

Statistical Data Analysis

Assist. Prof. Dr. Zeyneb KURT

(Slides have been prepared by
Prof. Dr. Nizamettin AYDIN,
updated by Zeyneb KURT)

zeyneb@yildiz.edu.tr

<http://avesis.yildiz.edu.tr/zeyneb/>

Hypothesis Testing

Hypothesis

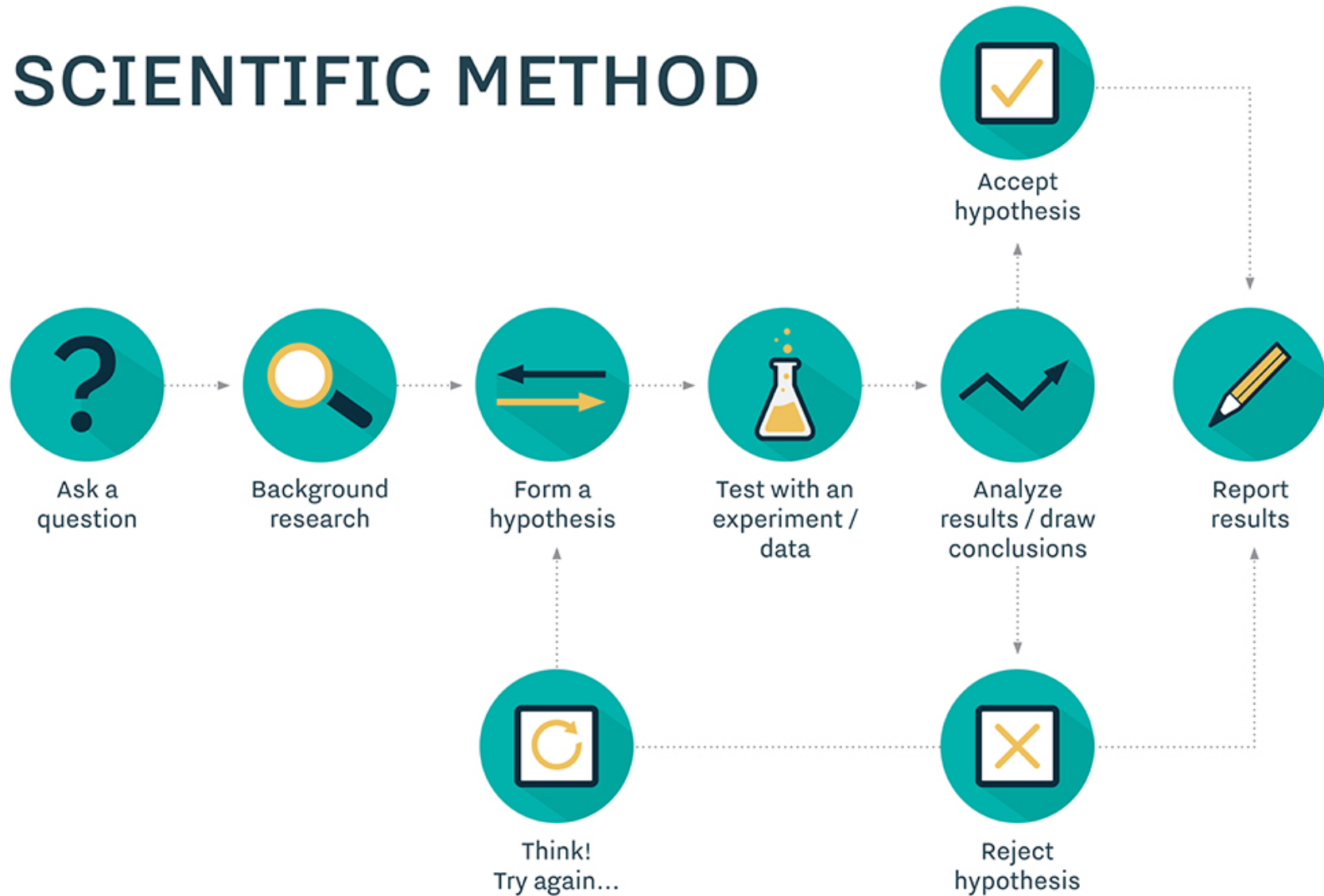
- A **hypothesis** (*plural: hypotheses*), in a scientific context, is a testable statement about the relationship between two or more variables or a proposed explanation for some observed phenomenon.
- In a scientific experiment or study, the hypothesis is a brief summation of the researcher's prediction of the study's findings, which may be supported or not by the outcome.
- Hypothesis testing is the core of the **scientific method**.

Hypothesis

- **Scientific method** is an approach to seeking knowledge that involves forming and testing a hypothesis.
- This methodology is used to answer questions in a wide variety of disciplines outside of science, including business.
- Scientific method provides a logical, systematic way to answer questions and removes subjectivity by requiring each answer to be authenticated with objective evidence that can be reproduced.
- Goal of scientific method is to gather data that will validate or invalidate a cause and effect relationship.
- Scientific method is often carried out in a linear manner, but the approach can also be cyclical, because once a conclusion has been reached, it often raises more questions.

Hypothesis

SCIENTIFIC METHOD



Hypothesis

- In general, many scientific investigations start by expressing a hypothesis.
- To evaluate hypotheses, we rely on
 - estimators,
 - their sampling distributions,
 - their specific valuesfrom observed data.
- For example,
 - Mackowiak et al.* hypothesized that the average normal (i.e., for healthy people) body temperature is less than the widely accepted value of 98.6°F.
 - If we denote the population mean of normal body temperature as μ , then we can express this hypothesis as $\mu < 98.6$.

