COLUMBIA UNIVERSITY
Statistical Inference and Modeling
Quiz 4

Prof. Marco Avella
TA: Ian Kinsella
November 7, 2019

Name:

UNI:

You have 20 minutes to answer the following questions. Good luck!

Question 1 (4 points)

Assume that we have an i.i.d sample of pairs $(\mathbf{x}_1, Y_1), \ldots, (\mathbf{x}_n, Y_n)$ where $\mathbf{x}_i \in \mathbb{R}^p$. Consider the linear model

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \ i = 1, \ldots, n$$

where ε_i, are i.i.d. $N(0, \sigma^2)$. Let $\hat{\boldsymbol{\beta}}$ be the least squares estimator and $\hat{\boldsymbol{\beta}}_\lambda = (\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^\top \mathbf{y}$ the ridge regression estimator. Answer the following questions assuming the design matrix $\mathbf{X}$ is fixed.

1. Compute the mean and variance of $\hat{\boldsymbol{\beta}}_\lambda$, and give its distribution.
2. Consider the partition $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ and respective slope parameter partition $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$, and define the restricted least squares estimator $\hat{\boldsymbol{\beta}}_1^R = (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1 \mathbf{y}$. Give an exact test and and asymptotic test for the null hypothesis $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$.

1) $\mathbb{E}[\hat{\beta}_\lambda] = (X^T X + \lambda I)^{-1} X^T \mathbb{E}[y] = (X^T X + \lambda I)^{-1} X^T X \beta =: \beta_\lambda$

1.5 pts

$\text{Var}[\hat{\beta}_\lambda] = (X^T X + \lambda I)^{-1} X^T \underbrace{\text{Var}(\varepsilon)}_{=\sigma^2 I} X (X^T X + \lambda I)^{-1}$

$= (X^T X + \lambda I)^{-1} X^T X (X^T X + \lambda I)^{-1} =: V_\lambda$

0,5 pts

$\Rightarrow \hat{\beta}_\lambda = N(\beta_\lambda, V_\lambda)$

2) Exact: F-test, $\frac{p-q}{n-p} \cdot \frac{\|Y - X_1 \hat{\beta}_1^R\|_2^2 - \|Y - X\hat{\beta}\|_2^2}{\|Y - X\hat{\beta}\|_2^2} > \underbrace{G^{-1}(1-\alpha)}$

CDF of $F_{p-q, n-p}$

$1-\alpha$ quantile of $F_{p-q, n-p}$

Asymptotic: LRT, $n \log\left(\frac{\|Y - X_1 \hat{\beta}_1^R\|_2^2}{\|Y - X\hat{\beta}\|_2^2}\right) > H^{-1}(1-\alpha)$

H is CDF of χ^2_{p-q}

$(1-\alpha)$-quantile of a χ^2_{p-q}

Question 2 (6 points)

We consider data come from a study that examined the correlation between the level of prostate specific antigen and a number of clinical measures in men who were about to receive a radical prostatectomy. The list of variables is detailed below. We also have access to an R output.

- `lcavol`: log(cancer volume)
- `lweight`: log(prostate weight)
- `age`: age
- `lbph`: log(benign prostatic hyperplasia amount)
- `svi`: seminal vesicle invasion
- `lcp`: log(capsular penetration)
- `gleason`: Gleason score
- `pgg45`: percentage Gleason scores 4 or 5
- `lpsa`: log(prostate specific antigen)

```
Call:
lm(formula = y ~ X)

Residuals:
     Min       1Q   Median       3Q      Max
-1.73316 -0.37133 -0.01702  0.41414  1.63811

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.669399   1.296381   0.516  0.60690
Xlcavol      0.587023   0.087920   6.677 2.11e-09 ***
Xlweight     0.454461   0.170012   2.673  0.00896 **
Xage        -0.019637   0.011173  -1.758  0.08229 .
Xlbph        0.107054   0.058449   1.832  0.07040 .
Xsvi         0.766156   0.244309   3.136  0.00233 **
Xlcp        -0.105474   0.091013  -1.159  0.24964
Xgleason     0.045136   0.157464   0.287  0.77506
Xpgg45       0.004525   0.004421   1.024  0.30885
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7084 on 88 degrees of freedom
Multiple R-squared:  0.6548,Adjusted R-squared:  0.6234
F-statistic: 20.86 on 8 and 88 DF,  p-value: < 2.2e-16
```

1. What was the sample size n in these data?
2. Which 3 covariates that seem to be the most predictive? Give a two lines interpretation of the meaning of this.
3. Interpret the sign and significance of age. How would a one sided test change the conclusion?
4. Do we have evidence against $H_0 : \sigma^2 > 0.5$?

