Full CREDIT on Confidence Interval Problems

1) Conditions stated & checked (3 pts)

2) State:
"Conditions met, use Normal Model for 1 prop Z Int" (2 pts)

3) Formula with numbers in it (2 pts)

4) Interval (a, b) (1 pt)

5) Sentence interpretation (2 pts)

Tests Of Significance

- **Confidence intervals** are **one** of the **two** most common types of statistical inference. The second type of inference is called a ***Test of Significance*** (also called ***Hypothesis Test***).
- ***Use a confidence interval when your goal is to estimate a population parameter.***
- ***Use a hypothesis test when you wish to assess a claim concerning a population using the evidence provided***

Full CREDIT on Test of Significance

1) Hypotheses (2 pts)

2) Conditions stated and checked (3 pts)

3) Statement:
"Conditions met, Normal Model, 1 prop. Z test" (2 pts)

4) Mechanics: Z-score formula, with numbers filled in (2 pts)
and Z-score correctly calculated (1 pt)

5) Mechanics: P-value notation (2 pts) and correct value (1 pt)

6) Conclusion: 2 sentences (4 pts total)

The Reasoning of Tests of Significance

- Statistical tests are based on determining the probability of obtaining a particular outcome if we repeated an experiment many times using a specified sample size – sampling distributions.
- The underlying question posed by a test of significance is: **What's the probability that I received such data just by chance?**
- If a hypothesis seem consistent with what we would expect from natural sampling variability, we retain our hypothesis; however, if the results are far from what we would expect, we reject the hypothesis.

Introduction to Inference

Hypothesis Testing:
Drawing Conclusions

3

The Reasoning of Tests of Significance

- Of course, when the probability is very low, it may be that the hypothesis is actually true and we just witnessed something that is extremely unlikely by mere chance. When it comes to data though, statisticians don't believe in miracles. Instead, they say that if the data is unlikely **enough**, then it is OK to reject the hypothesis.

The Reasoning of Tests of Significance

- Consider the following example:
  - A manufacturer creates huge pieces of metal called ingots. Unfortunately, the ingots must be produced completely crack-fee. 80% of the ingots that they make are crack-free – this is unacceptable because of the high cost of recycling the defective ingots. Engineers develop a new process that is suppose to reduce the cracking proportion slightly. Using the new process, an SRS of 400 ingots were created and only 17% of them cracked. **Should the manufacturer declare victory? Has the cracking rate really decreased or was 17% just pure luck?**
- Significance tests try to answer questions like this. Others may be: Has the president's approval rating changed since last month? Has teenage smoking decreased in the last 5 years? Is the global temperature really increasing?

p-value & Statistical Significance

- ***p*-value** - (upper critical value, z*) is the probability (computed assuming that $\mathbf{H_0}$ is true) that the test statistic would take a value as extreme or more extreme than that actually observed. It is the probability of obtaining a test statistic equal to or more extreme than the observed value if $\mathbf{H_0}$ is true.
- Also called the **"observed" level of significance**.
- The smaller the ***p*-value** is, the stronger is the evidence against $\mathbf{H_0}$ provided by the data.

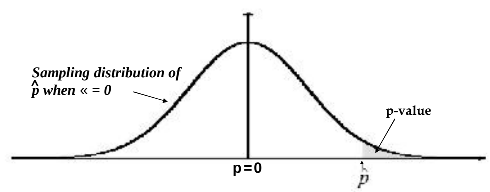
The Null Hypothesis

- A hypothesis is something that is not proved, but assumed to be true for the purpose of argument.
- In order to begin a hypothesis test, we need a hypothesis that we consider to be true – we call this the **Null Hypothesis**. The null hypothesis, which we denote $\mathbf{H_0}$, specifies a population model parameter of interest and proposes a value for that parameter. When we create a hypothesis test, we develop a dichotomy where either the hypothesis is true or the **Alternative Hypothesis, $\mathbf{H_a}$**, is true. The following demonstrates the notation that we will use:

