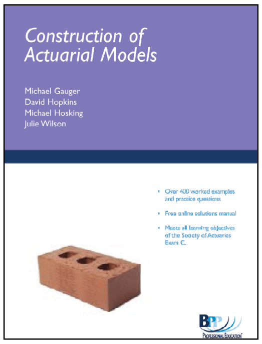



# Construction of Actuarial Models

**By J. Wilson, M. Hosking, D. Hopkins, and M. Gauger**

**Published by BPP Professional Education**

## *Solutions to practice questions – Chapter 2*

### *Solution 2.1*

The sample mean of the data is 0.31.
The sample size is 6.
So the $z$ statistic is:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{0.31 - 0}{1 / \sqrt{6}} = 0.759$$

---

### *Solution 2.2*

The null hypothesis is $H_0 : \sigma^2 = 29$. So $\sigma_0^2 = 29$.
The sample size is $n = 8$.
The sample variance is $s^2 = 30$.
So the $\chi^2$ statistic is:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{7 \times 30}{29} = 7.241$$

---

### *Solution 2.3*

Since the sample size is large, we can use the normal distribution even though we don't know the population variance.
The value of the $z$ statistic is:

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{583 - 550}{257 / \sqrt{500}} = 2.8712$$

The $p$ value for a two-sided test is:

$$2\Pr\left(Z > 2.8712\right) = 1\left[1 - \Phi(2.8712)\right] = 2\left(1 - 0.9979\right) = 0.0042$$

## *Solution 2.4*

The probability of making a Type I error is equal to $\alpha$, the significance level. (In Section 2.5 we defined $\alpha$ to be the probability of rejecting the null hypothesis when it is true.)
*The probability of making a Type I (Type II) error is sometimes called the size of the Type I (Type II) error.*

---

## *Solution 2.5*

The probability of a Type I error is given by:

$$\begin{aligned}\Pr(\bar{X} > 11 \mid \mu = 10) &= \Pr\left(N(0,1) > \frac{11-10}{\sqrt{25/100}}\right) \\ &= 1 - \Phi(2) \\ &= 1 - 0.9772 \\ &= 0.0228\end{aligned}$$

The probability of a Type II error is:

$$\begin{aligned}\Pr(\bar{X} \le 11 \mid \mu = 12) &= \Pr\left(N(0,1) \le \frac{11-12}{\sqrt{25/100}}\right) \\ &= \Phi(-2) \\ &= 1 - \Phi(2) \\ &= 1 - 0.9772 \\ &= 0.0228\end{aligned}$$

NB: Comparing this with Example 2.7, you will see that we have reduced the size of the Type II error, but this is at the expense of increasing the size of the Type I error.

---

## *Solution 2.6*

The power of the test is:

$$1 - \beta = 1 - 0.1587 = 0.8413$$

---

## *Solution 2.7*

Increasing the sample size decreases the probability of making either a Type I or a Type II error.

## *Solution 2.8*

The null and alternative hypotheses are:
$H_0$ : the $Bin(4,0.5)$ model is correct
$H_1$ : the $Bin(4,0.5)$ model is not correct
Under the suggested model the expected frequencies are:

| Number of boys | 0 | 1 | 2 | 3 | 4 | Total |
|---|---|---|---|---|---|---|
| Expected frequency | 9.375 | 37.5 | 56.25 | 37.5 | 9.375 | 150 |

These figures have been calculated using the formula:

$$150\Pr(N=n)=150\binom{4}{n}0.5^n\,0.5^{4-n}=150\binom{4}{n}0.5^4$$

The test statistic is:

$$\chi^2=\frac{(8-9.375)^2}{9.375}+\frac{(36-37.5)^2}{37.5}+\frac{(61-56.25)^2}{56.25}+\frac{(32-37.5)^2}{37.5}+\frac{(13-9.375)^2}{9.375}$$

$$=2.871$$

We compare this with $\chi_4^2$. The upper 5% point of $\chi_4^2$ is 9.49. As the test statistic is less than this critical value, we have insufficient evidence to reject the null hypothesis and we conclude that the binomial model is correct.

---

## *Solution 2.9*

We would have compared the value of the test statistic with $\chi_3^2$, instead of $\chi_4^2$.

