Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

LØSNINGSFORSLAG
EKSAMEN I ST1201/ST6201 STATISTISKE METODER
Lørdag 8. juni 2013
Tid: 09:00–13:00

Oppgave 1

77 goals were scored on the 2012 European Football Championship. The table shows the number of games where 0,1,2 etc. goals were scored.

The number of goals in a game	The number of games
0	2
1	6
2	10
3	6
4	3
5	3
6	1
7+	0

a) Test the hypothesis that the number of goals, scored in a game, has the Poisson distribution with parameter $\lambda = 1$. The significance level is 0.05.

b) Test the hypothesis that the number of goals, scored in a game, has a Poisson distribution (with unknown parameter λ). The significance level is 0.05.

Solution.

a) Poisson probabilities ($\lambda = 1$) are

$$p_k = \frac{e^{-1}}{k!},\ k = 0, 1, ...,$$

i.e.

$p_0 = 0.37$

$p_1 = 0.37$

$p_2 = 0.18$

$p_3 = 0.06$

$p_4 = 0.015$

$p_{5+} = 0.005$

To apply the Pearson goodness-of-fit test, we use the partition (of nonnegative integers)

$$\{0\},\ \{1\},\ \{2, 3, ...\}$$

with probabilities (under H_0)

$$p_{10} = 0.37,\ p_{20} = 0.37,\ p_{30} = 0.26.$$

This partition is chosen because the condition $np_{i0} \geq 5$ must be satisfied. Then

$$k_1 = 2,\ k_2 = 6,\ k_3 = 23;$$

$$np_{10} = 11.47,\ np_{20} = 11.47,\ np_{30} = 8.06.$$

The observed value of the test statistic is

$$d = \sum_{i=1}^{3} \frac{(k_i - np_{i0})^2}{np_{i0}} = 38.1 > 5.991 = \chi^2_{0.95,2}.$$

The hypothesis is rejected.

b) Maximum likelihood estimate of λ is $\hat{\lambda} = 2.48... \approx 2.5$ and estimates of Poisson probabilities are

$$\hat{p}_k = \frac{e^{-2.5} 2.5^k}{k!},\ k = 0, 1, ...,$$

i.e.

$\hat{p}_0 = 0.08$

$\hat{p}_1 = 0.21$

$\hat{p}_2 = 0.26$

$\hat{p}_3 = 0.21$

$\hat{p}_4 = 0.13$

$\hat{p}_5 = 0.07$

$\hat{p}_{6+} = 0.04$

To apply the Pearson goodness-of-fit test, we use the partition (of nonnegative integers)

$$\{0,1\},\ \{2\},\ \{3\},\ \{4,5,...\}$$

with probabilities (under H_0)

$$\hat{p}_{10} = 0.29,\ \hat{p}_{20} = 0.26,\ \hat{p}_{30} = 0.21,\ \hat{p}_{40} = 0.24.$$

This partition is chosen because the condition $n\hat{p}_{i0} \geq 5$ must be satisfied. Then

$$k_1 = 8,\ k_2 = 10,\ k_3 = 6,\ k_4 = 7;$$

$$n\hat{p}_{10} = 8.99,\ n\hat{p}_{20} = 8.06,\ n\hat{p}_{30} = 6.51,\ n\hat{p}_{40} = 7.44.$$

The observed value of the test statistic is

$$d_1 = \sum_{i=1}^{4} \frac{(k_i - n\hat{p}_{i0})^2}{n\hat{p}_{i0}} = 0.65 < 5.991 = \chi^2_{0.95,2}.$$

The hypothesis is accepted.

Oppgave 2

Let $X_1, ..., X_n$ be a random sample (independent identically distributed random variables) from an exponential distribution with unknown parameter λ, i.e. from the distribution with the probability density

$$f(x) = \lambda e^{-\lambda x},\ x \geq 0,\ \lambda > 0.$$

The null hypothesis $H_0 : \lambda = 1$ is tested versus the alternative $H_1 : \lambda > 1$. There are two tests. Both are based on the test statistic

$$T = \min\{X_1, ..., X_2\},$$

but the first test rejects H_0 for large values of T while the second one rejects H_0 for small values of T. In other words

Test 1: if $T \geq c_1$, then H_0 is rejected.

Test 2: if $T \leq c_2$, then H_0 is rejected.

The significance level is α.

a) Find c_1 and c_2.

b) Find the power function of each test.

c) Which test is better? Why?

d) Even the best of these two tests is bad. Explain why.

Solution.

a)

$$\alpha = P_{\lambda=1}(\min X_i \geq c_1) = P_{\lambda=1}(X_1 \geq c_1, ..., X_n \geq c_1) = (1-(1-e^{-c_1}))^n = e^{-nc_1},$$

therefore

$$c_1 = \frac{1}{n}\ln\frac{1}{\alpha}.$$

$$\alpha = P_{\lambda=1}(\min X_i \leq c_2) = 1 - P_{\lambda=1}(\min X_i > c_2) =$$
$$= 1 - P_{\lambda=1}(X_1 > c_2, ..., X_n > c_2) = 1-(1-(1-e^{-c_2}))^n = 1-e^{-nc_2},$$

therefore

$$c_2 = \frac{1}{n}\ln\frac{1}{1-\alpha}.$$

b) The power function of the first test

$$\pi_1(\lambda) = 1-\beta_1(\lambda) = P_\lambda\left(\min X_i \geq \frac{1}{n}\ln\frac{1}{\alpha}\right) = e^{\lambda\ln\alpha} = \alpha^\lambda.$$

