UC Berkeley Stat 135, Spring 2020 – Final Exam

Instructor: Adam Lucas

Date/time: Tuesday, May 12 2020, 7:10PM to 10:00PM PST

You have until 10:15PM to scan and upload the test to Gradescope.

- You have 170 minutes to complete this final exam. Do not start the test until 7:10PM. Stop punctually at 10:00PM.
- There are 5 problems. I suggest that you start with the problem that you find the easiest. If you are stuck, move on to the next problem!
- **You are allowed two double-sided cheat sheet with formulas and definitions only (no worked out examples), plus your calculator. Any other resources will be strictly prohibited.**
- **Show your work and explain your reasoning thoroughly.** Ambiguous or otherwise unreadable answers will be marked incorrect.

Your Name (LAST, First): ______________________________

I certify that the work appearing on this exam is completely my own ______________________________

(Your signature)

Problem #	Points	Your score
1	10	
2	20	
3	25	
4	10	
5	25	
Total	90	

Problem 1. (10 pts total, 5 pts each part) ANOVA/Bonferroni test

In a parallel universe, a pharmaceutical factory invented two different pills ("Bowtruckle", "Thunderbird") to cure cancer. However, you, as the Chief Biostatistician, suspect whether any of the two pills actually work. You decide to conduct a randomized controlled trial including "Bowtruckle", "Thunderbird", and a third kind of pill ("Niffler") that has no effect. You randomly selected twelve individuals for each kind of pill, and measure their health indices after two months. The higher the health index is, the healthier the individual is. Below is some summary statistics from the data you collected.

Now, you are going to conduct some test and decide whether there is any difference between the mean health index of the three kinds of pills.

Pill	n	**Mean Health Index**	**SE of Mean**	**SD**
Bowtruckle	12	155	15	51.96
Thunderbird	12	160	18	62.35
Niffler	12	40	19	65.82

(a) After verifying some assumptions, you decide that an ANOVA test is appropriate. Perform the ANOVA at a 5% significance level, and fully state the assumptions that you verified.

(b) To decide which pair of pills is different, you decide to further apply the Bonferroni t-test. Explain what assumption would be needed to perform Bonferroni t-test, implement the Bonferroni t-test, and draw your conclusion.
Note: We have given you a Bonferroni table that gives you the cutoff for rejecting the null given the degrees of freedom and the number of tests performed (i.e. the number of treatment groups).

Solution

Part (a) We make a one way ANOVA test which assumes that the data in each treatment group is normal with equal variance. Our null hypothes is that the mean of each group is the same and the alternative is that at least one of the means is different.

We have $I = 3$ and $J = n = 12$

$$\text{SSW} = (J-1)\sum_{i=1}^{I} s_i^2 = 11 * (51.96^2 + 62.35^2 + 65.82^2) = 120116,$$

The total mean is

$$\bar{Y}_{..} = \frac{155 + 160 + 40}{3} = 118.3.$$

Then

$$\text{SSB} = J\sum_{i=1}^{I}(\bar{Y}_{i.} - \bar{Y}_{..})^2 = 12[(155 - 118.3)^2 + (160 - 118.3)^2 + (40 - 118.3)^2] = 110600$$

Then,

$$\text{F} = \frac{\text{SSB}/(I-1)}{\text{SSW}/(I(J-1))} = \frac{110600/2}{120116/33} = 15.19281$$

According to the F-table attached, $F_{.05,2,33}$ is between 3.23 and 3.32, which is smaller than 15.19281 so we reject the null that all of the tests are the same.

Part (b)

$$S_p^2 = \frac{\text{SSW}}{I(J-1)} = \frac{120116}{33} = 3639.879$$

Thus, we know that

$$\text{Bonferroni t-test for Bowtruckle versus Thunderbird} : t_{I(J-1)} = \frac{160 - 155}{\sqrt{3639.879}\sqrt{\frac{2}{12}}} = 0.2030029$$

$$\text{Bonferroni t-test for Bowtruckle versus Niffler} : t_{I(J-1)} = \frac{155 - 40}{\sqrt{3639.879}\sqrt{\frac{2}{12}}} = 4.669066$$

$$\text{Bonferroni t-test for Thunderbird versus Niffler} : t_{I(J-1)} = \frac{160 - 40}{\sqrt{3639.879}\sqrt{\frac{2}{12}}} = 4.872069$$

The degree of freedom of t-test is $I(J-1) = 33$, and the number of tests is (3 chooses $2 = 3$).The test statistic is between 2.499 and 2.536. Thus, we reject the null hypothesis for the pair of Bowtruckle-versus-Niffler, and the pair of Thunderbird-versus-Niffler, but we fail to reject for pair of Bowtruckle-versus-Thunderbird.

