## ANOVA F-test

1

2

3

4

5

/\* Drug Example on Page 4 of Note Outline 8 \*/

**data** drug\_example;

infile datalines dlm = '09'x;

input drug $ response;

datalines;

A 7.3

A 8.2

A 10.1

A 6

A 9.5

B 7.1

B 10.6

B 11.2

B 9

B 8.5

B 10.9

B 7.8

C 5.8

C 6.5

C 8.8

C 4.9

C 7.9

C 8.5

C 5.2

;

**run**;

**proc** **anova** data=drug\_example order=data;

class drug;

model response=drug;

means drug / tukey cldiff alpha=**.05**;

**run**;

1. INFILE is normally used to read in data tables, but following it with DATALINES tells SAS that data will be provided directly. DLM=’09’X specifies that the delimiter (i.e. what separates values each variable) is a tab; if the data does not run with this line, comment it out by putting an asterisk (\*) in front of INFILE.

2. PROC ANOVA is the procedure used to conduct an analysis of variance in SAS;

ORDER=DATA tells SAS to order groups in the results in the same order in which they appear in the data (e.g. A before B before C); the default ordering is alphabetical (which doesn’t make a difference for this example)

2. CLASS tells SAS that the variable “drug” identifies the groups;

3. MODEL SAS we want to see how “drug” impacts the variable “response”;

4. MEANS <group> / TUKEY CLDIFF ALPHA=<level> statement requests the Tukey multiple comparisons procedure to be performed at the specified alpha level (5% is the default level) and that the output shown be the confidence intervals as shown on page 6 of Note Outline 8)

\* For more on PROC ANOVA see <https://documentation.sas.com/doc/en/statug/15.2/statug_anova_syntax01.htm>

The ANOVA Procedure

Shows variable that defines the groups

| **Class Level Information** | | |
| --- | --- | --- |
| **Class** | **Levels** | **Values** |
| **drug** | 3 | A B C |

|  |  |
| --- | --- |
| **Number of Observations Read** | 19 |
| **Number of Observations Used** | 19 |

The ANOVA Procedure

Dependent Variable: response

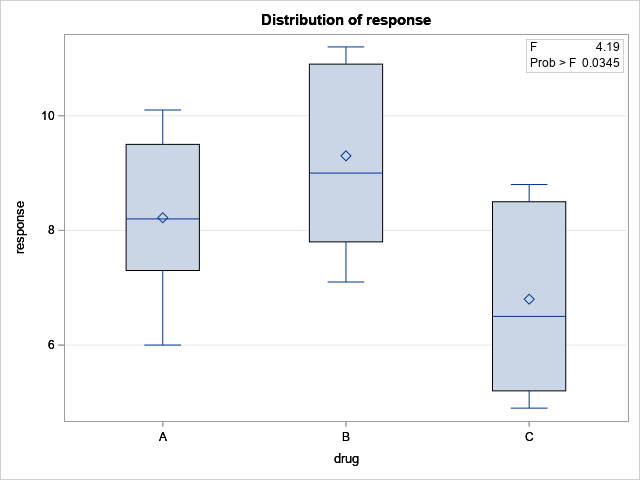
ANOVA Table (see pg 2 of Note Outline 8)

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Model** | 2 | 21.98147368 | 10.99073684 | 4.19 | 0.0345 |
| **Error** | 16 | 41.98800000 | 2.62425000 |  |  |
| **Corrected Total** | 18 | 63.96947368 |  |  |  |

Model summaries; will revisit these later

| **R-Square** | **Coeff Var** | **Root MSE** | **response Mean** |
| --- | --- | --- | --- |
| 0.343624 | 20.01243 | 1.619954 | 8.094737 |

| **Source** | **DF** | **Anova SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **drug** | 2 | 21.98147368 | 10.99073684 | 4.19 | 0.0345 |



