

Homework Due Friday, October 27:

p. 207: 5.21

p. 221: 5.26, 5.32,

Homework Due Friday, November 3:

p. 222: 5.33, 5.38 (you will need to compute the summary statistics \bar{Y} and s)

p. 228: 5.42, 5.43, 5.46

Exam III will be Monday, November 13

Covers: 4.13, 5.1 - 5.8 *including* 5.8

Some Review Problems:

p. 224: 5.36, 5.39

p. 240: 5.52, 5.53, 5.54

p. 253: 5.79, 5.80, 5.90, 5.94, 5.95 (a-d), 5.96

Chapter 5 Inferences about Population Central Values

- **Example** (from Ramsey, F.L. and Schafer, D.W., 1997. *The Statistical Sleuth*, p. 517). Obesity is known to be associated with increased risk of cardiovascular disease in western societies. Is this because of the strain of excessive weight or social stigma?
- A study, D.E. Crews. 1988. *Cardiovascular mortality in American Samoa*, Human Biology, **60**, 417-433, addressed this question by comparing the proportions of cardiovascular death among two samples of obese and non-obese American Samoan women.
- The proportion of deaths attributable to CV (between 1976 and 1981) were

$$\hat{\pi}_o = \frac{16}{2061} = 0.00776$$

and

$$\hat{\pi}_n = \frac{7}{1051} = 0.00666.$$

- What do these statistics say?

1. The incidence of CV death was small in both samples
2. $\hat{\pi}_n$ is $100(\hat{\pi}_n/\hat{\pi}_o) = 100 \times (666/776) = 85.8\%$ of $\hat{\pi}_o$. This is suggestive of a real difference, *but*
 - These estimates are based on relatively small numbers: 16 and 7. Can we trust them?
 - Our level of trust corresponds to the precision of these values 0.00776 and 0.00666 as estimates of the population death rates. That is, Crews is really after the rate of CV death rates in the population of *all* obese and non-obese American Samoan women, not just the 3112 in this study
 - To be more precise, we are interested in the difference $\hat{\pi}_o - \hat{\pi}_n$ and whether or not the true difference is $\pi_o - \pi_n$ is zero (or nearly zero), though we may prefer to ask whether the true ratio π_n/π_o is one (or nearly one)
 - In the first case, we are interested in the precision of the difference $\hat{\pi}_o - \hat{\pi}_n$; in the second, we are interested in the precision of the ratio $\hat{\pi}_n/\hat{\pi}_o$

Our **main objectives** are to develop formal statistical methods for

- 1) Estimation; e.g., what is the best estimate of μ and $\sigma_{\bar{y}}$?
- 2) Hypothesis testing; e.g., is it true that $\pi_o = \pi_n$? Specifically, what is the strength of evidence arguing against the hypothesis that $\pi_o = \pi_n$?

Estimation

- The usual *point estimate* of μ is \bar{Y}
- An *interval estimate* of μ is a set of hypothetical values for μ that are consistent with the data.

Confidence Intervals

- A confidence interval is an interval estimate with particular properties.
- We choose a small number α (most often $\alpha = 0.05$), and call the interval a $100(1 - \alpha)\%$ confidence interval. When $\alpha = 0.05$, we get a 95% confidence interval
- $1 - \alpha$ is the confidence coefficient
- A *confidence interval* is a set of values between two endpoints L (lower) and U (upper)

- The procedure (or process) of constructing a $100(1 - \alpha)\%$ CI will contain the true unknown mean μ $100(1 - \alpha)\%$ of the time, if the process is repeated many times.
- To determine a 90% confidence interval for μ , find the 5th and 95th percentiles of the distribution of \bar{Y} . The 5th percentile is L and the 95th percentile is U . 90% of distribution of \bar{Y} lies between L and U
- For example, suppose that we assume that the distribution of adult male heights is normal and has a standard deviation of $\sigma = 2$. A sample of $n = 25$ is selected and the sample mean height is $\bar{Y} = 69.6$.
- The point estimate of μ is $\bar{Y} = 69.6$.
- To determine a 90% confidence interval for μ , we find the 5th and 95th percentiles of the distribution of \bar{Y} , namely, $N(\mu, \sigma_{\bar{y}})$ where $\sigma_{\bar{y}} = \sigma / \sqrt{n}$ is the standard deviation of the distribution of \bar{Y}

Step 1) Find a 90% interval for a standard normal random variable. That is find $z_{0.05}$ and $z_{0.95}$ satisfying $P(z_{0.05} < Z < z_{0.95}) = 0.05$.

- Inspection of Table A gives $z_{0.05} = -1.645$; by symmetry, $z_{0.95} = 1.645$

Step 2) Reverse the normal transformation. Solve for $y_{0.05}$ within the standard normal formula

$$z_{0.05} = \frac{y_{0.05} - \mu}{\sigma_{\bar{y}}}$$

- Plug in the known quantities $z = -1.645$, $\bar{Y} = 69.6$ for μ , and $\sigma = 2$. This yields $\sigma_{\bar{y}} = 2 / \sqrt{25} = 0.4$, and

$$-1.645 = \frac{y_{0.05} - 69.6}{0.4}$$

- Solve for $y_{0.05}$:

$$L = y_{0.05} = 69.6 + 0.4 \times (-1.645) = 68.94.$$

- To find the 95th percentile, we repeat this calculation using $z_{0.95} = 1.645$:

$$1.645 = \frac{y_{0.95} - 69.6}{0.4}$$

$$\Rightarrow U = y_{0.95} = 69.6 + 0.4 \times 1.645 = 70.26.$$

- The 90% confidence interval for μ is 68.94 to 70.26.
- We say that we are 90% confident that the true mean is in the interval
- We cannot say that there is a 0.9 probability that μ is 68.94 between 70.26. In fact, $\mu = 69.7$ inches for adult male Americans, so the probability that μ is in the interval is 1.
- The confidence interval formed by L and U is
 - 1) all those values of μ that are consistent with the data. Any value outside the interval is not consistent with the data
 - 2) constructed using a method that captures μ 90% of the time. The reliability of the confidence interval procedure is relatively high (90%)

A formula for a $100(1 - \alpha)$ confidence interval for μ

- We need to find L and U such that $100(1 - \alpha)\%$ of the sampling distribution of \bar{Y} lies between L and U
- Combining steps 1) and 2) gives us

$$L = \bar{Y} - z_{\alpha/2}\sigma_{\bar{y}}$$

where $z_{\alpha/2}$ has an area of $\alpha/2$ to its *right*, and

$$U = \bar{Y} + z_{\alpha/2}\sigma_{\bar{y}}$$

- Then, a compact formula for the interval is

$$\bar{Y} \pm z_{\alpha/2}\sigma_{\bar{y}}$$

- Ott and Longnecker (p. 201) have a convenient table showing popular values of α and the corresponding values of $z_{\alpha/2}$
- Note that we have assumed knowledge of σ . Later, a slightly modified version will be given that allows an estimate of σ to be used in place of the true value
- For the time being, we will substitute the sample standard deviation s for σ (the population standard deviation) if σ is unknown and s is available.

Factors Affecting the Width of a Confidence Interval

- The width of the interval is the distance from U to L . Specifically, the width is

$$\begin{aligned} w &= U - L \\ &= \bar{Y} + z_{\alpha/2}\sigma_{\bar{y}} - (\bar{Y} - z_{\alpha/2}\sigma_{\bar{y}}) \\ &= 2z_{\alpha/2}\sigma_{\bar{y}} \end{aligned}$$

- There is a tradeoff between the width of the CI and the confidence interval. A greater level of confidence requires a wider interval. How wide must a 100% CI be?
- Wide intervals are undesirable because they are less specific about what values of μ are consistent with the data

- When capturing the true mean is of critical importance, then a high level of confidence are necessary. For instance, suppose that μ is the expected number of O-rings that fail on a space shuttle flight. If the true value of μ is larger than all values in the CI, then the consequences may be disastrous
- There is one way to decrease confidence interval width without decreasing the confidence level: take a larger sample. This is true because the interval width is

$$w = 2z_{\alpha/2}\sigma_{\bar{y}} = \frac{2z_{\alpha/2}\sigma}{\sqrt{n}}.$$

- By making n large, the width will shrink

5.3 Choosing n for estimating μ

- How large a sample is large enough? That depends on the desired level of precision.
- In practice, we choose an confidence level, say 90%, and a maximum tolerable width w
- The width of the interval is

$$w = 2z_{\alpha/2}\sigma_{\bar{y}} = \frac{2z_{\alpha/2}\sigma}{\sqrt{n}}.$$

- Given knowledge of σ , or at least an estimate of σ , we can solve for an approximate minimum sample size as follows:

$$\begin{aligned} w &= \frac{2z_{\alpha/2}\sigma}{\sqrt{n}} \\ \Rightarrow \sqrt{n} &= \frac{2z_{\alpha/2}\sigma}{w} \\ \Rightarrow n &= \left(\frac{2z_{\alpha/2}\sigma}{w} \right)^2 \end{aligned}$$

- Therefore, the minimum necessary sample size is

$$n = \frac{4z_{\alpha/2}^2\sigma^2}{w^2}.$$

- Ott and Longnecker have an alternate formula where they replace w by the half-width of the interval

$$E = \frac{w}{2}$$

- If we substitute $2E$ for w in the formula, we get Ott and Longnecker's formula on page 205.

$$n = \frac{4z_{\alpha/2}^2\sigma^2}{(2E)^2} = \frac{z_{\alpha/2}^2\sigma^2}{E^2}$$

Example Suppose that the objective is to estimate mean (human) lifetime μ in the Middle Ages.

