Exam in Statistical Methods, 2016-10-17 - Solutions

Time allowed: kl: 8-12

Allowed aids: Calculator. One handwritten A4 paper (both sides) with the students own notes.

Assisting teacher: Hector Rodriguez-Deniz

Grades: A=19-20 points, B=17-18p, C=14-16p, D=12-13p, E=10-11p, F=0-9p

Provide a detailed report that shows motivation of the results.

1

Let $f(y) = \begin{cases} cy^2 + y, & 0 \leq y \leq 1 \\ 0, & elsewhere \end{cases}$ be the density function of the random variable Y.

a) Find the value of c that makes $f(y)$ a density function. 1p

b) Find the cumulative distribution function $F(y)$. 1p

c) Calculate the probability $P\left(\frac{1}{2} \leq Y \leq 1\right)$. 1p

a)

$$1 = \int_0^1 cy^2 + y\,dy$$

$$1 = \left[\frac{cy^3}{3} + \frac{y^2}{2}\right]_0^1$$

$$1 = \frac{c}{3} + \frac{1}{2}$$

$$\frac{1}{2} = \frac{c}{3}$$

$$c = \frac{3}{2}$$

b)

$$F(y) = \int_0^y \frac{3}{2}y^2 + y\,dy$$

$$F(y) = \left[\frac{3}{2}\frac{y^3}{3} + \frac{y^2}{2}\right]_0^y$$

$$F(y) = \frac{y^3 + y^2}{2}$$

c)

$$P\left(\frac{1}{2} \leq Y \leq 1\right) = F(1) - F(\frac{1}{2})$$

$$P\left(\frac{1}{2} \le Y \le 1\right) = \frac{1^3 + 1^2}{2} - \frac{\left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^2}{2}$$

$$P\left(\frac{1}{2} \le Y \le 1\right) = 1 - \frac{\frac{1}{8} + \frac{1}{4}}{2}$$

$$P\left(\frac{1}{2} \le Y \le 1\right) = 1 - \frac{3}{16}$$

$$P\left(\frac{1}{2} \le Y \le 1\right) = \frac{13}{16} = 0.8125$$

2

Let the bivariate random variable (Y_1, Y_2) have the joint density function:

$$f(y_1, y_2) = \begin{cases} 6(1 - y_2), & 0 \le y_1 \le y_2 \le 1 \\ 0, & elsewhere \end{cases}$$

a) Find the marginal distributions $f(y_1)$ and $f(y_2)$. Are Y_1 and Y_2 independent? 2p

b) Find $E[y_1]$ 1p

c) Find $P(Y_1 | Y_2 = y_2)$. Can you identify the result as a known distribution? 1p

a)

Marginal for y1:

$$f(y_1) = 6\int_{y_1}^{1} 1 - y_2 \, dy_2$$

$$f(y_1) = 6\left[y_2 - \frac{{y_2}^2}{2}\right]_{y_1}^{1}$$

$$f(y_1) = 6\left(1 - \frac{1^2}{2} - y_1 + \frac{{y_1}^2}{2}\right)$$

$$f(y_1) = 3 - 6y_1 + 3{y_1}^2$$

Marginal for y2:

$$f(y_2) = 6\int_{0}^{y_2} 1 - y_2 \, dy_1$$

$$f(y_2) = 6[y_1 - y_2 y_1]_0^{y_2}$$

$$f(y_2) = 6y_2 - 6y_2^2$$

Now for independence we have to check if $f(y_1)f(y_2) = f(y_1, y_2)$

$$f(y_1)f(y_2) = (3 - 6y_1 + 3y_1^2)(6y_2 - 6y_2^2)$$

$$f(y_1)f(y_2) = 18y_2 - 18y_2^2 - 36y_1y_2 + 36y_1y_2^2 + 18y_1^2y_2 - 18y_1^2y_2^2 \neq 6(1-y_2)$$

Since $f(y_1)f(y_2) \neq f(y_1, y_2)$, we conclude Y_1, Y_2 are not independent

b)

