643</td>
<td>0.3085</td>
<td>0.3372</td>
<td>0.6628</td>
<td>0.6915</td>
<td>0.7357</td>
<td>0.7967</td>
<td>0.8944</td>
<td>1.0</td>
</tr>
</tbody>
</table>

6. Suppose that the lifetimes, in operating hours, of Phillips Model 12 CD recorders follow a skewed distribution with a mean of 4,000 hours and a standard deviation of 3,500 hours.

If we randomly select and test a sample of 40 of the recorders, what is the probability that our sample average lifetime will be less than 3,300 hours?

*It's a question about a sample mean $\bar{X}$. Since $n > 30$, $\bar{X}$ is approximately normally distributed with mean equal to 4,000 and standard deviation equal to $3,500/\sqrt{40} = 553.4$. Draw a sketch: we need the area below 3,300 (lower tail). Z-value is $(3,300 - 4000)/553.4 = -1.26$. Solution is $0.5000 - 0.3962 = 0.1038.*

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0000</td>
<td>0.0351</td>
<td>0.0516</td>
<td>0.0735</td>
<td><strong>0.1038</strong></td>
<td>0.1401</td>
<td>0.3859</td>
<td>0.3974</td>
<td>0.4090</td>
<td>.04207</td>
<td>.04207</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>K</th>
<th>L</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4325</td>
<td>1.000</td>
<td>Cannot be determined from the info given</td>
</tr>
</tbody>
</table>
7.  Probability problems involving the sample mean $\bar{X}$ or the sample proportion $p$ are fairly easy to solve is we know that $\bar{X}$ or $p$ has, either exactly or approximately, a normal distribution. Consider the following problems:

(i) Find $P\{\bar{X} < 50\}$ where $\bar{X}$ is based on a random sample of size $n = 8$ from a normally distributed population, i.e. individual $X$’s are normally distributed.
(ii) Find $P\{\bar{X} < 50\}$ where $\bar{X}$ is based on a random sample of size $n = 45$ from a skewed population.
(iii) Find $P\{\bar{X} < 50\}$ where $\bar{X}$ is based on a random sample of size $n = 22$ from a skewed population.
(iv) Find $P\{p < 0.20\}$ where $p$ is based on a random sample of size $n = 80$ from a population whose proportion of “successes” is given by $\pi = 0.5$.
(v) Find $P\{p < 0.20\}$ where $p$ is based on a random sample of size $n = 400$ from a population whose proportion of “successes” is given by $\pi = 0.12$.

For the $\bar{X}$-problems, you need a normal population OR a sample size larger than 30. For the proportion problems, you need $n\pi$ and $n(1 - \pi) > 5$. The only scenario that doesn’t meet this criteria is the third one.

For which of these problems is it NOT OK to use the normal distribution to calculate the desired probability?

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) only</td>
<td>(ii) only</td>
<td>(iii) only</td>
<td>(iv) only</td>
<td>(v) only</td>
<td>(i) &amp; (ii)</td>
<td>(i) &amp; (iii)</td>
</tr>
<tr>
<td>H</td>
<td>I</td>
<td>J</td>
<td>K</td>
<td>L</td>
<td>M</td>
<td>N</td>
</tr>
<tr>
<td>(i) &amp; (iv)</td>
<td>(i) &amp; (v)</td>
<td>(ii) &amp; (iii)</td>
<td>(ii) &amp; (iv)</td>
<td>(ii) &amp; (v)</td>
<td>(iii) &amp; (iv)</td>
<td>(iv) &amp; (v)</td>
</tr>
</tbody>
</table>

8.  The results of an extensive customer survey commissioned by uppermost management showed that 37% of all customers of Internet-R-Us, a nationwide ISP, rated the service as being poor due to slow response times when surfing the Web during peak hours. As a result, the head of operations has suggested investing in new servers that the manufacturer claims will substantially improve the response times and lead to fewer dissatisfied customers. Since this is a major investment, she has decided to first try the new servers in ten randomly selected cities served by Internet-R-Us. She will only agree to install the new servers nationwide if the ten city trial provides data that strongly supports the server manufacturer’s claim.

Let $\pi$ represent the proportion of customers who feel that the service is poor with the new servers. An appropriate hypothesis test setup for this situation is:

The manufacturer’s claim is that a smaller proportion of customers will be dissatisfied. That’s the alternative!