- A 90% CI is desired, and a tolerable width is 3 years.
- An examination of a modern lifetimes indicates that a reasonable estimate of σ is 8 years
- Then, $z_{\alpha/2} = 1.645$, and

$$\begin{aligned} n &= \frac{4z_{\alpha/2}^2\sigma^2}{w^2} \\ &\approx \frac{4 \times 1.645^2 \times 64}{9} = 76.9. \end{aligned}$$

Hence, we need at least 77 observations to obtain a 90% C.I. that is approximately of width 3 years

5.4 Hypothesis Testing

Example From Roberts, H.V. 1979. "Harris Trust and Savings Bank: An Analysis of Employee Compensation," Report 7946, Center for Mathematical Studies in Business and Economics, University of Chicago Graduate School of Business.

- A lawsuit was filed against the Harris Trust and Savings Bank claiming that as a group, new male employees received larger starting salaries than new women employees. Data are beginning salaries for 32 male and 61 female skilled, entry-level clerical workers hired by the bank from 1969 to 1977.

Statistical summaries and Approximate 95% Confidence Intervals

	n	\bar{y} (dollars)	s (dollars)	95% Confidence Interval (dollars)
Females	61	5139	691	5000.1 to 5277.1
Males	32	5956	540	5707.8 to 6205.9

- Do these data support the contention that female salaries are consistently and systematically lower than males? That is, is the mean starting salary μ of skilled, entry-level female clerical workers less than the mean starting salary of the males?
- A quick test is to determine if \$5956 is one of those values that is consistent with data for the females. That is, is 5956 between 5000.1 and 5277.1?
- It's not. What can we say? We are 95% confident that the mean level of females is not \$5956
- This test has several limitations (to be explained later). We will conduct a better test at that time

Example Mendelian (or traditional) genetic theory states that half the seed pods for a particular variety of peas will be wrinkled, and half will be smooth.

- New theory says the proportion of wrinkled seeds in the population of all seeds is 0.35.
- We believe the new theory is plausible and want to test it
- A random sample of 20 pods has $X = 6$ wrinkled seeds.
- Let $\pi =$ probability that a randomly sampled pod is wrinkled
- There are two competing hypotheses:

$$H_0 : \pi = 0.5 \text{ and } H_a : \pi = 0.35.$$

- If H_0 were true, then $X \sim B(20, 0.5)$ and the probability of getting so few wrinkled seeds as 6 is

$$P(X \leq 6 \mid \pi = 0.5) = \sum_{k=0}^6 \binom{20}{k} 0.5^k 0.5^{20-k} = 0.056.$$

- If H_a were true, then $X \sim B(20, 0.35)$ and the probability of getting so few wrinkled seeds as 6 is

$$P(X \leq 6 \mid \pi = 0.35) = \sum_{k=0}^6 \binom{n}{k} 0.35^k 0.35^{20-k} = 0.418.$$

- The probability of getting a count of 6 or fewer is quite improbable if H_0 is true. Either we were unlucky, or π is not 0.5. Since luck has no place in science, we conclude that $\pi = 0.35$.

Hypothesis Testing for μ

- Hypothesis testing is a methodology for measuring the strength of evidence (contained in a sample) contradicting a hypothesis (called the null hypothesis), and favoring an alternative, or research, hypothesis
- H_0 (null) and H_a (research) are set up so that they are completely contradictory
- H_a is the hypothesis which we favor (think to be true), and H_0 is its opposite (in some sense) which we hope to be able to reject. Rejecting H_0 implies that H_a is correct

Components of a Hypothesis Test for μ

1) The hypotheses. The test only provides statistical evidence *against* H_0 , so H_0 and H_a should be chosen accordingly.

- If you believe that the true mean μ of lifetimes in the middle ages is less than 45, and the conventional belief is $\mu \geq 45$, then set

$$H_a : \mu < 45 \text{ and } H_0 : \mu \geq 45.$$

- Generically, the value of μ specified by H_0 is denoted by μ_0 . The generic form of the hypotheses is

$$H_a : \mu < \mu_0 \text{ and } H_0 : \mu \geq \mu_0.$$

2) The test statistic. The test statistic measures the strength of evidence against H_0 . The test statistic is

$$Z = \frac{\bar{Y} - \mu_0}{\sigma_{\bar{y}}}.$$

- We use Z because we know that is $N(0,1)$ in distribution, provided that H_0 is true

3) The rejection region R . If the value of the test statistic falls R (i.e., $Z \in R$), then we reject H_0 in favor of H_a

- To determine R , consider these facts:

1) $\bar{Y} - \mu_0$ will probably be negative if the true value of μ is less than μ_0 because the expected value of \bar{Y} will probably be less than μ_0

2) Z will be negative when $\bar{Y} - \mu_0$ is negative and vice versa

- Thus, we chose R so that if H_0 is true, then $P(Z \in R)$ is small. In addition, the probability of $P(Z \in R)$ should be larger if H_a is true.

- In other words,

$$P(Z \in R \mid H_0 \text{ is true}) < P(Z \in R \mid H_a \text{ is true})$$

- For example, if we decide that $P(Z \in R \mid H_0 \text{ is true})$ should be 0.05, then

$$R = \{z \mid z < -1.645\}$$

because $0.05 = P(Z < -1.645) = P(Z \in R \mid H_0 \text{ is true})$.

4) The conclusion. We either

a) reject H_0 in favor of H_a , or

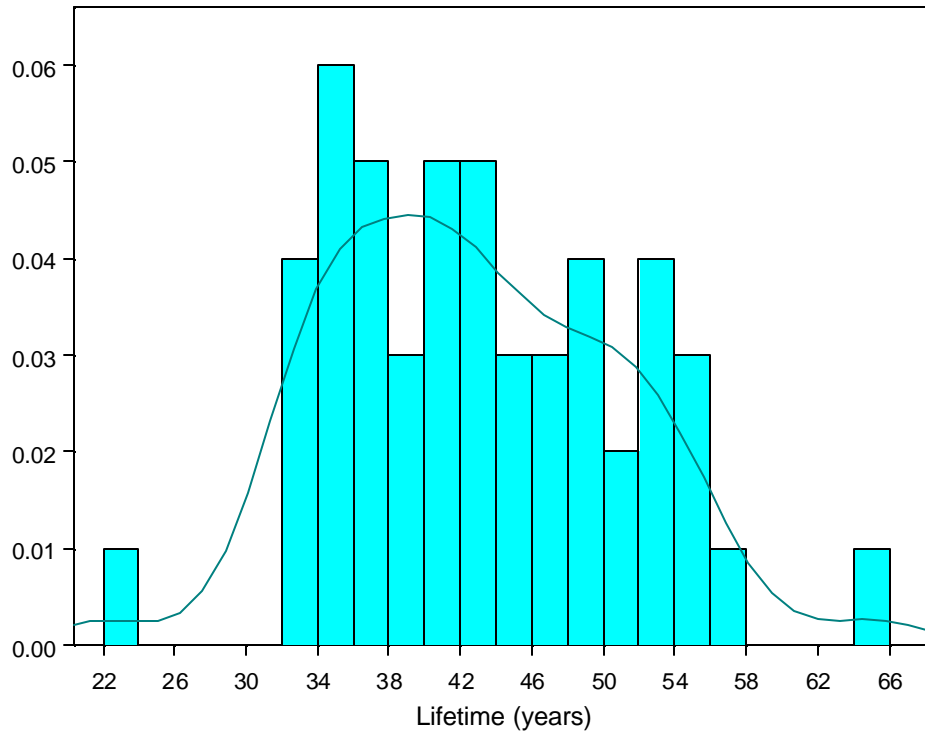
b) conclude that the data provide insufficient evidence to reject H_0 .

- If $Z \notin R$, then we do not accept H_a because our only objective is to assess the strength of evidence against H_0 and in favor of H_a . *We are not* assessing the strength of evidence against H_a and in favor of H_0

Example

- A graduate student, M., was interested in longevity during the Medieval age, and in particular, felt that a published value for mean lifetime of $\mu = 45$ for humans living in Europe between 400 and 1100 A.D. was too large. He went to Europe, and visited a sample of graveyards associated with churches that were in existence between 400 A.D. and today, and recorded the ages at death of individuals that died between 400 and 1100 A.D.

- The sample distribution of $n = 50$ values is shown below



- The sample statistics were $\bar{Y} = 42.64$ and $s = 8.17$. Hence, an estimate of $\sigma_{\bar{y}}$ is

$$\frac{s}{\sqrt{n}} = \frac{8.17}{\sqrt{50}} = 1.155.$$

The Hypothesis Test

- 1) The hypotheses: $H_0 : \mu = 45$ and $H_a : \mu < 45$.
- 2) The test statistic is

$$Z = \frac{\bar{Y} - \mu_0}{\sigma_{\bar{y}}},$$

where μ_0 is the value of μ specified by H_0 . In this example, $\mu_0 = 45$.

