Short Answer

Concepts to Understand: Exploratory Research

1. Definition and need for a research design

The design is the researcher’s plan of action for conducting the study. The design is needed so that there are guidelines for how to proceed with the study.

2. Exploratory research

Before testing an explicit hypothesis, the underlying patterns in the data need to be explored and initially understood at least at a general level. Exploratory research seeks to just “explore” to uncover patterns and relationships.

3. Qualitative vs quantitative research

Qualitative research produces non-numerical results, focused on gathering information through techniques such as open-ended questionnaires, interviews, and observations, and is usually exploratory. Quantitative research generates numerical results amenable to statistical hypothesis testing.

4. Focus groups and their primary purpose

A small group of people, in the presence of a moderator, discuss a potential product, and generally provide feedback about issues related to product and company. Surveys and other techniques provide marketers useful information, but sometimes just talking with your customers or potential customers is one of the best ways to learn about customer perceptions.

5. Focus group vs. an immersion group

This technique is used to understand the subconscious ideas of a brand. Examples of this technique include asking participants what their thoughts or feelings are on a particular brand or product. They are then asked why they chose that feeling or thought about the brand.

6. Projective technique with an example

Projective techniques delve into the subconscious perceptions of a product or brand, to uncover deeper motivations and feelings that perhaps the customer cannot clearly articulate without some kind of prompting. Word association and sentence completion are two examples.
Concepts to Understand: Data Analysis

1. *What is the relation between the null hypothesis and the alternative hypothesis?*

The alternative hypothesis is the range of values not specified by the null hypothesis.

2. *Mean and standard deviation of the distribution of the sample mean*

You only observe one sample mean, \( m \), but, hypothetically, there are as many possible sample means as there are possible samples. The mean of this distribution of sample means is the population mean. The standard deviation of this hypothetical distribution is the standard error, the standard deviation of the data divided by the square root of the sample size.

3. *Relation of the \( t \)-cutoff, \( t.025 \), to the obtained \( t \)-value*

The \( t \)-cutoff is a mathematical property of the \( t \)-distribution, which exists as a mathematical statement without any data. The obtained \( t \)-value is how many estimated standard errors your sample mean is from the hypothesized mean. If the obtained value exceeds the cutoff value of about 2, an unusual event has occurred given your assumption of the null hypothesis, so reject the null value.

4. *Relation of the \( p \)-value to the \( t \)-value*

If the null hypothesis is true, what is the probability of the obtained \( m \) being as far or farther from the hypothesized mean? This is the \( p \)-value. The actual distance from the mean is the \( t \)-value, expressed in terms of estimated standard errors. If this number is large, then the probability is high, so the result is consistent with the hypothesized value. If this result is small, then the \( m \) is a low probability event assuming the null hypothesis.

5. *Relation of size of the \( p \)-value to the decision regarding the null hypothesis.*

The \( p \)-value is the probability of the sample value being as far or further from the null value, on either side of the null value, assuming the null value is the true mean. If \( p \)-value is less than the criterion of \( \alpha = 0.05 \), then an unlikely event occurred given the truth of the null, so the null value is rejected. Otherwise, it is supported, but not proved.

6. *Relation of the size of the confidence interval to the size of the acceptance region for the corresponding hypothesis test*

Both regions are the same size, about ± two estimated standard errors, to defined 95% range of variability of the sample mean over hypothetical repeated samples. The confidence interval is centered about the sample mean and the acceptance region is centered about the hypothesized mean.
Worked Problems

Note: Remember that you do not need to memorize R commands for the tests. There is nothing about doing R on the tests. Instead the R output, the same as below but with different data, will be given to you as part of the midterm. You answer exactly the same questions as below on the midterm.

1. Interpretation of the confidence interval from last week.

The item for which the responses are analyzed:

   I want to learn more of how to apply statistical techniques to marketing research.

a. With 95% confidence, the true population mean attitude of PSU Marketing Research 460 students for learning data analysis applied to marketing research, as assessed on a 100 point scale from maximum disagreement (0) to maximum agreement (100), is somewhere between 57.5 to 71.7. There is a general level of moderate agreement with a desire to learn how to do data analysis regarding marketing research.

b. [Depends on your answer.]

c. [Comparison depends on your answer.]

Note: Interpretation of statistical output is one of the key aspects of doing marketing research, which also means that interpretation is emphasized on the tests. You should always compare your answer to each homework question with my answer, but here we formally do the comparison as part of the homework.

2. Real survey data regarding satisfaction from The SFG restaurant in Dallas, TX.

The variable with name x23 contains the responses to the following item.