Problem 2. (20 points total, 5 points each part) Short answer

The following four questions may be answered independently of each other.

1. You are given a probability density function

$$f_{T|\theta}(t) = \frac{1}{\pi}(1 - \theta \sin t), \quad t \in [-\frac{\pi}{2}, \frac{\pi}{2}]$$

where $\theta \in [-4, 4]$ is unknown, and $T \in [-\frac{\pi}{2}, \frac{\pi}{2}]$.

(a) First find the MoM estimate of θ given observations $t_1, ..., t_n$.

(b) After finding the MoM estimate, we need to perform bootstrap to calculate the 95% confidence interval of the MoM estimate. Here is the R code you need to do such calculation. Please explain whether this is a parametric or non-parametric bootstrap, find the line(s) of code that supports your argument and fully explain your reasoning.

```
# Input Data:  X is a vector of observations corresponding to t_1, ..., t_n
# Self-defined Function:  Calculate_MoM(observed_data) returns the MoM
    estimate from the data vector observed_data
# Self-defined function:  Generate_Data(sample_size, theta) generates
    a vector of size sample_size using the given probability density
    function with the parameter theta
n <- length(X)
B <- 5000
theta_MoM <- Calculate_MoM(X)
# Create an empty vector to record bootstrap estimates
theta_star <- rep(NA, B)
for(i in 1:B){
  X_star <- Generate_Data(n, theta_MoM)
  theta_star[i] <- Calculate_MoM(X_star)
}
r_star <- theta_star - theta_MoM
r_star_upper <- quantile(r_star, 0.975)
r_star_lower <- quantile(r_star, 0.025)
conf_int <- c(theta_MoM - r_star_upper, theta_MoM - r_star_lower)
```

2. In a parallel universe, scientists developed a new screening test to screen people for COVID-19. Each person taking the screening test either has or does not have COVID-19. The screening outcome can be positive (classifying the person as having COVID-19) or negative (classifying the person as not having COVID-19). The screening results for each subject may or may not match the subject's actual status. To decide whether this screening test works or not, you as the Chief Biostatistian, decides to randomly select 82 individuals and perform the screening test for each of them. And here is the summary statistics that you got.

	Individual does not have COVID-19	individual has COVID-19
screening outcome positive	10	31
screening outcome negative	23	18

If the null hypothesis for each individual is that the individual does not have COVID-19, please compute the following from the table: (i) Type I error, (ii) Type II error, (iii) power of the test.
Note: To receive full points, you need to fully specify the formulas you used to do such computation.

3. In a parallel universe, Yutong is graduating in May and decides to start PEARSON, the next million dollar startup based around performing hypothesis tests. For her first employee, she hires Adam to perform 4 of these tests per day. However, Adam doesn't actually know what he's doing, so for each test, he rejects at random independently with probability 1/2. After a period of declining profit, Yutong catches wind of this nonsense and fires Adam in order to hire Fitch. In the first 240 days, Fitch rejects 0 tests 10 times, 1 test 62 times, 2 tests 74 times, 3 tests 59 times, and 4 tests 35 times, and Yutong's startup has now tripled in valuation.

   Is Fitch a very good employee or is he guessing like Adam and PEARSON's success is due to chance? For full credit say which statistical test you are using and indicate the null and alternative hypothesis.

4. The following table gives statistics for the Quantitative Reasoning portion of the GRE for test takers below 18 years old in back-to-back years.

Period	**N**	**Mean**	**SD**
July '15 - June '16	147	158.3	10.3
July '16 - June '17	113	161.6	8.9

   Assume that no test takers took the test in both years and that the test scores of the two age groups are normally distributed from populations with equal variances. Specify the degrees of freedom if necessary and test if the means of the two groups are significantly different at the $\alpha = 0.01$ level. You do not need to report a p-value.

