1. There are $\binom{50}{4}$ total number of ways to pick 4 tickets among the 50. Let X denote the random variable for number of tickets won.

(a) If you want to win all 3 prizes with your 4 tickets, you must have 3 tickets chosen among those 3 prizes, and $4-3=1$ ticket chosen among the rest 47 non-prize tickets. That is, $\binom{3}{3} * \binom{47}{1}$ cases can win all 3 prizes.

The probability is then $P(X=3) = \binom{3}{3} * \binom{47}{1} / \binom{50}{4} = 1/4900 = 0.0002$

(b) $\binom{3}{2} * \binom{47}{2}$ cases can win exactly 2 prizes. $P(X=2) = 0.0141$. The largest possible number of prizes we can win is 3, thus:

$P(X \geq 2) = P(X=3) + P(X=2) = 1/70 = 0.0143$

2. The number of recoveries X follows a binomial distribution with $n=20$ and $p=0.8$.

(a) $E(X) = 20 * 0.8 = 16$

(b) $Var(X) = 20 * 0.8 * 0.2 = 3.2$

$sd(X) = \sqrt{3.2} = 1.7889$

3. Let X denote the random variable for the number of interviews until first applicant with advanced training: X follows a geometric distribution with $p = 0.3$.

$P(X=5) = 0.7^{(5-1)} * 0.3 = 0.0720$

4. Suppose the number of customers arrive follows a Poisson distribution with parameter λ. Then $\frac{\lambda^0 e^{-\lambda}}{0!} = \frac{\lambda^1 e^{-\lambda}}{1!}$: $\lambda = 1$.

(a) $P(X=2) = \frac{\lambda^2 e^{-\lambda}}{2!} = 0.1839$

(b) $P(X>1) = 1 - P(X=1) - P(X=0) = 0.2642$

5. Here,

(a) $P(Y < 0.5) = \int_0^{0.5} (1.5y^2 + y)dy = 3/16 = 0.1875$

(b) $E(Y) = \int_0^1 y(1.5y^2 + y)dy = \frac{17}{24}$

(c) $E(Y^2) = \int_0^1 y^2(1.5y^2 + y)dy = \frac{11}{20}$

$Var(Y) = E(Y^2) - E(Y)^2 = 0.0487$

(a) $f_2(y_2) = 2y_2, 0 \leq y_2 \leq 1$

(b)

$$
\begin{aligned}
P(Y_2 \geq 3/4) &= \int_{3/4}^1 f_2(y_2)dy_2 \\
&= \frac{7}{16} \\
P(Y_1 \leq 1/2, Y_2 \geq 3/4) &= \int_{3/4}^1 \int_0^{1/2} f(y_1, y_2)dy_1 dy_2 \\
&= \frac{7}{64} \\
P(Y_1 \leq 1/2 | Y_2 \geq 3/4) &= \frac{P(Y_1 \leq 1/2, Y_2 \geq 3/4)}{P(Y_2 \geq 3/4)} \\
&= \frac{1}{4}
\end{aligned}
$$

(c) $f_{1|2}(y_1|y_2) = \frac{f(y_1,y_2)}{f_2(y_2)} = 2y_1, 0 \leq y_1 \leq 1, \forall y_2$

(d)

$$
\begin{aligned}
E(Y_1 Y_2) &= \int_0^1 \int_0^1 y_1 y_2 f(y_1, y_2) dy_1 dy_2 \\
&= 4 \int_0^1 \int_0^1 y_1^2 y_2^2 dy_1 dy_2 = \frac{4}{9} \\
E(Y_1) &= \int_0^1 y_1 f_1(y_1) dy_1 = \frac{2}{3} \\
E(Y_2) &= \frac{2}{3} \\
Cov(y_1, y_2) &= E(Y_1 Y_2) - E(Y_1)E(Y_2) = 0
\end{aligned}
$$

Y_1, Y_2 are independent (since marginal densities are the same as conditional densities), thus the covariance is zero.