- The value of Z is

$$Z = \frac{42.64 - 45}{1.155} = -2.033.$$

- 3) The rejection region is $R = \{z \mid z < -1.645\}$. The probability of rejecting H_0 , given that H_0 is true, is $P(Z \in R \mid H_0 \text{ is true}) = 0.05$.
 - 4) Conclusion: reject H_0 in favor of H_a because $Z = -2.033 < -1.645$
- Note that the I cheated on the test statistic because I used the sample standard deviation $s = 8.17$ where I should have used the true value σ . I'm not worried because the test statistic -2.033 is not near the critical threshold of -1.645

A Quick Review of Hypothesis Testing

- The aim is limited: to determine the strength of evidence against one hypothesis (H_0) and favoring the other (H_a). For now, these hypotheses say something about the population mean μ
- We approach the problem of determining the strength of evidence conservatively. Specifically, we assume that H_0 is true, and compute the value of the test statistic $Z = (\bar{Y} - \mu_0) / \sigma_{\bar{y}}$
- We identify a rejection region consisting of values of Z that are improbable if H_0 is true, and relatively more likely if H_a is true
- If the value of the test statistic is in the rejection region, we reject H_0 and conclude that H_a is true. If the value of the test statistic is *not* in the rejection region, then we conclude that there is insufficient evidence to reject H_0

Choosing the Rejection Region (R)

- R is chosen so that it contains *all* of values of Z that are unlikely if H_0 is true
- The probability of rejecting H_0 is the probability of getting a value of Z which is in R . This probability, which is assumed H_0 to be true, is denoted by α , and it should be small
- α is central to the test, as it determines the rejection region. We choose α
- The event that H_0 is rejected, given that H_0 is true, is called a *Type I error*. Its probability is

$$\alpha = P(\text{Type I error}) = P(\text{reject } H_0 \mid H_0 \text{ is true})$$

- In the previous example, the hypotheses were $H_0 : \mu = 45$ and $H_a : \mu < 45$. If H_0 is false and H_a is true, then we expect that $\bar{Y} < 45$. Hence, small values of

$$Z = \frac{\bar{Y} - \mu_0}{\sigma_{\bar{y}}} = \frac{\bar{Y} - 45}{1.15}$$

are evidence against H_0 and in favor of H_a

- I choose $\alpha = 0.05$. That means I must determine the value of z_α satisfying

$$0.05 = P(Z < z_\alpha \mid H_0 \text{ is true})$$

- From Table A, the value of z_α is -1.645 . Therefore, R is the set of all values of Z that are less than -1.645

In-Class Exercise

- An answer to the question "how much sleep is necessary?" is posted on *SleepNet.Com*, a web-site dedicated to sleep and sleep-disorders. The answer is:

"...The largest group of people seem to report a need for approximately 7-8 hours of sleep. The best way to determine your sleep need is to evaluate alertness in the waking period. If fatigue, sleepiness, or drowsiness creeps into the waking hours this may be an indication that more sleep is needed. Another way to tell if your getting enough sleep is to examine the role of an alarm clock. People who are getting sufficient, good quality sleep at the correct time for their own body will wake up spontaneously - without the need for an alarm clock."¹

- It seems unlikely that the average college student (or instructor) gets that much sleep. You will test whether it is true that the mean amount of nightly sleep of college students is 7 hours versus a lesser amount.
- Everyone that is willing should write down the time (in hours) of their sleep last night on the board.
- Form 6 groups, and using the data on the board, carry out a hypothesis test that assesses the strength of evidence against the hypothesis that the mean amount of sleep is 7 hours or more.
- Put your results on the board. Give the 4 components of the test (hypotheses, test statistic, rejection region, and conclusion).

Hypothesis Testing - Correct and Incorrect Conclusions

- To understand the potential errors involved in hypothesis testing, we define a Type II error as the event that we fail to reject H_0 when H_0 is false
- The probability of a Type II error is denoted by β . Hence,
$$\beta = P(\text{fail to reject } H_0 \mid H_0 \text{ is false}).$$
- The decision-making process by tabulating the possible outcomes, and the probabilities of the outcomes, given the true state of nature are shown.

¹<http://www.sleepnet.com/wwwboard/messages/37.html>

The True State of Nature

Decision	H_0 is true	H_0 is False
<i>Reject H_0</i>	Type I error (prob. = α)	Correct (prob. = $1 - \beta$)
<i>Fail to reject H_0</i>	Correct (prob. = $1 - \alpha$)	Type II error (prob. = β)

- α and β are controlled by the researcher through the choice of R .
- β will increase if α is decreased because small α requires a small R .
Therefore, it is more difficult to reject H_0 when H_0 is false, and you are more likely to make a Type II error
- α is chosen by the experimenter (common values are 0.1, 0.05 and 0.01)
- It is possible to control β by manipulating the sample size n

One-sided Alternative Hypotheses

- Almost always, we formulate the alternative (or research) hypothesis according to our research questions. For example, we may state that μ is greater than (or less than) some value μ_0 . Generically, H_a is either
 - 1) $H_a : \mu > \mu_0$, or
 - 2) $H_a : \mu < \mu_0$
- Note that in both cases, H_a states that μ is on one side of μ_0
- The null hypothesis is counter to the research hypothesis, hence, it should be $H_0 : \mu \leq \mu_0$ in the first case, and $H_0 : \mu \geq \mu_0$ in the second case
- Occasionally, there is some pre-established value of μ (not a range), and H_0 is $H_0 : \mu = \mu_0$
- To compute the test statistic, we must state *some* value for μ_0 because

$$Z = \frac{\bar{Y} - \mu_0}{\sigma_{\bar{y}}}$$

Two-sided Alternative Hypotheses

- Sometimes, the alternative hypothesis states that μ is not equal to some null hypothesis value μ_0 (rather than, say, less than μ_0)
- Then, we *must* give a specific value for μ_0
- For example, it is thought by some that the global climate is changing. At a specific point on the earth, mean temperature during the last decade may be more or less than the historical average, though conventional wisdom

says that temperatures are increasing at most locations. To establish evidence for climate *change* in Missoula, the following test may be used

1) Test $H_0 : \mu = 43.2$ versus $H_a : \mu \neq 43.2$,

where μ is the mean annual temperature, and 43.2 F was the mean temperature in Missoula between 1900 and 1990

- $H_a : \mu \neq 43.2$ is called a *two-tailed* alternative because the rejection region is comprised of the two tails of the $N(0, 1)$ distribution

2) The test statistic is

$$Z = \frac{\bar{Y} - \mu_0}{\sigma_{\bar{y}}},$$

where $\mu_0 = 43.2$. Suppose that $\bar{Y} = 44.8$ ($n = 10$) and $\sigma = 2.1$. Then

$$\sigma_{\bar{y}} = \frac{2.1}{\sqrt{10}} = 0.664,$$

and

$$Z = \frac{\bar{Y} - \mu_0}{\sigma_{\bar{y}}} = \frac{44.8 - 43.2}{0.664} = 2.41$$

- If H_0 is true, then $Z \sim N(0, 1)$.
- 3) I choose R so that the probability of rejecting H_0 , given that H_0 is true, is α . Large values *and* small values of Z contradict H_0
- The *acceptance* region, \bar{R} , is the complement of R . Almost always, the acceptance region is symmetric.
 - For this two-tailed hypothesis test, the *acceptance* region is the set of all values of Z between $-z_{\alpha/2}$ and $z_{\alpha/2}$, i.e., $\bar{R} = \{z \mid -z_{\alpha/2} < z < z_{\alpha/2}\}$
 - R is the set of values of Z that are greater than $z_{\alpha/2}$ or less than $-z_{\alpha/2}$, i.e.,

$$\begin{aligned} R &= \{z \mid z \geq z_{\alpha/2}\} \cup \{z \mid z \leq -z_{\alpha/2}\} \\ &= \{z \mid |z| \geq z_{\alpha/2}\} \end{aligned}$$

- For example, if $\alpha = 0.05$, then $\alpha/2 = 0.025$ and $z_{\alpha/2} = 1.96$ (from p. 202 in Ott and Longnecker), and $R = \{z \mid |z| \geq 1.96\}$. The decision rule is: reject H_0 if $Z \geq 1.96$ or $Z \leq -1.96$
- 4) Conclusion: reject H_0 in favor of H_a because $z = 2.41$ is in R . Specifically, $Z = 2.41 > 1.96$.
- We will not make much use of two-sided alternative hypotheses until later

Power

- The concept of power is important. Power is the probability of rejecting H_0 given that H_0 is false.
- Power is directly related to the probability of a Type II error because "fail to reject H_0 " is the complement of "reject H_0 ". Therefore,

$$\begin{aligned} \text{Power} &= P(\text{reject } H_0 \mid H_0 \text{ is false}) \\ &= 1 - P(\text{fail to reject } H_0 \mid H_0 \text{ is false}) \\ &= 1 - P(\text{Type II error}) \end{aligned}$$

