BME 3005
Biostatistics

Lecture 5: *t-test, post hoc tests*

Asst. Prof. Burcu Tunç Çamlıbel

Introduction

- So far, we have learned
  - how to summarize the data
    - Mean, variance, std, median, percentiles
    - Standard error of the mean to estimate the precision with which a sample mean estimates the population mean
  - ANOVA
    - If the hypothesis that all the samples were drawn from a single population is true, then the within group or between group or the real population variances should be almost EQUAL & F should be close to 1.

$$F = \frac{\text{population variance estimated from sample means}}{\begin{array}{c}\text{population variance estimated as average} \\ \text{of sample variances}\end{array}}$$

$$F = \frac{s^2_{\text{bet}}}{s^2_{\text{wit}}}$$

Summary of Statistical Methods (table at the cover)

Summary of Some Statistical Methods to Test Hypotheses

Scale of measurement	Type of experiment: Two treatment groups consisting of different individuals	Three or more treatment groups consisting of different individuals	Before and after a single treatment in the same individuals	Multiple treatments in the same individuals	Association between two variables
Interval (and drawn from normally distributed populations*)	Unpaired *t* test (Chapter 4)	Analysis of variance (Chapter 3)	Paired *t* test (Chapter 9)	Repeated-measures analysis of variance (Chapter 9)	Linear regression, Pearson product-moment correlation, or Bland-Altman analysis (Chapter 8)
Nominal	Chi-square analysis-of-contingency table (Chapter 5)	Chi-square analysis-of-contingency table (Chapter 5)	McNemar's test (Chapter 9)	Cochrane Q†	Relative rank or odds ratio (Chapter 5)
Ordinal‡	Mann-Whitney rank-sum test (Chapter 10)	Kruskal-Wallis statistic (Chapter 10)	Wilcoxon signed-rank test (Chapter 10)	Friedman statistic (Chapter 10)	Spearman rank correlation (Chapter 8)
Survival time	Log-rank test or Gehan's test (Chapter 11)				

*If the assumption of normally distributed populations is not met, rank the observations and use the methods for data measured on an ordinal scale.

†Not covered in this text.

‡Or interval data that are not necessarily normally distributed.

t-test OR Student's t-test

- t-test should be used to compare ONLY 2 groups.
- When there are more than 2 groups, ANOVA should be used to test if all the data consisting different groups were drawn from the same "single" population.
- $F=t^2$
- t-test is the most common statistical test in the literature. More than %50 papers in medical literature use it.
- t-test is widely applied INCORRECTLY to compare multiple groups by doing all the pairwise comparisons.

Fig 4.1

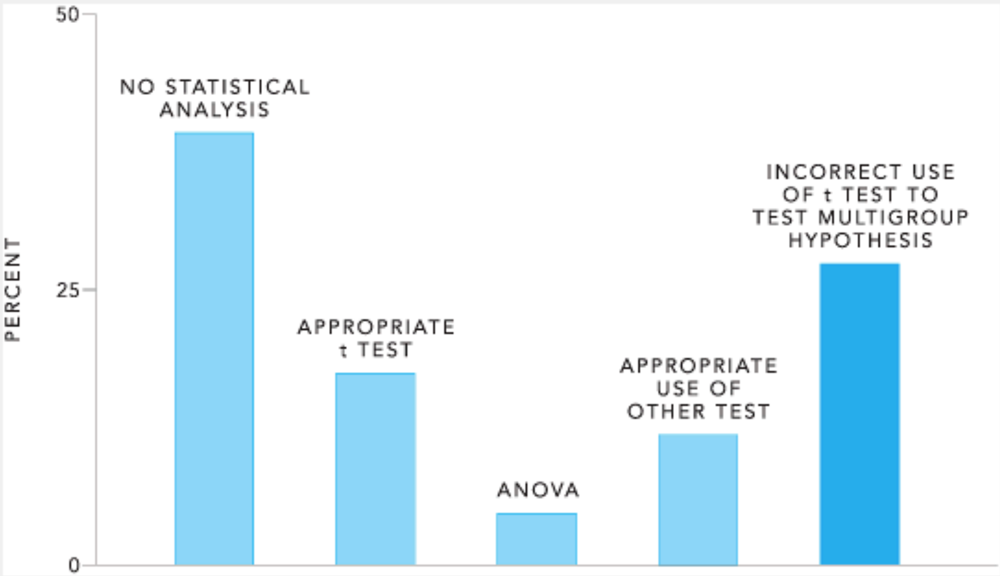


Figure 4–1 Of 142 original articles published in Vol. 56 of *Circulation* (excluding radiology, clinicopathologic, and case reports), 39 percent did not use statistics; 34 percent used a *t* test appropriately to compare two groups, analysis of variance (ANOVA), or other methods; and 27 percent used the *t* test incorrectly to compare more than two groups with each other. Twenty years later, misuse of the *t* test to compare more than two groups remained a common error in the biomedical literature. *(From S. A. Glantz, "How to Detect, Correct, and Prevent Errors in the Medical Literature,"* Circulation, **61:***1–7, 1980. By permission of the American Heart Association, Inc.)*

