RWTH BUSINESS SCHOOL

Mathematics & Statistics
M.Sc. Data Analytics and Decision Science

Prof. Dr. Thomas S. Lontzek

AACSB ACCREDITED

BUSINESS SCHOOL | RWTH AACHEN UNIVERSITY

Problem 1 – Z Scores

Given that z is a standard normal random variable, compute the following probabilities.

a. $P(z \leq 1)$

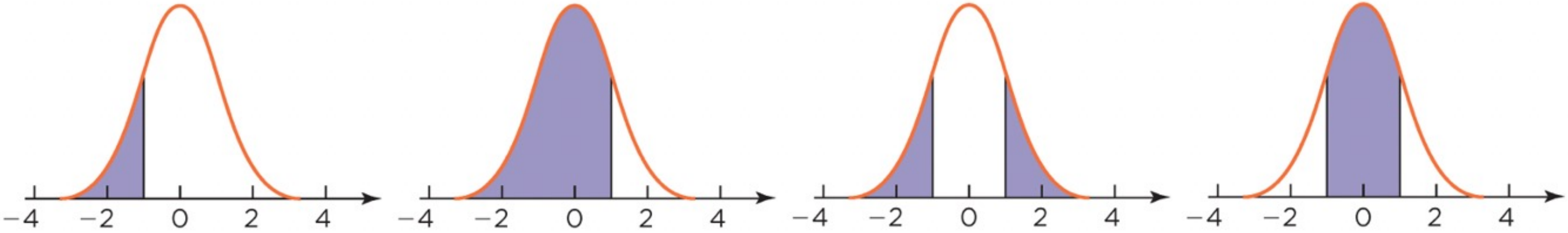
Hint: Use NORM.DIST in Excel or the z Tables on next slide

b. $P(z \geq -1)$

c. $P(-3 < z \leq 0)$

Problem 1 – Z Scores

Recap:



z	$P(Z \leq -z)$	$P(Z \leq z)$	$P(\lvert Z \rvert > z)$	$P(\lvert Z \rvert \leq z)$
1	0.1587	0.8413	0.3173	**0.6827**
2	0.02275	0.97725	0.04550	**0.95450**
3	0.00135	0.99865	0.00270	**0.99730**

Problem 1 – Z Scores (Solution)

Given that z is a standard normal random variable, compute the following probabilities.

a. $P(z \leq 1)$

$$P(z \leq 1.0) = .8413 \ = \text{NORM.S.DIST}(1,\text{TRUE})$$

Hint: Use NORM.DIST in Excel or the z Tables on next slide

b. $P(z \geq -1)$

$$P(z \geq -1) = 1 - P(z < -1) = 1 - .1587 = .8413 = 1 - \text{NORM.S.DIST}(-1,\text{TRUE})$$

c. $P(-3 < z \leq 0)$

$$P(-3 < z \leq 0) = P(z \leq 0) - P(z \leq -3) = .5000 - .0013 = .4987$$

$$\text{OR} = \text{NORM.S.DIST}(0,\text{TRUE}) - \text{NORM.S.DIST}(-3,\text{TRUE})$$

Problem 2 – Amazon Alexa App Download

Alexa is the popular virtual assistant developed by Amazon. Alexa interacts with users using artificial intelligence and voice recognition. It can be used to perform daily tasks such as making to-do lists, reporting the news and weather, and interacting with other smart devices in the home. In 2018, the Amazon Alexa app was downloaded some 2,800 times per day from the Google Play store. Assume that the number of downloads per day of the Amazon Alexa app is normally distributed with a mean of 2,800 and standard deviation of 860.