- Thus, $\text{Power} = 1 - \beta$ because $\beta = P(\text{Type II error})$
- Computing power is not done routinely, largely because it is difficult

Example A machine produces computer hard disks. If the average asymmetry of hard disks produced by the machine is greater than 0.29, then the machine is considered to be malfunctioning

- Suppose that the population of hard disk asymmetries is normal in distribution with mean $\mu = 0.28$ and standard deviation $\sigma = 0.005$ (this implies that the machine is functioning properly)
- A process control engineer is monitoring the process for errors. She wants to find them as soon as possible by drawing a random sample and determining the strength of evidence indicating a process error. She will test these hypotheses

$$H_0 : \mu = \mu_0 \text{ versus } H_a : \mu > \mu_0 \text{ where } \mu_0 = 0.28$$

- To compute power, a specific alternative stating that $\mu = \mu_a$, where $\mu_a > \mu_0$ is needed. My test is set up using a conservative choice, $\mu_a = 0.29$:

1) $H_0 : \mu = \mu_0$ and $H_a : \mu = \mu_a$

where $\mu_0 = 0.28$ and $\mu_a = 0.29$

2) The test statistic is

$$Z = \frac{\bar{Y} - \mu_0}{\sigma_{\bar{y}}}$$

3) Set $\alpha = 0.05$. We reject for large values of Z , specifically, $R = \{z \mid z > z_\alpha = 1.645\}$.

• We are computing power. Therefore, we assume $H_a : \mu = \mu_a$ is true. There are several consequences:

a)

$$\frac{\bar{Y} - \mu_a}{\sigma_{\bar{y}}} \sim N(0,1),$$

b) The test statistic Z no longer has a $N(0,1)$ distribution. Lets call it Z^* to emphasis that its distribution is *not* $N(0,1)$ if H_a is true. It's not $N(0,1)$ because the mean is

$$\begin{aligned} E(Z^*) &= E\left(\frac{\bar{Y} - \mu_0}{\sigma_{\bar{y}}} \mid \mu = \mu_a\right) . \\ &= \frac{E(\bar{Y} \mid \mu = \mu_a) - \mu_0}{\sigma_{\bar{y}}} \\ &= \frac{\mu_a - \mu_0}{\sigma_{\bar{y}}} \end{aligned}$$

• The standard deviation of Z^* is 1

- To compute power, I need to standardize Z^* . Hence, we apply the standard normal transformation to Z^* so that the transformed variable is $N(0,1)$. The transformation is

$$\frac{Z^* - \frac{\mu_a - \mu_0}{\sigma_{\bar{y}}}}{1} = Z^* - \frac{\mu_a - \mu_0}{\sigma_{\bar{y}}} \sim N(0,1)$$

- The power using a sample of $n = 25$ disks is the probability that Z^* will be in R , that is, larger than 1.645 **given** that $H_a : \mu = \mu_a = 0.29$ is true. Specifically,

$$\begin{aligned} \text{Power} &= P(Z^* > z_\alpha) = P\left(Z^* - \frac{\mu_a - \mu_0}{\sigma_{\bar{y}}} > z_\alpha - \frac{\mu_a - \mu_0}{\sigma_{\bar{y}}}\right) \\ &= P\left(Z > z_\alpha - \frac{\mu_a - \mu_0}{\sigma_{\bar{y}}}\right), \end{aligned}$$

where this Z really is $N(0,1)$ in distribution. The standard deviation of \bar{Y} is

$$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}} = \frac{0.005}{\sqrt{25}} = 0.001$$

- Then,

$$\begin{aligned} P(Z^* > z_\alpha) &= P\left(Z > 1.645 - \frac{0.29 - 0.28}{0.001}\right) \\ &= P(Z > -8.35) > 0.9999 \end{aligned}$$

Another Look at Power

- Ott and Longnecker (p. 214) have a different approach. In their example, there is a question of whether the annual mean number of improperly issued parking tickets has increased relative to the historical average of 380 per year. The distribution of improperly issued parking tickets has mean and standard deviation

$$\mu = 380 \text{ and } \sigma = 35.2$$

- To test $H_0 : \mu \leq 380$ versus $H_a : \mu > 380$, they use the value $\mu_0 = 380$ as this value is closest to what is specified under H_a . This is the usual conservative approach.
- The test statistic is

$$Z = \frac{\bar{Y} - \mu_0}{\sigma_{\bar{y}}}$$

- They set $\alpha = 0.01$, and reject for large values of Z . Specifically, $R = \{z \mid z > z_\alpha = 2.33\}$
- The decision rule is

$$\text{Reject } H_0 \text{ if } Z = \frac{\bar{Y} - 380}{\sigma_{\bar{y}}} > z_\alpha$$

- The rule also can be stated as

$$\text{Reject } H_0 \text{ if } \bar{Y} > 380 + z_\alpha \sigma_{\bar{y}}$$

- The power of the test is $P(\text{reject } H_0 \mid H_0 \text{ is false})$, which in this case is also

$$\text{Power} = P(\bar{Y} > 380 + z_\alpha \sigma_{\bar{y}} \mid \mu = \mu_a)$$

where μ_a is a specific value (not yet identified)

- They choose 3 different values for μ_a because there is no best single alternative value
- The power calculation assumes $H_a : \mu = \mu_a$ is true. In this case, the mean of the distribution of \bar{Y} (also called the expected value of \bar{Y}) is μ_a
- Assume $n = 50$, so that $\sigma_{\bar{y}} = 35.2 / \sqrt{50} = 4.98$, and let's use $\mu_a = 395$. Then, the power is

$$\begin{aligned} \text{Power} &= P(\bar{Y} > 380 + z_\alpha \sigma_{\bar{y}} \mid \mu = 395) \\ &= P(\bar{Y} > 380 + 2.33 \times 4.98 \mid \mu = 395) \\ &= P(\bar{Y} > 391.6 \mid \mu = 395) \end{aligned}$$

- Standardize both sides and look up the value in Table A:

$$\begin{aligned} \text{Power} &= P(\bar{Y} > 391.6 \mid \mu = 395) \\ &= P\left(\frac{\bar{Y} - \mu}{\sigma_{\bar{y}}} > \frac{391.6 - 395}{4.98} \mid \mu = 395\right) \\ &= P(Z > -0.682) \\ &= 0.752 \end{aligned}$$

- Ott and Longnecker's figure on p. 215 summarizes the situation *in terms of* β , rather than power, for the three values of μ_a (395, 387 and 400)
- Recall that the acceptance region is the complement of rejection region
- Figure 5.12 from Ott and Longenecker (p. 215) is another way of looking at power
- It is useful to have formulas for β . Ott and Longnecker (p. 216) provide convenient formulas for the one- and two-tails cases of alternative hypotheses

1) One tail:
$$\beta = P\left(Z \leq z_\alpha - \frac{|\mu_a - \mu_0|}{\sigma_{\bar{y}}}\right)$$

- This formula works for both types of alternatives, $\mu_0 < \mu_a$ and $\mu_0 > \mu_a$

2) Two tails:
$$\beta = P\left(Z \leq z_{\alpha/2} - \frac{|\mu_a - \mu_0|}{\sigma_{\bar{y}}}\right)$$

- In either case, Power = $1 - \beta$

Example Let's try the last example using formula 1). Recall the last power (and β) computation which used

$$z_\alpha = 2.33, \mu_a = 395, \mu_0 = 380, \sigma_{\bar{y}} = 4.98$$

- Then,

$$\begin{aligned} \beta &= P\left(Z \leq z_\alpha - \frac{|\mu_a - \mu_0|}{\sigma_{\bar{y}}}\right) \\ &= P\left(Z \leq 2.33 - \frac{|395 - 380|}{4.98}\right) \\ &= P(Z \leq 2.33 - 3.01) \\ &= P(Z < -0.68) = 0.248 \end{aligned}$$

- Finally, Power = $1 - \beta = 1 - 0.248 = 0.752$

Choosing the Sample Size

- The formulas for β can be used to estimate the minimum sample size necessary to achieve a Type II error rate = β *provided* that α , μ_a , μ_0 and σ are specified

- If H_a is one-tailed, then

$$\beta = P\left(Z \leq z_\alpha - \frac{|\mu_a - \mu_0|}{\sigma_{\bar{y}}}\right)$$

