3.81. Since the last cell in the data set has only 4 observations, we combine the last two cells to find that there are 14 one second intervals in which 6 or more gamma rays occurred. The null hypothesis is that the distribution is Poisson with $\lambda = 2.4$. Thus the probability that there are $n$ gamma rays in a one second interval is $\frac{(2.4)^n}{n!}e^{-2.4}$. Hence, for $n = 0, 1, 2, \ldots, 5$, the expected number of observations is $e_n = 300 \cdot \frac{(2.4)^n}{n!}e^{-2.4}$. The expected number of observations in the last cell is $e_6 = 300(1 - \sum_0^5 \frac{(2.4)^n}{n!}e^{-2.4})$. The test statistic is $\chi^2 = \sum_{0}^{6} \frac{(f_i - e_i)^2}{e_i}$. No parameters are estimated in the null hypothesis, which just asserts $\lambda = 2.4$, so under the null hypothesis, the test statistic should have approximately a chi-square distribution with $7 - 1 = 6$ degrees of freedom, since there are 7 terms in the sum. Our test at the 0.05 level will be to reject $H_0$ if the test statistic exceeds $\chi^2_{0.05, 6} = 12.592$. We have

<table>
<thead>
<tr>
<th>No. of rays</th>
<th>Frequency</th>
<th>Expected frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>19</td>
<td>27.22</td>
</tr>
<tr>
<td>1</td>
<td>48</td>
<td>65.32</td>
</tr>
<tr>
<td>2</td>
<td>66</td>
<td>78.38</td>
</tr>
<tr>
<td>3</td>
<td>74</td>
<td>62.70</td>
</tr>
<tr>
<td>4</td>
<td>44</td>
<td>37.62</td>
</tr>
<tr>
<td>5</td>
<td>35</td>
<td>18.06</td>
</tr>
<tr>
<td>$\geq 6$</td>
<td>14</td>
<td>10.70</td>
</tr>
</tbody>
</table>

The value of the test statistic $\chi^2$ for this data is 29.1. As this is greater than 12.592, we reject $H_0$.

13.83. (a) We do not have enough data to compute a sample mean and a sample variance. To approximate them, let us assume that each observation lands on the mid-point of the interval it is recorded to fall in. Then

$$\bar{x} = \frac{7 + 10 \cdot 12 + 37 \cdot 17 + 36 \cdot 22 + 13 \cdot 27 + 2 \cdot 32 + 37}{100} = 20.$$ 

$$s^2 = \frac{1}{100} \left[ (7 - 20)^2 + 10 \cdot (12 - 20)^2 + 37 \cdot (17 - 20)^2 + \cdots + (37 - 20)^2 \right] = 25.$$ 

(b) Let $W$ be normal with mean $\mu = 20$ and variance $\sigma^2 = 25$. Let $Z = (W - 20)/5$, so that $Z$ is standard normal. Then (I have rounded to three decimal places),

$$P(W \leq 9.5) = P(Z \leq -2.1) = 0.018, \quad P(9.5 < W \leq 14.5) = P(-2.1 < Z \leq -1.1) = 0.118,$n
$$P(14.5 < W \leq 19.5) = P(-1.1 < Z \leq -0.1) = 0.325, \quad P(19.5 < W \leq 24.5) = P(-0.1 < Z \leq 0.9) = 0.356,$n
$$P(24.5 < W \leq 29.5) = P(0.9 < Z \leq 1.9) = 0.155, \quad P(29.5 < W \leq 34.5) = P(1.9 < Z \leq 2.9) = 0.027,$n
$$P(34.5 < W) = P(2.9 < Z) = 0.002)$$
(c) We combine the first two cells into one and the last three cells into one and compute expected
frequencies under the hypothesis of normality using $\hat{\mu} = \bar{x} = 20$ and $\hat{\sigma}^2 = 25$. With the cells labeled
in increasing order, $e_1 = 100 \cdot (0.018 + 0.0118) = 13.6$, $e_2 = 32.5$, $e_3 = 35.6$, $e_4 = 18.3$. (We adjusted
the roundoff of $e_4$ so that $e_1 + \cdots + e_4 = 100$.) The chi-square test statistic will have $4 - 1 - 2 = 1$
dergree of freedom, since two parameters $\mu$ and $\sigma^2$ are being estimated. We will reject if the test
statistic is greater than $\chi^2_{0.05,1} = 3.841$.