For the expected value of y_1, we use $f(y_1)$

$$E[y_1] = \int_0^1 y_1 f(y_1) dy_1$$

$$E[y_1] = \int_0^1 y_1(3 - 6y_1 + 3y_1^2) dy_1$$

$$E[y_1] = \int_0^1 3y_1 - 6y_1^2 + 3y_1^3 \, dy_1$$

$$E[y_1] = \left[\frac{3y_1^2}{2} - \frac{6y_1^3}{3} + \frac{3y_1^4}{4}\right]_0^1 = \frac{3}{2} - \frac{6}{3} + \frac{3}{4} = \frac{18 - 24 + 9}{12} = \frac{3}{12} = \frac{1}{4}$$

c)

$$P(Y_1|Y_2 = y_2) = \frac{f(y_1, y_2)}{f(y_2)}$$

$$P(Y_1|Y_2 = y_2) = \frac{6(1-y_2)}{6y_2 - 6y_2^2}$$

$$P(Y_1|Y_2 = y_2) = \frac{6(1-y_2)}{6y_2(1-y_2)}$$

$$P(Y_1|Y_2 = y_2) = \frac{1}{y_2} = \frac{1}{y_2 - 0} \sim Uniform(0, y_2)$$

3

Let a sample of size *n* from a random variable $Y \sim Exponential(\lambda)$, where $\lambda = \frac{1}{\beta}$ is the rate parameter and $\lambda > 0$. The density is given by $f(y) = \lambda e^{-\lambda y}$. Note that $E[Y] = \frac{1}{\lambda}$

a) Estimate λ using Method of Moments. 1p

b) Estimate λ using Maximum Likelihood. 2p

c) Estimate λ using Bayesian method with a conjugate Gamma prior $\lambda \sim$Gamma(8,4).
What is the actual value of the estimate for $n = 10$ and $\sum y = 30$? 2p

a)

We have to solve $\mu'_k = m'_k$, and in this case we have only one parameter so k=1.

$$\mu'_1 = E[Y^1] = E[Y] = \frac{1}{\lambda}$$

$$m'_1 = \frac{1}{n}\sum_{i=1}^{n} Y_i^1 = \bar{Y}$$

Now solving $\frac{1}{\lambda} = \bar{Y}$ we get that the methods-of-moments estimate is $\hat{\lambda}_{MoM} = \frac{1}{\bar{Y}}$

b)

The likelihood is

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda y_i}$$

$$L(\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} y_i}$$

Now the log-likelihood is

$$l(\lambda) = \ln\left(\lambda^n e^{-\lambda \sum_{i=1}^{n} y_i}\right) = nln(\lambda) - \lambda \sum_{i=1}^{n} y_i$$

Taking derivative with respect to λ

$$\frac{d\, l(\lambda)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} y_i$$

After equaling to zero we have

$$\frac{n}{\lambda} = \sum_{i=1}^{n} y_i$$

And finally

$$\hat{\lambda}_{MLE} = \frac{n}{\sum_{i=1}^{n} y_i} = \frac{1}{\bar{Y}}$$

c)

This is the Gamma-Exponential model for λ

We know from b) the likelihood for λ is $L(\lambda) = p(y|\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} y_i}$

The prior Gamma(8,4) for λ is $f(\lambda|\alpha = 8, \beta = 4) = p(\lambda) = \frac{1}{\Gamma(8)4^8}\lambda^{8-1}e^{\frac{-\lambda}{4}}$

Now the posterior for λ, $p(\lambda|y)$, is proportional to Likelihood x prior, so multiplying we get

$$p(\lambda|y) \propto \lambda^n \lambda^7 e^{-\lambda \sum_{i=1}^{n} y_i} e^{\frac{-\lambda}{4}}$$

$$p(\lambda|y) \propto \lambda^{n+7} e^{-\lambda(\frac{1}{4} + \sum_{i=1}^{n} y_i)} \propto Gamma(8 + n, \frac{4}{1 + 4\sum_{i=1}^{n} y_i})$$

The Bayes estimate is the mean of the posterior density $p(\lambda|y)$ of the parameter; in the case of $Y \sim Gamma(\alpha, \beta)$ we know that $E[Y] = \alpha\beta$ so our estimator for λ is

$$\hat{\lambda}_{BAYES} = \frac{4(8+n)}{1 + 4\sum_{i=1}^{n} y_i}$$

For the specific case of $n = 10$ and $\sum y = 30$, we replace in the formula above

$$\hat{\lambda}_{BAYES} = \frac{4(8+10)}{1+4(30)} = \frac{72}{121} \approx 0.595$$

4

We collect the following random sample (y_i, x_i) of size n = 6 where Y is the dependent (output) variable and X the independent (input) variable.

