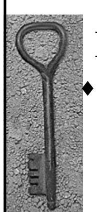
Hypothesis Test: Comparing Multiple Groups (ANOVA)

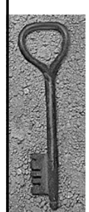
Review

- One-sample hypothesis test:
  - H_0: μ=constant, H_1: $\mu\neq$constant
  - H_0: π=constant, H_1: $\pi\neq$constant
- Two-sample hypothesis test:
  - H_0: $\mu_1=\mu_2$, H_1: $\mu_1\neq\mu_2$
  - H_0: $\pi_1=\pi_2$, H_1: $\pi_1\neq\pi_2$
  - Dependent samples H_0: $\mu_D=0$
- One-tailed vs. two-tailed tests



Issues

- What if we have more than two groups?
  - different ethnic groups;
  - difference classes in a school;
  - multiple years of data
- H_0: All groups are identical
  - E.g. $\mu_1 = \mu_2 = \mu_3 = \mu_4$
- H_1: One or more groups differ



Option #1

- Two-sample t-test for every combination of groups
  - $\mu_1 = \mu_2$, $\mu_1 = \mu_3$, $\mu_1 = \mu_4$, $\mu_2 = \mu_3$, $\mu_2 = \mu_4$, $\mu_3 = \mu_4$
- But, the possibility of a Type I error proliferates… 5% for each test.
  - With only 4 groups, 6 two-sample tests, chance of error reaches 6*5%=30%

Option #2: ANOVA

- ANOVA = "**AN**alysis **O**f **VA**riance"
  - "Oneway ANOVA" : The simplest form
- Only one test is needed, test whether all groups are the same ($\mu_1 = \mu_2 = \mu_3 = \mu_4$)
  - But, doesn't distinguish which specific group(s) differ
  - Maybe only μ_2 differs, or maybe all differ from others

ANOVA: Example

- Suppose you suspect that a firm is engaging in wage discrimination based on ethnicity
  - Certain ethnic groups might be getting paid more…
- The company counters: "We pay entry-level workers all **about** the same amount of money. No group gets preferential treatment."
- Given data on a sample of employees, ANOVA lets you test this hypothesis.
  - Are observed group differences just due to chance?
  - Or do they reflect differences in the underlying population? (i.e., the whole company)

ANOVA: Example

- The company has workers of three ethnic groups:
  - Whites, African-Americans, Asian-Americans
- Based on a sample of workers:
  - Y-bar$_{White}$ = $8.78 / hour
  - Y-bar$_{AfAm}$ = $8.52 / hour
  - Y-bar$_{AsianAm}$ = $8.91 / hour
- What can we conclude?
  - Nothing! Sample means differ randomly even if all groups had the same population mean ($\mu_{White} = \mu_{AfAm} = \mu_{AsianAm}$).
- Q: Are the observed differences so large it is unlikely that they are due to random error?
  - Thus, it is unlikely that: $\mu_{White} = \mu_{AfAm} = \mu_{AsianAm}$

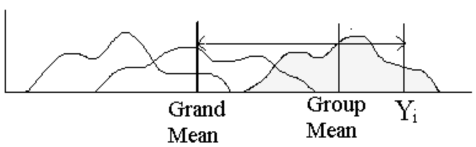
ANOVA: Concepts & Definitions

- Previously: μ, Y-bar; μ_1, μ_2, Y-bar$_1$, Y-bar$_2$
- The **grand mean** is the mean of all groups/cases
  - ex: mean of all entry-level workers = $8.70/hour
- The **group mean** is the mean of a particular sub-group of the population
- We hope to make inferences about **population** grand mean and group means, even though we only have **sample** grand mean and group means
  - We know Y-bar, Y-bar$_{White}$, Y-bar$_{AfAm}$, Y-bar$_{AsianAm}$
  - We want to infer about: μ, μ_{White}, μ_{AfAm} , $\mu_{AsianAm}$

ANOVA: Concepts & Definitions

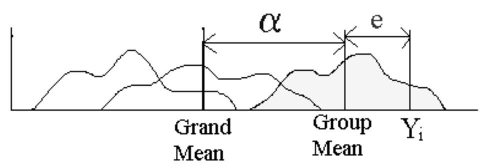
- Recall: variance, standard deviation are based on deviations, which is the distance of a point from the grand mean: $d_i = Y_i - \bar{Y}$