<table>
<thead>
<tr>
<th>No. of particles</th>
<th>Frequency</th>
<th>Expected frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 – 14</td>
<td>11</td>
<td>13.6</td>
</tr>
<tr>
<td>15 – 19</td>
<td>37</td>
<td>32.5</td>
</tr>
<tr>
<td>20 – 24</td>
<td>36</td>
<td>35.6</td>
</tr>
<tr>
<td>25 – 39</td>
<td>16</td>
<td>18.3</td>
</tr>
</tbody>
</table>

$$\chi^2 = \frac{(11 - 13.6)^2}{13.6} + \frac{(37 - 32.5)^2}{32.5} + \frac{(36 - 35.6)^2}{35.6} + \frac{(16 - 18.3)^2}{18.3} = 1.4$$

This is less than 3.841 and hence we accept $H_0$.

14.3. The joint density of $(X, Y)$ is $f(x, y) = 6x1_{0 < x < y < 1}$. The marginals are

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \begin{cases} \int_{x}^{1} 6x \, dy = 6x(1 - x), & \text{if } 0 < x < 1; \\ 0, & \text{otherwise.} \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx = \begin{cases} \int_{0}^{y} 6x \, dx = 3y^2, & \text{if } 0 < y < 1; \\ 0, & \text{otherwise.} \end{cases}$$

$\mu_{Y|x}$ is defined where $f_X(x) > 0$, namely on $0 < x < 1$, and there

$$\mu_{Y|x} = \int_{0}^{1} y \cdot \frac{f(x, y)}{f_X(x)} \, dy = \int_{0}^{1} y \cdot \frac{6x}{6x(1 - x)} \, dy = \frac{(1 - x^2)/2}{1 - x} = \frac{1 + x}{2}.$$ 

$\mu_{X|y}$ is defined where $f_Y(y) > 0$, namely on $0 < y < 1$, and there

$$\mu_{X|y} = \int_{0}^{1} x \cdot \frac{f(x, y)}{f_Y(y)} \, dx = \int_{0}^{y} \frac{6x}{3y^2} \, dx = \frac{2y}{3}.$$ 

14.7. (a) For the joint density $f(x, y) = 2_{0 < y < x < 1}$, the marginals are $f_X(x) = 2x1_{0 < x, 1}$ and
$f_Y(y) = 2(1 - y)1_{0 < y < 1}$. Thus, for $0 < x, y < 1$,

$$\mu_{Y|x} = \int_{0}^{x} \frac{2y}{2x} \, dx = \frac{x}{2}, \quad \mu_{X|y} = \int_{y}^{1} \frac{2x}{2(1 - y)} \, dy = \frac{1 + y}{2}.$$ 

(b) $E[X^m Y^n] = \int_{0}^{1} \int_{0}^{x} 2x^m y^n \, dy \, dx = \int_{0}^{1} \frac{2x^{m+n+1}}{n + 1} \, dx = \frac{2}{(n + 1)(n + m + 2)}.$
(c) Using the formula of (b), \( \mu_1 = E[X] = \frac{2}{3} \), \( \mu_2 = \frac{1}{3} \), \( \sigma_1^2 = E[X^2] - \mu_2^2 = \frac{1}{3} \), \( \sigma_2^2 = E[X^2] - \mu_2^2 = \frac{1}{2} - (\frac{2}{3})^2 = \frac{1}{18} \), \( \sigma_2^2 = 1/6 - 1/9 = 1/18 \), \( \text{Cov}(X,Y) = E[XY] = \mu_1 \mu_2 = 1/4 - 2/9 = 1/36 \), \( \rho = (1/36)/(\sqrt{1/18})^2 = 1/2 \).