The power function of the second test

$$\pi_2(\lambda) = 1-\beta_2(\lambda) = P_\lambda\left(\min X_i \leq \frac{1}{n}\ln\frac{1}{1-\alpha}\right) =$$
$$= 1-e^{\lambda\ln(1-\alpha)} = 1-(1-\alpha)^\lambda.$$

c)

$$\frac{\partial}{\partial\lambda}\pi_1(\lambda) = e^{\lambda \ln \alpha} \ln \alpha < 0$$

i.e. $\pi_1(\lambda)$ decreases for $\lambda \geq 1$.

$$\frac{\partial}{\partial\lambda}\pi_2(\lambda) = -e^{\lambda \ln(1-\alpha)} \ln \frac{1}{1-\alpha} > 0$$

i.e. $\pi_2(\lambda)$ increases for $\lambda \geq 1$. But

$$\pi_1(1) = \pi_2(1) = \alpha,$$

therefore

$$\pi_1(\lambda) < \pi_2(\lambda)$$

for all $\lambda > 1$. This means that probability of type 2 error of the second test is less than probability of type 2 error of the first test for all λ from the alternative. The second test is better than the first one.

d) The power of both tests does not depend on n i.e. the probability of type 2 error does not decrease when the sample size increases.

Oppgave 3

Let Y be measured percentage of fat in a certain type of sausages. A laboratory has measured the fat percentage in 15 sausages and results $y_1, y_2, ..., y_{15}$ are supposed to be realisations of independent continuous random variables with a symmetric (about the unknown expectation) distribution. The results are

$$19.2, 27.6, 25.6, 32.2, 17.7, 20.5, 23.9, 20.2, 24.2, 26.1, 32.0, 24.8, 28.9, 16.2, 18.7.$$

Fat percentage 20.0 is considered as normal. We wish to test the hypothesis that the fat percentage is normal (i.e. equals 20.0) versus the alternative that it is greater than 20.0. The significance level is 0.05.

a) Test the hypothesis using the large-sample sign test.

b) Test the hypothesis using the large-sample Wilcoxon signed rank test.

c) Suppose that in addition to the conditions above it is known that the distribution of the fat percentage is approximately normal. Test the hypothesis using the $t-$test (for simple calculations you can use that $\sum_{i=1}^{15} y_i = 357.8$ and $\sum_{i=1}^{15} y_i^2 = 8888.62$).

Solution.

a) Let X be the number of observations greater than 20.0. The large-sample sign test: H_0 is rejected if

$$Z = \frac{X - n/2}{\sqrt{n/4}} \geq z_\alpha.$$

In our case ($X = 11$, $n = 15$)

$$Z = 1.8 > 1.645 = z_\alpha$$

therefore H_0 is rejected.

b) The large-sample Wilcoxon signed rank test: H_0 is rejected if

$$z = \frac{w - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \geq z_\alpha,$$

where $w = \sum_{i=1}^{n} r_i z_i$, r_i – rank of $|y_i - 20|$, $z_i = 1$ if $y_i > 20$ and 0 otherwise.

In our case

i	1	2	3	4	5	6	7	8
$\lvert y_i - 20\rvert$	0.8	7.6	5.6	12.2	2.3	0.5	3.9	0.2
r_i	3	12	10	15	5	2	7	1
z_i	0	1	1	1	0	1	1	1
i	9	10	11	12	13	14	15	
$\lvert y_i - 20\rvert$	4.2	6.1	12.0	4.8	8.9	3.8	1.3	
r_i	8	11	14	9	13	6	4	
z_i	1	1	1	1	1	0	0	

$$w = 102$$

and

$$z = 2.39 > 1.645 = z_\alpha.$$

Thus H_0 is rejected.

c) Let

$$T = \sqrt{n}\frac{\bar{y} - \mu_0}{s}$$

where

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2.$$

The t-test for testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$: H_0 is rejected if $T \geq t_{\alpha,n-1}$. In our case $\mu_0 = 20$, $\alpha = 0.05$, $n = 15$, $t_{\alpha,n-1} = 1.761$, $T = 2.97$.

H_0 is rejected.

Oppgave 4

The following table is an ANOVA table in which some entries are lost (stars).

Source	df	SS	MS	F
Treatment	3	$\star$	$\star$	1.6
Error	$\star$	$\star$	5.1	
Total	43	$\star$		

a) Find lost values and fill in the ANOVA table. Show how you calculate values where there are $\star$ in the table.

b) Test the hypothesis that

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4.$$

The significance level $\alpha = 0.05$.

Solution.

a)

$$MSTR = F \cdot MSE = 1.6 \cdot 5.1 = 8.16$$
$$SSTR = MSTR \cdot df = 8.16 \cdot 3 = 24.48$$
$$SSE = MSE \cdot df = 5.1 \cdot 40 = 204$$
$$SSTOT = SSTR + SSE = 204 + 24.48 = 228.48$$

Thus the filled ANOVA table is

Source	df	SS	MS	F
Treatment	3	24.48	8.16	1.6
Error	40	204	5.1	
Total	43	228.48		

b) $F_{0.95,3,40} = 2.84 > 1.6$ (observed F). Therefore H_0 is not rejected.