Part (1a): By integration by parts, we know that

$$\int t \sin t \, dt = \sin t - t \cos t + \text{constant}$$

$$\mathbb{E}(t) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{t}{\pi}(1 - t \sin t) \, dt = 0 - \frac{2\theta}{\pi} = -\frac{2\theta}{\pi}$$

Denote $\bar{t} = \frac{t_1 + t_2 + \ldots + t_n}{n}$, we have that $\hat{\theta}_{\text{MoM}} = -\frac{\pi \bar{t}}{2}$

Part (1b): It is parametric bootstrap. The line of code is `X_star <- Generate_Data(n, theta_MoM)`. Because it generates simulated data, by sampling from the model parametrized by $\hat{\theta}_{\text{MoM}}$ as opposed to sampling from the empirical distribution of the data X. Therefore it is a parametric bootstrap.

Part (2a): $\alpha = \frac{n_{11}}{n_{11} + n_{21}} = 10/(10 + 23) = 0.3030303$

Part (2b): $\beta = \frac{n_{22}}{n_{12} + n_{22}} = 18/(31 + 18) = 0.3673469$

Part (2c): Power of the test $= 1 - \frac{n_{22}}{n_{12} + n_{22}} = \frac{n_{12}}{n_{12} + n_{22}} = 0.6326531$

Part (3): We perform a χ^2 goodness of fit test. The null hypothesis is that Fitch rejects according to a Binomial(4,1/2) distribution just like Adam. The alternative is that Fitch isn't rejecting according to this distribution.

If Fitch were rejecting at random like Adam, the probabilities of him rejecting $0, 1, 2, 3,$ and 4 tests would be $1/16, 4/16, 6/16$, $4/16$, and $1/16$. Fitch performed 4 test for 240 days so we should expect him to reject $0, 1, 2, 3,$ and 4 tests on $15, 60, 90, 60,$ and 15 days respectively. Hence, our contingency table will look as follows

Number of Rejections	Observed	Expected
0	10	15
1	62	60
2	74	90
3	59	60
4	35	15

The χ^2 test statistic would be

$$X^2 = \frac{(10 - 15)^2}{15} + \frac{(62 - 60)^2}{60} + \frac{(74 - 90)^2}{90} + \frac{(59 - 60)^2}{60} + \frac{(35 - 15)^2}{15} = 31.3$$

With $5 - 1 = 4$ degrees of freedom, $31.3 > 13.28$ so $p < .01$ and hence we reject the null with strong significance, Fitch is not guessing randomly and he is definitely a good employee!

Part (4):

$$s_p = \sqrt{\frac{146 \times 10.3^2 + 112 \times 8.9^2}{258}} \approx 9.72.$$

The t-statistic for this test is

$$t = \frac{161.6 - 158.3}{s_p \sqrt{\frac{1}{113} + \frac{1}{147}}} \approx 2.71$$

which has 258 degrees of freedom. The degrees of freedom is quite large, so the null t-distribution is a approximately normal. Looking at the t-distribution table we see that as *df* goes to ∞, the critical value of significance at the $\alpha = 0.01$ level for a two-sided test is $t_{.995} = 2.576$. We're well above that threshold and conclude that there is a significant difference between the two groups.

Problem 3. (25 points total, 5 points each part) MLE & Sufficiency & GLRT

Let $X \sim MN(m, p_1, p_2, p_3)$, the multinomial distribution with three categories. We write $X = (X_1, X_2, X_3)$ where X_i is a count of the i^{th} category and p_i is the probability of the i^{th} category. Note that you can treat X as a single sample drawn from $MN(m, p_1, p_2, p_3)$ and thus the sample size in this setting is 1. Suppose that $p_1 = \theta^2$, $p_2 = 2\theta(1-\theta)$, and $p_3 = (1-\theta)^2$, where $0 < \theta < 1$ and m is known.

(a) Express the likelihood function for θ.

(b) Find the MLE estimator $\hat{\theta}_{MLE}$.