Assuming that the regression functions are linear, Theorem 4.1 implies that

\[
\begin{align*}
\mu_Y|x &= \mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x - \mu_1) = \frac{1}{3} \cdot \frac{1}{2} (x - \frac{2}{3}) = \frac{x}{2} \\
\mu_X|y &= \mu_1 + \rho \frac{\sigma_1}{\sigma_2} (y - \mu_2) = \frac{2}{3} \cdot \frac{1}{2} (y - \frac{1}{3}) = \frac{1+y}{2}.
\end{align*}
\]

This agrees with the answer derived by direct calculation in part (a).

14.16. To fit \( y = \alpha + \beta x + \gamma x^2 \) by least squares to \( \{(x_1, y_1), \ldots, (x_n, y_n)\} \) choose \( \hat{\alpha}, \hat{\beta}, \hat{\gamma} \) to minimize

\[
D(\alpha, \beta, \gamma) = \sum_{i=1}^{n} (y_i - (\alpha + \beta x_i + \gamma x_i^2))^2.
\]

This is done by setting the first derivatives of \( D \) equal to zero:

\[
\begin{align*}
0 &= \frac{\partial D}{\partial \alpha}(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = 2 \cdot \sum_{i=1}^{n} [y_i - (\hat{\alpha} + \hat{\beta} x_i + \hat{\gamma} x_i^2)] \\
0 &= \frac{\partial D}{\partial \beta}(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = 2 \cdot \sum_{i=1}^{n} [y_i x_i - (\hat{\alpha} x_i + \hat{\beta} x_i^2 + \hat{\gamma} x_i^3)] \\
0 &= \frac{\partial D}{\partial \gamma}(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = 2 \cdot \sum_{i=1}^{n} [y_i x_i^2 - (\hat{\alpha} x_i^2 + \hat{\beta} x_i^3 + \hat{\gamma} x_i^4)]
\end{align*}
\]

These lead directly to the normal equations,

\[
\begin{align*}
n \hat{\alpha} + (n \bar{x}) \hat{\beta} + (\sum_{i=1}^{n} x_i^2) \hat{\gamma} &= n \bar{y} \\
n \bar{x} \hat{\alpha} + (\sum_{i=1}^{n} x_i^2) \hat{\beta} + (\sum_{i=1}^{n} x_i^3) \hat{\gamma} &= \sum_{i=1}^{n} x_i y_i \\
(\sum_{i=1}^{n} x_i^2) \hat{\alpha} + (\sum_{i=1}^{n} x_i^3) \hat{\beta} + (\sum_{i=1}^{n} x_i^4) \hat{\gamma} &= \sum_{i=1}^{n} x_i^2 y_i.
\end{align*}
\]

14.52. In the data of Exercise 14.44, the humidity measurements are the \( x_i \) variables and the moisture measurements are observations \( y_i \) of random variables \( Y_i, i = 1, \ldots, 12 \). The model is that for each \( i \), \( Y_i \) is a \( N(\alpha + \beta x_i, \sigma^2) \) random variable and \( Y_1, \ldots, Y_{12} \) are independent. For this data, we calculate, \( \bar{x} = 50/7/12 = 42.25 \), \( \bar{y} = 12 \), \( S_{xx} = 844.25 \), \( S_{yy} = 74 \), \( S_{xy} = 230 \). We obtain the least squares coefficients, \( \hat{\beta} = S_{xy}/S_{xx} = 0.272 \), and \( \hat{\alpha} = \bar{y} - \hat{\beta} \cdot \bar{x} = 0.490 \). Also the maximum likelihood estimate of the standard deviation is \( \hat{\sigma} = \sqrt{(S_{yy} - \hat{\beta} \cdot S_{xy})/12} = 0.972 \).
According to Theorem 14.4, under the null hypothesis that \( \beta = 0.350 \),
\[
t = \frac{\hat{\beta}(x, Y_1, \ldots, Y_{12}) - 0.350}{\hat{\sigma}} \sqrt{\frac{10/12}{S_{xx}}} = \frac{\hat{\beta} - \beta}{\hat{\sigma}} \cdot 26.52
\]
has the \( t \)-distribution with 10 degrees of freedom. If the alternative hypothesis is \( H_1 : \beta < 0.350 \), the appropriate rejection region for a level 0.05 test is \( \{ t \leq -t_{0.05, 10} \} = \{ t < -1.812 \} \). For the data of the problem \( t = -2.12 \). Therefore we reject \( H_0 \).