•ANOVA is based on partitioning deviation into different components



ANOVA: Concepts & Definitions

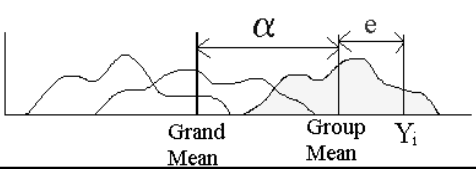
- The deviation of **any case** is determined by:
  - the distance between a group mean and the grand mean: the "group effect" (α), common to group members
  - the distance from group mean to a case's value: the within-group deviation (e): called "error",



ANOVA: Concepts & Definitions

- Initially we calculated deviation as the distance of a point from the grand mean: $d_i = Y_i - \bar{Y}$
- The total deviation can be partitioned into a_j (group effect) and e_{ij} (case errors, case i in group j)

$$d_{ij} = \alpha_j + e_{ij} = (\bar{Y}_j - \bar{Y}) + (Y_{ij} - \bar{Y}_j) = Y_{ij} - \bar{Y}$$

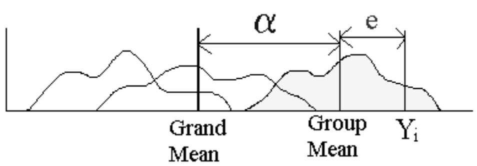


Sum of Squared Deviation

- The group effects: α_j $\alpha_j = \bar{Y}_j - \bar{Y}$
  - Deviation of the group from the grand mean
- Individual case error: e_{ij} $e_{ij} = Y_{ij} - \bar{Y}_j$
  - Deviation of the individual from the group mean
- Each are deviations that can be squared, and "summed up"–> sum of squared deviation
  - Recall: variance is sum of squared deviation

$$s_Y^2 = \frac{\sum_{i=1}^{n} d_i^2}{n-1} = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n-1}$$

Sum of Squared Deviation



- The total variance (SS_{total}) is made up of:
  - **between group** variance ($SS_{between}$)
  - **within group** variance (SS_{within})
  - $SS_{total} = SS_{between} + SS_{within}$

Sum of Squared Deviation

- Given a sample with j sub-groups:

Grand Mean $= \bar{Y}$

Group Means $= \bar{Y}_1, \bar{Y}_2, \bar{Y}_3, ..., \bar{Y}_j$

- "**Total Sum of Squares**" (SS_{total})

$$SS_{total} = \sum_{i=1}^{N} (Y_i - \bar{Y})^2 = \sum_{j=1}^{J} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y})^2$$

Sum of Squared Deviation

- The **between group** variance is the distance from the grand mean to each group mean (summed for all cases):

$$SS_{Between} = \sum_{j=1}^{J} n_j (\bar{Y}_j - \bar{Y})^2$$

- The **within group** variance is the distance from each case to its group mean (summed):

$$SS_{Within} = \sum_{j=1}^{J} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2$$

Sum of Squared Variance

- The sum of squares grows as n gets larger.
  - To derive a more comparable measure, we "average" it, just as with the variance: i.e, divided by n-1

$$s_Y^2 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n-1}$$

- For similar reasons, it is desirable to "average" the between/within Sum of Squares
- Result the "Mean Square" variance
  - $MS_{between}$ and MS_{within}

Sum of Squared Variance

- Divide Sum of Squares by degree of freedom:

$$MS_{Between} = \frac{\sum_{j=1}^{J} n_j (\bar{Y}_j - \bar{Y})^2}{J-1} \quad \text{(df=J-1)}$$

$$MS_{Within} = \frac{\sum_{j=1}^{J}\sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2}{N-J}$$

(df=$(n_1-1)+(n_2-1)+ \ldots + (n_j-1)$=N-J)

Mean Squares and Group Differences

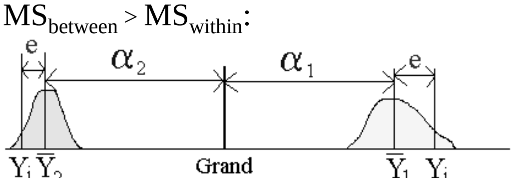
- Q: Which suggests that group means are quite different?

  $MS_{between} > MS_{within}$ or

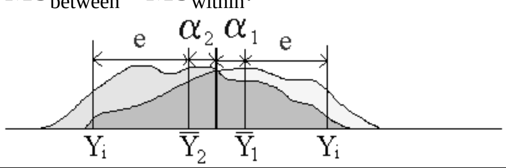
  $MS_{between} < MS_{within}$

Mean Squares and Group Differences

$MS_{between} > MS_{within}$:



$MS_{between} < MS_{within}$:



Mean Squares and Group Differences

- Q: Which suggests that group means are quite different:
  - $MS_{between} > MS_{within}$ or $MS_{between} < MS_{within}$
- Answer: If between group variance is greater than within, the groups are quite distinct
  - It is unlikely that they came from a population with the same mean
- If within is greater than between, the groups aren't very different – they overlap a lot
  - It is plausible that $\mu_1 = \mu_2 = \mu_3 = \mu_4$