1pt

1) $n = 98 = DF + \#$ parameters

1,5 pts

2) log (cancer volume), log (prostate weight), log(capsular penetration)

These tree covariates are positively associated with higher prostate antigen levels. The significantly predict the severity of the prostate cancer.

1,5 pts

3) Age is not significant at the 0,05 level based on a two sided alternative. It would however reject the one sided null $H_0 : \beta_{age} < 0$. This hypothesis is not obvious at first but reflects the idea that our immune system gets weaker when men age.

2pts

4) $\hat{\sigma}^2 = 0{,}7084^2 \approx 0{,}49$

$$\hat{\sigma}^2 \sim \frac{\sigma^2}{88} \chi^2_{88}$$

$$\Rightarrow \mathbb{P}_{\sigma^2=0{,}5}\left(\chi^2_{88} > \underbrace{\frac{88}{0{,}5} \cdot 0{,}7084^2}_{\approx \mathbb{E}\left[\chi^2_{88}\right] = 88} \right)$$ is not a small probability

$\Rightarrow$ cannot reject H_0

Table of the *t* Distribution

If X has a t distribution with m degrees of freedom, the table gives the value of x such that $\Pr(X \le x) = p$.

m	$p = .55$	.60	.65	.70	.75	.80	.85	.90	.95	.975	.99	.995
1	.158	.325	.510	.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	.142	.289	.445	.617	.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	.137	.277	.424	.584	.765	.978	1.250	1.638	2.353	3.182	4.541	5.841
4	.134	.271	.414	.569	.741	.941	1.190	1.533	2.132	2.776	3.747	4.604
5	.132	.267	.408	.559	.727	.920	1.156	1.476	2.015	2.571	3.365	4.032
6	.131	.265	.404	.553	.718	.906	1.134	1.440	1.943	2.447	3.143	3.707
7	.130	.263	.402	.549	.711	.896	1.119	1.415	1.895	2.365	2.998	3.499
8	.130	.262	.399	.546	.706	.889	1.108	1.397	1.860	2.306	2.896	3.355
9	.129	.261	.398	.543	.703	.883	1.100	1.383	1.833	2.262	2.821	3.250
10	.129	.260	.397	.542	.700	.879	1.093	1.372	1.812	2.228	2.764	3.169
11	.129	.260	.396	.540	.697	.876	1.088	1.363	1.796	2.201	2.718	3.106
12	.128	.259	.395	.539	.695	.873	1.083	1.356	1.782	2.179	2.681	3.055
13	.128	.259	.394	.538	.694	.870	1.079	1.350	1.771	2.160	2.650	3.012
14	.128	.258	.393	.537	.692	.868	1.076	1.345	1.761	2.145	2.624	2.977
15	.128	.258	.393	.536	.691	.866	1.074	1.341	1.753	2.131	2.602	2.947
16	.128	.258	.392	.535	.690	.865	1.071	1.337	1.746	2.120	2.583	2.921
17	.128	.257	.392	.534	.689	.863	1.069	1.333	1.740	2.110	2.567	2.898
18	.127	.257	.392	.534	.688	.862	1.067	1.330	1.734	2.101	2.552	2.878
19	.127	.257	.391	.533	.688	.861	1.066	1.328	1.729	2.093	2.539	2.861
20	.127	.257	.391	.533	.687	.860	1.064	1.325	1.725	2.086	2.528	2.845
21	.127	.257	.391	.532	.686	.859	1.063	1.323	1.721	2.080	2.518	2.831
22	.127	.256	.390	.532	.686	.858	1.061	1.321	1.717	2.074	2.508	2.819
23	.127	.256	.390	.532	.685	.858	1.060	1.319	1.714	2.069	2.500	2.807
24	.127	.256	.390	.531	.685	.857	1.059	1.318	1.711	2.064	2.492	2.797
25	.127	.256	.390	.531	.684	.856	1.058	1.316	1.708	2.060	2.485	2.787
26	.127	.256	.390	.531	.684	.856	1.058	1.315	1.706	2.056	2.479	2.779
27	.127	.256	.389	.531	.684	.855	1.057	1.314	1.703	2.052	2.473	2.771
28	.127	.256	.389	.530	.683	.855	1.056	1.313	1.701	2.048	2.467	2.763
29	.127	.256	.389	.530	.683	.854	1.055	1.311	1.699	2.045	2.462	2.756
30	.127	.256	.389	.530	.683	.854	1.055	1.310	1.697	2.042	2.457	2.750
40	.126	.255	.388	.529	.681	.851	1.050	1.303	1.684	2.021	2.423	2.704
60	.126	.254	.387	.527	.679	.848	1.046	1.296	1.671	2.000	2.390	2.660
120	.126	.254	.386	.526	.677	.845	1.041	1.289	1.658	1.980	2.358	2.617
∞	.126	.253	.385	.524	.674	.842	1.036	1.282	1.645	1.960	2.326	2.576