Out of 142 articles published in Circulation journal:

%54 used t-test

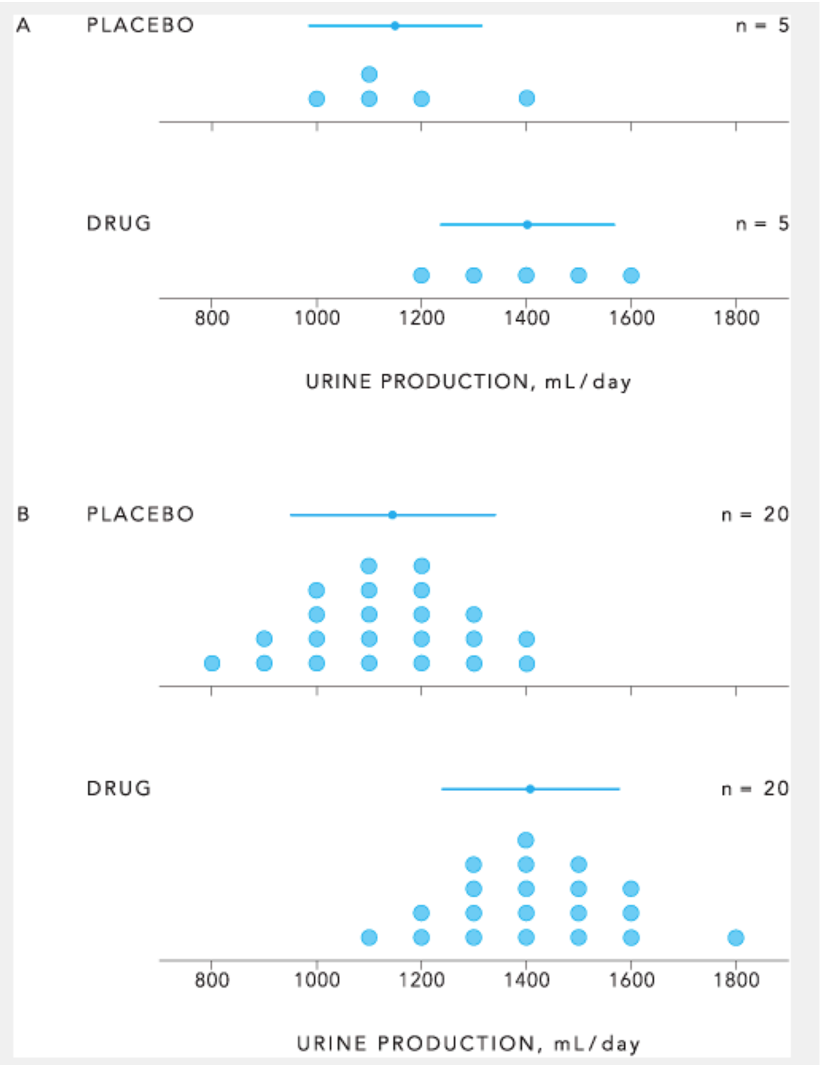
%27 used the t-test to compare more than 2 groups

t-test Problem in a Nutshell

- The incorrect use of t-test increases the chance of rejecting the NULL hypothesis.

- It increases the chance of reporting that a therapy had an effect when the evidence does not support it.

Figure 4.2



10 people
Randomly divided into 2 groups Urine production after 24h

Drug group has average 240 ml higher output

40 people
The means and std's are similar
to 4.2A
But now we are more confident to report that drug has an effect. WHY?

Sample Size

- As sample size increases observers become more confident in their estimate of the population mean.
- As sample size increases, SEM decreases as,

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

where n is the sample size and is the population std.

Sample Size (cont)

- As the sample size increases, the uncertainty in the estimate of the difference of the means between placebo and drug groups decreases relative to the difference of the means.

