Chapter 2 Review of Statistical Inference

Section 2.1 and 2.2 Statistical Inference (Review)

Population and sample

Sample mean: $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

Sample variance:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \quad \left(\text{unbiased estimator of } \sigma^2\right)$$

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \quad \left(\text{biased estimator of } \sigma^2\right)$$

Sample standard deviation:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$$

$$S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$$

Order statistics

Empirical distribution function

Let $x_1, x_2, \cdots, x_n$ be a random sample.

$S(x) =$ fraction of x_i's that are less than or equal to x $\left(-\infty < x < \infty\right)$

$$S(x) = \begin{cases} 0 & x < x_{(1)} \\ 1/n & x_{(1)} \le x < x_{(2)} \\ 2/n & x_{(2)} \le x < x_{(3)} \\ & \vdots \\ (n-1)/n & x_{(n-1)} \le x < x_{(n)} \\ 1 & x \ge x_{(n)} \end{cases}$$

Data: 4, 2, 1, 7, 10, 6

Statistical inference about population parameters

- Point estimation
- Interval estimation
- Testing hypotheses

Statistical inference about population distributions

Statistical procedures used in elementary statistics:

Confidence interval for μ:

$$\left(\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right)$$ Requirement: large sample size

$$\left(\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right)$$ Requirement: normality

Statistical test about μ: $\left(H_0 : \mu = \mu_0 \quad \text{vs. A.} H_a : \mu > \mu_0 \quad \text{B.} H_a : \mu < \mu_0 \quad \text{C.} H_a : \mu \neq \mu_0\right)$

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$ Requirement: large sample size

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$ Requirement: normality

Confidence interval for p:

$$\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$ Requirement: large sample size

Statistical test about p: $(H_0 : p = p_0 \quad \text{vs. A.} H_a : p > p_0 \quad \text{B.} H_a : p < p_0 \quad \text{C.} H_a : p \neq p_0)$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$ Requirement: large sample size

Section 2.3 Hypothesis Testing (Review)

Alternative Hypothesis (Research Hypothesis)

The alternative hypothesis is the hypothesis that the researcher wishes to support.

Null Hypothesis

The null hypothesis is the negation of the alternative hypothesis

Test Statistic

The test statistic is the sample statistic used to determine whether or not to reject null hypothesis.

Example Suppose the building specifications in a city requires that the average strength of residential sewer pipe be more than 2400 pounds per foot of length. Each manufacture who wants to sell pipe in the city must demonstrate that its product meets the specification. Does a manufacturer's pipe meet building code?

Define μ as the average strength of residential sewer pipes produced by the manufacturer.

Example We want to show that the average hourly wage of construction workers in the state of California is different from \$14, which is the national average.

Define μ as the average hourly wage of construction workers in the state of California.

Example A milling process currently produces an average of 3% defectives. We are interested in showing that a simple adjustment will decrease p, the proportion of defectives produced in the milling process.

Rejection Region

The rejection region of a statistical test is the set consisting of values that will lead to the null hypothesis being rejected.

Decision and Conclusion

Decision to (or not to) reject the null hypothesis based on a comparison of test statistic to rejection region.

Type I Error

Rejecting the null hypothesis when it is true.

Type II Error

Failing to reject the null hypothesis when it is false.

Type I Error vs. Type II Error

	True State of Nature	
	H_0 True	H_0 False
Reject H_0	Type I Error	Correct Decision
Fail to reject H_0	Correct Decision	Type II Error

α – Probability of committing a Type I error

β – Probability of committing a Type II error

Level of Significance (Significance Level)

The upper bound of the probability of committing Type I error.

Commonly used level of significance:

0.1, 0.05, 0.025, 0.01, 0.005, …

Example

It is known that there are three balls in a bag. The balls may have red or green color.

H_0 : There are more red balls then green balls in the bag. (2 or 3 red ball)

H_a : There are more green balls than red balls in the bag. (0 or 1 red ball)

One ball is drawn from the bag at random. Reject H_0 is the ball is green.

$\alpha = P_{H_0}\left(\text{Rejecting } H_0\right)$

Level of significance: maximum probability of rejecting H_0 when H_0 is true

$\beta = P_{H_a}\left(\text{Failing to reject } H_0\right)$

Example 1 (Page 96)

A certain machine manufactures parts. The machine is considered to be operating properly if 5% or less of the manufactured parts are defective. If more than 5% of the parts are defective, the machine needs remedial attention. A random sample of ten parts is selected to conduct the test.

Level of significance: maximum probability of rejecting H_0 when H_0 is true.

Level of significance $= P(T > 2 \mid p = 0.05)$

$\beta(p) = P(\text{Type II error}) = P(T \le 2 \mid p) = 1 - P(T \le 2 \mid p)$

Power $= 1 - \beta(p) = P(T > 2 \mid p)$

When $p = 0.1$,

power $= 1 - \beta(p)$

When $p = 0.2$,

power $= 1 - \beta(p)$

$p-$value of a test: the smallest significance level at which H_0 is rejected

Reject H_0 if p-value $< \alpha$.

Example 1 (P. 96) Continue

When the observed value of T is 3,

p-value $= P_{H_0}(T \geq 3)$

Formulas Used in Elementary Statistics To Find p-values:

Setting A $H_0 : \mu = \mu_0$

$H_a : \mu > \mu_0$

$p\text{-value} = P\left(Z \geq z_{(\text{obs})}\right)$

Setting B $H_0 : \mu = \mu_0$

$H_a : \mu < \mu_0$

$p\text{-value} = P\left(Z \leq z_{(\text{obs})}\right)$

Setting C $H_0 : \mu = \mu_0$

$H_a : \mu \neq \mu_0$

$p\text{-value} = 2P\left(Z \geq |z_{(\text{obs})}|\right)$

Example

A random sample of size 100 has been drawn from a population distribution. It is known that the sample means is 61 and the sample standard deviations are is 10. Find the p-value for the test $H_0: \mu = 60$ vs. $H_a: \mu > 60$.

Example

A bottling company distributes beer in bottles labeled 32 oz. The local Bureau of Weights and Measures randomly selects 50 of these bottles, measures their contents, and obtains a sample mean of 31 oz and a standard deviation of 0.75 oz. Is it valid at the 0.01 significance level to conclude that the bottling company is cheating consumers?

Example

The daily yield of a chemical plant, recorded for 50 days, possesses sample mean of 871 tons and standard deviation of 21 tons. Test the hypothesis that the average daily yield of the chemical is 880 tons per day against the alternative that it is either greater than or less than 880 tons per day. (α=0.05)