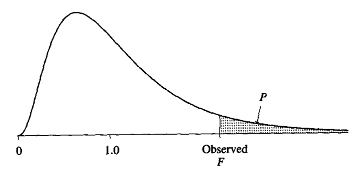
The F Ratio

$$F_{J-1,N-J} = \frac{MS_{Between}}{MS_{Within}}$$

- If $MS_{between} > MS_{within}$ then $F > 1$
- If $MS_{between} < MS_{within}$ then $F < 1$
- Larger F indicates that groups are more separate

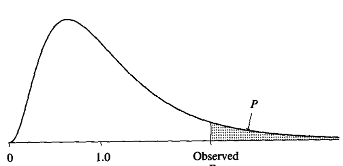
The F Ratio

- The F ratio has a sampling distribution (F-distribution)
- Again, this sampling distribution has known properties that can be looked up in a table
  - So, we can test hypotheses



F- Distribution

- Assume only positive values
- Skewed to the right
- Shape is determined by two degrees of freedom:
  - J-1, one for number of groups
  - N-J, one for total sample size N



The F-test

- Assumptions required for hypothesis testing using an F-statistic
  - Population distributions for groups are normal
  - Variance for population groups are equal
  - Independent random samples from population groups

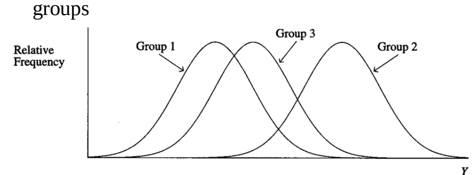
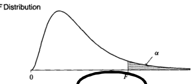


Figure 12.1 Assumptions About Population Distributions: Normal with Equal Standard Deviations

The F-test

- If these assumptions hold, the F statistic can be looked up in an F-distribution table

TABLE D *F* Distribution



$\alpha = .05$

df_2 \ df_1	1	2	3	4	5	6	8	12	24	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	243.9	249.0	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.38	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
∞	3.84	2.99	2.60	2.37	2.21	2.09	1.94	1.75	1.52	1.00

Source: From Table V of R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, published by Longman Group Ltd., London, 1974. (Previously published by Oliver & Boyd, Edinburgh.) Reprinted by permission of the authors and publishers.

One table for each significance level

Example

- Wage discrimination in a firm, 3 groups of workers
  - Whites, African-Americans, Asian-Americans
- You observe in a sample of 300 employees:
  - n1=100, Y-bar$_{White}$ = \$8.78 /hour, s_1=1.5
  - n2=100, Y-bar$_{AfAm}$ = \$8.52 /hour, s_2=1.2
  - n3=100,Y-bar$_{AsianAm}$ = \$8.91 /hour, s_3=0.9
  - Y-bar = \$8.74 /hour (grand mean)

Example

- 1. Assumption
  - Wage for each group is normally distributed
  - Assume same variance for each group
  - Independent random sample from each ethnic group
- 2. H_0: $\mu_{white} = \mu_{AfAm} = \mu_{AsianAm}$
  H_1: one or more group mean is different
- 3. Calculate F-statistic

$$F_{J-1,N-J} = \frac{MS_{Between}}{MS_{Within}}$$

Calculating Mean Sum of Squares

$$\text{MS}_{\text{Between}} = \frac{\sum_{j=1}^{J} n_j (\bar{Y}_j - \bar{Y})^2}{J - 1}$$

$$\text{MS}_{\text{Between}} = \frac{100(8.78-8.74)^2 + 100(8.52-8.74)^2 + 100(8.91-8.74)^2}{3-1}$$
$$= 3.95$$

Calculating Mean Sum of Squares

$$\text{MS}_{\text{Within}} = \frac{\sum_{j=1}^{J}\sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)}{N - J} = \frac{\sum_{j=1}^{J} (n_j - 1) S_j^2}{N - J}$$

S: Group standard deviation

(Because $s_Y^2 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n-1}$)

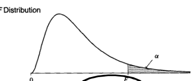
$$\text{MS}_{\text{Within}} = \frac{(100-1)1.5^2 + (100-1)1.2^2 + (100-1)0.9^2}{300-3}$$
$$= 1.5$$

Example

$$\text{F}_{J-1,N-J} = \frac{MS_{Between}}{MS_{Within}} = \frac{3.95}{1.5} = 2.63$$

- Recall that N = 300, J = 3
  - df_1=J-1 = 2, df_2=N-J = 297
- 5. If α = .05, the critical F value=?