Y	X
7	1
8	3
9	5
11	7
15	9
22	11

a) Fit the model, $Y = \beta_0 + \beta_1 x + \varepsilon$, to the data above using ordinary least-squares (OLS). Present the equation of the fitted line. Also, plot the points and sketch the fitted line.
2p

b) Calculate the sum of squared (SSE) and mean squared (MSE) errors.
1p

c) Test if the intercept is significant ($H_0: \beta_0 = 0$) at 95% confidence level. Conclusions?
2p

a)

We know the OLS solutions for β_0, β_1 are

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

First we need to calculate means for both variables

$$\bar{y} = \frac{7+8+9+11+15+22}{6} = \frac{72}{6} = 12$$
$$\bar{x} = \frac{1+3+5+7+9+11}{6} = \frac{36}{6} = 6$$

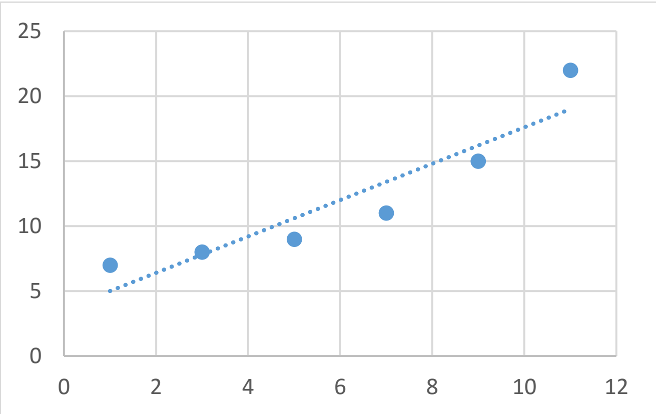
Now is useful to construct a table like this

y	x	$y - \bar{y}$	$x - \bar{x}$	$(x - \bar{x})^2$	$(x - \bar{x})(y - \bar{y})$
7	1	-5	-5	25	25
8	3	-4	-3	9	12
9	5	-3	-1	1	3
11	7	-1	1	1	-1
15	9	3	3	9	9
22	11	10	5	25	50

Summing the last two columns we get $S_{xx} = 70,\ S_{xy} = 98$.

$$\hat{\beta}_1 = \frac{98}{70} = 1.4$$
$$\hat{\beta}_0 = 12 - 1.4(6) = 3.6$$

The equation of the fitted line is then $\hat{y} = 3.6 + 1.4x$, and the points plus line looks like



b) For the SSE and MSE we will need residuals, i.e.

$e_i = y_i - \hat{y}_i$, where $\hat{y} = 3.6 + 1.4x$

So we should add to the table above three columns such as

y	$\hat{y}$	e	e^2
7	5	2	4
8	7.8	0.2	0.04
9	10.6	-1.6	2.56
11	13.4	-2.4	5.76
15	16.2	-1.2	1.44
22	19	3	9

Now $SSE = \sum_{i=1}^{6} e_i^2 = 22.8$, and $MSE = \frac{SSE}{n-2} = \frac{22.8}{4} = 5.7$

c) This is a two-tailed test in the form $H_0: \beta_0 = 0, H_a: \beta_0 \neq 0$, and $\alpha = 0.05$. The statistic for β_0 is:

$$T = \frac{\widehat{\beta}_0 - \beta_0}{S\sqrt{\frac{\sum x_i^2}{nS_{xx}}}} \sim t(n-2)$$

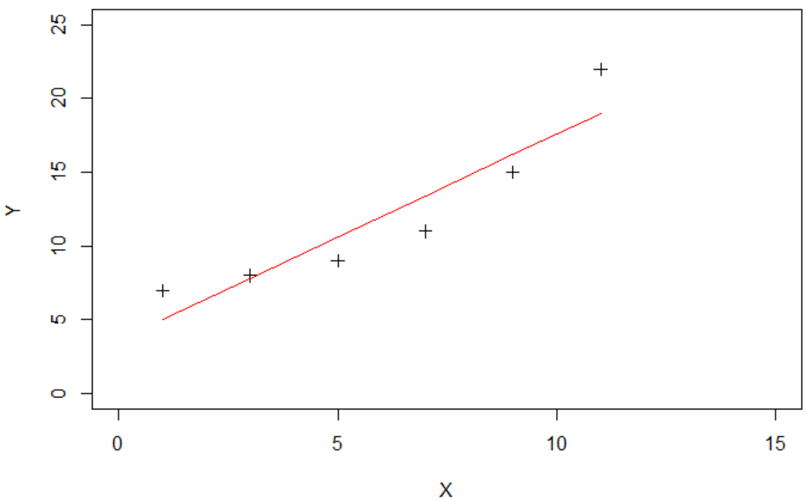
We have n=6, and we can approximate $S^2 = MSE$, so $S = \sqrt{5.7} = 2.38$. Also from calculations above we know $S_{xx} = 70, \hat{\beta}_0 = 3.6$, and from the table is easy to calculate $\sum x_i^2 = 1 + 9 + 25 + 49 + 81 + 121 = 286$. Replacing all the previous into the statistic we get