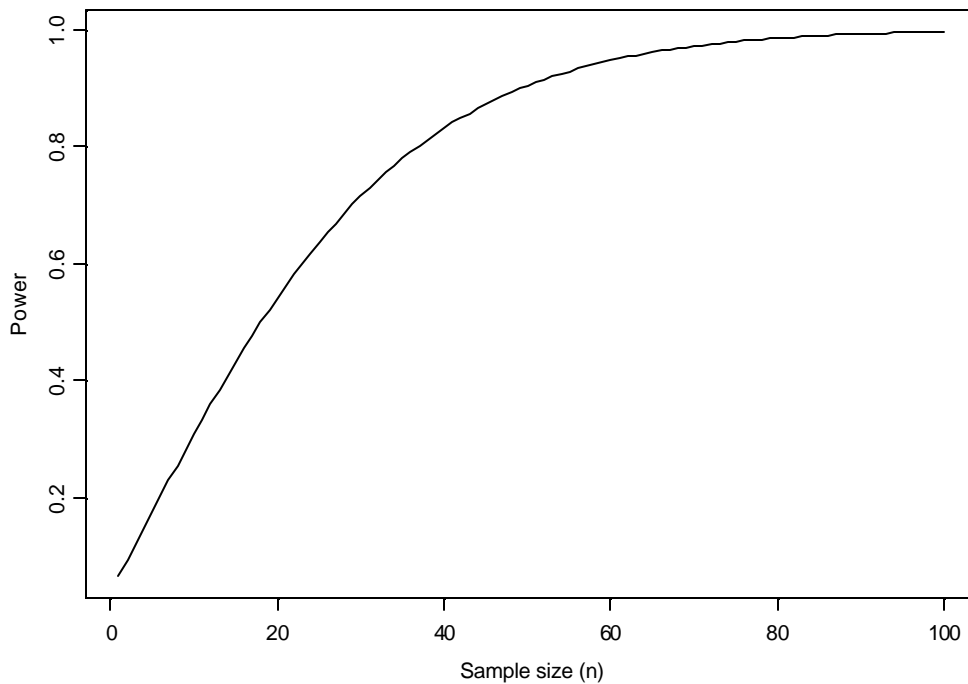
and power is

$$\text{Power} = P\left(Z > z_\alpha - \frac{|\mu_a - \mu_0|}{\sigma_{\bar{y}}}\right)$$

- First, let's expand this formula to see the effect of sample size (n) on power by replacing $\sigma_{\bar{y}}$ by σ/\sqrt{n} :

$$\text{Power} = P\left(Z > z_\alpha - \sqrt{n} \frac{|\mu_a - \mu_0|}{\sigma}\right)$$

- Now, plug in $n = 1, 2, \dots, 100$, compute, and plot power against n :



- This figure indicates that n should be at least 25 to insure acceptable power (i.e. Power > 0.5), and that n should be at least 40 to insure good power (Power > 0.8)
- Rather than compute all these Power values, it is usually easier to compute the minimum n necessary to insure some level of Power.

- For example, suppose that I choose $\beta = 0.1$, $\alpha = 0.05$, $\mu_a = 395$, $\mu_0 = 380$ and $\sigma = 35.2$.
- To achieve Power = 0.9, it must be true that $\beta = 0.1$ and

$$0.1 = P\left(Z \leq z_\alpha - \sqrt{n} \frac{|\mu_a - \mu_0|}{\sigma}\right),$$

and this will occur only if $z_\alpha - \sqrt{n} \frac{|\mu_a - \mu_0|}{\sigma}$ is the 10th percentile of the $N(0,1)$ distribution. Because the 10th percentile is -1.28 , β will be 0.1 if the equation

$$-1.28 = z_\alpha - \sqrt{n} \frac{|\mu_a - \mu_0|}{\sigma}$$

is correct (or true). Specifically, we must have

$$\begin{aligned} -1.28 &= 1.645 - \sqrt{n} \frac{15}{35.2} \\ \Rightarrow n &= \left[\frac{35.2(-1.28 - 1.645)}{15} \right]^2 = 47.1 \end{aligned}$$

- In general, a Type II error rate of β will be achieved if

$$n = \sigma^2 \frac{(z_\alpha + z_\beta)^2}{(\mu_a - \mu_0)^2}$$

where z_β is the value that has an area of β to its *right*

- Ott and Longnecker have a similar formula on p. 219
- A similar formula is used when the alternative is two-tailed:

$$n = \sigma^2 \frac{(z_{\alpha/2} + z_\beta)^2}{(\mu_a - \mu_0)^2}$$

Example Let's determine n necessary to achieve $\beta = 0.05$ when $\alpha = 0.01$.
Recall

$$\mu_a = 395, \mu_0 = 380, \sigma = 35.2$$

- In addition, $\alpha = 0.01 \Rightarrow z_\alpha = 2.33$ and $\beta = 0.05 \Rightarrow z_\beta = 1.645$ so

$$\begin{aligned}
 n &= \sigma^2 \frac{(z_\alpha + z_\beta)^2}{(\mu_a - \mu_0)^2} \\
 &= 35.2^2 \frac{(2.33 + 1.645)^2}{(395 - 380)^2} \\
 &= 1239.0 \times \frac{15.80}{225} \\
 &= 87.0
 \end{aligned}$$

- At least 87 are necessary to achieve $\beta = 0.05$ when $\alpha = 0.01$

5.6 The Level of Significance of a Statistical Test

- Motivation: Type I error rate (α) of a test is chosen by the experimenter, and there are different opinions regarding appropriate values
- To remove the arbitrariness in choosing α , it is helpful to report the *observed significance level or p-value* of the test
- The p-value of the test can be described in any of these three terms:
 1. It is the smallest possible α that would allow the null hypothesis to be rejected (technical definition)
 2. It is the probability of obtaining as much or more evidence (than observed) against H_0 and favoring H_a , given that H_0 is true (interpretative definition)
 3. It is the weight of evidence against H_0 and for H_a (informal definition)

Example of a Hypothesis Test that uses the P-value

1. Hypotheses: $H_0 : \mu = 45$ versus $H_a : \mu < 45$.

2. The test statistic is

$$Z = \frac{\bar{Y} - \mu_0}{\sigma_{\bar{y}}},$$

where $\mu_0 = 45$, $\bar{Y} = 42.64$, and $\sigma_{\bar{y}} = 1.155$

• The value of Z is

$$z = \frac{42.64 - 45}{1.155} = -2.033.$$

• To determine the p-value, note that if α were **smaller** than -2.033 , then we could not reject H_0 . If α were larger than -2.033 , then we could reject H_0 . Therefore, the observed significance level is

$$\text{p-value} = P(Z \leq -2.033) = 0.021.$$

4. Interpretation: because p-value = 0.021 is small, we have found strong evidence against H_0 and in favor of H_a .

Example of a P-value Calculation for a Two-Tailed Alternative

1. Test $H_0 : \mu = 45$ versus $H_a : \mu \neq 45$.

2. The test statistic is

$$Z = \frac{\bar{Y} - \mu_0}{\sigma_{\bar{y}}},$$

where $\mu_0 = 45$.

• The value of Z is

$$z = \frac{42.64 - 45}{1.155} = -2.033.$$

3. To determine the p-value, we need to find the largest α that allows us to reject H_0 . R is two-tailed because $H_a : \mu \neq 45$.

• The critical value of Z for this data-determined test is $z_{\alpha/2} = 2.033$, and the area to the right of $z_{\alpha/2} = 2.033$ is 0.021. The α -level is $2 \times 0.021 = 0.042$, because there are two tails, both with areas of 0.021.

• In summary,

$$\text{p-value} = P(Z \leq -2.033) + P(Z \geq 2.033) = 0.021 + 0.021 = 0.042.$$

• Generally, for two-tailed tests, if z is the value of the test statistic, then

$$\text{p-value} = 2P(Z \geq |z|),$$

where $Z \sim N(0,1)$

• Example: given $z = -2.033$,

$$\text{p-value} = 2P(Z \geq |z|) = 2P(Z \geq 2.033) = 2 \times 0.021.$$

Warnings About CI's and Tests Based on the Normal Assumption

• In this setting, the normal assumption specifies that \bar{Y} is normal in distribution.

• \bar{Y} is normal if the sampled population is normal; \bar{Y} is approximately normal if the sampled population is not normal

- Differences between the actual distribution of \bar{Y} and the normal distribution are a source of error
- These differences originate from the sampled population
- It is important to assess the validity of the normal distribution as a model for the population. Not knowing the population, we must use the sample
- Things to check (roughly in order of importance):
 - 1) Symmetry - examine a histogram of the sample
 - 2) Outliers - examine a boxplot of the sample
 - 3) Heavy tails - examine a normal P-P, normal probability, or Q-Q plot

Methods of Assessing the Normal Distribution Assumption

- A normal P-P plot shows the proportion of the sample, y_1, \dots, y_n that is less than or equal to each sample value plotted against the *expected* proportion *if* the data were normal.
- Let $y_{(i)}$ denote the i th largest observation among y_1, \dots, y_n
- The (estimated) expected proportion less than or equal to $y_{(i)}$ is

$$p_i = P\left(Z \leq \frac{y_{(i)} - \bar{Y}}{s}\right)$$

- For example, for the SAT data, the smallest of 50 observations on state mean SAT is

$$y_{(1)} = 790,$$

and the proportion of the sample that is less than or equal to $y_{(1)}$ is $1/50 = 0.02$.