(c) Find the asymptotic distribution of $\hat{\theta}_{MLE}$.

(d) Find a minimal sufficient statistic of θ and give your justifications.

(e) One might be skeptical about the structure of $(p_1, p_2, p_3) = (\theta^2, 2\theta(1-\theta), (1-\theta)^2)$ above. We can apply GLRT to test the following:
H_0: $X \sim MN(m, p_1, p_2, p_3)$ where $(p_1, p_2, p_3) = (\theta^2, 2\theta(1-\theta), (1-\theta)^2)$
H_1: $X \sim MN(m, p_1, p_2, p_3)$ where $(p_1, p_2, p_3) \neq (\theta^2, 2\theta(1-\theta), (1-\theta)^2)$
Express the test statistic in terms of (X_1, X_2, X_3) and determine its limiting distribution.

Solution

(a)

$$\begin{aligned} L(\theta) &= Pr(X_1, X_2, X_3; \theta) \\ &= \frac{m!}{X_1!X_2!X_3!} p_1^{X_1} p_2^{X_2} p_3^{X_3} \\ &= \frac{m!}{X_1!X_2!X_3!} 2^{X_2} \theta^{2X_1+X_2} (1-\theta)^{X_2+2X_3} \end{aligned}$$

(b)

$$\frac{\partial \log L(\theta)}{\partial \theta}\Big|_{\theta=\hat{\theta}_{MLE}} = \frac{2X_1+X_2}{\hat{\theta}_{MLE}} - \frac{2X_3+X_2}{1-\hat{\theta}_{MLE}} = 0$$

$$\hat{\theta}_{MLE} = \frac{2X_1+X_2}{2(X_1+X_2+X_3)} = \frac{2X_1+X_2}{2m}$$

(c)

$$\begin{aligned} I(\theta) &= E\left(-\frac{\partial^2 \log L(\theta)}{\partial \theta^2}\right) \\ &= E\left(\frac{2X_1+X_2}{\theta^2} + \frac{2X_3+X_2}{(1-\theta)^2}\right) \\ &= \frac{2m\theta^2 + 2m\theta(1-\theta)}{\theta^2} + \frac{2m(1-\theta)^2 + 2m\theta(1-\theta)}{(1-\theta)^2} \\ &= 2m\left(\frac{1}{\theta} + \frac{1}{1-\theta}\right) \end{aligned}$$

Thus,

$$\hat{\theta}_{MLE} \overset{A}{\sim} N(\theta, \frac{1}{2mn(\frac{1}{\theta} + \frac{1}{1-\theta})})$$

(d) A natural minimal sufficient statistic is $2X_1 + X_2$. Proof:

$$Pr(X_1, X_2, X_3; \theta) = \frac{m!}{X_1!X_2!X_3!} 2^{X_2} \theta^{2X_1+X_2} (1-\theta)^{2m-(2X_1+X_2)} = f(X)g(X,\theta)$$

One natural sufficient statistic is $T(X) = 2X_1 + X_2$ by factorization theorem.

$$\frac{Pr(X_1, X_2, X_3; \theta)}{Pr(Y_1, Y_2, Y_3; \theta)} = h(X,Y)\theta^{(2X_1+X_2)-(2Y_1+Y_2)}(1-\theta)^{(2Y_1+Y_2)-(2X_1+X_2)}$$

Where $h(X,Y)$ is some function not involving θ. As we can judge from the formula above, the ratio does not depend on θ if and only if $2X_1 + X_2 = 2Y_1 + Y_2$. Thus, $T(X) = 2X_1 + X_2$ is minimally sufficient.