Side-by-side boxplots for checking normality and equal variance assumptions

The ANOVA Procedure

Output on this page requested by the TUKEY option in the MEANS statement

Tukey's Studentized Range (HSD) Test for response

|  |  |
| --- | --- |
| Note: | This test controls the Type I experimentwise error rate. |

|  |  |
| --- | --- |
| **Alpha** | 0.05 |
| **Error Degrees of Freedom** | 16 |
| **Error Mean Square** | 2.62425 |
| **Critical Value of Studentized Range** | 3.64914 |

CI for:

\*

\*

\*

| **Comparisons significant at the 0.05 level are indicated by \*\*\*.** | | | | |
| --- | --- | --- | --- | --- |
| **drug Comparison** | **Difference Between**  [Sample] **Means** | **Simultaneous 95% Confidence Limits** | |  |
| **B - A** | 1.0800 | -1.3676 | 3.5276 |  |
| **B - C** | 2.5000 | 0.2657 | 4.7343 | \*\*\* |
| **A - B** | -1.0800 | -3.5276 | 1.3676 |  |
| **A - C** | 1.4200 | -1.0276 | 3.8676 |  |
| **C - B** | -2.5000 | -4.7343 | -0.2657 | \*\*\* |
| **C - A** | -1.4200 | -3.8676 | 1.0276 |  |

\* These CIs are redundant with others in the table

## Chi-square Tests

1

2

1

/\* Satisfaction with Appearance by Age Example on Page 9 \*/

**data** appearance;

input age $ satisfied $ count;

datalines;

1 Yes 38

2 Yes 30

3 Yes 34

1 No 10

2 No 29

3 No 9

;

**proc** **freq** data=appearance;

tables satisfied \* age / chisq expected nocol norow nocum nopercent;

weight count;

**run**;

/\* Vacation Destination Preference Example on Page 11 \*/

**data** vacation;

input place $ count;

datalines;

Hawaii 119

Europe 394

Caribbean 130

Other 285

;

**proc** **freq** data=vacation;

tables place / chisq testp=(**18** **40** **12** **30**);

weight count;

**run**;

/\* Chicken Pox Vaccine Side-effects Example on Page 12 \*/

**data** side\_effects;

input trt $ swelling $ count;

datalines;

vaccine major 54

vaccine minor 42

vaccine none 134

placebo major 16

placebo minor 32

placebo none 142

;

**proc** **freq** data=side\_effects;

tables trt \* swelling / chisq expected nocol norow nocum nopercent;

weight count;

**run**;

1. Chi-square tests are run via PROC FREQ (which was discussed in annotated code for 1-sample CI and HT); the CHISQ option for the TABLES command requests the Chi-square test (test statistic and p-value);

EXPECTED requests the expected counts in the table; other useful\* options are:

DEVIATION, which requests the difference (observed – expected) for each cell in the table, and

CELLCHI2, which requests the contribution of each cell to the test statistic: (observed – expected)^2 / expected;

\* DEVIATION and CELLCHI2 can be helpful for determining which groups differ from what is expected, if the Chi-square p-value is statistically significant

NOCOL … NOPERCENT suppress the column, row, cumulative frequencies and percentages (this was done for the sake of shortening the output)

2. TESTP= specifies the null proportions for the null distribution in a test of goodness of fit; notice that they are ordered to match the alphabetical order of the categories (e.g. Caribbean, Europe, Hawaii, Other)

The FREQ Procedure

**Output for Appearance Example (Test of Independence)**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | |  | | --- | | **Frequency** | | **Expected** | | | | **Table of satisfied by age** | | | | | | --- | --- | --- | --- | --- | | **satisfied** | **age** | | | | | **1** | **2** | **3** | **Total** | | **No** | |  | | --- | | 10 | | 15.36 | | |  | | --- | | 29 | | 18.88 | | |  | | --- | | 9 | | 13.76 | | |  | | --- | | 48 | |  | | | **Yes** | |  | | --- | | 38 | | 32.64 | | |  | | --- | | 30 | | 40.12 | | |  | | --- | | 34 | | 29.24 | | |  | | --- | | 102 | |  | | | **Total** | |  | | --- | | 48 | | |  | | --- | | 59 | | |  | | --- | | 43 | | |  | | --- | | 150 | |  * 1st number in each cell is the observed count (; “Frequency”); * 2nd number is the expected count (; “Expected”) * For this test, * Conduct follow-up multiple comparisons (if appropriate!) by comparing observed to expected counts in each cell |