Table of the Standard Normal Distribution Function

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{1}{2}u^2\right) du$$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0.5000	0.60	0.7257	1.20	0.8849	1.80	0.9641	2.40	0.9918
0.01	0.5040	0.61	0.7291	1.21	0.8869	1.81	0.9649	2.41	0.9920
0.02	0.5080	0.62	0.7324	1.22	0.8888	1.82	0.9656	2.42	0.9922
0.03	0.5120	0.63	0.7357	1.23	0.8907	1.83	0.9664	2.43	0.9925
0.04	0.5160	0.64	0.7389	1.24	0.8925	1.84	0.9671	2.44	0.9927
0.05	0.5199	0.65	0.7422	1.25	0.8944	1.85	0.9678	2.45	0.9929
0.06	0.5239	0.66	0.7454	1.26	0.8962	1.86	0.9686	2.46	0.9931
0.07	0.5279	0.67	0.7486	1.27	0.8980	1.87	0.9693	2.47	0.9932
0.08	0.5319	0.68	0.7517	1.28	0.8997	1.88	0.9699	2.48	0.9934
0.09	0.5359	0.69	0.7549	1.29	0.9015	1.89	0.9706	2.49	0.9936
0.10	0.5398	0.70	0.7580	1.30	0.9032	1.90	0.9713	2.50	0.9938
0.11	0.5438	0.71	0.7611	1.31	0.9049	1.91	0.9719	2.52	0.9941
0.12	0.5478	0.72	0.7642	1.32	0.9066	1.92	0.9726	2.54	0.9945
0.13	0.5517	0.73	0.7673	1.33	0.9082	1.93	0.9732	2.56	0.9948
0.14	0.5557	0.74	0.7704	1.34	0.9099	1.94	0.9738	2.58	0.9951
0.15	0.5596	0.75	0.7734	1.35	0.9115	1.95	0.9744	2.60	0.9953
0.16	0.5636	0.76	0.7764	1.36	0.9131	1.96	0.9750	2.62	0.9956
0.17	0.5675	0.77	0.7794	1.37	0.9147	1.97	0.9756	2.64	0.9959
0.18	0.5714	0.78	0.7823	1.38	0.9162	1.98	0.9761	2.66	0.9961
0.19	0.5753	0.79	0.7852	1.39	0.9177	1.99	0.9767	2.68	0.9963
0.20	0.5793	0.80	0.7881	1.40	0.9192	2.00	0.9773	2.70	0.9965
0.21	0.5832	0.81	0.7910	1.41	0.9207	2.01	0.9778	2.72	0.9967
0.22	0.5871	0.82	0.7939	1.42	0.9222	2.02	0.9783	2.74	0.9969
0.23	0.5910	0.83	0.7967	1.43	0.9236	2.03	0.9788	2.76	0.9971
0.24	0.5948	0.84	0.7995	1.44	0.9251	2.04	0.9793	2.78	0.9973
0.25	0.5987	0.85	0.8023	1.45	0.9265	2.05	0.9798	2.80	0.9974
0.26	0.6026	0.86	0.8051	1.46	0.9279	2.06	0.9803	2.82	0.9976
0.27	0.6064	0.87	0.8079	1.47	0.9292	2.07	0.9808	2.84	0.9977