- We become more confident that the drug has an effect.

t-test

$$t = \frac{\text{difference in sample means}}{\text{standard error of difference of sample means}}$$

- We are comparing the relative magnitude of the differences in the sample means with the amount of variability that would be expected from looking within the samples.
- When this ratio is small, we will conclude that the data are compatible with the hypothesis that both samples were drawn from the same population.
- When it is large, we will conclude otherwise.

The standard deviation of a difference or a sum

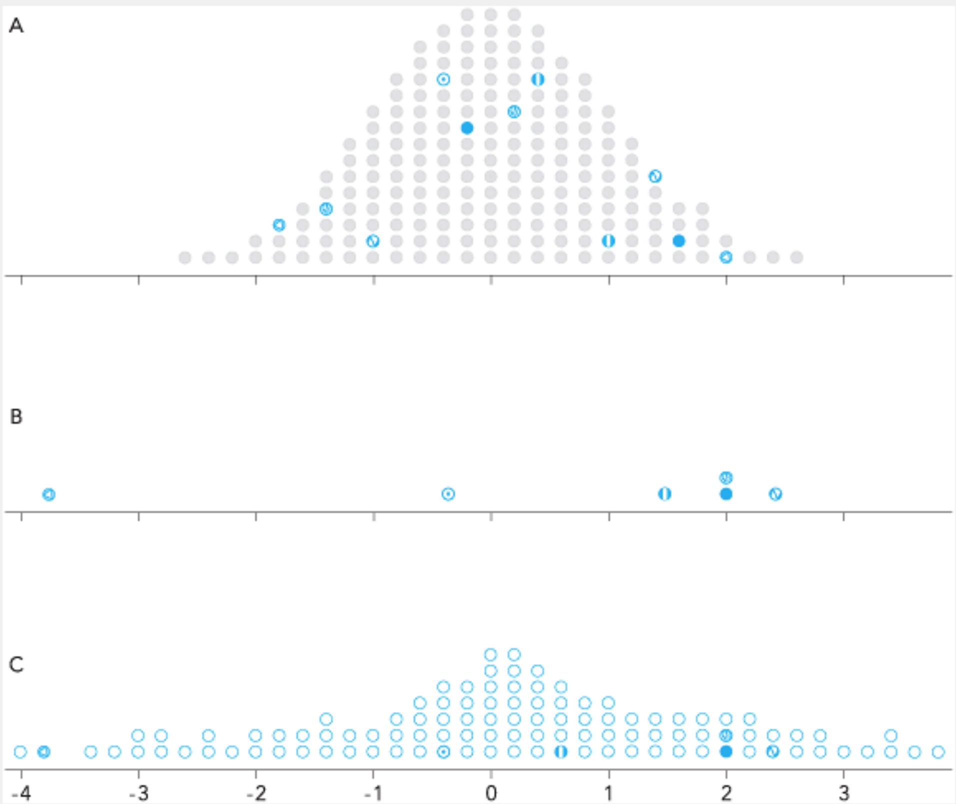


Figure 4–3 If one selects pairs of members of the population in panel ***A*** at random and computes the difference, the population of differences, shown in panel ***B***, has a wider variance than the original population. Panel ***C*** shows another 100 values for differences of pairs of members selected at random from the population in ***A*** to make this point again.

A population of 200 people
mean=0, std=1

We draw two samples at random and look at their difference

B- 5 such sample differences

C- the differences of 50 pairs of randomly selected numbers -more variability in differences than samples themselves

The standard deviation of a difference or a sum (cont)

- The standard deviation of the population of differences is about 40 percent larger than the standard deviation of the population from which the samples were drawn.

- Mathematically, the variance of the difference (or sum) of 2 variables equals to the sum of the variances of the 2 populations from which the samples were drawn.

$$\sigma^2_{X-Y} = \sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$$

SEM and population std relation

- Suppose we draw n numbers randomly from a single population with an std of σ

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + X_3 + \cdots + X_n)$$

- The mean of the values will be

- so $n\bar{X} = X_1 + X_2 + X_3 + \cdots + X_n$

- The variance of n$\bar{X}$ will be $\sigma^2_{n\bar{X}} = \sigma^2 + \sigma^2 + \sigma^2 + \cdots + \sigma^2 = n\sigma^2$

- and the standard deviation of means will be

$$\sigma_{n\bar{X}} = \sqrt{n}\sigma \quad \Rightarrow \quad \sigma_{\bar{X}} = \sqrt{n}\sigma/n = \sigma/\sqrt{n}$$

Standard deviation of difference (cont)