$$T = \frac{3.6 - 0}{2.38\sqrt{\frac{286}{6(70)}}} = \frac{3.6}{1.9697} \approx 1.8276$$

At 95% confidence, if we check the table we get that critical values are at $\pm t_{\frac{\alpha}{2}}(n-2) = \pm t_{0.025}(4) = \pm 2.776$, so acceptance region is defined by $[-2.776 < T < 2.776]$, and the rejection region by $[|2.776| < T]$. Finally, since the value of the statistic $T = 1.8276$ is within the acceptance region, we cannot reject the null hypothesis $H_0: \beta_0 = 0$.

(Extra) Below we present a simple code in R to fit the model:

```
> Y <- c(7,8,9,11,15,22)
> X <- c(1,3,5,7,9,11)
> mod <- lm(Y~X)
> summary(mod)

Call:
lm(formula = Y ~ X)

Residuals:
   1    2    3    4    5    6 
 2.0  0.2 -1.6 -2.4 -1.2  3.0 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept)   3.6000     1.9701   1.827  0.14167   
X             1.4000     0.2854   4.906  0.00801 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.387 on 4 degrees of freedom
Multiple R-squared:  0.8575,	Adjusted R-squared:  0.8219 
F-statistic: 24.07 on 1 and 4 DF,  p-value: 0.008009

> plot(Y~X,ylim=c(0,25),xlim=c(0,15),pch=3)
> lines(fitted(mod)~X,col=2)
```



5

Show, by using the properties of the expected value and variance, that the expected value of a Chi-square distributed variable is equal to the number of degrees of freedom. That is, show that if $Y \sim \chi^2(v)$ then $E[Y] = v$.

Hint: $Y = \sum_{i=1}^{v} \left(\frac{X_i - \mu}{\sigma}\right)^2$ and $X_i \sim N(\mu, \sigma^2)$

3p

We have to show that $E[Y] = v$, and we know that $Y = \sum_{i=1}^{v} \left(\frac{X_i - \mu}{\sigma}\right)^2$, so we have to show that

$$E\left[\sum_{i=1}^{v} \left(\frac{X_i - \mu}{\sigma}\right)^2\right] = v$$

Working with the square of the sum inside the summatory yields

$$E\left[\sum_{i=1}^{v} \left(\frac{{X_i}^2 + \mu^2 - 2\mu X_i}{\sigma^2}\right)\right]$$

Sigma^2 is constant so we take out of both sum and expectation so we have

$$\frac{1}{\sigma^2} E\left[\sum_{i=1}^{v} \left({X_i}^2 + \mu^2 - 2\mu X_i\right)\right]$$

The expected value of the sum is the sum of the expected values:

$$\frac{1}{\sigma^2}\left[\sum_{i=1}^{v} E[{X_i}^2] + \sum_{i=1}^{v} E[\mu^2] - 2\mu \sum_{i=1}^{v} E[X_i]\right]$$

From the hint we know that $E[X_i] = \mu$, and for $E[{X_i}^2]$ we use the relation $V[X] = E[X^2] - E[X]^2$ and we get that $E[X^2] = V[X] + E[X]^2$, which in the present case results in $E[{X_i}^2] = \sigma^2 + \mu^2$. Remember also that the expected value of a constant is the constant itself so $E[\mu^2] = \mu^2$. Replacing all the previous in the expression above we have

$$\frac{1}{\sigma^2}\left[\sum_{i=1}^{v} (\sigma^2 + \mu^2) + \sum_{i=1}^{v} \mu^2 - 2\mu \sum_{i=1}^{v} \mu\right]$$

Now everything inside the sums is constant so we multiply by v

$$\frac{1}{\sigma^2}[v\sigma^2 + v\mu^2 + v\mu^2 - 2v\mu^2] = \frac{1}{\sigma^2}[v\sigma^2] = v$$

Good luck!
Lycka till!