---

## *Solution 2.10*

The following values are realizations of a random variable $Y$ :
100 150 225 290 300 500
You want to test whether these data come from a Pareto distribution with parameters $\alpha=3$ and $\theta=600$.
The hypothesized distribution function is:

$$F(y)=1-\left(\frac{600}{y+600}\right)^3$$

For the given data we have:

| $j$ | $y_{(j)}$ | $F\left(y_{(j)}\right)$ | $\frac{j}{6}-F\left(y_{(j)}\right)$ | $F\left(y_{(j)}\right)-\frac{j-1}{6}$ |
|---|---|---|---|---|
| 1 | 100 | 0.370 | –0.204 | 0.370 |
| 2 | 150 | 0.488 | –0.155 | 0.321 |
| 3 | 225 | 0.615 | –0.115 | 0.282 |
| 4 | 290 | 0.694 | –0.027 | 0.194 |
| 5 | 300 | 0.703 | 0.130 | 0.037 |
| 6 | 500 | 0.838 | 0.162 | 0.004 |

So the Kolmogorov-Smirnov test statistic is 0.370.

## *Solution 2.11*

We would expect the variance of the fitted distribution to be greater than the sample variance in this case.
We have fitted an exponential distribution that has mean 10 and variance 100.
If we calculate the variance of the sample data, we have:

$$s^2 = \frac{1}{n-1}\left(\sum X_i^2 - n\bar{x}^2\right) = \frac{1}{8}\left(624.1 - 9(7.4)^2\right) = 16.408$$

So this confirms what we suspected. The fitted distribution has a variance that is much greater than the sample variance, in other words we have fitted a distribution that is "too spread out" compared with the empirical data. This is what is suggested by a *p-p* plot that lies above the diagonal line to start with, then below.

---

## *Solution 2.12*

Type (3) – This suggests that there is not enough probability in either of the tails of the fitted distribution, *ie* that the fitted distribution is not spread out enough. It has too small a variance.

Type (4) – This suggests that there is too much probability in both of the tails of the distribution, and not enough in the center. This was the result we obtained in the previous example. The fitted distribution has too large a variance.

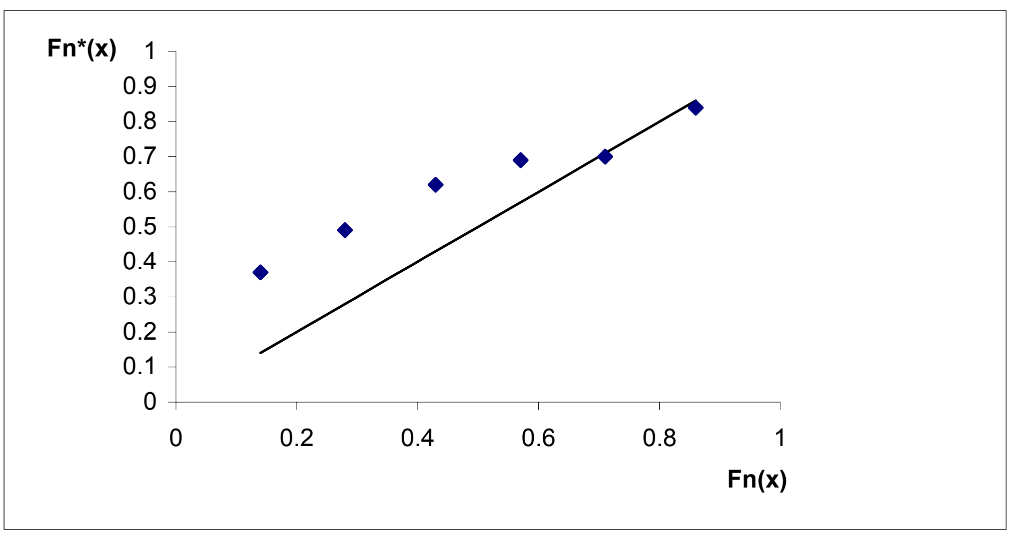
---

## *Solution 2.13*

The sample is size 6, so the empirical distribution function has values 1/7, 2/7, ..., 6/7.

The values of the model distribution function are (from the previous solution) .37, .49, .62, .69, .70, .84.

So the *p-p* plot looks like this:




We see that the *p-p* plot suggests that again the fitted distribution is too spread out, and has too large a sample variance. We could again compare the sample variance with the fitted distribution variance if we wanted to confirm this.