(e)

$$\begin{aligned}\Lambda &= \frac{\max_{H_0} Pr(X_1, X_2, X_3)}{\max_{\Omega} Pr(X_1, X_2, X_3)} \\ &= \frac{2^{X_2}\hat{\theta}^{2X_1+X_2}(1-\hat{\theta})^{X_2+2X_3}}{\hat{p}_1^{X_1}\hat{p}_2^{X_2}\hat{p}_3^{X_3}}\end{aligned}$$

Where

$$\hat{\theta} = \hat{\theta}_{MLE} = \frac{2X_1 + X_2}{2(X_1 + X_2 + X_3)} = \frac{2X_1 + X_2}{2m}$$

$$(\hat{p}_1, \hat{p}_2, \hat{p}_3) = (\frac{X_1}{m}, \frac{X_2}{m}, \frac{X_3}{m})$$

Under H_0, the dimension of parameter space is 1 since we can only determine the value of θ. Under Ω, we have dimension of parameter space of 2 since we can determine two values of (p_1, p_2, p_3). Thus, $df = 2 - 1 = 1$. Limiting distribution:

$$-2\log\Lambda \xrightarrow{d} \chi_1^2$$

Problem 4. (10 points total, 2 points each part) Bayesian Statistics

Consider a Bayesian model in which, conditional on unknown parameter $\lambda > 0$, $X_1, ..., X_n$ are iid with likelihood

$$f_{X|\Lambda}(x) = \frac{1}{\lambda} e^{-x/\lambda}$$

for $x > 0$ and the prior distribution is the inverse gamma distribution, with known parameters (a, b) such that

$$f(\lambda) = \frac{b^a}{\Gamma(a)} \lambda^{-a-1} e^{-b/\lambda}$$

for $\lambda > 0$.

(a) Find the posterior distribution for λ conditioning on $X_1, ..., X_n$. You do not need to compute the pdf explicitly, but you must write the correct distribution and its parameters.

(b) Is the prior and likelihood density, a conjugate pair? Explain.

(c) What is the MLE for λ?

(d) Show that the posterior mean can be written as a weighted average of the prior mean and the MLE for λ. The mean of an inverse gamma distribution with parameters a and b is $\frac{b}{a-1}$

(e) What happens to the posterior mean as $n \to \infty$?

Solution (a) Using the law of total probability, we have that

$$f_{\Lambda|X_1=x_1,...,X_n=x_n}(\lambda) = \frac{f_{X_1,...,X_n|\Lambda=\lambda}(x) f_\Lambda(\lambda)}{\int f_{X_1,...,X_n|\Lambda=\lambda}(x) f_\Lambda(\lambda) d\lambda}$$

So for the numerator, we have:

$$\prod_{i=1}^{n} \frac{1}{\lambda} e^{-x_i/\lambda} \frac{b^a}{\Gamma(a)} \lambda^{-a-1} e^{-b/\lambda}$$

$$\frac{b^a}{\Gamma(a)} exp\{\frac{-1}{\lambda}(\sum x_i + b)\} \lambda^{-(n+a)-1}$$

We know the denominator is just a scaling factor, so we know the pdf we are looking for is proportional to the numerator, which when grouped carefully as above, we can see is the inverse gamma distribution with parameters $(n + a, \sum x_i + b)$

(b) Yes, the prior and likelihood density, a conjugate pair because the both the prior and posterior are inverse gamma distributions (with different parameters).

(c) $\bar{X_n}$. take log likelihood and derviative.

(d) Choose weights c_1, c_2 such that

$$\frac{\sum x_i + b}{n + a - 1} = c_1 \frac{b}{a - 1} + c_2 \frac{\sum x_i}{n}$$

.

$$c_1 = \frac{a - 1}{n + a - 1}, c_2 = \frac{n}{n + a - 1}$$

(e) The prior mean goes to zero, and the posterior mean tends toward the MLE.

Problem 5. (25 points total, 5 points each part) Linear Regression

Consider a linear regression $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{2i}x_{3i} + e_i$, where $x_{2i}x_{3i}$ is the product of x_{2i} and x_{3i}. The predictor, x_{3i}, is a binary variable, either 0 or 1 and when $x_{3i} = 1$, we denote the ith observation is in the treatment group; when $x_{3i} = 0$, we denote the ith observation is in the control group. As usual the error e_i follows $i.i.d$ $N(0, \sigma^2)$. After we have some data, we ran a linear regression in R. The regression output is as below:

```
ols_fit = lm(y~x1+x2+x2*x3)
summary(ols_fit)

Call:
lm(formula = y ~ x1 + x2 + x2 * x3)

Residuals:
     Min       1Q   Median       3Q      Max
-2.48890 -0.59790  0.05647  0.59271  2.51871

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.01106    0.22912  -0.048 0.961594
x1           0.98880    0.02204  44.872  < 2e-16 ***
x2           1.05248    0.06845  15.375  < 2e-16 ***
x3           1.10425    0.28091   3.931 0.000161 ***
x2:x3        1.02610    XXXXXXX   XXXXX 3.67e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.9829 on 95 degrees of freedom
Multiple R-squared:  0.9714,    Adjusted R-squared:  0.9702
F-statistic: 805.9 on 4 and 95 DF,  p-value: < 2.2e-16
```

(a) What's the predicted response value of y if $x_1 = 5$, $x_2 = 1$ and $x_3 = 0$?

