FINAL EXAM review (draft: 2019/05/05-11:40:52)

The final exam 3 is comprehensive over the entire course

My recommendation is to study the all quizzes and exams. Make sure you understand everything on those perfectly. Make sure you have looked over any example problems included in the course notes that I have posted online. Study all review materials I have posted over the entire semester. Make sure you understand the webwork homework problems. Of course, you should also understand everything and all examples covered in class. Sometimes these vary somewhat form what is seen on homework or in my online notes. You should also pay attention to what we have done with R as well.

See the tables at the end of this document. Since we will not have the capability to use R on the exam, you will be given tables like them. Make sure you know how to read them.

Example/practice problems:

1. Patients arrive at a hospital on average five every two hours. Model the number of patients in a given time interval as Poisson and the time between successive patients as exponential.

   (a) How many patients are expected in a 24 hour period?

   Solution:

   Using Poisson, we scale the arrival rate: $5/2 \cdot 24 = 60$ so we use $\lambda = 60$ and this is the expected number of patients in a 24 hours period.

   (b) What is the probability that no patients arrive during a 1 hour break?

   Solution:

   Using Poisson, we scale the arrival rate: $5/2 \cdot 1 = 2.5$. The probability of zero patients is $e^{-\lambda} = e^{-2.5} \approx 0.082085$

   (c) If there are only 3 open beds, what is the probability that all patients over the next hour will each get a bed immediately? Assume each patient is assigned to a bed instantaneously if there is one available, and that no patients leave.

   Solution:

   Using Poisson, we scale the arrival rate: $5/2 \cdot 1 = 2.5$. We want to know the probability that there are 3 or fewer patients over the next hour.

   $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$ for Poisson. We calculate the probabilities for $x = 0, 1, 2, 3$ and sum them. If there were a 4th patient, they would not get a bed.

   $P(X \leq 3) = e^{-\lambda} + e^{-\lambda}\lambda + \frac{e^{-\lambda}\lambda^2}{2} + \frac{e^{-\lambda}\lambda^3}{6} \approx 0.0820850 + 0.2052125 + 0.2565156 + 0.2137630$

   Be sure you can do such a calculation in a basic calculator.

   In R we can do this with `ppois(3,lambda=5/2*1)` = 0.7575761.

   Also, we could use `sum(dpois(0:3,lambda=5/2*1))`.

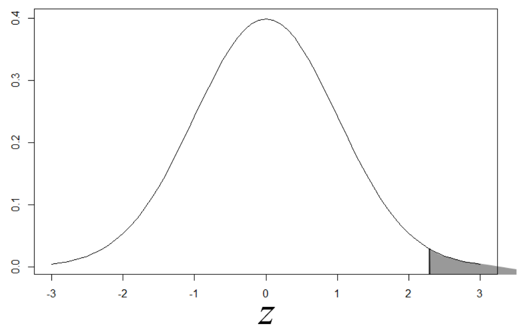
   (d) What is the probability that the next patient arrives between after 30 but before 60 minutes.

   Solution:

   We'll use the exponential distribution here. Let X be the time in hours between successive patient arrivals. We wish to know $P(0.5 < X < 1)$.

   $P(0.5 < X <) = \int_{0.5}^{1} \frac{5}{2} e^{-\frac{5}{2}x}\, dx = e^{-\frac{5}{2}\cdot 1} - e^{-\frac{5}{2}\cdot 0.5} \approx 0.2044198$

2. The graph below is given for a particular hypothesis test for a population mean. The test statistic is the edge of the shaded region. Answer the given questions.



(a) Which of the following is true, false, or cannot be determined from the information given.

i. This is a 2-sided test for a single population mean.

ii. We know the population variance.

iii. The sample size is large.

Solution:

i. False, it is clearly 1-sided.

ii. Uncertain. z-test can be used for any sample if we know σ^2 and assume a normal population. z-test can generally still be used if we do not know the variance, but the sample size is large (generally $n > 30$). So we can't be sure from the information given.

iii. Uncertain. As stated above, the sample size could be large or small, depending on information we are not given.

(b) Write the null and alternative hypotheses.