  The Null Hypothesis: $\mathbf{H_0: p = p_0}$ **(original proportion)**

  Alternative Hypothesis $\mathbf{H_a: p > p_0}$ or $\mathbf{H_a: p < p_0}$ or $\mathbf{H_a: p \neq p_0}$

p-value & Statistical Significance



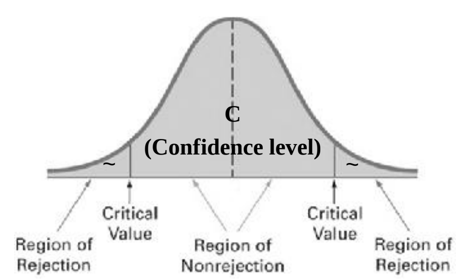
Compare the p-value with α (level of significance)

- If p-value < α, reject $\mathbf{H_0}$.
- If p-value > α, do not reject $\mathbf{H_0}$.
- "If p-value is low, then $\mathbf{H_0}$ must go."

Outline of a test

- Describe the effect you are searching for in terms of a **population parameter** like the mean μ or p.
  - **Never** state a hypothesis in terms of a sample statistic like $\hat{p}$
- The **null hypothesis $\mathbf{H_0}$** is the statement that this effect is not present in the population, whereas the **alternative hypothesis $\mathbf{H_a}$** states that it is.
- If the sample statistic is far from the parameter value stated by $\mathbf{H_0}$, then we have evidence **to REJECT $\mathbf{H_0}$** and if the sample is relatively close to the parameter value, we **FAIL TO REJECT $\mathbf{H_0}$**
- The **p-value** says how unlikely (or likely) a result at least as extreme as the one observed would be if $\mathbf{H_0}$ were true. Results with small P-values would rarely occur if the $\mathbf{H_0}$ were true.

p-value & Statistical Significance



Interpreting a *p*-value

Could random variation alone account for the difference between the null hypothesis and observations from a random sample?

- A **small *p*-value** implies that random variation because of the sampling process alone is not likely to account for the observed difference.
- With a **small *p*-value**, we **reject H_0**. The true property of the population is **significantly** different from what was stated in H_0.

Thus, small *p*-values are strong evidence AGAINST H_0.

But how small is small…?

Does the packaging machine need revision?

- H_0: μ = 227 g versus H_a: $\mu \neq$ 227 g
- What is the probability of drawing a random sample such as yours if H_0 is true?

$\bar{x} = 222g \quad \sigma = 5g \quad n = 4$

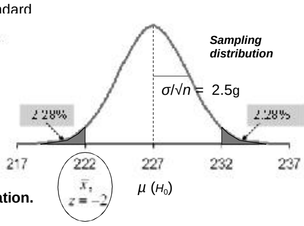
$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{222 - 227}{5/\sqrt{4}} = -2$$

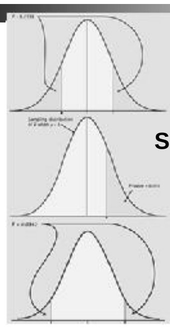
From Table A, the area under the standard normal curve to the left of z is **0.0228**.

Thus, ***p*-value = 2*0.0228 = 4.56%.**

The probability of getting a random sample average so different from μ is so low that we reject H_0.

➔**The machine does need recalibration.**



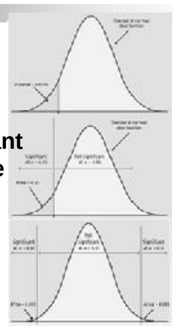


p = 0.2758

p = 0.1711

p = 0.0892

Significant *p*-value ???



p = 0.0735

p = 0.05

p = 0.01

When the shaded area becomes very small, the probability of drawing such a sample at random gets very slim. Oftentimes, a *p*-value of 0.05 or less is considered **significant:** The phenomenon observed is unlikely to be entirely due to chance event from the random sampling.

Significance Tests and OJ Simpson

- If you think about the OJ Simpson case, it brings to mind several ideas that are parallel to significance tests. First, OJ was considered innocent until proven guilty beyond a reasonable doubt. For significance, our Null hypothesis is true until rejected beyond a "reasonable doubt." Second, it was up to the prosecutor to have enough evidence to convict OJ. Similarly, it is our burden to provide enough evidence to reject the Null Hypothesis. However, since the prosecutor was not able to provide enough evidence to convict him, they declared him "not guilty." Likewise, if we do not have enough "evidence" to reject the Null Hypothesis, we "fail to reject the Null Hypothesis." Note: Just as the jury did not say that OJ was innocent, we do not say that the Null Hypothesis is true!!!