(b) What's the sample size n? What's the degree of freedom of t distributions(t-value) in the output table?

(c) Given $(X^T X)^{-1}$ as below:

$$\begin{bmatrix} 0.054335785 & XXXXXX & XXXXXX & -0.0391027885 & 0.0102493987 \\ -0.002927820 & XXXXXX & XXXXXX & 0.0003129854 & -0.0001385643 \\ -0.009798354 & XXXXXX & XXXXXX & 0.0094802532 & -0.0048591864 \\ -0.039102788 & XXXXXX & XXXXXX & 0.0816732747 & -0.0216209556 \\ 0.010249399 & XXXXXX & XXXXXX & -0.0216209556 & 0.0112598306 \end{bmatrix}$$

What's the estimator for $corr(\hat{\beta}_0, \hat{\beta}_4)$, the correlation of $\hat{\beta}_0$ and $\hat{\beta}_4$?

(d) What's the value of $MSS = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$, the model sum of squares?

(e) If $\beta_4 = 0$, what's the interpretation of β_3? Hint: Calculate $E(y_i|x_{1i}, x_{2i}, x_{3i} = 1) - E(y_i|x_{1i}, x_{2i}, x_{3i} = 0)$.

Solutions

(a) $\hat{y}_i = -0.01106 + 0.98880 * 5 + 1.0524 = 5.98534$.

(b) $n = 95 + 5 = 100$, $df = 95$.

(c) $\widehat{SD}(\hat{\beta}_4) = 0.9829 * \sqrt{0.011259830} = 0.1043$ and $\widehat{SD}(\hat{\beta}_0) = 0.9829 * \sqrt{0.054335785} = 0.22912$.

$$\widehat{corr}(\hat{\beta}_0, \hat{\beta}_4) = \frac{\widehat{cov}(\hat{\beta}_0, \hat{\beta}_4)}{\widehat{SD}(\hat{\beta}_0)\widehat{SD}(\hat{\beta}_4)} = \frac{0.010249399 * 0.9829^2}{0.22912 * 0.1043} = 0.414$$

(d) We know

$$F = \frac{\frac{MSS}{4}}{\frac{RSS}{95}}$$

and and from the output $F = 805.9$ and $RSS = 0.983^2 * 95 = 91.80$. It follows that $MSS = 3114.93$.

Altererrnatively,

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + 95\hat{\sigma}^2}$$

Thus,

$$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \frac{95\hat{\sigma}^2 R^2}{1 - R^2} = 3117$$

(e) If $\beta_4 = 0$ then $\beta_3 = E(y_i|x_{1i}, x_{2i}, x_{3i} = 1) - E(y_i|x_{1i}, x_{2i}, x_{3i} = 0)$. Thus, if $\beta_4 = 0$, then β_3 is the expectation difference of y between treatment group and control group, holding other covariates constant.

Discrete

name and range	$P(k) = P(X = k)$ for $k \in$ range	mean	variance
uniform on $\{a, a+1, \ldots, b\}$	$\frac{1}{b-a+1}$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$
Bernoulli (p) on $\{0, 1\}$	$P(1) = p$; $P(0) = 1 - p$	p	$p(1-p)$
binomial (n, p) on $\{0, 1, \ldots, n\}$	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
Poisson (μ) on $\{0, 1, 2, \ldots\}$	$\frac{e^{-\mu}\mu^k}{k!}$	μ	μ
hypergeometric (n, N, G) on $\{0, \ldots, n\}$	$\frac{\binom{G}{k}\binom{N-G}{n-k}}{\binom{N}{n}}$	$\frac{nG}{N}$	$n\left(\frac{G}{N}\right)\left(\frac{N-G}{N}\right)\left(\frac{N-n}{N-1}\right)$
geometric (p) on $\{1, 2, 3 \ldots\}$	$(1-p)^{k-1}p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
geometric (p) on $\{0, 1, 2 \ldots\}$	$(1-p)^k p$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
negative binomial (r, p) on $\{0, 1, 2, \ldots\}$	$\binom{k+r-1}{r-1} p^r (1-p)^k$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$

Continuous

† undefined.