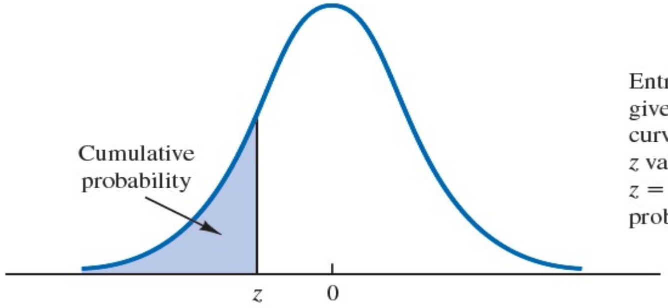
Hint: Use NORM.DIST in Excel or the z Tables on next slide

a. What is the probability there are 2,000 or fewer downloads of Amazon Alexa in a day?

b. What is the probability there are between 1,500 and 2,500 downloads of Amazon Alexa in a day?

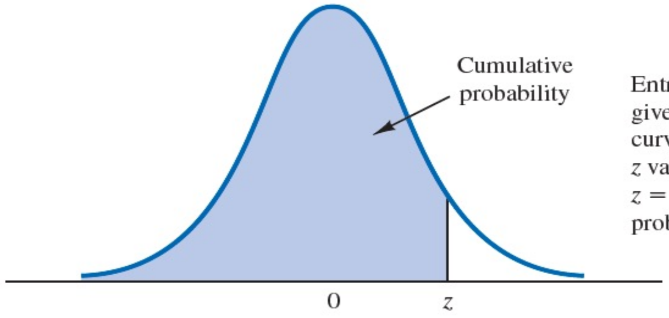
c. What is the probability there are more than 3,000 downloads of Amazon Alexa in a day?

d. Assume that Google has designed its servers so there is probability .01 that the number of Amazon Alexa app downloads in a day exceeds the servers' capacity and more servers have to be brought online. How many Amazon Alexa app downloads per day are Google's servers designed to handle?



Entries in the table give the area under the curve to the left of the z value. For example, for $z = -.85$, the cumulative probability is .1977.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641



Entries in the table give the area under the curve to the left of the z value. For example, for $z = 1.25$, the cumulative probability is .8944.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

Problem 2 – Amazon Alexa App Download (Solution)

a. What is the probability there are 2,000 or fewer downloads of Amazon Alexa in a day?

$\mu = 2800$ and $\sigma = 860$

$$z = \frac{x-\mu}{\sigma} = \frac{2000-2800}{860} = -.93$$

$P(x \leq 2000) = P(z \leq -.93) = .1762$

Using Excel: NORM.DIST(2000, 2800, 860, TRUE) = .1761

b. What is the probability there are between 1,500 and 2,500 downloads of Amazon Alexa in a day?

For $x = 2500$,

$$z = \frac{x-\mu}{\sigma} = \frac{2500-2800}{860} = -.35$$

For $x = 1500$,

$$z = \frac{x-\mu}{\sigma} = \frac{1500-2800}{860} = -1.51$$

$$P(1500 \leq x \leq 2500) = P(z \leq -.35) - P(z \leq -1.51) = .3632 - .0655 = .2977$$

Using Excel: NORM.DIST(2500, 2800, 860, TRUE) – NORM.DIST(1500, 2800, 860, TRUE) = .2983

Problem 2 – Amazon Alexa App Download (Solution)

c. What is the probability there are more than 3,000 downloads of Amazon Alexa in a day?

For $x = 3000$,

$$z = \frac{x-\mu}{\sigma} = \frac{3000-2800}{860} = .23$$

$$P(x > 3000) = 1 - P(z \leq .23) = 1 - .5910 = .4090$$

Using Excel: 1 – NORM.DIST(3000, 2800, 860, TRUE) = .4081

d. Assume that Google has designed its servers so there is probability .01 that the number of Amazon Alexa app downloads in a day exceeds the servers' capacity and more servers have to be brought online. How many Amazon Alexa app downloads per day are Google's servers designed to handle?

The z value for 1 – .01 = .99 is 2.33. Therefore, Google must design their servers to handle $\mu + z\sigma = 2800 + 2.33(860) = 4803.8$ (or about 4804) downloads per day.