*Mackowiak, P.A., Wasserman, S.S., Levine, M.M.: A critical appraisal of 98.6°F, the upper limit of the normal body temperature, and other legacies of Carl Reinhold August Wunderlich. JAMA 268, 1578–1580 (1992)

Null and Alternative hypotheses

- The **null hypothesis** usually reflects the “status quo” or “nothing of interest”.
- In contrast, we refer to our hypothesis (i.e., the hypothesis we are investigating through a scientific study) as the **alternative hypothesis** and denote it as H_A .
- Consider the body temperature example, where we want to examine the null hypothesis $H_0 : \mu = 98.6$ against the alternative hypothesis $H_A : \mu < 98.6$.
- The procedure for evaluating a hypothesis is called **hypothesis testing**, and it arises in many scientific problems.
- For hypothesis testing, we focus on the null hypothesis since it tends to be simpler.
- To this end, we examine the evidence that the observed data provide against the null hypothesis H_0 .
 - If the evidence against H_0 is strong, we reject H_0 .
 - If not, we state that the evidence provided by the data is not strong enough to reject H_0 , and we fail to reject it.

Null and Alternative hypotheses

- With respect to our decision regarding the null hypothesis H_0 , we might make two types of errors:
 - Type I error:
 - we reject H_0 when it is true and should not be rejected.
 - Type II error:
 - we fail to reject H_0 when it is false and should be rejected.
- We denote the probability of making type I error as α and the probability of making type II error as β .

| Decision Made | Actual Validity of H_0 | | |
|---------------|--------------------------|----------------------------------|-----------------------------------|
| | | H_0 is true | H_0 is false |
| | Accept H_0 | True Negative | False Negative (Type II Error) |
| | Reject H_0 | False Positive (Type I Error) | True Positive |

Null and alternative hypotheses

- Now suppose that we have a hypothesis testing procedure that fails to reject the null hypothesis when it should be rejected with probability β .
 - This means that our test correctly rejects the null hypothesis with probability $1 - \beta$.
 - Note that the two events are complementary.
 - We refer to this probability (i.e., $1 - \beta$) as the power of the test.
- In practice, it is common to first agree on a tolerable type I error rate α , such as 0.01, 0.05, and 0.1.
- Then try to find a test procedure with the highest power among all reasonable testing procedures.

Hypothesis testing for the population mean

- To decide whether we should reject the null hypothesis, we quantify the empirical support (provided by the observed data) against the null hypothesis using some statistics.
- We use statistics to evaluate our hypotheses.
 - We refer to them as test statistics.
 - To evaluate hypotheses regarding the population mean, we use the sample mean \bar{X} as the test statistic
- For a statistic to be considered as a test statistic, its sampling distribution must be fully known (exactly or approximately) under the null hypothesis.
 - We refer to the distribution of test statistics under the null hypothesis as the null distribution.
 - For the sample mean, the CLT states that the sampling distribution is approximately normal when the sample size is large.

Hypothesis testing for the population mean

- Consider the body temperature example, where we want to examine the null hypothesis $H_0 : \mu = 98.6$ against the alternative hypothesis $H_A : \mu < 98.6$.
- To start with, suppose that $\sigma^2 = 1$ is known and we have randomly selected a sample of 25 healthy people from the population and measured their body temperature.
- Using the CLT, the sampling distribution of \bar{X} is approximately normal as follows:

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

- For the above example,

$$\bar{X} \sim N(\mu, 1/25)$$

- If the null hypothesis is true and the population mean is $\mu = 98.6$,
- the sampling distribution of \bar{X} becomes

$$\bar{X} | H_0 \sim N(98.6, 0.04)$$

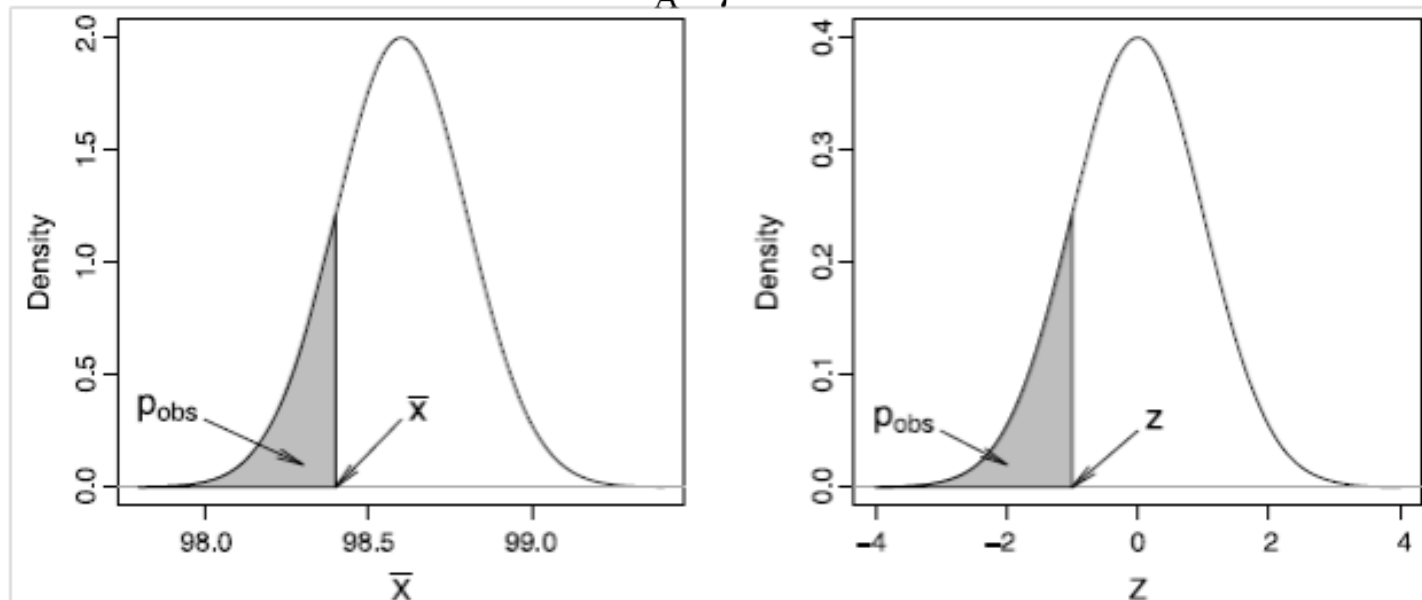
- Note that the distribution of \bar{X} is obtained conditional (hence the notation for conditional probability) on the assumption that the null hypothesis is true

Hypothesis testing for the population mean

- In reality, we have one value, \bar{x} , for the sample mean.
- We can use this value to quantify the evidence of departure from the null hypothesis.
- Suppose that from our sample of 25 people we find that the sample mean is $\bar{x} = 98.4$
- To evaluate the null hypothesis $H_0 : \mu = 98.6$ versus the alternative $H_A : \mu < 98.6$,
 - we use the lower tail probability of this value from the null distribution.