1

(a) (a) 0.9243, (b) 2.3263, (c) 0

(b) $P(-2 < X < 5.74) = P(\frac{-2-1}{2} < Z < \frac{5.74-1}{2}) = 0.9243$

3

(a)

$$
\begin{aligned}
F_X(x) &= P(X < x) \\
&= P(-(1/3)\log Y < x) = P(Y > e^{-3x}) \\
&= 1 - F_Y(e^{-3x}) = 1 - e^{-3x} \\
f_X(x) &= 3e^{-3x}, x \geq 0
\end{aligned}
$$

(b)

$$
E(e^X) = \int_0^\infty e^x * 3e^{-3x} dx = \frac{3}{2}
$$

$$
\begin{aligned}
F(y) &= 1 - e^{-y/\beta} \\
F_{(n)}(y) &= F(y)^n = (1 - e^{-y/\beta})^n \\
P(Y_{(n)} \geq 4) &= 1 - F_{(n)}(4) \\
&= 1 - (1 - e^{-4/2})^5 = 0.5167
\end{aligned}
$$

5

$$
\begin{aligned}
P(|\bar{X} - \mu| < 1) &= P(\frac{|\bar{X} - \mu|}{\sigma/\sqrt{n}} < \frac{1}{\sigma/\sqrt{n}}) \\
&= P(|Z| < \frac{1}{2/\sqrt{25}}) \\
&= P(|Z| < \frac{5}{2}) \\
&= 0.9876
\end{aligned}
$$

8.62

(a) $24.8-21.3 \pm Z_{0.995} * \sqrt{\frac{7.1^2}{34}+\frac{8.1^2}{41}} = 3.5 \pm 4.52 = (-1.02, 8.02)$

(b) We are 99% confident that the difference in mean molt time for normal males versus those split from their mates is between $(-1.02, 8.02)$.

8.70

(a) $Z_{0.975}\sqrt{\frac{p(1-p)}{n}} = 0.05, p = 0.9$: $n = 138.29 = 139$

(b) $Z_{0.975}\sqrt{\frac{p(1-p)}{n}} = 0.05, p = 0.5$: $n = 384.15 = 385$

8.80

$26.6 \pm T_{21-1,0.975} * \frac{7.4}{\sqrt{21}} = 26.6 \pm 3.37 = (23.23, 29.97)$

8.90

(a)

$$S_p^2 = \frac{(15-1)*42^2+(15-1)*45^2}{15+15-2} = 1894.5$$

$$\begin{aligned} CI &= 446-534 \pm t_{28,0.975} * \sqrt{(\frac{1}{15}+\frac{1}{15})S_p} \\ &= -88 \pm 32.55 \\ &= (-120.55, -55.45) \end{aligned}$$

(b)

$$S_p^2 = \frac{(15-1)*57^2+(15-1)*52^2}{15+15-2} = 2976.5$$

$$\begin{aligned} CI &= 548-517 \pm t_{28,0.975} * \sqrt{(\frac{1}{15}+\frac{1}{15})S_p} \\ &= -31 \pm 40.8 \\ &= (-9.8, 71.8) \end{aligned}$$

10.18

$H_0 : \mu = 13.2$ vs. $H_a : \mu < 13.2$. $Z = \frac{12.2-13.2}{2.5/\sqrt{40}} = -2.53 < Z_{0.01} = -2.326$.

Reject the null hypothesis, there is evidence that the company is paying less than average.

10.21

$H_0 : \mu_1 = \mu_2$ vs. $H_a : \mu_1 \neq \mu_2$

$Z = \frac{1.65-1.43}{\sqrt{0.26^2/30+0.22^2/35}} = 3.648 > Z_{0.995} = 2.576$

Reject the null hypothesis, the soils don't have equal mean shear strengths.

10.33

Define p_1: proportion of republicans; p_2: proportion of democrats.

$H_0 : p_1 = p_2$ vs. $H_a : p_1 > p_2$

$\hat{p}_1 = 0.23, \hat{p}_2 = 0.17, \hat{\sigma}_1^2 = 0.23(1-0.23), \hat{\sigma}_2^2 = 0.17(1-0.17)$

$Z = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{\sigma}_1^2/n_1+\hat{\sigma}_2^2/n_2}} = 1.504 < Z_{0.95} = 1.645$

We fail to reject the null hypothesis, under 95% confidence level there is no evidence that proportion for republicans are higher.