Using Excel: NORM.INV(.99, 2800, 860) = 4800.7 (or about 4801) downloads per day

Problem 3 – University Costs

After deducting grants based on need, the average cost to attend the University of Southern California (USC) is $27,175. Assume the population standard deviation is $7400. Suppose that a random sample of 60 USC students will be taken from this population.

a. What is the value of the standard error of the mean?

b. What is the probability that the sample mean will be more than $27,175?

c. What is the probability that the sample mean will be within $1000 of the population mean?

d. How would the probability in part (c) change if the sample size were increased to 100?

Problem 3 – University Costs (Solution)

After deducting grants based on need, the average cost to attend the University of Southern California (USC) is $27,175. Assume the population standard deviation is $7400. Suppose that a random sample of 60 USC students will be taken from this population.

$$\mu = 27{,}175 \quad \sigma = 7400$$

a. What is the value of the standard error of the mean?

$$\sigma_{\bar{x}} = 7400 / \sqrt{60} = 955$$

b. What is the probability that the sample mean will be more than $27,175?

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{0}{955} = 0$$

$P(\bar{x} > 27{,}175) = P(z > 0) = .50$

Using Excel: 1-NORM.DIST(27175,27175,7400/SQRT(60),TRUE) = .5000

Note: This could have been answered easily without any calculations; 27,175 is the expected value of the sampling distribution of $\bar{x}$.

BUSINESS SCHOOL | RWTHAACHEN UNIVERSITY

Problem 3 – University Costs (Solution)

After deducting grants based on need, the average cost to attend the University of Southern California (USC) is \$27,175. Assume the population standard deviation is \$7400. Suppose that a random sample of 60 USC students will be taken from this population.

c. What is the probability that the sample mean will be within \$1000 of the population mean?

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{1000}{955} = 1.05 \qquad P(z \leq 1.05) = .8531$$

$$P(z < -1.05) = .1469$$

$$P(26{,}175 \leq \bar{x} \leq 28{,}175) = P(-1.05 \leq z \leq 1.05) = .8531 - .1469 = .7062$$

Using Excel: NORM.DIST(28175,27175,7400/SQRT(60),TRUE) –
NORM.DIST(26175,27175,7400/SQRT(60),TRUE) = .7048

Problem 3 – University Costs (Solution)

After deducting grants based on need, the average cost to attend the University of Southern California (USC) is \$27,175. Assume the population standard deviation is \$7400. Suppose that a random sample of 60 USC students will be taken from this population.

d. How would the probability in part (c) change if the sample size were increased to 100?

$$\sigma_{\bar{x}} = 7400 / \sqrt{100} = 740$$

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{1000}{740} = 1.35 \qquad P(z \leq 1.35) = .9115$$

$$P(z < -1.35) = .0885$$

$$P(26{,}175 \leq \bar{x} \leq 28{,}175) = P(-1.35 \leq z \leq 1.35) = .9115 - .0885 = .8230$$

Problem 4 – Household Grocery Expenditures

The Food Marketing Institute shows that 17% of households spend more than $100 per week on groceries. Assume the population proportion is p=0.17 and a simple random sample of 800 households will be selected from the population.

a. Show the sampling distribution of p, the sample proportion of households spending more than $100 per week on groceries.

b. What is the probability that the sample proportion will be within 6.02 of the population proportion?

c. Answer part (b) for a sample of 1600 households.

Problem 4 – Household Grocery Expenditures (Solution)

The Food Marketing Institute shows that 17% of households spend more than \$100 per week on groceries. Assume the population proportion is p=0.17 and a simple random sample of 800 households will be selected from the population.

a. Show the sampling distribution of $\bar{p}$, the sample proportion of households spending more than \$100 per week on groceries.

$$E(\bar{p}) = p = .17$$

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(.17)(1-.17)}{800}} = .0133$$

Distribution is approximately normal because $np = 800(.17) = 136 > 5$

and $n(1-p) = 800(.83) = 664 > 5$

Problem 4 – Household Grocery Expenditures (Solution)