- In Fig 4.3 $$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y = 1 + 1 = 2$$
- Std is the square root of the variance, so it is 1.4. %40 larger than the original standard deviations
- For samples, $$s^2_{X-Y} = s^2_X + s^2_Y$$
- Standard error of the mean is the standard deviation of population of all sample means. Similarly, standard error of the difference of means is,

$$s_{\bar{X}-\bar{Y}} = \sqrt{s^2_{\bar{X}} + s^2_{\bar{Y}}}$$

Caution

- Don’t confuse
  - s_X --- standard deviation of a SAMPLE X
  - $s_{\overline{X}}$ --- standard error of the mean calculated from a sample X

$$s_{\overline{X}} = \frac{s}{\sqrt{n}}$$

Back to t-test

- Recall: $$t = \frac{\text{difference in sample means}}{\text{standard error of difference of sample means}}$$
- So, $$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s^2_{\bar{X}_1} + s^2_{\bar{X}_2}}}$$
- Or, we can use sample stds instead of standard errors of the mean, (recall $s_{\bar{X}} = \frac{s}{\sqrt{n}}$)

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{(s_1^2/n) + (s_2^2/n)}}$$

Pooled variance estimate

- If the two samples were drawn from the same population, then their variances will be equal to a population variance s.

- We replace the two estimates of the population variance with a single “pooled variance” estimate by averaging these 2 separate estimates.

$$s^2 = {}^1\!/_2\,(s_1^2 + s_2^2)$$

t-test

- t-test statistic based on the pooled variance estimate is

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{(s^2/n) + (s^2/n)}}$$

Unequal Sample Sizes

$$s_{\bar{X}_1}^2 = \frac{s_1^2}{n_1} \quad \text{and} \quad s_{\bar{X}_2}^2 = \frac{s_2^2}{n_2}$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{(s_1^2/n_1) + (s_2^2/n_2)}}$$

When the two sample sizes are unequal, the pooled variance is calculated as,

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

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{(s^2/n_1) + (s^2/n_2)}}$$

Significance

- There will be a range of t values based on the samples picked similar to ANOVA and F test.
- How big a “t” is BIG?
- We use two tailed t-test. %95 of the time the “t” value we estimated will be within these boundaries, and we have an error possibility of %5. --> p<0.05
- There is one tailed t-test that has lower t cutoff values, and leads to incorrect significances -- use two tailed

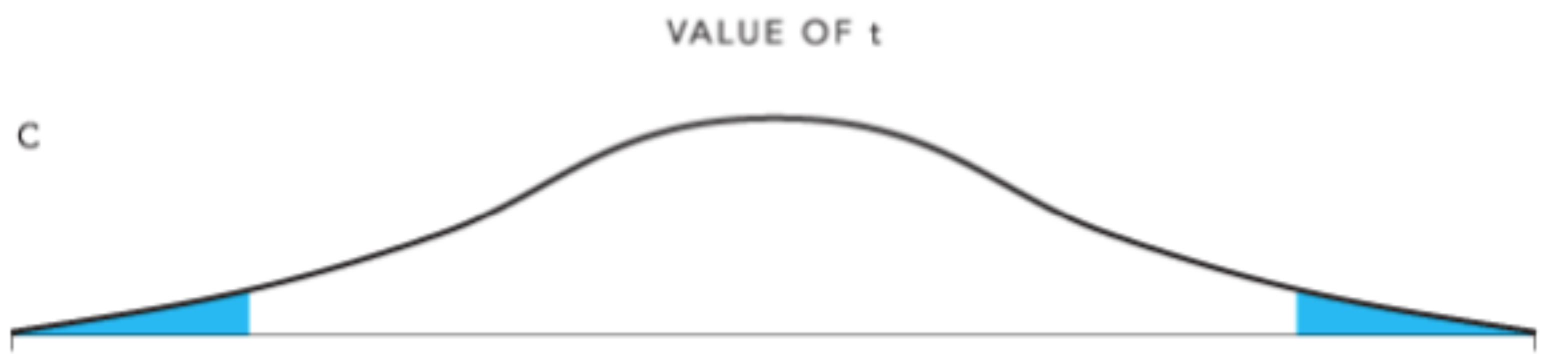
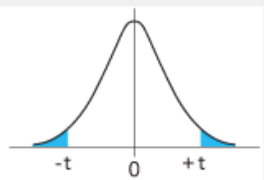


Table 4-1

Table 4–1 Critical Values of *t* (Two-Tailed)



	Probability of greater value, *P*								
ν	0.50	0.20	0.10	0.05	0.02	0.01	0.005	0.002	0.001
1	1.000	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	1.415	1.895	2.365	2.998	3.449	4.029	4.785	5.408
8	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850

ν is the degrees of freedom

$\nu = 2(n-1)$ when the sample sizes are equal

$\nu = n_1+n_2 - 2$ when the samples sizes are unequal

Small Homework

- The t-test is actually an analysis of variance.
- Read pg 84-85 and do the math to prove it to yourselves.
- $F=t^2$.
- Note: Remember, if there are more than two groups, you need to use F test or so called ANOVA and NOT the t-test.