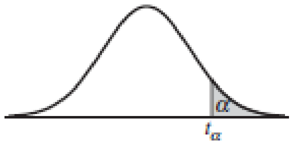
Table 1 Discrete Distributions

Distribution	Probability Function	Mean	Variance	Moment-Generating Function
Binomial	$p(y) = \binom{n}{y} p^y (1-p)^{n-y};$ $y = 0, 1, \ldots, n$	np	$np(1-p)$	$[pe^t + (1-p)]^n$
Geometric	$p(y) = p(1-p)^{y-1};$ $y = 1, 2, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t}$
Hypergeometric	$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}};$ $y = 0, 1, \ldots, n$ if $n \le r$, $y = 0, 1, \ldots, r$ if $n > r$	$\frac{nr}{N}$	$n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$	does not exist in closed form
Poisson	$p(y) = \frac{\lambda^y e^{-\lambda}}{y!};$ $y = 0, 1, 2, \ldots$	λ	λ	$\exp[\lambda(e^t - 1)]$
Negative binomial	$p(y) = \binom{y-1}{r-1} p^r (1-p)^{y-r};$ $y = r, r+1, \ldots$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	$\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$

Table 2 Continuous Distributions

Distribution	Probability Function	Mean	Variance	Moment-Generating Function
Uniform	$f(y) = \frac{1}{\theta_2 - \theta_1};\ \theta_1 \le y \le \theta_2$	$\frac{\theta_1 + \theta_2}{2}$	$\frac{(\theta_2 - \theta_1)^2}{12}$	$\frac{e^{t\theta_2} - e^{t\theta_1}}{t(\theta_2 - \theta_1)}$
Normal	$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\left(\frac{1}{2\sigma^2}\right)(y-\mu)^2\right]$ $-\infty < y < +\infty$	μ	σ^2	$\exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)$
Exponential	$f(y) = \frac{1}{\beta} e^{-y/\beta};\quad \beta > 0$ $0 < y < \infty$	β	β^2	$(1 - \beta t)^{-1}$
Gamma	$f(y) = \left[\frac{1}{\Gamma(\alpha)\beta^\alpha}\right] y^{\alpha-1} e^{-y/\beta};$ $0 < y < \infty$	$\alpha\beta$	$\alpha\beta^2$	$(1 - \beta t)^{-\alpha}$
Chi-square	$f(y) = \frac{(y)^{(v/2)-1} e^{-y/2}}{2^{v/2}\Gamma(v/2)};$ $y > 0$	v	$2v$	$(1 - 2t)^{-v/2}$
Beta	$f(y) = \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right] y^{\alpha-1}(1-y)^{\beta-1};$ $0 < y < 1$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	does not exist in closed form

Percentage Points of the *t* Distributions



$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	df
3.078	6.314	12.706	31.821	63.657	1
1.886	2.920	4.303	6.965	9.925	2
1.638	2.353	3.182	4.541	5.841	3
1.533	2.132	2.776	3.747	4.604	4
1.476	2.015	2.571	3.365	4.032	5
1.440	1.943	2.447	3.143	3.707	6
1.415	1.895	2.365	2.998	3.499	7
1.397	1.860	2.306	2.896	3.355	8
1.383	1.833	2.262	2.821	3.250	9
1.372	1.812	2.228	2.764	3.169	10
1.363	1.796	2.201	2.718	3.106	11
1.356	1.782	2.179	2.681	3.055	12
1.350	1.771	2.160	2.650	3.012	13
1.345	1.761	2.145	2.624	2.977	14
1.341	1.753	2.131	2.602	2.947	15
1.337	1.746	2.120	2.583	2.921	16
1.333	1.740	2.110	2.567	2.898	17
1.330	1.734	2.101	2.552	2.878	18
1.328	1.729	2.093	2.539	2.861	19
1.325	1.725	2.086	2.528	2.845	20
1.323	1.721	2.080	2.518	2.831	21
1.321	1.717	2.074	2.508	2.819	22
1.319	1.714	2.069	2.500	2.807	23
1.318	1.711	2.064	2.492	2.797	24
1.316	1.708	2.060	2.485	2.787	25
1.315	1.706	2.056	2.479	2.779	26
1.314	1.703	2.052	2.473	2.771	27
1.313	1.701	2.048	2.467	2.763	28
1.311	1.699	2.045	2.462	2.756	29
1.282	1.645	1.960	2.326	2.576	inf.