TABLE D *F* Distribution



$\alpha = .05$

df_2 \ df_1	1	2	3	4	5	6	8	12	24	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	243.9	249.0	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.38	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
∞	3.84	2.99	2.60	2.37	2.21	2.09	1.94	1.75	1.52	1.00

Source: From Table V of R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, published by Longman Group Ltd., London, 1974. (Previously published by Oliver & Boyd, Edinburgh.) Reprinted by permission of the authors and publishers.

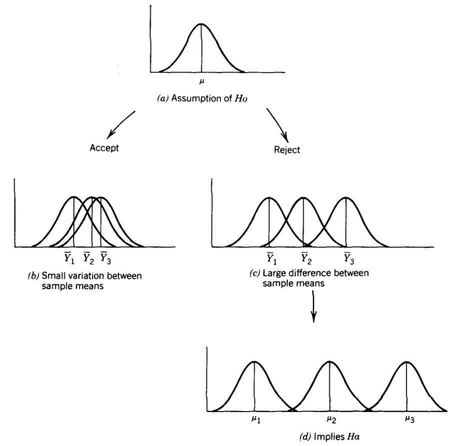
df_1=2

df_2=297

Example

$$F_{J-1,N-J} = \frac{MS_{Between}}{MS_{Within}} = \frac{3.95}{1.5} = 2.63$$

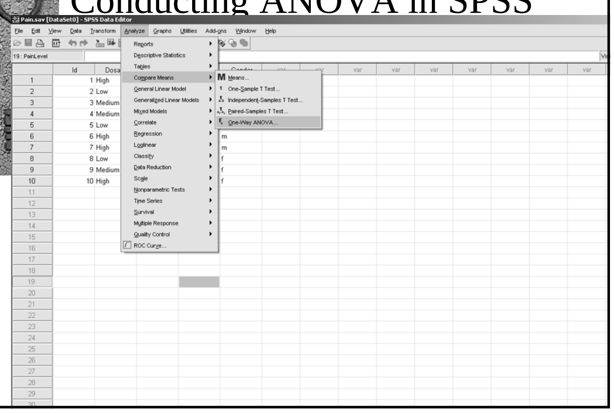
- Recall that N = 300, J = 3
  - df_1=J-1 = 2, df_2=N-J = 297
- 5. If α = .05, the critical F value for 2, 297 is about 3.00
- 6. Conclusion: the observed F < 3, so we fail to reject H0; we can conclude that the groups have the same population mean→ no racial discrimination in wage

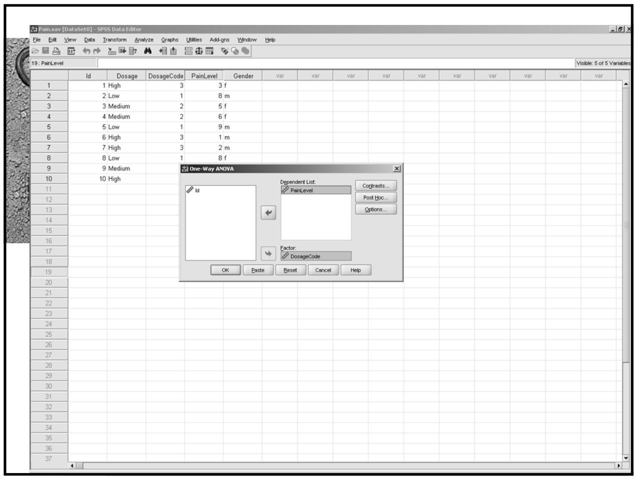


Summary

- Concepts:
  - Grand vs. group means
  - Between vs. within group sum of squared deviation (variance) ($SS_{between}$, SS_{within})
  - Between vs. within group mean squares ($MS_{between}$, Ms_{within})
- ANOVA (F-test)
  - Assumptions
  - H0: all group means are the same
  - H1: one or more group means are different
  - F statistic: $F=MS_{between} / Ms_{within}$
  - Critical value from F-distribution table, df_1=J-1, df_2=N-J
  - Conclusion: if F>C.V., reject H0; if F<C.V., fail to reject H0

Conducting ANOVA in SPSS





Output

Descriptives

PainLevel

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean: Lower Bound	95% Confidence Interval for Mean: Upper Bound	Minimum	Maximum
1	3	8.33	.577	.333	6.90	9.77	8	9
2	3	5.00	1.000	.577	2.52	7.48	4	6
3	4	2.25	.957	.479	.73	3.77	1	3
Total	10	4.90	2.767	.875	2.92	6.88	1	9

ANOVA

PainLevel

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	63.483	2	31.742	41.020	.000
Within Groups	5.417	7	.774		
Total	68.900	9			

GOG502/PLN504 Youqin Huang 38