Solution:

$H_0 : \mu = \mu_0$, $H_a : \mu > \mu_0$

We don't know what the null mean μ_0 is, but we know that the hypotheses will be of this form.

(c) Estimate the test statistic and p-value.

Solution:

The test statistic looks to be approximately $z \approx 2.2$ or 2.3. Using the table at the end of this study guide, we see that `pnorm(2.3)`=0.9893. We subtract this from one to get $p = 0.0107$.

Remember we can use `pnorm(-2.3)` or `1-pnorm(2.3)` for such a 1-sided test.

(d) What would we conclude?

Solution:

We are not given a particular significance level α, but we would reject the null hypothesis at any $alpha \geq p$. E.g. at the popular $\alpha = 0.05$ level, we would reject H_0. But at $\alpha = 0.01$ we would not reject H_0.

3. Let X be the number of heads in three flips of a biased coin with $P(H) = \frac{3}{4}$.

(a) Write the pmf for X.

Solution:

X is binomial with 3 trials and probability of success $\frac{3}{4}$ so

$$f_X(x) = \binom{3}{x}(0.75)^x(0.25)^{3-x} = \binom{3}{x}\frac{3^x}{4^3}.$$

$$f_X(x) = \begin{cases} 1/64 & \text{for } x = 0 \\ 9/64 & \text{for } x = 1 \\ 27/64 & \text{for } x = 2 \\ 27/64 & \text{for } x = 3 \\ 0 & \text{otherwise} \end{cases}$$

(b) Write the cdf for X.

Solution:

$$F_X(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1/64 & \text{for } 0 \leq x < 1 \\ 10/64 & \text{for } 1 \leq x < 2 \\ 37/64 & \text{for } 2 \leq x < 3 \\ 1 & \text{for } 3 \leq x \end{cases}$$

(c) Using your cdf above, calculate $P(0 < X < 3)$.

Solution:

Recall that for discrete random variables we need to be careful with inequalities. $P(0 < X < 3) = P(0 < X \leq 2) = F(2) - F(0) = 37/64 - 1/64 = 36/64$.

Notice that this is just the probability of $P(X = 1) + P(X = 2)$ in this case.

4. Find the pdf given cdf $F_X(x) = x^3$ for $0 \leq x \leq 1$.

Solution:

$f_X(x) = \frac{d}{dx} F_X(x) = 3x^2$ for $0 \leq x \leq 1$.

5. By hand calculate the linear regression for data in table below.

(a) Fill in the entire table.

x	y	$\overline{x}$	$\overline{y}$	$x - \overline{x}$	$y - \overline{y}$	$(x - \overline{x})^2$	$(y - \overline{y})^2$	$(x - \overline{x})(y - \overline{y})$
10	100							
13	120							
19	150							
22	170							

Solution:

x	y	$\overline{x}$	$\overline{y}$	$x - \overline{x}$	$y - \overline{y}$	$(x - \overline{x})^2$	$(y - \overline{y})^2$	$(x - \overline{x})(y - \overline{y})$
10	100	16	135	-6	-35	36	1225	210
13	120	16	135	-3	-15	9	225	45
19	150	16	135	3	15	9	225	45
22	170	16	135	6	35	36	1225	210

(b) Find the formula for the regression line.

Solution:

slope$=\hat{b} = cov(x, y)/var(x) = \frac{\frac{1}{n-1}\sum(x_i - \overline{x})(y_i - \overline{y})}{\frac{1}{n-1}\sum(x_i - \overline{x})^2} = \frac{210+45+45+210}{36+9+9+36} = 51/9 \approx 5.67$

intercept $= \hat{a} = \overline{y} - \hat{b}\overline{x} = 135 - 51/9 \cdot 16 \approx 44.3$

(c) Calculate the correlation coefficient and interpret it.

Solution:

$$R = cor(x,y) = \frac{cov(x,y)}{sd(x)sd(y)} = \frac{\frac{1}{n-1}\sum(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\frac{1}{n-1}\sum(x_i-\overline{x})^2}\cdot\sqrt{\frac{1}{n-1}\sum(y_i-\overline{y})^2}}$$
$$= \frac{210+45+45+210}{\sqrt{(36+9+9+36)(1225+225+225+1225)}} \approx 0.9982744$$

Thus X and Y are highly correlated. There is very little deviation from the regression line.