Common t-test errors and compensation

- If we use the t-test statistic to test the differences between more than two groups by comparing all possible pairs of means, the actual p value will be much higher than $p<0.05$.
- If we repeat t-test 3 times, $p=3x0.05 = 0.15$! and so on…
- Multiple comparison tests is the solution if the t-test is used for more then two groups.

Figure 4.6

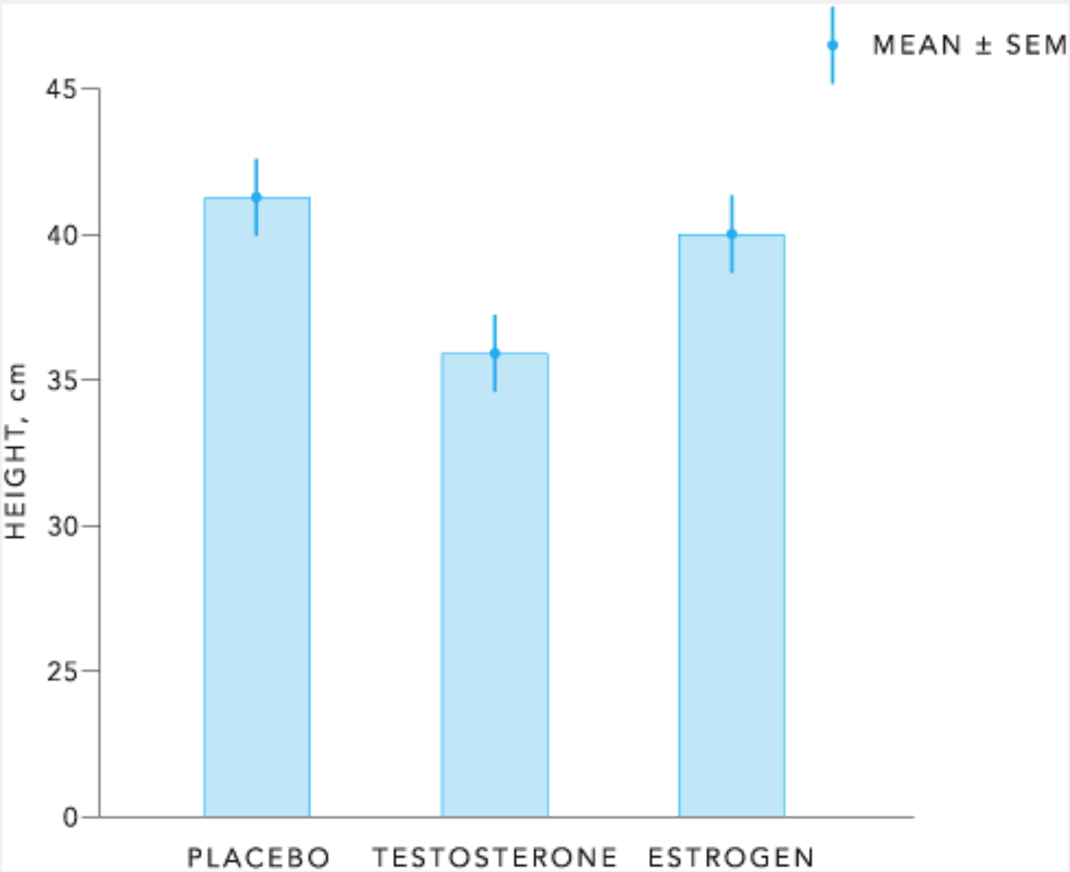


Figure 4–6 Results of a study of human hormones on Martians as it would be commonly presented in the medical literature. Each large bar has a height equal to the mean of the group; the small vertical bars indicate 1 standard error of the mean on either side of the mean (not 1 standard deviation)

- Pick 10 Martians -- placebo, testosterone or estrogen
- Measure their heights
- Three t-tests

placebo vs testosterone-2.39
placebo vs estrogen- 0.93
estrogen vs testosterone-1.34

- t cutoff for 2(10-1)=18 degrees of freedom is 2.101.
- Testosterone produces shorter Martians while estrogen does not and estrogen and testosterone do not differ in their height effects ????
- Some authors write a creative discussion for such problems ☺

Figure 4.6 (cont)

- ANOVA of the same data yields F=2.74 [cutoff 3.35]
  - the treatments did not have an effect.