(a) $H_0: \pi = 0.37$ vs. $H_1: \pi \neq 0.37$
(b) $H_0: \pi \leq 0.37$ vs. $H_1: \pi > 0.37$
(c) $H_0: \pi < 0.37$ vs. $H_1: \pi \geq 0.37$
(d) $H_0: \pi \geq 0.37$ vs. $H_1: \pi < 0.37$
(e) $H_0: \pi > 0.37$ vs. $H_1: \pi \leq 0.37$
9. In hypothesis testing, a Type II error is made if we: This is just the definition.

(a) Fail to reject the null hypothesis when in fact it is false
(b) Take too few samples to reach a meaningful conclusion
(c) Take a sample which is not random, leading to the possibility of a biased conclusion
(d) Reject the null hypothesis when in fact it is true

10. Using the current system, it takes an average of $\mu = 2.1$ minutes to serve a customer at the drive-up window at First Midstates Bank, with a standard deviation of $\sigma = 0.7$ minutes. Most of the time is devoted to the teller using the bank’s computer system for a variety of transaction-related activities.

The bank is considering switching to a touch screen system which the bank’s IS department head claims will reduce the average service time. The switch would be expensive, so the bank isn’t going to switch unless evidence from a series of trial runs indicates that switching to touch screens does indeed reduce service times. To test the IS department head’s claim, several branches are set up with touch screens and $n = 50$ randomly selected service times are observed. The appropriate test setup is: $H_0: \mu_{new} \geq 2.1$ vs. $H_1: \mu_{new} < 2.1$.

Assume that if the new system proves itself the bank will switch.... otherwise, they won’t. What is the implication, to the bank, of making a Type I error?

Follows directly from the definition of Type I error.

(a) They’d conclude that switching to touch screens would indeed reduce the average service time when in fact it would not. They’d switch to the touch screens and in the long run would learn that there was no value in switching.
(b) They’d conclude that switching to touch screens would not reduce the average service time when in fact it would. They’d miss the opportunity to improve the quality of their customer service.

Consider testing $H_0: \mu \leq 150$ vs. $H_1: \mu > 150$ based on a random sample of size $n = 50$ from a population having known standard deviation $\sigma = 12.0$.

11. At the $\alpha = 0.01$ level of significance, this test would reject the null hypothesis if the calculated $z$-statistic is:

You need the upper 1% tail cutoff on the $z$-distribution. Sketch it: the ‘inside’ area is 0.49. That corresponds, on the $z$-table, to $z = 2.33$. 

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; -1.28</td>
<td>&gt; 1.28</td>
<td>&lt; -1.645</td>
<td>&gt; 1.645</td>
<td>&lt; -1.96</td>
<td>&gt; 1.96</td>
<td>&lt; -2.33</td>
<td>&gt; 2.33</td>
<td>&lt; -2.575</td>
<td>&gt; 2.575</td>
</tr>
</tbody>
</table>
12. If the test is performed and the resulting sample mean is $\overline{X} = 152.9$, then the value of the calculated z-statistic is:

$$Z = \frac{152.9 - 150}{12/\sqrt{50}} = 1.71.$$ 

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.24</td>
<td>0.30</td>
<td>0.34</td>
<td>0.39</td>
<td>0.45</td>
<td>1.71</td>
<td>2.12</td>
<td>2.42</td>
<td>2.77</td>
<td>3.18</td>
</tr>
</tbody>
</table>

13. In March of 2000 an extensive nationwide survey of venture capitalists conducted by *Entrepreneur* magazine indicated that 80% would consider investing funds in a hypothetical start-up e-business company where the company’s management had no previous e-business experience, were all under 30 years old, and wore jeans and sandals to work. Considering the recent drop in value on both the New York Stock Exchange and NASDAQ, a researcher for the magazine believes that this percentage may have decreased. He selects a random sample of $n = 120$ executives and observes that 91 of the 120 would currently consider the investment described above.

Let $\pi_{\text{new}}$ represent the current proportion of all venture capitalists who would consider this investment. The researcher sets up the test as: $H_0$: $\pi_{\text{new}} \geq 0.80$ vs. $H_1$: $\pi_{\text{new}} < 0.80$.