14.55. From the data in Exercise 14.45, \( S_{xx} = 70 \), \( \hat{\beta} = -0.0857 \) and \( \hat{\sigma} = 0.0655 \). From Theorem 14.4 or from Theorem 14.5,
\[
\hat{\beta} \pm \hat{\sigma} \sqrt{\frac{6}{4S_{xx}}} t_{0.01, 4}
\]
yields a 98% confidence interval for \( \beta \). Using \( t_{0.01, 4} = 3.747 \), and the values computed for \( \hat{\beta} \) and \( \hat{\sigma} \) this confidence interval is
\[
[-0.0857 - (0.0655) \sqrt{\frac{6}{280}} \cdot 3.747, -0.0857 + (0.0655) \sqrt{\frac{6}{280}} \cdot 3.747] = [-0.122, -0.050].
\]

6.45. (a) A bivariate normal pair \((X, Y)\), with means \( \mu_1, \mu_2 \), variances \( \sigma_1^2 \) and \( \sigma_2^2 \) and correlation \( \rho \) between \( X \) and \( Y \) has the density,
\[
f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} \exp \left\{ -\frac{1}{2(1 - \rho^2)} \left[ \left( \frac{x - \mu_1}{\sigma_1} \right)^2 - 2\rho \frac{x - \mu_1}{\sigma_1} \frac{y - \mu_2}{\sigma_2} + \left( \frac{y - \mu_2}{\sigma_2} \right)^2 \right] \right\}.
\]
Thus if the exponent of a bivariate normal is \(-(1/102) [ (x + 2)^2 - 2.8(x + 2)(y - 1) + 4(y - 1)^2] \), we must have \( \mu_1 = -2, \mu_2 = 1 \), and
\[
\frac{1}{102} = \frac{1}{2(1 - \rho^2)\sigma_1^2}, \quad \frac{2.8}{102} = \frac{2\rho}{2(1 - \rho^2)\sigma_1\sigma_2}, \quad \frac{4}{102} = \frac{1}{2(1 - \rho^2)\sigma_2^2}.
\]
By dividing the first fraction by the third and then the second by the third,
\[
\frac{\sigma_2^2}{\sigma_1^2} = \frac{1}{4}, \quad 2\rho \frac{\sigma_2}{\sigma_1} = 0.7.
\]
It follows that \( \rho = 0.7, \sigma_1^2 = 102/2(1 - (0.7)^2) = 100 \), and \( \sigma_2^2 = 25 \).

6.47. For a bivariate normal random variable, \( \mu_{Y|x} = \mu_2 + (\rho\sigma_2/\sigma_1)(x - \mu_1) \) and \( \sigma^2_{Y|x} = (1 - \rho^2)\sigma_2^2 \). So if \( \mu_1 = 2, \mu_2 = 5, \sigma_1 = 3, \sigma_2 = 6, \rho = 2/3 \),
\[
\mu_{Y|1} = 5 + (4/3)(1 - 2) = 11/3, \quad \sigma_{Y|1} = \sqrt{(1 - (4/9))36} = 2\sqrt{5}.
\]

14.31. First observe by simple algebra that if \( B > 0 \) and \( |\rho| < 1 \)
\[
\frac{1 + \rho}{1 - \rho} \leq B \quad \text{is equivalent to} \quad \rho \leq \frac{B - 1}{B + 1}.
\]
It follows that if $A > 0$ and $|\rho| < 1$

$$\frac{1 + \rho}{1 - \rho} \geq A$$

is equivalent to

$$\rho \geq \frac{A - 1}{A + 1}.$$ 

Let $z$ be defined as on page 456. Then, $z = \sqrt{n - 3} \ln \left[ \frac{1 + r}{1 - r} \frac{1 - \rho}{1 + \rho} \right]$, and by exponentiating $-z_{\alpha/2} \leq z \leq z_{\alpha/2}$ is equivalent to

$$\frac{1 + r}{1 - r} e^{-2z_{\alpha/2}/\sqrt{n-3}} \leq \frac{1 + \rho}{1 - \rho} \leq \frac{1 + r}{1 - r} e^{2z_{\alpha/2}/\sqrt{n-3}}.$$ 