Hypothesis testing for the population mean

- For the normal body temperature example, examining the hypotheses $H_0 : \mu = 98.6$ versus the alternative $H_A : \mu < 98.6$.



- *Left panel:* The shaded area shows the lower-tail probability of the observed sample mean, $\bar{x} = 98.4$. This is the observed significance level, p -value, which is denoted as p_{obs} .
- *Right panel:* After standardizing, the p -value corresponds to the lower tail probability of $z = -1$ based on the standard normal distribution

Observed significance level

- The observed significance level for a test is the probability of values as or more extreme than the observed value, based on the null distribution in the direction supporting the alternative hypothesis.
- This probability is also called the **p-value** and denoted as p_{obs} .

- For the above example,

$$p_{obs} = P(\bar{X} \leq \bar{x} | H_0) = P(\bar{X} \leq 98.4) = 0.16$$

- To find the p -value in R-Commander,
 - click *Distributions* → *Continuous distributions* → *Normal distribution* → *Normal probabilities*.
 - Then set the *Variable value* to 98.4 and the parameters for the null distribution ($\mu = 98.6$ and $\sigma = 0.2$).

z-Tests of the Population Mean

- In practice, it is more common to use the standardized version of the sample mean as our test statistic.
- We know that if a random variable is normally distributed (as it is the case for \bar{X}), subtracting the mean and dividing by standard deviation creates a new random variable with standard normal distribution, $Z \sim N(0,1)$
- We refer to the standardized value of the observed test statistic as the **z-score**,

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{98.4 - 98.6}{0.2} = -1$$
$$p_{obs} = P(Z \leq -1) = 0.16$$

- We refer to the corresponding hypothesis test of the population mean as the **z-test**.
- In a z-test, instead of comparing the observed sample mean \bar{x} to the population mean according to the null hypothesis, we compare the z-score to 0

Interpretation of p-value

- The p-value is the conditional probability of extreme values (as or more extreme than what has been observed) of the test statistic assuming that the null hypothesis is true.
 - When the p-value is small, say 0.01 for example, it is rare to find values as extreme as what we have observed (or more so).
- As the p-value increases, it indicates that there is a good chance to find more extreme values (for the test statistic) than what has been observed.
 - Then, we would be more reluctant to reject the null hypothesis.
- The common significance levels are 0.01, 0.05, and 0.1.

Interpretation of p-value

- If p_{obs} is less than the assumed cutoff, we say that the data provides statistically significant evidence against H_0
- For the body temperature example where $p_{\text{obs}} = 0.16$, if we set the significance level at 0.05, we say that there is not significant evidence against the null hypothesis $H_0: \mu = 98.6$ at the 0.05 significance level, so we do not reject the null hypothesis.
- If we had set the cutoff at 0.05 and we had observed $\bar{x} = 98.25$ instead of 98.4, then $p_{\text{obs}} = 0.04$, and we could reject the null hypothesis.
- In this case, we say that the result is statistically significant and the data provide enough evidence against $H_0: \mu = 98.6$.

One-Sided Hypothesis Testing

- In general, for one-sided hypothesis testing, we evaluate the null hypothesis $H_0 : \mu = \mu_0$ by using the following standardized test statistic:

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

- We find the sample mean \bar{x} and calculate the observed value of Z called z-score (assuming σ is known):

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

One-Sided Hypothesis Testing

- We then use the standard normal distribution to find the p -value.
- If the alternative hypothesis regarding the population mean is $H_A : \mu < \mu_0$, we use the standard normal distribution to find lower tail probability of the z -score: $P(Z \leq z)$.
- If the alternative hypothesis regarding the population mean is $H_A : \mu > \mu_0$, we use $P(Z \geq z)$ instead.
- The resulting probability, p_{obs} , is the observed significance level, which can be compared to several significance levels such as 0.01, 0.05, and 0.1.