The Food Marketing Institute shows that 17% of households spend more than $100 per week on groceries. Assume the population proportion is p=0.17 and a simple random sample of 800 households will be selected from the population.

b. What is the probability that the sample proportion will be within +-0.02 of the population proportion?

$$z = \frac{.19 - .17}{.0133} = 1.51 \qquad P(z \leq 1.51) = .9345$$

$$P(z < -1.51) = .0655$$

$$P(.15 \leq \bar{p} \leq .19) = P(-1.51 \leq z \leq 1.51) = .9345 - .0655 = .8690$$

Using Excel: NORM.DIST(.19,.17,SQRT(.17*.83/800),TRUE) –
NORM.DIST(.15,.17,SQRT(.17*.83/800),TRUE) = .8679

Problem 4 – Household Grocery Expenditures (Solution)

The Food Marketing Institute shows that 17% of households spend more than \$100 per week on groceries. Assume the population proportion is p=0.17 and a simple random sample of 800 households will be selected from the population.

c. Answer part (b) for a sample of 1600 households.

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(.17)(1-.17)}{1600}} = .0094$$

$$z = \frac{.19-.17}{.0094} = 2.13 \qquad P(z \leq 2.13) = .9834$$

$$P(z < -2.13) = .0166$$

$$P(.15 \leq \bar{p} \leq .19) = P(-2.13 \leq z \leq 2.13) = .9834 - .0166 = .9668$$

Using Excel: NORM.DIST(.19,.17,SQRT(.17*.83/1600),TRUE) –
NORM.DIST(.15,.17,SQRT(.17*.83/1600),TRUE) = .9668

Problem 5 – Black Panther Sales

According to Box Office Mojo, the film Black Panther achieved the top opening weekend gross domestic box office ticket sales in 2018 with $202,033,951. The ticket sales revenue in dollars for a sample of 30 theatres is provided in the file Black Panther.

a. What is the 95% confidence interval estimate for the mean ticket sales revenue per theatre? Interpret this result.

b. Using the movie ticket price of $9.11 per ticket, what is the estimate of the mean number of customers per theatre?

c. The movie was shown in 4020 theatres during its opening weekend. Estimate the total number of customers who saw Black Panther and the total box office ticket sales for the weekend.

Problem 5 – Black Panther Sales (Solution)

a. What is the 95% confidence interval estimate for the mean ticket sales revenue per theatre? Interpret this result.

The Excel output from using the Descriptive Statistics analysis tool with the Black Panther file is shown:

	Revenue ($)
Mean	23100
Standard Error	726.991162
Median	22950
Mode	#N/A
Standard Deviation	3981.89458
Sample Variance	15855484.5
Kurtosis	1.87574748
Skewness	0.47956961
Range	20200
Minimum	13700
Maximum	33900
Sum	693000
Count	30
Confidence Level (95.0%)	1486.86387

$\bar{x} \pm t_{\alpha/2}(s/\sqrt{n})$ with $n = 30$ $df = 29$ $s = 3981.89$ $t = 2.045$

$23100 \pm 2.045\,(3981.89/\sqrt{30})$

23100 ± 14686.86

The sample mean is 23,100 and the margin of error (Confidence Level) is 1486.86.

The 95% confidence interval for the population mean is \$21,613.14 to \$24,586.86. We are 95% confident that the population mean two-day ticket sales revenue per theater is between \$21,613.14 and \$24,586.86.

Problem 5 – Black Panther Sales (Solution)

b. Using the movie ticket price of $9.11 per ticket, what is the estimate of the mean number of customers per theatre?

Mean number of customers per theater = 23,100/8.11 = 2848.34

c. The movie was shown in 4020 theatres during its opening weekend. Estimate the total number of customers who saw Black Panther and the total box office ticket sales for the weekend.

Total number of customers = 4080(2848.34) = 11,621,227 (or 11,621,208 if calculated from the unrounded prior calculations) ≈ $11.6 million customers

Total box office ticket sales for the three-day weekend = 4080(23,100) = $94,248,000 ≈ $94 million

Problem 6 – Time in Supermarket Checkout Lines.