The value of the calculated z-statistic is equal to:

$\text{Sample proportion } = p = \frac{91}{120} = 0.7583$. Calculate

$z = \frac{0.7583 - 0.80}{\sqrt{0.8(0.2)/120}} = -1.14$.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3.09</td>
<td>-2.97</td>
<td>-2.81</td>
<td>-2.74</td>
<td>-1.66</td>
<td>-1.14</td>
<td>-1.02</td>
<td>-0.91</td>
<td>-0.77</td>
<td>-0.68</td>
</tr>
</tbody>
</table>

14. Consider testing $H_0$: $\pi = 0.80$ vs. $H_1$: $\pi \neq 0.80$ at the $\alpha = 0.05$ level of significance. A sample is taken, the z-statistic is calculated, and the corresponding p-value is found to equal 0.035. The test conclusion is to:

*Since$p < alpha$ we reject $H_0$. That’s the ‘universal decision rule.’*

(a) Reject $H_0$
(b) Do not reject $H_0$
(c) Cannot be determined from the information given
15. Consider testing $H_0$: $\mu = 40$ vs. $H_1$: $\mu \neq 40$ at the $\alpha = 0.01$ level of significance. A sample of size $n = 50$ is taken and a 99% confidence interval for $\mu$ is calculated as (33.4, 37.1). This means that we should.....

_Solution will be (a) or (b), depending on your exam version. The rule is to reject $H_0$ if the CI for the mean EXCLUDES the value specified in $H_0$, otherwise don’t reject. On this version the CI excludes 40 so we reject._

(a) Reject $H_0$
(b) Do not reject $H_0$
(c) Cannot be determined from the information given

16. The average GMAT score for applicants to the full-time MBA program at a large Midwestern university has historically been 620. The program has recently moved into _Newsday_’s top 100 MBA program list, and an admissions officer would like to test to see if this positive media exposure has led to an increase in the quality of the applicants. She sets up the test: $H_0$: $\mu \leq 620$ vs. $H_1$: $\mu > 620$ where $\mu$ represents the mean score of current applicants. The test is based on a random sample of $n = 22$ recent applicants and is to be conducted at the $\alpha = 0.01$ level of significance. Scores are assumed to be normally distributed. If the sample results are: $X = 642.0$ and $s = 42.0$, the test conclusion is to:

*It’s a $t$-test (with 21 DF) since the standard deviation $\sigma$ is not known and is estimated by $s$. You reject in the upper 1% tail.*

(a) reject $H_0$ since $z = 2.46 > 1.96$
(b) reject $H_0$ since $z = 2.46 > 2.33$
(c) reject $H_0$ since $t = 2.46 > 1.717$
(d) reject $H_0$ since $t = 2.46 > 1.721$
(e) do not reject $H_0$ since $t = 2.46 < 2.472$
(f) _do not reject $H_0$ since $t = 2.46 < 2.518_

17. Consider testing: $H_0$: $\mu = 25.0$ vs. $H_1$: $\mu \neq 25.0$ where $\mu$ represents the mean of a normally distributed population having known standard deviation $\sigma = 4.8$. A sample of size $n = 30$ is taken and the resulting sample mean is $\bar{X} = 23.6$. Based on this result, we calculate a $z$-statistic of $z = -1.60$. The $p$-value corresponding to this result is:

_Since it’s a two-sided test, the $p$-value is twice the smaller of the area to the right and to the left of the $z$-statistic. Sketch it if you need to! The smaller area is the tail to the left of -1.60, which has an area of $0.5000 - 0.4452 = 0.0548$. Double it to obtain 0.1096._

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0548</td>
<td>0.0721</td>
<td>0.0968</td>
<td><strong>0.1096</strong></td>
<td>0.1292</td>
<td>0.1442</td>
<td>0.1788</td>
<td>0.1936</td>
<td>0.2584</td>
<td>0.3576</td>
</tr>
</tbody>
</table>
18. Consider a test involving the mean of a normal distribution: \( H_0: \mu \geq 100 \) vs. \( H_1: \mu < 100 \). The test is to be based on a sample of size \( n = 12 \). The sample is taken and the following sample results are obtained: \( \bar{X} = 95, s = 10.2 \). These results lead to a calculated \( t \)-statistic of \( t = -1.70 \).

The \( p \)-value corresponding to this sample result is….

**One-sided test with rejection in the lower tail, so the \( p \)-value is the area to the LEFT of the \( t \)-stat. Sketch it! Note that that’s the same as the area to the right of \( +1.70 \). Going to the \( t \)-table with 11 DF, I see that this tail area must be between 0.05 and 0.10. It’ll be different for different versions!**

- (a) less than 0.005
- (b) between 0.005 and 0.01
- (c) between 0.01 and 0.02
- (e) between 0.01 and 0.025
- (f) between 0.02 and 0.05
- (g) between 0.025 and 0.05
- (i) between 0.10 and 0.20
- (j) greater than 0.20
- (k) cannot be determined based on the information given
- (d) less than 0.01
- (h) between 0.05 and 0.10