- To center things, we need to subtract $1/(2n) = 0.01$ from the observed proportion, so 0.02 becomes 0.01
- Also, $\bar{Y} = 947.94$ and $s = 70.8$. Then,

$$p_1 = P\left(Z \leq \frac{790 - 947.94}{70.8}\right) = P(Z \leq -2.23) = 0.013$$

- We plot the pair (0.01, 0.013)

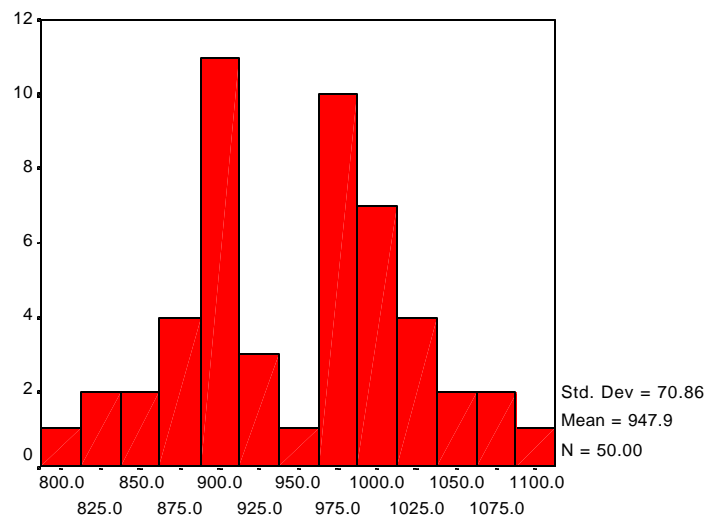
- The 25th largest value is 958, so

$$p_{25} = P\left(Z \leq \frac{958 - 947.94}{70.8}\right) = P(Z \leq 0.142) = 0.556,$$

- The pair to be plotted is (0.49, 0.556)
- If the data are a sample from a normal distribution, then the points should fall on a straight line (aside from some random variation)

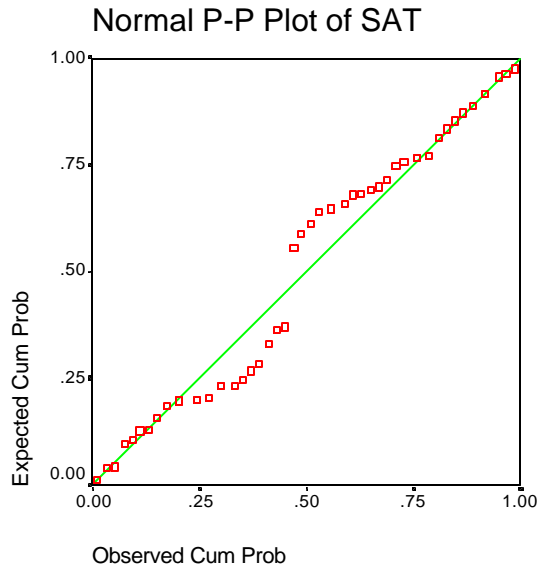
Example

- The SAT data. State mean SAT scores



SAT

- The normal probability plot reflects the bimodal aspect of the histogram. There is no evidence of skewness or heavy tails



5.7 Inferences about μ for a Normal Population When σ is Unknown

- Almost always, the population standard deviation σ is not known, and estimated by the sample standard deviation s
- The aim is to adjust hypothesis tests and confidence intervals to accurately reflect the uncertainty in the estimate s
- Previously, both confidence intervals and hypothesis tests were based on the Z statistic

$$Z = \frac{\bar{Y} - \mu}{\sigma_{\bar{y}}} = \sqrt{n} \frac{\bar{Y} - \mu}{\sigma} \sim N(0, 1).$$

- When σ is replaced by s , we get the T statistic

$$T = \sqrt{n} \frac{\bar{Y} - \mu_0}{s}.$$

- T has a t -distribution with $n - 1$ degrees of freedom
- I will write $T \sim \mathcal{T}_{n-1}$ to say that T has a t -distribution with $n - 1$ degrees of freedom

The \mathcal{T}_{n-1} Distribution

- This distribution is similar to the $N(0, 1)$ distribution because it is symmetrically distributed about its mean of 0

- The standard deviation of the \mathcal{T}_{n-1} distribution is

$$\sigma_T = \sqrt{\frac{\text{df}}{\text{df} - 2}},$$

where df are the *degrees of freedom*. For now, $\text{df} = n - 1$. For example, if $n = 10$, then $\text{df} = 9$ and

$$\sigma_T = \sqrt{\frac{9}{7}} = 1.13.$$

- For moderately large n , (say $n > 30$), there are no practical differences between the \mathcal{T}_{n-1} and $N(0,1)$ distributions
- An abbreviated table of percentage points of the \mathcal{T}_{n-1} distribution for important values of n is given on page 1093 of Ott and Longnecker. For maximum confusion, Table 2 is reversed compared to Table 1 for the $N(0,1)$.
- The tabled values are percentiles of the t -distributions, but the top row gives the *area to the right* of the percentile. For example, if $n = 10$, $\text{df} = 9$, and $T = 1.833$, Table 2 tells us that

$$P(T > 1.833) = 0.05$$

- To get the area to the left, we must use subtraction. E.g.,

$$P(T \leq 1.833) = 1 - 0.05 = 0.95$$

Degrees of Freedom

- This is a concept from theoretical statistics. The main idea is that we are estimating the average distance from a random observation to μ , and that we are forced to use \bar{y} as an estimate of μ
- s (the estimator of σ) is biased downwards because \bar{y} is computed from the sample and that makes \bar{y} closer to the sample observations than μ
- The bias in s is corrected for by using the appropriate degrees of freedom
- The degrees of freedom ($n - 1$) are a measure of the real amount of information, versus the apparent amount of information (n)
- From now on, we will use the \mathcal{T}_{n-1} for hypothesis testing and confidence intervals if σ is unknown, and n is less than, say 50

- If n is large, say at least 50, we will use the standard normal distribution as before

Hypothesis Testing

- Let's compactly summarize the possible hypotheses:

$$H_0 : \mu = \mu_0$$

- There are three forms of the alternative hypothesis

- 1) $H_a : \mu \neq \mu_0$,

- 2) $H_a : \mu > \mu_0$, and

- 3) $H_a : \mu < \mu_0$.

- If σ is unknown, then the test statistic is

$$T = \sqrt{n} \frac{\bar{Y} - \mu}{s}.$$

Under H_0 , $T \sim \mathcal{T}_{n-1}$.

- For an α -level test, determine the degrees of freedom (df) and look up $t_{\alpha/2}$ or t_α for a two- or one-sided test, respectively,

- For each of the three cases, the rejection regions are

- 1) All values of T that are larger than $t_{\alpha/2}$ or smaller than $-t_{\alpha/2}$

- 2) All values of T that are larger than t_α

- 3) All values of T that are smaller than $-t_\alpha$

- The p-value is obtained by approximation or interpolation

Example

$$H_0 : \mu = \mu_0 \text{ versus } H_a : \mu > \mu_0.$$

Suppose that $n = 10$ and

$$T = \sqrt{n} \frac{\bar{Y} - \mu_0}{s} = 1.86.$$

- What is the approximate p-value (from the \mathcal{T} table)?
- First, find the values that bracket $T = 1.86$ on the line for $df = 9$ in Table 2. These values are 1.833 and 2.262

- The first row tells us that

$$0.025 = P(T > 2.262) \text{ and } P(T > 1.833) = 0.05.$$

- We can say either:

1) p-value ≈ 0.05 , or

2) $0.025 < \text{p-value} < 0.05$

- In any case, we conclude that there is moderate statistical evidence against $H_0 : \mu = \mu_0$ and supporting $H_a : \mu < \mu_0$.

Example 2

- Suppose instead that we want to test $H_0 : \mu = \mu_0$ versus $H_a : \mu \neq \mu_0$.
- Again, suppose that $n = 10$ and $T = 1.86$.
- We have already determined that

$$0.025 < P(T > 1.86) < 0.05$$

- Because this is a two-sided alternative, when we determine the p-value, we need to determine the probability of obtaining as much or more evidence against H_0 and favoring H_a .
- Values of T that are greater than 1.86 are more evidence, **and** values of T that are less than -1.86 constitute more evidence. Because \mathcal{T}_{n-1} is symmetric in distribution,

$$\text{p-value} = P(T < -1.86) + P(T > 1.86) = 2P(T > 1.86)$$

- Because $0.025 < P(T > 1.86) < 0.05$, it is true that

$$2 \times 0.025 < 2 \times P(t > 1.86) < 2 \times 0.05$$

- Therefore, $0.05 < \text{p-value} < 0.10$.

