Math 5090, Assignment 5, Chapter 12, Exercises 3, 9, 12, and a problem from Mood, Graybill and Boes.

3. (a)

$$\chi^2 = \frac{(10-50(0.25))^2}{50(0.25)} + \frac{(40-50(0.75))^2}{50(0.75)} + 2\frac{(50-100(0.5))^2}{100(0.5)} + \frac{(30-50(0.5))^2}{50(0.5)} + \frac{(20-50(0.5))^2}{50(0.5)}$$
$$= \frac{6.25}{12.5} + \frac{6.25}{37.5} + 2\cdot 0 + 2\frac{25}{25} = \frac{1}{2} + \frac{1}{6} + 2 = \frac{8}{3} \approx 2.67 \not> \chi^2_{0.90}(3) = 6.25.$$

We do not reject H_0.

(b)

$$\chi^2 = \frac{(10-50(0.5))^2}{50(0.5)} + \frac{(40-50(0.5))^2}{50(0.5)} + 2\frac{(50-100(0.5))^2}{100(0.5)} + \frac{(30-50(0.5))^2}{50(0.5)} + \frac{(20-50(0.5))^2}{50(0.5)}$$
$$= 2\frac{225}{25} + 2\cdot 0 + 2\frac{25}{25} = 2\cdot 9 + 2 = 20 > \chi^2_{0.90}(3) = 6.25.$$

We reject H_0.

(c) Under H_0, $\hat{p} = 90/200 = 0.45$, so

$$\chi^2 = \frac{(10-50(0.45))^2}{50(0.45)} + \frac{(40-50(0.55))^2}{50(0.55)} + \frac{(50-100(0.45))^2}{100(0.45)} + \frac{(50-100(0.55))^2}{100(0.55)}$$
$$+ \frac{(30-50(0.45))^2}{50(0.45)} + \frac{(20-50(0.55))^2}{50(0.55)}$$
$$= \frac{156.25}{22.5} + \frac{156.25}{27.5} + \frac{25}{45} + \frac{25}{55} + \frac{56.25}{22.5} + \frac{56.25}{27.5} = 18.18 > \chi^2_{0.90}(2) = 4.61.$$

We reject H_0.

9. Under H_0, we estimate the multinomial parameters, $\hat{p}_1 = (36+31+58)/(152+140+225) = 125/517 = 0.241779$, $\hat{p}_2 = (67+60+87)/517 = 214/517 = 0.413926$, and $\hat{p}_3 = 1 - \hat{p}_1 - \hat{p}_2 = 178/517 = 0.344294$.

$$\chi^2 = \frac{(36-152\hat{p}_1)^2}{152\hat{p}_1} + \frac{(67-152\hat{p}_2)^2}{152\hat{p}_2} + \frac{(49-152\hat{p}_3)^2}{152\hat{p}_3}$$
$$+ \frac{(31-140\hat{p}_1)^2}{140\hat{p}_1} + \frac{(60-140\hat{p}_2)^2}{140\hat{p}_2} + \frac{(49-140\hat{p}_3)^2}{140\hat{p}_3}$$
$$+ \frac{(58-225\hat{p}_1)^2}{225\hat{p}_1} + \frac{(87-225\hat{p}_2)^2}{225\hat{p}_2} + \frac{(80-225\hat{p}_3)^2}{225\hat{p}_3}$$
$$= 1.543 \not> \chi^2_{0.90}(4) = 7.78.$$

We do not reject H_0.

12. The estimated counts under H_0 are (row sum)(col sum)/750.

$$\begin{aligned}
\chi^2 = {} & \frac{(100-(200)(270)/750)^2}{(200)(270)/750} + \frac{(50-(200)(310)/750)^2}{(200)(310)/750} + \frac{(50-(200)(170)/750)^2}{(200)(170)/750} \\
& + \frac{(50-(320)(270)/750)^2}{(320)(270)/750} + \frac{(200-(320)(310)/750)^2}{(320)(310)/750} + \frac{(70-(320)(170)/750)^2}{(320)(170)/750} \\
& + \frac{(120-(230)(270)/750)^2}{(230)(270)/750} + \frac{(60-(230)(310)/750)^2}{(230)(310)/750} + \frac{(50-(230)(170)/750)^2}{(230)(170)/750} \\
= {} & 125.7 > \chi^2_{0.90}(4) = 7.78
\end{aligned}$$

We reject H_0, so income and stature are not independent.

50. According to a genetic model, the proportions of blood types should have the form O: q^2; A: p^2+2pq; B: r^2+2qr; AB: $2pr$, where $p+q+r=1$. Given the sample O: 374; A: 436; B: 132; AB: 58, how would you test the correctness of the model?