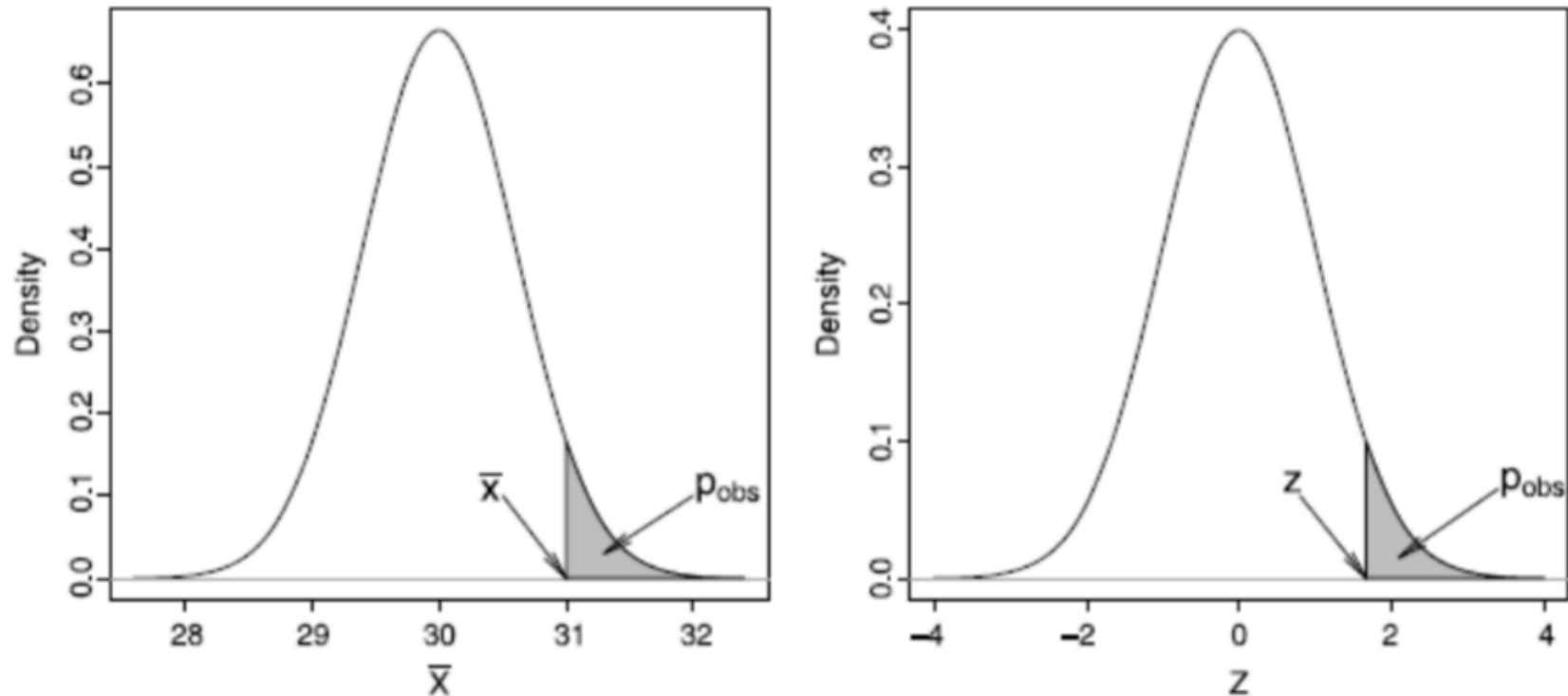
One-Sided Hypothesis Testing

- In some situations, we might hypothesize that the population mean is greater than a specific value and express our hypothesis as $H_A : \mu > \mu_0$.
- Our null hypothesis is still $H_0 : \mu = \mu_0$.
- This is also a one-sided test since the departure from the null is still in one direction: toward values larger than μ_0 .
- For example, suppose that we have observed that many Pima Indian women suffer from diabetes.
- We know that obesity and diabetes are related; we might therefore hypothesize that this population is obese on average, where obesity is defined as BMI higher than 30.
- If we denote the population mean of BMI for Pima Indian women, we can then express our hypothesis as $\mu > 30$.
- In this case, the null hypothesis is $H_0 : \mu = 30$; that is, $\mu_0 = 30$.

One-Sided Hypothesis Testing

- As before, we use the sample mean as the test statistic. For illustrative purposes,
 - suppose that we have obtained a sample of size $n = 100$ from the population of Pima
 - Indian women. Further, suppose we know that the population variance is $\sigma^2 = 6^2$.
 - If the null hypothesis is true and the population mean is $\mu = 30$, then the sampling distribution is
$$\bar{X}|H_0 \sim N(30, 6^2/100)$$
- The distribution is shown in the left panel of the figure in the next slide

One-Sided Hypothesis Testing



- *Left panel:* The sampling distribution for the test statistic under the null hypothesis $H_0 : \mu = 30$. The p -value, which is the probability of values as or more extreme than the observed value of the test statistic $\bar{x} = 31$, is shown as the *shaded area*.
- *Right panel:* Obtaining the upper tail probability using one-sided z -test

One-Sided Hypothesis Testing

- If the null hypothesis is indeed true, then we would expect to see the value of sample mean near the population mean according to the null distribution (here, 30).
- In contrast, if the null hypothesis is false, then the null distribution does not represent the sampling distribution of the test statistics, and we would expect to see the value of the sample mean away from 30, in this case, larger than 30 according to the alternative hypothesis.
- Suppose that from our sample of 100 Pima Indian women we find that the sample mean is $\bar{x} = 31$.
- As before, we find the observed significance level, p -value, to measure the amount of evidence provided by the data in support for H_0 .

One-Sided Hypothesis Testing

- Recall that we defined p -value as the probability of values as or more extreme than the observed value of the test statistic (here, $\bar{x} = 31$) based on the null distribution, in the direction specified by the alternative hypothesis.
- If the null distribution is in fact true and $\mu = 30$, then values larger than $\bar{x} = 31$ would seem more extreme than what we have observed.
- Therefore, since $H_A : \mu > \mu_0$

$$p_{\text{obs}} = P(\bar{X} \geq \bar{x} | H_0)$$

- If we drop H_0 for simplicity,

$$p_{\text{obs}} = P(\bar{X} \geq 31)$$

- This probability is shown as the shaded area in the left panel of the figure (in slide 20).

One-Sided Hypothesis Testing

- We can standardize the test statistic by subtracting the mean and dividing the result by the standard deviation:

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X} - 30}{\frac{6}{\sqrt{100}}} = \frac{\bar{X} - 30}{0.6} \sim N(0, 1)$$

- The corresponding z-score is obtained as follows:

$$z = \frac{\bar{x} - \mu_0}{0.6} = \frac{31 - 30}{0.6} = 1.67$$

- Now, to find the p -value, we can find the upper tail probability of $z = 1.67$ from the null distribution $N(0, 1)$:

$$p_{\text{obs}} = P(Z \geq 1.67)$$

- This probability is shown as the shaded area in the right panel of the figure (in slide 20).
 - This is the upper tail probability at 1.67 based on the standard normal distribution.

One-Sided Hypothesis Testing

- The upper tail probability is by convention $P(Z > 1.67)$.
- However, for continuous random variables, $P(Z > 1.67) = P(Z \geq 1.67)$ since the probability of any specific value (here, 1.67) is zero.
- For this example, $p_{\text{obs}} = 0.048$.
- We can reject the null hypothesis at 0.05 level but not at 0.01 level.
- At 0.05 level, we can conclude that the population mean of BMI for Pima Indian women is higher than 30 and the difference is statistically significant.