From the first remarks, this is equivalent to

$$\frac{1 + r}{1 - r} e^{-2z_{\alpha/2}/\sqrt{n-3}} - 1 \leq \rho \leq \frac{1 + r}{1 - r} e^{2z_{\alpha/2}/\sqrt{n-3}} - 1.$$ 

By multiplying numerator and denominator in both expressions by $1 - r$, this can be expressed as

$$\frac{(1 + r)e^{-2z_{\alpha/2}/\sqrt{n-3}} - (1 - r)}{(1 + r)e^{-2z_{\alpha/2}/\sqrt{n-3}} + (1 + r)} \leq \rho \leq \frac{(1 + r)e^{2z_{\alpha/2}/\sqrt{n-3}} - (1 - r)}{(1 + r)e^{2z_{\alpha/2}/\sqrt{n-3}} + (1 + r)}.$$ 

(The book’s answer seems to have reversed the proper order of the inequalities.)

14.65. From the data given, we calculate,

$$\begin{align*}
\sum_{i=1}^{20} x_i &= 688, & \sum_{i=1}^{20} y_i &= 703, & \sum_{i=1}^{20} x_i^2 &= 24282, & \sum_{i=1}^{20} x_i y_i &= 24582, & \sum_{i=1}^{20} y_i^2 &= 25555.
\end{align*}$$

Using the formulas on page 442 and 443,

$$S_{xx} = 614.8, \quad S_{xy} = 398.8, \quad S_{yy} = 844.55.$$ 

The estimate of the correlation coefficient is therefore $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{398.8}{\sqrt{(614.8)(844.55)}} = 0.55$ (rounding to two decimal places.)

We are to test $H_0 : \rho = 0$ against $H_1 : \rho \neq 0$ at the 0.05 significance level. According to section 14.5, assuming the null hypothesis,

$$\frac{1}{2} \ln \frac{1 + R}{1 - R}$$

is approximately normal with mean $(1/2) \ln \frac{1 + 0}{1 - 0} = 0$ and variance $1/(n - 3) = 1/17$.

(Here $R$ is the statistic $S_{xy}/\sqrt{S_{xx}S_{yy}}$ treated as a random variable.) In other words,

$$z = \frac{1}{2} \ln \frac{1 + r}{1 - r} = \sqrt{\frac{17}{2}} \ln \frac{1 + r}{1 - r}$$

is approximately a standard normal random variable.
Therefore to test $H_0$ at significance level 0.05, we should reject $H_0$ if $|z| \geq z_{0.05/2} = 1.960$. Since, in fact, $z = (\sqrt{17}/2) \ln(1.55/0.45) = 2.550$, we reject $H_0$ at the 0.05 significance level.

14.66. From the fact that $z = \frac{\sqrt{n-3}}{2} \ln \frac{(1+r)(1-\rho)}{(1-r)(1+\rho)}$ is approximately a standard normal, it follows that the probability that $-z_{\alpha/2} < z < z_{\alpha/2}$ is approximately $1-\alpha$. We can solve this to get a $(1-\alpha)100\%$ confidence interval for $\rho$. The algebra is a bit complicated so we just give the solution, as found in the problem answers at the back of the text, except that that answer reverses the upper and lower endpoints. Here is the correct confidence interval:

$$\frac{(1+r)e^{2z_{\alpha/2}/\sqrt{n-3}} - (1-r)}{(1+r)e^{2z_{\alpha/2}/\sqrt{n-3}} + (1+r)} \leq \rho \leq \frac{(1+r)e^{2z_{\alpha/2}/\sqrt{n-3}} - (1-r)}{(1+r)e^{2z_{\alpha/2}/\sqrt{n-3}} + (1+r)}.$$

The answer to this problem is found simply by plugging in $r = 0.55$, $n = 20$, and $z_{0.025} = 1.96$. We find the 95% confidence interval: $0.14 < \rho < 0.80$. 

6