(d) Calculate the SSE.

Solution:

The fitted y values are $\{101, 118, 152, 169\}$ so

$SSE = \sum(y_i - \hat{y}_i)^2 = (100-101)^2 + (120-118)^2 + (150-152)^2 + (170-169)^2 = 10$

z	pnorm(z)
0.0	0.5000
0.5	0.6915
1.0	0.8413
1.1	0.8643
1.2	0.8849
1.3	0.9032
1.4	0.9192
1.5	0.9332
1.6	0.9452
1.7	0.9554
1.8	0.9641
1.9	0.9713

z	pnorm(z)
2.0	0.9772
2.1	0.9821
2.2	0.9861
2.3	0.9893
2.4	0.9918
2.5	0.9938
2.6	0.9953
2.7	0.9965
2.8	0.9974
2.9	0.9981
3.0	0.9987
3.1	0.9990

p	qnorm(p)
0.5	0.0000
0.6	0.2533
0.7	0.5244
0.8	0.8416
0.9	1.2816
0.95	1.6449
0.975	1.9600
0.98	2.0537
0.99	2.3263
0.995	2.5758
0.9995	3.2905

	qt(p,df) given in table for df degrees of freedom listed below									
p	df = 1	2	3	4	5	6	7	8	9	10
0.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.6	0.325	0.289	0.277	0.271	0.267	0.265	0.263	0.262	0.261	0.260
0.7	0.727	0.617	0.584	0.569	0.559	0.553	0.549	0.546	0.543	0.542
0.8	1.376	1.061	0.978	0.941	0.920	0.906	0.896	0.889	0.883	0.879
0.9	3.078	1.886	1.638	1.533	1.476	1.440	1.415	1.397	1.383	1.372
0.95	6.314	2.920	2.353	2.132	2.015	1.943	1.895	1.860	1.833	1.812
0.975	12.706	4.303	3.182	2.776	2.571	2.447	2.365	2.306	2.262	2.228
0.98	15.895	4.849	3.482	2.999	2.757	2.612	2.517	2.449	2.398	2.359
0.99	31.821	6.965	4.541	3.747	3.365	3.143	2.998	2.896	2.821	2.764
0.995	63.657	9.925	5.841	4.604	4.032	3.707	3.499	3.355	3.250	3.169
0.9995	636.619	31.599	12.924	8.610	6.869	5.959	5.408	5.041	4.781	4.587

	pt(t,df) given in table for df degrees of freedom listed below									
t	df = 1	2	3	4	5	6	7	8	9	10
0	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
0.5	0.6476	0.6667	0.6743	0.6783	0.6809	0.6826	0.6838	0.6847	0.6855	0.6861
1.0	0.7500	0.7887	0.8045	0.8130	0.8184	0.8220	0.8247	0.8267	0.8283	0.8296
1.5	0.8128	0.8638	0.8847	0.8960	0.9030	0.9079	0.9114	0.9140	0.9161	0.9177
2.0	0.8524	0.9082	0.9303	0.9419	0.9490	0.9538	0.9572	0.9597	0.9617	0.9633
2.5	0.8789	0.9352	0.9561	0.9666	0.9728	0.9767	0.9795	0.9815	0.9831	0.9843
3.0	0.8976	0.9523	0.9712	0.9800	0.9850	0.9880	0.9900	0.9915	0.9925	0.9933
3.5	0.9114	0.9636	0.9803	0.9876	0.9914	0.9936	0.9950	0.9960	0.9966	0.9971
4.0	0.9220	0.9714	0.9860	0.9919	0.9948	0.9964	0.9974	0.9980	0.9984	0.9987
4.5	0.9304	0.9770	0.9898	0.9946	0.9968	0.9979	0.9986	0.9990	0.9993	0.9994
5.0	0.9372	0.9811	0.9923	0.9963	0.9979	0.9988	0.9992	0.9995	0.9996	0.9997