Decision Making Rules in Hypothesis Testing

- **Rule one:** If the p-value is less than or equal to the significance level then reject the null hypothesis and conclude that the research finding is statistically significant.
- **Rule two:** If the p-value is greater than the significance level then you "fail to reject" the null hypothesis and conclude that the finding is not statistically significant.

Significance Level, ~

- The decisive value of P is called the **Significance level ~**. The value of α is related to the confidence level C by ~ **= 1 – C**

  Therefore, a **5% significance level corresponds to a 95% Confidence level**.

  i.e $\alpha = 1 - 0.95 = 0.05$

  If the p-value is less than or equal to α, then the result is considered significant at level α, p-value ≤ α, the result is significant.

  What z* would represent 95% confidence for a one-sided-test? **1.645**

  What about 95% confidence for a two-sided test? **1.96**

Tests for a Population Parameter

Steps to follow:

- **Step 1**: Identify population **Parameter**, state the null and alternative **Hypotheses**, determine what you are trying to do (and determine what the question is asking).
- **Step 2**: Verify the **Assumptions** by checking the conditions.
- **Step 3**: If conditions are met, **Name the inference procedure** and carry out the inference:
  1) **C**alculate the **Test statistic** and
  2) **Obtain the p-value**
- **Step 4**: **Make a decision** (reject or fail to reject H_0). **State your conclusion** in context of the problem using p-value.
- If you haven't noticed, this spells out **PHANTOMS**

z Test for a Population Parameter

- To test the hypothesis **H_0: $p = p_0$** based on an SRS of size n from a population with unknown proportion p and standard error **SE_p**, compute the **one-proportion z statistic**

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

← **This is your Test statistic!!!**

In terms of a variable Z having the standard normal distribution, the P-value for a test of H_0 against

H_a: $p > p_0$ is $P(Z \geq z)$ **(one sided to the right)**

H_a: $p < p_0$ is $P(Z \leq z)$ **(one sided to the left)**

H_a: $p \neq p_0$ is $2P(Z \geq |z|)$ **(two sided)**

These p-values assume a normal distribution by the CLT and can be applied if the ***sample is large enough***

Examples

Write the hypotheses, in symbols and in context, then determine if it's a one-side or two sided test.

11c)

- The sample of customers may not be representative of all customers, so we will proceed cautiously. A Normal model can be used to model the sampling distribution of the proportion. (SHOW WORK)

11d)

- If his dowsing has the same success rate as standard drilling methods, there is more than a 23% chance of seeing results as good as those of the dowser, or better, by natural sampling variation.

Tests with Fixed Significance Level, ~

- A level of significance α says how much evidence we require. In terms of P-value, the outcome of a test is significant at level α if P-value ≤ α.

α = 1 – C

  - Where C is the confidence
- **When Making a decision, use the following:**
  - **We reject H_0 if p-value ≤ α**
  - **We fail to reject H_0 if p-value > α**

- With an SRS of size n, from a N(p_0, SE), p = unknown. We want to test the hypothesis that p has a specified value p_0.

The null hypothesis is

H_0: $p = p_0$.

The test is based on the sample proportion, $\hat{p}$.

Normal calculations require standardized variables $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$

It's important to note that we use p_0, not $\hat{p}$ to calculate SE_p

This **one-proportion z statistic** has the standard normal distribution when **H_0** is true. If the alternative is one-sided on the high side: Ha: $p > p_0$

then the P-value is the probability that a standard normal variable **Z** takes a value at least as large as the observed **z**. That is, **$P = P(Z \geq z)$**

Tests from CI's

- A two-sided test at significance level α can be carried out directly from a confidence interval with confidence level C = 1 - α.

CI and Two-Sided Tests ($p \neq p_0$)

- A level α two-sided significance test rejects a hypothesis H_0: $p = p_0$ exactly when the value p_0 falls outside a level 1 - α confidence interval for p.