0.28	0.6103	0.88	0.8106	1.48	0.9306	2.08	0.9812	2.86	0.9979
0.29	0.6141	0.89	0.8133	1.49	0.9319	2.09	0.9817	2.88	0.9980
0.30	0.6179	0.90	0.8159	1.50	0.9332	2.10	0.9821	2.90	0.9981
0.31	0.6217	0.91	0.8186	1.51	0.9345	2.11	0.9826	2.92	0.9983
0.32	0.6255	0.92	0.8212	1.52	0.9357	2.12	0.9830	2.94	0.9984
0.33	0.6293	0.93	0.8238	1.53	0.9370	2.13	0.9834	2.96	0.9985
0.34	0.6331	0.94	0.8264	1.54	0.9382	2.14	0.9838	2.98	0.9986
0.35	0.6368	0.95	0.8289	1.55	0.9394	2.15	0.9842	3.00	0.9987
0.36	0.6406	0.96	0.8315	1.56	0.9406	2.16	0.9846	3.05	0.9989
0.37	0.6443	0.97	0.8340	1.57	0.9418	2.17	0.9850	3.10	0.9990
0.38	0.6480	0.98	0.8365	1.58	0.9429	2.18	0.9854	3.15	0.9992
0.39	0.6517	0.99	0.8389	1.59	0.9441	2.19	0.9857	3.20	0.9993
0.40	0.6554	1.00	0.8413	1.60	0.9452	2.20	0.9861	3.25	0.9994
0.41	0.6591	1.01	0.8437	1.61	0.9463	2.21	0.9864	3.30	0.9995
0.42	0.6628	1.02	0.8461	1.62	0.9474	2.22	0.9868	3.35	0.9996
0.43	0.6664	1.03	0.8485	1.63	0.9485	2.23	0.9871	3.40	0.9997
0.44	0.6700	1.04	0.8508	1.64	0.9495	2.24	0.9875	3.45	0.9997
0.45	0.6736	1.05	0.8531	1.65	0.9505	2.25	0.9878	3.50	0.9998
0.46	0.6772	1.06	0.8554	1.66	0.9515	2.26	0.9881	3.55	0.9998
0.47	0.6808	1.07	0.8577	1.67	0.9525	2.27	0.9884	3.60	0.9998
0.48	0.6844	1.08	0.8599	1.68	0.9535	2.28	0.9887	3.65	0.9999
0.49	0.6879	1.09	0.8621	1.69	0.9545	2.29	0.9890	3.70	0.9999
0.50	0.6915	1.10	0.8643	1.70	0.9554	2.30	0.9893	3.75	0.9999
0.51	0.6950	1.11	0.8665	1.71	0.9564	2.31	0.9896	3.80	0.9999
0.52	0.6985	1.12	0.8686	1.72	0.9573	2.32	0.9898	3.85	0.9999
0.53	0.7019	1.13	0.8708	1.73	0.9582	2.33	0.9901	3.90	1.0000
0.54	0.7054	1.14	0.8729	1.74	0.9591	2.34	0.9904	3.95	1.0000
0.55	0.7088	1.15	0.8749	1.75	0.9599	2.35	0.9906	4.00	1.0000
0.56	0.7123	1.16	0.8770	1.76	0.9608	2.36	0.9909		
0.57	0.7157	1.17	0.8790	1.77	0.9616	2.37	0.9911		
0.58	0.7190	1.18	0.8810	1.78	0.9625	2.38	0.9913		
0.59	0.7224	1.19	0.8830	1.79	0.9633	2.39	0.9916		