CCN and ActMedia provided a television channel targeted to individuals waiting in supermarket checkout lines. The channel showed news, short features, and advertisements. The length of the program was based on the assumption that the population mean time a shopper stands in a supermarket checkout line is 8 minutes. A sample of actual waiting times will be used to test this assumption and determine whether actual mean waiting time differs from this standard.

a. Formulate the hypotheses for this application.

b. A sample of 120 shoppers showed a sample mean waiting time of 8.4 minutes. Assume a population standard deviation of σ =3.2 minutes. What is the p-value?

c. At a $\alpha = .05$, what is your conclusion?

d. Compute a 95% confidence interval for the population mean. Does it support your conclusion?

Problem 6 – Time in Supermarket Checkout Lines (Solution)

CCN and ActMedia provided a television channel targeted to individuals waiting in supermarket checkout lines. The channel showed news, short features, and advertisements. The length of the program was based on the assumption that the population mean time a shopper stands in a supermarket checkout line is 8 minutes. A sample of actual waiting times will be used to test this assumption and determine whether actual mean waiting time differs from this standard.

a. Formulate the hypotheses for this application.

H_0: $\mu = 8$

H_a: $\mu \neq 8$ Research hypothesis

Problem 6 – Time in Supermarket Checkout Lines.

CCN and ActMedia provided a television channel targeted to individuals waiting in supermarket checkout lines. The channel showed news, short features, and advertisements. The length of the program was based on the assumption that the population mean time a shopper stands in a supermarket checkout line is 8 minutes. A sample of actual waiting times will be used to test this assumption and determine whether actual mean waiting time differs from this standard.

b. A sample of 120 shoppers showed a sample mean waiting time of 8.4 minutes. Assume a population standard deviation of σ =3.2 minutes. What is the p-value?

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{8.4 - 8.0}{3.2 / \sqrt{120}} = 1.37$$

Because $z > 0$, p-value is two times the upper tail area

Using normal table with $z = 1.37$: p-value = 2(1 – .9147) = .1706
Using Excel: p-value = 2*(1 – NORM.S.DIST(1.37,TRUE)) = .1707
Using unrounded test statistic via Excel with cell referencing, p-value = .1709

Problem 6 – Time in Supermarket Checkout Lines.

CCN and ActMedia provided a television channel targeted to individuals waiting in supermarket checkout lines. The channel showed news, short features, and advertisements. The length of the program was based on the assumption that the population mean time a shopper stands in a supermarket checkout line is 8 minutes. A sample of actual waiting times will be used to test this assumption and determine whether actual mean waiting time differs from this standard.

c. At a $\alpha = .05$, what is your conclusion?

p-value > .05; do not reject H_0. Cannot conclude that the population mean waiting time differs from 8 minutes.

Problem 6 – Time in Supermarket Checkout Lines.

CCN and ActMedia provided a television channel targeted to individuals waiting in supermarket checkout lines. The channel showed news, short features, and advertisements. The length of the program was based on the assumption that the population mean time a shopper stands in a supermarket checkout line is 8 minutes. A sample of actual waiting times will be used to test this assumption and determine whether actual mean waiting time differs from this standard.

d. Compute a 95% confidence interval for the population mean. Does it support your conclusion?

$$\bar{x} \pm z_{.025}(\sigma / \sqrt{n})$$

$$8.4 \pm 1.96(3.2/\sqrt{120})$$

$$8.4 \pm .57 \quad (7.83 \text{ to } 8.97)$$

Yes; $\mu = 8$ is in the interval. Do not reject H_0.

RWTH BUSINESS SCHOOL

Mathematics & Statistics
M.Sc. Data Analytics and Decision Science

Prof. Dr. Thomas S. Lontzek

AACSB ACCREDITED

BUSINESS SCHOOL | RWTH AACHEN UNIVERSITY