Two-Sided Hypothesis Testing

- For many hypothesis testing problems, we might be indifferent to the direction of departure from the null value.
 - In such cases, we can express the null and alternative hypotheses as $H_0 : \mu = \mu_0$ and $H_A : \mu \neq \mu_0$, respectively.
 - Then we consider both large positive values and small negative values of z-score as evidence against the null hypothesis, and our alternative hypothesis is referred to as two-sided.
- The p -value for the two-sided hypothesis test is calculated as follows (assuming σ is known):
 - Determine the observed z-score: $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$
 - Take the absolute value of the z score: $|z|$.
 - Obtain the upper tail probability: $P(Z \geq |z|)$.
 - Double the resulting probability: $p_{\text{obs}} = 2 \times P(Z \geq |z|)$.

Two-Sided Hypothesis Testing

- For example, suppose we believe that the average normal body temperature is different from the accepted value 98.6°F, but we are not sure whether it is higher or lower than 98.6.
- Then the null hypothesis remains $H_0 : \mu = 98.6$, but the alternative hypothesis is expressed as $H_A : \mu \neq 98.6$.
- We calculate the sample mean $\bar{x} = 98.4$ and standardize it to obtain the z-score,

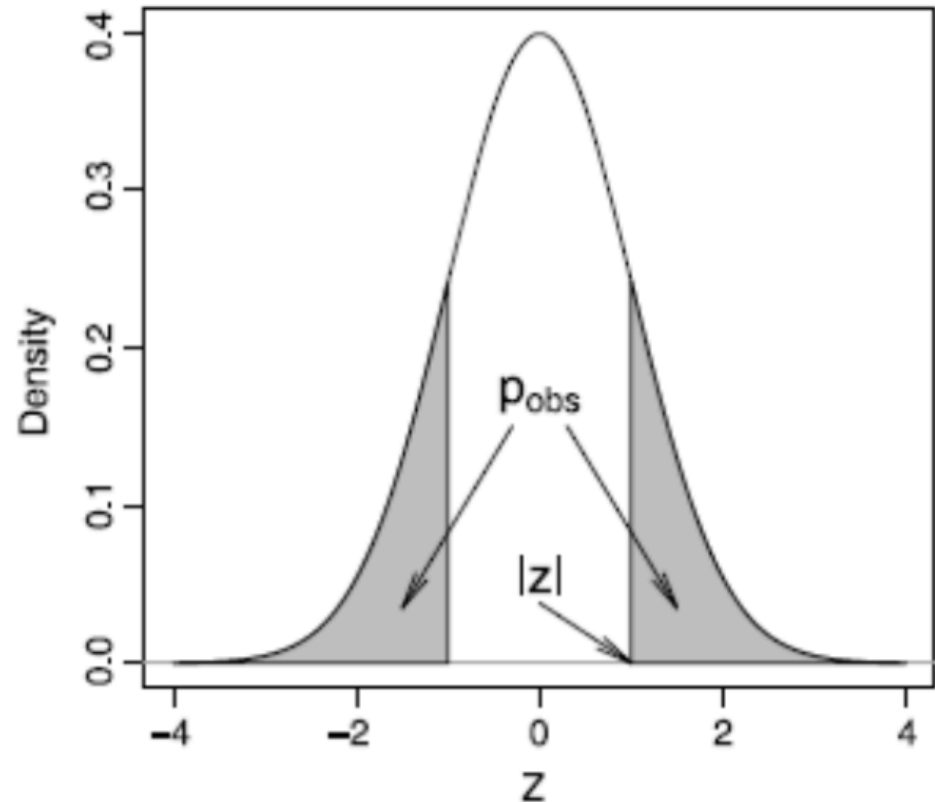
$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{98.4 - 98.6}{1/\sqrt{25}} = \frac{98.4 - 98.6}{0.2} = \frac{-0.2}{0.2} = -1$$

Two-Sided Hypothesis Testing

- The p -value is still calculated as the probability of values as or more extreme than the observed z -score.
 - However, in this case, extreme values are those whose distance from 0 is more than the distance of -1 from zero.
 - These are values that are either less than -1 or greater than 1 .
- Therefore, to find the observed significance level, we need to add the probabilities for $Z \leq -1$ and $Z \geq 1$:
$$p_{\text{obs}} = P(Z \leq -1) + P(Z \geq 1)$$
- This probability is equal to the shaded area in the following figure

Two-Sided Hypothesis Testing

- Illustrating the p -value for a two-sided hypothesis test of average normal body temperature, where $H_0 : \mu = 98.6$ and $H_A : \mu \neq 98.6$.



- After standardizing,
$$p_{\text{obs}} = P(Z \leq -1) + P(Z \geq 1) = 2 \times 0.16 = 0.32$$
- The p -value is greater than typical significance levels such as 0.01, 0.05, and 0.1, so we cannot reject it at these levels.
- Therefore, we conclude that the observed difference is not statistically significant, and could be due to chance alone.

Hypothesis testing using t-tests

- So far, we have assumed that the population variance σ^2 is known.
 - In reality, σ^2 is almost always unknown, and we need to estimate it from the data.
- As before, we estimate σ^2 using the sample variance S^2 .
- Similar to our approach for finding confidence intervals, we account for this additional source of uncertainty by using the t-distribution with $n-1$ degrees of freedom instead of the standard normal distribution.
 - The hypothesis testing procedure is then called the t -test.

Hypothesis testing using t-tests

- To perform a t -test , we use the following test statistic:

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

- The test statistic, T , has a t -distribution with $n - 1$ degrees of freedom under the null.

$$T \sim t(n - 1)$$

- Using the observed values of \bar{X} and S , the observed value of the test statistic is obtained as follows:

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

- We refer to t as the t -score.

Hypothesis testing using t-tests

- To assess the null hypothesis $H_0 : \mu = \mu_0$ using the t - test,
 - we first calculate the t -score based on the observed sample mean \bar{x} and sample standard deviation.
- We then calculate the corresponding p -value as follows:
 - if $H_A : \mu < \mu_0$, $p_{\text{obs}} = P(T \leq t)$,
 - if $H_A : \mu > \mu_0$, $p_{\text{obs}} = P(T \geq t)$,
 - if $H_A : \mu \neq \mu_0$, $p_{\text{obs}} = 2 \times P(T \geq |t|)$,where T has a t -distribution with $n - 1$ degrees of freedom, and t is our observed t -score.
- This is known as the **single-sample t -test**.