## *Solution 2.14*

It will have a saw-tooth pattern. The value of $F(x)$ jumps up as we pass a sample value. So $D(x)$ will increase sharply as we move past each sample value. Our fitted model however will be (usually) a smooth curve, and so the value of $D(x)$ will decrease gently as $x$ increases (until we meet the next sample value). We hope that the values of $D(x)$ will be small if the fit is good. Positive values of $D(x)$ imply that our model underestimates the empirical probabilities and vice versa.

---

## *Solution 2.15*

The geometric distribution is a special case of the negative binomial distribution with the restriction that $r = 1$. As we are given the maximized log-likelihoods, the likelihood ratio test statistic is:

$$-2(-578.9 + 573.6) = 10.6$$

---

## *Solution 2.16*

Using the suggested percentiles, we have the following results:

| Range | Observed | Expected |
|---|---|---|
| 0 - 1,000 | 3 | 4 |
| 1,000 - 3,000 | 7 | 8 |
| 3,000 - 5,000 | 10 | 6 |
| 5,000 - 8,000 | 9 | 10 |
| 8,000 - 11, 000 | 5 | 8 |
| 11,000 + | 6 | 4 |

Combining the first two groups and the last two groups to make the expected values at least 5, we obtain the following test statistic:

$$\chi^2 = \frac{(10-12)^2}{12} + \frac{(10-6)^2}{6} + \frac{(9-10)^2}{10} + \frac{(11-12)^2}{12} = 3.183$$

## *Solution 2.17*

We first have to calculate the cdf. If $0 < x \le ½$, then:

$$F(x) = \int_0^x 4t\,dt = \left[2t^2\right]_0^x = 2x^2$$

If $½ < x < 1$, then:

$$F(x) = ½ + \int_{½}^{x} (4-4t)\,dt = ½ + \left[4t - 2t^2\right]_{½}^{x} = ½ + \left(4x - 2x^2\right) - (2 - ½) = 4x - 2x^2 - 1$$

So we have:

| $j$ | $x_{(j)}$ | $F(x_{(j)})$ | $\frac{j}{4} - F(x_{(j)})$ | $F(x_{(j)}) - \frac{j-1}{4}$ |
|---|---|---|---|---|
| 1 | 0.16 | 0.0512 | 0.1988 | 0.0512 |
| 2 | 0.39 | 0.3042 | 0.1958 | 0.0542 |
| 3 | 0.62 | 0.7112 | 0.0388 | 0.2112 |
| 4 | 0.75 | 0.8750 | 0.1250 | 0.1250 |

The maximum of the numbers in the last two columns is 0.2112, and this is the Kolmogorov-Smirnov test statistic.

---

## *Solution 2.18*

The ranking criterion is $\ln L - \frac{r}{2}\ln n$, where $r$ is the number of parameters in the model and $n$ is the number of data points. The figures are as follows:

Inverse Burr model: $\ln L - \frac{r}{2}\ln n = -180.2 - \frac{3}{2}\ln 200 = -188.1$

Inverse Gamma model: $\ln L - \frac{r}{2}\ln n = -181.4 - \frac{2}{2}\ln 200 = -186.7$

Weibull model: $\ln L - \frac{r}{2}\ln n = -181.6 - \frac{2}{2}\ln 200 = -186.9$

Exponential model: $\ln L - \frac{r}{2}\ln n = -183.0 - \frac{1}{2}\ln 200 = -185.6$

The largest of these numbers is –185.6, so the best model according to the Schwarz-Bayesian criterion is the Exponential.

---

## *Solution 2.19*

From Example 1.5, we have:

$$\hat{\mu} = 5.35307$$

and:

$$\text{var}(\tilde{\mu}) \approx 0.01981$$

So the test statistic is:

$$\frac{\hat{\mu} - 5}{\sqrt{\text{var}(\tilde{\mu})}} \approx 2.51$$

This exceeds the upper 5% point of the standard normal distribution, so we reject the null hypothesis at the 5% significance level.

---

## *Solution 2.20*

The $p$ value of the test is:

$$\Pr\left(N(0,1) > 2.51\right) = 1 - \Phi(2.51) = 1 - 0.99396 = 0.00604$$

*Since this is less than 0.01, the null hypothesis would also have been rejected at the 1% significance level.*