Use the chi-squared test, with expected cell frequencies given in terms of the maximum-likelihood estimators of p, q, r. We will find the MLEs $\hat{p}$ and $\hat{q}$ since there are only two free parameters ($\hat{r} = 1-\hat{p}-\hat{q}$). The likelihood function is

$$L(p,q) = C_1\,(q^2)^{374}(p^2+2pq)^{436}((1-p-q)^2+2q(1-p-q))^{132}(2p(1-p-q))^{58}$$

so the log-likelihood is

$$\begin{aligned}
\log L(p,q) &= C_2 + 748\log q + 436\log(p^2+2pq) \\
&\qquad + 132\log((1-p-q)^2+2q(1-p-q)) + 58\log(2p(1-p-q)) \\
&= C_2 + 748\log q + 436\log p + 436\log(p+2q) + 132\log(1-p-q) \\
&\qquad + 132\log(1-p+q) + 58\log 2 + 58\log p + 58\log(1-p-q) \\
&= C_3 + 748\log q + 494\log p + 436\log(p+2q) + 190\log(1-p-q) \\
&\qquad + 132\log(1-p+q)
\end{aligned}$$

So we want to maximize the function

$$f(p,q) = 748\log q + 494\log p + 436\log(p+2q) + 190\log(1-p-q) + 132\log(1-p+q)$$

over the region $p \geq 0$, $q \geq 0$, and $p+q \leq 1$. The partial derivatives

$$f_p(p,q) = \frac{494}{p} + \frac{436}{p+2q} - \frac{190}{1-p-q} - \frac{132}{1-p+q}$$

and

$$f_q(p,q) = \frac{748}{q} + \frac{872}{p+2q} - \frac{190}{1-p-q} + \frac{132}{1-p+q}$$

must both be 0. We could use an iteration method (e.g., Newton's method) or an equation solver such as NSolve in Mathematica. We find that

$$\hat{p} = 0.288632, \qquad \hat{q} = 0.611379, \qquad \hat{r} = 0.099989$$

though four decimal places is probably enough. Our chi-squared statistic is

$$\chi^2 = \frac{(374 - n\hat{q}^2)^2}{n\hat{q}^2} + \frac{(436 - n(\hat{p}^2 + 2\hat{p}\hat{q}))^2}{n(\hat{p}^2 + 2\hat{p}\hat{q})} + \frac{(132 - n(\hat{r}^2 + 2\hat{q}\hat{r}))^2}{n(\hat{r}^2 + 2\hat{q}\hat{r})} + \frac{(58 - n(2\hat{p}\hat{r}))^2}{n(2\hat{p}\hat{r})}$$
$$= 0.002121,$$

which is much smaller than $\chi^2_{0.95}(1) = 3.84$, so we cannot reject H_0. Notice that the expected cell frequencies are

$$n\hat{q}^2 = 373.785, \quad n(\hat{p}^2+2\hat{p}\hat{q}) = 436.235, \quad n(\hat{r}^2+2\hat{q}\hat{r}) = 132.260, \quad n(2\hat{p}\hat{r}) = 57.720$$

so the observed cell frequencies are the estimated cell frequencies rounded to the nearest integer. This suggests that the data are phony.

Here is the Mathematica code I used.

```
(*Plot3D[748Log[q]+494Log[p]+436Log[p+2q]+190Log[1-p-q]+132Log[1-p+q],
 {p,0.0001,0.9999},{q,0.0001,1-p},BoxRatios→{1,1,1},AxesLabel→Automatic]*)

NSolve[
 {494/p + 436/(p + 2 q) - 190/(1 - p - q) - 132/(1 - p + q) == 0, 748/q + 872/(p + 2 q) - 190/(1 - p - q) + 132/(1 - p + q) == 0}, {p, q}]
{{p → 1.56083, q → -0.650923}, {p → 0.722065, q → -0.19536},
 {p → 0.209806, q → -0.719762}, {p → 0.288632, q → 0.611379}}

p = 0.28863167504812326`;
q = 0.6113792562688001`;
494/p + 436/(p + 2 q) - 190/(1 - p - q) - 132/(1 - p + q)
748/q + 872/(p + 2 q) - 190/(1 - p - q) + 132/(1 - p + q)
6.13909 × 10^-12

7.27596 × 10^-12

n = 374 + 436 + 132 + 58
p = 0.2886316750
q = 0.6113792563
r = 1 - p - q
chisquared = (374 - n q^2)^2 / (n q^2) + (436 - n (p^2 + 2 p q))^2 / (n (p^2 + 2 p q)) +
  (132 - n (r^2 + 2 q r))^2 / (n (r^2 + 2 q r)) + (58 - n (2 p r))^2 / (n (2 p r))
1000

0.288632

0.611379

0.0999891

0.00212115
```