Hypothesis testing using t-tests - example

- Suppose we hypothesize that the population mean of BMI among Pima Indian women is above 30: $H_A : \mu > 30$.
- The corresponding null hypothesis is $H_0 : \mu = 30$.
- To test this hypothesis, we use the Pima.tr data set from the MASS package.
- The sample size is $n = 200$.
- The sample mean and standard deviation are $\bar{x} = 32.31$ and $s = 6.13$, respectively. The t -score is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{32.31 - 30}{6.13/\sqrt{200}} = 5.33$$

Hypothesis testing using t-tests - example

- For the above example, $p_{\text{obs}} = P(T \geq 5.33)$, which we obtain from the t -distribution with $200 - 1 = 199$ degrees of freedom.
- To obtain this probability in R-Commander,
 - click *Distributions* → *Continuous Distributions* → *t distribution* → *t probabilities*.
 - Then enter 5.33 for *Variable value* and 199 for *Degrees of freedom*, and select *Upper tail*.
 - The resulting probability is 1.33×10^{-07} .
- At any reasonable significance level, there is strong evidence to reject the null hypothesis and conclude that the population mean of BMI among Pima Indian women is in fact greater than 30.
 - Therefore, on average, the population is obese.

Hypothesis testing for population proportion

- For a binary random variable X with possible values 0 and 1, we are typically interested in evaluating hypotheses regarding the population proportion of the outcome of interest, denoted as $X = 1$.
- As discussed before, the population proportion is the same as the population mean for such binary variables.
- So we follow the same procedure as described above.
- More specifically, we use the z -test for hypothesis testing.
- Note that we do not use t -test, because for binary random variable, population variance is $\sigma^2 = \mu(1 - \mu)$.

Hypothesis testing for population proportion

- In general, to assess the null hypothesis $H_0 : \mu = \mu_0$, where μ is the population proportion (mean) of a binary random variable,
 - we first calculate z-score based on the observed sample proportion p :

$$z = \frac{p - \mu_0}{\sqrt{\mu_0 (1 - \mu_0)/n}}$$

- Then we determine the support for the null hypothesis as:
 - if $H_A : \mu < \mu_0$, $p_{\text{obs}} = P(Z \leq z)$,
 - if $H_A : \mu > \mu_0$, $p_{\text{obs}} = P(Z \geq z)$,
 - if $H_A : \mu \neq \mu_0$, $p_{\text{obs}} = 2 \times P(Z \geq |z|)$,

where Z has the standard normal distribution, and z is the observed z-score.

Hypothesis testing for population proportion

- Consider the **Melanoma** dataset available from the MASS package.
- Suppose that we hypothesize that less than 50% of cases ulcerate: $\mu < 0.5$.
- Then the null hypothesis can be expressed as $H_0 : \mu = 0.5$.
- Using the **Melanoma** data set, we can test the above null hypothesis.
- The number of observations in this data set is $n = 205$, of which 90 patients had ulceration.
- Therefore,

$$p = 90/205 = 0.44$$

Hypothesis testing for population proportion

- Next, we can find the z -score for our test statistic as follows:

$$z = \frac{p - \mu_0}{\sqrt{\mu_0 (1 - \mu_0)/n}} = \frac{0.44 - 0.5}{\sqrt{0.5 (1 - 0.5)/202}} = -1.72$$

- Because $H_A : \mu < 0.5$, the observed significance level based on this z -score is the lower tail probability $P(Z \leq -1.72)$.
- Using R-Commander, we find the p -value to be $p_{\text{obs}} = 0.043$.
- Therefore, we can reject the null hypothesis at 0.05 level but not at 0.01 level.

Test of Normality

- Visual inspection of the distribution may be used for assessing normality, although this approach is usually unreliable and does not guarantee that the distribution is normal.
 - The frequency distribution (histogram),
 - stem-and-leaf plot,
 - boxplot,
 - P-P plot (probability-probability plot),
 - Q-Q plot (quantile-quantile plot)are used for checking normality visually

Test of Normality

- The appropriateness of the normality assumption can be evaluated formally using a testing procedure such as
 - Kolmogorov-Smirnov (K-S) test
 - Lilliefors corrected K-S test
 - Shapiro-Wilk test
 - Anderson-Darling test
 - Cramer-von Mises test
 - D'Agostino skewness test
 - Anscombe-Glynn kurtosis test
 - D'Agostino-Pearson omnibus test
 - the Jarque-Bera test

Test of Normality - Example

- More specifically, this test evaluates the null hypothesis that the distribution of the random variable is normal.
- As usual, we then either reject this hypothesis and conclude that the normality assumption is not appropriate, or fail to reject it and conclude that there is no strong evidence of deviation from normality.
- Suppose we assume that the *bmi* variable in **Pima.tr** has normal distribution.

Test of Normality - Example

- To evaluate this assumption, In R-Commander,
 - click *Statistics* → *Summaries* → *Test of normality* and chose *Shapiro-Wilk*,
 - then select the *bmi*.
- The p -value for this test is 0.25.
- Therefore, we do not reject the null hypothesis (which states that the distribution is normal) and conclude that the deviation of the distribution from normality is not statistically significant.

Test of Normality - Example

- For comparison, repeat the above steps to test the normality assumption for the *age* variable in the **Pima.tr** data set.
- Using the Shapiro–Wilk test, the p -value is 1.853×10^{-12} , which is very small.
- Therefore, we can comfortably reject the null hypothesis and conclude that the deviation from normality is statistically significant.

Test of Normality - Example

Test of Normality

Variable (pick one)

- age
- bmi**
- bp
- glu
- npreg
- ped

Normality Test

☒ Shapiro-Wilk

☐ Anderson-Darling

☐ Cramer-von Mises

☐ Lilliefors (Kolmogorov-Smirnov)

☐ Shapiro-Francia

☐ Pearson chi-square

Number of bins for Pearson chi-square <

Test by groups...