How likely are you to return to the SFG in the future?

The responses are on a scale of the first 7 integers, which vary from Definitely Will Not Return, coded as a 1, to Definitely Will Return, coded as a 7.

First, read the data into R:

   > mydata <- Read("http://web.pdx.edu/~gerbing/data/SFGsfg.csv")

Preliminaries

a. Identify the response variable (i.e., the variable of interest) in this analysis. Identify by its

   i. Meaningful description [a phrase or sentence]

   The item in a survey: How likely are you to return to the SFG in the future?
ii. Variable name as it is referred to in the data set for analysis [see above] x23

b. Specify the type of values that are in the data file for the response variable and provide an example data value. [look at the data file, or the output from the Read statement]

c. How many data values are present for the responses? Any missing data?
From the Read function, isolate just the variable of interest:

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Type</th>
<th>Missing</th>
<th>Unique</th>
<th>Values</th>
<th>Values</th>
<th>Values</th>
<th>First and last values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x23</td>
<td>integer</td>
<td>253</td>
<td>0</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>... 5 4 3</td>
</tr>
</tbody>
</table>

So, 253 responses to the item and no missing values. That is, all who responded to the survey responded to this item.

d. skip

e. If needed, evaluate the assumption of the normality of sample mean.
The sample size is well larger than \( n = 30 \), so the sample mean, \( \mu \), would be approximately normal over repeated sampling. No evaluation of the histogram of the data values for this variable is needed, because regardless of how normal is the data, or not, the sample mean, \( \mu \), is normal.

Computer output.

```
> tt.brief(x23, mu0=4)

<table>
<thead>
<tr>
<th>t-cutoff: tcut</th>
<th>1.969</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Error of Mean:</td>
<td>SE = 0.05</td>
</tr>
<tr>
<td>Hypothesized Value H0:</td>
<td>( \mu = 4 )</td>
</tr>
<tr>
<td>Hypothesis Test of Mean:</td>
<td>t-value = 3.115, df = 252, p-value = 0.002</td>
</tr>
<tr>
<td>Margin of Error for 95% Confidence Level:</td>
<td>0.10</td>
</tr>
<tr>
<td>95% Confidence Interval for Mean:</td>
<td>4.06 to 4.27</td>
</tr>
</tbody>
</table>
```

f. Estimated standard error of sample mean. Show calculations.
The standard error estimated from the data is

\[
s_m = \frac{s}{\sqrt{n}} = \frac{0.85}{\sqrt{253}} = 0.053
\]
Hypothesis Test

g. skip

h. Cutoff (critical) values

The normal curve cutoff for the number of true, population standard deviations on either side of the population mean that contains 95% of the distribution is 1.96. A true fact always, but we must use the $t$-distribution to calculate probabilities, so we will have a cut-off value *slightly* larger than 1.96. We pay a penalty for not knowing the population standard deviation, using its estimate (from the data) instead. That penalty is a less precise estimate of the population mean, that is, a wider confidence interval.

The output of tt.brief provides the cutoff (critical) value that defines the 95% range of variation.

| t-cutoff: $t_{cut} = 1.969$ |

i. Null and alternative hypotheses

Null hypothesis: $\mu = 4$, that is, the population mean response to $x_{23}$ is neutral, 4

Alternative hypothesis: $\mu \neq 4$, that is, whatever is not the null value

j. Explain in words the calculation of this obtained $t$-value, can include the formula but explain what it means

The test statistic is the observed $t$-value, the number of estimated standard errors that separates the sample mean from the hypothesized value.

The standardized distance of the sample mean, $m$, from the hypothesized mean, $\mu_0$, follows.

$$t_m = \frac{m - \mu_0}{s_m} = \frac{4.166 - 4.0}{0.053} = 3.115$$

The sample mean of 4.166 is 3.115 estimated standard errors above the hypothesized mean of 4.0.

k. Present the $p$-value for the hypothesis test and the related statistical decision with $\alpha = 0.05$.

$p$-value $= 0.002$.

*If* the null hypothesis of $\mu = 4$ is true, then the probability of obtaining a sample mean that is 0.17 units or beyond 4, both above and below, is only 0.002.

l. Statistical decision.
\( p\text{-value} = 0.002 < \alpha = 0.05 \), so the statistical decision is to reject the null hypothesis. An event with a probability of only 0.002 is sufficiently unusual so as to render the assumption of \( \mu = 4 \) as unreasonable.

m. **Interpret** the hypothesis test.