10.40

$H_0 : p_1 = p_2$ vs. $H_a : p_1 > p_2$

$\beta = 0.2$, minimum detection $= 0.1$: $\frac{\hat{p}_1-\hat{p}_2-0.1}{\sqrt{p_1(1-p_1)/n+p_2(1-p_2)/n}} = Z_{0.2} = -0.842$

$\alpha = 0.05$: $\frac{\hat{p}_1-\hat{p}_2}{\sqrt{p_1(1-p_1)/n+p_2(1-p_2)/n}} = Z_{0.95} = 1.645$

Thus $\frac{0.1}{\sqrt{p_1(1-p_1)/n+p_2(1-p_2)/n}} = 1.645+0.842$

Plugging in $p_1 = p_2 = 0.5$: $n = 308.76 = 309$

10.43

(a) $H_0 : \mu_1 = \mu_2$ vs. $H_a : \mu_1 > \mu_2$

$Z = \frac{32.19-31.68}{\sqrt{4.34^2/37+4.56^2/37}} = 0.49 < Z_{0.95} = 1.645$

Do not reject the null hypothesis, where is no difference.

(b) $\bar{Y}_1 - \bar{Y}_2 > Z_{0.95}\sqrt{4.34^2/37+4.56^2/37} = 1.702$

Rejection region: $(1.702, \infty)$

$\beta = P(\bar{Y}_1 - \bar{Y}_2 \leq 1.702 | \mu_1 - \mu_2 = 3) = P(Z \leq \frac{1.702-3}{\sqrt{4.34^2/37+4.56^2/37}}) = 0.105$

10.17

(a) $H_0 : \mu_1 = \mu_2$ vs. $H_a : \mu_1 > \mu_2$

(b) Reject if $Z > 2.326$, where Z is the test statistic

(c) $Z = \frac{9017-5853}{\sqrt{7162^2/130+1961^2/80}} = 4.756$

(d) Reject H_0, there is sufficient evidence to indicate that the average number of meters per week spent practicing breaststroke is greater for exclusive breaststrokers than it is for those swimming individual medley

(e) Two groups have very different sample means.

8.95

$$\begin{aligned} s^2 &= 0.503 \\ \frac{(6-1)s^2}{\sigma^2} &\sim \chi_5^2 \\ \chi^2_{5,0.05} &= 1.145 \\ \chi^2_{5,0.95} &= 11.071 \\ \sigma^2 &\in [0.227, 2.196] \end{aligned}$$

We are 90% confident that σ^2 is within $[0.227, 2.196]$.

9.19

$$\begin{aligned} E(\bar{Y}) &= E(Y_1) \\ &= \int_0^1 y*\theta y^{\theta-1}dy \\ &= \frac{\theta}{\theta+1} \\ E(Y_1^2) &= \frac{\theta}{\theta+2} \\ Var(\bar{Y}) &= Var(Y_1)/n \\ &= [E(Y_1^2)-E(Y_1)^2]/n \\ &= \frac{\theta}{n(\theta+1)^2(\theta+2)} \end{aligned}$$

Since (1) $E(\bar{Y}) = \frac{\theta}{\theta+1}$ and (2) $Var(\bar{Y}) \to 0$ as $n \to \infty$, $\bar{Y}$ is a consistent estimator of $\frac{\theta}{\theta+1}$.

9.37

Likelihood $l(p) = p^{\sum X_i}(1-p)^{n-\sum X_i} = g(\sum X_i, p)*h(\underline{X})$, where $g(\sum X_i, p) = p^{\sum X_i}(1-p)^{n-\sum X_i}, h(\underline{X}) = 1$.

Thus $\sum X_i$ is sufficient for p.

9.63

(a)

$$\begin{aligned} F(y|\theta) &= (y/\theta)^3, 0 \leq y \leq \theta \\ F_{(n)}(y|\theta) &= (\frac{y}{\theta})^{3n} \\ f_{(n)}(y|\theta) &= \frac{3ny^{3n-1}}{\theta^{3n}}, 0 \leq y \leq \theta \end{aligned}$$

(b)

$$\begin{aligned} E(Y_{(n)}) &= \frac{3n}{3n+1}\theta \\ E(\frac{3n+1}{3n}Y_{(n)}) &= \theta \end{aligned}$$

$\frac{3n+1}{3n}Y_{(n)}$ is unbiased and sufficient of θ: MVUE.