Help Reset OK Cancel

Shapiro-Wilk normality test

```
data:  bmi
W = 0.99104, p-value = 0.2523
```

Test of Normality

Variable (pick one)

- age**
- bmi
- bp
- glu
- npreg
- ped

Normality Test

☒ Shapiro-Wilk

☐ Anderson-Darling

☐ Cramer-von Mises

☐ Lilliefors (Kolmogorov-Smirnov)

☐ Shapiro-Francia

☐ Pearson chi-square

Number of bins for Pearson chi-square <

Test by groups...

Help Reset OK Cancel

Shapiro-Wilk normality test

```
data:  age
W = 0.86268, p-value = 1.853e-12
```

Hypothesis Testing with R Programming

- To perform the z-test in R, use the function `pnorm()`
- For the body temperature example discussed earlier, the z-score was -1 .
 - For the one-sided hypothesis of the form $H_A : \mu < \mu_0$, we find the lower tail probability of -1 as follows:

```
> pnorm(-1, mean = 0, sd = 1, lower.tail = TRUE)
```

```
[1] 0.1586553
```
- For the BMI example, z-score was 1.67 .
 - For the one-sided hypothesis of the form $H_A : \mu > \mu_0$, we need to find the upper tail probability of 1.67 as follows:

```
> pnorm(1.67, mean = 0, sd = 1, lower.tail = FALSE)
```

```
[1] 0.04745968
```

Hypothesis Testing with R Programming

- When σ^2 is unknown and we need to use the data to estimate it separately, we use the t -test to evaluate hypotheses regarding the mean of a normal distribution.
- For the BMI example discussed earlier, the t -score was $t = 5.33$.
 - For the one-sided hypothesis of the form $H_A : \mu > \mu_0$, we find the upper tail probability of 5.33 from a t distribution with $n-1$ degrees of freedom, where $n = 200$ in this example.
 - We use the `pt()` function:

```
> pt(5.33, df = 199, lower.tail = FALSE)  
[1] 0.0000001324778
```

Statistical Inference for the Relationship Between Two Variables

Relationship Between a Numerical Variable and a Binary Variable

- In general, we can denote the means of the two groups as μ_1 and μ_2 .
- The null hypothesis indicates that the population means are equal, $H_0 : \mu_1 = \mu_2$.
- In contrast, the alternative hypothesis is one the following:
 - if $H_A : \mu_1 > \mu_2$, if we believe the mean for group 1 is greater than the mean for group 2.
 - if $H_A : \mu_1 < \mu_2$, if we believe the mean for group 1 is less than the mean for group 2.
 - if $H_A : \mu_1 \neq \mu_2$, if we believe the means are different but we do not specify which one is greater.
- We can also express these hypotheses in terms of the difference in the means:
$$H_A : \mu_1 - \mu_2 > 0, H_A : \mu_1 - \mu_2 < 0, \text{ or } H_A : \mu_1 - \mu_2 \neq 0$$
- Then the corresponding null hypothesis is that there is no difference in the population means, $H_0 : \mu_1 - \mu_2 = 0$

Relationship Between a Numerical Variable and a Binary Variable

- Previously, we used the sample mean \bar{X} to perform statistical inference regarding the population mean μ .
- To evaluate our hypothesis regarding the difference between two means, $\mu_1 - \mu_2$, it is reasonable to choose the difference between the sample means, $\bar{X}_1 - \bar{X}_2$, as our statistic.
- We use μ_{12} to denote the difference between the population means μ_1 and μ_2 , and use \bar{X}_{12} to denote the difference between the sample means \bar{X}_1 and \bar{X}_2 :

$$\mu_{12} = \mu_1 - \mu_2 \qquad \bar{X}_{12} = \bar{X}_1 - \bar{X}_2$$

Relationship Between a Numerical Variable and a Binary Variable

- By the Central Limit Theorem,

$$\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right), \quad \bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

where n_1 and n_2 are the number of observations.

- Therefore,

$$\bar{X}_{12} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

- We can rewrite this as

$$\bar{X}_{12} \sim N(\mu_{12}, SD_{12}^2) \text{ where } SD_{12} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Relationship Between a Numerical Variable and a Binary Variable

- We want to test our hypothesis that $H_A : \mu_{12} \neq 0$ (i.e., the difference between the two means is not zero) against the null hypothesis that $H_0 : \mu_{12} = 0$.
- To use \bar{X}_{12} as a test statistic, we need to find its sampling distribution under the null hypothesis (i.e., its null distribution).
- If the null hypothesis is true, then $\mu_{12} = 0$.
- Therefore, the null distribution of \bar{X}_{12} is
$$\bar{X}_{12} \sim N(0, SD_{12}^2)$$
- As before, however, it is more common to standardize the test statistic by subtracting its mean (under the null) and dividing the result by its standard deviation.

$$Z = \frac{\bar{X}_{12}}{SD_{12}}$$

- where Z is called the z-statistic, and it has the standard normal distribution: $Z \sim N(0, 1)$.

Two-sample z-test

- To test the null hypothesis $H_0 : \mu_{12} = 0$, we determine the z-score,

$$Z = \frac{\bar{x}_{12}}{SD_{12}}$$

- Then, depending on the alternative hypothesis, we can calculate the p-value, which is the observed significance level, as:
 - if $H_A : \mu_{12} > 0$, $p_{\text{obs}} = P(Z \geq z)$,
 - if $H_A : \mu_{12} < 0$, $p_{\text{obs}} = P(Z \leq z)$,
 - if $H_A : \mu_{12} \neq 0$, $p_{\text{obs}} = 2 \times P(Z \geq |z|)$,
- The above tail probabilities are obtained from the standard normal distribution.