Confidence Intervals

- The old form of the $100(1 - \alpha)\%$ confidence interval for μ was

$$\begin{aligned} \bar{Y} \pm z_{\alpha/2} \sigma_{\bar{y}} \\ \bar{Y} \pm z_{\alpha/2} \sigma / \sqrt{n} \end{aligned}$$

- The $100(1 - \alpha)\%$ confidence interval for μ when s estimates σ is

$$\bar{Y} \pm t_{\alpha/2} s / \sqrt{n},$$

where $t_{\alpha/2}$ is the $100(1 - \alpha/2)$ th percentile of the \mathcal{T}_{n-1} distribution

Warnings About t -procedures

- The validity of t -procedures depend on the distribution of \bar{Y}
- Things to check (roughly in order of importance):
 - 1) Symmetry - use a histogram
 - 2) Outliers - use a boxplot
 - 3) Heavy tails - use a P-P, normal probability, or Q-Q plot

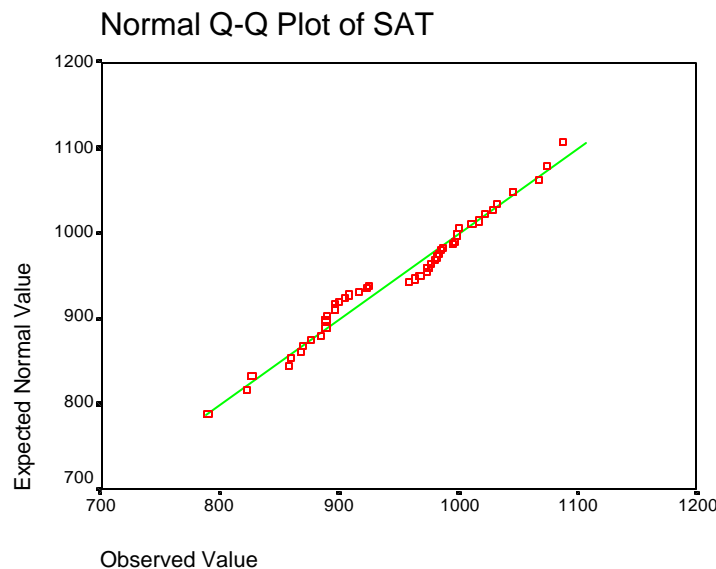
- A normal Q-Q plot shows the sample quantiles plotted against the predicted quantiles assuming the data were normal. I.e., we plot $y_{(i)}$ against

$$\hat{y}_{(i)} = \bar{Y} + z_{(2i-1)/2n} s,$$

where $y_{(i)}$ is the i th largest observation among the sample, and $z_{(2i-1)/2n}$ is the $(2i-1)/2n$ percentile of the Z distribution

- For example, we plot $y_{(1)} = 790$ against

$$\hat{y}_{(1)} = \bar{Y} + z_{0.01} s = 947.94 + (-2.33)70.8 = 783.0$$



- A straight line is super-imposed on the graph to reveal deviations indicative of a non-normal population

Caveats

- It's difficult to quantify the amount of non-normality that can be tolerated
- The effect of using the t -test when the distribution of \bar{Y} is *not normal* is that the adopted Type I error rate is wrong, and the Type II rate is not the computed value (if you compute it)
- For large values of n (> 50), the effects of non-normality are usually minor. t -procedures are robust or resistant against deviations from normality
- When $n < 20$, the effects of non-normality may be very substantial

- Ott and Longnecker have a nice discussion on pages 236 to 238

Presentation of a Data Analysis (Key Components)

- Introduction - clear statement of objectives
- Description of data collection - how it was done, what variables were measured, substantive problems
- Methods - description of statistical methods (usually it is sufficient to give a citation) but an explanation of why they were chosen is important
- Results
 - 1) Describe the distribution of the sample data. Use graphs such as histograms, boxplots, PP plots to support your description
 - 2) Give point and interval estimates of μ , and test statistics, p-values and conclusions if tests are used. Only present results that are relevant to the objectives
- Discussion - interpret statistical results in plain language

5.8 Inferences About the Median

- When the population distribution is not symmetric and the sample size is not large, then t - and Z -procedures are inaccurate
- Sometimes a transformation such as log or square root will yield a symmetric distribution, and we can analyze the transformed data
- Sometimes transformations don't work, or are undesirable, and we shift our attention to the population *median* instead of the mean μ
- The center of population is often best described by the median when the population is not symmetric
- The *order statistics* are the ordered data, from smallest to largest. E.g., if the data are $X = \{65, 23, 41, 23, 3\}$, then the order statistics are $x_{(1)} = 3$, $x_{(2)} = 23$, $x_{(3)} = 23$, $x_{(4)} = 41$, and $x_{(5)} = 65$
- The *median* is the middle value of the data when the values are arranged in order

Finding the Median M

- If n is odd, then the $(n+1)/2$ largest value is M , that is Then, $n = 5$,

$$(n+1)/2 = (5+1)/2 = 3,$$

and $M = x_{(3)}$

- If n is even, then M is defined to be the average of the $n/2$ and $n/2 + 1$ largest values, i.e.,

$$M = \frac{x_{(n/2)} + x_{(n/2+1)}}{2}$$

- Ott and Longnecker (p. 243) define $m = n/2$ so that M can be defined as

$$M = \frac{x_{(m)} + x_{(m+1)}}{2}$$

when n is even

Hypothesis Tests About the Population Median

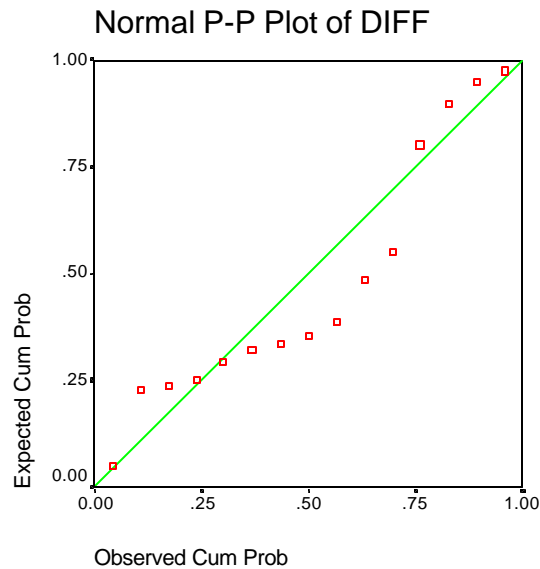
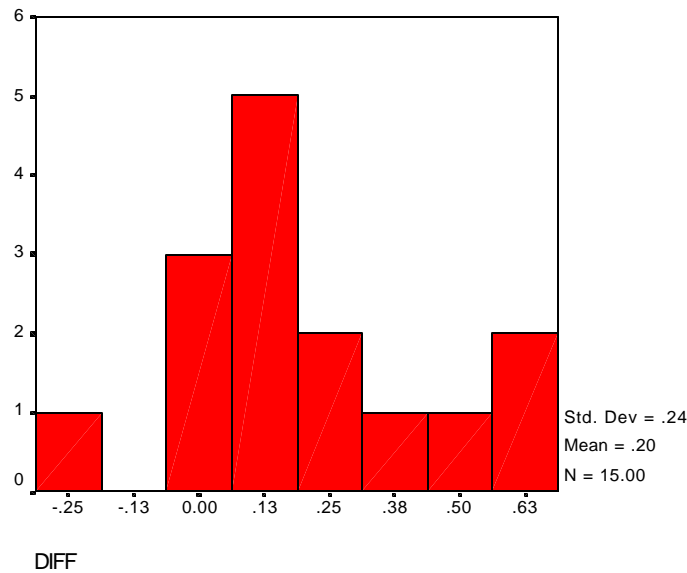
Case Study (from Ramsey and Schafer 1997, p. 29). Early studies based on post-mortem analysis reported that the brains of persons afflicted with schizophrenia were morphologically different from the brains of individuals free of schizophrenia

- Suddath, et al. 1990, "Anatomical abnormalities associated in the brains of monozygotic twins discordant for schizophrenia", New England Journal of Medicine, **322**(12): 789-93, found 15 twins where one was schizophrenic and the other was not. An MRI was used to measure the left hippocampus of the subjects
- Data are differences in volumes (cm^3) in the left hippocampus in 15 sets of monozygotic twins where one is affected with schizophrenia

Pair	Unaffected	Affected	Difference
1	1.94	1.27	0.67
2	1.44	1.63	-0.19
3	1.56	1.47	0.09
4	2.06	1.93	0.13
5	1.58	1.39	0.19
6	1.66	1.26	0.40
7	1.75	1.71	0.04
8	1.77	1.67	0.10
9	1.78	1.28	0.50
10	1.92	1.85	0.07
11	1.25	1.02	0.23
12	1.93	1.34	0.59
13	2.04	2.02	0.02
14	1.62	1.59	0.03
15	2.08	1.97	0.11

- The *differences* between the two measurements on each pair are the data we should use to investigate the question of anatomical abnormalities. Why?