8.6 Suppose that $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$, $V(\hat{\theta}_1) = \sigma_1^2$, $V(\hat{\theta}_2) = \sigma_2^2$. Consider estimator $\hat{\theta}_3 = a\hat{\theta}_1 + (1-a)\hat{\theta}_2$

Show that $\hat{\theta}_3$ is an unbiased estimator for θ

(a) $E(\hat{\theta}_3) = aE(\hat{\theta}_1) + (1-a)E(\hat{\theta}_2) = \theta(a+1-a) = \theta$

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are indepen, how should the constant a be chosen in order to maximize variance of $\hat{\theta}_3$

(b) $V(\hat{\theta}_3) = V(a\hat{\theta}_1 + (1-a)\hat{\theta}_2) = a^2\sigma_1^2 + (1-a)^2\sigma_2^2 = (\sigma_1^2+\sigma_2^2)a^2 - 2\sigma_2^2 a + \sigma_2^2$

Minimize: $a = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$

8.8

$f(y) = \begin{cases} (1/\theta)e^{-y/\theta}, & y>0 \\ 0 & \text{otherwise} \end{cases}$

$\hat{\theta}_1 = Y_1$, $\hat{\theta}_2 = (Y_1+Y_2)/2$, $\hat{\theta}_3 = (Y_1+2Y_2)/3$, $\hat{\theta}_4 = \min(Y_1, Y_2, Y_3)$, $\hat{\theta}_5 = \bar{Y}$

Which estimators are unbiased?

(a) $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_5$ are unbiased.

Among unbiased, which has smallest variance?

(b) $V(\hat{\theta}_1) = \theta^2, V(\hat{\theta}_2) = \theta^2/2, V(\hat{\theta}_3) = 5\theta^2/9, V(\hat{\theta}_5) = \theta^2/3$

It's $\hat{\theta}_5$

8.12

The reading on voltage meter connected to test circuit is uniformly dist. over interval $(\theta, \theta+1)$, where θ is true but unknown voltage of the

Show that $\bar{Y}$ is a biased estimator of θ and find the bias

(a) $E(\bar{Y}) = \theta + 0.5$: $B(\bar{Y}) = E(\bar{Y}) - \theta = 0.5$

Find function of $\bar{Y}$ that is an unbiased estimator of θ

(b) Unbiased estimator is $\bar{Y} - 0.5$

Find $MSE(\bar{Y})$ when $\bar{Y}$ is used as an estimator of θ

(c) $V(\bar{Y}) = V(Y_1)/n = \frac{1}{12n}$: $MSE = \frac{1}{12n} + 0.25$

9.81 Suppose that $Y_1, \ldots Y_n$ denote random sample from an exponentially distributed population with mean θ. Find MLE of the population variance θ^2

The log-likelihood:

$$l(\theta) = -n\log\theta - \frac{1}{\theta}\sum y_i$$
$$l'(\hat{\theta}) = 0$$
$$\hat{\theta} = \bar{Y}$$

$\bar{Y}$ is the MLE for θ. By invariance property of MLE, $\bar{Y}^2$ is the MLE for θ^2.

9.97

The geometric probability mass function is given by $p(y \mid p) = p(1-p)^{y-1}$, $y = 1, 2, 3 \ldots$ A random sample of size n is taken from a population with a geometric distribution

Find the method of moments estimator for p

(a) $E(Y) = 1/p$: $\hat{p} = 1/m_1' = \frac{1}{\bar{Y}}$

(b) Find the MLE for p

$$l(p) = p^n(1-p)^{\sum y_i - n}$$
$$l'(p) = np^{n-1}(1-p)^{\sum y_i - n} - (\sum y_i - n)p^n(1-p)^{\sum y_i - n - 1}$$
$$= p^{n-1}(1-p)^{\sum y_i - n - 1}[n(1-p) - (\sum y_i - n)p]$$
$$= 0$$
$$\hat{p} = \frac{1}{\bar{Y}}$$