Example

- Suppose that our sample includes $n_1 = 25$ women and $n_2 = 27$ men.
- The sample mean of body temperature is $\bar{x}_1 = 98.2$ for women and $\bar{x}_2 = 98.4$ for men.
- Then, our point estimate for the difference between population means is $\bar{x}_{12} = -0.2$.
- We assume that $\sigma_1^2 = 0.8$ and $\sigma_2^2 = 1$.
- The variance of the sampling distribution is $(0.8/25) + (1/27) = 0.07$, and the standard deviation is $SD_{12} = \sqrt{0.07} = 0.26$.
- The z-score is $z = \frac{\bar{x}_{12}}{SD_{12}} = \frac{-0.2}{0.26} = -0.76$

Example

- $H_A : \mu_{12} \neq 0$ and $z = -0.76$.
- Therefore, $p_{\text{obs}} = 2P(Z \geq |-0.76|) = 2 \times 0.22 = 0.44$.
- For the body temperature example, $p_{\text{obs}} = 0.44$ is greater than the commonly used significance levels (e.g., 0.01, 0.05, and 0.1).
- Therefore, the test result is not statistically significant, and we cannot reject the null hypothesis (which states that the population means for the two groups are the same) at these levels.
- That is, any observed difference could be due to chance alone.

Two-Sample t-test

- In practice, SD_{12} is not known since σ_1 and σ_2 are unknown.
- As before, we can use the sample variances S_1^2 and S_2^2 to estimate σ_1^2 and σ_2^2 , and take this additional source of uncertainty into account by using t -distributions instead of the standard normal distribution.
- We use s_1^2 and s_2^2 (point estimates for population variances σ_1^2 and σ_2^2) to estimate the standard deviation,

$$SE_{12} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where SE_{12} is the standard error of \bar{X}_{12} .

- Then, instead of the standard normal distribution, we need to use t -distributions to find p -values.
- For this, we can use R or R-Commander.

Two-Sample t-test

- Using the specific value of \bar{X}_{12} , which is denoted \bar{x}_{12} , as our point estimate for the difference between the two population means, $\mu_{12} = \mu_1 - \mu_2$, along with the standard error SE_{12} of \bar{X}_{12} , we find confidence intervals for μ_{12} as follows:

$$[\bar{x}_{12} - t_{crit} \times SE_{12}, \bar{x}_{12} + t_{crit} \times SE_{12}]$$

where t_{crit} is the t -critical value from a t -distribution for the desired confidence level c .

- When comparing the population means for two groups, the formula for finding the degrees of freedom is as follows:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1} \left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1} \left(\frac{s_2^2}{n_2}\right)^2}$$

Two-Sample t-test

- For testing a hypothesis regarding $\mu_{12} = \mu_1 - \mu_2$ when the population variances are unknown, we follow similar steps as above, but we use SE_{12} instead of SD_{12} and use the following t -statistic instead of the z -statistic to account for the additional source of uncertainty involved in estimating the population variances:

$$T = \frac{\bar{X}_{12}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

where $\bar{X}_{12} = \bar{X}_1 - \bar{X}_2$ as before.

Two-Sample t-test

- Using the observed data, we obtain $\bar{x}_{12} = \bar{x}_1 - \bar{x}_2$ as the observed value of \bar{X}_{12} . We also use the observed data to obtain s_1 and s_2 as the observed values of sample variances.
- Then, we calculate the observed value of the test statistic T as follows:

$$t = \frac{\bar{x}_{12}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{\bar{x}_{12}}{SE_{12}}$$

which is called the t -score.

- Depending on the alternative hypothesis, we calculate p_{obs} as
 - if $H_A : \mu_{12} > 0$, $p_{\text{obs}} = P(T \geq t)$,
 - if $H_A : \mu_{12} < 0$, $p_{\text{obs}} = P(T \leq t)$,
 - if $H_A : \mu_{12} \neq 0$, $p_{\text{obs}} = 2 \times P(T \geq |t|)$,

where T has a t -distribution with the degrees of freedom obtained as above

Example

- For the body temperature example, suppose that the sample variances based on our sample of $n_1 = 25$ women and $n_2 = 27$ men are $s_1^2 = 1.1$ and $s_2^2 = 1.2$, respectively.
- The standard error of \bar{X}_{12} is

$$SE_{12} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{1.1}{25} + \frac{1.2}{27}} = 0.3$$

- Degrees of freedom is

Example

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{1}{n_1 - 1} \left(\frac{s_1^2}{n_1} \right)^2 + \frac{1}{n_2 - 1} \left(\frac{s_2^2}{n_2} \right)^2} = \frac{\left(\frac{1.1}{25} + \frac{1.2}{27} \right)^2}{\frac{1}{25 - 1} \left(\frac{1.1}{25} \right)^2 + \frac{1}{27 - 1} \left(\frac{1.2}{27} \right)^2} = 49.9$$

- To find the corresponding t_{crit} , we follow similar steps as before.
- Suppose that we are interested in 95% confidence interval for $\mu_1 - \mu_2$.
- We find t_{crit} from the t –distribution with $df = 49.9$ degrees of freedom.
- In R-Commander,
 - click *Distributions* → *t distribution* → *t quantiles*.
 - Then enter $(1 - 0.95)/2 = 0.025$ for Probabilities, 49.9 for Degrees of freedom, and check the option Upper tail.
- The corresponding t-critical value is 2.01.

Example

- This results in the following 95% confidence interval:

$$[\bar{x}_{12} - t_{crit} \times SE_{12}, \bar{x}_{12} + t_{crit} \times SE_{12}]$$

$$[-0.2 - 2.01 \times 0.30, -0.2 + 2.01 \times 0.30] = [-0.80, 0.40].$$

- Therefore, at 0.95 confidence level, we believe that the true difference between the two means falls between -0.80 and 0.40 .
- The t –score is

$$t = \frac{\bar{x}_{12}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{\bar{x}_{12}}{SE_{12}} = \frac{-0.2}{0.3} = -0.67$$

Example

- The alternative hypothesis is $H_A : \mu_{12} \neq 0$.
- Using the t -distribution with $df = 49.9$ degrees of freedom, the upper tail probability of $|-0.67| = 0.67$ is $P(T > 0.67) = 0.25$.
- The observed significance level is $p_{\text{obs}} = 2 \times 0.25 = 0.50$, which is considered to be large (compared to commonly used significance levels).
- Therefore, the result is not statistically significant, and we cannot reject the null hypothesis, which indicates that the two populations (men and women) have the same mean body temperature.