Rules of t-test

- The t-test can be used to test the hypothesis that two group means are not different.
- When the experimental design involves multiple groups, analysis of variance should be used.
- When t-tests are used to test for differences between multiple groups, you can estimate the true P value by multiplying the reported P value with the number of possible t-tests.

Using t-test to Isolate Differences Between Groups in ANOVA

- ANOVA only tests for the global hypothesis that all the samples were drawn from a single population.
- It does not provide any information as to which sample or samples differed from the others.
- Multiple comparison tests are methods based on the t-test but include appropriate corrections for applying multiple pairs of t-tests.
- *Approach:* First apply an ANOVA to see whether anything is different, then use a multiple comparison test to see which treatment(s) produce the difference.

Bonferroni t-test

- Bonferoni inequality states that:
  "If k statistical tests are performed with the cutoff value for the test statistics at the α significance level, then the likelihood of observing a value of the test statistic exceeding the cutoff value at least once when the treatments did not have an effect is no greater than k times α."

$$\alpha_T < k\alpha \qquad \frac{\alpha_T}{k} < \alpha$$

Bonferroni t-test

- So, if we do each of the t-tests using the critical value of t corresponding to T/k, then the error rate for all the comparisons taken as a group will be at most .

- So if we do 5 t-tests, we need to use the cutoff value of 0.05/5=0.01, or in other words - be more strict, to not detect a false significance.

Bonferroni t-test

- This test works well for small number of t-tests, upto 8-10.
- When the number of t-tests is higher then the Bonferroni test gets too strict. Holm test might be more appropriate for such cases.

Bonferroni t-test

- One way of making Bonferroni less restrictive is to use the estimate of the population variance computed from within the groups in the analysis of variance

$$s_{\text{wit}}^2 = 1/4\,(s_{\text{con}}^2 + s_{\text{spa}}^2 + s_{\text{st}}^2 + s_{\text{f}}^2)$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{(s^2/n_1) + (s^2/n_2)}} \quad \longrightarrow \quad t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{(s_{\text{wit}}^2/n_1) + (s_{\text{wit}}^2/n_2)}}$$

- For equal sample sizes:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{2s_{\text{wit}}^2/n}}$$

Bonferroni t-test

- The degrees of freedom for this modified Bonferroni test will be same as the denominator degrees of freedom of F test.
- $\nu_d = m(n-1)$ (4 groups, size 7 each, 4(7-1)=24)
- So the number will be larger than 2(n-1) if there are more than 2 groups.
- Since the critical value of t decreases as the degrees of freedom increases, it will be possible to detect smaller differences.

Holm t-test

- As easy to compute as Bonferroni t-test, but is more powerful
- It is a sequentially rejective, or step-down, procedure
- It applies a reject/accept criterion to a set of ordered null hypothesis starting with the smallest p value and proceeds until it fails to reject a null hypothesis.

Holm t-test

To perform Holm t-test

– Compute pairwise comparisons with t-test using the pooled variance estimate like we did with less strict ANOVA and detect *unadjusted* P values.

– Compare these t values (or the p values) to critical values that have been adjusted to allow for the fact that we are doing multiple comparisons.

– In contrast to Bonferroni, we take into account how many tests we have already done and become less conservative with each subsequent comparison.

Holm t-test

- Let’s say we make k pairwise comparisons.
- Order these k uncorrected P values from smallest to largest so that P_1 is the smallest P value in the sequence and P_k is the largest.
- For the jth hypothesis, use $\alpha_j = \alpha_T/(k-j+1)$ and j=1…k
- So allow for a smaller cutoff value (or larger P value) for each consecutive step that allows for a more powerful test.

Student-Newman-Keuls (SNK) Test

- It uses q statistic.

- It is more powerful than Bonferroni test because it uses a more realistic estimate of the total true probability of errorneously concluding that a difference exists.

- First: Do an ANOVA. If F is significant, arrange all the means in descending order and compute SNK test statistic q.