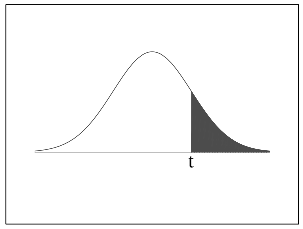
name	range	density $f(x)$ for $x \in$ range	c.d.f. $F(x)$ for $x \in$ range	Mean	Variance
uniform (a, b)	(a, b)	$\frac{1}{b-a}$	$\frac{x-a}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
normal $(0, 1)$	$(-\infty, \infty)$	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$	$\Phi(x)$	0	1
normal (μ, σ^2)	$(-\infty, \infty)$	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}$	$\Phi\left(\frac{x-\mu}{\sigma}\right)$	μ	σ^2
exponential (λ) = gamma $(1, \lambda)$	$(0, \infty)$	$\lambda e^{-\lambda x}$	$1 - e^{-\lambda x}$	$1/\lambda$	$1/\lambda^2$
gamma (r, λ)	$(0, \infty)$	$\Gamma(r)^{-1}\lambda^r x^{r-1} e^{-\lambda x}$	$1 - e^{-\lambda x}\sum_{k=0}^{r-1}\frac{(\lambda x)^k}{k!}$ for integer r	r/λ	r/λ^2
chi-square (n) =gamma $(\frac{n}{2}, \frac{1}{2})$	$(0, \infty)$	$\Gamma(\frac{n}{2})^{-1}(\frac{1}{2})^{\frac{n}{2}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}$	as above for $\lambda = \frac{1}{2}$, $r = \frac{n}{2}$ if n is even	n	$2n$
Rayleigh	$(0, \infty)$	$xe^{-\frac{1}{2}x^2}$	$1 - e^{-\frac{1}{2}x^2}$	$\sqrt{\frac{\pi}{2}}$	$\frac{4-\pi}{2}$
beta (r, s)	$(0, 1)$	$\frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} x^{r-1}(1-x)^{s-1}$	see Exercise 4.6.5 for integer r and s	$\frac{r}{r+s}$	$\frac{rs}{(r+s)^2(r+s+1)}$
Cauchy	$(-\infty, \infty)$	$\frac{1}{\pi(1+x^2)}$	$\frac{1}{2} + \frac{1}{\pi}\arctan(x)$	†	†
arcsine =beta $(1/2, 1/2)$	$(0, 1)$	$\frac{1}{\pi\sqrt{x(1-x)}}$	$\frac{2}{\pi}\arcsin(\sqrt{x})$	$\frac{1}{2}$	$\frac{1}{8}$

TABLE 11 Bonferroni Multipliers for 95% Confidence Intervals

The values given in the table are $t_{df,0.025/k}$ where k is the number of tests.