- Paired-data procedures exploit similarities between twins by using the differences, rather than the original data, as the basis for inference
- Let θ denote the *population median* difference (unaffected – affected) between all monozygotic twins (one with schizophrenia, the other without) in left hippocampus volume
- The hypotheses of interest are
 $H_0 : \theta = 0$ versus $H_a : \theta \neq 0$.
- Graphical summaries of the data - histogram and P-P plot



- These $n = 15$ observations do not appear to have been sampled from a normal distribution so the t -statistic is not appropriate (if we specify the α -level, the Type I error rate probably will be different)

The Sign Test

- We assume that the data $X = \{x_1, \dots, x_n\}$ are a random sample from the population of interest
- If $H_0 : \theta = 0$ is true, then the data should be distributed about 0 in equal portions. That is, 50% should be to the left and 50% should be to the right.
- Let

$$X = \# \text{ of observations greater than } \theta_0$$

where θ_0 is the median specified by H_0 . In this case, $\theta_0 = 0$

- If H_0 is true, then $X \sim \text{Bin}(n, \pi = 0.5)$, because each observation is either above or below θ_0 with probability 0.5, and the observations are independent
- The observed value of X is 14 (we expect 7.5 if H_0 were true). The probability of obtaining as much or more evidence (than observed) against H_0 and favoring H_a , given that H_0 is true is

$$\begin{aligned} p\text{-value} &= P(X \geq 14 \mid \pi = 0.5) + P(X \leq 1 \mid \pi = 0.5) \\ &= 2P(X \geq 14 \mid \pi = 0.5) \end{aligned}$$

- Also,

$$\begin{aligned} P(X \geq 14 \mid \pi = 0.5) &= P(14) + P(15) \\ &= \binom{15}{14} \left(\frac{1}{2}\right)^{14} \frac{1}{2} + \binom{15}{15} \left(\frac{1}{2}\right)^{15} \frac{1}{2} \\ &= (15 + 1) \frac{1}{2^{15}} = 0.00048 \end{aligned}$$

- There is very strong evidence that these populations differ with respect to mean left hippocampus volume because $p\text{-value} = 0.00096$
- We cannot say that the difference is caused by schizophrenia. Why not?
- When $n\pi$ and $n(1 - \pi)$ are both larger than 5, then we can use the normal approximation to the binomial to compute approximate binomial probabilities
- In the case of the sign test, $\pi = 0.5$, so we only need $n > 10$

- For example, with $n = 15$, and $\pi = 0.5$,

$$X \sim N(n\pi, \sqrt{n\pi(1-\pi)})$$

with $n\pi = 7.5$ and $\sqrt{n\pi(1-\pi)} = \sqrt{3.75} = 1.94$. Then, using the continuity correction

$$\begin{aligned} P(X \geq 14) &\approx P\left(Z \geq \frac{14 - 7.5 - 0.5}{1.94}\right) \\ &= P\left(Z \geq \frac{14 - 7.5 - 0.5}{1.94}\right) \\ &= 0.00097 \end{aligned}$$

and $p\text{-value} = 2(0.00097) = 0.0019$

Confidence Intervals for θ , the Median

- Recall how a 90% confidence interval for μ was determined. The lower bound L is the 5th percentile of the distribution of \bar{Y} , and the upper bound U is the 95th percentile. The idea is that all values near the estimate \bar{Y} are consistent with the data
- The interval was constructed using the $N(0,1)$ distribution. Specifically, we found all values near the center of the $N(0,1)$ distribution, e.g., all values between the $\alpha/2$ and the $1 - \alpha/2$ percentiles of the $N(0,1)$ distribution
- These percentiles are $-z_{\alpha/2}$ and $z_{\alpha/2}$, and they are inverted using the formulas

$$\begin{aligned} L &= \bar{Y} - z_{\alpha/2}\sigma/\sqrt{n} \\ U &= \bar{Y} + z_{\alpha/2}\sigma/\sqrt{n} \end{aligned}$$

- The same idea is used here
- The interval for θ is constructed using the $\text{Bin}(n, \pi = 0.5)$. Specifically, we
 - 1) find all values near the center of the $\text{Bin}(n, 0.5)$ distribution, e.g., all values between the $\alpha/2$ and $1 - \alpha/2$ percentiles of the distribution of $\text{Bin}(n, 0.5)$
- The $\alpha/2$ percentile is denoted by $C_{\alpha(2),n}$
- The $1 - \alpha/2$ percentile is $n + 1 - C_{\alpha(2),n}$

- $\alpha(2)$ is short-hand for "alpha, two-sided". We use Table 4, p. 1096 to find $C_{\alpha(2),n}$.

2) Invert the values using the formulas

$$L = y_{(C_{\alpha(2),n})}$$

$$U = y_{(n+1-C_{\alpha(2),n})}$$

where $y_{(C_{\alpha(2),n})}$ is the $C_{\alpha(2),n}$ largest value in the sample and $y_{(n+1-C_{\alpha(2),n})}$ is the $n + 1 - C_{\alpha(2),n}$ largest value in the sample

3) Interpret the interval: The procedure (or process) of constructing a $100(1 - \alpha)\%$ CI will contain the true unknown median θ $100(1 - \alpha)\%$ of the time, if the process is repeated many times

Example

- A 90% CI for the median difference in hippocampus volume requires the 0.05 and 0.95 percentiles of the Bin(15,0.5) distribution. From Table 4, p. 1096,

$$C_{\alpha(2),n} = C_{0.05,15} = 3$$

and $n + 1 - C_{\alpha(2),n} = 15 + 1 - 3 = 13$.

- Going back to the original data, we find that $L = y_{(3)} = 0.03$ and $U = y_{(13)} = 0.50$
- The 90% CI for θ is 0.03 to 0.50. I am 90% confident that the true value of θ is in the interval
- If $C_{\alpha(2),n} = 0$, then there is insufficient data to construct a confidence interval for θ

Large-Sample Approximation

- The lower limit is determined by

$$C_{\alpha(2),n} = \frac{n}{2} - z_{\alpha/2} \sqrt{\frac{n}{4}}$$

and the upper limit is determined by

$$n + 1 - C_{\alpha(2),n}$$

Example: $n = 15$, $\alpha = 0.10$ implies that $z_{\alpha/2} = 1.645$,

$$C_{\alpha(2),n} \approx \frac{15}{2} - 1.645 \sqrt{\frac{15}{4}} = 4.3$$

- Our convention is to round down, so $C_{\alpha(2),n} \approx 4$ and $n + 1 - C_{\alpha(2),n} = 15 - 4 = 11$. Then $L = y_{(4)} = 0.04$ and $U = y_{(12)} = 0.40$.

- The 90% CI for θ is 0.04 to 0.40.

Table 5. (p. 248 of Ott and Longnecker) shows that when $n \leq 20$, the sign test has substantially larger power than the t -statistic when the population distribution is heavy-tailed or is highly skewed. The t -statistic is comparable to the sign test with respect to power when the population is lightly skewed.

Example 5.20 - The Sign Test (Ott and Longnecker, p. 247)

- Instead of directly computing the p -value, as discussed in class, Ott and Longnecker have a different method. We will use example 5.20 to illustrate their method

- The hypotheses of interest are $H_0 : M \leq 5$ and $H_a : M > 5$. M represents the population median. I use θ instead of M .

- The test statistic is

$$B = \# \text{ of observations larger than } M,$$

though, Ott and Longnecker have a longer definition: Let i index the observations Y_1, \dots, Y_n , and let $W_i = Y_i - M$. Then, $B = \#$ of W_i 's that are positive.

- The value of B is 13 based on $n = 25$ observations. Note that under H_0 we expect to get $n\pi = 25 \times 0.5 = 12.5$ observations larger than M

- There really is no point in continuing- there is no evidence in favor of H_a . However, let's continue.

- The rejection region is determined as follows. If H_a is true, then we expect that more observations will be larger than M than are smaller than M , hence, B will be larger than $n/2$ if H_a is true.

- The larger B is, the more evidence against H_0 . We reject H_a if B is sufficiently large
- Given $\alpha = 0.05$, say, look up the critical value $C_{\alpha(2),n}$ in Table 5. Because $n = 25$ for this example, and H_a is one-sided, we look up $C_{\alpha(1),n} = C_{0.05,25}$ in Table 5
- The entries in the Table are values of B satisfying

$$P(B \leq b) \approx \alpha$$

The entry in the Table is 7. We need the other tail of the distribution to form the rejection region. Because the distribution of B is symmetric,

- The rejection region is all values of B that are larger than $n - C_{\alpha(1),n} = 25 - 7 = 18$. Because $B = 13 < 18$, we do not have sufficient evidence to reject H_0 in favor of H_a .
- If H_a were $H_a : M < 5$, then the rejection region would be all values of B that are smaller than the critical value $C_{\alpha(1),n}$
- If H_a were $H_a : M \neq 5$, then the rejection region would be all values of B that are smaller than the lower critical value $C_{\alpha(2),n} = 7$ and all values of B that are larger than the upper critical value $n - C_{\alpha(2),n} = 25 - 7 = 18$.

Review Problems

p. 256, 5.90

Let μ represent the mean time (minutes) to fill an order (recently). I will test

$$H_0 : \mu = 25 \text{ versus } H_a : \mu > 25.$$

The test statistic is

$$t = \frac{\bar{Y} - \mu_0}{s/\sqrt{n}},$$

where $\mu_0 = 25$. Then,

$$t = \sqrt{15} \frac{28.2 - 25}{11.44} = 1.08.$$

Using $df = 14$ produces $p\text{-value} = P(T > 1.08) = 0.148$. I conclude that there is insufficient evidence to conclude that μ is larger than 25 minutes.

p. 257, 5.94

a) F, b) F, c) T, d) F, e) T, f) F, g) T h) F

p. 257, 5.95

a) T, b) F c) T, d) T

p. 257, 5.96

a) approximately normal, to the population mean

b) smaller

c) larger

d) Central Limit Theorem e) Type II