Apparently the true mean response to this Likert item is different from 4, probably above 4 because the corresponding sample mean is above 4. Any value larger than 4 indicates that an intention to return to the restaurant, so, on average, there is at least some tendency to return.

n. skip

**Confidence Interval**

o. What the obtained confidence interval estimates

The confidence interval of the mean estimates the unknown population mean attitude regarding the likelihood to return to the restaurant, the SFG.

p. Explain in words the calculation of this obtained margin of error, can include the formula but explain what it means

The distance from the sample mean \( m = 4.166 \) to either end of the confidence interval is the margin of error, \( E \), which is \( t_{0.025} = 1.97 \) standard errors. One standard error is

\[
sm = \frac{s}{\sqrt{n}} = \frac{0.85}{\sqrt{253}} = 0.053
\]

So 1.97 standard errors is,

\[
E = (t_{0.025})(sm) = 1.97(0.053) = 0.105
\]

or, express as half the width of the confidence interval,

\[
(4.27 - 4.06)/2 = 0.105
\]

q. Explain in words the calculation of this obtained confidence interval, can include the formula but explain what it means

The confidence interval is the 95% range of sampling variation, plus and minus the margin of error, about the sample mean.

Lower Bound: \( m - E = 4.166 - 0.105 = 4.06 \)
Upper Bound: \( m + E = 4.166 + 0.105 = 4.27 \)

r. **Interpret** the confidence interval.

With 95% confidence, the true average response for customers of the SFG to the 7-pt Likert response format is in the Will Return range, somewhere between 4.06 and 4.27, where 4 represents Neutral and 7 represents Definitely Will Return.
Relation of HT to CI

t. **Relate** the findings of the confidence interval and the hypothesis test.

Both forms of statistical inference provide the same conclusion: The true mean, $\mu$, whatever its value, is larger than the hypothesized value of $\mu = 4$. The value of 4 is outside of the confidence interval and the value of 4 is rejected by the hypothesis test.

u. What is the **managerial relevance** of these findings?

The good news is that, on average, there is a tendency, on average, to wish to return to the SFG. The bad news is that this tendency, although statistically significant, is weak, barely above 4. And many specific customers are not desiring to return.

3. A market survey of 386 people has shown that people intend to spend an average of $14.73 each for your product next year, with a standard deviation of $2.12. The goal was to achieve an average of $15. Was the goal plausibly achieved?

ą. What does the analysis of just descriptive statistics imply for answering the question: Was the goal plausibly achieved?

According to the obtained sample mean, $m = 14.83$, the goal was not achieved in this particular sample. Descriptive statistics cannot answer this question for the population as a whole which is of the primary interest since this one particular sample will not ever be obtained again.
b. Now answer the question with inferential statistics: Hypothesis test

\[ t_m = \frac{m - \mu_0}{s_m} = \frac{14.83 - 15}{0.108} = -1.58 \]

This \( t \)-value yields a \( p \)-value of -1.58.

\[ p\text{-value} = 0.116 > \alpha = 0.05 \]

Fail to reject the value of \( \mu = \$15 \), so no average difference detected from \$15.

c. Now answer the question with inferential statistics: Confidence interval

The confidence interval follows.

Lower Bound: \( m - E = 14.830 - 0.212 = 14.62 \)

Upper Bound: \( m + E = 14.830 + 0.212 = 15.04 \)

With 95% confidence, the true average of intention to spend is somewhere between \$14.62\ and \$15.04.\n
The goal of a mean of \$15\ is a plausible value because it is in the confidence interval of the mean. That conclusion does not mean that the goal has been achieved because values less than \$15\ are also in the confidence interval. But at least it is a plausible value.

d. Relate the confidence interval and hypothesis test to each other.

The hypothesis could not reject the value of \( \mu = 15 \) and the same value is within the confidence interval so it is plausible.

e. What is the managerial relevance of these findings?

\$15\ is a plausible value for the intended amount of money people will spend on this product next year. However, the confidence interval just barely includes \$15, and so although plausible, there is a distinct possibility that the true mean value is below \$15. Only the analysis of a larger sample size will yield a narrower confidence interval, and thus a more precise estimate, that may, or may not, include \$15.

f. Forget the actual confidence interval you obtained. Pretend the confidence interval was from 14.1 to 14.9? What do you conclude the goal of achieving an average of \$15?\n
No values of a mean of \$15\ are plausible and all plausible values are below \$15, so the goal was not achieved.

g. Forget the actual confidence interval you obtained. Pretend the confidence interval was from 15.1 to 15.9? What do you conclude the goal of achieving an average of \$15?\n
No values of a mean of \$15\ are plausible and all plausible values are above \$15, so the goal was achieved.