df	NUMBER OF TESTS 1	2	3	4	5	6	8	10	15	20
1	12.706	25.452	38.185	50.923	63.657	76.384	101.856	127.321	190.946	254.647
2	4.303	6.205	7.648	8.860	9.925	10.885	12.590	14.089	17.275	19.963
3	3.182	4.177	4.857	5.392	5.841	6.231	6.895	7.453	8.575	9.465
4	2.776	3.495	3.961	4.315	4.604	4.851	5.261	5.598	6.254	6.758
5	2.571	3.163	3.534	3.810	4.032	4.219	4.526	4.773	5.247	5.604
6	2.447	2.969	3.287	3.521	3.707	3.863	4.115	4.317	4.698	4.981
7	2.365	2.841	3.128	3.335	3.499	3.636	3.855	4.029	4.355	4.595
8	2.306	2.752	3.016	3.206	3.355	3.479	3.677	3.833	4.122	4.334
9	2.262	2.685	2.933	3.111	3.250	3.364	3.547	3.690	3.954	4.146
10	2.228	2.634	2.870	3.038	3.169	3.277	3.448	3.581	3.827	4.005
11	2.201	2.593	2.820	2.981	3.106	3.208	3.370	3.497	3.728	3.895
12	2.179	2.560	2.779	2.934	3.055	3.153	3.308	3.428	3.649	3.807
13	2.160	2.533	2.746	2.896	3.012	3.107	3.256	3.372	3.584	3.735
14	2.145	2.510	2.718	2.864	2.977	3.069	3.214	3.326	3.529	3.675
15	2.131	2.490	2.694	2.837	2.947	3.036	3.177	3.286	3.484	3.624
16	2.120	2.473	2.673	2.813	2.921	3.008	3.146	3.252	3.444	3.581
17	2.110	2.458	2.655	2.793	2.898	2.984	3.119	3.222	3.410	3.543
18	2.101	2.445	2.639	2.775	2.878	2.963	3.095	3.197	3.380	3.510
19	2.093	2.433	2.625	2.759	2.861	2.944	3.074	3.174	3.354	3.481
20	2.086	2.423	2.613	2.744	2.845	2.927	3.055	3.153	3.331	3.455
25	2.060	2.385	2.566	2.692	2.787	2.865	2.986	3.078	3.244	3.361
30	2.042	2.360	2.536	2.657	2.750	2.825	2.941	3.030	3.189	3.300
40	2.021	2.329	2.499	2.616	2.704	2.776	2.887	2.971	3.122	3.227
50	2.009	2.311	2.477	2.591	2.678	2.747	2.855	2.937	3.083	3.184
60	2.000	2.299	2.463	2.575	2.660	2.729	2.834	2.915	3.057	3.156
70	1.994	2.291	2.453	2.564	2.648	2.715	2.820	2.899	3.039	3.137
80	1.990	2.284	2.445	2.555	2.639	2.705	2.809	2.887	3.026	3.122
100	1.984	2.276	2.435	2.544	2.626	2.692	2.793	2.871	3.007	3.102
140	1.977	2.266	2.423	2.530	2.611	2.676	2.776	2.852	2.986	3.079
1000	1.962	2.245	2.398	2.502	2.581	2.643	2.740	2.813	2.942	3.031
∞	1.960	2.241	2.394	2.498	2.576	2.638	2.734	2.807	2.935	3.023

t-Distribution Table



The shaded area is equal to α for $t = t_\alpha$.

df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
32	1.309	1.694	2.037	2.449	2.738
34	1.307	1.691	2.032	2.441	2.728
36	1.306	1.688	2.028	2.434	2.719
38	1.304	1.686	2.024	2.429	2.712
∞	1.282	1.645	1.960	2.326	2.576

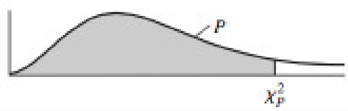
Standard Normal Distribution Table (Right-Tail Probabilities)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002

F Values for $\alpha = 0.05$

	d_1								
d_2	1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.3	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
inf	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

TABLE 3 Percentiles of the χ^2 Distribution—Values of χ^2_P Corresponding to P



df	$\chi^2_{.005}$	$\chi^2_{.01}$	$\chi^2_{.025}$	$\chi^2_{.05}$	$\chi^2_{.10}$	$\chi^2_{.90}$	$\chi^2_{.95}$	$\chi^2_{.975}$	$\chi^2_{.99}$	$\chi^2_{.995}$
1	.000039	.00016	.00098	.0039	.0158	2.71	3.84	5.02	6.63	7.88
2	.0100	.0201	.0506	.1026	.2107	4.61	5.99	7.38	9.21	10.60
3	.0717	.115	.216	.352	.584	6.25	7.81	9.35	11.34	12.84
4	.207	.297	.484	.711	1.064	7.78	9.49	11.14	13.28	14.86
5	.412	.554	.831	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	.676	.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	.989	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
120	83.85	86.92	91.58	95.70	100.62	140.23	146.57	152.21	158.95	163.64