1 Chi-Square Test for Goodness of Fit

1.1 Hypotheses:

H_0: sample was taken from a population with proportions p_1, p_2, …

H_A: sample was taken from a population with proportions <u>different</u> from p_1, p_2, …

1.2 Test Procedure:

1.2.1 Arrange data into frequencies observed for each category.

1.2.2 Calculate frequencies expected for sample size n if H_0 true.

1.2.3 Calculate chi-square statistic:

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - \hat{f}_i)^2}{\hat{f}_i}$$

f_i = frequency observed in category i

$\hat{f}_i$ = frequency expected in category i if H_0 true

1.2.4 Compare with critical value, $\chi^2_{\alpha,\nu}$ degrees of freedom: $\nu = k - 1$

1.2.5 Reject H_0 if $\chi^2 \geq \chi^2_{\alpha,\nu}$

1.3 Example: gender composition of ESCI 340: 26 students; 11 women, 15 men.
gender composition of WWU undergraduates: 54.5% women; 45.5% men
"expected" gender composition of ESCI 340: 14.2 women, 11.8 men.

$$\chi^2 = \frac{(11-14.2)^2}{14.2} + \frac{(15-11.8)^2}{11.8} = 0.709 + 0.849 = 1.559$$

$$\nu = k - 1 = 2-1 = 1$$

$$\chi^2_{0.05,1} = 3.841$$

Do not reject H_0 $(0.25 > P > 0.10)$

1.4 Chi-Square Correction for Continuity

$$\chi_c{}^2 = \sum_{i=1}^{2} \frac{\left(\left|f_i - \hat{f}_i\right| - 0.5\right)^2}{\hat{f}_i}$$

do not use when $k > 2$

Application to ESCI 340 gender composition:

$$\chi_c{}^2 = \frac{\left(\left|11-14.2\right| - 0.5\right)^2}{14.2} + \frac{\left(\left|15-11.8\right| - 0.5\right)^2}{11.8} = 0.503 + 0.603 = 1.106$$

Do not reject H_0 $(0.50 > P > 0.25)$

2 Komogorov-Smirnov Goodness of Fit Test for Discrete Data

1.1 Data sorted into categories.

1.2 Same hypotheses as above.

1.3 Expected frequencies $\left(\hat{f}_i\right)$ calculated as in Chi-squared goodness of fit test.

1.4 Calculate cumulative observed frequencies (F_i) and cumulative expected frequencies ($\hat{F}_i$):
cumulative frequency for i is sum of frequencies 1 through i.

2 Komogorov-Smirnov Goodness of Fit Test (continued)

1.5 For each category, *i*, determine absolute difference:

$$|d_i| = |F_i - \hat{F}_i|$$

1.6 Test statistic, d_{max} , is largest $|d_i|$.

Note: critical value depends on sample size (n) and number of categories (k).

1.7 Kolmorgorov-Smirnov test more powerful than chi-square test

when n small or $\hat{f}_i$ values small.

1.8 Example: Grizzly bear age distribution in Greater Yellowstone Ecosystem.

H_0: The age distribution of female Yellowstone grizzly bears prior to ESA listing was equal to the proportions of the stable age distribution: cubs of the year (0.198), yearlings (0.152), two year-olds (0.123), three year-olds (0.099), and adults (0.428).*

The Craighead brothers† reported age structure data on the 83 GYE female grizzly bears prior to ESA listing ("observed frequency," below).

Age	Cubs	1yr	1yr	3yr	Adults
Observed frequency (f_i)	14	8	8	6	47
Expected frequency ($\hat{f}_i$)	16.4	12.6	10.2	8.2	35.5
Cumulative obs.freq.(F_i)	14	22	30.	36	83
Cumulative exp.freq. ($\hat{F}_i$)	16.4	29.1	39.3	47.5	83
$\|d_i\| = \|F_i - \hat{F}_i\|$	2.4	7.1	9.3	11.5	0

$d_{max} = 11.5$

From Zar Table B.8, $(d_{max})_{0.05,5,83} = 11$ and $(d_{max})_{0.02,5,83} = 12$.
Conclude that age proportions of GYE grizzly bears differed from the stable age distribution prior to delisting ($0.02 < P < 0.05$). Younger bears were disproportionately less abundant and adult bears were disproportionately more abundant than the stable age distribution.

*Stable age distribution derived from: Pease CM and Mattson DJ. 1999. Demography of Yellowstone grizzly bears. *Ecology* 80(3):957-975.

†Craighead JJ, et al. 1995. *The Grizzly Bears of Yellowstone*. Island Press, Washington, D.C.

3 Chi-Square Analysis of Independence (Contingency Tables)

3.1 Hypotheses:

H_0: In sampled population, factors are independent. H_A: Factors are <u>not</u> independent.

3.2 Test Procedure:

→ analogous to chi-square test for goodness of fit.

Chi-square statistic:

$$\chi^2 = \sum\sum\frac{(f_{ij} - \hat{f}_{ij})^2}{\hat{f}_{ij}}$$

f_{ij} = frequency observed in row *i* & column *j*

$\hat{f}_{ij}$ = # expected in row *i*, column *j* if H_0 true

$$\hat{f}_{ij} = \frac{(R_i)(C_j)}{n}$$

DF = $(r - 1)(c - 1)$