BAU
BAHÇEŞEHİR ÜNİVERSİTESİ

SNK Test

- q is,

$$q = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{s^2_{\text{wit}}}{2}\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}$$

- where X_A and X_B are the means compared, $s^2{}_{wit}$ is the within variance estimated in ANOVA, n_a and n_b are the sample sizes.
- Compare q to Table 4-3.

Table 4-3

Table 4-3
Critical Values of q

	$\alpha_T = 0.05$								
ν_d	$p = 2$	3	4	5	6	7	8	9	10
1	17.97	26.98	32.82	37.08	40.41	43.12	45.40	47.36	49.07
2	6.085	8.331	9.798	10.88	11.74	12.44	13.03	13.54	13.99
3	4.501	5.910	6.825	7.502	8.037	8.478	8.853	9.177	9.462
4	3.927	5.040	5.757	6.287	6.707	7.053	7.347	7.602	7.826
5	3.635	4.602	5.218	5.673	6.033	6.330	6.582	6.802	6.995
6	3.461	4.339	4.896	5.305	5.628	5.895	6.122	6.319	6.493
7	3.344	4.165	4.681	5.060	5.359	5.606	5.815	5.998	6.158
8	3.261	4.041	4.529	4.886	5.167	5.399	5.597	5.767	5.918
9	3.199	3.949	4.415	4.756	5.024	5.244	5.432	5.595	5.739
10	3.151	3.877	4.327	4.654	4.912	5.124	5.305	5.461	5.599
11	3.113	3.820	4.256	4.574	4.823	5.028	5.202	5.353	5.487
12	3.082	3.773	4.199	4.508	4.751	4.950	5.119	5.265	5.395

- ν_d is the denominator degrees of freedom of ANOVA
  m(n-1)
- p is the number of means being tested

When comparing the largest and smallest of four means, p=4
When comparing the second smallest and smallest means p=2

SNK Test

- Procedure:
  - First compare the largest mean with the smallest, then the largest with the second smallest until the largest has been compared with the second largest. Then go on to the second largest.
  - If no significant difference exists between 2 means, conclude that no difference exists between any means enclosed by the two without testing for them.
  - So, if no significant difference between means 3 and 1, then don’t test for a difference between means 3 and 2 and means 2 and 1.
  - 3 [2] 1

Tukey test

- It is computed exactly like the SNK test, the only difference is the critical value used to test whether a given difference is significant.
- In Tukey test, the number of parameter p is set to m, the number of groups in the study, for all comparisons.
- Some believe Tukey test is overly conservative because it requires all the groups to be compared as though they are separated by the maximum number of steps while SNK allows for a comparison in reference to the exact number of steps that separate the two means.

Which Multiple Comparison Procedure to Use?

- Bonferroni is too conservative.
- SNK tends to over-detect significant differences because it controls the error rate among all comparisons spanning a fixed number of means rather than all pairwise comparisons.
- Tukey tends to under-detect significant differences. -- more conservative
- SNK is preferred over Tukey as a good compromise between sensitivity and caution.
- Holm test is less conservative than Tukey or Bonferroni while controlling the overall risk of a false positive.

Multiple Comparisons Against a Single Group

- Compare multiple treatment groups to a single control group
- Use Bonferroni, SNK or Tukey to do all the pairwise comparisons then only consider the ones that involve the control group.
- But this approach results in many more than necessary comparisons and more strict comparison for each pairwise comparison.

2nd Bonferroni t-test

- The t-test statistic is computed as before, and the number of comparisons is reduced to reflect the total number of comparisons made.

- Instead of control-testosterone, control- estrogen, testosterone-estrogen (3) just do control-testosterone, control-estrogen (2) and use p to represent the case. p=0.05/2 instead of p=0.05/3

.

2^{nd} Holm t-test

- Similarly, calculate the t-test statistic as before and use a cutoff value to represent the number of comparisons made.

$$\underline{\alpha_1 = \alpha_T/(k - j + 1) = .05/(2 - 1 + 1)}$$

$$\alpha_2 = \alpha_T/(k - j + 1) = .05/(2 - 2 + 1)$$

Dunnett's test

- Analog of SNK for multiple comparisons against a single control group

$$q' = \frac{\bar{X}_{\text{con}} - \bar{X}_A}{\sqrt{s^2_{\text{wit}}\left(\frac{1}{n_{\text{con}}} + \frac{1}{n_A}\right)}}$$

- First order the means, then do the comparisons from the largest to smallest difference.
- Table 4-4

Table 4-4

Table 4-4

Critical Values of q'