Relevant test statistic and p-value

|  |
| --- |
| **Statistics for Table of satisfied by age** |

| **Statistic** | **DF** | **Value** | **Prob** |
| --- | --- | --- | --- |
| **Chi-Square** | 2 | 13.1493 | 0.0014 |
| **Likelihood Ratio Chi-Square** | 2 | 13.0387 | 0.0015 |
| **Mantel-Haenszel Chi-Square** | 1 | 0.0181 | 0.8930 |
| **Phi Coefficient** |  | 0.2961 |  |
| **Contingency Coefficient** |  | 0.2839 |  |
| **Cramer's V** |  | 0.2961 |  |

|  |
| --- |
| **Sample Size = 150** |

**Output for Vacation Destination (Test of Goodness of Fit)**

The FREQ Procedure

Expected counts not shown in this table, but would be calculated as

| **place** | **Frequency**  () | **Percent**  ( as a  percent) | **Test Percent**  ( as a  percent) | **Cumulative Frequency** | **Cumulative Percent** |
| --- | --- | --- | --- | --- | --- |
| **Caribbea** | 130 | 14.01 | 18.00 | 130 | 14.01 |
| **Europe** | 394 | 42.46 | 40.00 | 524 | 56.47 |
| **Hawaii** | 119 | 12.82 | 12.00 | 643 | 69.29 |
| **Other** | 285 | 30.71 | 30.00 | 928 | 100.00 |

Relevant test statistic and p-value

| **Chi-Square Test for Specified Proportions** | |
| --- | --- |
| **Chi-Square** | 10.2944 |
| **DF** | 3 |
| **Pr > ChiSq** | 0.0162 |



Deviation of observed count from expected; can be used for follow-up multiple comparisons of groups

|  |
| --- |
| **Sample Size = 928** |

**Output for Side Effects Example (Test of Homogeneity)**

The FREQ Procedure

* 1st number in each cell is the observed count (; “Frequency”);
* 2nd number is the expected count (; “Expected”)
* For this test,
* Conduct follow-up multiple comparisons (if appropriate!) by comparing observed to expected counts in each cell

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | |  | | --- | | **Frequency** | | **Expected** | | | | **Table of trt by swelling** | | | | | | --- | --- | --- | --- | --- | | **trt** | **swelling** | | | | | **major** | **minor** | **none** | **Total** | | **placebo** | |  | | --- | | 16 | | 31.667 | | |  | | --- | | 32 | | 33.476 | | |  | | --- | | 142 | | 124.86 | | |  | | --- | | 190 | |  | | | **vaccine** | |  | | --- | | 54 | | 38.333 | | |  | | --- | | 42 | | 40.524 | | |  | | --- | | 134 | | 151.14 | | |  | | --- | | 230 | |  | | | **Total** | |  | | --- | | 70 | | |  | | --- | | 74 | | |  | | --- | | 276 | | |  | | --- | | 420 | | |

Relevant test statistic and p-value

|  |
| --- |
| **Statistics for Table of trt by swelling** |

| **Statistic** | **DF** | **Value** | **Prob** |
| --- | --- | --- | --- |
| **Chi-Square** | 2 | 18.5707 | <.0001 |
| **Likelihood Ratio Chi-Square** | 2 | 19.5565 | <.0001 |
| **Mantel-Haenszel Chi-Square** | 1 | 17.6963 | <.0001 |
| **Phi Coefficient** |  | 0.2103 |  |
| **Contingency Coefficient** |  | 0.2058 |  |
| **Cramer's V** |  | 0.2103 |  |

|  |
| --- |
| **Sample Size = 420** |