						$\alpha_T = 0.05$								
ν_d	$p = 2$	3	4	5	6	7	8	9	10	11	12	13	16	21
5	2.57	3.03	3.29	3.48	3.62	3.73	3.82	3.90	3.97	4.03	4.09	4.14	4.26	4.42
6	2.45	2.86	3.10	3.26	3.39	3.49	3.57	3.64	3.71	3.76	3.81	3.86	3.97	4.11
7	2.36	2.75	2.97	3.12	3.24	3.33	3.41	3.47	3.53	3.58	3.63	3.67	3.78	3.91
8	2.31	2.67	2.88	3.02	3.13	3.22	3.29	3.35	3.41	3.46	3.50	3.54	3.64	3.76
9	2.26	2.61	2.81	2.95	3.05	3.14	3.20	3.26	3.32	3.36	3.40	3.44	3.53	3.65

- p is equal to the number of means in the study.
- The degrees of freedom is the denominator degrees of freedom in ANOVA

p-value revisited

- The p-value is the probability of obtaining a value of the test statistic as large as or larger than the value computed from the data when in reality there is no difference between the different treatments.
- The p-value is the probability of being wrong when asserting that a true difference exists.
- If $p<=0.05$, in the long run we accept to be wrong 1/20 times.
- 2 ways an investigator can reach a wrong conclusion based on data:
  - Type I or α error: erroneously concluding that the treatment had an effect – p-value quantifies this
  - Type II or β error : erroneously concluding that the

treatment had no effect when in reality it did - Ch 6

Problem 4.1, Primer of Biostatistics, Glantz (6th Edition)

4-1 Conahan and associates also measured the mean arterial pressure and total peripheral resistance (a measure of how hard it is to produce a given flow through the arterial bed) in 9 patients who were anesthetized with halothane and 16 patients who were anesthetized with morphine. The results are summarized in Table 4–2. Is there evidence that these two anesthetic agents are associated with differences in either of these two variables?

	Halothane (n = 9)		Morphine (n = 16)	
	Mean	SD	Mean	SD
Best cardiac index, induction to bypass, L/m^2 Â· min	2.08	1.05	1.75	.88
Mean arterial blood pressure at time of best cardiac index, mmHg	76.8	13.8	91.4	19.6
Total peripheral resistance associated with best cardiac index, dyn Â· s/cm^5	2210	1200	2830	1130

Problem 4.2, Primer of Biostatistics, Glantz (6th Edition)

Mean Arterial Pressure (mmHg) after Receiving Cocaine	
Placebo	Nifedipine
156	73
171	81
133	103
102	88
129	130
150	106
120	106
110	111
112	122
130	108
105	99

4-2 Cocaine has many adverse effects on the heart, to the point that when people under 40 years of age appear in an emergency room with a heart attack, it is a good guess that it was precipitated by cocaine. In experiments, cocaine has been shown to constrict coronary arteries and reduce blood flow to the heart muscle as well as depress the overall mechanical function of the heart. A class of drugs know as calcium channel blockers has been used to treat problems associated with coronary artery vasoconstriction in other contexts, so Sharon Hale and colleagues ("Nifedipine Protects the Heart from the Acute Deleterious Effects of Cocaine if Administered Before but Not After Cocaine," *Circulation*, 83: 1437–1443, 1991) hypothesized that the calcium channel blocker nifedipine could prevent coronary artery vasoconstriction and the attendant reduction in blood flow to the heart and mechanical function. If true, nifedipine might be useful for treating people who had heart problems brought on by cocaine use. They measured mean arterial pressure in two groups of dogs after administering cocaine, one of whom was treated with nifedipine and the other of which received a placebo.

Does treatment with nifedipine after administering cocaine affect mean arterial pressure?

Problems 4.8 & 4.11, Primer of Biostatistics, Glantz (6th Edition)

4-8 Suppose that we were just interested in comparisons of the joggers and the marathon men with the inactive adults (as the control group). Use the data in Prob. 3–3 and make these comparisons with Holm *t* tests.

4-11 Repeat Prob. 4–10 using the SNK and Holm tests. Compare the results with those of Prob. 4–10 and explain any differences.

Problem 4.13, Primer of Biostatistics, Glantz (6th Edition)

4-13 In a test of significance, the *P* value of the test statistic is .063. Are the data statistically significant at

a. both the $\alpha = .05$ and $\alpha = .01$ levels?

b. the $\alpha = .05$ level but not at the $\alpha = .01$ level?

c. the $\alpha = .01$ level but not at the $\alpha = .05$ level?

d. neither the $\alpha = .05$ nor the $\alpha = .01$ levels?